Computing in High Energy and Nuclear Physics (CHEP) 2012

Monday 21 May 2012 - Friday 25 May 2012

New York City, NY, USA

Book of Abstracts

Contents

Computer Facilities, Production Grids and Networking / 1

Intercontinental Multi-Domain Monitoring for the LHC Optical Private Network

Author: Domenico Vicinanza[1]

[1] *DANTE*

Corresponding Author: domenico.vicinanza@dante.net

The Large Hadron Collider (LHC) is currently running at CERN in Geneva, Switzerland. Physicists are using LHC to recreate the conditions just after the Big Bang, by colliding two beams of particles and heavy ions head-on at very high energy. The project is expected to generate 27 TB of raw data per day, plus 10 TB of "event summary data". This data is sent out from CERN to eleven Tier 1 academic institutions in Europe, Asia, and North America using a multi-gigabits Optical Private Network (OPN), the LHCOPN.

Network monitoring on such complex network architecture to ensure robust and reliable operation is of crucial importance. The chosen approach for monitoring the OPN is based on the perfSONAR MDM framework (http://perfsonar.geant.net), which is designed for multi-domain monitoring environments.

perfSONAR (www.perfsonar.net) is an infrastructure for performance monitoring data exchange between networks, making it easier to solve performance problems occurring between network measurement points interconnected through several network domains. It contains a set of services delivering performance measurements in a multi-domain environment. These services act as an intermediate layer, between the performance measurement tools and the visualization applications. This layer is aimed at exchanging performance measurements between networks, using well defined protocols. perfSONAR is web-service based, modular, and it uses NM-WG OGF standards.

perfSONAR MDM is the perfSONAR version built by GÉANT (www.geant.net), the consortium operating the European Backbone for research and education.

Given the quite particular structure of the LHCOPN, a specially customised version of the perfSONAR MDM was prepared by an international consortium for the specific monitoring of IP and circuits of the LHC Optical Private Network.

The proposed presentation will introduce the main points of the LHCOPN structure, provide an introduction about perfSONAR framework (software, architecture, service structure) and finally describe the way the whole monitoring infrastructure is monitored and how the support is organised.

Summary:

The presentation submitted will be organised as follows:
- Introduction to the LHCOPN: structure, motivation, challenges
- The perfSONAR Multi-Domain Monitoring framework: software, architecture, service structure
- Monitoring the monitoring infrastructure (including support)

Poster Session / 2

The Pandora Software Development Kit for Particle Flow Calorimetry

Author: John Marshall[1]

Co-author: Mark Thomson [1]

[1] *University of Cambridge (GB)*

Corresponding Author: marshall@hep.phy.cam.ac.uk

Pandora is a robust and efficient framework for developing and running pattern-recognition algorithms. It was designed to perform particle flow calorimetry, which requires many complex pattern-recognition techniques to reconstruct the paths of individual particles through fine granularity detectors. The Pandora C++ software development kit (SDK) consists of a single library and a number of carefully designed application programming interfaces (APIs). A client application can use the Pandora APIs to pass details of tracks and hits/cells to the Pandora framework, which then creates and manages named lists of self-describing objects. These objects can be accessed by Pandora algorithms, which perform the pattern-recognition reconstruction. Development with the Pandora SDK promotes the creation of small, re-usable algorithms containing just the kernel of a specific operation. The algorithms are configured via XML and can be nested to perform complex reconstruction tasks. As the algorithms only access the Pandora objects in a controlled manner, via the APIs, the framework can perform most book-keeping and memory-management operations. The Pandora SDK has been fully exploited in the implementation of PandoraPFA, which uses over 60 algorithms to provide the state of the art in particle flow calorimetry for ILC and CLIC.

Poster Session / 3

Bolting the Door

Author: Mark Mitchell[1]

Co-author: David Crooks [1]

[1] *University of Glasgow/GridPP*

Corresponding Author: mark.mitchell@glasgow.ac.uk

This presentation will cover the work conducted within the ScotGrid Glasgow Tier-2 site. It will focus on the multi-tiered network security architecture developed on the site to augment Grid site server security and will discuss the variety of techniques used including the utilisation of Intrusion Detection systems, logging and optimising network connectivity within the infrastructure.

Also the analysis of the limitations of this approach and the potential for future research in this area will be investigated and discussed

Poster Session / 4

Engaging with IPv6: addresses for all

Author: Mark Mitchell[1]

Co-author: David Crooks [2]

[1] *University of Glasgow*

[2] *University of Glasgow/GridPP*

Corresponding Author: mark.mitchell@glasgow.ac.uk

Due to the changes occurring within the IPv4 address space, the utilisation of IPv6 within Grid Technologies and other IT infrastructure is becoming a more pressing solution for IP addressing. The employment and deployment of this addressing scheme has been discussed widely both at the academic and commercial level for several years. The uptake is not as advanced as was predicted and the potential of this technology hasn't been fully utilised. Presently, an investigation into this technology is underway as it may offer solutions to the future of IP addressing for collaborative environments.

As part of the HEPIX IPv6 Working Group we investigate the test deployments of IPv6 at the University of Glasgow Tier-2 within Scot Grid and report on both the enablement of Grid services within this framework and also possible configuration solutions for Tier-2 network environments housed within University networks.

Drawing upon various test scenarios enabled within the University of Glasgow, areas such as DNS, Monitoring and security mechanisms will also be touched upon.

Poster Session / 5

Calibration and performance monitoring of the LHCb Vertex Locator

Author: Tomasz Szumlak[1]

Co-author: Karol Hennessy [2]

[1] *AGH Univ. of Science & Technology (PL)*

[2] *Liverpool*

Corresponding Authors: karol.hennessy@cern.ch, szumlak@agh.edu.pl

The LHCb experiment is dedicated to searching for New Physics effects in the heavy flavour sector, precise measurements of CP violation and rare heavy meson decays. Precise tracking and vertexing around the interaction point is crucial in achieving these physics goals.
The LHCb VELO (VErtex LOcator) silicon micro-strip detector is the highest precision vertex detector at the LHC and is located at only 8 mm from the proton beams. The high spatial resolution (up to 4 microns single hit precision) is obtained by a complex chain of processing algorithms to suppress noise and reconstruct clusters. These are implemented in large FPGAs, with over one million parameters that need to be individually optimised. Previously we presented a novel approach that has been developed to optimise the parameters and integrating their determination into the full software framework of the LHCb experiment. Presently we report on the experience gained from regular operation of the calibration and monitoring software with the collision data taken in 2011 by the LHCb experiment. Both the VELO performance and its impact on the physics results will be detailed.

Poster Session / 6

Optimization of HEP Analysis activities using a Tier2 Infrastructure

Authors: Giuseppe Bagliesi[1]; Tommaso Boccali[1]

[1] *INFN Sezione di Pisa*

Corresponding Authors: giuseppe.bagliesi@cern.ch, tommaso.boccali@cern.ch

While the model for a Tier2 is well understood and implemented within the HEP Community, a refined design for Analysis specific sites has not been agreed upon as clearly. We aim to describe the solutions adopted at the INFN Pisa, the biggest Tier2 in the Italian HEP Community. A Standard Tier2 infrastructure is optimized for GRID CPU and Storage access, while a more interactive oriented use of the resources is beneficial to the final data analysis step. In this step, POSIX file storage access is easier for the average physicist, and has to be provided in a real or emulated way. Modern analysis techniques use advanced statistical tools (like RooFit and RooStat), which can make use of multi core

systems. The infrastructure has to provide or create on demand computing nodes with many cores available, above the existing and less elastic Tier2 flat CPU infrastructure. At last, the users do not want to have to deal with data placement policies at the various sites, and hence a transparent WAN file access, again with a POSIX layer, must be provided, making use of the just-installed 10 GBit/s regional lines.
Even if standalone systems with such features are possible and exist, the implementation of an Analysis site as a virtual layer over an existing Tier2 requires novel solutions; the ones used in Pisa are described here.

Poster Session / **7**

Investigation of Storage Systems for use in Grid Applications

Authors: Gabriele Garzoglio[1]; Ted Hesselroth[1]

[1] *Fermi National Accelerator Laboratory*

In recent years, several new storage technologies, such as Lustre, Hadoop, OrangeFS, and BlueArc, have emerged. While several groups have run benchmarks to characterize them under a variety of configurations, more work is needed to evaluate these technologies for the use cases of scientific computing on Grid clusters and Cloud facilities. This paper discusses our evaluation of the technologies as deployed on a test bed at FermiCloud, one of the Fermilab infrastructure-as-a-service Cloud facilities. The test bed consists of 4 server-class nodes with 40 TB of disk space and up to 50 virtual machine clients, some running on the storage server nodes themselves. With this configuration, the evaluation compares the performance of some of these technologies when deployed on virtual machines and on "bare metal" nodes. In addition to running standard benchmarks such as IOZone to check the sanity of our installation, we have run I/O intensive tests using physics-analysis applications. This paper presents how the storage solutions perform in a variety of realistic use cases of scientific computing. One interesting difference among the storage systems tested is found in a decrease in total read throughput with increasing number of client processes, which occurs in some implementations but not others.

Poster Session / **8**

Design and implementation of a reliable and cost-effective cloud computing infrastructure: the INFN Naples experience

Authors: Francesco Taurino[1]; Gennaro Tortone[2]; Rosario Esposito[2]; Silvio Pardi[2]; Vincenzo Capone[3]

[1] *CNR SPIN (IT)*

[2] *INFN (IT)*

[3] *Universita e INFN (IT)*

Over the last few years we have seen an increasing number of services and applications needed to manage and maintain cloud computing facilities. This is particularly true for computing in high energy physics which often requires complex configurations and distributed infrastructures. In this scenario a cost effective rationalization and consolidation strategy is the key to success in terms of scalability and reliability.
In this work, we describe an IaaS (Infrastructure as a Service) cloud computing system, with high availability and redundancy features which is currently in production at INFN-Naples and ATLAS Tier-2 data centre.
The main goal we intended to achieve was a simplified method to manage our computing resources and deliver reliable user services, reusing existing hardware without incurring heavy costs.
A combined usage of virtualization and clustering technologies allowed us to consolidate our services on a small number of physical machines, reducing electric power costs.
As a results of our efforts we developed a complete solution for data and computing centers that can

be easily replicated using commodity hardware.
Our architecture mainly consists of 2 subsystems: a clustered storage solution, built on top of disk servers running Gluster file system, and a virtual machines execution environment. The hypervisor hosts use Scientific Linux and KVM as virtualization technology and run both Windows and Linux guests. Virtual machines have their root file systems on qcow2 disk-image files, stored on a Gluster network file system. Gluster is able to perform parallel writes on multiple disk servers (two, in our system), providing this way live replication of data. A failure of a disk server doesn't cause glitches or stops any of the running virtual guests as each hypervisor host still has full access to disk-image files. When the failing disk server returns to normal activity Gluster self-healing integrated mechanism performs a background transparent reconstruction of missing replicas.
High availability is also achieved via a network configuration using redundant switches and multiple paths between hypervisor hosts and disk servers. Linux channel bonding provides adaptive load balancing of network traffic over multiple links and dedicated VLANs guarantee isolation of the storage subsystem from the general-purpose network.
We also developed a set of management scripts to easily perform basic system administration tasks such as automatic deployment of new virtual machines, adaptive scheduling of virtual machines on hypervisor hosts, live migration and automated restart in case of hypervisor failures.

Summary:

The work is organized as follows:
In the first part we identify the main requirements and the goal we want to achieve in terms of system reliability and availability. Then we introduce a set of currently available open-source technologies and we discuss the motivation of our choice. After that, we describe our cloud computing model: the architecture, all the features and the main aspects.

In the second part we show our implementation at INFN-Naples describing the hardware, the network topology, the storage configuration and the migration process of our services from physical machines to cloud infrastructure. The description is accompanied by some stress test benchmark results and a technical analysis of the system utilization during the last year.
In the last part we illustrate other possible application scenarios with a set of recommendations based on our local experience.

Poster Session / **9**

glideinWMS experience with glexec

Authors: Burt Holzman[1]; Claudio Grandi[2]; Dan Bradley[3]; Frank Würthwein[4]; Igor Sfiligoi[5]; Igor Sfiligoi[4]; Igor Sfiligoi[6]; Jeffrey Michael Dost[4]; Kenneth Bloom[7]; Zachary Miller[3]

[1] *Fermi National Accelerator Laboratory*

[2] *INFN Bologna*

[3] *University of Wisconsin-Madison*

[4] *University of California San Diego*

[5] *INFN LABORATORI NAZIONALI DI FRASCATI*

[6] *Univ. of California San Diego (US)*

[7] *University of Nebraska-Lincoln*

Corresponding Authors: sfiligoi@lnf.infn.it, isfiligoi@ucsd.edu, sfiligoi@fnal.gov

Multi-user pilot infrastructures provide significant advantages for the communities using them, but also create new security challenges.
With Grid authorization and mapping happening with the pilot credential only, final user identity is not properly addressed in the classic Grid paradigm.
In order to solve this problem, OSG and EGI have deployed glexec, a privileged executable on the worker nodes that allows for final user authorization and mapping from inside the pilot itself.
The glideinWMS instances deployed on OSG have been now using glexec on OSG sites for several years, and have started using it on EGI resources in the past year.

The user experience of using glexec has been mostly positive, although there are still some edge cases where things could be improved.
This talk provides both the usage statistics as well as a description of the still remaining problems and the expected solutions.

Poster Session / **10**

Preparing the ALICE DAQ upgrade

Author: Pierre Vande Vyvre[1]

Co-authors: Adriana Telesca [1]; Alexandru Grigore [2]; Barthelemy von Haller [1]; Bartolomeu Andre Rodrigues Fernandes Rabacal [3]; Csaba Soos [1]; Ervin Denes [4]; Filippo Costa [1]; Franco Carena [1]; Giuseppe Simonetti [5]; Roberto Divia [1]; Sylvain Chapeland [1]; Ulrich Fuchs [1]; Vasco Chibante Barroso [1]; Wisla Carena [1]

[1] *CERN*

[2] *Polytechnic University of Bucharest*

[3] *Instituto Superior Tecnico (IST)*

[4] *Hungarian Academy of Sciences (HU)*

[5] *Universita e INFN*

Corresponding Author: pierre.vande.vyvre@cern.ch

In November 2009, after 15 years of design and installation, the ALICE experiment started to detect and record the first collisions produced by the LHC. It has been collecting hundreds of millions of events ever since with both proton-proton and heavy ion collision. The future scientific programme of ALICE has been refined following the first year of data taking. The physics targeted beyond 2016 will be the study of rare signals. Several detectors will be upgraded, modified, or replaced to prepare ALICE for future physics challenges. An upgrade of the triggering and readout system is also required to accommodate the needs of the upgraded ALICE and to better select the data of the rare physics channels. The ALICE upgrade will have major implications in the detector electronics and controls, data acquisition, event triggering, offline computing and storage systems. Moreover, the experience accumulated during more than two years of operation has also lead to new requirements for the control software. We will review all these new needs and the current R&D activities to address them.
Several papers of the same conference present in more details some elements of the ALICE DAQ system.

Summary:

A review of the ALICE DAQ R&D activities in view of addressing the future scientific programme of ALICE following the first year of data taking.

Poster Session / **11**

Improvements in ROOT I/O

Author: Philippe Canal[1]

[1] *FERMILAB*

Corresponding Author: philippe.canal@cern.ch

In the past year, the development of ROOT I/O has focused on improving the existing code and increasing the collaboration with the experiments' experts. Regular I/O workshops have been held

to share and build upon the varied experiences and points of view. The resulting improvements in ROOT I/O span many dimensions including reduction and more control over the memory usage, drastic reduction in CPU usage as well as optimization of the file size and the hardware I/O utilization.

Poster Session - Board: 1 / **12**

Virtualizing A Large Cluster at Brookhaven

Authors: Christopher Hollowell[1]; Costin Caramarcu[2]; Tony Wong[1]; William Strecker-Kellogg[3]

[1] *Brookhaven National Laboratory*

[2] *Horia Hulubei National Institute of Physics and Nuclear Enginee*

[3] *Brookhaven National Lab*

Corresponding Author: willsk@bnl.gov

In this presentation we will address the development of a prototype virtualized worker node cluster, using Scientific Linux 6.x as a base
OS, KVM for virtualization, and the Condor batch software to manage virtual machines. The discussion provides details on our experiences
with building, configuring, and deploying the various components from bare metal, including the base OS, the virtualized OS images and the
integration of batch services with the virtual machines.

We also discuss benefits and drawbacks of widespread deployment of virtualized clusters in support of private clouds in a distributed
computing environment. We show that under certain computing models the visualization of worker nodes is of limited value.

Summary:

Worker node virtualization using Condor as a VM manager.

Poster Session / **13**

Triggering on hadronic tau decays in ATLAS: algorithms and performance

Author: Cristobal Cuenca Almenar[1]

Co-author: Patrick Czodrowski [2]

[1] *Yale University (US)*

[2] *Technische Universitaet Dresden (DE)*

Corresponding Authors: patrick.czodrowski@cern.ch, cristobal.cuenca@cern.ch

Hadronic tau decays play a crucial role in taking Standard Model measurements as well as in the search for physics beyond the Standard Model. However, hadronic tau decays are difficult to identify and trigger on due to their resemblance to QCD jets. Given the large production cross section of QCD processes, designing and operating a trigger system with the capability to efficiently select hadronic tau decays, while maintaining the rate within the bandwidth limits, is a difficult challenge.

This contribution will summarize the algorithms and performance of the ATLAS tau trigger system during the 2011 data taking period. The use of resources and implementation of trigger algorithms in the ATLAS trigger architecture will be shown in detail. Moreover, comparisons of data and simulation results, studies of the correlation of the variable definitions at different trigger stages as well as efficiency versus rate analyses are the key elements to describe the performance of the tau trigger Finally, in light of the vast statistics collected in 2011, future prospects for triggering on hadronic tau decays in this exciting new period of increased instantaneous luminosity will be presented.

Summary:

Hadronic tau decays play a crucial role in taking Standard Model measurements as well as in the search for physics beyond the Standard Model. However, hadronic tau decays are difficult to identify and trigger on due to their resemblance to QCD jets. Given the large production cross section of QCD processes, designing and operating a trigger system with the capability to efficiently select hadronic tau decays, while maintaining the rate within the bandwidth limits, is a difficult challenge.

This contribution will summarize the algorithms and performance of the ATLAS tau trigger system during the 2011 data taking period. The use of resources and implementation of trigger algorithms in the ATLAS trigger architecture will be shown in detail. Moreover, comparisons of data and simulation results, studies of the correlation of the variable definitions at different trigger stages as well as efficiency versus rate analyses are the key elements to describe the performance of the tau trigger Finally, in light of the vast statistics collected in 2011, future prospects for triggering on hadronic tau decays in this exciting new period of increased instantaneous luminosity will be presented.

Poster Session / **14**

b-jet triggering in ATLAS: from algorithm implementation to physics analyses

Authors: Andrea Coccaro[1]; Per Ola Hansson[2]

Co-author: Alexander Oh [3]

[1] *Universita e INFN (IT)*

[2] *SLAC National Accelerator Laboratory (US)*

[3] *University of Manchester (GB)*

Corresponding Authors: alexander.oh@cern.ch, andrea.coccaro@cern.ch

The online event selection is crucial to reject most of the events containing uninteresting background collisions while preserving as much as possible the interesting physical signals. The b-jet selection is part of the trigger strategy of the ATLAS experiment and a set of dedicated triggers is in place from the beginning of the 2011 data-taking period and is contributing to keep the total bandwidth to an affordable rate. The b-jets acceptance is increased and the background reduced by lowering jet transverse energy thresholds at the first trigger level and applying b-tagging techniques at the subsequent levels. Different physics channels, especially topologies containing more than one b-jet where higher rejection factors are achieved, benefit from requesting this trigger to be fired. An overview of the status-of-art of the b-jet trigger menu and the performance on real data is presented in this contribution. Data-driven techniques to extract the online b-tagging efficiency and mis-tag rate, key ingredients for all analyses relying on such triggers, are also discussed and results presented.

Poster Session / **16**

Scientific Cluster Deployment & Recovery: Using puppet to simplify cluster management

Author: Valerie Hendrix[1]

Co-authors: Doug Benjamin [2]; Yushu Yao [3]

[1] *Lawrence Berkeley National Lab. (US)*

[2] *Duke University (US)*

[3] *LBNL*

Corresponding Author: vchendrix@lbl.gov

Deployment, maintenance and recovery of a scientific cluster, which has complex, specialized services, can be a time consuming task requiring the assistance of Linux system administrators, network engineers as well as domain experts. Universities and small institutions that have a part-time FTE with limited knowledge of the administration of such clusters can be strained by such maintenance tasks.

This current work is the result of an effort to maintain a data analysis cluster with minimal effort by a local system administrator. The realized benefit is the scientist, who is the local system administrator, is able to focus on the data analysis instead of the intricacies of managing a cluster. Our work provides a cluster deployment and recovery process based on the puppet configuration engine allowing a part-time FTE to easily deploy and recover entire clusters with minimal effort.

Puppet is a configuration management system (CMS) used widely in computing centers for the automatic management of resources. Domain experts use Puppet's declarative language to define reusable modules for service configuration and deployment.

Our deployment process has three actors: domain experts, a cluster designer and a cluster manager. The domain experts first write the puppet modules for the cluster services. A cluster designer would then define a cluster. This includes the creation of cluster roles, mapping the services to those roles and determining the relationships between the services. Finally, a cluster manager would acquire the resources (machines, networking), enter the cluster input parameters (hostnames, IP addresses) and automatically generate deployment scripts used by puppet to configure it to act as a designated role. In the event of a machine failure, the originally generated deployment scripts along with puppet can be used to easily reconfigure a new machine.

The cluster definition produced in our cluster deployment process is an integral part of automating cluster deployment in a cloud environment. Our future cloud efforts will further build on this work.

Poster Session / **17**

Experience of BESIII data production with local cluster and distributed computing model

Author: ziyan Deng[1]

Co-authors: huaimin Liu [1]; weidong Li [1]; yongzhao Sun [1]

[1] *Institute of High Energy Physics, Beijing, China*

Corresponding Author: dengzy@ihep.ac.cn

The BES III detector is a new spectrometer which works on the upgraded high-luminosity collider, the Beijing Electron-Positron Collider (BEPCII). The BES III experiment studies physics in the tau-charm energy region from 2GeV to 4.6GeV . Since spring 2009, BEPCII has produced large scale data samples. All the data samples were processed successfully and many important physics results have been achieved based on these samples. Doing data production correctly and efficiently with limited CPU and storage resources is a big challenge. This paper will describe the implementation of the experiment-specific data production for BESIII in detail, including data calibration with event-level parallel computing model, data reconstruction, inclusive Monte Carlo generation, random trigger background mixing and multi-stream data skimming. Now, with the data sample increasing rapidly,

there is a growing demand to move from solely using a local cluster to a more distributed computing model. A distributed computing environment is being set up and expected to go into production use in 2012. The experience of BESIII data production, both with a local cluster and with a distributed computing model, is presented here.

Summary:

Large scale data samples from BESIII have been successfully processed. And more data is accumulated. The experience of BESIII data production, both with a local cluster and with a distributed computing model is presented.

Poster Session / **18**

A scalable low-cost Petabyte scale storage for HEP using Lustre

Author: Christopher John Walker[1]

Co-author: Alex Martin [2]

[1] *University of London (GB)*

[2] *QUEEN MARY, UNIVERSITY OF LONDON*

Corresponding Authors: christopher.john.walker@cern.ch, a.j.martin@qmul.ac.uk

We describe a low-cost Petabyte scale Lustre filesystem deployed for High Energy Physics. The use of commodity storage arrays and bonded ethernet interconnects makes the array cost effective, whilst providing high bandwidth to the storage. The filesystem is a POSIX filesytem, presented to the Grid using the StoRM SRM. The system is highly modular. The building blocks
of the array, the Lustre Object Storage Servers (OSS) each have 12*2TB SATA disks configured as a RAID6 array, delivering 18TB of storage. The network bandwidth from the storage servers is designed to match that from the compute
servers within each module of 6 storage servers and 12 compute servers. The modules are connect together by a 10Gbit core network to provide balanced overall performance. We present benchmarks demonstrating the performance and scalability of the filesystem.

Software Engineering, Data Stores and Databases / **19**

GOoDA: The Generic Optimization Data Analyzer

Authors: David Levinthal[1]; Paolo Calafiura[2]; Roberto Agostino Vitillo[3]; Stephane Eranian[1]

[1] *Google*

[2] *Lawrence Berkeley National Lab. (US)*

[3] *LBNL*

Modern superscalar, out-of-order microprocessors dominate large scale server computing. Monitoring their activity, during program execution, has become complicated due to the complexity of the microarchitectures and their IO interactions. Recent processors have thousands of performance monitoring events. These are required to actually provide coverage for all of the complex interactions and performance issues that can occur. Knowing which data to collect and how to interpret the results has become an unreasonable burden for code developers whose tasks are already hard enough. It becomes the task of the analysis tool developer to bridge this gap.
To address this issue, a generic decomposition of how a microprocessor is using the consumed cycles

allows code developers to quickly understand which of the myriad of microarchitectural complexities they are battling, without requiring a detailed knowledge of the microarchitecture. When this approach is intrinsically integrated into a performance data analysis tool, it enables software developers to take advantage of the microarchitectural methodology that has only been available to experts.
The Generic Optimization Data Analyzer (GOoDA) project integrates this expertise into a profiling tool in order to lower the required expertise of the user and, being designed from the ground up with large-scale object-oriented applications in mind, it will be particularly useful for large HENP codebases

Distributed Processing and Analysis on Grids and Clouds / 20

Computing at Belle II

Authors: Takanori HARA[1]; Thomas Kuhr[2]

[1] *KEK*

[2] *KIT - Karlsruhe Institute of Technology (DE)*

Corresponding Author: thomas.kuhr@cern.ch

The Belle II experiment, a next-generation B factory experiment at KEK, is expected to record a two orders of magnitude larger data volume than its predecessor, the Belle experiment. The data size and rate are comparable to or more than the ones of LHC experiments and requires to change the computing model from the Belle way, where basically all computing resources were provided by KEK, to a more distributed scheme. The Belle II distributed computing system is based on DIRAC which provides an interface to grid and cloud resources, and AMGA for the management of file metadata. A common software framework is used in the whole chain from the data acquisition up to the analysis. It has a modular design, is steered via python files, and supports parallel execution on multi-core nodes.
In this talk the status and plans of the Belle II computing system and its main components are presented.

Poster Session / 21

Computing On Demand: Dynamic Analysis Model

Author: Anar Manafov[1]

Co-author: Peter Malzacher [1]

[1] *GSI - Helmholtzzentrum fur Schwerionenforschung GmbH (DE)*

Corresponding Author: a.manafov@gsi.de

Constant changes in computational infrastructure like the current interest in Clouds, imply conditions on the design of applications. We must make sure that our analysis infrastructure, including source code and supporting tools, is ready for the on demand computing (ODC) era.

This presentation is about a new analysis concept, which is driven by users needs, completely disentangled from the computational resources, and scalable.

What does it take for an analysis code to be performed on any resource management system? How can one achieve goals of on demand analysis, using PROOF on Demand (PoD)? These questions and such topics as preferable location of data files as well as tools and software development techniques for on demand data analysis are covered. Also analysis implementation requirements and comparisons of traditional and “on demand”facilities will be discussed during this talk.

Distributed Processing and Analysis on Grids and Clouds / 22

PoD: dynamically create and use remote PROOF clusters. A thin client concept.

Author: Anar Manafov[1]

Co-author: Peter Malzacher [1]

[1] *GSI - Helmholtzzentrum fur Schwerionenforschung GmbH (DE)*

Corresponding Author: a.manafov@gsi.de

PROOF on Demand (PoD) is a tool-set, which dynamically sets up a PROOF cluster at a user's request on any resource management system (RMS). It provides a plug-in based system, in order to use different job submission front-ends.
PoD is currently shipped with gLite, LSF, PBS (PBSPro/OpenPBS/Torque), Grid Engine (OGE/SGE), Condor, LoadLeveler, and SSH plug-ins. It makes it possible just within a few seconds to get a private PROOF cluster on any RMS. If there is no RMS, then SSH plug-in can be used, which dynamically turns a bunch of machines to PROOF workers.

In this presentation new developments and use cases will be covered.

Recently a new major step in PoD development has been made. It can now work not only with local PoD servers, but also with remote ones.
PoD's newly developed "pod-remote"command made it possible for users to utilize a thin client concept. In order to create dynamic PROOF clusters, users are now able to select a remote computer, even behind a firewall, to control a PoD server on it and to submit PoD jobs. In this case a user interface machine is just a lightweight control center and could run on different OS types or mobile devices.
All communications are secured and provided via SSH channels. Additionally PoD automatically creates and maintains SSH tunnels for PROOF connections between a user interface and PROOF muster.

PoD will create and manage remote and local PROOF clusters for you. Just two commands of PoD will provide you with the full functional PROOF cluster and a real computing on demand.

The talk will also include several live demos from real life use cases.

Computer Facilities, Production Grids and Networking / 23

A strategy for load balancing in distributed storage systems

Author: Gerd Behrmann[1]

Co-author: Erik Mattias Wadenstein [2]

[1] *NDGF*

[2] *Unknown*

Corresponding Authors: mattias.wadenstein@cern.ch, behrmann@ndgf.org

Distributed storage systems are critical to the operation of the WLCG. These systems are not limited to fulfilling the long term storage requirements. They also serve data for computational analysis and other computational jobs. Distributed storage systems provide the ability to aggregate the storage and IO capacity of disks and tapes, but at the end of the day IO rate is still bound by the capabilities of the hardware, in particular the hard drives. Throughput of hard drives has increased dramatically over the decades, however for computational analysis IOPS is typically the limiting factor. To maximize return of investment, balancing IO load over available hardware is crucial. The task is made

complicated by the common use of heterogeneous hardware and software environments that results from combining new and old hardware into a single storage system.

This paper describes recent advances made in load balancing in the dCache distributed storage system. We describe a set of common requirements for load balancing policies. These requirements include considerations about temporal clustering, resistance to disk pool divisioning, evolution of the hardware portfolio as data centers get extended and upgraded, the non-linearity of disk performance, garbage collection, age of replicas, stability of control decisions, and stability of tuning parameters. We argue that the existing load balancing policy in dCache fails to satisfy most of these requirements.

An alternative policy is proposed, the weighted available space selection policy. The policy incorporates ideas we have been working on for years while observing dCache in production at NDGF and at other sites. At its core the policy uses weighted random selection, but it incorporates many different signals into the weight. We argue that although the algorithm is technically more complicated, it is in our experience easier to predict the effect of parameter changes and thus the parameters are easier to tune than in the previous policy. It many cases it may not even require manual tuning, although we need more empirical data to conclude that.

The new policy has been integrated into dCache 2.0 using a new load balancing plugin system developed by NDGF. It has been used in production at NDGF since end of August 2011. We will report on our experiences. A qualitative and quantitative analysis of the policy will be presented augmented by simulations and empirical data.

Although our algorithm has been developed and is used in the context of dCache, the ideas are universal and could be applied to many storage systems.

Computer Facilities, Production Grids and Networking / 24

Analysing I/O bottlenecks in LHC data analysis on grid storage resources

Author: Wahid Bhimji[1]

Co-authors: Ilija Vukotic [2]; Martin Philipp Hellmich [3]; Matthew Doidge [4]; Philip Clark [1]

[1] *University of Edinburgh (GB)*

[2] *Universite de Paris-Sud 11 (FR)*

[3] *University of Edinburgh*

[4] *Lancaster University*

Corresponding Author: wahid.bhimji@cern.ch

We describe recent I/O testing frameworks that we have developed and applied within the UK GridPP Collaboration, the ATLAS experiment and the DPM team, for a variety of distinct purposes. These include benchmarking vendor supplied storage products, discovering scaling limits of SRM solutions, tuning of storage systems for experiment data analysis, evaluating file access protocols, and exploring IO read patterns of experiment software and their underlying event data models. With multiple grid sites now dealing with petabytes of data, such studies are becoming increasingly essential. We describe how the tests build, and improve, on previous work and contrast how the use-cases differ. We also detail the results obtained and the implications for storage hardware, middleware and experiment software.

Poster Session / 25

File and Metadata Management for BESIII Distributed Computing

Authors: Caitriana Nicholson[1]; Lei Lin[2]; Weidong Li[3]; ziyan deng[4]

Co-author: Yangheng Zheng [5]

[1] *Graduate University of the Chinese Academy of Sciences*

[2] *Suzhou University*

[3] *Weidong Li*

[4] *Institute of High Energy Physics, Beijing*

[5] *GUCAS*

Corresponding Author: caitriana@gucas.ac.cn

The BES III experiment at the Institute of High Energy Physics (IHEP), Beijing, uses the high-luminosity BEPC II e+e- collider to study physics in the τ-charm energy region around 3.7 GeV; BEPC II has produced the world's largest samples of J/ψ and ψ'events to date. An order of magnitude increase in the data sample size over the 2011-2012 data-taking period demanded a move from a very centralized to a distributed computing environment, as well as the development of an efficient file and metadata management system. While BES III is on a smaller scale than some other HEP experiments, this poses particular challenges for its distributed computing and data management system. These constraints include limited resources and manpower, and low quality of network connections to IHEP. Drawing on the rich experience of the HEP community, an AMGA-based system has been developed which meets these constraints. The design and development of the BES III distributed data management system, including its integration with other BES III distributed computing components, such as job management, are presented here.

Poster Session / **26**

Clustering induced Pattern Recognition in a TPC for the Linear Collider

Author: Frank-Dieter Gaede[1]

[1] *Deutsches Elektronen-Synchrotron (DE)*

Corresponding Author: frank-dieter.gaede@cern.ch

ILD is a proposed detector concept for a future linear collider, that envisages a Time Projection Chamber (TPC) as the central tracking detector. The ILD TPC will have a large number of voxels that have dimensions that are small compared to the typical distances between charged particle tracks. This allows for the application of simple nearest neighbor type clustering algorithms to find clean track segments.
Clupatra is a TPC pattern recognition algorithm that uses such clustering methods to find track seeds and then a Kalman Filter to extend these segments to form complete tracks. We present the algorithm and it's performance and track finding efficiency for the ILC, including machine induced backgrounds, as well as for the case of CLIC with much more challenging occupancies that are comparable to those of the ALICE TPC.
Clupatra is written in the iLCSoft framework based on LCIO and Marlin and will be used for the massive Monte Carlo production for the Conceptual Design Report of ILD in 2012.

Poster Session / **27**

Implementing Parallel Algorithms

Author: Julius Hrivnac[1]

[1] *Universite de Paris-Sud 11 (FR)*

Corresponding Author: julius.hrivnac@cern.ch

The possible implementation of parallel algorithms will be described.
- The functionality will be demonstrated using Swarm - a new experimental interactive parallel framework.
- The access from several parallel-friendly scripting languages will be shown.
- The benchmarks of the typical tasks used in High Energy Physics code will be provided.

The talk will concentrate on using the "Fork and Join" approach, which is the default solution included in the Java 7 environment. The comparison of that approach with other alternatives will be given too.

Poster Session / **28**

FAZIA DATA ACQUISITION: STATUS, DESIGN AND CONCEPT

Author: Gennaro Tortone[1]

Co-authors: Alfonso Boiano [1]; Antonio Ordine [1]; Elio Rosato [2]; Giulio Spadaccini [2]; Mariano Vigiliante [2]

[1] *INFN Napoli*

[2] *Dip. Scienze Fisiche Federico II*

The FAZIA project groups together several institutions in Nuclear Physics, which are working in the domain of heavy-ion induced reactions around and below the Fermi energy. The aim of the project is to build a 4Pi array for charged particles, with high granularity and good energy resolution, with A and Z identification capability over the widest possible range.

It will use the up-to-date techniques concerning detection, signal processing and data flow, with full digital electronics. The FAZIA data acquisition system introduces various issues about high data flow bandwith (˜600 MB/s) and design of nested data event format (up to five level).

In this poster DAQ design and architecture will be described focusing on event data model, software trigger and NARVAL, a novel event transport framework. Overall benchmarks and first results will be also discussed.

Poster Session / **30**

Multi-platform masterclass and data analysis application

Authors: Joao Antunes Pequenao[1]; Neng Xu[2]

Co-author: Gerardo Ganis [3]

[1] *Lawrence Berkeley National Lab. (US)*

[2] *University of Wisconsin (US)*

[3] *CERN*

Corresponding Author: joao.pequenao@cern.ch

New types of hardware, like smartphones and tablets, are becoming more available, affordable and popular in the market. Furthermore with the advent of Web2.0 frameworks, Web3D and Cloud computing, the way we interact, produce and exchange content is being dramatically transformed.

How can we take advantage of these technologies to produce engaging applications which can be conveniently used both by physicists and the general public?

We will demonstrate the development of a platform independent application for data analysis and educational scenarios. This application should enable educators to conduct a novel type of masterclasses, as well as facilitate the collaboration between physicists due to its inherent simplicity, lightness and aesthetic appeal. Users will be able to run it on different hardware such as laptops, smart phones or tablets, and have access to the data everywhere.
The application can also run within a web browser.
Based on one of the most popular graphic engines, people can view 2D histograms, animated 3D event displays and do event analysis. The heavy processing jobs will be sent to the Cloud via a master server, in such a way that people can run multiple complex jobs simultaneously.
All of this can be automated and shared with the community trough XML files describing a succession of actions.

After having introduced the new system structure and the way the new application will fit in the overall picture, we will describe the current progress of the development and the test facility and discuss further technical difficulties that we expect to be confronted to, like the security (user authentication and authorization) data discovery and load balancing.

Poster Session / **31**

Offline software for the Resistive Plate Chambers in the Daya Bay Antineutrino Experiment

Author: Miao HE[1]

[1] *Institute of High Energy Physics, Chinese Academy of Sciences*

Corresponding Author: hem@ihep.ac.cn

Neutrino flavor oscillation is characterized by three mixing angles. The Daya Bay reactor antineutrino experiment is designed to determine the last unknown mixing angle θ_{13}. The experiment is located in southern China, near the Daya Bay nuclear power plant. Eight identical liquid scintillator detectors are being installed in three experimental halls, to detect antineutrinos released in nuclear fission. The Water Cherenkov detector and the Resistive Plate Chambers (RPC) deployed in each experimental hall form a muon system to veto cosmic muons which are the main source of backgrounds. The combined muon veto efficiency is designed as 99.5% with 0.25% uncertainties. Offline software for the Daya Bay experiment is being developed in the framework of Gaudi. In this presentation, we will give a brief introduction to the Daya Bay experiment and the software framework. Then we will focus on the simulation, calibration and reconstruction of the Resistive Plate Chambers.

Poster Session / **32**

Simultaneous Operation and Control of about 100 Telescopes for the Cherenkov Telescope Array

Author: Peter Wegner[1]

Co-authors: A. Lopatin [2]; C. Stegmann [2]; D. Hoffmann [3]; D. Melkumyan [1]; E. Lyard [4]; G. Lamanna [5]; H. Koeppel [1]; I. Oya [6]; J. Colomé [7]; J.-L. Panazol [5]; R. Walter [4]; S. Schlenstedt [1]; T. Le-Flour [5]; T. Schmidt [1]; U. Schwanke [6]

[1] *Deutsches Elektronen–Synchrotron, DESY, Platanenallee 6, D-15738 Zeuthen, Germany*

[2] *Erlangen Centre for Astroparticle Physics (ECAP), University of Erlangen-Nuremberg, Department of Physics, Erwin-Rommel-Str. 1, D-91058 Erlangen, Germany*

[3] *Centre de Physique des Particules de Marseille 163, avenue de Luminy Case 902, 13288 Marseille cedex 09*

[4] *ISDC, Geneva Observatory, University of Geneva, Chemin d'Ecogia 16, CH-1290 Versoix, Geneva, Switzerland*

[5] *Laboratoire d'Annecy-le-Vieux de Physique des Particules, Universit´e de Savoie, CNRS/IN2P3, F-74941 Annecy-le-Vieux, France*

[6] *Institut für Physik, Humboldt-Universität zu Berlin, Newtonstrasse 15, D-12489 Berlin, Germany*

[7] *Institut de Ciències de l'Espai, IEEC-CSIC, Campus UAB, Facultat de Ciències, Torre C5 par-2, Bellaterra, 08193, Spain*

Corresponding Author: peter.wegner@desy.de

The CTA (Cherenkov Telescope Array) project is an initiative to build the next generation ground-based very high energy (VHE) gamma-ray instrument. Compared to current imaging atmospheric Cherenkov telescope experiments CTA will extend the energy range and improve the angular resolution while increasing the sensitivity by a factor of 10. With these capabilities it is expected that CTA will increase the number of known VHE gamma-ray sources from O(100) to O(1000), and will raise the field of ground based VHE gamma-ray astronomy to the level of astronomy with radio waves or X-rays. With about separate 100 telescopes it will be operated as an observatory open to a wide astrophysics and particle physics community, providing a deep insight into the non-thermal high-energy universe. The presentation will give an overview on the principles of the CTA Array Control system (ACTL), responsible for several essential control tasks including the evaluation, selection, preparation, scheduling, and finally the execution of observations with the array.

A possible basic distributed software framework for ACTL being considered is the ALMA Common Software (ACS). Used by several projects, this open-source software was originally developed for the Atacama Large Millimeter Array (ALMA), a joint project between astronomical organizations in Europe, North America, and Asia for a millimeter and sub-millimeter array. ALMA is presently being commissioned in Chile and will consist of at least fifty-four 12 meter antennas and a further twelve 7 meter antennas.

The ACS framework follows a container component model and contains a high level abstraction layer to integrate different types of device. To achieve a low-level consolidation of connecting control hardware, OPC UA client functionality is integrated directly into ACS, thus allowing interaction with other OPC UA capable hardware.

In addition to the presentation of the ACS middleware, new techniques for automatic code generation based on an UML representation of the ACS components will be introduced and illustrated with first examples.

Poster Session / 33

E-Center: collaborative platform for the Wide Area network users

Author: Maxim Grigoriev[1]

Co-authors: Andrew Lake [2]; Brian Tierney [2]; David Eads [1]; Michael Frey [3]; Philip DeMar [4]; Prasad Calyam [5]; joe metzger [2]

[1] *Fermilab*

[2] *ESnet*

[3] *Bucknell University*

[4] *FERMILAB*

[5] *Ohio SC OARNet*

Corresponding Author: maxim@fnal.gov

The LHC computing model relies on intensive network data transfers.
The E-Center is a social collaborative web based platform for Wide Area

network users. It is designed to give user all required tools to isolate, identify and resolve any network performance related problem.

Summary:

Fermilab is a leading Tier1 facility for US CMS data storage and analysis. It applies extra requirements on the Wide Area network connectivity and expected performance between Tier1 and Tier2 sites. Historically user expectations for data transfer performance across Wide Area networks are rarely met. In order to isolate the probable cause of the sub-optimal performance one needs to obtain network monitoring data from multiple network domains and aggregate these data at some centralized location. Extra steps are desired as well for the transparent network path visualization and advanced anomalous conditions analysis along the network path(s).
The E-Center project, funded by Department's of Energy Office of Science, is designed to become a centralized collaborative platform for the Wide Area network users. This is the place where network user may find answers, identify and isolate any network related problem, exchange information with other users or network experts. It is built on top of webservices architecture and most advanced open source Drupal Content Management System. In the following presentation we cover E-Center design, distributed architecture, Data Retrieval Service, novel network performance visualization ideas, Anomalous Network Events detection and Forecasting services. The E-Center is deployed at https://ecenter.fnal.gov. The most general use cases will be outlined as well.

Poster Session / **34**

Implementation of Intensity Frontier Beam Information Database

Author: Igor Mandrichenko[1]

Co-authors: Andrew Norman [1]; Andrey Petrov [1]; Vladimir Podstavkov [1]

[1] *Fermilab*

Corresponding Author: ivm@fnal.gov

Neutrino physics research is an important part of FNAL scientific program in post Tevatron era. Neutrino experiments are taking advantage of high beam intensity delivered by the FNAL accelerator complex. These experiments share a common beam infrastructure, and require detailed information about the operation of the beam to perform their measurements. We have designed and implemented a system to capture, store and deliver this common beam data to all of the neutrino experiments in real-time. The solution that we designed and built is a robust, high reliability, high performance system that is capable of providing both real-time and historic beam conditions data to the experiments at different stages in their data acquisition and analysis chains. This system is currently being integrated into the online data collection, online monitoring and off-line data processing for each of the experiments. The presentation will cover the design and implementation of this system, its interfaces.

Poster Session / 35

BESIII and SuperB: Distributed job management with Ganga

Authors: Andrea Galvani[1]; Armando Fella[2]; Eleonora Luppi[3]; Luca Tomassetti[3]; Matteo Manzali[3]; Vincenzo Spinoso[4]; Xiaomei Zhang[5]

Co-authors: Alex Richards [6]; Ivan Antoniev Dzhunov [7]; Jakub JakubMoscicki [8]; Johannes Ebke [9]; Mark Slater [10]; Mike Kenyon [8]; Ulrik Egede [11]; Yanliang Han [12]; Ziyan Deng [12]

[1] *Infn - sezione di Ferrara*

[2] *INFN, Italy*

[3] *University of Ferrara and INFN, Italy*

[4] *INFN and Università degli Studi di Bari*

[5] *IHEP, China*

[6] *Imperial College London, UK*

[7] *cern*

[8] *CERN*

[9] *Ludwig-Maximilians-Univesity Muenchen, Germany*

[10] *Birmingham University, UK*

[11] *Imperial College Sci., Tech. & Med. (GB)*

[12] *IHEP*

A job submission and management tool is one of the necessary components in any distributed computing system. Such a tool should provide a user-friendly interface for physics production group and ordinary analysis users to access heterogeneous computing resources, without requiring knowledge of the underlying grid middleware. Ganga, with its common framework and customizable plug-in structure, is such a tool.
This paper will describe how experiment-specific job-management tools for BESIII and SuperB were developed as Ganga plugins, and discuss our experiences of using Ganga.
The BESIII experiment studies electron-positron collisions in the tau-charm threshold region at BEPCII, located in Beijing. The SuperB experiment will take data at the new generation High Luminosity Flavor Factory, under construction in Rome. With its extremely high targeted luminosity (100 times more than previously achieved) it will provide a uniquely important source of data about the details of the New Physics uncovered at hadron colliders. To meet the challenge of rapidly increasing data volumes in the next few years, BESIII and SuperB are both now developing their own distributed computing environments.
For both BESIII and SuperB, the experiment-specific Ganga plugins are described and their integration with the wider distributed system shown. For BESIII, this includes integration with the software system (BOSS) and the Dirac based distributed environment. Interfacing with the BESIII metadata and file catalog for dataset discovery is one of the key parts and is also described. The SuperB experience includes the development of a plugin capable of managing users'analysis and Monte Carlo production jobs and integration of the Ganga job management features with two SuperB-specific information systems: the simulation production bookkeeping database and the data placement database.
The experiences of these two different experiments in developing Ganga plugins to meet their own unique requirements are compared and contrasted, highlighting lessons learned.

Poster Session / 36

FAZIA FRONT-END ELECTRONICS, GLOBAL SYNCHRONIZATION AND TRIGGER DESIGN

Author: Alfonso Boiano[1]

Co-authors: Antonio Ordine [1]; Elio Rosato [2]; Gennaro Tortone [3]; Giulio Spadaccini [4]; Mariano Vigilante [4]; Raffaele Giordano [1]

[1] *INFN*
[2] *INFN UNINA*
[3] *INFN Napoli*
[4] *INFN UNINA*

Corresponding Author: boiano@na.infn.it

FAZIA stands for the Four Pi A and Z Identification Array. This is a project which aims at building a new 4pi particle detector for charged particles. It will operate in the domain of heavy-ion induced reactions around the Fermi energy. It puts together several international institutions in Nuclear Physics.

It is planned to be operating with both stable and radioactive nuclear beams. A large effort on research and development is currently made, especially on digital electronics and pulse shape analysis, in order to improve the detection capabilities.

This contribution will describe electronic layout from detector signal conversion to data transport through optical fiber. System synchronization for “time-of-flight” measurement of particles, trigger development and overall tests will be also dicussed.

Poster Session / **37**

Using Hadoop File System and MapReduce in a small/medium Grid site

Author: Hassen Riahi[1]

[1] *Universita e INFN (IT)*

Corresponding Author: hassen.riahi@cern.ch

Data storage and access represent the key of CPU-intensive and data-intensive high performance Grid computing. Hadoop is an open-source data processing framework that includes, fault-tolerant and scalable, distributed data processing model and execution environment, named MapReduce, and distributed file system, named Hadoop distributed file system (HDFS).

HDFS was deployed and tested within the Open Science Grid (OSG) middleware stack. Efforts have been taken to integrate HDFS with gLite middleware. We have tested the file system thoroughly in order to understand its scalability and fault-tolerance while dealing with small/medium site environment constraints. To benefit entirely from this file system, we made it working in conjunction with Hadoop Job scheduler to optimize the executions of the local physics analysis workflows. The performance of the analysis jobs which used such architecture seems to be promising, making it useful to follow up in the future.

Student? Enter ’yes’. See http://goo.gl/MVv53:

yes

Poster Session / **38**

Multi-threaded Event Reconstruction with JANA

Author: David Lawrence[1]

[1] *Jefferson Lab*

Corresponding Author: davidl@jlab.org

The JANA framework has been deployed and in use since 2007 for development of the GlueX experiment at Jefferson Lab. The multi-threaded reconstruction framework is routinely used on machines with up to 32 cores with excellent scaling. User feedback has also helped to develop JANA into a user-friendly environment for development of reconstruction code and event playback. The basic design of JANA will be presented along with results of scaling tests on many-core machines.

Poster Session / **39**

Workload management in the EMI project

Author: Marco Cecchi[1]

[1] *Istituto Nazionale Fisica Nucleare (IT)*

Corresponding Author: marco.cecchi@cnaf.infn.it

The EU-funded project EMI, now at its second year, aims at providing a unified, high quality middleware distribution for e-Science communities. Several aspects about workload management over diverse distributed computing environments are being challenged by the EMI roadmap: enabling seamless access to both HTC and HPC computing services, implementing a commonly agreed framework for the execution of parallel computations and supporting interoperability models between Grids and Clouds. Besides, a rigourous requirements collection process, involving the WLCG and various NGIs across Europe, assures that the EMI stack is always committed to serving actual needs. With this background, the gLite Workload Management System (WMS), the metascheduler service delivered by EMI, is augmenting its functionality and scheduling models according to the aforementioned project roadmap and the numerous requirements collected over the first project year. This paper is about present and future work of the WMS in EMI, reporting on design changes, implementation choices and long-term vision.

Poster Session / **40**

STEPtoRoot - from CAD to monte carlo simulation

Author: Tobias Stockmanns[1]

[1] *Forschungszentrum Jülich GmbH*

Corresponding Author: t.stockmanns@fz-juelich.de

Modern experiments in hadron and particle physics are searching for more and more rare decays which have to be extracted out of a huge background of particles. To achieve this goal a very high precision of the experiments is required which has to be reached also from the simulation software. Therefore a very detailed description of the hardware of the experiment is needed including also tiny details.
To help the programmer of the simulation software to achieve the required level of detail a semi-automatic tool was developed which is able to convert geometry descriptions coming from CAD programs into root geometries which can be used directly in any root based simulation software.
The features of the conversion program will be presented and results from its use for the PANDA experiment will be shown.

Poster Session / **41**

The Offline Software Framework of the NA61/Shine Experiment

Author: Roland Sipos[1]

Co-authors: Andras Laszlo [1]; Antoni Jerzy Marcinek [2]; Darko Veberic [3]; Marek Szuba [4]; Michael Unger [4]; Oskar Wyszynski [2]; Tom Paul [5]

[1] *Hungarian Academy of Sciences (HU)*

[2] *Jagiellonian University (PL)*

[3] *University of Nova Gorica (SI)*

[4] *KIT - Karlsruhe Institute of Technology (DE)*

[5] *Department of Physics*

Corresponding Author: roland.sipos@cern.ch

NA61/SHINE (SHINE = SPS Heavy Ion and Neutrino Experiment) is an experiment at the CERN SPS using the upgraded NA49 hadron spectrometer. Among its physics goals are precise hadron production measurements for improving calculations of the neutrino beam flux in the T2K neutrino oscillation experiment as well as for more reliable simulations of cosmic-ray air showers. Moreover, p+p, p+Pb and nucleus+nucleus collisions will be studied extensively to allow for a study of properties of the onset of deconfinement and search for the critical point of strongly interacting matter.

Currently NA61/SHINE uses the old NA49 software framework for reconstruction, simulation and data analysis. The core of this legacy framework was developed in the early 1990s. It is written in different programming languages (C, pgi-Fortran) and provides several concurrent data formats including obsolete parts in the data model.

In this contribution we will introduce the new software framework, called Shine, that is written in C++ and designed to comprize three principal parts: a collection of processing modules which can be assembled and sequenced by the user using XML, an event data model which contains all simulation and reconstruction information based on ROOT, and a detector description which provides data on the configuration and state of experiment. To assure a quick migration from to the Shine framework, wrappers were introduced that allow to run legacy code parts as modules in the new framework and we will present first results on the cross validation of the two frameworks.

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Poster Session / **42**

Identification of charmed particles using Multivariate analysis in STAR experiment

Author: Jonathan Bouchet[1]

[1] *Kent State University*

Corresponding Author: bouchet@rcf.rhic.bnl.gov

Due to their production at the early stages, heavy flavor particles are of interest to study the properties of the matter created in heavy ion collisions at RHIC.
Previous measurements of D and B mesons at RHIC[1, 2] using semi-leptonic probes show a suppression similar to that of light quarks, which is in contradiction with theoretical models only including gluon radiative energy loss mechanism[3].
A direct topological reconstruction is then needed to obtain a precise measurement of charm meson

decays. This method leads to a substantial combinatorial background which can be reduced by using modern multivariate techniques (TMVA) which make optimal use of all the information available. Comparison with classical methods and performances of some classifiers will be presented for the reconstruction of D^0 decay vertex ($D^0 \rightarrow K^-\pi^+$) and its charge conjugate from Au+Au collisions at $\sqrt{s_{NN}}$ = 200 GeV.

[1]Adare A. et al., PHENIX Collaboration, arXiv:1005.1627\newline
[2]B.I. Abelev et al., STAR Collaboration, arXiv:0607012v3\newline
[3]Dokshitzer, Yuri L. and Kharzeev, D. E., Phys. Lett. B{\bf 519} \newline

Poster Session / 43

ALICE's detectors safety and efficiency optimization with automatic beam-driven operations

Author: Ombretta Pinazza[1]

Co-authors: Alexander Kurepin [2]; Andre Augustinus [3]; Mateusz Lechman [3]; Peter Chochula [3]; Peter Matthew Bond [4]

[1] *Universita e INFN (IT)*

[2] *Moscow Physical Engineering Institute (MePhl)*

[3] *CERN*

[4] *University of the West of England*

Corresponding Author: ombretta.pinazza@cern.ch

ALICE is one of the four main experiments at the CERN Large Hadron Collider (LHC) in Geneva. The Alice Detector Control System (DCS) is responsible for the operation and monitoring of the 18 detectors of the experiment and of central systems, for collecting and managing alarms, data and commands. Furthermore, it is the central tool to monitor and verify the beam mode and conditions in order to ensure the safety of the detectors.
Experience with systems and beams has allowed for a continuous evolution of the DCS in the direction of automatizing actions based on detector status and beam conditions, which otherwise were left to the judgement of the shift crew. Both the safety of the detectors and the data taking efficiency of the experiment benefits from this strategy.
This paper shows how the DCS is interpreting the daily operations from a beam-driven point of view. A tool is implemented, where automatic actions can be set and monitored through expert panels, with a custom level of automation. Several routine operations are already in place in a fully automatized fashion: e.g. the transition to a safe state of the detectors during critical beam modes such as injection, as communicated by the LHC, to avoid potentially unsafe situations and unnecessary delays to the accelerator procedures.

Event Processing / 45

A GPU-based multi-jet event generator for the LHC

Author: Gerben Stavenga[1]

Co-author: Walter Giele [1]

[1] *Fermilab*

Corresponding Author: giele@fnal.gov

We present a GPU-based parton level event generator for multi-jet events at the LHC. The current implementation generates up to 10 jets with a possible vector boson.

At leading order the speed increase over a single core CPU is in excess of a factor of 500 using a single desktop based NVIDIA Fermi GPU.
We will also present results for the next-to-leading order implementation.

Poster Session / **46**

The ALICE EMCal High Level Triggers

Author: Federico Ronchetti[1]

[1] *Istituto Nazionale Fisica Nucleare (IT)*

Corresponding Author: federico.ronchetti@gmail.com

The ALICE detector yields a huge sample of data, via millions of channels from different sub-detectors. On-line data processing must be applied to select and reduce the data volume in order to increase the significant information in the stored data.

ALICE applies a multi-level hardware trigger scheme where fast detectors are used to feed a three-level deep chain, L0-L2. The High-Level Trigger (HLT) is a fourth filtering stage sitting logically between the L2 trigger and the DAQ (Data AcQuisition) event building.

The EMCal detector comprises a large area electromagnetic calorimeter that extends the measured particle momenta up to pT=200GeV/c, thus ALICE capability to perform jet reconstruction is improved by measuring the neutral energy component of jets, photons and neutral mesons.

An online reconstruction and trigger chain has been developed within the HLT framework to sharpen the EMCal hardware triggers, by combining the central barrel tracking information with the shower reconstruction (clusters) in the calorimeter, thus allowing to obtain a clear and unbiased sample of electron and jet events, both in p-p and A-A LHC runs.

In the present talk the functionalities of the software components of the EMCal/HLT online reconstruction and trigger chain will be discussed. The status and results of the development work will be shown with particular reference to the online chain's physics performance.

Poster Session / **50**

Online Metadata Collection and Monitoring Framework for the STAR Experiment at RHIC

Authors: Dmitry Arkhipkin[1]; Gene Van Buren[1]; Jerome LAURET[2]; Wayne Betts[1]

[1] *Brookhaven National Laboratory*

[2] *BROOKHAVEN NATIONAL LABORATORY*

Corresponding Author: arkhipkin@bnl.gov

The STAR Experiment further exploits scalable message-oriented model principles to achieve a high level of control over online data
streams. In this report we present an AMQP-powered Message Interface and Reliable Architecture framework (MIRA), which allows STAR to orchestrate the activities of Metadata Collection, Monitoring, Online QA and several Run-Time / Data Acquisition system components in a very efficient manner. The very nature of the reliable message bus suggests parallel usage of multiple independent storage mechanisms for our metadata. We describe our experience of a robust data-taking setup employing MySQL and HyperTable based archivers for metadata processing. In addition, MIRA has an AJAX-enabled web GUI, which allows real-time visualisation of online process flow and detector subsystem states, and doubles as a sophisticated alarm system when combined with complex event

processing engines like Esper, Borealis or Cayuga. Reported data and suggested path forward are based on our experience during the 2011-2012 running of STAR.

Poster Session / 51

RECAST

Authors: Kyle Stuart Cranmer[1]; itay Yavin[2]

[1] *New York University (US)*

[2] *New-York University*

Corresponding Author: iy5@nyu.edu

Searches for new physics by experimental collaborations represent a significant investment in time and resources. Often these searches are sensitive to a broader class of models than they were originally designed to test. It is possible to extend the impact of existing searches through a technique we call 'recasting'. We present RECAST, a framework designed to facilitate the usage of this technique.

Summary:

We discuss the general concept of the framework, as well as recent work done on its implementation. The framework consists of the front-end, the back-end, and the supporting APIs each of which is separately discussed and explained.

Poster Session / 52

Analysis of DIRAC's behavior using model checking with process algebra

Author: Daniela Remenska[1]

Co-authors: Henri Bal [2]; Jeff Templon [1]; Kees Verstoep [3]; Tim Willemse [4]

[1] *NIKHEF (NL)*

[2] *Professor of Computer Science*

[3] *Scientific Programmer*

[4] *Assistant Professor*

Corresponding Author: daniela.remenska@cern.ch

DIRAC is the Grid solution designed to support LHCb production activities as well as user data analysis. Based on a service-oriented architecture, DIRAC consists of many cooperating distributed services and agents delivering the workload to the Grid resources. Services accept requests from agents and running jobs, while agents run as light-weight components, fulfilling specific goals. Services maintain database back-ends to store dynamic state information of entities such as jobs, queues, staging requests, etc. Agents use polling to check for changes in the service states, and react to these accordingly. A characteristic of DIRAC's architecture is the relatively low complexity in the logic of each agent; the main source of complexity lies in their cooperation. These agents run concurrently, and communicate using the services' databases as a shared memory for synchronizing the state transitions.

Although much effort is invested in making DIRAC reliable, entities occasionally get into inconsistent states, leading to a potential loss of efficiency in both resource usage and manpower. Tracing

and fixing the root of such encountered behaviors becomes a formidable task due to the inherent parallelism present. In this paper we propose the use of rigorous methods for improving software quality. Model checking is one such technique for analysis of an abstract model of a system , and verification of certain properties of interest. Unlike conventional testing, it allows full control over the execution of parallel processes and also supports exhaustive state-space exploration.

We used the mCRL2 language and toolset to model the behavior of two critical and related DIRAC subsystems:the workload management and the storage management system. mCRL2 is based on process algebra, and is able to deal with generic data types as well as user-defined functions for data transformation. This makes it particulary suitable for modeling the data manipulations made by DIRAC's agents. By visualizing the state space and replaying scenarios with the toolkit's simulator, we have detected critical race-conditions and livelocks in these systems, which we have confirmed to occur in the real system. We further formalized and verified several properties that were considered relevant. Our future direction is exploring to what extent a (pseudo)automatic extraction of a formal model from DIRAC's implementation is feasible. Given the highly dynamic features of the implementation platform (Python), this is a challenging task.

Student? Enter 'yes'. See http://goo.gl/MVv53:

Yes

Online Computing / 53

ALICE moves into warp drive.

Author: Vasco Chibante Barroso[1]

Co-authors: Adriana Telesca [1]; Alexandru Grigore [2]; Barthelemy von Haller [1]; Bartolomeu Andre Rodrigues Fernandes Rabacal [1]; Csaba Soos [1]; Ervin Denes [3]; Filippo Costa [1]; Franco Carena [1]; Giuseppe Simonetti [4]; Pierre Vande Vyvre [1]; Roberto Divia [1]; Sylvain Chapeland [1]; Ulrich Fuchs [1]; Wisla Carena [1]

[1] *CERN*

[2] *Polytechnic University of Bucharest (RO) and CERN*

[3] *Hungarian Academy of Sciences (HU)*

[4] *Universita e INFN (IT)*

Corresponding Author: vasco.chibante.barroso@cern.ch

A Large Ion Collider Experiment (ALICE) is the heavy-ion detector designed to study the physics of strongly interacting matter and the quark-gluon plasma at the CERN Large Hadron Collider (LHC). Since its successful start-up in 2010, the LHC has been performing outstandingly, providing to the experiments long periods of stable collisions and an integrated luminosity that greatly exceeds the planned targets.

To fully explore these privileged conditions, we aim at maximizing the experiment's data taking productivity during stable collisions. We present in this paper the evolution of the online systems in order to spot reasons of inefficiency and address new requirements.

This paper describes the features added to the ALICE Electronic Logbook (eLogbook) to allow the Run Coordination team to identify, prioritize, fix and follow causes of inefficiency in the experiment. Thorough monitoring of the data taking efficiency provides reports for the collaboration to portray its evolution and evaluate the measures (fixes and new features) taken to increase it. In particular, the eLogbook helps decision making by providing quantitative input, which can be used to better balance risks of changes in the production environment against potential gains in quantity and quality of physics data. It will also present the evolution of the Experiment Control System (ECS) to allow on-the-fly error recovery actions of the detector apparatus while limiting as much as possible the loss of integrated luminosity.

The paper will conclude with a review of the ALICE efficiency so far and the future plans to improve its monitoring.

This paper will describe how the ALICE Electronic Logbook (eLogbook) is used to recognize the main causes of inefficiency, allowing the Run Coordination team to identify, prioritize, address and follow them. It will also explain how the eLogbook is used to monitor the data taking efficiency, providing reports that allow the collaboration to portray its evolution and evaluate the measures taken to increase it. Finally, it will present the ALICE efficiency since the start-up of the LHC and the future plans to improve its monitoring.

Poster Session / 54

Evaluating the Control Software for CTA in a Medium Size Telescope Prototype.

Author: Igor Oya[1]

Co-authors: Bagmeet Behera [2]; David Melkumyan [2]; Emrah Birsin [1]; Koeppel Hendryk [2]; Michael Winde [1]; Peter Wegner [1]; Stephan Wiesand [1]; Torsten Schmidt [2]; Ullrich Schwanke [1]

[1] *Institut für Physik, Humboldt-Universität zu Berlin, Newtonstrasse 15, D-12489 Berlin, Germany*

[2] *Deutsches Elektronen–Synchrotron, DESY, Platanenallee 6, D-15738 Zeuthen, Germany*

Corresponding Author: oya@physik.hu-berlin.de

CTA (Cherenkov Telescope Array) is one of the largest ground-based astronomy projects being pursued and will be the largest facility for ground-based gamma-ray observations ever built. CTA will consist of two arrays (one in the Northern hemisphere and one in the Southern hemisphere) composed of several different sizes of telescopes. A prototype for the Medium Size Telescope (MST) type of a diameter of 12 m will be installed in Berlin by the beginning of 2012. This MST prototype will be composed of the mechanical structure, drive system, mirror facets mounted with an active mirror control system. Four CCD cameras and a weather station will allow measurement of the performance of the instrument. The ALMA Common Software (ACS) distributed control framework is currently being considered by the CTA consortium to serve as the array control middleware. In order to evaluate the ACS software, it has been decided to implement an ACS-based readout and control system for the MST prototype. The design of the control software is following the concepts and tools under evaluation within the CTA consortium, like the use of a UML based code generation framework for ACS component modeling, and the use of OPC Unified Architecture (OPC UA) for hardware access. In this contribution the progress in the implementation of the control system for this CTA prototype telescope is described.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Poster Session / 56

The ALICE DAQ Detector Algorithms framework

Author: Sylvain Chapeland[1]

Co-authors: Adriana Telesca [1]; Alexandru Grigore [2]; Barthelemy von Haller [1]; Bartolomeu Andre Rodrigues Fernandes Rabacal [3]; Csaba Soos [1]; Ervin Denes [4]; Filippo Costa [1]; Franco Carena [1]; Giuseppe Simonetti [5]; Pierre Vande Vyvre [1]; Roberto Divia [1]; Ulrich Fuchs [1]; Vasco Chibante Barroso [1]; Wisla Carena [1]

[1] *CERN*

[2] *Polytechnic University of Bucharest (RO)*

[3] *Instituto Superior Tecnico (IST)*

[4] *Hungarian Academy of Sciences (HU)*

[5] *Universita e INFN (IT)*

Corresponding Author: sylvain.chapeland@cern.ch

ALICE (A Large Ion Collider Experiment) is the heavy-ion detector studying the physics of strongly interacting matter and the quark-gluon plasma at the CERN LHC (Large Hadron Collider). The 18 ALICE sub-detectors are regularly calibrated in order to achieve most accurate physics measurements. Some of these procedures are done online in the DAQ (Data Acquisition System) so that calibration results can be directly used for detector electronics configuration before physics data taking, at run time for online event monitoring, and offline for data analysis.

A framework was designed to collect statistics and compute calibration parameters, and has been used in production since 2008. This paper focuses on the recent features developed to benefit from the multi-cores architecture of CPUs, and to optimize the processing power available for the calibration tasks. It involves some C++ base classes to effectively implement detector specific code, with independent processing of events in parallel threads and aggregation of partial results. We present benchmarks showing the performance improvements, and some results of investigations conducted with CUDA and GPUs to push the speed-up further.

The Detector Algorithm (DA) framework provides utility interfaces for handling of input and output (configuration, monitored physics data, results, logging), and self-documentation of the produced executable. New algorithms are created quickly by inheritance of base functionality and implementation of few ad-hoc virtual members, while the framework features are kept expandable thanks to the isolation of the detector calibration code. The DA control system also handles unexpected processes behavior, logs execution status, and collects performance statistics.

Poster Session / **57**

Orthos, an alarm system for the ALICE DAQ operations

Author: Sylvain Chapeland[1]

Co-authors: Adriana Telesca [1]; Alexandru Grigore [2]; Barthelemy von Haller [1]; Bartolomeu Andre Rodrigues Fernandes Rabacal [3]; Csaba Soos [1]; Ervin Denes [4]; Filippo Costa [1]; Franco Carena [1]; Giuseppe Simonetti [5]; Pierre Vande Vyvre [1]; Roberto Divia [1]; Ulrich Fuchs [1]; Vasco Chibante Barroso [1]; Wisla Carena [1]

[1] *CERN*

[2] *Polytechnic University of Bucharest (RO)*

[3] *Instituto Superior Tecnico (IST)*

[4] *Hungarian Academy of Sciences (HU)*

[5] *Universita e INFN (IT)*

Corresponding Author: sylvain.chapeland@cern.ch

ALICE (A Large Ion Collider Experiment) is the heavy-ion detector studying the physics of strongly interacting matter and the quark-gluon plasma at the CERN LHC (Large Hadron Collider). The DAQ (Data Acquisition System) facilities handle the data flow from the detectors electronics up to the mass storage. The DAQ system is based on a large farm of commodity hardware consisting of more than 600 devices (Linux PCs, storage, network switches), and controls hundreds of distributed hardware and software components interacting together.

This paper presents Orthos, the alarm system used to detect, log, report, and follow-up abnormal situations on the DAQ machines at the experimental area.

The main objective of this package is to integrate alarm detection and notification mechanisms with a full-featured issues tracker, in order to prioritize, assign, and fix system failures optimally. This tool relies on a database repository with a logic engine, SQL interfaces to inject or query metrics, and dynamic web pages for user interaction. We describe the system architecture, the technologies used for the implementation, and the integration with existing monitoring tools.

Poster Session / 58

Belle II Data Handling System

Authors: Junghyun Kim[1]; Kihyeon Cho[1]; Soonwook Hwang[1]; Sunil Ahn[1]; Taegil Bae[1]; Taesang Huh[1]

[1] *KISTI*

Corresponding Authors: cho@kisti.re.kr, hyun@kisti.re.kr, siahn@kisti.re.kr, tshuh@kisti.re.kr, hwang@kisti.re.kr, esrevinu@kisti.re.kr

In order to search for new physics beyond the standard model, the next generation of B-factory experiment, Belle II will collect a huge data sample that is a challenge for computing systems. The Belle II experiment, which should commence data collection in 2015, expects data rates 50 times higher than that of Belle. In order to handle this amount of data, we need a new data handling system based on a new computing model, which is a distributed computing model including grid farms as opposed to the central computing model using clusters at the Belle experiment.

The existing Belle data handling system has problems with performance, scalability, and robustness at the projected Belle II data rate, which makes it inappropriate for the Belle II experiment. Moreover, the solution applied by Belle is not intended to be used in a distributed environment. Therefore, the goal of the Belle II data handling system is to make a reliable and efficient metadata system based on grid farms.

In this talk, we explain the architecture, characteristics, components and interactions of them for the Belle II data handling system. We also show the user scenario for the data handling system. To determine where the files are located on the grid and thus to which sites the jobs that process these files should be submitted, we uses the LCG File Catalog (LFC).

Distributed Processing and Analysis on Grids and Clouds / 59

Scalable proxy cache for Grid Data Access

Author: Cristian Cirstea[1]

Co-authors: David Groep [2]; Jan Just Keijser [2]; Jeff Templon [3]; Oscar Arthur Koeroo [4]; Ronald Starink [5]

[1] *Technische Universtiteit Eindhoven*

[2] *NIKHEF*

[3] *NIKHEF (NL)*

[4] *Sticht. Fund. Onderzoek der Materie (FOM)*

[5] *Unknown*

Corresponding Author: templon@nikhef.nl

This contribution describes a prototype grid proxy cache system developed at Nikhef, motivated by a desire to construct the first
building block of a future https-based Content Delivery Network for multiple-VO grid infrastructures. Two goals drove the project:
firstly to provide a “native view” of the grid for desktop-type users, and secondly to improve performance for physics-analysis type use cases, where multiple passes are made over the same set of data (residing on the grid). We further constrained the design by
requiring that the system should be made of standard components wherever possible.

The prototype that emerged from this exercise is a horizontally-scalable, cooperating system of web server / cache nodes, fronted by a customized webDAV server. The webDAV server is custom only in the sense that it supports HTTP redirects (providing horizontal scaling) and that the authentication module has, as back
end, a proxy delegation chain that can be used by the cache nodes to retrieve files from the grid.

The prototype was deployed at Nikhef and tested at a scale of several terabytes of data and approximately one hundred fast cores of
computing. Both small and large files were tested, in a number of scenarios, and with various numbers of cache nodes, in order to
understand the scaling properties of the system. For properly-dimensioned cache-node hardware, the system showed speedup of
several integer factors for the analysis-type use cases. These results and others are presented and discussed in this contribution.

Poster Session / **60**

FlyingGrid : from volunteer computing to volunteer cloud

Author: oelg lodygensky[1]

Co-authors: Derrick Kondo [2]; Etienne Urbah [3]; Gilles Fedak [2]; Laurent Duflot [4]; Simon Dadoun [5]; Simon Delamare [2]; Xavier Garrido [5]

[1] *LAL - IN2P3 - CNRS*

[2] *INRIA*

[3] *Lab. de l'Accelerateur Lineaire (IN2P3) (LAL) - Universite de Pa*

[4] *Universite de Paris-Sud 11 (FR)*

[5] *CNRS*

Corresponding Author: lodygens@lal.in2p3.fr

Desktop grid (DG) is a well known technology aggregating volunteer computing resources donated by individuals to dynamically construct a virtual cluster. A lot of efforts are done these last years to extend and interconnect desktop grids to other distributed computing resources, especially focusing on so called "service grids"middleware such as "gLite", "ARC"and "Unicore".
In the former "EDGeS"european project (http://edges-grid.eu/), work has been done on standardizing and securing desktop grids to propose, since 2010, a new platform exposing an uniformed view of resources aggregated from DG run by Boinc (http://boinc.berkeley.edu/) or XtremWeb-HEP (http://www.xtremweb-hep.org/), and resources aggregated from EGEE (http://www.eu-egee.org/). Today, the current "EDGI"european project (http://edgi-project.eu/) extends the EDGeS platform by integrating "ARC"and "Unicore"middleware. This project also includes cloud related research topics. In this paper we present our first results on integrating cloud technology into desktop grid. This work has two goals. First goal is to permit to desktop grid users to deploy and use their own virtual machines over a set of volunteer resources aggregated over DG. Second goal is to continue to propose a standardized view to the user who would wish to submit jobs as well as virtual machines

Summary:

This paper first introduces standardization efforts done in EDGeS and EDGI. Cloud and virtualization over DG are then presented. We present our solution over XtremWeb-HEP and standardization effort to transparently submit jobs to both grid and cloud, as well a to transparently submit virtual machines to both grid and cloud. Finally we present some use cases where our platform is used by ATLAS and SuperNemo users.

Poster Session / **61**

Taking Global Scale Data Handling to the Fermilab Intensity Frontier

Author: Adam Lyon[1]

Co-authors: Andrew Norman [1]; Frederick Snider [2]; Marc Mengel [1]; Robert Illingworth [1]

[1] *Fermilab*

[2] *FERMILAB*

Corresponding Author: lyon@fnal.gov

Fermilab Intensity Frontier experiments like Minerva, NOvA, g-2 and Mu2e currently operate without an organized data handling system, relying instead on completely manual management of files on large central disk arrays at Fermilab. This model severely limits the computing resources that the experiments can leverage to those tied to the Fermilab site, prevents the use of coherent staging and caching of files from tape and other mass storage media, and produces an onerous burden on the individuals responsible for data processing.

The SAM data handling system[1], used by the Fermilab Tevatron experiments CDF and D0 for Run II (2002-2011), solves these problems by providing data set abstraction, automated file cataloging and management, global delivery, and processing tracking. It has been a great success at CDF and D0 achieving global delivery rates of ˜1.5 PB/week/experiment for raw data, Monte Carlo, production and analysis activities. However, SAM has been heavily tailored for integration in both CDF and D0 analysis frameworks, making it difficult and time-consuming to repeat that work for new experiments. The command line user interface is also complex, non-intuitive and represents a tall barrier for new and casual users. These issues have slowed the adoption of SAM by Intensity Frontier experiments. The Fermilab Computing Sector is improving SAM with a generic "deployment-less" HTTP based client for analysis framework integration and an intuitive FUSE[2] based user interface to permit universal adoption of SAM across the Intensity Frontier.

We will describe these solutions in detail, their technical implementation, and their impact on the adoptability of SAM for new experiments.

[1] http://projects.fnal.gov/samgrid/
[2] http://http://fuse.sourceforge.net/

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Poster Session / 62

EMI-european Middleware Initiative

Author: giuseppina salente[1]

Co-author: Emidlo Giorgio [2]

[1] *INFN*

[2] *Istituto Nazionale Fisica Nucleare (IT)*

Corresponding Author: emidio.giorgio@ct.infn.it

The EMI project intends to receive or rent an exhibition spot nearby the main and visible areas of the event (such as coffee-break areas), to exhibit the projects goals and the latest achievements, such as the EMI1 release.
The means used will be posters, video and distribution of flyers, sheets or brochures. It would be useful to have a 2x3 booth with panels available to post on posters, and some basic furniture as table, 2 chairs, a lamp, a wired/wi-fi connection, electrical outlet.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Summary:

A joint effort of the major European distributed computing middleware providers. Distributed, secure compute and data management services to support and evolve the research infrastructures and allow academic and industrial researchers to access resources, data and applications across the world. EMI improves the existing middleware services and harmonizes them, realizing a common framework for building, certifying and distributing with the result of rendering the middleware to be simpler and easier to use. EMI reduces and aims to solve the interoperability problems faced by the distributed computing infrastructure communities. http://www.eu-emi.eu/
The exhibition booth will showcase the latest achievements, such as the first Release (EMI1) and its features, a video/demo showing the advantage of using the EMI products for research, distribution of brochures and/or leaflets.

Poster Session / **63**

MARDI-Gross - Data Management Design for Large Experiments

Author: Roger Jones[1]

Co-authors: Brian Matthews [2]; Juan Bicarregui [2]; Norman Gray [3]; Robert Henderson [4]; Simon Lambert [2]

[1] *Lancaster University (GB)*

[2] *STFC-RAL*

[3] *The University of Glasgow*

[4] *Lancaster University*

Corresponding Author: roger.jones@cern.ch

MARDI-Gross builds on previous work with the LIGO collaboration, using the ATLAS experiment as a use case to develop a tool-kit on data management for people making proposals for large High Energy Physics experiments, as well a experiments such as LIGO and LOFAR, and also for those assessing such proposals. The toolkit will also be of interest to those in the active data management for new and current experiments.

Summary:

Data management and data preservation in science has moved from an issue for projects to a matter for public discussion. Citizen science and public access to public data have joined outreach, education, long-term data archival and analysis in the afterlife of collaborations as major items. Accordingly, research funding agencies are introducing data management policies that make far greater demands than before.

MARDI-Gross builds on previous work with the LIGO collaboration, using the ATLAS experiment as a use case to develop a tool-kit on data management for people making proposals for large High Energy Physics experiments, as well a experiments such as LIGO and LOFAR, and also for those assessing such proposals. The toolkit will also be of interest to those in the active data management for new and current experiments.

Poster Session / **64**

New data visualization of the LHC Era Monitoring (Lemon) system

Author: Ivan Fedorko[1]

Co-authors: Daniel Lenkes [1]; Omar Pera Mira [1]

[1] *CERN*

Corresponding Author: ivan.fedorko@cern.ch

In the last few years, new requirements have been received for visualization of monitoring data: advanced graphics, flexibility in configuration and decoupling of the presentation layer from the monitoring repository.
Lemonweb is the data visualization component of the LHC Era Monitoring (Lemon) system. Lemonweb consists of two sub-components: a data collector and a web visualization interface.
The data collector is a daemon, implemented in Python, responsible for data gathering from the central monitoring repository and storing into time series data structures. Data are stored on disk in Round Robin Database (RRD) files: one file per monitored entity, with all the available monitoring data. Entities may be grouped into a hierarchical structure, called "clusters"and supporting mathematical operations over entities and clusters (e.g. cluster A + cluster B /clusters C –entity XY). Using the configuration information, a cluster definition is evaluated in the collector engine and, at runtime, a sequence of data selects is built, to optimize access to the central monitoring repository.
An overview of the design and architecture as well as highlights of some implemented features will be presented. The CERN Computer Centre instance, visualizing ˜17k entries, will be described, with an example of the advanced cluster configuration and integration with the CLUMAN (a job management and visualization system) visualization module.

Event Processing / **66**

Track Reconstruction in Belle 2

Author: Moritz Nadler[1]

Co-authors: Jakob Lettenbichler [2]; Rudolf Fruhwirth [1]

[1] *Austrian Academy of Sciences (AT)*

[2] *HEPHY Vienna, Austria*

Corresponding Authors: moritz_nadler@gmx.de, rudolf.fruehwirth@oeaw.ac.at, jkl@jodoschka.com

The Silicon Vertex Detector (SVD) of the Belle II experiment is a newly developed device with four measurement layers. The detector is designed to enable track reconstruction down to the lowest momenta possible, in order to significantly increase the effective data sample and the physics potential of the experiment. Both track finding and track fitting have to deal with these requirements. We describe the outline of the track finding procedure and details of the track fit. An immportant aspect of the latter is the correct treatment of material effects such as multiple Coulomb scattering and energy loss by ionization at very low particle energies. As the SVD is an ultra-light design, non-Gaussian tails in the multiple scattering distributions are non-negligible and have to be dealt with. We present results from a Deterministic Annealing Filter (DAF) and compare its performance to the baseline Kalman filter. Both methods are implemented using the GENFIT package. We describe the various modifications and improvements of GENFIT that are required for a successful application in the Belle II environment.

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Collaborative tools / 67

From EVO to SeeVogh

Author: Philippe Galvez[1]

Co-author: Harvey Newman [2]

[1] *CALTECH*

[2] *California Institute of Technology (US)*

Corresponding Author: galvez@caltech.edu

Collaboration Tools, Videoconference, support for large scale scientific collaborations, HD video

Summary:

The EVO (Enabling Virtual Organizations) system is based on a new distributed and unique architecture, leveraging the 14+ years of unique experience of developing and operating large distributed production based collaboration systems. The primary objective being to provide to the High Energy and Nuclear Physics experiments a system/service that meet their unique requirements of usability, quality, scalability, reliability, and cost necessary for nationally and globally distributed research organizations. Today, he EVO system is heavily use by the LHC and more generally by High Energy and Nuclear Physics community and the LIGO community with more than 5,000 meetings a month.

As more features/functionality as been added to the system to better support the research community, we developed a new advanced and unified client called SeeVogh fully compatible with previous version that will be available to the community via authenticated portal (CERN, LIGO, etc..) using unified SSO.

The new service model and SeeVogh client will be described during this talk.

Poster Session / 68

Service Oriented Tracking: A Package For CLAS12 Reconstruction Using Clara Framework

Author: Sebouh Paul[1]

Co-author: Vardan Gyurjyan [2]

[1] *Jefferson Lab*

[2] *JEFFERSON LAB*

Corresponding Author: sebouh.paul@gmail.com

In the advent of the 12 GeV upgrade at CEBAF, it becomes necessary to create new detectors to accommodate the more powerful beam-line. It follows that new software is needed for tracking, simulation and event display. In the case of CLAS12, the new detector to be installed in Hall B, development has proceeded on new analysis frameworks and runtime environments, such as the Clara (CLAS12 Reconstruction and Analysis) framework. Our goal is to create a tracking program for the forward components of the CLAS12 which takes advantage of the service oriented architecture provided by the Clara framework. The tracking program must group together hits from the detector that were caused by the same particle, determine which type of particle it was (particle identification), and estimate its vector momentum and a point on its trajectory. We have an additional requirement of timing: the program must be fast enough that the reconstruction rate is comparable to the data acquisition rate. Also, due to the complexity of these sorts of programs, modularity is necessary. The purpose of our study is to create a program that meets all of the requirements of a tracking program,

while using a service oriented architecture to enhance timing as well as flexibility (software agility and scaling).

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Poster Session / 69

The Version Control Service for ATLAS Data Acquisition System Configuration Files

Author: Igor Soloviev[1]

[1] *University of California Irvine (US)*

Corresponding Author: igor.soloviev@cern.ch

To configure data taking run the ATLAS systems and detectors store more than 150 MBytes of data acquisition related configuration information in OKS[1] XML files. The total number of the files exceeds 1300 and they are updated by many system experts. In the past from time to time after such updates we had experienced problems with configuring of a run caused by XML syntax errors or inconsistent state of files from overall ATLAS configuration point of view. It was not always possible to know who made a modification caused problem or how to go back to previous version of modified file.

Few years ago the special service for XML files addressing the issues has been implemented and deployed on ATLAS Point-1. It excludes direct write access to XML files stored in central database repository. Instead for an update the files are copied into user repository, validated after modifications and committed using CVS server. The server's callback updates the central repository. Also, the CVS keeps track of all modifications allowing Web interface for browsing details of the modifications or restoring any previous version of files. The paper provides details of implementation and exploitation experience that maybe interesting for others using various files for configuration purposes.

---

[1] "The ATLAS DAQ system online configurations database service challenge", I.Soloviev et al., CHEP 2007 and J. Phys.: Conf. Ser. 119:022004

Poster Session / 70

experience with the custom-developed ATLAS trigger monitoring and reprocessing infrastructure

Authors: Martin Erik Gerd Zur Nedden[1]; Valeria Bartsch[2]

Co-authors: Carsten Kendziorra [1]; Diego Casadei [3]; Simon George [4]

[1] *Humboldt-Universitaet zu Berlin (DE)*

[2] *University of Sussex (GB)*

[3] *New York University (US)*

[4] *University of London (GB)*

Corresponding Authors: diego.casadei@cern.ch, valeria.bartsch@cern.ch, carsten.kendziorra@cern.ch

After about two years of data taking with the ATLAS detector manifold experience with the custom-developed trigger monitoring and reprocessing infrastructure could be collected.

The trigger monitoring can be roughly divided into online and offline monitoring. The online monitoring calculates and displays all rates at every level of the trigger and evaluates up to 3000 data quality histograms. The physics analysis relevant data quality information is being checked and recorded automatically. The offline trigger monitoring provides information depending of the physics motivated different trigger streams after a run has finished. Experts are checking the information being guided by the assessment of algorithms checking the current histograms with a reference. The experts are recording their assessment in a so-called data quality defects database which is being used to build a good run list of data good enough for physics analysis. In the first half of 2011 about three percent of all data had an intolerable defect resulting from the ATLAS trigger system.

To keep the percentage of data with defects low any changes of trigger algorithms or menus must be tested reliabely. A recent run with a sufficient statistics (in the order of one million events) is being reprocessed to check that the changes do not introduce any unexpected side-effects. The current framework for the reprocessing is a GRID production system custom built for ATLAS requirements called PANDA [1]. The reprocessed datasets are being checked in the same offline trigger monitoring framework that is being used for the offline trigger data quality. It turned out, that the current system works very reliable and all potential problems could be faced.

[1] PANDA: T. Maeno [ATLAS Collaboration], PanDA: Distributed production and distributed analysis system for ATLAS, J.Phys.Conf.Ser.119(2008)

Poster Session / 71

A System for Monitoring and Tracking the LHC Beam Spot within the ATLAS High Level Trigger

Authors: Andrey Salnikov[1]; Emanuel Alexandre Strauss[1]; Frank Winklmeier[2]; Josh Cogan[1]; Rainer Bartoldus[1]

Co-author: Chris Bee [3]

[1] *SLAC*

[2] *CERN*

[3] *Universite d'Aix - Marseille II (FR)*

Corresponding Authors: chris.bee@cern.ch, bartoldu@slac.stanford.edu

The parameters of the beam spot produced by the LHC in the ATLAS interaction region are computed online using the ATLAS High Level Trigger (HLT) system. The high rate of triggered events is exploited to make precise measurements of the position, size and orientation of the luminous region in near real-time, as these parameters change significantly even during a single data-taking run. We present the challenges, solutions and results for the online determination, monitoring and beam spot feedback system in ATLAS. A specially designed algorithm, which uses tracks registered in the silicon detectors to reconstruct event vertices, is executed on the HLT processor farm of several thousand CPU cores. Monitoring histograms from all the cores are sampled and aggregated across the farm every 60 seconds. The reconstructed beam values are corrected for detector resolution effects, measured in situ from the separation of vertices whose tracks have been split into two collections. Furthermore, measurements for individual bunch crossings have allowed for studies of single-bunch distributions as well as the behavior of bunch trains, calibrated to the beam average. Run control invokes a comparison of the nominal and measured beam spot values, and when threshold conditions are satisfied the farm configuration is updated. To achieve sharp time boundaries across the event stream, which is triggered at rates of several kHz, a special datagram is injected into the event path via the Central Trigger Processor that signals the pending update to the trigger nodes. Thousands of clients then fetch the same set of values from the conditions database in a fraction of a second via an efficient near-simultaneous access made possible through a dedicated CORAL Server and Proxy tree.

Poster Session / 72

The Electronic Logbook for the Information Storage of ATLAS Experiment at LHC

Author: Alina Corso Radu[1]

Co-authors: Giovanna Lehmann Miotto [2]; Luca Magnoni [2]

[1] *University of California Irvine (US)*

[2] *CERN*

Corresponding Authors: luca.magnoni@cern.ch, alina.radu@cern.ch

A large experiment like ATLAS at LHC (CERN), with over three thousand members and a shift crew of 15 people running the experiment 24/7, needs an easy and reliable tool to gather all the information concerning the experiment development, installation, deployment and exploitation over its lifetime. With the increasing number of users and the accumulation of stored information since the experiment start-up, the electronic logbook actually in use, ATLOG, started to show its limitations in terms of speed and usability. Its monolithic architecture makes the maintenance and implementation of new functionality a hard-to-almost-impossible process. A new tool ELisA has been developed to replace the existing ATLOG. It is based on modern web technologies: the Spring framework using a Model-View-Controller architecture was chosen, thus helping building flexible and easy to maintain applications. The new tool implements all features of the old electronic logbook with increased performance and better graphics: it uses the same database back-end for portability reasons. In addition, several new requirements have been accommodated which could not be implemented in ATLOG. This paper describes the architecture, implementation and performance of ELisA, with particular emphasis on the choices which allowed to have a scalable and very fast system and on the aspects that could be re-used in different contexts to build a similar application.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Poster Session / 73

Architecture and performance of the ATLAS Inner Detector Trigger software

Authors: Jiri Masik[1]; Nikolaos Konstantinidis[2]

Co-author: Pauline Bernat [3]

[1] *University of Manchester (GB)*

[2] *University College London (GB)*

[3] *LAL*

Corresponding Authors: bernat@hep.ucl.ac.uk, nikolaos.konstantinidis@cern.ch, pauline.bernat@gmail.com

The rising instantaneous luminosity of the LHC poses an increasing challenge to the pattern recognition algorithms for track reconstruction at the ATLAS Inner Detector Trigger. We will present the performance of these algorithms in terms of signal efficiency, fake tracks and execution time, as a function of the number of proton-proton collisions per bunch-crossing, in 2011 data and in simulation.

The strict time requirements at the Level-2 Trigger, where the average execution time per event is expected to be around 40 millisecs, make the pattern recognition particularly challenging. ATLAS has so far used both histogramming-based and combinatorial algorithms for the task of Level-2 track

reconstruction. In light of the experience from the data taking in 2011, a new software framework is being developed that will provide a suite of configurable tools, based on modularising the existing code and increasing the re-use of components, to provide the optimal solution in the various trigger signatures at higher luminosities. This new framework, as well as the work to optimise the overall performance of the Inner Detector Trigger software, will also be presented.

Poster Session / **74**

low momentum track finding in Belle 2

Author: Jakob Lettenbichler[1]

Co-authors: Moritz Nadler ; Rudi Frühwirth [2]

[1] *HEPHY Vienna, Austria*

[2] *Institut fuer Hochenergiephysik (HEPHY)*

Corresponding Authors: jkl@jodoschka.com, rudolf.fruehwirth@oeaw.ac.at, moritz_nadler@gmx.de

The Silicon Vertex Detector (SVD) of the Belle II experiment is a newly developed device with four measurement layers. Track finding in the SVD will be done both in conjunction with the Central Drift Chamber and in stand-alone mode. The reconstruction of very-low-momentum tracks in stand-alone mode is a big challenge, especially in view of the low redundancy and the large expected background. We describe two approaches for track finding in this domain, a cellular automaton and a combinatorial Kalman filter. Both methods are combined with a Hopfield network which finds an optimal subset of non-overlapping tracks. We present results on simulated data and compare the two methods in terms of efficiency, purity and speed

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Poster Session / **75**

The ATLAS Level-1 Trigger System

Author: Gabriel Anders[1]

Co-author: Will Buttinger [2]

[1] *Ruprecht-Karls-Universitaet Heidelberg (DE)*

[2] *University of Cambridge (GB)*

Corresponding Authors: will@cern.ch, gabriel.anders@cern.ch

The ATLAS Level-1 Trigger is the first stage of event selection for the ATLAS experiment at the LHC. In order to identify the interesting collisions events to be passed on to the next selection stage within a latency of less than 2.5 us, it is based on custom-built electronics. Signals from the Calorimeter and Muon Trigger System are combined in the Central Trigger Processor which processes the overall L1 Accept (L1A) decision. The Level-1 Trigger identifies event features such as missing transverse energy, candidate electrons, photons, jets and muons. This talk will present how the Level-1 Trigger System has performed with increasing LHC luminosity and discuss problems encountered during operations. We will also give an overview of the challenges and plans with respect to the increasingly demanding LHC running conditions.

Poster Session / 76

GPU-based algorithms for ATLAS High-Level Trigger

Author: Dmitry Emeliyanov[1]

Co-author: Jacob Russell Howard [2]

[1] *STFC - Science & Technology Facilities Council (GB)*

[2] *University of Oxford (GB)*

Corresponding Authors: jacob.howard@cern.ch, dmitry.emeliyanov@stfc.ac.uk

One possible option for the ATLAS High-Level Trigger (HLT) upgrade for higher LHC luminosity is to use GPU-accelerated event processing. In this talk we discuss parallel data preparation and track finding algorithms specifically designed to run on GPUs. We present a "client-server" solution for hybrid CPU/GPU event reconstruction which allows for the simple and flexible integration of the specific GPU-accelerated algorithms into existing ATLAS HLT software. The resulting speed-up of event processing times obtained with high-luminosity simulated data are presented and discussed.

Poster Session / 77

Automated Inventory and Monitoring of the ALICE HLT Cluster Resources with the SysMES Framework

Author: Jochen Ulrich[1]

Co-authors: Camilo Ernesto Lara Martinez [1]; Dieter Roehrich [2]; Oystein Haaland [2]; Stefan Boettger [3]; Udo Wolfgang Kebschull [1]

[1] *Johann-Wolfgang-Goethe Univ. (DE)*

[2] *University of Bergen (NO)*

[3] *Kirchhoff-Institut fuer Physik (KIP)-Ruprecht-Karls-Universitaet*

Corresponding Author: j.ulrich@iri.uni-frankfurt.de

The High-Level-Trigger (HLT) cluster of the ALICE experiment is a computer cluster with about 200 nodes and 20 infrastructure machines. In its current state, the cluster consists of nearly 10 different configurations of nodes in terms of installed hardware, software and network structure. In such a heterogeneous environment with a distributed application, information about the actual configuration of the nodes is needed to automatically distribute and adjust the application accordingly. An inventory database provides a unified interface to such information. To be useful, the data in the inventory has to be up to date, complete and consistent with itself. Manual maintenance of such databases is error-prone and data tends to become outdated. The inventory module of the ALICE HLT cluster overcomes these drawbacks by automatically updating the actual state periodically and, in contrast to existing solutions, it allows the definition of a target state for each node. A target state can simply be a fully operational state, i.e. a state without malfunctions, or a dedicated configuration of the node. The target state is then compared to the actual state to detect deviations and malfunctions which could induce severe problems when running the application. The inventory module of the ALICE HLT cluster has been integrated into the monitoring and management framework SysMES in order to use existing functionality like transactionality, monitors and clients. Additionally, SysMES allows to solve detected problems automatically via its rule-system. To describe the heterogeneous environment with all its specifics, like custom hardware, the inventory module uses an object-oriented model which is based on the Common Information Model. To summarize, the inventory module provides an automatically updated actual state of the cluster, detects discrepances between the actual and the target state and is able to solve detected problems automatically. This contribution presents the current implementation state of the inventory module as well as the future development.

Software Engineering, Data Stores and Databases / 78

Massively parallel Markov chain Monte Carlo with BAT

Author: Frederik Beaujean[1]

Co-authors: Allen Caldwell [1]; Daniel Kollar [1]; Julia Grebenyuk [2]; Kevin Alexander Kroeninger [3]; Shabnaz Pashapouralamdari [3]

[1] *Max Planck Institute for Physics*

[2] *DESY*

[3] *Georg-August-Universitaet Goettingen (DE)*

Corresponding Author: beaujean@mpp.mpg.de

The Bayesian Analysis Toolkit (BAT) is a C++ library designed to analyze data through the application of Bayes' theorem.
For parameter inference, it is necessary to draw samples from the posterior distribution within the given statistical model. At its core, BAT uses an adaptive Markov Chain Monte Carlo (MCMC) algorithm.

As an example of a challenging task, we consider the analysis of rare B-decays in a global fit involving about 20
observables measured at the B-factories and by the CDF and LHCb collaborations.
A single evaluation of the likelihood requires approximately 1 s.
In addition to the 3 − 12 parameters of interest, there are on the order of 25 nuisance parameters describing uncertainties from standard model parameters as well as
from unknown higher order theory corrections and non-perturbative QCD effects.
The resulting posterior distribution is multi-modal and shows significant
correlation between parameters as well as pronounced degeneracies, hence
the standard MCMC methods fail to produce accurate results.

Parallelization is the only solution to obtain a sufficient number of samples in reasonable time.
We present an enhancement of existing MCMC algorithms, including the ability for massive parallelization
on a computing cluster and, more importantly,
a general scheme to induce rapid convergence even in the face complicated posterior distributions.

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Poster Session / 79

The ATLAS Muon Trigger at high instantaneous luminosities

Author: Alexander Oh[1]

[1] *University of Manchester (GB)*

Corresponding Author: alexander.oh@cern.ch

The ATLAS experiment at CERN's Large Hadron Collider (LHC) has taken data with colliding beams at instantaneous luminosities of *2\10^33 cm^-2 s^-1. The LHC targets to deliver an integrated luminosity 5-fb in the run period 2011 at luminosities of up to 5*10^33 cm^-2 s^-1, which requires dedicated strategies to guard the highest physics output while reducing effectively the event rate.

The muon system is the largest sub-detector of the ATLAS experiment and has the capability to reconstruct muons in standalone mode, as well as in combination with the Inner Detector tracking systems. The L1 muon trigger system gets its input from fast muon trigger detectors. Fast sector logic boards select muon candidates, which are passed via an interface board to the central trigger processor and then to the High Level Trigger (HLT). The Muon HLT is purely software based and encompasses a level 2 trigger followed by an event filter for a staged trigger approach. It has access to the data of the precision muon detectors and other detector elements to refine the muon hypothesis.

The Muon HLT has successfully adapted to the changing environment of the low luminosity running of LHC in 2010 to the intensities encountered in 2011. The selection strategy has been optimized for the various physics analysis involving muons in the final state. This includes the use of isolation at the level 2 and event filter, combined trigger signatures with electron and jet trigger objects, and so-called full-scan triggers, which make use of the full event information to search for di-lepton signatures, seeded by single lepton objects.

This note reports about efficiency, resolution, and general performance of the muon trigger in the context of the physics goals of ATLAS.

Poster Session / **80**

Experience with highly-parallel software for the storage system of the ATLAS experiment at CERN

Authors: Tommaso Colombo[1]; Wainer Vandelli[2]

Co-author: Marius Tudor Morar [3]

[1] *Universita e INFN (IT)*

[2] *CERN*

[3] *University of Manchester (GB)*

Corresponding Authors: tudor.morar@cern.ch, wainer.vandelli@cern.ch

The ATLAS experiment is observing proton-proton collisions delivered by the LHC accelerator at a centre of mass energy of 7 TeV. The ATLAS Trigger and Data Acquisition (TDAQ) system selects interesting events on-line in a three-level trigger system in order to store them at a budgeted rate of several hundred Hz, for an average event size of ˜1.2 MB.
This paper focuses on the TDAQ data-logging system and in particular on the implementation and performance of a novel SW design, reporting on the effort of exploiting the full power of recently installed multi-core hardware. In this respect, the main challenge presented by the data-logging workload is the conflict between the largely parallel nature of the event processing, especially the recently introduced on-line event-compression, and the constraint of sequential file writing and checksum evaluation. This is additionally complicated by the necessity of operating in a fully data-driven mode, to cope with continuously evolving trigger and detector configurations.
The novel SW design is based on a thread-pool, implemented in C++ using modern parallel programming tools and techniques, as provided by libraries like TBB(1) and Boost(2). Lock-less patterns, atomic operations and concurrent containers have been employed to provide an efficient implementation able to cope with the above requirements.
In this paper we report on the design of the new ATLAS on-line storage software. In particular we will discuss our development experience using recent concurrency-oriented libraries. Finally we will show the new system performance with respect to the old, single-threaded software design.

Online Computing / **81**

Performance of the ATLAS trigger system

Author: Brian Petersen[1]

Co-author: Diego Casadei [2]

[1] *CERN*

[2] *New York University (US)*

Corresponding Authors: diego.casadei@cern.ch, brian.petersen@cern.ch

The ATLAS trigger has been used very successfully to collect collision data during 2009-2011 LHC running at centre of mass energies between 900 GeV and 7 TeV. The three-level trigger system reduces the event rate from the design bunch-crossing rate of 40 MHz to an average recording rate of about 300 Hz. The first level uses custom electronics to reject most background collisions, in less than 2.5 us, using information from the calorimeter and muon detectors. The upper two trigger levels are software-based triggers. The trigger system selects events by identifying signatures of muon, electron, photon, tau lepton, jet, and B meson candidates, as well as using global event signatures, such as missing transverse energy. We give an overview of the performance of these trigger selections based on extensive online running during the 2011 LHC run and discuss issues encountered during 2011 operations. Distributions of key selection variables are shown calculated at the different trigger levels and are compared with offline reconstruction. Trigger efficiencies with respect to offline reconstructed signals are shown and compared to simulation, illustrating a very good level of understanding of the detector and trigger performance. We describe how the trigger has evolved with increasing LHC luminosity coping with pileup conditions close to LHC design luminosity.

Poster Session / 83

Monitoring the data quality of the real-time event reconstruction in the ALICE High Level Trigger.

Author: Hege Austrheim Erdal[1]

[1] *Bergen University College (NO)*

Corresponding Author: hege.austrheim.erdal@cern.ch

ALICE (A Large Ion Collider Experiment) is a dedicated heavy ion experiment at the Large Hadron Collider (LHC). The High Level Trigger (HLT) for ALICE is a powerful, sophisticated tool aimed at compressing the data volume and filtering events with desirable physics content. Several of the major detectors in ALICE are incorporated into HLT to compute real-time event reconstruction, for instance the Inner Tracking System (ITS), the Time Projection Chamber (TPC), the electromagnetic calorimeters (EMCAL), the Transition Radiation Detector (TRD) and the muon spectrometer.

The HLT is used for real-time event reconstruction which provides the input for trigger algorithms. It is necessary to monitor the quality of the reconstruction where one focuses on track and event properties. Also, HLT implements data compression for the TPC in the heavy ion data taking in 2011 to reduce the data rate from the ALICE detector. The key for the data compression is to store clusters calculated by HLT rather than storing raw data. It is thus very important to monitor the cluster finder performance as a way to monitor the data compression.

The data monitoring is divided into two stages. The first stage is performed during data taking. A part of the HLT production chain is dedicated to perform online monitoring and facilities are available in the HLT production cluster to have real-time access to the reconstructed events in the ALICE control room. This includes track and event properties, and in addition this facility gives a way to display a small fraction of the reconstructed events in an online display. The second part of the monitoring is

performed after the data has been transferred to permanent storage. After a post-process of the real-time reconstructed data, one can look in more detail at the cluster finder performance, the quality of the reconstruction of tracks, vertices and vertex position. The monitoring solution will be presented in detail, with special attention to the heavy ion data taking of 2010 and 2011.

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Poster Session / 84

The Alignment of the BESIII Drift Chamber Using Cosmic-ray Data

Author: Linghui Wu[None]

Corresponding Author: wulh@ihep.ac.cn

BESIII/BEPCII is a major upgrade of the BESII experiment at the Beijing Electron-Positron Collider (BEPC) for studies of hadron spectroscopy and tau-charm physics. The BESIII detector adopts a small cell helium-based drift chamber (MDC) as the cetral tracking detector. The momentum resolution was deteriorated due to misalignment in the data taking. In order to improve the momentum resolution, a software alignment is necessary to reduce the effect of mechanical imperfection on the reconstruction. The BESIII alignment software was developed in the framework of the BESIII Offline Software System (BOSS). It was applied in the alignment of the drift chamber using cosmic-ray data successfully. The momentum resolution was improved significantly after the alignment. The report will show the alignment method. The alignment results will also be reported.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Poster Session / 85

Agents and Daemons, automating Data Quality Monitoring operations.

Author: Luis Ignacio Lopera Gonzalez[1]

[1] *Universidad de los Andes (CO)*

Corresponding Author: luis.ignacio.lopera.gonzalez@cern.ch

Since 2009 when the LHC came back to active service, the Data Quality Monitoring (DQM) team was faced with the need to homogenize and automate operations across all the different environments within which DQM is used for data certification.

The main goal of automation is to reduce operator intervention at the minimum possible level, especially in the area of DQM files management, where long-term archival presented the greatest challenges. Manually operated procedures cannot cope with the constant increase in luminosity, datasets and time of operation of the CMS detector. Therefore a solid and reliable set of agents has been designed since the beginning to manage all DQM-data related work-flows. This allows to fully exploit all available resources in every condition, maximizing the performance and reducing the latency in making data available for validation and certification. The agents can be easily fine-tuned to adapt to current and future hardware constraints and they proved to be flexible enough to include unforeseen features, like an ad-hoc quota management and a real time sound alarm system.

Poster Session / 86

Resource Utilization by the ATLAS High Level Trigger during 2010 and 2011 LHC running

Authors: Douglas Michael Schaefer[1]; Elliot Lipeles[1]; Rustem Ospanov[1]

[1] *University of Pennsylvania (US)*

Corresponding Author: douglas.michael.schaefer@cern.ch

Since starting in 2010, the Large Hadron Collider (LHC) has produced collisions at an ever increasing rate. The ATLAS experiment
successfully records the collision data with high efficiency and excellent data quality. Events are selected using a three-level trigger system, where each level makes a more rened selection. The level-1 trigger (L1) consists of a custom-designed hardware trigger which seeds two higher software based trigger levels. Over 300 triggers compose a trigger menu which selects physics signatures such as electrons, muons, particle jets, etc. Each trigger consumes computing resources of the ATLAS trigger system and oine storage. The LHC instantaneous luminosity conditions, desired physics goals of the collaboration, and the limits of the trigger infrastructure determine the composition of the ATLAS trigger menu. We describe a trigger monitoring framework for computing the costs of individual trigger algorithms such as data request rates and CPU consumption. This framework has been used to prepare the ATLAS trigger for data taking during increases of more than six orders of magnitude in the LHC luminosity and has been influential in guiding ATLAS Trigger computing upgrades.

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Summary:

Since starting in 2010, the Large Hadron Collider (LHC) has produced collisions at an ever increasing rate. The ATLAS experiment
successfully records the collision data with high efficiency and excellent data quality. I will discuss a framework which monitors the ATLAS trigger and has been used to make predictions for future data taking.

Poster Session / 87

The First Prototype for the FastTracker Processing Unit

Authors: Agostino Lanza[1]; Marco Piendibene[2]; Mauro Citterio[1]

Co-authors: Alberto Annovi [3]; Alberto Stabile [4]; Alessandro Andreani [4]; Andrea Negri [1]; Daniel Magalotti [5]; Fabrizio Alberti [4]; Fukun Tang [6]; Guido Volpi [3]; Lauren Alexandra Tompkins [7]; Matteo Mario Beretta [3]; Mel Shochet [6]; Mircea Bogdan [8]; Paola Giannetti [9]

[1] *Universita e INFN (IT)*

[2] *Universita di Pisa-Sezione di Pisa (INFN)*

[3] *Istituto Nazionale Fisica Nucleare (IT)*

[4] *INFN - Milano*

[5] *INFN - Perugia*

[6] *University of Chicago (US)*

[7] *Lawrence Berkeley National Lab. (LBNL)*

[8] *The University of Chicago*

[9] *Sezione di Pisa (IT)*

Corresponding Authors: andrea.negri@pv.infn.it, paola.giannetti@pi.infn.it

Modern experiments search for extremely rare processes hidden in much larger background levels. As the experiment
complexity and the accelerator backgrounds and luminosity increase we need increasingly complex and exclusive selections.
We present the first prototype of a new Processing Unit, the core of the FastTracker processor for Atlas, whose computing
power is such that a couple of hundreds of them will be able to reconstruct all the tracks with transverse momentum above 1
GeV in the ATLAS events up to Phase II instantaneous luminosities (5×1034 cm-2 s-1) with an event input rate of 100 kHz and
a latency below hundreds of microseconds. We plan extremely powerful, very compact and low consumption units for the far
future, essential to increase efficiency and purity of the Level 2 selected samples through the intensive use of tracking.
This strategy requires massive computing power to minimize the online execution time of complex tracking algorithms.
The time consuming pattern recognition problem, generally referred to as the "combinatorial challenge", is beat by the
Associative Memory (AM) technology [2] exploiting parallelism to the maximum level: it compares the event to precalculated
"expectations"or "patterns"(pattern matching) at once looking for candidate tracks called "roads". This approach
reduces to linear the typical exponential complexity of the CPU based algorithms. The problem is solved by the time data are
loaded into the AM devices.
We describe the board prototypes that face the very challenging aspects of the Processing Unit: a huge amount of detector
clusters ("hits") must be distributed at high rate with very large fan-out to all patterns (10 Millions of patterns will be located
on 128 chips placed on a single board) and a huge amount of roads must be collected and sent back to the FTK post-patternrecognition
functions. The Processing Unit consists of a 9U VME board, the AMBoard, controlled by an AUX card on the
back of the crate. The AMBoard has a modular structure consisting of 4 mezzanines, the Local Associative Memory Banks
(LAMB). Each LAMB contains 32 Associative Memory (AM) chips, 16 per side. The proto - AUX card provides hits on 8
buses for a total of 12 Gbits/sec to the AMBoard through 12 high frequency serial links and will sink the found roads trough
other 16 high frequency serial links (24 Gbits/sec). A special P3 connector allows the communication between the front and
rear boards placed on the same VME slot. A custom board profile has been studied and simulated at the CAD to guarantee a
perfect board-to-board closure of the P3 connector without a backplane support in that region. A network of high speed serial
links characterize the bus distribution on the AMBoard. The hit buses are fed to the four LAMBs and distributed to the 32 AM
chips on the LAMB, through fanout chips. The LAMB realization has represented a significant technological challenge, due to
the high density of chips allocated on both sides, and to the use of advanced packages and high frequency serial links.
We report on the design and first tests of the Processing Unit.
[1] A. Andreani et al., The FastTracker Real Time Processor and Its Impact on Muon Isolation, Tau and b-Jet Online
Selections at ATLAS, Conference Record 17th IEEE NPSS Real Time Conference Record of the 17th Real Time Conference,
Lisbon, Portugal, 24 - 28 May 2010.
[2] M. Dell'Orso and L. Ristori, "VLSI structures for track finding", Nucl. Instr. and Meth., vol. A278, pp. 436-440, (1989).

Poster Session / **88**

An Information System to Access Status Information of the LHCb Online

Author: Markus Frank[1]

Co-author: Clara Gaspar [1]

[1] *CERN*

Corresponding Author: markus.frank@cern.ch

The LHCb collaboration consists of roughly 700 physicists from 52 institutes and universities. Most of the collaborating physicists - including subdetector experts - are not permanently based at CERN. This paper describes the architecture used to publish data internal to the LHCb experiment control- and data acquisition system to the world wide web. Collaborators can access the online (sub-)system status and the system performance directly from the institute abroad, from home or from a smart phone without the need of direct access to the online computing infrastructure. The information is presented to them in form of web pages with a similar look and feel as it is provided by the experiment controls system.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Poster Session / **89**

Optimization of the HLT Resource Consumption in the LHCb Experiment

Author: Markus Frank[1]

[1] *CERN*

Corresponding Author: markus.frank@cern.ch

Today's computing elements for software based high level trigger processing (HLT) are based on nodes with multiple cores. Using process based parallelisation to filter particle collisions from the LHCb experiment on such nodes leads to expensive consumption of read-only memory and hence significant cost increase. In the following an approach is presented to fork multiple identical processes from a master process. This approach facilitated to minimize the resource consumption of the filter applications and to reduce the startup time. Described is the duplication of threads and the handling of files open in read-write mode when duplicating filter processes and the possibility to bootstrap the event filter applications directly from preconfigured checkpoint files. Emphasis was put on the condition, that the trigger code itself is agnostic to this process. The approach led to a reduced memory consumption of roughly 60 % in each worker node of the LHCb HLT farm and an overall reduced startup time of roughly 70 %.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Poster Session / **90**

Dynamic parallel ROOT facility clusters on the Alice Environment

Author: Cinzia Luzzi[1]

Co-authors: Anar Manafov [2]; Costin Grigoras [3]; Federico Carminati [4]; Latchezar Betev [4]; Pablo Saiz [4]

[1] *CERN - University of Ferrara*

[2] *GSI - Helmholtzzentrum fur Schwerionenforschung GmbH (DE)*

[3] *Conseil Europeen Recherche Nucl. (CERN)*

[4] *CERN*

Corresponding Author: cinzia.luzzi@cern.ch

The ALICE collaboration has developed a production environment (AliEn) that implements several components of the Grid paradigm needed to simulate, reconstruct and analyze data in a distributed way.
In addition to the Grid-like analysis, ALICE, as many experiments, provides a local interactive analysis using the Parallel ROOT Facility (PROOF).
PROOF is part of the ROOT analysis framework used by ALICE. It enables physicists to analyze and understand much larger datasets on a shorter time scale, allowing analysis of data in parallel on remote computer clusters.
The default installation of PROOF is a static shared cluster provided by administrators. However, using a new framework, PoD (Proof on Demand), PROOF can be used in a more user-friendly and convenient way, giving the possibility to dynamically set up a cluster after the user request.
Integrating PoD in the AliEn environment, different sets of machines can become workers allowing the system to react to an increasing number of requests for PROOF sessions by starting an higher number of proofd processes.
This paper will describe the integration of PoD framework in AliEn in order to provide private dynamic PROOF clusters. This functionality is transparent to the user who will only need to perform a job submission to the AliEn environment.

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Poster Session / 91

Balancing the resources of the High Level Trigger farm of the ATLAS experiment

Author: Marius Tudor Morar[1]

Co-authors: Nicoletta Garelli [2]; Wainer Vandelli [2]

[1] *University of Manchester (GB)*

[2] *CERN*

Corresponding Author: tudor.morar@cern.ch

The ATLAS High Level Trigger (HLT) is organized in two trigger levels running different selection algorithms on heterogeneous farms composed of off-the-shelf processing units. The processing units have varying computing power and can be integrated using diverse network connectivity. The ATLAS working conditions are changing mainly due to the constant increase of the LHC instantaneous luminosity, and consequently requiring the rolling expansion and replacement of the HLT hardware. Therefore, balancing the available resources is essential for optimizing the HLT
farm exploitation. In this paper, a tool for managing the HLT resources will be presented. The tool allows for showing, modifying and generating the HLT farm configuration, keeping the resource balance across the farms in terms of computing power and bandwidth under control.

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Poster Session / **92**

Scaling the AFS service at CERN

Author: Arne Wiebalck[1]

[1] *CERN*

Corresponding Author: arne.wiebalck@cern.ch

Serving more than 3 billion accesses per day, the CERN AFS cell is one of the most active installations in the world. Limited by overall cost, the ever increasing demand for more space and higher I/O rates drive an architectural change from small high-end disks organised in fibre-channel fabrics towards external SAS based storage units with large commodity drives. The presentation will summarise the challenges to scale the AFS service at CERN, discuss the approach taken, and highlight some of the applied techniques, such as SSD block level caching or transparent AFS server failover.

Software Engineering, Data Stores and Databases / **93**

Status and Future Perspectives of CernVM-FS

Author: Jakob Blomer[1]

Co-authors: Artem Harutyunyan [2]; Dag Larsen [3]; Predrag Buncic [2]; Rene Meusel [2]

[1] *Ludwig-Maximilians-Univ. Muenchen (DE)*

[2] *CERN*

[3] *University of Bergen (NO)*

Corresponding Author: jakob.blomer@cern.ch

The CernVM File System (CernVM-FS) is a read-only file system used to access HEP experiment software and conditions data. Files and directories are hosted on standard web servers and mounted in a universal namespace. File data and meta-data are downloaded on demand and locally cached. CernVM-FS has been originally developed to decouple the experiment software from virtual machine hard disk images and to be used as a replacement of the shared software area at Grid sites. Here it allows for the provision of an essentially zero-maintenance software service. CernVM-FS solves the scalability issues of network file systems such as AFS, NFS, or Lustre, which are traditionally used for shared software areas.

Currently, CernVM-FS distributes around 30 million files and directories. It is installed on a large portion of the Worldwide LHC Computing Grid (WLCG) worker nodes supporting the ATLAS and LHCb experiments. In order to scale to the order of 10ˆ5 worker nodes, CernVM-FS uses replicated repository servers and a hierarchy of web caches. Repository replica servers are operated at CERN, BNL, RAL, and ASGC Tier 1 sites. We will report on the lessons learned from the HEP community feedback and the experience from large-scale deployment.

For the server side, we present a new, streamlined and improved toolset to maintain repositories. The new toolset is supposed to reduce the delay for distributing new software releases to less than an hour. It provides parallel preprocessing of files and it introduces "push replication" of updates by means of a replication manager. The simplified repository maintenance also lowers the bar for small collaborations to distribute their software on the Grid.

Finally, we present the roadmap for the further development of CernVM-FS. The roadmap includes Mac OS X support, variable algorithms for file compression and content hashing, as well as a distributed shared memory cache for diskless server farms.

Summary:

The CernVM File System (CernVM-FS) provides a scalable, reliable and essentially zero-maintenance software distribution service. It was developed to assist HEP collaborations to deploy their software on the worldwide-distributed computing infrastructure used to run their data processing applications. CernVM-FS is deployed on a wide range of computing resources, ranging from powerful worker nodes at Tier 1 grid sites to simple virtual appliances running on volunteer computers. The key contribution is a new approach to stage updates and changes into the file system, which aims to reduce the delay in distributing a software release to less than an hour. In addition, it significantly reduces the complexity with respect to both required capabilities of the master storage as well as installation and maintenance. We will report on key scalability figures gathered from normal operational in production use cases. Furthermore, we will discuss new requirements for additional features that have been arisen from HEP community feedback and present the road map for the future development of the file system.

Distributed Processing and Analysis on Grids and Clouds / 94

CernVM Co-Pilot: an Extensible Framework for Building Scalable Cloud Computing Infrastructures

Author: Artem Harutyunyan[1]

Co-authors: Dag Larsen [2]; Ioannis Charalampidis [3]; Jakob Blomer [4]; Predrag Buncic [1]

[1] *CERN*

[2] *University of Bergen (NO)*

[3] *Aristotle Univ. of Thessaloniki (GR)*

[4] *Ludwig-Maximilians-Univ. Muenchen (DE)*

Corresponding Author: artem.harutyunyan@cern.ch

CernVM Co-Pilot is a framework for instantiating an ad-hoc computing infrastructure on top of distributed computing resources. Such resources include commercial computing clouds (e.g. Amazon EC2), scientific computing clouds (e.g. CERN lxcloud), as well as the machines of users participating in volunteer computing projects (e.g. BOINC). The framework consists of components that communicate using the Extensible Messaging and Presence protocol (XMPP), allowing for new components to be developed in virtually any programming language and interfaced to existing Grid and batch computing infrastructures exploited by the High Energy Physics community. Co-Pilot has been used to execute jobs for both the ALICE and ATLAS experiments at CERN.

CernVM Co-Pilot is also one of the enabling technologies behind the LHC@home 2.0 volunteer computing project, which is the first such project that exploits virtual machine technology. The use of virtual machines eliminates the necessity of modifying existing applications and adapting them to the volunteer computing environment. After start of the public testing in August 2011 LHC@home 2.0 quickly gained popularity, and as of October 2011 it had about 9000 registered volunteers. Resources provided by volunteers are used for running Monte-Carlo generator applications that simulate interactions between the colliding proton beams at the LHC.

In this contribution we present the latest developments and the current status of the system, discuss how the framework can be extended to suit the needs of a particular scientific community, describe

the operational experience using the LHC@home 2.0 volunteer computing infrastructure, as well as introduce future development plans.

Summary:

CernVM Co-Pilot is a framework for instantiating an ad-hoc computing infrastructure on top of academic and commercial computing clouds, or on the machines of users participating in volunteer computing projects.
The framework consists of components that communicate using the Extensible Messaging and Presence protocol (XMPP), allowing for new components to be developed in virtually any programming language and interfaced to existing Grid and batch computing infrastructures.
In this contribution we present the latest developments and the current status of the system, discuss how the framework can be extended to suit the needs of a particular scientific community, describe the operational experience using the LHC@home 2.0 volunteer computing infrastructure, as well as introduce future development plans.

Poster Session / **95**

Jigsaw: A runtime-configurable HEP analysis framework

Author: Riccardo Di Sipio[1]

Co-author: Marino Romano

[1] *Universita e INFN (IT)*

Corresponding Author: riccardo.di.sipio@cern.ch

Jigsaw provides a collection of tools for high-energy physics analyses. In Jigsaw's paradigm input data, analyses and histograms are factorized so that they can be configured and put together at run-time to give more flexibility to the user.

Analyses are focussed on physical objects such as particles and event shape quantities. These are distilled from the input data and brought to the analysis via ntuple wrappers, for which a base-class and some use-case examples are provided.

Manipulators can be applied to the events in order to calculate analysis-specific quantities and objects such as decayed particles and polarization angles. Jigsaw is shipped with a comprehensive collection of event cuts that can be composed at run-time via xml to build a cut-based analysis. Finally, histograms are defined externally via xml and filled at each stage of the analysis automatically. As for now still a work in progress, an infrastructure is also present for the creation of ROOT trees and multivariate analyses.

Jigsaw was designed and coded by R. Di Sipio (disipio@bo.infn.it) and M. Romano (marino.romano@bo.infn.it). The code is publicly available on CERN SVN:
https://svnweb.cern.ch/cern/wsvn/atlasgrp/Institutes/Bologna/AnalysisFramework

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Poster Session / **96**

High Speed Data Receiver Card for Future Upgrade of Belle II DAQ

Authors: Nobuhiko Katayama[1]; Takeo Higuchi[2]

[1] *HIGH ENERGY ACCELERATOR RESEARCH ORGANIZATION*

[2] *KEK*

Corresponding Author: higuchit@post.kek.jp

We present performance study of a high-speed RocketIO receiver card implemented as PCI-express device intended for the use in future luminosity-frontier HEP experiment.

To search for a new physics beyond the Standard Model, we start Belle II experiment from 2015 in KEK, Japan. In Belle II, the detector signals are digitized in or nearby the detector complex, and the digitized signals are transmitted to VME-9U sized data receiving boards located about 10m away from the detector over RocketIO optical links. The data receiving board is responsible to provide pipeline, online data processor, and Ethernet outlet connected to external event building PC. For a possible future upgrade of the data receiving board, we design a RocketIO receiver card to be attached to a PC as a PCI-express device.

In addition to above, the device is a backup solution for a data receiver from DEPFET pixel detectors of Belle II. In the backup solution, we plan to process the pixel data using GPUs.

We study firmware performance implemented in a prototype device, which has (up to) four optical input and eight lanes of PCI-express output. Data transfer throughputs for input line cases of one and four are measured 3.7Gbps and 11.8Gbps, respectively.

The first version card next to the prototype is under development and will be delivered by March 2012. Performance study of the first version card will also be presented as well.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Poster Session / **97**

Building a Prototype of LHC Analysis Oriented Computing Centers

Authors: Giacinto Donvito[1]; Giuseppe Bagliesi[2]

Co-authors: Giuseppe Della Ricca [3]; Marco Paganoni [4]; Tommaso Boccali [2]

[1] *Universita e INFN (IT)*

[2] *Sezione di Pisa (IT)*

[3] *University & INFN, Trieste*

[4] *Univ. degli Studi Milano-Bicocca (IT)*

Corresponding Authors: giacinto.donvito@cern.ch, tommaso.boccali@cern.ch

A Consortium between four LHC Computing Centers (Bari, Milano, Pisa and Trieste) has been formed in 2010 to prototype Analysis-oriented facilities for CMS data analysis, using a grant from the Italian Ministry of Research. The Consortium aims to the realization of an

ad-hoc infrastructure to ease the analysis activities on the huge data set collected by the CMS Experiment, at the LHC Collider. While "Tier2" Computing Centres, specialized in organized processing tasks like Monte Carlo simulation, are nowadays a well established concept, with years of running experience, site specialized towards end user chaotic analysis activities do not yet have a de-facto standard implementation. In our effort, we focus on all the aspects which can make the analysis tasks easier for a physics user not expert in computing. On the storage side, we are experimenting on storage techniques allowing for remote data access and on storage optimization on the typical analysis access patterns. On the networking side, we are studying the differences between flat and tiered LAN architecture, also using virtual partitioning of the same physical networking for the different use patterns. Finally, on the user side, we are developing tools and instruments to allow for an exhaustive monitoring of their processes at the site, and for an efficient support system in case of problems.
We will report about the results of the test executed on different subsystem and give a description of the layout of the infrastructure in place at the site participating to the consortium.

Poster Session / **98**

The Fermi-LAT Dataprocessing Pipeline

Author: Stephan Zimmer[1]

Co-authors: Andrei Tsaregorodtsev [2]; Claudia Lavalley [3]; Luisa Arrabito [3]; Tom Glanzmann [4]; Tony Johnsson [4]

[1] *OKC/ Stockholm University, on behalf the Fermi-LAT Collaboration*

[2] *Universite d'Aix - Marseille II (FR)*

[3] *IN2P3, LUPM*

[4] *SLAC, Fermi-LAT Collaboration*

The Data Handling Pipeline ("Pipeline") has been developed for the Fermi Gamma-Ray Space Telescope (Fermi) Large Area Telescope (LAT) which launched in June 2008. Since then it has been in use to completely automate the production of data quality monitoring quantities, reconstruction and routine analysis of all data received from the satellite and to deliver science products to the collaboration and the Fermi Science Support Center. In addition it receives heavy use in performing production MonteCarlo tasks. In daily use it receives a new data download every 3 hours and launches about 2000 jobs to process each download, typically completing the processing of the data before the next download arrives. The need for manual intervention has been reduced to less than <.01% of submitted jobs.

The Pipeline software is written almost entirely in Java and comprises several modules. The software comprises web-services that allow online monitoring and provides AIDA charts summarizing work flow aspects and performance information. The server supports communication with several batch systems such as LSF and BQS and recently also Sun Grid Engine and Condor. This is accomplished through dedicated JobControlDaemons that for Fermi are running at SLAC and the other computing site involved in this large scale framework, the Lyon computing center of IN2P3. While being different in the logic of a task, we evaluate a separate interface to the Dirac system in order to communicate with EGI sites to utilize Grid resources, using dedicated Grid optimized systems rather than developing our own.

More recently the pipeline and its associated data catalog have been generalized for use by other experiments, and are currently being used by the Enriched Xenon Observatory (EXO), Cryogenic Dark Matter Search (CDMS) experiments as well as for MonteCarlo simulations for the future Cherenkov Telescope Array (CTA).

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Poster Session / 99

A business model approach for a sustainable Grid infrastructure in Germany

Authors: Achim Streit[1]; Andreas Heiss[2]; Holger Marten[3]; Ruediger Berlich[4]; Torsten Antoni[2]; Wilhelm Buehler[5]

[1] *KIT - Karlsruhe Institute of Technology (KIT)*

[2] *KIT - Karlsruhe Institute of Technology (DE)*

[3] *Forschungszentrum Karlsruhe GmbH (FZK)*

[4] *FORSCHUNGSZENTRUM KARLSRUHE*

[5] *KIT*

Corresponding Author: torsten.antoni@kit.edu

After a long period of project-based funding,during which the improvement of the services provided to the user communities was the main focus, distributed computing infrastructures (DCIs), having reached and established production quality, now need to tackle the issue of long-term sustainability.
With the transition from EGEE to EGI in 2010 the major part of the responsibility (especially financially) now is on the national grid initiatives (NGIs). It is their duty not only to ensure the unobstructed continuation of scientific work on the grid, but also to cater for the needs of the user communities to be able to utilise a broader range of middlewares and tools.
Sustainability in grid computing therefore must take into account the integration of this variety of technical developments. Newer developments like cloud computing need to be taken into account and integrated into the usage scenarios of the grid infrastructure, leading to a distributed computing infrastructure encompassing the positive aspects of both.
On the whole a strategy for sustainability must focus on the three main aspects of technical integration, core services and business development and must make concrete statements how the respective efforts can be financed. Although not common in science, it seems necessary to use a business model approach to create a business plan to enable the long-term sustainability of the NGIs and international DCIs, like EGI.

Summary:

This talk presents a business plan as suggested for the national German Grid initiative NGI-DE. It is based on quantitative calculations, making it possible to forecast profits and losses, according to a set of mandatory services and “products”. The presentation also wants to solicit input from the relevant grid user communities, like WLCG, with the goal of creating a common basis for and common understanding of sustainability strategies.

Poster Session / 100

xGUS - a helpdesk template for grid user support

Author: Torsten Antoni[1]

[1] *KIT - Karlsruhe Institute of Technology (DE)*

Corresponding Author: torsten.antoni@kit.edu

The xGUS helpdesk template is aimed at NGIs, DCIs and user communities wanting to structure their user support and integrate it with the EGI support.
xGUS contains all basic helpdesk functionalities. It is hosted and maintained at KIT in Germany. Portal administrators from the client DCI or user community can customize the portal to their specific needs. Via web, they can edit the support units, variables which are used for the classification of tickets like 'Type of problem', 'VO'etc. and the hyperlinks to related web pages displayed on the portal.

The xGUS portal is a template framework for a helpdesk system. It is based on BMC Remedy ARS with an Oracle database for the tickets and a MySQL database for news and user administration.
The portal contains various features needed to provide effective user support. Users can access the portal using their grid certificate imported into their browser or via login and password. They can submit tickets via a web form and classify their problem by setting e.g. a 'type of problem', an 'affected site' or a priority. The tickets get assigned to the appropriate support unit by the first level support. The responsible support unit gets informed via email about open tickets. The user can choose whether he wants to stay up to date about every step of process or only get notified once the problem is solved. Users and support staff can also use an email interface to add comments to the ticket. When replying to an email received from the helpdesk, the answer text is added to the ticket history.
Support staff can create relations between different tickets. If several tickets depend on the solution of another one, they can be marked as slaves. When the master ticket is solved, the solution is transferred to the slaves and they are solved automatically. If one ticket depends on the solution of several other tickets, these tickets are marked as children of this ticket. Only when all of the child tickets are solved, the parent ticket can be solved, too.
With the news module, which is included in the portal, events and news can be announced via the portal. Registered users can add tickets of their personal interest to their dashboard.
Subscription to a ticket triggers email notifications about ticket updates for interested users who are not the submitter.
Tickets which can not be solved within the helpdesk instance, can be duplicated to GGUS, the EGI helpdesk. All changes which are made in GGUS are synchronized to the original ticket.
The xGUS instance is a tool to track and document problems. It can also be used to collect statistics on the problem solving process.
Based on the same technology, adjustments the interface between GGUS and xGUS can be made quickly and efficiently. Clients need not care about technical details of their helpdesk system. All server related issues are handled at KIT, as well as the operation and maintenance of the helpdesk portal itself.

Summary:

With the xGUS framework DCIs and user communities have easy access to their own, independent helpdesk system with many helpful features. They can benefit of the experience gained over several years in the GGUS team instead of starting from scratch with a new helpdesk system. Their user support can be integrated into the existing and well-established structure with GGUS at the center. All problems described in the tickets are stored in databases as well as the steps that have been done to obtain a solution. Each helpdesk system becomes a problem database which can help to solve similar or related problems.
The helpdesk system gives project leaders and users the possibility to gain an overview of problems and to find out where improvements could be necessary.
xGUS offers a comfortable way for user communities, who need a user support infrastructure, to obtain an independent helpdesk portal which provides all necessary functionality to track and classify problems. It enables a quick and easy communication between the user and the support staff. Helpdesk administrators can customize the helpdesk to the specific needs of the community. They can set links on the portal which are helpful for users like domumentations or other relevant web pages.
The use of xGUS guarantees a consistent user support infrastructure linked into the central support systems of the major grid infrastructures.

Poster Session / **101**

Designing the ATLAS trigger menu for high luminosities

Author: Monica Dunford[1]

Co-author: Yu.nakahama Higuchi [1]

[1] *CERN*

Corresponding Authors: yu.nakahama@cern.ch, monica.dunford@cern.ch

The LHC, at design capacity, has a bunch-crossing rate of 40 MHz whereas the ATLAS detector has an average recording rate of about 300 Hz. To reduce the rate of events but still a maintain high efficiency of selecting rare events such as Higgs Boson decays, a three-level trigger system is used in ATLAS. Events are selected based on physics signatures such as events with energetic leptons, photons, jets or large missing energy. In total, the ATLAS trigger systems consists of more than 300 different individual triggers.

The ATLAS trigger menu specifies which triggers are used during data taking and how much rate a given trigger is allocated. This menu must reflect not only the physics goals of the collaboration but also take into consideration the instantaneous luminosity of the LHC and the design limits of the ATLAS detector. We describe the criteria for designing the trigger menu for different LHC luminosities that spanned many orders of magnitude during the 2010 and 2011 running periods. We discuss how the trigger menu is tested and validated before being used for data taking, how the prescale values for different triggers are determined and how the menu as a whole is monitored during data taking itself.

Poster Session / **102**

Handling of network and database instabilities in CORAL

Authors: Alexander Kalkhof[1]; Andrea Valassi[1]; Raffaello Trentadue[2]

[1] *CERN*

[2] *Universita e INFN (IT)*

Corresponding Author: andrea.valassi@cern.ch

The CORAL software is widely used by the LHC experiments for storing and accessing data using relational database technologies. CORAL provides a C++ abstraction layer that supports data persistency for several backends and deployment models, including local access to SQLite files, direct client access to Oracle and MySQL servers, and read-only access to Oracle through the FroNTier/Squid and CoralServer/!CoralServerProxy server/cache systems.

During 2010, several problems were reported by the LHC experiments using CORAL, involving application hangs or crashes after the network or the database servers became temporarily unavailable. CORAL already provided some level of handling of these instabilities, which are due to external causes and cannot be avoided, but this proved to be insufficient in some cases and to be itself the cause of other problems, such as the hangs mentioned before, in other cases. As a consequence, a major redesign of the CORAL plugins was implemented, with the aim of making the software more robust against these network glitches. The new implementation ensures that CORAL automatically reconnects to the database in a transparent way whenever possible and gently terminates the application when this is not possible. Internally, it takes care of resetting all relevant parameters of the underlying backend technology (such as OCI, the Oracle Call Interface). This presentation will report on the status of this work at the time of the CHEP2012 conference, covering the design and implementation of these new features and the results from the first experience with their use.

Poster Session / **103**

Monitoring in CORAL

Authors: Alexander Kalkhof[1]; Alexander Loth[2]; Andrea Valassi[1]; Raffaello Trentadue[3]

[1] *CERN*

[2] *CERN/University of the West of England*

[3] *Universita e INFN (IT)*

Corresponding Author: andrea.valassi@cern.ch

The CORAL software is widely used by the LHC experiments for storing and accessing data using relational database technologies. CORAL provides a C++ abstraction layer that supports data persistency for several backends and deployment models, including local access to SQLite files, direct client access to Oracle and MySQL servers, and read-only access to Oracle through the FroNTier/Squid and CoralServer/!CoralServerProxy server/cache systems.

Given the huge amount of operations executed by several CORAL clients at the same time on several database servers, it was crucial to develop a monitoring system with two main goals: first, to allow individual CORAL users to study and optimize the performance of the relational operations executed by their applications; second, to check whether the whole system is properly working and well configured. Client-level monitoring functionalities already existed in CORAL, but they have recently been reviewed and significantly improved, especially for the Oracle and Frontier plugins, and the same functionality are also being integrated into the CORAL server component (itself a CORAL-based application) and its client plugin. Work is in progress also on the monitoring of the CoralServerProxy components and on the aggregation of the monitoring information these proxies provide when deployed in a hierarchical structure, such as that used by the ATLAS High Level Trigger system. This presentation will report on the status of this work at the time of the CHEP2012 conference, covering the design and implementation of these new features and the results from the first experience with their use.

Software Engineering, Data Stores and Databases / **104**

LCG Persistency Framework (POOL, CORAL, COOL) - Status and Outlook

Authors: Alexander Kalkhof[1]; Alexander Loth[2]; Andrea Valassi[1]; Andrey Salnikov[3]; Dave Dykstra[4]; David Front[5]; Marcin Nowak[6]; Marco Clemencic[1]; Markus Frank[1]; Martin Wache[7]; Raffaello Trentadue[8]

[1] *CERN*

[2] *CERN/University of the West of England*

[3] *SLAC National Accelerator Laboratory (US)*

[4] *Fermi National Accelerator Lab. (US)*

[5] *Weizmann Institute of Science (IL)*

[6] *Brookhaven National Laboratory (US)*

[7] *Institut fur Physik-Johannes-Gutenberg-Universitaet-Unknown*

[8] *Universita e INFN (IT)*

Corresponding Authors: raffaello.trentadue@cern.ch, andrea.valassi@cern.ch

The LCG Persistency Framework consists of three software packages (POOL, CORAL and COOL) that address the data access requirements of the LHC experiments in several different areas. The project is the result of the collaboration between the CERN IT Department and the three experiments (ATLAS, CMS and LHCb) that are using some or all of the Persistency Framework components to access their data. The POOL package is a hybrid technology store for C++ objects, using a mixture of streaming and relational technologies to implement both object persistency and object metadata catalogs and collections. POOL provides generic components that can be used by the experiments to store both their event data and their conditions data. The CORAL package is an abstraction layer with an SQL-free API for accessing data stored using relational database technologies. It is used directly by experiment-specific applications and internally by both COOL and POOL. The COOL

package provides specific software components and tools for the handling of the time variation and versioning of the experiment conditions data.

This presentation will report on the status and outlook in each of the three sub-projects at the time of CHEP2012. It will focus on COOL and POOL, as several new features of CORAL are the subject of other presentations at this conference.

Poster Session / **105**

Operational Experience with the ALICE High Level Trigger

Author: Artur Szostak[1]

[1] *University of Bergen (NO)*

Corresponding Author: artur.szostak@cern.ch

The ALICE High Level Trigger (HLT) is a dedicated real-time system for on-line event reconstruction and triggering. Its main goal is to reduce the large volume of raw data that is read out from the detector systems, up to 25 GB/s, by an order of magnitude to fit within the available data acquisition bandwidth. This is accomplished by a combination of data compression and triggering. When a reconstructed event is selected by the HLT trigger algorithms as interesting for physics then it is recorded, otherwise the raw data for that event is discarded. The combination of both approaches allows for flexible strategies for data reduction.

A second but equally vital function of the HLT is on-line monitoring. The HLT has access to all raw data and status information from the detectors during data taking. Combined with on-line event reconstruction the HLT becomes a powerful monitoring tool for ensuring data quality. Many problems can only be spotted easily when looking at the high level information on the physics level. In addition, on-line compression and triggering must be monitored live during data taking to ensure stability of the system and quality of recorded data.

A very high computational load is placed on the HLT to perform its tasks, in particular during event reconstruction and compression. A large dedicated computing cluster for on-line operations is used, which comprises 206 individual machines, 2744 CPU cores, 64 GPUs, 5.24 TB of distributed memory; all interconnected with an InfiniBand network and Gigabit Ethernet for management. There are an additional 43 machines which provide a development and testing environment, infrastructure support and storage.

Running a large complex system like the HLT in production data taking mode proves to be a challenge. During the 2010 pp and Pb-Pb running period many problems were experienced that lead to a sub-optimal operational efficiency. Lessons were learned and certain crucial changes were made early in 2011 to prepare for the 2011 Pb-Pb run, in which HLT would have a vital role performing data compression for the largest detector in ALICE, the Time Projection Chamber (TPC). Key changes such as separation of the production part of the system from the supporting infrastructure and upgrading to a mass storage system more suited to the HLT performance requirements has lead to higher stability, improved operational efficiency and reduction in startup latency of the system during runs.

A overview of the status of the HLT, experience from 2010 and 2011 production runs and important lessons learned are presented. Emphasis is given to the overall performance, showing a overall reduction in failure rates between 2010 and 2011, attributed to the significant improvements made to the system. Finally, further opportunities for improvement are identified and discussed, based on the experience gained in the 2011 Pb-Pb run.

Poster Session / **106**

Software design and implementation for the ATLAS Muon Cathode Strip Chamber ROD

Author: Raul Murillo Garcia[1]

Co-authors: Andrew James Lankford [1]; Andy Nelson [1]; James Panetta [2]; Jianrong Deng [1]; Leonid Sapozhnikov [2]; Michael Huffer [2]; Michael Schernau [1]; Richard Claus [2]; Ryan Herbst [2]

[1] *University of California Irvine (US)*

[2] *SLAC*

Corresponding Author: raul.murillo.garcia@cern.ch

The ATLAS Cathode Strip Chamber system consists of two end-caps with 16 chambers each. The CSC Readout Drivers (RODs) are purpose-built boards encapsulating 13 DSPs and around 40 FPGAs. The principal responsibility of each ROD is for the extraction of data from two chambers at a maximum trigger rate of 75 kHz. In addition, each ROD is in charge of the setup, control and monitoring of the on-detector electronics. This paper introduces the design and implementation of the CSC ROD firmware and software. The main features of this design include an event flow schema that decentralizes the different dataflow streams, which can thus operate asynchronously at its own natural rate; a ROD communication interface designed for high I/O throughput by minimizing the number of cycles necessary to move event data in and out of the DSPs; an event building mechanism that associates data transferred by the asynchronous streams but belongs to the same event; and a sparcification algorithm that discards uninteresting events and thus reduces the data occupancy volume, a crucial feature due to bandwidth limitations. The time constraints imposed by the high trigger rate have made paramount the use of optimization techniques such as the curiously recurrent template pattern and the programming of critical code in assembly language. The behaviour of the CSC RODs has been characterized in order to validate the performance of the software implementation.

Software Engineering, Data Stores and Databases / 107

Improving Software Quality of the ALICE Data-Acquisition System through Program Analysis

Author: Jianlin Zhu[1]

Co-authors: Daicui Zhou [1]; Guoping Zhang [2]; Jin Huang [3]; Sylvain Chapeland [4]

[1] *Huazhong Normal University (CN)*

[2] *Huazhong Normal University*

[3] *Huazhong University of Science and Technology*

[4] *CERN*

Corresponding Author: jianlin.zhu@cern.ch

The Data-Acquisition System designed by ALICE , which is the experiment dedicated to the study of strongly interacting matter and the quark-gluon plasma at the CERN LHC(Large Hadron Collider), handles the data flow from the sub-detector electronics to the archiving on tape. The software framework of the ALICE data-acquisition system is called DATE (ALICE Data Acquisition and Test Environment) and consists of a set of software packages grouped into main logic packages and utility packages.

In order to assess the software quality of DATE, and review possible improvements, we implement PAF (Program Analysis Framework) to analyze the software architecture and software modularity. The basic idea about PAF is recording the call relationships information among the important elements (i.e., functions, global variables, complex structures) firstly and then using the different analysis algorithms to find the Crosscutting Concerns which could destroy the modularity of the software from this recording information.

The PAF is based on the API of Eclipse C/C++ Development Tooling(CDT) because the source codes of DATE framework is written in C language. The CDT project based on the Eclipse platform provides a fully functional C and C++ Integrated Development Environment. The PAF for DATE could

also be used for the analysis of other projects written in C language.
Finally we evaluate our framework through analyzing the software system of DATE. The analysis result proves the effectiveness and efficiency of our framework. PAF has pinpointed a number of possible optimizations which could be applied to DATE and help maximizing the software quality.

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Poster Session / **108**

Evolution and performance of electron and photon triggers in ATLAS in the year 2011

Authors: Alessandro Tricoli[1]; Takanori Kono[2]

Co-author: Liam Duguid [3]

[1] *CERN*

[2] *Deutsches Elektronen-Synchrotron (DE)*

[3] *University of London (GB)*

Corresponding Authors: liam.duguid@cern.ch, takanori.kohno@cern.ch

The electron and photon triggers are among the most widely used triggers in ATLAS physics analyses. In 2011, the increasing luminosity and pile-up conditions demanded higher and higher thresholds and the use of tighter and tighter selections for the electron triggers. Optimizations were performed at all three levels of the ATLAS trigger system. At the high-level trigger (HLT), many variables from the calorimeters and tracking detectors are used to achieve high efficiency and large rejection power. The use of isolation criteria at the HLT has also been investigated. At L1, the thresholds were raised and optimised to account for η-dependence and hadronic isolation was implemented. In addition to physics triggers, dedicated triggers for collecting a large number of control samples of J/psi->ee, W->enu and jet background, for calibration, efficiency and fake rate measurements were developed. This contribution summarizes the algorithms and performance of ATLAS electron and photon triggers used in 2011 data taking.

Poster Session / **109**

Physics Data Processing with Google Protocol Buffers

Author: Johannes Ebke[1]

Co-author: Peter Waller [2]

[1] *Ludwig-Maximilians-Univ. Muenchen (DE)*

[2] *University of Liverpool*

Corresponding Author: johannes.ebke@physik.uni-muenchen.de

Historically, HEP event information for final analysis is stored in
Ntuples or ROOT Trees and processed using ROOT I/O, usually resulting in
a set of histograms or tables.

Here we present an alternative data processing framework, leveraging the Protocol Buffer open-source library, developed and used by Google Inc. for loosely coupled interprocess communication and serialization.

We save event information as a stream of Protocol Messages, which can be read and written using high-performance code generated by the Protocol Buffer software. No seeks are performed in write mode, and during processing, making easy deployment over streaming network connections possible.

The performance of our code on an example mock-physics analysis is then compared with a ROOT analysis on the same data, showing the gain obtained by leveraging current developments from outside HEP.

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Distributed Processing and Analysis on Grids and Clouds / 110

The "Common Solutions" Strategy of the Experiment Support group at CERN for the LHC Experiments

Author: Maria Girone[1]

Co-authors: Alessandro Di Girolamo [1]; Andrea Sciaba [1]; Andrea Valassi [1]; Daniel Colin Van Der Ster [1]; Daniele Spiga [1]; David Kingsley Tuckett ; Domenico Giordano [1]; Edward Karavakis [1]; Elisa Lanciotti [1]; Fernando Harald Barreiro Megino [2]; Guidone Negri [1]; Jamie Shiers [1]; Julia Andreeva [1]; Lukasz Kokoszkiewicz [1]; Maria Dimou [1]; Maria Dolores Saiz Santos [3]; Mattia Cinquilli [4]; Michael John Kenyon [1]; Nicolo Magini [1]; Pablo Saiz [1]; Raffaello Trentadue [5]; Simone Campana [1]; Stefan Roiser [1]

[1] *CERN*

[2] *Universidad Autonoma de Madrid (ES)*

[3] *Conseil Europeen Recherche Nucl. (CERN)*

[4] *Univ. of California San Diego (US)*

[5] *Universita e INFN (IT)*

Corresponding Author: maria.girone@cern.ch

After two years of LHC data taking, processing and analysis and with numerous changes in computing technology, a number of aspects of the experiments'computing as well as WLCG deployment and operations need to evolve. As part of the activities of the Experiment Support group in CERN' s IT department, and reinforced by effort from the EGI-InSPIRE project, we present work aimed at common solutions across all LHC experiments. Such solutions allow us not only to optimize development manpower but also offer lower long-term maintenance and support costs. The main areas cover Distributed Data Management, Data Analysis, Monitoring and the LCG Persistency Framework. Specific tools have been developed including the HammerCloud framework, automated services for data placement, data cleaning and data integrity (such as the data popularity service for CMS, the common Victor cleaning agent for ATLAS and CMS and tools for catalogue/storage consistency), the Dashboard Monitoring framework (job monitoring, data management monitoring, File Transfer monitoring) and the Site Status Board. This talk focuses primarily on the strategic aspects of providing such common solutions and how this relates to the overall goals of long-term sustainability and the relationship to the various WLCG Technical Evolution Groups

Summary:

Common Solutions for the LHC experiments provided by the CERN Experiment Support group of the IT department

Poster Session / 111

Status and evolution of CASTOR (Cern Advanced STORage)

Author: Sebastien Ponce[1]

Co-authors: Dennis Waldron [1]; Elvin Alin Sindrilaru [1]; Eric Cano [1]; Giuseppe Lo Presti [1]; Ignacio Reguero [1]; Jan Iven [1]; John Hefferman [1]; Massimo Lamanna [1]; Reece Madison [1]; Stefano Alberto Russo [1]; Steven Murray [1]

[1] *CERN*

Corresponding Author: sebastien.ponce@cern.ch

This is an update on CASTOR (CERN Advanced Storage) describing the recent evolution and related experience in production during the latest high-intensity LHC runs.
In order to handle the increasing data rates (10GB/s average for 2011), several major improvements have been introduced.
We describe in particular the new scheduling system that has replaced the original CASTOR one. It removed the limitations ATLAS and CMS were hitting in terms of file openings rates (from 20 Hz to 200+ Hz) while simplifying the code and operations at the same time.
We detail how the usage of the internal database has been optimized to improve efficiency by a factor 3 and cut opening file latency by orders of magnitude (from O(1s) to O(1ms)).
Finally, we will report on the evolution of the CASTOR monitoring and give the roadmap for the future.

Poster Session / 112

Flexible event reconstruction software chains with the ALICE High-Level Trigger

Author: Dinesh Ram[1]

[1] *Johann-Wolfgang-Goethe Univ. (DE)*

Corresponding Author: dinesh.ram@cern.ch

The ALICE High-Level Trigger (HLT) is a complex real-time system, whose primary objective is to scale down the data volume read out by the ALICE detectors to at most 4 GB/sec before being written to permanent storage. This can be achieved by using a combination of event filtering, selection of the physics regions of interest and data compression, based on detailed on-line event reconstruction. ALICE's largest detector - the Time Projection Chamber (TPC) - alone can easily reach data rates of upto 15 GB/sec which exceeds the available mass-storage bandwidth. Hence the ALICE HLT is a critical system logically sitting in between the detector readout electronics and the DAQ event building network.

The ALICE HLT has a large high-performance computing cluster at CERN consisting of 2752 CPU cores supported by 64 GPUs and 246 FPGAs. Data-flow in this cluster is controlled by a custom designed software framework. It consists of a set of components which can communicate with each other via a common control interface. The software framework also supports the creation of different configurations based on the detectors participating in the HLT. These configurations define a logical data processing "chain"of detector data-analysis components. Readout data passes through these software components in a pipelined fashion so that several events are processed in the software chain at the same time. An instance of such a chain can run and manage a few thousand physics analysis and data-flow components.

As more detectors participate in the HLT and with the increasing data challenges posed by ALICE, from the computing point of view, it translates into a need to efficiently manage an even higher number of software components communicating with each other and competing for the same resources in the cluster.

In this contribution the experience of running the HLT software and the configuration scheme used in 2011 –with special emphasis on the heavy ion period of ALICE - will be discussed. The current status of the software would be presented and the improvements made, based on past experience of running the software would be reviewed.

- Dinesh Ram for the ALICE HLT Collaboration

Poster Session / 113

A new communication framework for the ALICE Grid

Authors: Alina Gabriela Grigoras[1]; Costin Grigoras[1]; Steffen Schreiner[2]

Co-authors: Federico Carminati [1]; Latchezar Betev [1]; Pablo Saiz [1]

[1] *CERN*

[2] *Technische Universitaet Darmstadt (DE)*

Corresponding Author: costin.grigoras@cern.ch

Since the ALICE experiment began data taking in late 2009, the amount of end user jobs on the AliEn Grid has increased significantly. Presently 1/3 of the 30K CPU cores available to ALICE are occupied by jobs submitted by about 400 distinct users. The overall stability of the AliEn middleware has been excellent throughout the 2 years of running, but the massive amount of end-user analysis and its specific requirements and load has revealed few components which can be improved. One of them is the interface between users and central AliEn services (catalogue, job submission system) which we are currently re-implementing in Java. The interface provides persistent connection with enhanced data and job submission authenticity. In this paper we will describe the architecture of the new interface, the ROOT binding which enables the use of a single interface in addition to the standard UNIX-like access shell and the new security-related features.

Collaborative tools / 114

A New Information Architecture, Web Site and Services for the CMS Experiment

Author: Lucas Taylor[1]

Co-authors: Eleanor Rusack [1]; Vidmantas Zemleris [2]

[1] *Fermi National Accelerator Lab. (US)*

[2] *Faculty of Mathematics and Informatics-Vilnius University*

Corresponding Authors: lucas.taylor@cern.ch, eleanor.m.rusack@cern.ch

The age and size of the CMS collaboration at the LHC means it now has many hundreds of inhomogeneous web sites and services and more than 100,000 documents.

We describe a major initiative to create a single coherent CMS internal and public web site. This uses the Drupal web Content Management System (now supported by CERN/IT) on top of a standard LAMP stack (Linux, Apache, MySQL, and php/perl). The new navigation, content and search services

are coherently integrated with numerous existing CERN services (CDS, EDMS, Indico, phonebook, Twiki) as well as many CMS internal Web services.

We describe the information architecture; the system design, implementation and monitoring; the document and content database; security aspects; and our deployment strategy which ensured continual smooth operation of all systems at all times.

Poster Session / **115**

Track and Vertex Reconstruction Strategies in the ATLAS Inner Detector in the High Multiplicity LHC Environment

Authors: Heather Gray[1]; Simone Pagan Griso[2]

Co-author: Christoph Wasicki [3]

[1] *CERN*

[2] *Lawrence Berkeley National Lab. (US)*

[3] *Deutsches Elektronen-Synchrotron (DE)*

Corresponding Authors: heather.gray@cern.ch, simone.pagan.griso@cern.ch, christoph.wasicki@cern.ch

The track and vertex reconstruction algorithms of the ATLAS Inner Detector have demonstrated excellent performance in the early data from the LHC. However, the rapidly increasing number of interactions per bunch crossing introduces new challenges both in computational aspects and physics performance. We will discuss the strategy adopted by ATLAS in response to this increasing multiplicity by balancing physics requirements with the available computing resources. In addition the performance of the track and vertex reconstruction algorithms in this challenging environment will be demonstrated.

Collaborative tools / **117**

Talking Physics: Can Social Media Teach HEP to Converse Again?

Author: Steven Goldfarb[1]

[1] *University of Michigan (US)*

Corresponding Author: steven.goldfarb@cern.ch

Og, commonly recognized as one of the earliest contributors to experimental particle physics, began his career by smashing two rocks together, then turning to his friend Zog and stating those famous words "oogh oogh". It was not the rock-smashing that marked HEP's origins, but rather the sharing of information, which then allowed Zog to confirm the important discovery, that rocks are indeed made of smaller rocks.

Over the years, Socrates and other great teachers developed the methodology of this practice. Yet, as small groups of friends morphed into large classrooms of students, readers of journals, and audiences of television viewers, science conversation evolved into lecturing and broadcasting. While information is still conveyed in this manner, the invaluable, iterative nature of question/response is often lost or limited in duration.

The birth of Web 2.0 and the development of Social Media tools, such as Facebook, Twitter and Google +, are allowing iterative conversation to reappear in nearly every aspect of communication. From comments on public articles and publications to "wall"conversations and tweets, physicists are finding themselves interacting with the public before, during and after publication. I discuss both

the danger and the powerful potential of this phenomenon, and present methods currently used in HEP to make the best of it.

Poster Session / 118

ATLAS Virtual Visits: Bringing the World into the ATLAS Control Room

Author: Steven Goldfarb[1]

[1] *University of Michigan (US)*

Corresponding Author: steven.goldfarb@cern.ch

The newfound ability of Social Media to transform public communication back to a conversational nature provides HEP with a powerful tool for Outreach and Communication. By far, the most effective component of nearly any visit or public event is that fact that the students, teachers, media, and members of the public have a chance to meet and converse with real scientists.

While more than 30,000 visitors passed through the ATLAS Visitor Centre in 2011, nearly 7 billion did not have a chance to make the trip. Clearly this is not for lack of interest. Rather, the costs of travel, in terms of time and money, and limited parking, put that number somewhat out of reach. On the other hand, during the LHC "First Physics"event of 2010, more than 2 million visitors joined the experiment control rooms via webcast for the celebration.

I present a project developed for the ATLAS Experiment's Outreach and Education program that complements the webcast infrastructure with video conferencing and wireless sound systems, allowing the public to interact with hosts in the control room with minimal disturbance to the shifters. These "Virtual Visits"have included high school classes, LHC Masterclasses, conferences, expositions and other events in Europe, USA, Japan and Australia, to name a few. I will discuss the technology used, potential pitfalls (and ways to avoid them), and our plans for the future.

Poster Session / 119

ALICE Grid Computing at the GridKa Tier-1 Center

Author: Christopher Jung[1]

Co-authors: Andreas Petzold [2]; Christoph-Erdmann Pfeiler [2]; Kilian Schwarz [3]

[1] *KIT - Karlsruhe Institute of Technology (DE)*

[2] *KIT*

[3] *GSI - Helmholtzzentrum fur Schwerionenforschung GmbH (DE)*

Corresponding Author: christopher.jung@kit.edu

The GridKa center at the Karlsruhe Institute for Technology is the largest ALICE Tier-1 center. It hosts 40,000 HEPSEPC'06, approximately 2.75 PB of disk space and 5.25 PB of tape space for for A Large Ion Collider Experiment (ALICE), at the CERN LHC. These resources are accessed via the AliEn middleware. The storage is divided into two instances, both using the storage middleware xrootd.
We will focus on the set-up of these resources and on the topic of monitoring. The latter serves a vast number of purposes, ranging from efficiency statistics for process and procedure optimization to alerts for on-call duty engineers.

Poster Session / 120

Neural network based cluster creation in the ATLAS silicon pixel detector

Authors: Andreas Salzburger[1]; Giacinto Piacquadio[1]

[1] *CERN*

Corresponding Authors: andreas.salzburger@cern.ch, giacinto.piacquadio@cern.ch

The read-out from individual pixels on planar semi-conductor sensors are grouped into clusters to reconstruct
the location where a charged particle passed through the sensor. The resolution given by individual pixel sizes
is significantly improved by using the information from the charge sharing between pixels.
Such analog cluster creation techniques have been used by the ATLAS experiment for many years to obtain an excellent performance.
However, in dense environments, such as those inside high-energy jets, clusters have an increased probability of merging
the charge deposited by multiple particles.
Recently, a neural network based algorithm which estimates both the cluster position and whether a cluster should be split
into sub-cluster has been developed for the ATLAS pixel detector. The algorithm significantly reduces ambiguities
in the assignment of pixel detector measurement to tracks within jets and improves the position accuracy with respect
to standard interpolation techniques by taking into account the 2-dimensional charge distribution.
The implementation of the neural network, the training parameters and performance of the new clustering will be presented.
Significant improvements to the track and vertex resolution obtained using this new method will be presented
based on Monte Carlo simulated data and the results will be compared to data recorded with the ATLAS detector.
Finally, the resulting improvements to the identification of jets containing b-quarks will be discussed.

Poster Session / 121

Service management at CERN with Service-Now

Author: Zhechka Toteva[1]

[1] *CERN*

Corresponding Author: zhechka.toteva@cern.ch

The Information Technology (IT) and the General Services (GS) departments at CERN have decided to combine their extensive experience in support for IT and non-IT services towards a common goal –to bring the services closer to the end user based on ITIL best practice. The collaborative efforts have so far produced definitions for the incident and the request fulfillment processes which are based on a unique two-dimensional service catalogue that combines both the user and the support team view of all services.

After an extensive evaluation of the available industrial solutions, Service-now was selected as the tool to implement the CERN Service-Management processes. The initial release of the tool made provided an attractive web portal for the users and successfully implemented two basic ITIL processes; the incident management and the request fulfilment processes. It also integrated with the CERN personnel databases and the LHC GRID ticketing system.

Subsequent releases continued to integrate with other third-party tools like the facility management systems of CERN as well as to implement new processes such as change management. Independently from those new development activities it was decided to simplify the request fulfillment process in order to achieve easier acceptance by the CERN user community.

We believe that due to the high modularity of the Service-now tool, the parallel design of ITIL processes e.g., event management and non-ITIL processes, e.g., computer centre hardware management, will be easily achieved.

This presentation will describe the experience that we have acquired and the techniques that were followed to achieve the CERN customization of the Service-Now tool.

Poster Session / **122**

Track Based Alignment of the ATLAS Inner Detector: Implementation and Performance

Author: Anthony Morley[1]

Co-author: Salvador Marti I Garcia [2]

[1] *CERN*

[2] *IFIC-Valencia (UV/EG-CSIC)*

Corresponding Authors: anthony.morley@cern.ch, martis@ific.uv.es

The Large Hadron Collider (LHC) at CERN is the world's largest particle accelerator, which collides proton beams at an unprecedented centre of mass energy of 7 TeV.

ATLAS is a multipurpose experiment that records the products of the LHC collisions. In order to reconstruct the trajectories of charged particles produced in these collisions,
ATLAS is equipped with a tracking system (Inner Detector) built on two different technologies: silicon planar (pixel and microstrip) sensors and drift-tube based detectors.
The goal of the Inner Detector alignment is to determine accurately the position and orientation of its sensors with a precision better than 10 micrometers,
such the tracker performance is not degraded far beyond its intrinsic resolution. This requires the determination of over 700,000 degrees of freedom (DoF) with high accuracy.

The implementation of the track based alignment within the ATLAS software framework unifies different alignment approaches and allows the alignment of all tracking subsystems together.
The alignment specific classes are directly linked with the track reconstruction software, which provides tools for computation of specific quantities (residuals, pulls, track derivatives,
covariance matrices, ...). The detector specific classes inherit from a common base allowing for a unified definition of the alignment geometry.

As the alignment algorithms are based on minimization of the track-hit residuals,
one has to solve a linear system with large number of DoF. The solving itself poses a real challenge as it involves inversion or diagonalization of a large matrix that may be dense.
Fast solving algorithms as well as full diagonalization have been implemented to calculate the results for the alignment.
The alignment software also has the ability to constrain the system either with constraints on the tracks
(beam spot, primary vertex, momentum from the ATLAS muon system, E/p, ...) or constraints on the alignment corrections.

The alignment is executed on a run by run basis at the ATLAS calibration loop using the CERN Analysis Facility with ˜200 CPUs running Scientific Linux CERN 5.
For these purposes, two independent data streams are selected online by the event filter (at a 50 Hz rate).
The first one consists of a collection of high momentum and isolated tracks.
The second is a set of cosmic-ray tracks triggered during the LHC empty bunches.

Detailed alignment runs on the GRID, where data corresponding to many data periods is analysed, allowing for thousands of CPU to be utilised simultaneously.
We will present an outline of the track based alignment approaches,
their implementation within the ATLAS software framework and their performance when aligning the ATLAS detector.

Summary:

The alignment of the Inner Detector of ATLAS poses a real computing challenge. There are more than 700 thousand degrees of freedom to align with high accuracy. Therefore many millions of tracks are needed. Besides the computing resources are a key ingredient for the alignment procedure. The ATLAS ID alignment has been adapted to run on the GRID. Several thousand jobs are submitted in parallel to the ATLAS tier 2 centers. There the data is processed and the ouptut is collected in form of alignment matrices and vectors, plus monitoring histograms. The whole alignment procedure is run iteratively till convergence is found. En each iteration, before solving the alignment, all the matrices and vector of individual jobs have to be added together. In order to obtain the alignment corrections one needs to solve a linear system with many thousands degrees of freedom. Thus the computing represents also a challenge. Fast matrix inversion with matrix conditioning and diagonalization of the full matrix are used. In the second case and in order to obtain sensible alignment corrections, on has to identify the singular and the near-singular modes of the alignment.
The whole alignment chain can also be run quasi-online in the calibration loop, where alignment constants are required to be derived run by run. There is also a limit of 36 hours to obtain the corrections, prior the bulk processing starts.

Poster Session / **124**

Bug Tracking in Open Source and High Energy Physics Software - A Comparative Study

Authors: Benedikt Hegner[1]; Hoda Khalafalla[2]

[1] *CERN*

[2] *Max-Planck-Gesellschaft (DE)*

Corresponding Author: benedikt.hegner@cern.ch

Bug tracking is a process which comprises activities of reporting, documenting, reviewing, planning, and fixing software bugs. While there exist many studies on the usage of bug tracking tools and procedures in open source software, the situation in high energy physics has never been looked at in a systematic way. In our study we have compared and analyzed several scientific and non-scientific software projects to define the similarity and variability in bug-tracking practices. We will present our findings, with emphasis on a comparison of the three projects ATLAS, Belle II and Eclipse. In addition, we aim at defining the problems and the specific needs of the development paradigm in high energy physics.

Poster Session / **125**

The LCG/AA integration build system

Authors: Alex Liam James Hodgkins[1]; Benedikt Hegner[2]; Victor Diez Gonzalez[3]

[1] *Loughborough University of Tech.*

[2] *CERN*

[3] *CERN fellow*

Corresponding Author: victor.diez.gonzalez@cern.ch

The LCG Applications Area relies on regular integration testing of the provided software stack. In the past, regular builds have been provided using a system which has been changed and developed constantly adding new features like server-client communication, long-term history of results and a summary web interface using present-day web technologies.

However, the ad-hoc style of software development resulted in a setup that is hard to monitor, inflexible and difficult to expand.

The new version of the infrastructure is based on the Django Python framework, which allows for a structured and modular design, making it easy to plug in later additions. Transparency in the workflows and ease of monitoring has been one of the priorities in the design. Formerly missing functionality like e.g. on-demand builds or release triggering will support the transition to a more agile development process.

Poster Session / 126

Managing operational documentation in the ALICE Detector Control System

Author: Mateusz Lechman[1]

Co-authors: Alexander Kurepin [2]; Andre Augustinus [1]; Ombretta Pinazza [3]; Peter Chochula [1]; Peter Matthew Bond [4]; Peter Rosinsky [5]

[1] *CERN*

[2] *Moscow Physical Engineering Institute (MePhl)*

[3] *Universita e INFN (IT)*

[4] *University of the West of England*

[5] *Department of Nuclear Physics-Comenius University*

Corresponding Author: mateusz.lechman@cern.ch

ALICE (A Large Ion Collider Experiment) is one of the big LHC (Large Hadron Collider) experiments at CERN in Geneve, Switzerland.

The experiment is composed of 18 sub-detectors controlled by an integrated Detector Control System (DCS) that is implemented using the commercial SCADA package PVSS. The DCS includes over 1200 network devices, over 1,000,000 input channels and numerous custom made software components that are prepared by over 100 developers from all around the world.

This complex system is controlled by a single operator via a central user interface. One of his/her main tasks is recovery of anomalies and errors that may appear during the operation. Therefore, clear, complete and easily accessible documentation is essential to guide the shifter through the expert interfaces of different subsystems.

This paper describes the idea of managing of the operational documentation in ALICE using a generic repository that is based on relational database and is integrated with the control system. The experience gained and the conclusions drawn from the project are also presented.

Distributed Processing and Analysis on Grids and Clouds / 127

The LHCb Data Management System

Author: Philippe Charpentier[1]

Co-author: Computing Group LHCb [2]

[1] *CERN*

[2] *LHCb*

Corresponding Author: philippe.charpentier@cern.ch

The LHCb Data Management System is based on the DIRAC Grid Community Solution. LHCbDirac provides extensions to the basic DMS such as a Bookkeeping System. Datasets are defined as sets of files corresponding to a given query in the Bookkeeping system. Datasets can be manipulated by CLI tools as well as by automatic transformations (removal, replication, processing). A dynamic handling of dataset replication is performed, based on disk space usage at the sites and dataset popularity. For custodial storage, an on-demand recall of files from tape is performed, driven by the requests of the jobs, including disk cache handling.
We shall describe all the tools that are available for Data Management, from handling of large datasets to basic tools for users as well as for monitoring the dynamic behaviour of LHCb Storage capacity.

Online Computing / 128

Applications of advanced data analysis and expert system technologies in ATLAS Trigger-DAQ Controls framework

Author: Andrei Kazarov[1]

Co-authors: Alina Corso Radu [2]; Giovanna Lehmann Miotto [3]; Giuseppe Avolio [2]; Luca Magnoni [3]

[1] *B.P. Konstantinov Petersburg Nuclear Physics Institute - PNPI (*

[2] *University of California Irvine (US)*

[3] *CERN*

Corresponding Authors: giuseppe.avolio@cern.ch, andrei.kazarov@cern.ch

The Trigger and DAQ (TDAQ) system of the ATLAS experiment is a very complex distributed computing system, composed of O(10000) of applications running on more than 2000 computers. The TDAQ Controls system has to guarantee the smooth and synchronous operations of all TDAQ components and has to provide the means to minimize the downtime of the system caused by runtime failures, which are inevitable for a system of such scale and complexity.

During data taking runs, streams of information messages sent or published by TDAQ applications are the main sources of knowledge about correctness of running operations. The huge flow of operational monitoring data produced (with an average rate of O(1-10KHz)) is constantly monitored by experts to detect problem or misbehavior.

Given the scale of the system and the rates of data to be analyzed, the automation of the Control system functionality in areas of operational monitoring, system verification, error detection and recovery is a strong requirement. It allows to reduce the operations man power needs and to assure a constant high quality of problem detection and following recovery.

To accomplish its objective, the Controls system includes some high-level components which are based on advanced software technologies, namely the rule-based expert system (ES) and the complex event processing (CEP) engines. The chosen techniques allow to formalize, to store and to reuse the TDAQ experts' knowledge in the

Control framework and thus to assist TDAQ shift crew to accomplish its task.

DVS (Diagnostics and Verification System) and Online Recovery components are responsible for the automation of system testing and verification, diagnostics of failures and recovery procedures. These components are built on top of a common technology of a forward-chaining ES framework (based on CLIPS expert system shell), that allows to program the behavior of a system in terms of "if-then" rules and to easily extend or modify the knowledge base.

The core of AAL (Automated monitoring and AnaLysis) component is a CEP (Complex Event Processing) engine implemented using ESPER in Java. The engine is loaded with a set of directives and it performs correlation and analysis of operational messages and events and produces operator-friendly alerts, assisting TDAQ operators to react promptly in case of problems or to perform important routine tasks. The component is known to shifters as "Shifter Assistant" (SA), and introduction of the SA allowed to reduce the number of shifters in the ATLAS control room. Design foresees a machine learning module to detect anomaly and problems that cannot be defined in advance.

The described components are constantly used for the ATLAS Trigger-DAQ system operations, and the knowledge base is growing as more expertise is acquired. By the end of 2011 the size of the knowledge base used for TDAQ operations was about 300 rules.

The paper presents the design and present implementation of the components and also the experience of its use in a real operational environment of the ATLAS experiment.

Summary:

The paper presents the design and implementation of some intelligent expert system based TDAQ Controls components and also the experience of their use in a real operational environment of the ATLAS experiment.

Software Engineering, Data Stores and Databases / 129

Advanced Modular Software Performance Monitoring

Author: Alexander Mazurov[1]

[1] *Universita di Ferrara (IT)*

Corresponding Author: alexander.mazurov@cern.ch

The LHCb software is based on the Gaudi framework, on top of which are built several large and complex software applications. The LHCb experiment is now in the active phase of collecting and analyzing data and significant performance problems arise in the Gaudi based software beginning from High Level Trigger (HLT) programs and ending with data analysis frameworks (DaVinci). It' s not easy to find hot spots in the code - only special tools can help to understand where CPU or memory usage is not reasonable. There exist many performance analyzing tools, but the main problem is that they show reports in terms of class and function names and such information usually is not very useful - the majority of algorithm developers use the Gaudi framework abstractions and usually do not know about functions which lie at the lower level. We will show a new approach which adds to performance reports a higher abstraction level based on knowledge of framework architecture and run-time object properties. A set of profiling tools (based on sampling and unwind library - a software for introspection of the program call-chain) and visualizing interfaces has been developed and deployed.

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Poster Session / 130

Application of the DIRAC framework in CTA: first evaluation

Author: Luisa Arrabito[1]

Co-authors: Bruno Khelifi [2]; Cecile Barbier [3]; Claudia Lavalley [4]; George Vasileiadis [4]; Giovanni Lamanna [3]; Jean Philippe Lenain [5]; Nukri Komin [3]; Ricardo Graciani Diaz [6]

[1] *IN2P3/LUPM on behalf of the CTA Consortium*

[2] *IN2P3/LLR, CTA Consortium*

[3] *IN2P3/LAPP, CTA Consortium*

[4] *IN2P3/LUPM, CTA Consortium*

[5] *INSU/Observatoire de Paris, CTA Consortium*

[6] *University of Barcelona, CTA Consortium*

Corresponding Author: arrabito@in2p3.fr

The Cherenkov Telescope Array (CTA) –an array of many tens of Imaging Atmospheric Cherenkov Telescopes deployed on an unprecedented scale –is the next generation instrument in the field of very high energy gamma-ray astronomy.
CTA will operate as an open observatory providing data products to the scientific community. An average data stream of some GB/s for about 1000 hours of observation per year, thus producing several PB/year, is expected. Large CPU time is required for data processing as well as for massive Monte Carlo simulations (MC) needed for detector calibration purposes and performance studies as a function of detectors and lay-out configurations.
Given these large storage and computing requirements, the Grid approach is well suited and massive MC simulations are already running on the EGI Grid.
In order to optimize resource usage and to handle in a coherent way all production and future analysis activities, a high level framework with advanced functionalities is needed.
For this purpose the DIRAC framework for distributed computing access implementing CTA workloads is evaluated. The benchmark test results of DIRAC as well as the extensions developed to cope with CTA specific needs are presented.

Distributed Processing and Analysis on Grids and Clouds / 131

End-To-End Solution for Integrated Workload and Data Management using glideinWMS and Globus Online

Author: Parag Mhashilkar[1]

Co-authors: Burt Holzman [2]; Cathrin Weiss [3]; Gabriele Garzoglio [4]; Lukasz Lacinski [5]; Raj Kettimuthu [6]; Xi Duan [7]; Zach Miller [3]

[1] *Fermi National Accelerator Laboratory*

[2] *Fermi National Accelerator Lab. (US)*

[3] *UW Madison*

[4] *FERMI NATIONAL ACCELERATOR LABORATORY*

[5] *University of Chicago*

[6] *Argonne National Laboratory*

[7] *Illinois Institute of Technology*

Corresponding Author: parag@fnal.gov

Grid computing has enabled scientific communities to effectively share computing resources distributed over many independent sites. Several such communities, or Virtual Organizations (VO), in the Open Science Grid and the European Grid Infrastructure use the glideinWMS system to run complex application work-flows. GlideinWMS is a pilot-based workload management system (WMS) that creates on demand, dynamically-sized overlay Condor batch system on Grid resources. While the WMS addresses the management of compute resources, however, data management in the Grid is still the responsibility of the VO. In general, large VOs have resources to develop complex custom solutions, while small VOs would rather push this responsibility to the infrastructure. The latter requires a tight integration of the WMS and the data management layers, an approach still not common in modern Grids. In this paper we describe a solution developed to address this shortcoming in the context of Center for Enabling Distributed Petascale Science (CEDPS) by integrating glideinWMS with Globus Online (GO). GO is a fast, reliable file transfer service that makes it easy for any user to move data. The solution eliminates the need for the users to provide custom data transfer solutions in the application by making this functionality part of the glideinWMS infrastructure. To achieve this, glideinWMS uses the file transfer plug-in architecture of Condor. The paper describes the system architecture and how this solution can be extended to support data transfer services other than GO when used with Condor or glideinWMS.

Online Computing / 132

The operational performance of the ATLAS trigger and data acquisition system and its possible evolution

Author: Andrea Negri[1]

Co-authors: Daniel Whiteson [2]; Ning Zhou [2]; Per Werner [3]; Reiner Hauser [4]; Werner Wiedenmann [5]

[1] *Universita e INFN (IT)*

[2] *University of California Irvine (US)*

[3] *CERN*

[4] *Michigan State University (US)*

[5] *University of Wisconsin (US)*

Corresponding Author: andrea.negri@pv.infn.it

The ATLAS experiment at the Large Hadron Collider at CERN relies on a complex and highly distributed Trigger and Data Acquisition (TDAQ) system to gather and select particle collision data at unprecedented energy and rates. The TDAQ is composed of three levels which reduces the event rate from the design bunch-crossing rate of 40 MHz to an average event recording rate of about 200 Hz.
The first part of this presentation will give an overview of the operational performance of the DAQ system during 2011 and the first months of data taking in 2012. It will describe how the flexibility inherent in the design of the system has be exploited to meet the changing needs of ATLAS data taking and in some cases push performance beyond the original design performance specification.

The experience accumulated in the ATLAS DAQ/HLT system operation during these years stimulated also interest to explore possible evolutions, despite the success of the current design. One attractive direction is to merge three systems - the second trigger level (L2), the Event Builder (EB), and the Event Filter (EF) - within a single homogeneous one in which each HLT node executes all the steps required by the trigger and data acquisition process. Each L1 event is assigned to an available HLT node which executes the L2 algorithms using a subset of the event data and, upon positive selection, builds the event, which is further processed by the EF algorithms. Appealing aspects of this design are: a simplification of the software architecture and of its configuration, a better exploitation of the computing resources, the caching of fragments already collected for L2 processing,

the automated load balancing between L2 and EF selection steps, the sharing of code and services on HLT nodes.
Furthermore, the full treatment of the HLT selection on a single node allows more flexible approaches, for example “incremental event building” in which trigger algorithms progressively enlarge the size of the analysed region of interest, before requiring the building of the complete event.
To spot possible limitations of the new approach and to demonstrate the benefits out-lined above, a prototype has been implemented. The preliminary measurements are positive and further tests are scheduled for the next months.

Appealing aspects of this design are: a simplification of the software architecture and of its configuration, a better exploitation of the computing resources, the caching of fragments already collected for L2 processing, the automated load balancing between L2 and EF selection steps, the sharing of code and services on HLT nodes.

Furthermore, the full treatment of the HLT selection on a single node allows more flexible approaches, for example “incremental event building” in which trigger algorithms progressively enlarge the size of the analysed region of interest, before requiring the building of the complete event.

To spot possible limitations of the new approach and to demonstrate the benefits out-lined above, a prototype has been implemented. The preliminary measurements are positive and further tests are scheduled for the next months. Their results are the subject of this paper.

Distributed Processing and Analysis on Grids and Clouds / 133

Deployment of Multifactor Authentication for Critical Services at CERN

Author: Stefan Lueders[1]

Co-authors: Remi Mollon [1]; Romain Wartel [1]

[1] *CERN*

Corresponding Author: stefan.lueders@cern.ch

Access protection is one of the cornerstones of security. The rule of least-privilege demands that any access to computer resources like computing services or web applications is restricted in such a way that only users with a need-to can access those resources. Usually this is done when authenticating the user asking her for something she knows, e.g. a (public) username and secret password. Unfortunately, passwords are regularly lost to attackers: Because of ignorance, users voluntarily reply to so-called Phishing emails that are specially crafted to steal passwords; attackers repeatedly succeeded to intercept passwords that are typed into compromised PCs···Adding a second factor to the authentication process, something the user is, like employing iris-scans, or has, like a hardware token, will prevent that the attacker can do any bad with the stolen password. He now also needs to get hold of the second factor.
In order to protect critical services and applications, the CERN Computer Security Team has evaluated several means for multi-factor authentication. Since there is no silver-bullet, three techniques have been selected: certificates stored in SmartChips embedded in the standard CERN access card, one-time passwords generated on USB sticks from Yubico (so-called yubi-Keys) and one-time passwords generated using mobile phone applications. This presentation will detail on the evaluation process, compare the different techniques, and outline the implementation and first experience in the field.

Student? Enter ’yes’. See http://goo.gl/MVv53:

no

Poster Session / 134

Managing Virtual Machine Lifecycle in CernVM Project

Author: Ioannis Charalampidis[1]

Co-authors: Artem Harutyunyan [2]; Dag Larsen [3]; Jakob Blomer [4]; Predrag Buncic [2]

[1] *Aristotle Univ. of Thessaloniki (GR)*

[2] *CERN*

[3] *University of Bergen (NO)*

[4] *Ludwig-Maximilians-Univ. Muenchen (DE)*

Corresponding Author: ioannis.charalampidis@cern.ch

The creation and maintenance of a Virtual Machine (VM) is a complex process. To build the VM image, thousands of software packages have to be collected, disk images suitable for different hypervisors have to be built, integrity tests must be performed, and eventually the resulting images have to become available for download. In the meanwhile, software updates for the older versions must be published, obsolete images must be revoked, and the clouds that use them must be updated. Initially, in the CernVM project we used several commercial solutions to drive this process. In addition to the cost, the drawback of such an approach was lack of a common and coherent framework that would allow for full control of every step in the process and easy adaptation to new technologies (hypervisors, clouds, APIs).

In an attempt to provide a complete lifecycle management solution for virtual machines, we collected a set of open-source tools, adapted them to our needs and combined them with our existing development tools in order to create an extensible framework that could serve as end to end solution for VM lifecycle management.

This new framework is based on the Archipel Open Source Project and shares some of its main principles, namely, every component of the system is a stand-alone agent; the front-end is a stand-alone application; all of them communicate over the same messaging network based on the Extensible Message and Presence Protocol (XMPP). Each component of the framework can thus interact with each other in order to perform automated tasks and all of them can be managed from a single User Interface. The agents that manage the Hypervisor infrastructure, as well as the agents that deploy and monitor the Virtual Machines and the web-based user interface, are provided by the Archipel Project. In CernVM, we developed iBuilder, a tool to instrument VM images for almost all popular hypervisors. We integrated Tapper, an open-source tool that tests the resulted images, and we developed all the appropriate agents to control the software repositories and the previously mentioned tools. All these agents now allow us to continuously build and test development images.

To complete the system, we plan to develop agents that will be capable of deploying contextualized CernVM images in various scenarios such as clouds that support the EC2 API Interface, or private/academic clouds using the native tools. Finally, a new lightweight front-end is under development aiming to provide access and complete control of the framework from the portable devices (smartphones and tablets). In this contribution we will present the details of this system, it's current status and future plans.

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Summary:

The maintenance of a Virtual Machine is a complex process that involves many software packages, many different phases and no standardized methodology.

For the CernVM project our first attempts to use several commercial solutions to drive this process failed as they were unable to provide a common and coherent framework that would allow for full control of every step. We therefore decided to combine existing open-source tools into an extensible framework that could serve as end-to-end solution for VM lifecycle management.

This new framework is based on the Archipel Project and shares some of its main principles. The whole system is stateless, there is no central server, and all the involved components plus the user interfaces are just inter-connected over the same messaging network. In this way, new components can be added on-the-fly, every component can communicate with each other to perform automated tasks, and all of them can be controlled from the same interface.

In this contribution we will present the architecture of this framework, our new tools that allow us to continuously build and test development images, some in-development highlights and our future plans.

Poster Session / 135

Long-term preservation of analysis software environment

Author: Dag Larsen[1]

Co-authors: Artem Harutyunyan [2]; Ioannis Charalampidis [3]; Jakob Blomer [4]; Predrag Buncic [2]

[1] *University of Bergen (NO)*

[2] *CERN*

[3] *Aristotle Univ. of Thessaloniki (GR)*

[4] *Ludwig-Maximilians-Univ. Muenchen (DE)*

Corresponding Authors: dag.larsen@cern.ch, artem.harutyunyan@cern.ch

Long-term preservation of scientific data represents a challenge to all experiments. Even after an experiment has reached its end of life, it may be necessary to reprocess the data. There are two aspects of long-term data preservation: "data" and "software". While data can be preserved by migration, it is more complicated for the software. Preserving source code and binaries is not enough; the full software and hardware environment is needed. Virtual machines (VMs) may offer a solution by "freezing" a virtual hardware platform "in software", where the legacy software can run in the original environment.

A complete infrastructure package is developed for easy deployment and management of such VMs. It is based on a dedicated distribution of Linux, CERNVM. Updated versions will be made available for new software, while older versions will still be available for legacy analysis software. Further, a HTTP-based file system, CVMFS, is used for the distribution of the software. Since multiple versions of both software and VMs are available, it is possible to process data with any software version, and a matching VM version. OpenNebula is used to deploy the VMs. Traditionally, there are many tools for managing clouds from a VM point-of-view. However, for experiments, it can be more useful to have a tool which is mainly centred around the data, but also allows for management of VMs. Therefore, a point-and-click web user interface is being developed that can (a) keep track of the processing status of all data; (b) select data to be processed and which type of processing, also selecting the version of software and matching VM; and (c) the configuration of the processing nodes, e.g. memory and number of nodes. It is preferable that the interface has an experiment-dependent module which will allow for easy adoption to various experiments. The complete package is designed to be easy to replicate on any processing site, and to scale well. Besides data preservation, this paradigm also allows for distributed cloud-computing on private and public clouds through the EC2 interface, for both legacy and contemporary experiments, e.g. NA61 and the LHC experiments.

Summary:

Long-term preservation of scientific data represents a challenge to experiments, especially with regard to the analysis software. Preserving source code and binaries is not enough; the full software and hardware environment is needed. Virtual machines (VMs) make it possible to preserve hardware "in software". A complete infrastructure package is developed for easy deployment and management of VMs, based on CERNVM Linux. Older CERNVM versions will still be available for legacy software. Further, a HTTP-based file system, CVMFS, is used for the distribution of the software. It is possible to process data with any software version, and a matching VM version. Most importantly, a point-and-click web user interface is being developed for setting up the complete processing chain, including VM/software versions, number/type of processing nodes, and the particular type of analysis and data. This paradigm

also allows for distributed cloud-computing on private and public clouds, for both legacy and contemporary experiments.

Poster Session / **136**

FermiGrid: High Availability Authentication, Authorization, and Job Submission.

Author: Steven Timm[1]

[1] *Fermilab*

Corresponding Author: timmsteve@yahoo.com

FermiGrid is the facility that provides the Fermilab Campus Grid with unified job submission, authentication, authorization and other ancillary services for the Fermilab scientific computing stakeholders.

We have completed a program of work to make these services resilient to high authorization request rates, as well as failures of building or network infrastructure.

We will present the techniques used, the response of the system against real world events and the performance metrics that have been achieved.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Summary:

Describes the tuning that was made to the FermiGrid SAZ and GUMS servers to use the XACML protocol more efficiently and to the clients to avoid timouts

Poster Session / **137**

FermiCloud - A Production Science Cloud for Fermilab

Author: Steven Timm[1]

[1] *Fermilab*

Corresponding Author: timmsteve@yahoo.com

FermiCloud is an Infrastructure-as-a-Service facility deployed at Fermilab based on OpenNebula that has been in production for more than a year. FermiCloud supports a variety of production services on virtual machines as well as hosting virtual machines that are used as development and integration platforms. This infrastructure has also been used as a testbed for commodity storage evaluations.

As part of the development work, an X.509 authentication plugins for OpenNebula were developed and deployed on FermiCloud. These X.509

plugins were contributed back to the OpenNebula project and were made generally available with the release of OpenNebula 3.0 in October 2011.

The FermiCloud physical infrastructure has recently been deployed across multiple physical buildings with the eventual goal of being resilient to a single building or network failure. Our current focus is the deployment of a distributed SAN with a shared and mirrored filesystem.

We will discuss the techniques being used and the progress to date as well as future plans for the project.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Summary:

Description of FermiCloud project

Poster Session / **138**

Comparison of the CPU efficiency of High Energy and Astrophysics applications on different multi-core processor types.

Author: Andreas Heiss[1]

[1] *KIT - Karlsruhe Institute of Technology (DE)*

Corresponding Author: andreas.heiss@kit.edu

GridKa, operated by the Steinbuch Centre for Computing at KIT, is the German regional centre for high energy and
astroparticle physics computing, supporting currently 10 experiments and serving as a Tier-1 centre for the four LHC
experiments. Since the beginning of the project in 2002, the total compute power is upgraded at least once per year to follow
the increasing demands of the experiments. The hardware is typically operated for about four years until it is replaced by
more modern machines. The GridKa compute farm thus consists of a mixture of several generations of compute nodes differing
in several parameters, e.g. CPU types, main memory, network connection bandwidth etc.
We compare the CPU efficiency (CPU time to wall time ratio) of high energy physics and astrophysics compute jobs on these different types of compute nodes and estimate the impact of the ongoing trend towards many-core CPUs.

Poster Session / **139**

Distributed error and alarm processing in the CMS data acquisition system

Author: Andrea Petrucci[1]

Co-authors: Alexander Flossdorf [2]; Andre Georg Holzner [3]; Andrei Cristian Spataru [1]; Attila Racz [1]; Aymeric Arnaud Dupont [1]; Christian Deldicque [1]; Christian Hartl [1]; Christoph Paus [4]; Christoph Schwick [1]; Dennis Shpakov [5]; Dominique Gigi [1]; Emilio Meschi [1]; Frank Glege [1]; Frans Meijers [1]; Gerry Bauer [4]; Giovanni Polese [1]; Hannes Sakulin [1]; James Branson [3]; Jeroen Hegeman [1]; Jose Antonio Coarasa Perez [1]; Konstanty Sumorok [4]; Lorenzo Masetti

[1]; Luciano Orsini [1]; Marc Dobson [1]; Marco Pieri [3]; Matteo Sani [3]; Matthew Bowen [6]; Michal Simon ; Olivier Raginel [4]; Remi Mommsen [5]; Robert Gomez-Reino Garrido [1]; Samim Erhan [7]; Sebastian Bukowiec [1]; Sergio Cittolin [3]; Ulf Behrens [8]; Vivian O'Dell [9]; Yi Ling Hwong [1]

[1] *CERN*

[2] *DESY*

[3] *Univ. of California San Diego (US)*

[4] *Massachusetts Inst. of Technology (US)*

[5] *Fermi National Accelerator Lab. (US)*

[6] *University of the West of England*

[7] *Univ. of California Los Angeles (US)*

[8] *Deutsches Elektronen-Synchrotron (DE)*

[9] *Fermi National Accelerator Laboratory (FNAL)*

Corresponding Author: andrea.petrucci@cern.ch

The Error and Alarm system for the data acquisition of the Compact Muon Solenoid (CMS) at CERN is successfully used for the physics runs at Large Hadron Collider (LHC) during the first three years of activities. Error and alarm processing entails the notification, collection, store and visualization of all exceptional conditions occurring in the highly distributed CMS online system using a uniform scheme. Alerts and reports are shown on-line by web application facilities that map them to graphical models of the system as defined by the user. A persistency service keeps history of all exceptions occurred, allowing subsequent retrieval of user defined time windows of events for later playback or analysis. The paper describes the architecture and the technologies used and deals with operational aspects during the first years of LHC. In particular it focuses on performance, stability and integration with the CMS sub-detectors.

Poster Session / 140

Mucura: your personal file repository in the cloud

Author: Fabio Hernandez[1]

Co-authors: Ran Du [2]; Wenjing Wu [3]

[1] *IN2P3/CNRS Computing Centre & IHEP Computing Centre*

[2] *Chinese Academy of Sciences (CN)*

[3] *Institute of High Energy Physics,Chinese Academy of Sciences (CN)*

Corresponding Author: fabio@in2p3.fr

By aggregating the storage capacity of hundreds of sites around the world, distributed data-processing platforms such as the LHC computing grid offer solutions for transporting, storing and processing massive amounts of experimental data, addressing the requirements of virtual organizations as a whole. However, from our perspective, individual workflows require a higher level of flexibility, ease of use and extensibility, which are not yet fully satisfied by the deployed storage systems.

In this contribution we report on our experience building Mucura, a prototype of a software system for building cloud-based file repositories of extensible capacity. Intended for individual scientists, the system allows you to store, retrieve, organize and share your remote files from your personal computer, by using both command line and graphical user interfaces.

Designed with usability, scalability and operability in mind, it exposes web-based standard APIs for storing and retrieving files and is compatible with the authentication mechanisms used by the existing grid computing platforms. At the core of the system there are components for managing file metadata and for secure storage of the files'contents, both implemented on top of highly available, distributed, persistent, scalable key-value stores. A front-end component is responsible for user

authentication and authorization and for handling requests from clients performing operations on the stored files.

We will present the selected open-source implementations for each component of the system and the integration work we have performed. In particular, we will present the rationale and findings of our exploration of key-value data stores as the central component of the system, as opposed to the usage of traditional networked file systems. We will also describe the pros and cons of our choices from the perspectives of both the end-user and the operator of the service. Finally, we will report on the feedback received from the early users and from the operators of the service.

This work is inspired not only by the increasing number of commercial services available nowadays to individuals for their personal storage needs (backup, file sharing, synchronization, ⋯) such as Amazon S3, Dropbox, SugarSync, bitcasa, etc., but also by several efforts in the same area in the academic and research worlds (NASA, SDSC, etc.). We are persuaded that the level of flexibility offered to individuals by this kind of systems adds value to the day-to-day work of scientists.

Poster Session / 141

High availability through full redundancy of the CMS detector controls system

Author: Giovanni Polese[1]

Co-authors: Alexander Flossdorf [2]; Andre Georg Holzner [3]; Andrea Petrucci [1]; Andrei Cristian Spataru [1]; Attila Racz [1]; Aymeric Arnaud Dupont [1]; Christian Deldicque [1]; Christian Hartl [1]; Christoph Paus [4]; Christoph Schwick [1]; Dennis Shpakov [5]; Dominique Gigi [1]; Emilio Meschi [1]; Frank Glege [1]; Frans Meijers [1]; Gerry Bauer [4]; Hannes Sakulin [1]; James Branson [3]; Jeroen Hegeman [1]; Jose Antonio Coarasa Perez [1]; Konstanty Sumorok [4]; Lorenzo Masetti [1]; Luciano Orsini [1]; Marc Dobson [1]; Marco Pieri [3]; Matteo Sani [3]; Matthew Bowen [6]; Michal Simon ; Olivier Raginel [4]; Remi Mommsen [5]; Robert Gomez-Reino Garrido [1]; Samim Erhan [7]; Sebastian Bukowiec [1]; Sergio Cittolin [3]; Ulf Behrens [8]; Vivian O'Dell [9]; Yi Ling Hwong [1]

[1] *CERN*

[2] *DESY*

[3] *Univ. of California San Diego (US)*

[4] *Massachusetts Inst. of Technology (US)*

[5] *Fermi National Accelerator Lab. (US)*

[6] *CERN, Geneva, Switzerland*

[7] *Univ. of California Los Angeles (US)*

[8] *Deutsches Elektronen-Synchrotron (DE)*

[9] *Fermi National Accelerator Laboratory (FNAL)*

Corresponding Author: giovanni.polese@cern.ch

The CMS detector control system (DCS) is responsible for controlling and monitoring the detector status and for the operation of all CMS sub detectors and infrastructure. This is required to ensure safe and efficient data taking, so that high quality physics data can be recorded. The current system architecture is composed of more than 100 servers, in order to provide the required processing resources. An optimization of the system software and hardware architecture is under development to ensure redundancy of all the controlled sub-systems and to reduce any downtime due to hardware or software failures. The new optimized structure is based mainly on powerful and highly reliable blade servers and makes use of a fully redundant approach, guaranteeing high availability and reliability. The analysis of the requirements, the challenges, the improvements and the optimized system architecture as well as its specific hardware and software solutions are presented.

Poster Session / 142

Legacy code: lessons from NA61/SHINE offline software upgrade adventure.

Author: Oskar Wyszynski[1]

Co-authors: Andras Laszlo [2]; Antoni Jerzy Marcinek [1]; Darko Veberic [3]; Marek Szuba [4]; Michael Unger [4]; Roland Sipos [2]; Tom Paul [5]

[1] *Jagiellonian University (PL)*

[2] *Hungarian Academy of Sciences (HU)*

[3] *University of Nova Gorica (SI)*

[4] *KIT - Karlsruhe Institute of Technology (DE)*

[5] *Department of Physics-Northeastern University*

Corresponding Author: oskar.wyszynski@cern.ch

Shine is the new offline software framework of the NA61/SHINE experiment at the CERN SPS for data reconstruction, analysis and visualization as well as detector simulation.
To allow for a smooth migration to the new framework, as well as to facilitate its validation, our transition strategy foresees to incorporate considerable parts of the old NA61/SHINE reconstruction chain which is based on the legacy code of NA49 experiment. Such a reuse of parts of old code, written mostly in C and Fortran, is an often arising problem in HEP experiments. Apart from the need to properly interface the old and new code, the migration task is complicated in our case due to the use of nonstandard commercial compilers in the NA49 code.
In this presentation we will describe the challenges faced during the porting of legacy code and discuss solutions that can help developers embarking on a similar adventure. In particular, we will describe the transition from scattered Makefiles to a monolithic CMake built system, the design of C++ interfaces to the legacy code and the semi-automatic conversion of non-standard PGI-Fortran constructs to code that compiles with GFortran. In addition, the validation of the physics output of the new framework will be discussed.

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Poster Session / 143

LHCb Conditions Database Operation Assistance Systems

Author: Illya Shapoval[1]

Co-authors: Hubert Degaudenzi [2]; Marco Clemencic [3]; Roberto Santinelli [3]

[1] *CERN, KIPT*

[2] *Ecole Polytechnique Federale de Lausanne (CH)*

[3] *CERN*

Corresponding Author: illya.shapoval@cern.ch

The Conditions Database of the LHCb experiment (CondDB) provides versioned, time dependent geometry and conditions data for all LHCb data processing applications (simulation, high level trigger, reconstruction, analysis) in a heterogeneous computing environment ranging from user laptops to the HLT farm and the Grid. These different use cases impose front-end support for multiple database technologies (Oracle and SQLite are used). Sophisticated distribution tools are required to ensure timely and robust delivery of updates to all environments. The content of the database has to be managed to ensure that updates are internally consistent and externally compatible with multiple versions of the physics application software. In this paper we describe three systems that we have developed to address these issues:

- an extension to the automatic content validation done by the "Oracle Streams"replication technology, to trap cases when the replication was unsuccessful;
- an automated distribution process for the SQLite-based CondDB, providing also smart backup and checkout mechanisms for the CondDB managers and LHCb users respectively;
- a system to verify and monitor the internal (CondDB self-consistency) and external (LHCb physics software vs. CondDB) compatibility.

These systems are used in production in the LHCb experiment and have achieved the desired goal of higher flexibility and robustness for the management and operation of the CondDB.

Poster Session / 144

Grid administration: towards an autonomic approach

Author: Federico Stagni[1]

Co-authors: Mario Ubeda Garcia [1]; Vincent Roger Yvan Bernardoff [2]

[1] *CERN*

[2] *Univ. P. et Marie Curie (Paris VI) (FR)*

Corresponding Author: federico.stagni@cern.ch

Within the DIRAC framework in the LHCb collaboration, we deployed an autonomous policy system acting as a central status information point for grid elements.
Experts working as grid administrators have a broad and very deep knowledge about the underlying system which makes them very precious. We have attempted to formalize this knowledge in an autonomous system able to aggregate information, draw conclusions, validate them, and take actions accordingly.
The DIRAC Resource Status System is a monitoring and generic policy system that enforces managerial and operational actions automatically. As an example, the status of a grid entity can be evaluated using a number of policies, each making assessments relative to specific monitoring information. Individual results of these policies can be combined to evaluate and propose a global status for the resource. This evaluation goes through a validation step driven by a state machine and an external validation system. Once validated, actions can be triggered accordingly.
External monitoring and testing systems such as Nagios or Hammercloud are used by policies for site commission and certification. This shows the flexibility of our system, and of what an autonomous policy system can achieve.

Poster Session / 145

LHCbDIRAC: distributed computing in LHCb

Author: Federico Stagni[1]

Co-authors: Alexey Zhelezov [2]; Andrei Tsaregorodtsev [3]; Joel Closier [1]; Krzysztof Ciba [1]; Mario Ubeda Garcia [1]; Matvey Sapunov [3]; Philippe Charpentier [1]; Ricardo Graciani Diaz [4]; Zoltan Mathe [5]

[1] *CERN*

[2] *Ruprecht-Karls-Universitaet Heidelberg (DE)*

[3] *Universite d'Aix - Marseille II (FR)*

[4] *University of Barcelona (ES)*

[5] *University College Dublin (IE)*

Corresponding Author: federico.stagni@cern.ch

We present LHCbDIRAC, an extension of the DIRAC community Grid solution to handle the LHCb specificities.

The DIRAC software has been developed for many years within LHCb only. Nowadays it is a generic software, used by many scientific communities worldwide. Each community wanting to take advantage of DIRAC has to develop an extension, containing all the necessary code for handling their specific cases.
LHCbDIRAC is an actively developed extension, implementing the LHCb computing model and workflows. LHCbDIRAC extends DIRAC to handle all the distributed computing activities of LHCb. Such activities include real data processing (reconstruction, stripping and streaming), Monte-Carlo simulation and data replication. Other activities are groups and user analysis, data management, resources management and monitoring, data provenance, accounting for user and production jobs. LHCbDIRAC also provides extensions of the DIRAC interfaces, including a secure web client, python APIs and CLIs. While DIRAC and LHCbDIRAC follow indpendent release cycles, every LHCbDIRAC is built on top of an existing DIRAC release. Before putting in production a new release candidate, a number of certification tests are run in a separate setup. This contribution highlights the versatility of the system, also presenting the experience with real data processing, data and resources management, monitoring for activities and resources.

Online Computing / 146

The Software Architecture of the LHCb High Level Trigger

Author: Hans Dijkstra[1]

Co-authors: Gerhard Raven [2]; J Michael Williams [3]; Johannes Albrecht [1]; Roel Aaij [4]; Vanya Belyaev [5]; Vladimir Gligorov [1]

[1] *CERN*

[2] *Free University (NL)*

[3] *Imperial College Sci., Tech. & Med. (GB)*

[4] *NIKHEF (NL)*

[5] *ITEP Institute for Theoretical and Experimental Physics (RU)*

Corresponding Authors: mariusz.witek@cern.ch, hans.dijkstra@cern.ch, gerhard.raven@nikhef.nl, vladimir.gligorov@cern.ch, johannes.albrecht@cern.ch, ivan.belyaev@cern.ch, roel.aaij@cern.ch, michael.williams@imperial.ac.uk

The LHCb experiment is a spectrometer dedicated to the study of heavy flavor at the LHC. The rate of proton-proton collisions at the LHC is 15 MHz, but disk space limitations mean that only 3 kHz can be written to tape for offline processing. For this reason the LHCb data acquisition system – trigger – plays a key role in selecting signal events and rejecting background. In contrast to previous experiments at hadron colliders like for example CDF or D0, the bulk of the LHCb trigger is implemented in software and deployed on a farm of 20k parallel processing nodes. This system, called the High Level Trigger (HLT) is responsible for reducing the rate from the maximum at which the detector can be read out, 1.1 MHz, to the 3 kHz which can be processed offline,and has 20 ms in which to process and accept/reject each event. In order to minimize systematic uncertainties, the HLT was designed from the outset to reuse the offline reconstruction and selection code, and is based around multiple independent and redundant, selection algorithms, which make it possible to trigger efficiently even in the case that one of the detector subsystems fails. Because of specific LHC running conditions, the HLT had to cope with three times higher event multiplicities than it was designed for in 2010 and 2011. This contribution describes the software architecture of the HLT, focusing on the code optimization and commissioning effort which took place during 2010 and 2011 in order to enable LHCb to trigger efficiently in these unexpected running conditions, and the flexibility and robustness of the LHCb software framework which allowed this reoptimization to be performed in a timely manner. We demonstrate that the software architecture of the HLT, in particular the concepts of algorithm redundancy and independence, were crucial to enable LHCb to deliver its nominal trigger signal efficiency and background rejection rate in these unexpected conditions, and outline lessons for future trigger design in particle physics experiments.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Computer Facilities, Production Grids and Networking / 147

From IPv4 to eternity - the HEPiX IPv6 working group

Authors: David Kelsey[1]; Edoardo Martelli[2]

Co-authors: Andreas Pfeiffer [2]; Bruno Hoeft [3]; David Foster [2]; Erik Mattias Wadenstein [4]; Francesco Prelz [5]; Julia Rohlfing [3]; Kars Ohrenberg [6]; Luuk Uljee [7]; Mario Reale [8]; Mark Mitchell [9]; Philip DeMar [10]; Ramiro Voicu [11]; Ronald van der Pol [7]; Simon Leinen [12]; Soumaya Lanouar [13]; Thomas Finnern [6]; Tony Wildish [14]; sabah salih [15]; sandor Rozsa [16]

[1] *STFC - Science & Technology Facilities Council (GB)*

[2] *CERN*

[3] *KIT*

[4] *Umea*

[5] *Sezione di Milano (INFN)-Universita e INFN*

[6] *DESY*

[7] *SARA*

[8] *GARR*

[9] *University of Glasgow*

[10] *FERMILAB*

[11] *California Institute of Technology (US)*

[12] *SWITCH (CH)*

[13] *EPFL*

[14] *Princeton University (US)*

[15] *Manchester*

[16] *California Institute of Technology (CALTECH)*

Corresponding Authors: edoardo.martelli@cern.ch, d.p.kelsey@rl.ac.uk

The much-heralded exhaustion of the IPv4 networking address space has finally started. While many of the research and education networks have been ready and poised for years to carry IPv6 traffic, there is a well-known lack of academic institutes using the new protocols. One reason for this is an obvious absence of pressure due to the extensive use of NAT or that most currently still have sufficient IPv4 addresses. More importantly though, the fact is that moving your distributed applications to IPv6 involves much more than the routing, naming and addressing solutions provided by your campus and national networks. Application communities need to perform a full analysis of their applications, middleware and tools to confirm how much development work is required to use IPv6 and to plan a smooth transition. A new working group of HEPiX (http://www.hepix.org) was formed in Spring 2011 to address exactly these issues for the High Energy Physics community.

The HEPiX IPv6 Working Group has been investigating the many issues which feed into the decision on the timetable for a transition to the use of IPv6 in HEP Computing, in particular for the Worldwide LHC Computing Grid (http://lcg.web.cern.ch/lcg/). The activities include the analysis and testing of the readiness for IPv6 and performance of many different components, including the applications, middleware, management and monitoring tools essential for HEP computing. A distributed IPv6 testbed has been deployed and used for this purpose and we have been working closely with the HEP experiment collaborations. The working group is also considering other operational issues such as the implications for security arising from a move to IPv6.

This paper describes the work done by the HEPiX IPv6 working group since its inception and presents our current conclusions and recommendations.

Event Processing / 148

”Swimming” : a data driven acceptance correction algorithm

Author: Vladimir Gligorov[1]

Co-authors: Gerhard Raven [2]; Hans Dijkstra [1]; Marco Gersabeck [1]; Roel Aaij [3]; Vanya Belyaev [4]

[1] *CERN*

[2] *Free University (NL)*

[3] *NIKHEF (NL)*

[4] *ITEP Institute for Theoretical and Experimental Physics (RU)*

Corresponding Authors: marco.cattaneo@cern.ch, vladimir.gligorov@cern.ch

The LHCb experiment is a spectrometer dedicated to the study of heavy flavor at the LHC. The rate of proton-proton collisions at the LHC is 15 MHz, but disk space limitations mean that only 3 kHz can be written to tape for offline processing. For this reason the LHCb data acquisition system – trigger – plays a key role in selecting signal events and rejecting background. Because the trigger efficiency, measured with respect to signal events selected by offline analysis algorithms, is not 100%, the trigger introduces biases in variables of interest. In particular, heavy flavor particles have a longer lifetime than background events, and the trigger exploits this information in its selections, introducing a bias in the lifetime distribution of offline selected heavy flavor particles. This bias must then be corrected for in order to perform measurements of heavy flavor lifetimes at LHCb, measurements which are particularly sensitive to physics beyond the Standard Model. This correction is accomplished by an algorithm called "swimming", which replays the passage of every offline selected event through the LHCb trigger, varying the lifetime of the signal at each step, and thus computes an event-by-event lifetime acceptance function for the trigger. This contribution describes the commissioning and deployment of the swimming algorithm during 2010 and 2011, and the world best lifetime and CP violation measurements accomplished using this method. In particular we focus on the key design decision in the architecture of the LHCb trigger which allows this method to work : the bulk of the triggering is implemented in software, reusing offline reconstruction and selection code to minimize systematics, and allowing the trigger selections to be re-executed offline exactly as they ran during data taking. We demonstrate the reproducibility of the LHCb trigger algorithms, show how the reuse of offline code and selections minimizes the biases introduced in the trigger, and show that the swimming method leads to an acceptance correction which contributes a negligible uncertainty to the measurements in question.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Software Engineering, Data Stores and Databases / 149

Optimizing Python-based ROOT I/O with PyPy's Tracing JIT

Author: Wim Lavrijsen[1]

[1] *Lawrence Berkeley National Lab. (US)*

Corresponding Author: wim.lavrijsen@cern.ch

The Python programming language allows objects and classes to respond dynamically to the execution environment. Most of this, however, is made possible through language hooks which by definition can not be optimized and thus tend to be slow. The PyPy implementation of Python includes a tracing just in time compiler (JIT), which allows similar dynamic responses but at the interpreter-, rather than the application-level. Therefore, it is possible to fully remove the hooks, leaving only the dynamic response, in the optimization stage for hot loops, if the types of interest are opened up to the JIT.

A general opening up of types to the JIT, based on reflection information, has already been developed (cppyy). The work described in this paper takes it one step further by customizing access to ROOT I/O to the JIT, allowing for automatic selective reading, judicious caching, and buffer tuning.

Poster Session / 151

Validation of Geant4 Releases with distributed resources

Author: Andrea Dotti[1]

[1] *CERN*

Corresponding Author: andrea.dotti@cern.ch

In this paper we present the Geant4 validation and testing suite.
The application is used to test any new Geant4 release. The simulation of a particularly demanding use-case (High Energy Physics calorimeters) is tested with different physics parameters.
The suite is integrated with a job submission system that allows for the generation of high statistics data-sets on distributed resources. The analysis of the data is also integrated and tools to store and visualize the results are provided.

Summary:

The simulation of calorimeters is particularly demanding: it challenges all aspects of physics simulation (tracking in magnetic field, electromagnetic and hadronic interactions). The Geant4 collaboration publishes a new version of Geant4 every year containing refinements of physics models and improvements in computing performance. In addition to the public releases, monthly development releases are used to monitor the developments of physics modeling.

To efficiently test all Geant4 versions a testing suite has been developed. A simplified version of HEP calorimeter has been implemented with Geant4. All LHC calorimeters materials and technologies have been implemented. The most important variables for calorimetric measurements are reconstructed and recorded for later analysis.

To increase the statistics being produced with this application, the testing suite has been recently extended to be run on distributed resources, being batch or GRID resources. Software is distributed to remote sites via a novel FUSE-based file system (CernVM-FS). The configuration of jobs, their submission, monitoring and collection of results is fully automated and integrated with GRID tools (DIANE/GANGA). Analysis of produced data is also performed in an automatic way and the relevant results are stored in a database. A simple web-interface (DRUPAL) has been developed to retrieve the data and produce interactively the plots (ROOT) to compare the physics performance between models or between versions of Geant4.

The testing suite is an example of the integration of different tools and technologies used in the HEP community that allows small Virtual Organizations to effectively use GRID resources.

Poster Session / 152

hBrowse - Generic framework for hierarchical data visualization

Author: Lukasz Kokoszkiewicz[1]

Co-authors: Ivan Antoniev Dzhunov [2]; Jakub Moscicki [1]; Julia Andreeva [1]; Laura Sargsyan [3]; Massimo Lamanna [1]

[1] *CERN*

[2] *University of Sofia*

[3] *A.I. Alikhanyan National Scientific Laboratory (AM)*

Corresponding Author: lukasz.kokoszkiewicz@cern.ch

The hBrowse framework is a generic monitoring tool designed to meet the needs of various communities connected to grid computing. It is strongly configurable and easy to adjust and implement accordingly to a specific community needs. It's a html/JavaScript client side application utilizing the latest web technologies to provide presentation layer to any hierarchical data structures. Each part of this software (dynamic tables, user selection etc.) is in fact a separate plugin which can be used separately from the main application. It was especially designed to meet the requirements of Atlas and CMS users as well as to use it as a bulked Ganga monitoring tool.

Summary:

The hBrowse Framework is a new kind of generic open source monitoring application. It's a html/javascript client that can be combined with any kind of server as long as it can send json formatted data. Whole application can be setup using just one settings file.

Poster Session / **153**

Scalability and performance improvements in Fermilab Mass Storage System.

Author: Alexander Moibenko[1]

Co-authors: Catalin Lucian Dumitrescu [2]; Dmitry Litvintsev [3]; Gene Oleynik [1]; Matt Crawford [4]

[1] *Fermilab*

[2] *Fermi National Accelerator Lab. (US)*

[3] *FNAL*

[4] *Fermi National Accelerator Lab. (Fermilab)*

Corresponding Author: moibenko@fnal.gov

By 2009 the Fermilab Mass Storage System had encountered several challenges:

1. The required amount of data stored and accessed in both tiers of the system (dCache and Enstore)had significantly increased.
2. The number of clients accessing Mass Storage System had also increased from tens to hundreds of nodes and from hundreds to thousands of parallel requests.

To address these challenges Enstore and the SRM part of dCache were modified to scale for performance, access rates, and capacity. This work increased the amount of simultaneously processed requests in a single Enstore Library instance from about 1000 to 30000. The rates of incoming request to Enstore increased from tens to hundreds per second.

Fermilab is invested in LTO4 tape technology and we have investigated both LTO5 and Oracle T10000C to cope with the increasing needs in capacity. We have decided to adopt T10000C, mainly due to its large capacity, which allows us to scale up the existing robotic storage space by a factor 6.

This paper describes the modifications and investigations that allowed us to meet these scalability and performance challenges and provided some perspectives of Fermilab Mass Storage System.

Poster Session / **154**

Geant4 electromagnetic physics for high statistic LHC simulation

Author: Vladimir Ivanchenko[1]

Co-authors: Alexander Bagulya [2]; Alexander Howard [3]; Alfonso Mantero ; Andreas Schaelicke [4]; Francisca Garay Walls [4]; Giovanni Santin [5]; Jean Jacquemier [6]; John Allison [7]; John Apostolakis [1]; Laszlo Urban [8]; Luciano Pandola [9]; Michel Maire [10]; Petteri Nieminen [5]; Sabine Elles [11]; Sebastien Laurent Incerti [11]; Vladimir Grichine [2]

[1] *CERN*

[2] *Russian Academy of Sciences (RU)*

[3] *Institut fuer Teilchenphysik-Eidgenoessische Tech. Hochschule Z*

[4] *University of Edinburgh (GB)*

[5] *ESA*

[6] *Laboratoire d'Annecy-le-Vieux de Physique des Particules (LAPP)*

[7] *University of Manchester (GB)*

[8] *Unknown*

[9] *INFN-LNGS*

[10] *LAPP*

[11] *Centre National de la Recherche Scientifique (FR)*

Corresponding Authors: francisca.garay.walls@cern.ch, vladimir.ivantchenko@cern.ch

An overview of the current status of electromagnetic physics (EM) of the Geant4 toolkit is presented. Recent improvements are focused on the performance of large scale production for LHC and on the precision of simulation results over a wide energy range. Significant efforts have been made to improve the accuracy and CPU speed for EM particle transport. New biasing options available for Geant4 EM physics are presented. It is shown that the performance of the EM sub-package is improved. We will report extensions of the testing suite for high statistics validation of EM physics. It includes validation of multiple scattering, bremsstrahlung and other models. The precision of simulation results on the shape of EM showers is discussed in detail. It includes the validation of both high and low energy components of a shower. Cross checks between standard and low-energy EM models have been performed using evaluated data libraries and reference benchmark results.

Poster Session / 155

Implementation of parallel processing in the basf2 framework for Belle II

Authors: Ryosuke ITOH[1]; Soohyung Lee[2]

Co-authors: Andreas Moll [3]; Martin Heck [4]; Nobuhiko Katayama [5]; Sogo Mineo [6]; Thomas Kuhr [7]

[1] *KEK*

[2] *Korea University*

[3] *MPI*

[4] *KIT*

[5] *HIGH ENERGY ACCELERATOR RESEARCH ORGANIZATION*

[6] *University of Tokyo*

[7] *KIT - Karlsruhe Institute of Technology (DE)*

Corresponding Author: ryosuke.itoh@kek.jp

Recent PC servers are equipped with multi-core CPUs and it is desired to utilize the full processing power of them for the data analysis in large scale HEP experiments. A software framework "basf2" is being developed for the use in the Belle II experiment, an upgraded B-factory experiment at KEK, and the parallel event processing is in its design. The framework accepts a set of plug-in functional modules and executes them in the specified order.

The parallel processing is implemented so that the execution of the partial portion of the module chain can be parallelized, while keeping the other modules to be executed in a single path. This implementation expands the capability of the parallel processing from the trivial event-by-event to the pipeline processing of a module chain, keeping the single input and output stream. The execution of one path is performed in a UNIX process forked from the main program of basf2.

The data passed between modules are a set of ROOT objects. In the parallel processing, whenever to pass the set to other process, they are streamed once and placed on a ring buffer implemented using UNIX IPC. The receiving process picks up the streamed packet from the ring buffer and restores the objects. The mechanism can be easily extended for the connection between PC servers over a network, which is used for the high level trigger application.

The details of the parallel processing implementation in the basf2 framework will be discussed at the conference, which includes a report of the realistic performance of the processing in various cases.

Summary:

The implementation of the parallel processing for the multi-core CPU in the Belle II software framework (basf2) is presented. It features the partial parallel execution of the module chain plugged in the framework, which enables the pipeline processing of modules in addition to the trivial event-by-event parallel processing.

Computer Facilities, Production Grids and Networking / **156**

The IceCube Computing Infrastructure Model

Author: Steve Barnet[1]

Co-author: Martin Merck [1]

[1] *University of Wisconsin Madison*

Corresponding Authors: barnet@wisc.edu, mmerck@icecube.wisc.edu

Besides the big LHC experiments a number of mid-size experiments is coming online which need to define new computing models to meet the demands on processing and storage requirements of those experiments. We present the hybrid computing model of IceCube which leverages GRID models with a more flexible direct user model as an example of a possible solution. In IceCube a central datacenter at UW-Madison servers as Tier-0 with a single Tier-1 datacenter at DESY Zeuthen. We describe the setup of the IceCube computing infrastructure and report on our experience in successfully provisioning the IceCube computing needs.

Event Processing / **157**

IceCubes GPGPU's cluster for extensive MC production

Author: Heath Skarlupka[1]

Co-authors: Dmitry Chirkin [1]; Juan Carlos Díaz Vélez [1]; Martin Merck [2]

[1] *UW Madison*

[2] *University of Wisconsin Madison*

Corresponding Authors: heath.skarlupka@icecube.wisc.edu, mmerck@icecube.wisc.edu

GPGPU computing offers extraordinary increases in pure processing power for parallelizable applications. In IceCube we use GPUs for ray-tracing of cherenkov photons in the antarctic ice as part of detector simulation. We report on how we implemented the mixed simulation production chain to include the processing on the GPGPU cluster for the IceCube Monte-Carlo production. We also present ideas to include GPGPU accelerated reconstructions into the IceCube data processing.

Poster Session / **158**

Enstore with Chimera namespace provider

Authors: Alexander Moibenko[1]; Dmitry Litvintsev[1]; Gene Oleynik[1]; Michael Zalokar[1]

[1] *Fermilab*

Corresponding Author: litvinse@fnal.gov

Enstore is a mass storage system developed by Fermilab that provides distributed access and management of the data stored on tapes. It uses namespace service, pnfs, developed by DESY to provide filesystem-like
view of the stored data. Pnfs is a legacy product and is being replaced by a new implementation, called Chimera, which is also developed by DESY. Chimera namespace offers multiple advantages over the pnfs in terms of performance and functionality. Enstore client component, encp, has been modified to work with Chimera or any other namespace provider. We performed high load end-to-end acceptance test of Enstore with chimera namespace. This paper describes modifications to Enstore, test procedure and results of the acceptance testing.

Poster Session / **159**

Building a local analysis center on OpenStack

Author: Martin Sevior[1]

[1] *University of Melbourne (AU)*

Corresponding Author: martines@unimelb.edu.au

The experimental high energy physics group at the University of Melbourne is a member of the ATLAS, Belle and Belle II collaborations. We maintain a local data centre which enables users to test pre-production code and to do final stage data analysis. Recently the Australian National eResearch Collaboration Tools and Resources (NeCTAR) organisation implemented a Research Cloud based on OpenStack middlewear.
This presentation details the development of an OpenStack-based local data analysis centre compromising hundreds of virtual machines with commensurate data storage.

Poster Session / **160**

CMS Analysis Deconstructed

Author: Sudhir Malik[1]

[1] *University of Nebraska-Lincoln*

Corresponding Author: sudhir.malik@cern.ch

The CMS Analysis Tools model has now been used robustly in a plethora of physics papers. This model is examined to investigate successes and failures as seen by the analysts of recent papers.

Poster Session / 161

Maintaining and improving of the training program on the analysis software in CMS

Author: Sudhir Malik[1]

[1] *University of Nebraska-Lincoln*

Corresponding Author: sudhir.malik@cern.ch

Since 2009, the CMS experiment at LHC has provided an intensive training on the use of Physics Analysis Tools (PAT), a collection of common analysis tools designed to share expertise and maximise the productivity in the physics analysis. More than ten one-week courses preceded by prerequisite studies have been organized and the feedback from the participants has been carefully analysed. This note describes how the training team designs, maintains and improves the course contents based on the feedback, the evolving analysis practices and the software development.

Software Engineering, Data Stores and Databases / 162

A CMake-based build and configuration framework

Author: Marco Clemencic[1]

Co-authors: Hubert Degaudenzi [2]; Pere Mato Vila [1]

[1] *CERN*

[2] *Ecole Polytechnique Federale de Lausanne (CH)*

Corresponding Author: marco.clemencic@cern.ch

The LHCb experiment has been using the CMT build and configuration tool for its software since the first versions, mainly because of its multi-platform build support and its powerful configuration management functionality. Still, CMT has some limitations in terms of build performance and the increased complexity added to the tool to cope with new use cases added latterly. Therefore, we have been looking for a viable alternative to it and we have investigated the possibility of adopting the CMake tool, which does a very good job for building and is getting very popular in the HEP community. The result of this study is a CMake-based framework which provides most of the special configuration features available natively only in CMT, with the advantages of better performances, flexibility and portability.

Software Engineering, Data Stores and Databases / 163

CMS experience with online and offline Databases

Author: Andreas Pfeiffer[1]

[1] *CERN*

Corresponding Author: andreas.pfeiffer@cern.ch

The CMS experiment is made of many detectors which in total sum up to more than 75 million channels. The online database stores the configuration data used to configure the various parts of the detector and bring it in all possible running states. The database also stores the conditions data, detector monitoring parameters of all channels (temperatures, voltages), detector quality information, beam conditions, etc. These quantities are used by the experts to monitor the detector performance in detail, as they occupy a very large space in the online database they cannot be used as-is for offline data reconstruction. For this, a "condensed" set of the full information, the "conditions data", is created and copied to a separate database used in the offline reconstruction.

The offline conditions database contains the alignment and calibrations data for the various detectors. Conditions data sets are accessed by a tag and an interval of validity through the offline reconstruction program CMSSW, written in C++. Performant access to the conditions data as C++ objects is a key requirement for the reconstruction and data analysis. About 200 types of calibration and alignment exist for the various CMS sub-detectors. Only those data which are crucial for reconstruction are inserted into the offline conditions DB. This guarantees a fast access to conditions during reconstruction and a small size of the conditions DB.

Calibration and alignment data are fundamental to maintain the design performance of the experiment. Very fast workflows have been put in place to compute and validate the alignment and calibration sets and insert them in the conditions database before the reconstruction process starts. Some of these sets are produced analyzing and summarizing the parameters stored in the online database. Others are computed using event data through a special express workflow. A dedicated monitoring system has been put up to monitor these time-critical processes.

The talk describes the experience with the CMS online and offline databases during the 2010 and 2011 data taking periods, showing some of the issues found and lessons learned.

Distributed Processing and Analysis on Grids and Clouds / 164

The Integration of CloudStack and OpenNebula with DIRAC

Authors: Ricardo Graciani Diaz[1]; Victor Manuel Fernandez Albor[2]; Victor Mendez Munoz[3]

Co-authors: Adrian Casajus Ramo [1]; Gonzalo Merino Arevalo [4]; Juan Jose Saborido Silva [2]

[1] *University of Barcelona (ES)*

[2] *Universidade de Santiago de Compostela (ES)*

[3] *Port d'Informació Científica (PIC)*

[4] *Centro de Investigaciones Energ. Medioambientales y Tecn. - (ES*

Corresponding Authors: victormanuel.fernandez@usc.es, vmendez@pic.es

The increasing availability of cloud resources is making the
scientific community to consider a choice between Grid and Cloud.
The DIRAC framework for distributed computing is an easy way
to obtain resources from both systems.
In this paper we explain the integration of DIRAC with a two
Open-source Cloud Managers, OpenNebula and CloudStack.
They are computing tools to manage the complexity and
heterogeneity of distributed data center infrastructures
which allow to create virtual clusters on demand including
public, private and hybrid clouds.
This approach requires to develop an extension to the
previous DIRAC Virtual Manager Server, developed for
Amazon EC2, allowing the connection with the cloud

managers.
In the OpenNebula case, the development has been based on the CERN Virtual Machine image with appropriate contextualization, while in the case of CloudStack, the infrastructure has been kept more general allowing other Virtual Machine sources and operating systems.
In both cases, CernVM File System has been used to facilitate software distribution to the computing nodes.
With the resulting infrastructure, users are allowed to use cloud resources transparently through a friendly interface like DIRAC Web Portal.

The main purpose of this integration is a system that can manage cloud and grid resources at the same time. Users from different communities do not need to care about the installation of the standard software that is available at the nodes, nor the operating system of the host machine which is transparent to the user. In this paper we analyse the overhead of the virtual layer, with some tests comparing the proposed approach with the existing Grid solution.

Poster Session / 165

MCPLOTS - a new tool for tuning and validation of Monte Carlo generators

Authors: Anton Karneyeu[1]; Anton Pytel[2]; Dmitri Konstantinov[3]; Liza Mijovic[4]; Michelangelo Mangano[5]; Peter Skands[5]; Stefan Prestel[None]; Witold Pokorski[5]

[1] *Russian Academy of Sciences (RU)*

[2] *Slovak Technical University*

[3] *Institute for High Energy Physics (RU)*

[4] *Deutsches Elektronen-Synchrotron (DE)*

[5] *CERN*

Corresponding Author: witold.pokorski@cern.ch

In this paper we present a new tool for tuning and validation of Monte Carlo (MC) generators, essential in order to have predictive power in the area of high-energy physics (HEP) experiments. With the first year of LHC data being now analyzed, the need for reliable MC generators is very clear. The tool, called MCPLOTS, is composed of a browsable repository of plots comparing HEP event generators to a wide variety of available experimental data, an underlying database, as well as a machinery for performing new analysis and validation and for producing new plots.

The browsable repository is the user entry point to the tool. It contains menus organized according to specific process types and observables. The plots show a comparison of different generators/tunes and experimental data. In a separate section, different versions of the same generator are compared to each other, to track the evolution of the implemented models. Future plans include an interactive service, where users can define and produce their comparisons and upload their own data plots for validation.

The underlying database uses MySQL technology and it holds histograms with the associated meta-data (beam type and energy, generator version, tune, etc). A PHP-based interface is used for the communication between the web page and the database. The use of a database allows for making specific queries to extract histograms for different ways of presentation (observables view, validation view). Links to the MC steering files and references to experimental data are also stored in the database which allows full reproducibility of results.

The machinery used to produce these plots is under continuous development and is touching different areas of computing, such as GRID, CLOUD or voluntary computing. The main MC analysis tool is Rivet. It allows to process the MC output in a way that the direct comparison to experimental published data is possible. Reliable comparisons sometimes mandate very large statistics of MC data. For this purpose interfaces to different generalized production farms have been implemented. In particular, the use of the CERN batch system and voluntary computing (LHC@home 2.0/BOINC project), have been envisaged.

After a year of being publicly available, the MCPLOTS tool has gained a lot of interest and positive feedback. It is constantly being developed and improved and its role for the LHC experiments is growing. The browsable repository can be accessed through http://mcplots.cern.ch.

Poster Session / **166**

Native ROOT graphics support on Apple devices (OSX and iOS)

Author: Timur Pocheptsov[1]

[1] *Joint Inst. for Nuclear Research (RU)*

Corresponding Author: timur.pocheptsov@cern.ch

ROOT's graphics works mainly via the TVirtualX class (this includes both GUI and non-GUI graphics). Currently, TVirtualX has two native implementations based on the X11 and Win32 low-level APIs. To make the X11 version work on
OS X we have to install the X11 server (an additional application), but unfortunately, there is no X11 for iOS and so no graphics for mobile devices from Apple - iPhone, iPad, iPod touch.

Apple provides developers with a very rich set of APIs and
frameworks, and in the area of GUI and 2D graphics these APIs are superior to the X11 API (e.g. we can easily add transparency, anti-aliasing, complex polygons and paths, blending, etc. etc.).
Using these APIs (mainly Quartz 2D) we have a new implementation of TVirtualX, which works both on OSX and iOS. The window management part for OSX is implemented using the Cocoa API. However, iOS has a completely different GUI model, which does not fit ROOT's GUI classes. In this case we provide our users with several classes, in (Objective-)C++, that allow the development of ROOT-based graphical applications for iOS.

Concerning 3D graphics, iOS does only support OpenGLES. OpenGLES is a sibling of OpenGL for mobile devices and browsers. We are porting the ROOT OpenGL based 3D graphics code to OpenGLES to bring ROOT's
3D graphics (event display, different math plots, etc.) to iOS.

Poster Session / **167**

XRootD client improvements

Author: Lukasz Janyst[1]

[1] *CERN*

Corresponding Author: lukasz.janyst@cern.ch

The XRootD server framework is becoming increasingly popular in the HEP community and beyond due to its simplicity, scalability and capability to construct distributed storage federations. With the growing adoption and new use cases emerging, it has become clear that the XRootD client code

has reached a stage, where a significant refactoring of the code base is necessary to remove, by now, unneeded functionality and further enhance scalability and maintainability. Areas of particular interest are consistent cache management and full support for multi-threading.

The cache support in ROOT has during the last year been re-implemented and generalized to laverage the application knowledge about future read locations lifting a consistent read-ahead strategy to the ROOT layer and thus making it available for all ROOT-supported protocols. This change allows to disable the XRootD-specific cache and read-ahead when XRootD is used from ROOT. Unfortunately, the current XRootD client design does not easily support this change, as the current cache is tightly coupled to the handling of asynchronous requests.

Also the current multi-threading support in the XRootD client is incomplete since file objects cannot be safely shared between multiple execution threads. Further the choice to use one thread per active socket limits scalability due to its resource consumption and makes it complex to synchronize parallel operations without the significant risk of dead-locks.

This contribution describes the developments that have been started in the XRootD project to address the above issues and presents some first scalability measurements obtained with the new client design.

Computer Facilities, Production Grids and Networking / **168**

A new data-centre for the LHCb experiment

Authors: Beat Jost[1]; Daniel Lacarrere[1]; Eric Thomas[1]; Laurent Roy[1]; Loic Brarda[1]; Niko Neufeld[1]; Rolf Lindner[1]

[1] *CERN*

Corresponding Author: niko.neufeld@cern.ch

The upgraded LHCb experiment, which is supposed to go into operation in 2018/19 will require a massive increase in its compute facilities. A new 2 MW data-centre is planned at the LHCb site. Apart from the obvious requirement of minimizing the cost, the data-centre has to tie in well with the needs of online processing, while at the same time staying open for future and offline use. We present our design and the evaluation process of various cooling and powering solutions as well as our ideas for fabric monitoring and control.

Poster Session / **169**

PEAC - A set of tools to quickly enable PROOF on a cluster

Authors: Gerardo GANIS[1]; Martin VALA[2]

[1] *CERN*

[2] *Slovak Academy of Sciences, Slovakia*

Corresponding Author: gerardo.ganis@cern.ch

With advent of the analysis phase of LHC data-processing, interest in PROOF technology has considerably increased. While setting up a simple PROOF cluster for basic usage is reasonably straightforward, exploiting the several new functionalities added in recent times may be complicated.

PEAC, standing for PROOF Enabled Analysis Cluster, is a set of tools aiming to facilitate the setup and management of a PROOF cluster. PEAC is based on the experience made by setting up PROOF for the ALICE analysis facilities. PEAC allows to easily build and configure

ROOT and the additional software needed on the cluster and features its own distribution system based on xrootd and Proof on Demand (PoD). The latter is also used for resource management (start, stop or daemons).

Finally, PEAC sets-up and configures dataset management (using the afdsmgrd daemon), as well as cluster monitoring (machine status and PROOF query summaries) using MonAlisa. In this respect, a Monalisa page has been dedicated to PEAC users, so that a cluster managed by PEAC can be automatically monitored.

In this talk we present and describe the main components of PEAC and show details of the existing facilities using it.

Software Engineering, Data Stores and Databases / 170

ROOT I/O in Javascript - Reading ROOT files in a browser

Author: Bertrand Bellenot[1]

[1] *CERN*

Corresponding Author: bertrand.bellenot@cern.ch

A JavaScript version of the ROOT I/O subsystem is being developed, in order to be able to browse (inspect) ROOT files in a platform independent way. This allows the content of ROOT files to be displayed in most web browsers, without having to install ROOT or any other software on the server or on the client. This gives a direct access to ROOT files from new (e.g. portable) devices in a light way. It will be possible to display simple graphical objects such as histograms and graphs (TH1, TH2, TH3, TProfile, TGraph, ...). The rendering will first be done with an external JavaScript graphic library, before investigating a way to produce graphics closer to what ROOT supports on other platforms (X11, Windows).

Poster Session / 171

Precision analysis of Geant4 condensed transport effects in detectors

Authors: Georg Weidenspointner[1]; Maria Grazia Pia[2]

Co-authors: Gabriela Hoff [3]; Matej Batic [4]; Stefanie Granato [1]

[1] *MPI Halbleiterlabor*

[2] *Universita e INFN (IT)*

[3] *CERN*

[4] *Jozef Stefan Institute*

Corresponding Authors: gabriela.hoff@cern.ch, maria.grazia.pia@cern.ch

Physics models and algorithms operating in the condensed transport scheme - multiple scattering and energy loss of charged particles - play a critical role in the simulation of energy deposition in detectors.

Geant4 algorithms pertinent to this domain involve a number of parameters and physics modeling approaches, which have evolved in the course of the years. Results in the literature document their effects on physics observables in detectors, but comparisons with experiment for model validation are relatively scarce, and a comprehensive overview of the problem domain is still missing, despite its relevance to experimental applications.

In-depth analysis of Geant4 models operating in the condensed transport scheme is reported. A simultaneous validation is performed to evaluate the accuracy of backscattering and energy deposition: accurate rendering of both observables through the same physics settings is a known issue in Monte Carlo simulation, and a sensitive test of the robustness of the algorithms. The analysis involves the contributions of Geant4 charged particle interaction models, energy loss and multiple scattering algorithms: quantitative results highlighting the role of the various components are presented.

Poster Session / **172**

A tool for Image Management in Cloud Computing

Author: qiulan huang[1]

Co-authors: sha li [1]; wenxiao kan [1]

[1] *Institute of High Energy Physics, Beijing*

Corresponding Author: huangql@ihep.ac.cn

Entering information industry, the most new technologies talked about are virtualization and cloud computing. Virtualization makes the heterogeneous resources transparent to users, and plays a huge role in large-scale data center management solutions. Cloud computing emerges as a revolution in computing science which bases on virtualization, demonstrating a gigantic advantage in resource sharing, resource utilization, resource flexibility and resource scalability. And the new technology comes with new problems in which IT infrastructure is deployed with virtual machines. Among these is the problem of managing the virtual machine images that are indispensible to cloud environment.
In order to deploying a large-scale cloud infrastructure within a tolerant time, how to distribute image to hypervisor quickly and make its validity and integrity is the most important thing to be considered. To address that, this paper proposes an image management system acting as image repository as well as image distributor that provides users a friendly portal and also effective standard commands. The system interfaces implement the operations like register, upload, download, unregister, delete and so on. Hence, some other features like access control rules for diverse users to guarantee security in cloud computing. To optimize the performance, different storage systems such as NFS, Lustre, AFS and gLusterFS are compared in detail as well as the image distribution protocol like peer-to-peer(P2P), http and scp are analyzed, which demonstrates the proper storage system and distribution protocol are essential to the performance of the system.
The workflow of the deployment of a cloud using virtual machine provisioning like Opennebula is introduced and the comparison between diverse storage systems and transfer protocols is discussed. The high performance and scalability of image distribution of the system in production are fully proved and one virtual machine is deployed quickly within minute in average. Some useful tips for image management are also proposed.

Summary:

This paper proposes an image management system acting as image repository as well as image distributor that provides users a friendly portal and also effective standard commands. The system interfaces implement the operations like register, upload, download, unregister, delete and so on. Hence, some other features like access control rules for diverse users to guarantee security in cloud computing.
Besides,the workflow of the deployment of a cloud using virtual machine provisioning like Opennebula is introduced and the comparison between diverse storage systems and transfer protocols is discussed.The high performance and scalability of image distribution of the system in production are fully proved. Some useful tips for image management are also proposed.

Poster Session / **173**

Evaluation of software based redundancy algorithms for the EOS storage system at CERN

Authors: Andreas Peters[1]; Elvin Alin Sindrilaru[1]

[1] *CERN*

Corresponding Author: andreas.joachim.peters@cern.ch

EOS is a new disk based storage system used in production at CERN since autumn 2011. It is implemented using the plug-in architecture of the XRootD software framework and allows remote file access via XRootD protocol or POSIX-like file access via FUSE mounting. EOS was designed to fulfill specific requirements of disk storage scalability and IO scheduling performance for LHC analysis use cases. This is achieved by following a strategy of decoupling disk and tape storage as individual storage systems. A key point of the EOS design is to provide high availability and redundancy of files via a software implementation which uses disk-only storage systems without hardware RAID arrays. All this is aimed at reducing the overall cost of the system and also simplifying the operational procedures. This paper presents advantages and disadvantages of redundancy by hardware (most classical storage installations) in comparison to redundancy by software. The latter is implemented in the EOS system and achieves its goal by spawning data and parity stripes via remote file access over nodes. The gain in redundancy and reliability comes with a trade-off in the following areas:

```
- Increased complexity of the network connectivity
- CPU intensive parity computations during file creation and recovery
- Performance loss through remote disk coupling
```

An evaluation and performance figures of several redundancy algorithms are presented for simple file mirroring, dual parity RAID, Reed-Solomon and LDPC codecs. Moreover, the characteristics and applicability of these algorithms are discussed in the context of reliable data storage systems. Finally, a summary of the current state of implementation is given, sharing some experiences on migration and operation of a new multi-PB disk storage system at CERN.

Poster Session / 174

Health and performance monitoring of the large and diverse online computing cluster of CMS

Author: Olivier Raginel[1]

Co-authors: Alexander Flossdorf [2]; Andre Georg Holzner [3]; Andrea Petrucci [4]; Andrei Cristian Spataru [4]; Attila Racz [4]; Aymeric Arnaud Dupont [4]; Christian Deldicque [4]; Christian Hartl [4]; Christoph Paus [1]; Christoph Schwick [4]; Dennis Shpakov [5]; Dominique Gigi [4]; Emilio Meschi [4]; Frank Glege [4]; Frans Meijers [4]; Gerry Bauer [1]; Giovanni Polese [4]; Hannes Sakulin [4]; James Branson [3]; Jeroen Hegeman [4]; Jose Antonio Coarasa Perez [4]; Konstanty Sumorok [1]; Lorenzo Masetti [4]; Luciano Orsini [4]; Marc Dobson [4]; Marco Pieri [3]; Marek Ciganek [4]; Matteo Sani [3]; Matthew Bowen [6]; Michal Simon ; Olivier Bouffet [4]; Remi Mommsen [5]; Robert Gomez-Reino Garrido [4]; Samim Erhan [7]; Sebastian Bukowiec [4]; Sergio Cittolin [3]; Ulf Behrens [8]; Vivian O'Dell [9]; Yi Ling Hwong [4]

[1] *Massachusetts Inst. of Technology (US)*

[2] *DESY*

[3] *Univ. of California San Diego (US)*

[4] *CERN*

[5] *Fermi National Accelerator Lab. (US)*

[6] *University of the West of England*

[7] *Univ. of California Los Angeles (US)*

[8] *Deutsches Elektronen-Synchrotron (DE)*

[9] *Fermi National Accelerator Laboratory (FNAL)*

Corresponding Author: olivier.raginel@cern.ch

The CMS experiment online cluster consists of 2300 computers and170 switches or routers operating on a 24 hour basis. This huge infrastructure must be monitored in a way that the administrators are proactively warned of any failures or degradation in the system, in order to avoid or minimize downtime of the system which can lead to loss of data taking. The number of metrics monitored per host varies from 20 to 40 and covers basic host checks (disk, network, load) to application specific checks (service running) in addition to hardware monitoring (through IPMI). The sheer number of hosts and checks per host in the system stretches the limits of many monitoring tools and requires careful usage of various configuration optimizations in order to work reliably. The initial monitoring system used in the CMS online cluster was based on Nagios, but suffered from various drawbacks and did not work reliably in the recently expanded cluster. The CMS cluster administrators investigated the different open source tools available and chose to use a fork of Nagios called Icinga, with several plugin modules to enhance its scalability. The Gearman module provides a queuing system for all checks and their results allowing easy load balancing across worker nodes. Supported modules allow the grouping of checks in one single request thereby significantly reducing the network overhead for doing a set of checks on a group of nodes. The PNP4nagios module provides the graphing capability to Icing, which uses files as round robin databases (RRD). Additional software (rrdcached) optimizes access to the RRD files and is vital in order to achieve the required number of operations. Furthermore, to make best use of the monitoring information to notify the appropriate communities of any issues with their systems, much work was put into the grouping of the checks according to, for example, the function of the machine, the services running, the sub-detectors they belong to, and the criticality of the computer. An automated system to generate the configuration of the monitoring system has been produced to facilitate its evolution and maintenance. The use of these performance enhancing modules and the work on grouping the checks has yielded impressive performance improvements over the pervious Nagios infrastructure allowing for the monitoring of X metrics per second (compared to Y on the previous system). Furthermore the design allows the easy growth of the infrastructure without the need to rethink the monitoring system as a whole.

Event Processing / 175

Pattern Recognition for a Continuously Operating GEM-TPC

Author: Johannes Rauch[1]

Co-authors: Bernhard Ketzer [2]; Felix Valentin Boehmer [2]; Sebastian Neubert [3]; Stephan Paul [4]

[1] *Technische Universität München*

[2] *Technische Universitaet Muenchen (DE)*

[3] *Technical University Munich*

[4] *Institut fuer Theoretische Physik*

Corresponding Author: johannes.rauch@mytum.de

A pattern recognition software for a continuously operating high rate Time Projection Chamber with
Gas Electron Multiplier amplification (GEM-TPC) has been designed and tested.
A track-independent clustering algorithm delivers space points. A true 3-dimensional track follower combines
them to helical tracks, without constraints on the vertex position.
Fast helix fits, based on a conformal mapping on the Riemann sphere, are the basis for deciding whether points belong to one track.

The software has been tested on simulated as well as on real data
taken in a physics run of the GEM-TPC prototype installed in the FOPI detector at GSI facility, Germany.
To assess the performance of the algorithm in a high-rate environment, ppbar-interactions
corresponding to a maximum average track density of 0.5 cm/cm^3 have been simulated.

The pattern recognition is capable of finding all kinds of track topologies with high efficiency
and provides excellent seed values for fitting or online event selection.

Computational costs are O(50) ms/track on a 3.1 GHz office PC.
Parallel implementation of the code on a graphics processing unit (GPU) is under investigation.

Structure, functioning and benchmark results of the algorithm will be presented.

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Distributed Processing and Analysis on Grids and Clouds / 176

Implementing data placement strategies for the CMS experiment based on a popularity mode

Authors: Daniele Spiga[1]; Domenico Giordano[1]; Edward Karavakis[1]; Fernando Harald Barreiro Megino[2]; Maria Girone[1]; Mattia Cinquilli[3]; Nicolo Magini[1]; Valentina Mancinelli[4]

[1] *CERN*

[2] *Universidad Autonoma de Madrid (ES)*

[3] *Univ. of California San Diego (US)*

[4] *Sezione di Perugia (INFN)-Universita e INFN*

Corresponding Authors: domenico.giordano@cern.ch, fernando.harald.barreiro.megino@cern.ch

During the first two years of data taking, the CMS experiment has collected over 20 PetaBytes of data and processed and analyzed it on the distributed, multi-tiered computing infrastructure on the WorldWide LHC Computing Grid. Given the increasing data volume that has to be stored and efficiently analyzed, it is a challenge for several LHC experiments to optimize and automate the data placement strategies in order to fully profit of the available network and storage resources and to facilitate daily computing operations.

Building on previous experience acquired by ATLAS, we have developed the CMS Popularity Service that tracks file accesses and user activity on the grid and will serve as the foundation for the evolution of their data placement. A fully automated, popularity-based site-cleaning agent has been deployed in order to scan Tier2 sites that are reaching their space quota and suggest obsolete, unused data that can be safely deleted without disrupting analysis activity. Future work will be to demonstrate dynamic data placement functionality based on this popularity service and integrate it in the data and workload management systems: as a consequence the pre-placement of data will be minimized and additional replication of hot datasets will be requested automatically.

This paper will give an insight into the development, validation and production process and will analyze how the framework has influenced resource optimization and daily operations in CMS.

Poster Session / 177

No file left behind - monitoring transfer latencies in PhEDEx

Authors: Natalia Ratnikova[1]; Nicolo Magini[2]

Co-authors: Alberto Sanchez Hernandez [3]; Andrea Sartirana [4]; Chih-Hao Huang [5]; Federica Moscato [6]; Markus Klute [7]; Mingming Yang [8]; Oliver Gutsche [6]; Paul Rossman [9]; Rapolas Kaselis [10]; Si Xie [8]; Stefan Piperov [11]; Thorsten Chwalek [12]; Tony Wildish [13]

[1] *KIT - Karlsruhe Institute of Technology (DE)*

[2] *CERN*

[3] *Centro Invest. Estudios Avanz. IPN (MX)*

[4] *Ecole Polytechnique (FR)*

[5] *Fermi National Accelerator Laboratory*

[6] *Fermi National Accelerator Lab. (US)*

[7] *Massachusettes Institute of Technology*

[8] *Massachusetts Inst. of Technology (US)*

[9] *Fermi National Accelerator Laboratory (FNAL)*

[10] *Vilnius University (LT)*

[11] *Bulgarian Academy of Sciences (BG)*

[12] *KIT - Karlsruhe Institute of Technology*

[13] *Princeton University (US)*

Corresponding Authors: natalia.ratnikova@cern.ch, nicolo.magini@cern.ch

The CMS experiment has to move Petabytes of data among dozens of computing centres with low latency in order to make efficient use of its resources. Transfer operations are well established to achieve the desired level of throughput, but operators lack a system to identify early on transfers that will need manual intervention to reach completion.

File transfer latencies are sensitive to the underlying problems in the transfer infrastructure, and their measurement can be used as prompt trigger for preventive actions. For this reason, PhEDEx, the CMS transfer management system, has recently implemented a monitoring system to measure the transfer latencies at the level of individual files. For the first time now, the system can predict the completion time for the transfer of a data set. The operators can detect abnormal patterns in transfer latencies early, and correct the issues while the transfer is still in progress. Statistics are aggregated for blocks of files, recording a historical log to monitor the long-term evolution of transfer latencies, which are used as cumulative metrics to evaluate the performance of the transfer infrastructure, and to plan the global data placement strategy.

In this contribution, we present the typical patterns of transfer latencies that have been identified in the operational experience acquired with the latency monitor. We show how we are able to detect the sources of latency arising from the underlying infrastructure (such as stuck files) which need operator intervention, and we identify the areas in PhEDEx where a development effort can reduce the latency. The improvement in transfer completion times achieved since the implementation of the latency monitoring in 2011 is demonstrated.

Poster Session / 178

Integrating PROOF Analysis in Cloud and Batch Clusters

Author: Ana Y. Rodríguez-Marrero[1]

Co-authors: Alberto Cuesta-Noriega [2]; Enol Fernández-del-Castillo [1]; Francisco Matorras-Weinig [1]; Isidro González-Caballero [2]; Jesús Marco-de-Lucas [1]; Álvaro López-García [1]

[1] *Instituto de Física de Cantabria (UC-CSIC)*

[2] *Universidad de Oviedo*

High Energy Physics (HEP) analysis are becoming more complex and demanding due to the large amount of data collected by the current experiments. The Parallel ROOT Facility (PROOF) provides researchers with an interactive tool to speed up the analysis of huge volumes of data by exploiting parallel processing on both multicore machines and computing clusters. The typical PROOF deployment scenario is a permanent set of cores configured to run the PROOF daemons. However, this approach is incapable of adapting to the dynamic nature of interactive usage. Several initiatives seek to improve the use of computing resources by integrating PROOF with a batch system, such as PoD or PROOF Cluster. These solutions are currently in production at Universidad de Oviedo and IFCA and are positively evaluated by users. Although they are able to adapt to the computing

needs of users, they must comply with the specific configuration, OS and software installed at the batch nodes. Furthermore, they share the machines with other workloads, which may cause disruptions in the interactive service for users. These limitations make PROOF a typical use-case for cloud computing. In this work we take profit from Cloud Infrastructure at IFCA in order to provide a dynamic PROOF environment where users can control the software configuration of the machines. The Proof Analysis Framework (PAF) facilitates the development of new analysis and offers a transparent access to PROOF resources. Several performance measurements are presented for the different scenarios (PoD, SGE and Cloud), showing a speed improvement closely correlated with the number of cores used.

Poster Session / **179**

Developing CMS software documentation system

Author: Mantas Stankevicius[1]

Co-authors: Kati Lassila-Perini [2]; Sudhir Malik [3]

[1] *Vilnius University (LT)*

[2] *Helsinki Institute of Physics (FI)*

[3] *University of Nebraska-Lincoln*

Corresponding Author: mantas.stankevicius@cern.ch

CMSSW (CMS SoftWare) is the overall collection of software and services needed by the simulation, calibration and alignment, and reconstruction modules that process data so that physicists can perform their analysie. It is a long term project, with a large amount of source code. In large scale and complex projects is important to have as up-to-date and automated software documentation as possible. The core of the documentation should be version-based and available online with the source code. CMS uses Doxygen and Twiki as main tools to provide automated and non-automated documentation. Both of them are heavily cross-linked to prevent duplication of information. Doxygen is used to generate functional documentation and dependency graphs from the source code. Twiki is divided into two parts: WorkBook and Software guide. WorkBook contains tutorial-type instructions on accessing computing resources and using the software to perform analysis within the CMS collaboration and Software guide gives further details. This note describes the design principles, the basic functionalities and the technical implementations of the CMSSW documentation.

Poster Session / **180**

OSG Ticket Synchronization: Keeping Your Home Field Advantage In A Distributed Environment

Author: Kyle Gross[1]

Co-author: Robert Quick [2]

[1] *Open Science Grid / Indiana University*

[2] *Indiana University*

Large distributed computing collaborations, such as the WLCG, face many issues when it comes to providing a working grid environment for their users. One of these is exchanging tickets between various ticketing systems in use by grid collaborations. Ticket systems such as Footprints, RT, Remedy, and ServiceNow all have different schema that must be addressed in order to provide a reliable exchange of information between support entities and users in different grid environments. To combat this problem, Open Science Grid (OSG) Operations has created a ticket synchronization interface called GOC-TX that relies on web services instead of error-prone email parsing methods of the past. Synchronizing tickets between different ticketing systems allows any user or support

entity to work on a ticket in their home environment, thus providing a familiar and comfortable place to provide updates without having to learn another ticketing system. The interface is built in a way that it is generic enough that it can be customized for nearly any ticketing system with a web-service interface with only minor changes. This allows us to be flexible and rapidly bring new ticket synchronization online. Synchronization can be triggered by different methods including mail, web services interface, and active messaging. GOC-TX currently interfaces with GGUS for WLCG, Remedy at BNL, and RT at VDT. Work is progressing on the FNAL ServiceNow synchronization. This paper will explain the problems faced by OSG and how they led OSG to create and implement this ticket synchronization system along with the technical details that allow synchronization to be preformed at a production level.

Student? Enter 'yes'. See http://goo.gl/MVv53:

No

Summary:

See Abstract Content

Poster Session / 181

The Event Notification and Alarm System for the Open Science Grid Operations Center

Author: Scott Teige[1]

Co-authors: Robert Quick [1]; Soichi Hayashi [1]

[1] *Indiana University*

The Open Science Grid Operations (OSG) Team operates a distributed set of services and tools that enable the utilization of the OSG by several HEP projects. Without these services users of the OSG would not be able to run jobs, locate resources, obtain information about the status of systems or generally use the OSG. For this reason these services must be highly available. This paper describes the automated monitoring and notification systems used to diagnose and report problems. Described here are the means used by OSG Operations to monitor systems such as physical facilities, network operations, server health, service availability and software error events.
Once detected, an error condition generates a message sent to, for example,
Email, SMS, Twitter, an Instant Message Server, etc.
The approach used to integrate these monitoring systems into a prioritized and configurable alarming
mechanism is particularly emphasized. This system along with the ability to quickly restore interrupted services has allowed consistent operation of critical services with near 100% availability.

Distributed Processing and Analysis on Grids and Clouds / 182

Towards a global monitoring system for CMS computing operations

Authors: Andrea Sciaba[1]; Lothar A.T. Bauerdick[2]

[1] *CERN*

[2] *Fermi National Accelerator Lab. (US)*

Corresponding Authors: andrea.sciaba@cern.ch, lothar.bauerdick@cern.ch

The operation of the CMS computing system requires a complex monitoring system to cover all its aspects: central services, databases, the distributed computing infrastructure, production and analysis workflows, the global overview of the CMS computing activities and the related historical information. Several tools are available to provide this information, developed both inside and outside of the collaboration and often used in common with other experiments. Despite the fact that the current monitoring allowed CMS to successfully perform its computing operations, an evolution of the system is clearly required, to adapt to the recent changes in the data and workload management tools and models and to address some shortcomings that make its usage less than optimal. Therefore, a recent and ongoing coordinated effort was started in CMS, aiming at improving the entire monitoring system by identifying its weaknesses and the new requirements from the stakeholders, rationalise and streamline existing components and drive future software development. This contribution gives a complete overview of the CMS monitoring system and a description of all the recent activities that have been started with the goal of providing a more integrated, modern and functional global monitoring system for computing operations.

Poster Session / 183

Fast Simulation of the CMS Detector at the LHC

Author: Rahmat Rahmat[1]

[1] *University of Mississippi (US)*

Corresponding Author: rahmat.rahmat@cern.ch

A framework for Fast Simulation of particle interactions in the CMS detector has been developed and implemented in the overall simulation, reconstruction and analysis framework of CMS. It produces data samples in the same format as the one used by the Geant4-based (henceforth Full) Simulation and Reconstruction chain; the output of the Fast Simulation of CMS can therefore be used in the analysis in the same way as other ones. The Fast Simulation has been used already for several physics analyses in CMS, in particular those requiring a generation of many samples to scan an extended parameter space of the physics model (e.g. SUSY). Other use cases dealt with by the Fast Simulation of CMS are those involving the generation of large cross-section backgrounds, and samples of manageable size can only be produced by events skimming based on the final reconstructed objects, or those for which in general a large computation time is foreseen. An important issue, related with the high luminosity achieved by the LHC accelerator, is the pileup. The Fast Simulation of CMS can further take into account the superposition of as many pileup events as the ones provided now or even expected in the LHC upgrades, in an extremely shorter computation time than the one required by the Full Simulation for the same task, with just a few shortcuts which will be also discussed here. Comparisons of the Fast Simulation results both with the Full Simulation and with the LHC data collected in the years 2010 and 2011 at the center of mass energy of 7 TeV will also be shown, to demonstrate the level of accuracy achieved so far.

Student? Enter 'yes'. See http://goo.gl/MVv53:

No

Poster Session / 184

Life in extra dimensions of database world or penetration of NoSQL in HEP community

Author: Valentin Kuznetsov[1]

[1] *Cornell University*

Corresponding Author: valentin.kuznetsov@cern.ch

The recent buzzword in IT world is NoSQL. Major players, such as Facebook, Yahoo, Google, etc. are widely adopted different "NoSQL" solutions for their needs. Horizontal scalability, flexible data model and management of big data volumes are only a few advantages of NoSQL. In CMS experiment we use several of them in production environment. Here we present CMS projects based on NoSQL solutions, their strengths and weaknesses as well as our experience with those tools and their coexistence with standard RDMS solutions in our applications.

Poster Session / **185**

A gLite FTS based solution for managing user output in CMS

Author: Daniele Spiga[1]

Co-authors: Eric Wayne Vaandering [2]; Hassen Riahi [3]; Marco Mascheroni [4]; Mattia Cinquilli [5]

[1] *CERN*

[2] *Fermi National Accelerator Lab. (US)*

[3] *Universita e INFN (IT)*

[4] *Nat. Inst. of Chem.Phys. & Biophys. (EE)*

[5] *Univ. of California San Diego (US)*

Corresponding Authors: daniele.spiga@cern.ch, mattia.cinquilli@cern.ch, hassen.riahi@cern.ch

The CMS distributed data analysis workflow assumes that jobs run in a different location to where their results are finally stored. Typically the user output must be transferred across the network from one site to another, possibly on a different continent or over links not necessarily validated for high bandwidth/high reliability transfer. This step is named stage-out and in CMS was originally implemented as a synchronous step of the job execution. However, our experience showed the weakness of this approach both in terms of low total job execution efficiency and failure rates, wasting precious CPU resources.
The nature of analysis data makes it inappropriate to use PhEDEx, CMS' core data placement system. As part of the new generation of CMS Workload Management tools, the Asynchronous Stage-Out system (AsyncStageOut) has been developed to enable third party copy of the user output. The AsyncStageOut component manages glite FTS transfers of data from a temporary store at the site where the job ran to the final location of the data on behalf of that data owner. The tool uses python daemons, built using the WMCore framework, talking to CouchDB, to manage the queue of work and FTS transfers. CouchDB also provides the platform for a dedicated operations monitoring system.
In this paper, we present the motivations of the asynchronous stage-out system. We give an insight into the design and the implementation of key features, describing how it is coupled with the CMS workload management system. Finally, we show the results and the commissioning experience.

Poster Session / **186**

Cloud based multi-platform data analysis application

Authors: Gerardo Ganis[1]; Joao Antunes Pequenao[2]; Neng Xu[3]

[1] *CERN*

[2] *Lawrence Berkeley National Lab. (US)*

[3] *University of Wisconsin (US)*

Corresponding Author: neng.xu@cern.ch

With the start-up of the LHC in 2009, more and more data analysis facilities have been built or enlarged at Universities and laboratories. In the mean time, new technologies, like Cloud computing and Web3D, and new types of hardware, like smartphones and tablets, have become available and popular in the market. Is there a way to integrate them into the existing data analysis models and allow physicists to do their daily work more conveniently and efficiently?

In this paper we will discuss the development of a platform independent thin client application for data analysis on Cloud based infrastructures. The goal of this new development is to allow physicists to be able to run their data analysis with different hardware, like laptop, smart
phone, tablet and access their data everywhere. The application can run within the web browser and smartphones without compatibility problems. Based on one of the most popular graphic engines, people can view 2D histograms, animated 3D event displays and even do event analysis. The heavy processing jobs will be sent to the Cloud via a master server, in such a way that people can run multiple complex jobs simultaneously.

After having introduced the new system structure and the way the new application will fit in the overall picture, we will describe the current progress of the development and the test facility and discuss further technical difficulties that we expect to be confronted to, like the security (user authentication and authorization) data discovery and load balancing.

Poster Session / **187**

Data Bookkeeping Service 3 - A new event data catalog for CMS

Author: Manuel Giffels[1]

Co-author: Yuyi Guo [2]

[1] *CERN*

[2] *Fermi National Accelerator Lab. (US)*

Corresponding Author: manuel.giffels@cern.ch

The Data Bookkeeping Service 3 (DBS 3) provides an improved event data catalog for Monte Carlo and recorded data of the CMS (Compact Muon Solenoid) experiment at the Large Hadron Collider (LHC). It provides the necessary information used for tracking datasets, like data processing history, files and runs associated with a given dataset on a scale of about 10ˆ5 datasets and more than 10ˆ7 files. All kinds of data processing in CMS are relying on the information stored in DBS. Thus, it is widely used within CMS, in Monte Carlo production, processing of recorded data as well as in physics analysis done by users.

DBS 3 has been completely re-designed and re-implemented in Python using a CherryPy based environment, and has as its basis RESTful (Representational State Transfer) web services, commonly used within the data management and workload management (DMWM) group of CMS. DBS 3 is using the Java Script Object Notation (JSON) dataformat for interchanging information and Oracle as database backend. Main focuses during the process of development were an adaptation of the database schema to better match the evolving CMS data processing model, the introduction of the Data Aggregation System in CMS, which is combining the information of a variety of database services (PhEDEx, SiteDB, DBS, etc.) in one user interface and the achievement of a better scalability to match the growing demands even in the future.

The design, the current status of the development and deployment as well as first experiences with that system during testing/operation will be described in this contribution.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Poster Session / 188

From toolkit to framework - the past and future evolution of PhEDEx

Author: Tony Wildish[1]

Co-authors: Alberto Sanchez Hernandez [2]; Chih-Hao Huang [3]; Natalia Ratnikova [4]; Nicolo Magini [5]

[1] *Princeton University*

[2] *Centro Invest. Estudios Avanz. IPN (MX)*

[3] *Fermi National Accelerator Laboratory*

[4] *KIT - Karlsruhe Institute of Technology (DE)*

[5] *CERN*

Corresponding Author: tony.wildish@cern.ch

PhEDEx is the data-movement solution for CMS at the LHC. Created in 2004, it is now one of the longest-lived components of the CMS dataflow/workflow world.

As such, it has undergone significant evolution over time, and continues to evolve today, despite being a fully mature system. Originally a toolkit of agents and utilities dedicated to specific tasks, it is becoming a more open framework that can be used in several ways, both within and beyond its original problem domain.

In this talk we describe how a combination of refactoring and adoption of new technologies that have become available over the years have made PhEDEx more flexible, maintainable, and scalable. Finally, we describe how we will guide the evolution of PhEDEx into the future.

Online Computing / 189

The Compact Muon Solenoid Detector Control System

Author: Robert Gomez-Reino Garrido[1]

Co-authors: Alexander Flossdorf [2]; Andre Georg Holzner [3]; Andrea Petrucci [1]; Andrei Cristian Spataru [1]; Attila Racz [1]; Aymeric Arnaud Dupont [1]; Christian Deldicque [1]; Christian Hartl [1]; Christoph Paus [4]; Christoph Schwick [1]; Denis Shpakov [5]; Dominique Gigi [1]; Emilio Meschi [1]; Frank Glege [1]; Frans Meijers [1]; Gerry Bauer [4]; Giovanni Polese [1]; Hannes Sakulin [1]; James Branson [6]; Jeroen Hegeman [1]; Jose Antonio Coarasa Perez [1]; Konstanty Sumorok [4]; Lorenzo Masetti [1]; Luciano Orsini [1]; Marc Dobson [1]; Marco Pieri [3]; Matteo Sani [3]; Matthew Bowen [7]; Michal Simon ; Olivier Raginel [4]; Remi Mommsen [8]; Samim Erhan [9]; Sebastian Bukowiec [1]; Sergio Cittolin [3]; Ulf Behrens [2]; Vivian O'Dell [10]; Yi Ling Hwong [1]

[1] *CERN*

[2] *Deutsches Elektronen-Synchrotron (DE)*

[3] *Univ. of California San Diego (US)*

[4] *Massachusetts Inst. of Technology (US)*

[5] *Fermi National Accelerator Lab. (Fermilab)-Unknown-Unknown*

[6] *UC San Diego*

[7] *University of the West of England*

[8] *Fermi National Accelerator Lab. (US)*

[9] *Univ. of California Los Angeles (US)*

[10] *Fermi National Accelerator Laboratory (FNAL)*

Corresponding Author: robert.gomez-reino@cern.ch

The Compact Muon Solenoid (CMS) is a CERN multi-purpose experiment that exploits the physics of the Large Hadron Collider (LHC). The Detector Control System (DCS) ensures a safe, correct and efficient experiment operation, contributing to the recording of high quality physics data. The DCS is programmed to automatically react to the LHC changes. CMS sub-detector's bias voltages are set depending on the machine mode and particle beam conditions. A protection mechanism ensures that the sub-detectors are locked in a safe mode whenever a potentially dangerous situation exists. The system is supervised from the experiment control room by a single operator. A small set of screens summarizes the status of the detector from the approximately 6M monitored parameters. Using the experience of nearly two years of operation with beam the DCS automation software has been enhanced to increase the system efficiency. The automation allows now for configuration commands that can be used to automatically pre-configure hardware for given beam modes, decreasing the time the detector needs to get ready when reaching physics modes. The protection mechanism was also improved so that sub-detectors could define their own protection response algorithms allowing, for example, tolerating a small proportion of channels out of the configured safe limits. From the infrastructure point of view the DCS will be subject to big modifications in 2012. The current rack mounted control PCs will be exchanged by a redundant pair of DELL Blade systems. These blades are a high-density modular solution that incorporates servers and networking into a single chassis that provides shared power, cooling and management. This infrastructure modification will challenge the DCS software and hardware factorization capabilities since the SCADA systems running currently in individual nodes will be combined in single blades. The undergoing studies allowing for this migration together with the latest modifications are discussed in the paper.

Summary:

CMS Detector Control System is preparing a computing hardware infrastructure upgrade. This upgrade will provide CMS with a highly compact and redundant controls computing system. There is a big impact in the architecture of the SCADA systems and the challenges, solutions and undergoing studies are presented in the paper.

Poster Session / 190

The PhEDEx next-gen website

Author: Tony Wildish[1]

Co-authors: Chih-Hao Huang [2]; Nicolo Magini [3]; Paul Rossman [4]

[1] *Princeton University (US)*

[2] *Fermi National Accelerator Laboratory*

[3] *CERN*

[4] *Fermi National Accelerator Lab. (US)*

Corresponding Author: tony.wildish@cern.ch

PhEDEx is the data-transfer management solution written by CMS. It consists of agents running at each site, a website for presentation of information, and a web-based data-service for scripted access to information.

The website allows users to monitor the progress of data-transfers, the status of site agents and links between sites, and the overall status and behaviour of everything about PhEDEx. It also allows uses to make and approve requests for data-transfers and for deletion of data. It is the main point-of-entry for all users wishing to interact with PhEDEx.

For several years, the website has consisted of a single perl program with about 10K SLOC. This program has limited capabilities for exploring the data, with only coarse filtering capabilities and no context-sensitive awareness. Graphical information is presented as static images, generated on the

server, with no interactivity. It is also not well connected to the rest of the PhEDEx codebase, since much of it was written before the data-service was developed. All this makes it hard to maintain and extend.

We are re-implementing the website to address these issues. The UI is being rewritten in Javascript, replacing most of the server-side code. We are using the YUI toolkit to provide advanced features and context-sensitive interaction, and will adopt a Javascript charting library for generating graphical representations client-side. This relieves the server of much of its load, and automatically improves server-side security. The Javascript components can be re-used in many ways, allowing custom pages to be developed for specific uses. In particular, standalone test-cases using small numbers of components make it easier to debug the Javascript than it is to debug a large server program.

Information about PhEDEx is accessed through the PhEDEx data-service, since direct SQL is not available from the clients browser. This provides consistent semantics with other, externally written monitoring tools, which already use the data-service. It also reduces redundancy in the code, yielding a simpler, consolidated codebase

Poster Session / 191

Combining virtualization tools for a dynamic, distribution agnostic grid environment for ALICE grid jobs in Scandinavia

Author: Boris Wagner[1]

Co-author: Bjarte Kileng [2]

[1] *University of Bergen (NO)*

[2] *Bergen University College (NO)*

Corresponding Author: boris.wagner@cern.ch

The Nordic Tier-1 for LHC is distributed over several, sometimes smaller, computing centers. In order to minimize administration effort, we are interested in running different grid jobs over one common grid middleware. ARC is selected as the internal middleware in the Nordic Tier-1. At the moment ARC has no
mechanism of automatic software packaging and deployment. The AliEn grid middleware, used by ALICE provides this functionality. We are investigating the possibilities to use modern virtualization technologies to make these
capabilities available for ALICE grid jobs on ARC.

The CernVM project is developing a virtual machine that can provide a common analysis environment for all LHC experiments. One of our interests is to investigate the use of CernVM as a base setup for a dynamical grid environment capable of running grid jobs. For this, performance comparisons between different virtualization technologies have been conducted.

CernVM needs an existing virtualization infrastructure, which is not always existing or wanted at some computing sites. To increase the possible application of dynamical grid environments to those sites, we describe several possibilities that are less invasive and have less specific Linux distribution requirements, at the cost of lower performance.

Different tools like user-mode Linux (UML), micro Linux distributions, a new software packaging project by Stanford university (CDE) and CernVM are under investigation for their invasiveness, distribution requirements and performance. Comparisons between the different methods with solutions that are closer to the
hardware will be presented.

Software Engineering, Data Stores and Databases / 192

Artificial Intelligence in the service of system administrators

Author: Christophe Haen[1]

Co-authors: Enrico Bonaccorsi [2]; Niko Neufeld [2]; Vincent BARRA [3]

[1] *Univ. Blaise Pascal Clermont-Fe. II (FR)*

[2] *CERN*

[3] *LIMOS, UMR 6158 CNRS, Univ. Blaise Pascal*

Corresponding Author: christophe.denis.haen@cern.ch

The LHCb online system relies on a large and heterogeneous IT infrastructure made from thousands of servers on which many different applications are running. They run a great variety of tasks : critical ones such as data taking and secondary ones like web servers. The administration of such a system and making sure it is working properly represents a very important workload for the small expert-operator team.
Research has been performed to try to automatize (some) system administration tasks, starting in 2001 when IBM defined the so-called "self objectives"supposed to lead to "autonomic computing" . In this context, we present a framework that makes use of artificial intelligence and machine learning to monitor and diagnose at a low level and in a non intrusive way Linux-based systems and their interaction with software. Moreover, the multi agent approach we use, coupled with a "object oriented paradigm" architecture should increase a lot our learning speed, and highlight relations between problems.

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Poster Session / 193

Methods to quantify the performance of the primary vertex reconstruction in the ATLAS experiment under high luminosity conditions

Authors: Andreas Wildauer[1]; Federico Meloni[2]; Kirill Prokofiev[3]; Simone Pagan Griso[4]

[1] *Universidad de Valencia (ES)*

[2] *Università degli Studi e INFN Milano (IT)*

[3] *New York University (US)*

[4] *Lawrence Berkeley National Lab. (US)*

Corresponding Authors: kirill.prokofiev@cern.ch, andreas.wildauer@cern.ch, simone.pagan.griso@cern.ch, federico.meloni@cern.ch

Presented in this contribution are methods currently developed and used by the ATLAS collaboration to measure the performance of the primary vertex reconstruction algorithms. These methods quantify the amount of additional pile up interactions and help to identify the hard scattering process (the so called primary vertex) in the proton-proton collisions with high accuracy. The correct identification of the primary vertex and
knowledge of the amount of pile up per bunch crossing is crucial for many physics analyses. With the increasing instantaneous luminosity at the LHC additional effects like splitting one vertex into many or reconstructing several pile up interactions as one become sizable effects. Statistical methods based on data and Monte Carlo simulation are applied to disentangle the different contributions. The mathematical methods, their software implementation and comparisons with independent luminosity measurements are presented.

Summary:

Because of the increasing amount of the pile up interactions, current and future LHC conditions represent a high challenge for data reconstruction and analysis. Presented in this contribution is the discussion of mathematical and computing methods to quantify the performance of primary vertex reconstruction in ATLAS. Methods to disentangle various pile up effects are presented, the influence of the these effects on data analysis is evaluated. The approaches to the vertex reconstruction in the high luminosity environment are shown.

Event Processing / **194**

Study of a Fine Grained Threaded Framework Design

Author: Christopher Jones[1]

[1] *Fermi National Accelerator Lab. (US)*

Corresponding Author: cdj@fnal.gov

Traditionally, HEP experiments exploit the multiple cores in a CPU by having each core process one event. However, future PC designs are expected to use CPUs which double the number of processing cores at the same rate as the cost of memory falls by a factor of two. This effectively means the amount of memory per processing core will remain constant. This is a major challenge for LHC processing frameworks since the LHC is expected to deliver more complex events (e.g. greater pile-up events) in the coming years while the LHC experiment's frameworks are already memory constrained. Therefore in the not so distant future we may need to be able to efficiently use multiple cores to process one event. In this presentation we will discuss a design for an HEP processing framework which can allow very fine grained parallelization within one event as well as supporting processing multiple events simultaneously while minimizing the memory footprint of the job. The design is built around the libdispatch framework created by Apple Inc. (a port for Linux is available) whose central concept is the use of task queues. This design also accommodates the reality that not all code will be thread safe and therefore allows one to easily mark modules or sub parts of modules as being thread unsafe. In addition, the design efficiently handles the requirement that events in one run must all be processed before starting to process events from a different run. After explaining the design we will provide measurements from simulating different processing scenarios where the CPU times used for the simulation are drawn from CPU times measured from actual CMS event processing. Our results have shown this design to be very promising and will be further pursued by CMS.

Poster Session / **195**

High-performance scalable information service for the ATLAS experiment.

Author: Serguei Kolos[1]

Co-author: Giuseppe Avolio [1]

[1] *University of California Irvine (US)*

Corresponding Authors: giuseppe.avolio@cern.ch, serguei.kolos@cern.ch

The ATLAS experiment is being operated by highly distributed computing system which is constantly producing a lot of status information which is used to monitor the experiment operational conditions as well as to access the quality of the physics data being taken. For example the ATLAS High Level Trigger(HLT) algorithms are executed on the online computing farm consisting from about 2000 nodes. Each HLT algorithm is producing few thousands histograms, which have to be summed over the whole cluster and carefully analysed in order to properly tune the event rejection.

In order to handle all such non-physics data the Information Service (IS) facility has been developed in the scope of the ATLAS TDAQ project. The IS provides high-performance scalable solution for information exchange in distributed environment. In the course of an ATLAS data taking session the IS handles about hundred gigabytes of information which is being constantly updated with the update interval varying from a second to few tens of seconds. IS provides access to any information item on request as well as distributing notification to all the information subscribers. In both cases the IS clients receive information in less then 1ms after it was updated. IS can handle arbitrary type of information including histograms produced by the HLT applications and provides C++, Java and Python API. The Information Service is a primarily and in most cases a unique source of information for the majority of the online monitoring analysis and GUI applications, used to control and monitor the ATLAS experiment.
Information Service provides streaming functionality allowing efficient replication of all or part of the managed information. This functionality is used to duplicate the subset of the ATLAS monitoring data to the CERN public network with the latency of the order of 1ms, allowing efficient real-time monitoring of the data taking from outside the protected ATLAS network. Each information item in IS has an associated URL which can be used to access that item online via HTTP protocol. This functionality is being used by many online monitoring applications which can run in a WEB browser, providing real-time monitoring information about ATLAS experiment over the globe.
This paper will describe design and implementation of the IS and present performance results which have been taken in the ATLAS operational environment.

Poster Session / 196

Upgrade and integration of the configuration and monitoring tools for the ATLAS Online farm

Author: Sergio Ballestrero[1]

Co-authors: Alexandr Zaytsev [2]; Diana Scannicchio [3]; Franco Brasolin [4]; Georgiana Lavinia Darlea [5]; Irina Dumitru [6]; Liviu Valsan [6]; Matthew Shaun Twomey [7]

[1] *University of Johannesburg (ZA)*

[2] *Budker Institute of Nuclear Physics (RU)*

[3] *University of California Irvine (US)*

[4] *Universita e INFN (IT)*

[5] *Polytechnic University of Bucharest (RO)*

[6] *University of Bucharest (RO)*

[7] *University of Washington (US)*

Corresponding Authors: georgiana.lavinia.darlea@cern.ch, sergio.ballestrero@cern.ch

The ATLAS Online farm is a non-homogeneous cluster of more than 3000 PCs which run the data acquisition, trigger and control of the ATLAS detector. The systems are configured and monitored by a combination of open-source tools, such as Quattor and Nagios, and tools developed in-house, such as ConfDB.

We report on the ongoing introduction of new provisioning and configuration tools, Puppet and ConfDB v2 which are more flexible and allow automation for previously uncovered needs, and on the upgrade and integration of the monitoring and alerting tools, including the interfacing of these with the TDAQ Shifter Assistant software and their integration with configuration tools.

We discuss the selection of the tools and the assessment of their functionality and performance, and how they enabled the introduction of virtualization for selected services.

Poster Session / 197

Centralized configuration system for a large scale farm of network booted computers

Author: Liviu Valsan[1]

Co-authors: Alexandr Zaytsev [2]; Diana Scannicchio [3]; Franco Brasolin [4]; Georgiana Lavinia Darlea [5]; Irina Dumitru [1]; Matthew Shaun Twomey [6]; Sergio Ballestrero [7]

[1] *University of Bucharest (RO)*

[2] *Budker Institute of Nuclear Physics (RU)*

[3] *University of California Irvine (US)*

[4] *Universita e INFN (IT)*

[5] *Polytechnic University of Bucharest (RO)*

[6] *University of Washington (US)*

[7] *University of Johannesburg (ZA)*

Corresponding Authors: georgiana.lavinia.darlea@cern.ch, liviu.valsan@cern.ch

In the ATLAS Online computing farm, the majority of the systems are network booted - they run an operating system image provided via network by a Local File Server. This method guarantees the uniformity of the farm and allows very fast recovery in case of issues to the local scratch disks. The farm is not homogeneous and in order to manage the diversity of roles, functionality and hardware of different nodes we developed a dedicated central configuration system, ConfDB v2. We describe the design, functionality and performance of this system and its web-based interface, including its integration with CERN and ATLAS databases and with the monitoring infrastructure.

Poster Session / **198**

Tools and strategies to monitor the ATLAS online computing farm

Author: Diana Scannicchio[1]

Co-authors: Alexandr Zaytsev [2]; Franco Brasolin [3]; Georgiana Lavinia Darlea [4]; Irina Dumitru [5]; Liviu Valsan [5]; Matthew Shaun Twomey [6]; Sergio Ballestrero [7]

[1] *University of California Irvine (US)*

[2] *Budker Institute of Nuclear Physics (RU)*

[3] *Universita e INFN (IT)*

[4] *Polytechnic University of Bucharest (RO)*

[5] *University of Bucharest (RO)*

[6] *University of Washington (US)*

[7] *University of Johannesburg (ZA)*

Corresponding Authors: georgiana.lavinia.darlea@cern.ch, diana.scannicchio@cern.ch

In the ATLAS experiment the collection, processing, selection and conveyance of event data from the detector front-end electronics to mass storage is performed by the ATLAS online farm consisting of more than 3000 PCs with various characteristics. To assure the correct and optimal working conditions the whole online system must be constantly monitored. The monitoring system should be able to check up to 100000 health parameters and provide alerts on a selected subset.

In this paper we present the assessment of a new monitoring and alerting system based on Icinga. This is an open source monitoring system derived from Nagios, granting backward compatibility with already known configurations, plugins and add-ons, while providing new features. We also report on the evaluation of different data gathering systems and visualization interfaces.

Distributed Processing and Analysis on Grids and Clouds / 199

Multi-core processing and scheduling performance in CMS

Authors: Dave Evans[1]; Jose Hernandez Calama[2]; Steve Foulkes[1]

[1] *Fermi National Accelerator Lab. (US)*

[2] *Centro de Investigaciones Energ. Medioambientales y Tecn. - (ES*

Corresponding Author: chema.hernandez.calama@gmail.com

Commodity hardware is going many-core. We might soon not be able to satisfy the job memory needs per core in the current single-core processing model in High Energy Physics. In addition, an ever increasing number of independent and incoherent jobs running on the same physical hardware not sharing resources might significantly affect processing performance. It will be essential to effectively utilize the multi-core architecture.

CMS has incorporated support for multi-core processing in the event processing framework and the workload management system. Multi-core processing jobs share common data in memory, such us the code libraries, detector geometry and conditions data, resulting in a much lower memory usage than standard single-core independent jobs.

Exploiting this new processing model requires a new model in computing resource allocation, departing from the standard single-core allocation for a job. The experiment job management system needs to have control over a larger quantum of resource since multi-core aware jobs require the scheduling of multiples cores simultaneously. CMS is exploring the approach of using whole nodes as unit in the workload management system where all cores of a node are allocated to a multi-core job. Whole-node scheduling allows for optimization of the data/workflow management (e.g. I/O caching, local merging) but efficient utilization of all scheduled cores is challenging. Dedicated whole-node queues have been setup at all Tier-1 centers for exploring multi-core processing workflows in CMS.

We will present the evaluation of the performance scheduling and executing multi-core workflows in whole-node queues compared to the standard single-core processing workflows.

Poster Session / 201

Evolution of the Distributed Computing Model of the CMS experiment at the LHC

Author: Claudio Grandi[1]

[1] *INFN - Bologna*

Corresponding Author: claudio.grandi@cern.ch

The Computing Model of the CMS experiment was prepared in 2005 and described in detail in the CMS Computing Technical Design Report. With the experience of the first years of LHC data taking and with the evolution of the available technologies, the CMS Collaboration identified areas where improvements were desirable. In this work we describe the most important modifications that have been, or are being implemented in the Distributed Computing Model of CMS. The Worldwide LHC computing Grid (WLCG) Project acknowledged that the whole distributed computing infrastructure is impacted by this kind of changes that are happening in most LHC experiments and decided to create several Technical Evolution Groups (TEG) aiming at assessing the situation and developing a strategy for the future. In this work we describe the CMS view on the TEG activities as well.

Poster Session / 202

Monitor and alarm system for time-critical conditions data handling

Author: Salvatore Di Guida[1]

[1] *CERN*

Corresponding Author: salvatore.di.guida@cern.ch

With LHC producing collisions at larger and larger luminosity, CMS must be able to take high quality data and process them reliably: these tasks need not only correct conditions, but also that those datasets must be promptly available. The CMS condition infrastructure relies on many different pieces, such as hardware, networks, and services, which must be constantly monitored, and any faulty situations must be recorded, and notified with different alarm scales.

In this talk, we describe EasyMon, a fast, simple, web-based application for monitoring CMS condition infrastructure. It is based on the Nagios framework, where all checks on the different pieces of the system are implemented, and from whence the web server retrieves their status. In case of failures, the Nagios backend evaluates the severity of the issue, and sends alarms or warnings via email and/or sms to the different stakeholders identified for each piece of the infrastructure. The EasyMon GUI, finally, allows to publish the results on the web, using jQuery plugins optimized also for browsing with mobile devices, without exposing any sensitive information. In this way, all experts involved in the CMS condition operations can be easily informed of the status of the system, and take actions as soon as an incident occurs.

Poster Session / **203**

Making Connections - Networking the distributed computing system with LHCONE for CMS

Author: Daniele Bonacorsi[1]

Co-authors: Andrea Sartirana [2]; James Letts [3]; José Flix ; Nicolo Magini [4]

[1] *Universita e INFN (IT)*

[2] *Ecole Polytechnique (FR)*

[3] *Univ. of California San Diego (US)*

[4] *CERN*

Corresponding Author: daniele.bonacorsi@bo.infn.it

The LHCONE project aims to provide effective entry points into a network infrastructure that is intended to be private to the LHC Tiers. This infrastructure is not intended to replace the LHCOPN, which connects the highest tiers, but rather to complement it, addressing the connection needs of the LHC Tier-2 and Tier-3 sites which have become more important in the new less-hierarchical computing models. LHCONE is intended to grow as a robust and scalable solution for a global system serving such needs, thus reducing the load on GPN infrastructures in different nations. The CMS experiment pioneered the commissioning of data transfer links between Tier-2 sites in 2010. During 2011, and in the context of preparing for LHCONE to go into production, the CMS Computing project has launched an activity to measure in detail the performance, quality and latency of large-scale data transfers among CMS Tier-2 sites. The outcome of this activity will be presented and its impact on the design and commissioning new transfer infrastructures will be discussed.

Poster Session / **205**

Monitoring techniques and alarm procedures for CMS services and sites in WLCG

Authors: Jorge Amando Molina-Perez[1]; Jose Flix Molina[2]; Peter Kreuzer[3]

Co-authors: Andrea Sciaba [4]; Diego Da Silva Gomes [5]; Ignas Butenas [6]; Nicolo Magini [4]; Rapolas Kaselis [6]; Weizhen Wang [7]

[1] *Univ. of California San Diego (US)*

[2] *Centro de Investigaciones Energ. Medioambientales y Tecn. - (ES*

[3] *Rheinisch-Westfaelische Tech. Hoch. (DE)*

[4] *CERN*

[5] *Universidade do Estado do Rio de Janeiro (BR)*

[6] *Vilnius University (LT)*

[7] *Chinese Academy of Sciences (CN)*

Corresponding Author: jorge.amando.molina-perez@cern.ch

The CMS offline computing system is composed of more than 50 sites and a number of central services to distribute, process and analyze data worldwide. A high level of stability and reliability is required from the underlying infrastructure and services, partially covered by local or automated monitoring and alarming systems such as Lemon and SLS; the former collects metrics from sensors installed on computing nodes and triggers alarms when values are out of range, the latter measures the quality of service and warns managers when service is affected. CMS has established computing shift procedures with personnel operating worldwide from remote Computing Centers, under the supervision of the Computing Run Coordinator on duty at CERN. This dedicated 24/7 computing shift personnel is contributing to detect and react timely on any unexpected error and hence ensure that CMS workflows are carried out efficiently and in a sustained manner. Synergy among all the involved actors is exploited to ensure the 24/7 monitoring, alarming and troubleshooting of the CMS computing sites and services. We review the deployment of the monitoring and alarming procedures, and report on the experience gained throughout the first 2 years of LHC operation. We describe the efficiency of the communication tools employed, the coherent monitoring framework, the pro-active alarming systems and the proficient troubleshooting procedures that helped the CMS Computing facilities and infrastructure to operate at high reliability levels.

Poster Session / 206

CRAB3: Establishing a new generation of services for distributed analysis at CMS

Author: Daniele Spiga[1]

Co-authors: Eric Wayne Vaandering [2]; Hassen Riahi [3]; Marco Mascheroni [4]; Mattia Cinquilli [5]

[1] *CERN*

[2] *Fermi National Accelerator Lab. (US)*

[3] *Universita e INFN (IT)*

[4] *Nat. Inst. of Chem.Phys. & Biophys. (EE)*

[5] *Univ. of California San Diego (US)*

Corresponding Author: daniele.spiga@cern.ch

In CMS Computing the highest priorities for analysis tools are the improvement of the end users' ability to produce and publish reliable samples and analysis results as well as a transition to a sustainable development and operations model. To achieve these goals CMS decided to incorporate analysis processing into the same framework as the data and simulation processing. This strategy foresees that all workload tools (Tier0, Tier1, production, analysis) share a common core which allows long term maintainability as well as the standardization of the operator interfaces. The re-engineered analysis workload manager, called CRAB3, makes use of newer technologies, such as RESTful based web services, NoSQL Databases aiming to increase the scalability and reliability of the system.
As opposed to CRAB2 in CRAB3 all work is centrally injected and managed in a global queue. A

pool of agents, which can be geographically distributed, consumes work from the central services, servicing the user tasks. The new architecture of CRAB substantially changes the deployment model and operations activities. In this paper we present the implementation of CRAB3 emphasizing how the new architecture improves the workflow automation and simplifies maintainability. We will highlight, in particular, the impact of the new design on daily operations.

Poster Session / **207**

Secure Wide Area Network Access to CMS Analysis Data Using the Lustre Filesystem

Author: Dimitri Bourilkov[1]

Co-authors: Bockjoo Kim [1]; Dave Dykstra [2]; Jorge Luis Rodriguez [3]; Paul Ralph Avery [1]

[1] *University of Florida (US)*

[2] *Fermi National Accelerator Lab. (US)*

[3] *Florida International University (US)*

Corresponding Author: bourilkov@phys.ufl.edu

This paper reports the design and implementation of a secure, wide area network, distributed filesystem by the ExTENCI project, based on the Lustre filesystem. The system is used for remote access to analysis data from the CMS experiment at the Large Hadron Collider, and from the Lattice Quantum ChromoDynamics (LQCD) project. Security is provided by Kerberos authentication and authorization with additional fine grained control based on Lustre ACLs (Access Control List) and quotas. We investigate the impact of using various Kerberos security flavors on the I/O rates of CMS applications on client nodes reading and writing data to the Lustre filesystem, and on LQCD benchmarks. The clients can be real or virtual nodes. We are investigating additional options for user authentication based on user certificates. We compare the Lustre performance to those obtained with other distributed storage technologies.

Poster Session / **208**

Using Virtual Lustre Clients on the WAN for Analysis of Data from High Energy Experiments

Author: Dimitri Bourilkov[1]

Co-authors: Bockjoo Kim [1]; Paul Ralph Avery [1]

[1] *University of Florida (US)*

Corresponding Author: bourilkov@phys.ufl.edu

We describe the work on creating system images of Lustre virtual clients in the ExTENCI project, using several virtual technologies (KVM, XEN, VMware). These virtual machines can be built at several levels, from a basic Linux installation (we use Scientific Linux 5 as an example), adding a Lustre client with Kerberos authentication, and up to complete clients including local or distributed (based on CernVM-FS) installations of the full CERN and project specific software stack for typical LHC experiments. The level, and size, of the images are determined by the users on demand. Various sites and individual users can just download and use them out of the box on Linux/UNIX, Windows and Mac OS X based hosts. We compare the performance of virtual clients with that of real physical systems for typical high energy physics applications like Monte Carlo simulations or analysis of data stored in ROOT trees.

Poster Session / 209

Alert Messaging in the CMS Distributed Workload System

Author: Zdenek Maxa[1]

[1] *California Institute of Technology (US)*

Corresponding Author: zdenek.maxa@hep.caltech.edu

WMAgent is the core component of the CMS workload management system. One of the features of this job managing platform is a configurable messaging system aimed at generating, distributing and processing alerts: short messages describing a given alert-worthy informational or pathological condition. Apart from the framework's sub-components running within the WMAgent instances, there is a stand-alone application collecting alerts from all WMAgent instances running across the CMS distributed computing environment. The alert framework has a versatile design that allows for receiving alert messages also from other CMS production applications, such as PhEDEx data transfer manager. We present implementation details of the system, including its python implementation using ZeroMQ, CouchDB message storage and future visions as well as operational experiences. Inter-operation with monitoring platforms such as Dashboard or Lemon is described.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Software Engineering, Data Stores and Databases / 210

Development and Evaluation of Vectorised and Multi-Core Event Reconstruction Algorithms within the CMS Software Framework

Authors: Danilo Piparo[1]; Thomas Hauth[2]; Vincenzo Innocente[1]

[1] *CERN*

[2] *KIT - Karlsruhe Institute of Technology (DE)*

Corresponding Authors: thomas.hauth@cern.ch, danilo.piparo@cern.ch

The processing of data acquired by the CMS detector at LHC is carried out with an object-oriented C++ software framework: CMSSW. With the increasing luminosity delivered by the LHC, the treatment of recorded data requires extraordinary large computing resources, also in terms of CPU usage. A possible solution to cope with this task is the exploitation of the features offered by the latest microprocessor architectures. Modern CPUs present several vector units, the capacity of which is growing steadily with the introduction of new processor generations. Moreover, an increasing number of cores per die is offered by the main vendors, even on consumer hardware. Most recent C++ compilers provide facilities to take advantage of such innovations, either by explicit statements in the programs'sources or automatically adapting the generated machine instructions to the available hardware, without the need of modifying the existing code base. Programming techniques to implement reconstruction algorithms and optimised data structures are presented, that aim to scalable vectorization and parallelization of the calculations. One of their features is the usage of new language features of the C++11 standard. Portions of the CMSSW framework are illustrated which have been found to be especially profitable for the application of vectorization and multi-threading techniques. Specific utility components have been developed to help vectorization and parallelization. They can easily become part of a larger common library. To conclude, careful measurements are described, which show the execution speedups achieved via vectorised and multi-threaded code in the context of CMSSW.

Summary:

The processing of data acquired by the CMS detector at LHC is carried out with an object-oriented C++ software framework: CMSSW. With the increasing luminosity delivered by the LHC, the treatment of recorded data requires extraordinary large computing resources, also in terms of CPU usage. A possible solution to cope with this task is the exploitation of the features offered by the latest microprocessor architectures. Modern CPUs present several vector units, the capacity of which is growing steadily with the introduction of new processor generations. Moreover, an increasing number of cores per die is offered by the main vendors, even on consumer hardware. Most recent C++ compilers provide facilities to take advantage of such innovations, either by explicit statements in the programs' sources or automatically adapting the generated machine instructions to the available hardware, without the need of modifying the existing code base. Programming techniques to implement reconstruction algorithms and optimised data structures are presented, that aim to scalable vectorization and parallelization of the calculations. One of their features is the usage of new language features of the C++11 standard. Portions of the CMSSW framework are illustrated which have been found to be especially profitable for the application of vectorization and multi-threading techniques. Specific utility components have been developed to help vectorization and parallelization. They can easily become part of a larger common library. To conclude, careful measurements are described, which show the execution speedups achieved via vectorised and multi-threaded code in the context of CMSSW.

Poster Session / 211

RelMon: A General Approach to QA, Validation and Physics Analysis through Comparison of large Sets of Histograms

Author: Danilo Piparo[1]

[1] *CERN*

Corresponding Author: danilo.piparo@cern.ch

The estimation of the compatibility of large amounts of histogram pairs is a recurrent problem in High Energy Physics. The issue is common to several different areas, from software quality monitoring to data certification, preservation and analysis. Given two sets of histograms, it is very important to be able to scrutinize the outcome of several goodness of fit tests, obtain a clear answer about the overall compatibility, easily spot the single anomalies and directly access the concerned histogram pairs. This procedure must be automated in order to reduce the human workload, therefore improving the process of identification of differences which is usually carried out by a trained human mind. Some solutions to this problem have been proposed, but they are experiment specific. RelMon depends only on ROOT and offers several goodness of fit tests (e.g. Chi-squared or Kolmogorov-Smirnov). It produces highly readable web reports, in which aggregations of the comparisons rankings are available as well as all the plots of the single histogram overlays. The comparison procedure is fully automatic and scales smoothly towards ensembles of millions of histograms. Examples of RelMon utilisation within the regular workflows of the CMS collaboration and the advantages therewith obtained are described. Its interplay with the data quality monitoring infrastructure is illustrated as well as its role in the QA of the event reconstruction code, its integration in the CMS software release cycle process, CMS user data analysis and dataset validation.

Summary:

The estimation of the compatibility of large amounts of histogram pairs is a recurrent problem in High Energy Physics. The issue is common to several different areas, from software quality monitoring to data certification, preservation and analysis. Given two sets of histograms, it is very important to be able to scrutinize the outcome of several goodness of fit tests, obtain a clear answer about the overall compatibility, easily spot the single anomalies and directly access the concerned histogram pairs. This procedure must be automated in order to reduce the human workload, therefore improving the process of identification of differences which is usually carried out by a trained human mind. Some solutions to this problem have been proposed, but they are experiment specific. RelMon depends only on ROOT and offers several goodness of fit tests (e.g. Chi-squared or Kolmogorov-Smirnov). It produces highly readable web reports, in which aggregations of the comparisons rankings are available as well as all the plots of the single histogram overlays. The comparison procedure is fully automatic and scales smoothly towards ensembles of millions of histograms. Examples of RelMon usage within the regular workflows

of the CMS collaboration and the advantages therewith obtained are described. Its interplay with the data quality monitoring infrastructure is illustrated as well as its role in the QA of the event reconstruction code, its integration in the CMS software release cycle process, CMS user data analysis and dataset validation.

Poster Session / 212

Supporting Shared Resource Usage for a Diverse User Community: the OSG experience and lessons learned

Authors: Chander Sehgal[1]; Gabriele Garzoglio[2]; Marko Slyz[1]; Mats Rynge[3]; Tanya Levshina[1]

[1] *Fermi National Accelerator Laboratory*

[2] *FERMI NATIONAL ACCELERATOR LABORATORY*

[3] *Information Sciences Institute (ISI)*

Corresponding Author: garzoglio@fnal.gov

The Open Science Grid (OSG) supports a diverse community of new and existing users to adopt and make effective use of the Distributed High Throughput Computing (DHTC) model. The LHC user community has deep local support within the experiments. For other smaller communities and individual users the OSG provides a suite of consulting and technical services through the User Support organization. We describe these sometimes successful and sometimes not so successful experiences and analyze lessons learned that are helping us improve our services. The services offered include forums to enable shared learning and mutual support, tutorials and documentation for new technology, and troubleshooting of problematic or systemic failure modes. For new communities and users, we bootstrap their use of the distributed high throughput computing technologies and resources available on the OSG by following a phased approach. We first adapt the application and run a small production campaign on a subset of "friendly" sites. Only then we move the user to run full production campaigns across the many remote sites on the OSG, where they face new hindrances including no determinism in the time to job completion, diverse errors due to the heterogeneity of the configurations and environments, lack of support for direct login to troubleshoot application crashes, etc. We cover recent experiences with image simulation for the Large Survey Synoptic Telescope (LSST), small-file large volume data movement for the Dark Energy Survey (DES), civil engineering simulation with the Network for Earthquake Engineering Simulation (NEES), and accelerator modeling with the Electron Ion Collider group at BNL. We will categorize and analyze the use cases and describe how our processes are evolving based on lessons learned.

Poster Session / 213

The DESY Grid Centre

Author: Andreas Haupt[1]

Co-authors: Andreas Gellrich [2]; Dmitry Ozerov [3]; Kai Leffhalm [1]; Peter Wegner [2]; Yves Kemp [3]

[1] *Deutsches Elektronen-Synchrotron (DE)*

[2] *DESY*

[3] *Deutsches Elektronen-Synchrotron (DE)*

Corresponding Author: andreas.haupt@desy.de

DESY is one of the world-wide leading centers for research with particle accelerators, synchrotron light and astroparticles. DESY participates in LHC as a Tier-2 center, supports on-going analyzes of HERA data, is a leading partner for ILC, and runs the National Analysis Facility (NAF) for LHC

and ILC in the framework of the Helmholtz Alliance, Physics at the Terascale. For the research with synchrotron light major new facilities are operated and
built (FLASH, PETRA-III, and XFEL). DESY furthermore acts as Data-Tier1 centre for the Neutrino detector IceCube.

Established within the EGI-project DESY operates a Grid infrastructure which supports a number of virtual Organizations (VO), incl. ATLAS, CMS, and LHCb. Furthermore, DESY is the home for some of HEP and non-HEP VOs, such as the HERA experiments and ILC as well as photon science communities. The support of the new astroparticle physics VOs IceCube and CTA is addressed.

As the global structure of the Grid offers huge resources which are perfect for batch-like computing, DESY has set up the National Analysis Facility (NAF) which complements the Grid to allow German HEP users for efficient data analysis. The Grid Infrastructure and the NAF are based on and coupled via the data which is distributed via the Grid.

We call the conjunction of Grid and NAF the DESY Grid centre.

In the contribution to CHEP2012 we will in depth discuss the conceptional and operational aspects of our multi-VO and multi-community Grid centre and present the system set-up. We will in particular focus on the interplay of Grid and NAF and present experiences of the operations.

Poster Session / **214**

Identifying gaps in Grid middleware on fast networks with the Advanced Network Initiative

Authors: Dave Dykstra[1]; Gabriele Garzoglio[2]

[1] *Fermi National Accelerator Laboratory*

[2] *FERMI NATIONAL ACCELERATOR LABORATORY*

Corresponding Author: garzoglio@fnal.gov

By the end of 2011, a number of US Department of Energy (DOE) National Laboratories will have access to a 100 Gb/s wide-area network backbone. The ESnet Advanced Networking Initiative (ANI) project is intended to develop a prototype network, based on emerging 100 Gb/s ethernet technology. The ANI network will support DOE's science research programs. A 100 Gb/s network testbed is a key component of the ANI project. The test bed offers the opportunity for early evaluation of 100Gb/s network infrastructure for supporting the high impact data movement typical of science collaborations and experiments. In order to make effective use of this advanced infrastructure, the applications and middleware currently used by the distributed computing systems of large-scale science need to be adapted and tested within the new environment, with gaps in functionality identified and corrected.

As a user of the ANI testbed, Fermilab aims to study the issues related to end-to-end integration and use of 100 Gb/s networks for the event simulation and analysis applications of physics experiments. In this paper we discuss our findings evaluating in the high-speed environment existing HEP Physics middleware and application components, including GridFTP, Globus Online, etc. These will include possible recommendations to the system administrators, application and middleware developers on changes that would make production use of the 100 Gb/s networks, including data storage, caching and wide area access.

Online Computing / **215**

Message Correlation Analysis Tool for NOvA

Author: Qiming Lu[1]

Co-authors: James Kowalkowski [2]; Kurt Biery [3]

[1] *Fermi National Accelerator Laboratory*

[2] *Fermi National Accelerator Laboratory (FNAL)*

[3] *CMS/Fermilab*

Corresponding Author: qlu@fnal.gov

A complex running system, such as the NOvA online data acquisition, consists of a large number of distributed but closely interacting components. This paper describes a generic realtime correlation analysis and event identification engine, named Message Analyzer. Its purpose is to capture run time abnormalities and recognize system failures based on log messages from participating components. The initial design of analysis engine is driven by the DAQ of the NOvA experiment. The Message Analyzer performs filtering and pattern recognition on the log messages and reacts to system failures identified by associated triggering rules. The tool helps the system maintain a healthy running state and to minimize data corruption. This paper also describes a domain specific language that allows the recognition patterns and correlation rules to be specified in a clear and flexible way. In addition, the engine provides a plugin mechanism for users to implement specialized patterns or rules in generic languages such as C++.

Poster Session / 216

An Active CAD Geometry Handling System for MAUS Software

Author: Matthew Littlefield[1]

Co-authors: Antony Wilson [2]; Chris Rogers [3]; On Behalf of the MICE Collaboration [4]; Paul Kyberd [5]; The HEP Group Brunel University [6]

[1] *Brunel University*

[2] *STFC - Science & Technology Facilities Council (GB)*

[3] *STFC*

[4] *MICE*

[5] *Departm.of Physics(QMW-Coll.)*

[6] *Brunel*

Corresponding Author: matthew.littlefield@brunel.ac.uk

The Mice Analysis User Software (MAUS) for the Muon Ionisation Cooling Experiment (MICE) is a new simulation and analysis framework based
on best-practice software design methodologies. It replaces G4MICE as it offers new functionality and incorporates an improved design structure. A
new and effective control and management system has been created for handling the simulation geometry within MAUS . The active CAD geometry handling
system translates a great level of detail of the experiment with over twenty beam line components from CAD drawings, which accurately represent the
on-going construction of the experiment into Geometry Description Markup Language (GDML). Due to the on-going construction the CAD drawings are
altered and improved at regular intervals. This is stored on the online Configuration Database (CDB). The CDB also stores field information and
specific details of each data run conducted. The geometry handling system allows users to download either a current representation of the experiment,
a previous representation of the experiment for a particular time frame or a geometry which relates to a particular run. The download process combines
all geometric, field and run data for the users to simulate. This paper describes the design and operation of the system.

Student? Enter 'yes'. See http://goo.gl/MVv53:

Yes

Poster Session / **217**

CMS Simulation Software

Author: Sunanda Banerjee[1]

[1] *Saha Institute of Nuclear Physics (IN)*

Corresponding Author: sunanda.banerjee@cern.ch

The CMS simulation, based on the Geant4 toolkit, has been operational within the new CMS software framework for more than four years. The description of the detector including the forward regions has been completed and detailed investigation of detector positioning and material budget has been carried out using collision data. Detailed modelling of detector noise has been performed and validated with the collision data. In view of the high luminosity runs of the Large
Hadron Collider, simulation of pile-up events has become a key issue. Challenges have raised from the point of view of providing a realistic luminosity profile and modelling of out-of-time pileup events, as well as computing issues regarding memory footprint and IO access. These will be especially severe in the simulation of collision events for the LHC upgrades; a new pileup simulation architecture has been introduced
to cope with these issues.

The CMS detector has observed anomalous energy deposit in the calorimeters and there has been a substantial effort to understand these anomalous signal events present in the collision data. Emphasis has also been given to validation of the simulation code including the physics of the underlying models of Geant4. Test beam as well as collision data are used for this purpose. Measurements of mean response, resolution, energy sharing between the electromagnetic and hadron calorimeters, shower shapes for single hadrons are directly compared with predictions from Monte Carlo. A suite of performance analysis tools has been put in place and has been used to drive several optimizations to allow the code to fit the constraints posed by the CMS computing model.

Student? Enter 'yes'. See http://goo.gl/MVv53:

No

Summary:

CMS has been validating the physics models inside Geant4 using its test beam as well as collision data. Several physics lists inside the most
recent version of Geant4 provide good agreement of the energy response, resolution of $\pi^{\pm}$ and protons. More work is needed to improve
the physics for charged kaons,anti-protons and hyperons.

Electromagnetic physics in Geant4 gives a good description of shower shapes for electron and photon candidates in the collision data. Isolated charged particles are used to measure calorimeter response of hadrons as a function of particle energy. These are used to compare data with Monte
Carlo predictions. There is an impressive agreement between Geant4 predictions and data in the barrel region. The agreement worsens in the endcap region. This is currently under investigation.

Rare anomalous hits in the calorimeter can be explained using the present transport codes in Geant4. Thus bulk as well as rare events can
be handled by the physics models within Geant4.

Software Engineering, Data Stores and Databases / **218**

Comparison of the Frontier Distributed Database Caching System with NoSQL Databases

Author: Dave Dykstra[1]

[1] *Fermi National Accelerator Lab. (US)*

Corresponding Author: dwd@fnal.gov

Non-relational "NoSQL" databases such as Cassandra and CouchDB are best known for their ability to scale to large numbers of clients spread over a wide area. The Frontier distributed database caching system, used in production by the Large Hadron Collider CMS and ATLAS detector projects, is based on traditional SQL databases but also has the same high scalability and wide-area distributability for an important subset of applications. This paper compares the architectures, behavior, performance, and maintainability of the two different approaches and identifies the criteria for choosing which approach to prefer over the other.

Poster Session / 219

The CMS High Level Trigger System: Experience and Future Development

Author: Andrei Cristian Spataru[1]

Co-authors: Alexander Flossdorf [2]; Andre Georg Holzner [3]; Andrea Petrucci [1]; Attila Racz [1]; Aymeric Arnaud Dupont [1]; Christian Deldicque [1]; Christian Hartl [1]; Christoph Paus [4]; Christoph Schwick [1]; Dennis Shpakov [5]; Dominique Gigi [1]; Emilio Meschi [1]; Frank Glege [1]; Frans Meijers [1]; Gerry Bauer [4]; Giovanni Polese [1]; Hannes Sakulin [1]; James Branson [3]; Jeroen Hegeman [1]; Jose Antonio Coarasa Perez [1]; Konstanty Sumorok [4]; Lorenzo Masetti [1]; Luciano Orsini [1]; Marc Dobson [1]; Marco Pieri [3]; Matteo Sani [3]; Matthew Bowen [6]; Michal Simon ; Olivier Raginel [4]; Remi Mommsen [5]; Robert Gomez-Reino Garrido [1]; Samim Erhan [7]; Sebastian Bukowiec [1]; Sergio Cittolin [3]; Ulf Behrens [8]; Vivian O'Dell [9]; Yi Ling Hwong [1]

[1] *CERN*

[2] *DESY*

[3] *Univ. of California San Diego (US)*

[4] *Massachusetts Inst. of Technology (US)*

[5] *Fermi National Accelerator Lab. (US)*

[6] *University of the West of England*

[7] *Univ. of California Los Angeles (US)*

[8] *Deutsches Elektronen-Synchrotron (DE)*

[9] *Fermi National Accelerator Laboratory (FNAL)*

Corresponding Author: andrei.cristian.spataru@cern.ch

The CMS experiment at the LHC features a two-level trigger system. Events accepted by the first level trigger, at a maximum rate of 100 kHz, are read out by the Data Acquisition system (DAQ), and subsequently assembled in memory in a farm of computers running a software high-level trigger (HLT), which selects interesting events for offline storage and analysis at a rate of order few hundred Hz. The HLT algorithms consist of sequences of offline-style reconstruction and filtering modules, executed on a farm of 0(10000) CPU cores built from commodity hardware. Experience from the operation of the HLT system in the collider run 2010/2011 is reported. The current architecture of the CMS HLT, its integration with the CMS reconstruction framework and the CMS DAQ, are discussed in the light of future development. The possible short- and medium-term evolution of the HLT software infrastructure to support extensions of the HLT computing power, and to address remaining performance and maintenance issues, are discussed.

Poster Session / 220

Operational Experience with the Frontier System in CMS

Authors: Ran Du[1]; Weizhen Wang[1]

Co-authors: Barry Jay Blumenfeld [2]; Dave Dykstra [3]; Peter Kreuzer [4]

[1] *Chinese Academy of Sciences (CN)*

[2] *Johns Hopkins University (US)*

[3] *Fermi National Accelerator Lab. (US)*

[4] *Rheinisch-Westfaelische Tech. Hoch. (DE)*

Corresponding Author: dwd@fnal.gov

The Frontier framework is used in the CMS experiment at the LHC to deliver conditions data to processing clients worldwide, including calibration, alignment, and configuration information. Each of the central servers at CERN, called a Frontier Launchpad, uses tomcat as a servlet container to establish the communication between clients and the central Oracle database. HTTP-proxy squid servers, located close to clients, cache the responses to queries in order to provide high performance data access and to reduce the load on the central Oracle database. Each Frontier Launchpad also has its own reverse-proxy squid for caching. The three central servers have been delivering about 10 million responses every day since the LHC startup, containing about 60 GB data in total, to more than one hundred Squid servers located worldwide, with an average response time on the order of 10 milliseconds. The squid caches deployed worldwide process many more requests per day, over 700 million, and deliver over 40 TB of data. Several monitoring tools of the tomcat log files, the accesses of the squid on the central Launchpad server, and the availability of remote squids have been developed to guarantee the performance of the service and make the system easily maintainable. Following a brief introduction of the Frontier framework, we describe the performance of this highly reliable and stable system, detail monitoring concerns and their deployment, and discuss the overall operational experience from the first two years of LHC data-taking.

Poster Session / 221

Maintaining and improving the control and safety systems for the Electromagnetic Calorimeter of the CMS experiment

Authors: Diogo Raphael Da Silva Di Calafiori[1]; Guenther Dissertori[1]; Oliver Holme[1]; Serguei Zelepukin[2]; Werner Lustermann[1]

[1] *Eidgenoessische Tech. Hochschule Zuerich (CH)*

[2] *University of Wisconsin (US)*

Corresponding Author: diogo.di.calafiori@cern.ch

This paper presents the current architecture of the control and safety systems designed and implemented for the Electromagnetic Calorimeter (ECAL) of the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC). A complete evaluation of both systems performance during all CMS physics data taking periods is reported, with emphasis on how software and hardware solutions have been used to overcome limitations whilst maintaining and improving reliability and robustness. The outcomes of the CMS ECAL DCS Software Analysis Project were a fundamental step towards the integration of all control system applications and the consequent piece-by-piece software improvements allowed a smooth transition to the latest revision of the system. The ongoing task of keeping the system in-line with the CMS DCS standards, as well as with new hardware technologies and software platforms is discussed. The structure of the comprehensive support service with detailed incident logging is presented in addition to a complete test setup used for reproducing failures and for testing solutions prior to deployment into production. A correlation between the acquired experience, the development of new software tools and a reduction in the DCS support load is highlighted.

Poster Session / **222**

The benefits and challenges of sharing glidein factory operations across nine time zones between OSG and CMS

Authors: Frank Wuerthwein[1]; Ignas Butenas[2]; Igor Sfiligoi[3]; Jeffrey Michael Dost[3]; Jose Hernandez Calama[4]; José Flix[None]; Marian Zvada[5]; Peter Kreuzer[6]; Rob Quick[7]; Scott Werner Teige[8]

[1] *Univ. of California San Diego (US)*

[2] *Vilnius University (LT)*

[3] *University of California San Diego*

[4] *Centro de Investigaciones Energ. Medioambientales y Tecn. - (ES*

[5] *KIT - Karlsruhe Institute of Technology (DE)*

[6] *Rheinisch-Westfaelische Tech. Hoch. (DE)*

[7] *OSG - Indiana University*

[8] *Indiana University (US)*

Corresponding Author: isfiligoi@ucsd.edu

OSG has been operating for a few years at UCSD a glideinWMS factory for several scientific communities, including CMS analysis, HCC and GLOW. This setup worked fine, but it had become a single point of failure. OSG thus recently added another instance at Indiana University, serving the same user communities. Similarly, CMS has been operating a glidein factory dedicated to reprocessing activities at Fermilab, with similar results. Recently, CMS decided to host another glidein factory at CERN, to increase the availability of the system, both for analysis, MC and reprocessing jobs. Given the large overlap between this new factory and the three factories in the US, and given that CMS represents a significant fraction of glideins going through the OSG factories, CMS and OSG formed a common operations team that operates all of the above factories. The reasoning behind this arrangement is that most operational issues stem from Grid-related problems, and are very similar for all the factory instances. Solving a problem in one instance thus very often solves the problem for all of them. This talk presents the operational experience of how we address both the social and technical issues of running multiple instances of a glideinWMS factory with operations staff spanning multiple time zones on two continents.

Poster Session / **224**

Data storage accounting and verification in LHC experiments

Authors: Cedric Serfon[1]; Elisa Lanciotti[2]; Natalia Ratnikova[3]

Co-authors: Alberto Sanchez Hernandez [4]; Chih-Hao Huang [5]; Nicolo Magini [2]; Tony Wildish [6]; xiaomei zhang [7]

[1] *Ludwig-Maximilians-Univ. Muenchen (DE)*

[2] *CERN*

[3] *KIT - Karlsruhe Institute of Technology (DE)*

[4] *Centro Invest. Estudios Avanz. IPN (MX)*

[5] *Fermi National Accelerator Laboratory*

[6] *Princeton University (US)*

[7] *IHEP,Beijing*

Corresponding Author: natalia.ratnikova@cern.ch

All major experiments at Large Hadron Collider (LHC) need to measure real storage usage at the Grid sites. This information is equally important for the resource management, planning, and operations.

To verify consistency of the central catalogs, experiments are asking sites to provide full list of files they have on storage, including size, checksum, and other file attributes. Such storage dumps provided at regular intervals give a realistic view of the storage resource usage by the experiments. Regular monitoring of the space usage and data verification serve as additional internal checks of the system integrity and performance. Both the importance and the complexity of these tasks increase with the constant growth of the total data volumes during the active data taking period at the LHC.

Developed common solutions help to reduce the maintenance costs both at the large Tier-1 facilities supporting multiple virtual organizations, and at the small sites that often lack manpower.

We discuss requirements and solutions to the common tasks of data storage accounting and verification, and present experiment-specific strategies and implementations used within the LHC experiments according to their computing models.

Summary:

Comparative analysis of the CMS, ATLAS, and LHCb solutions to the storage accounting and verification

Poster Session / **225**

Computing at Tier-3 sites in CMS

Author: Robert Snihur[1]

[1] *University of Nebraska (US)*

Corresponding Author: robert.snihur@cern.ch

There are approximately 60 Tier-3 computing sites located on campuses of collaborating institutions in CMS. We describe the function and architecture of these sites, and illustrate the range of hardware and software options. A primary purpose is to provide a platform for local users to analyze LHC data, but they are also used opportunistically for data production. While Tier-3 sites vary widely in size (number of nodes, users, support personnel), there are some common features. A site typically has a few nodes reserved for interactive use and to provide services such as an interface to the GRID. The remainder of the nodes are usually available for running CPU intensive batch jobs; a future plan will allow jobs to flock to other clusters on campus. In addition, data storage systems may be provided and we discuss several models in use, including the new paradigm of a diskless site with wide area access to data via a global XROOTD redirector. Compared to Tier-1 and Tier-2 sites, the Tier-3 sites are highly flexible and are designed for easy operation. Their ultimate configuration balances cost, performance, and reliability.

Online Computing / **226**

Modeling event building architecture for the triggerless data acquisition system for PANDA experiment at the HESR facility at FAIR/GSI

Author: Krzysztof Korcyl[1]

Co-authors: Igor Konorov [2]; Lars Schmitt [3]; Wolfgang Kuehn [4]

[1] *Polish Academy of Sciences (PL)*

[2] *Technische Universitat Munchen*

[3] *GSI Darmstadt*

[4] *Justus-Liebig-Universitaet Giessen (DE)*

Corresponding Author: krzysztof.korcyl@cern.ch

A novel architecture is being proposed for the data acquisition and trigger system for PANDA experiment at the HESR facility at FAIR/GSI. The experiment will run without the hardware trigger signal and use timestamps to correlate detector data from a given time window. The broad physics program in combination with high rate of 2 10ˆ7 interactions require very selective filtering algorithms which access information from almost all detectors. Therefore the effective filtering will happen later than it used to in today's systems ie after the event building.
To assess that, the complete architecture will be built of two stages: the data concentrator stage providing event building and the rate reduction stage. For the former stage, which allows to switch 100 GB/s of event fragments to perform event building, we propose two layers of ATCA crates filled with compute nodes - modules designed at University of Giessen for trigger and data acquisition systems. Each board is equipped with 5 Virtex4 FX60 FPGAs and high bandwidth connectivity is provided by four Gbit Ethernet links and 8 additional optical links connected to RocketIO ports.

Using the SystemC as the modeling platform we designed simplified models of the components of the architecture and demonstrated expected throughput. We also show impact of some architectural choices and key parameters on the architecture's performance.

Summary:

Key words: SystemC, modeling real-time systems, data acquisition and trigger systems

Poster Session / **227**

The WorkQueue project - a task queue for the CMS workload management system

Authors: Seangchan Ryu[1]; Stuart Wakefield[2]

Co-authors: Dave Evans [3]; Matthew Norman [4]; Simon Metson [5]; Stephen Foulkes [6]; Zdenek Maxa [7]

[1] *Fermilab*

[2] *Imperial College London*

[3] *Fermi National Accelerator Lab. (US)*

[4] *University of California at San Diego*

[5] *University of Bristol (GB)*

[6] *Fermi National Accelerator Lab. (Fermilab)*

[7] *California Institute of Technology (US)*

Corresponding Author: stuart.wakefield@imperial.ac.uk

We present the development and first experience of a new component (termed WorkQueue) in the CMS workload management system. This component provides a link between a global request system (Request Manager) and agents (WMAgents) which process requests at compute and storage resources (known as sites). These requests typically consist of creation or processing of a data sample (possibly terabytes in size).

Unlike the standard concept of a task queue, the WorkQueue does not contain fully resolved work units (known typically as jobs in HEP). This would require the WorkQueue to run computationally heavy algorithms that are better suited to run in the WMAgents. Instead the request specifies an algorithm that the WorkQueue uses to split the request into reasonable size chunks (known as elements). An advantage of performing lazy evaluation of an element is that expanding datasets can be accommodated by having job details resolved as late as possible.

The WorkQueue architecture consists of a global WorkQueue which obtains requests from the request system, expands them and forms an element ordering based on the request priority. Each WMAgent contains a local WorkQueue which buffers work close to the agent, this overcomes temporary unavailability of the global WorkQueue and reduces latency for an agent to begin processing. Elements are pulled from the global WorkQueue to the local WorkQueue and into the WMAgent based on the estimate of the amount of work within the element and the resources available to the agent.

WorkQueue is based on CouchDB, a document oriented no-sql database. WorkQueue uses the features of CouchDB (map/reduce views, bi-directional replication between distributed instances) to provide a scalable distributed system for managing large queues of work.

The project described here represents an improvement over the old approach to workload management in CMS which involved individual operators feeding requests into agents. This new approach allows for a system where individual WMAgents are transient and can be added or removed from the system as needed.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Summary:

We present the development and first experience of a new component (termed WorkQueue) in the CMS workload management system. This component provides a link between a global request system (Request Manager) and agents (WMAgents) which process requests at compute and storage resources (known as sites). These requests typically consist of creation or processing of a data sample (possibly terabytes in size).

WorkQueue is based on CouchDB, a document oriented no-sql database. WorkQueue uses the features of CouchDB (map/reduce views, bi-directional replication between distributed instances) to provide a scalable distributed system for managing large queues of work.

The project described here represents an improvement over the old approach to workload management in CMS which involved individual operators feeding requests into agents. This new approach allows for a system where individual WMAgents are transient and can be added or removed from the system as needed.

Poster Session / 228

DIRAC RESTful API

Author: Adrian Casajus Ramo[1]

Co-authors: Andrei Tsaregorodtsev [2]; Ricardo Graciani Diaz [1]

[1] *University of Barcelona (ES)*

[2] *Universite d'Aix - Marseille II (FR)*

Corresponding Author: adria@ecm.ub.es

The DIRAC framework for distributed computing has been designed as a flexible and modular solution that can be adapted to the requirements of any community. Users interact with DIRAC via

command line, using the web portal or accessing resources via the DIRAC python API. The current DIRAC API requires users to use a python version valid for DIRAC.

Some communities have developed their own software solutions for handling their specific workload, and would like to use DIRAC as their back-end to access distributed computing resources easily. Many of these solutions are not coded in python or depend on a specific python version. To solve this gap DIRAC provides a new language agnostic API that any software solution can use. This new API has been designed following the RESTful principles. Any language with libraries to issue standard HTTP queries may use it. GSI proxies can still be used to authenticate against the API services. However GSI proxies are not a widely adopted standard. The new DIRAC API also allows clients to use OAuth for delegating the user credentials to a third party solution. These delegated credentials allow the third party software to query to DIRAC on behalf of the users.

This new API will further expand the possibilities communities have to integrate DIRAC into their distributed computing models.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Poster Session / 229

Evolution of the Virtualized HPC Infrastructure of Novosibirsk Scientific Center

Authors: Alexandr Zaytsev[1]; Andrey Sukharev[1]

Co-authors: Alexey Adakin [2]; Alexey Anisenkov [3]; Dmitri Chubarov [2]; Nikolay Kuchin [4]; Sergey Belov [1]; Sergey Lomakin [4]; Victor Kaplin [1]; Vitaly Nikultsev [2]; Vladislav Kalyuzhny [5]

[1] *Budker Institute of Nuclear Physics*

[2] *Institute of Computational Technologies*

[3] *Budker Institute of Nuclear Physics (RU)*

[4] *Institute of Computational Mathematics and Mathematical Geophysics*

[5] *Novosibirsk State University*

Corresponding Author: alexey.anisenkov@cern.ch

Novosibirsk Scientific Center (NSC), also known worldwide as Akademgorodok, is one of the largest Russian scientific centers hosting Novosibirsk State University (NSU) and more than 35 research organizations of the Siberian Branch of Russian Academy of Sciences including Budker Institute of Nuclear Physics (BINP), Institute of Computational Technologies, and Institute of Computational Mathematics and Mathematical Geophysics (ICM&MG). Since each institute has specific requirements on the architecture of computing farms involved in its research field, currently we've got several computing facilities hosted by NSC institutes, each optimized for the particular set of tasks, of which the largest are the NSU Supercomputer Center, Siberian Supercomputer Center (ICM&MG), and a Grid Computing Facility of BINP. A dedicated optical network with the initial bandwidth of 10 Gbps connecting these three facilities was built in order to make it possible to share the computing resources among the research communities, thus increasing the efficiency of operating the existing computing facilities and offering a common platform for building the computing infrastructure for future scientific projects. Unification of the computing infrastructure is achieved by extensive use of virtualization technology based on XEN and KVM platforms. Our contribution gives a thorough review of the recent developments, present status and future plans for the NSC virtualized computing infrastructure focusing on its consolidation for the prospected deployment on other remote supercomputer sites and its applications for handling everyday data processing tasks of HEP experiments being carried out at BINP, the KEDR experiment in particular. We also present the results obtained while evaluating performance and scalability of the virtualized infrastructure following multiple hardware upgrades of the computing facilities involved over the last 2 years.

Poster Session / 230

An optimization of the ALICE XRootD storage cluster at the Tier-2 site in Czech Republic

Authors: Dagmar Adamova[1]; Jiri Horky[2]

[1] *Nuclear Physics Institute of the AS CR Prague/Rez*

[2] *Institute of Physics of the AS CR Prague*

Corresponding Authors: dagmar.adamova@cern.ch, horky@fzu.cz

ALICE, as well as the other experiments at the CERN LHC, has been building a distributed data management infrastructure since 2002. Experience gained during years of operations with different types of storage managers deployed over this infrastructure has shown that the most adequate storage solution for ALICE is the native XRootD manager developed within a CERN - SLAC collaboration. The XRootD storage clusters exhibit higher stability and availability in comparison with other storage solutions and demonstrate a number of other advantages like support of high speed WAN data access or no need for maintaining complex databases. Two of the operational charasteristics of XRootD data servers are a relatively high number of open sockets and a high Unix load. In this contribution we would like to describe our experience with the tuning/optimization of machines hosting the XRootD servers which are part of the ALICE storage cluster at the Tier-2 WLCG site in Prague, Czech Republic. The optimization procedure, in addition to boosting the read/write performance of the servers, also resulted in a reduction of the Unix load.

Poster Session / 231

Multiple-view, multiple-selection visualization of simulation geometry in CMS

Authors: Alja Mrak Tadel[1]; Matevz Tadel[1]

Co-authors: Avi Yagil [1]; Christopher Jones [2]; Dmytro Kovalskyi [3]; Giulio Eulisse [2]; Ianna Osborne [2]; Lothar A.T. Bauerdick [2]; Thomas Mc Cauley [2]

[1] *Univ. of California San Diego (US)*

[2] *Fermi National Accelerator Lab. (US)*

[3] *Univ. of California Santa Barbara (US)*

Corresponding Authors: alja.mrak.tadel@cern.ch, matevz.tadel@cern.ch, thomas.mccauley@cern.ch

Fireworks, the event-display program of CMS, was extended with an advanced geometry visualization package. ROOT's TGeo geometry is used as internal representation, shared among several geometry views. Each view is represented by a GUI list-tree widget, implemented as a flat vector to allow for fast searching, selection, and filtering by material type, node name, and shape type. Display of logical and physical volumes is supported. Color, transparency, and visibility flags can be modified for each node or for a selection of nodes. Further operations, like opening of a new view or changing of the root node, can be performed via a context menu. Node selection and graphical properties determined by the list-tree view can be visualized in any 3D graphics view of Fireworks. As each 3D view can display any number of geometry views, a user is free to combine different geometry-view selections within the same 3D view. Node-selection by proximity to a given point is possible. A visual clipping box can be set for each geometry view to limit geometry drawing into a specified region. Visualization of geometric overlaps, as detected by TGeo, is also supported.

The geometry visualization package is used for detailed inspection and display of simulation geometry with or without the event data. It also serves as a tool for geometry debugging and inspection, facilitating development of geometries for CMS detector upgrades and for SLHC.

Poster Session / 232

Controlled overflowing of data-intensive jobs from oversubscribed sites

Authors: Alja Mrak Tadel[1]; Brian Bockelman[2]; Daniel Bradley[3]; Frank Wuerthwein[1]; Igor Sfiligoi[4]; James Letts[1]; Kenneth Bloom[5]; Matevz Tadel[1]

[1] *Univ. of California San Diego (US)*

[2] *University of Nebraska*

[3] *University of Wisconsin - Madison*

[4] *University of California San Diego*

[5] *University of Nebraska (US)*

Corresponding Author: isfiligoi@ucsd.edu

The CMS analysis computing model was always relying on jobs running near the data, with data allocation between CMS compute centers organized at management level, based on expected needs of the CMS community. While this model provided high CPU utilization during job run times, there were times when a large fraction of CPUs at certain sites were sitting idle due to lack of demand, all while Terabytes of data were never accessed. To improve the utilization of both CPU and disks, CMS is moving toward controlled overflowing of jobs from sites that have data but are oversubscribed to others with spare CPU and network capacity, with those jobs accessing the data through real time xrootd streaming over WAN. The major limiting factor for remote data access is the ability of the source storage system to serve such data, so the number of jobs accessing it must be carefully controlled. The CMS approach to this is to implement the overflowing by means of glideinWMS, a Condor based pilot system, and by providing the WMS with the known storage limits and let it schedule jobs within those limits. This talk presents the detailed architecture of the overflow-enabled glideinWMS system, together with operational experience of the past 6 months.

Poster Session / 233

Xrootd Monitoring for the CMS experiment

Author: Matevz Tadel[1]

Co-authors: Alja Mrak Tadel [1]; Avi Yagil [1]; Brian Bockelman [2]; Daniel Charles Bradley [3]; Frank Wuerthwein [1]; Igor Sfiligoi [4]; Kenneth Bloom [5]; Lothar A.T. Bauerdick [6]; Sridhara Dasu [7]

[1] *Univ. of California San Diego (US)*

[2] *University of Nebraska*

[3] *High Energy Physics*

[4] *University of California San Diego*

[5] *University of Nebraska (US)*

[6] *Fermi National Accelerator Lab. (US)*

[7] *University of Wisconsin (US)*

Corresponding Authors: matevz.tadel@cern.ch, brian.bockelman@cern.ch

During spring and summer 2011 CMS deployed Xrootd front-end servers on all US T1 and T2 sites. This allows for remote access to all experiment data and is used for user-analysis, visualization, running of jobs at T2s and T3s when data is not available at local sites, and as a fail-over mechanism for data-access in CMSSW jobs.

Monitoring of Xrootd infrastructure is implemented on three levels. On the first level, service and data availability checks are performed by Nagios probes. The second level uses Xrootd report stream; a relatively simple stream processor is used to aggregate data from all sites and to feed the needed

data into MonALISA service and further into MonALISA repository providing web interface and long-term storage. The third level uses detailed monitoring stream of Xrootd servers configured to include detailed information about users, opened files and individual data transfers. A custom application was developed in C++ to process
this information and to, first, provide a real-time view of the system usage and, second, to store data into ROOT trees for detailed analysis. Detailed monitoring allows us to determine hot data-samples, to detect abuses of the system, including sub-optimal usage of the Xrootd protocol and ROOT tree-caching mechanism. Data from all three levels is also exported to CMS monitoring aggregators, Dashboard and Data Popularity Framework.

Poster Session / **234**

Calibration and reconstruction for the TOF system of BESIII

Author: Shengsen Sun[1]

[1] *Institute of High Energy Physics Chinese Academy of Scinences*

Corresponding Author: sunss@mail.ihep.ac.cn

The BESIII TOF detector system based on plastic scintillation counters consists of a double layer barrel and two single layer end caps. With the time calibration, the double-layer barrel TOF achieved 78ps time resolution for electrons, and end cap is about 110ps for muons. The attenuation length, effective velocity calibrations and TOF reconstruction are also described. The Kalman filter method is employed to calculate the predicted time instead of taking the track trajectory as a standard helix.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Distributed Processing and Analysis on Grids and Clouds / **235**

The HEPiX Virtualisation Working Group: Towards a "Grid of Clouds"

Author: Tony Cass[1]

[1] *CERN*

Corresponding Author: tony.cass@cern.ch

The HEPiX Virtualisation Working Group has sponsored the development of policies and technologies that permit Grid sites to safely instantiate remotely generated virtual machine images confident in the knowledge that they will be able to meet their obligations, most notably in terms of guaranteeing the accountability and traceability of any Grid Job activity at their site.

We will present the current status of the HEPiX Virtualisation Working Group technology and or links to related projects, notably StratusLab. We will also comment on the utility of our work in enabling a move from a Grid environment to a "Grid of Clouds"to provide a more responsive service to end users and reduce the service management load at participating sites.

Distributed Processing and Analysis on Grids and Clouds / **236**

The Reputation-Based Trust Model for AliEn2

Author: Jianlin Zhu[1]

Co-authors: Alina Gabriela Grigoras [2]; Costin Grigoras [2]; Daicui Zhou [3]; Federico Carminati [2]; Latchezar Betev [2]; Pablo Saiz [2]; Sergio Guinez-Molinos [4]; Steffen Schreiner [5]; guoping zhang [6]

[1] *Central China Normal University (CN)*

[2] *CERN*

[3] *Huazhong Normal University (CN)*

[4] *School of Bioinformatics Engineering, Universidad de Talca*

[5] *Technische Universitaet Darmstadt (DE)*

[6] *Huazhong Normal University*

Corresponding Author: jianlin.zhu@cern.ch

A Grid is a geographically distributed environment with autonomous sites that share resources collaboratively. In this context, the main issue within a Grid is encouraging site to site interactions, increasing the trust, confidence and reliability of the sites to share resources. To achieve this, the trust concept is vital component in every service transaction, and needs to be applied in the allocation and scheduling of jobs within a set of heterogeneous and dynamically changing resources .

In order to select a more reliable service is necessary monitoring and managing the behavior of the sites and your resources to build the trust and reputation between sites, considering the previous knowledge of their performance. All of this, for better the efficiency and delivery more and better information of the resources'performance.

As the running of the grid system for ALICE experiment, the reliability and efficiency attracts more concerns for jobs management and data management in the grid environment. We propose a Reputation-Based Trust Model (RBTM) for AliEn2 as a decision support to improve the reliability and efficiency of the grid platform.

Due to the highly dynamic, unpredictable characteristic of grid environments and the complexity of services, the Trust Model should make trust decision dynamically. With this consideration, the architecture of RBTM mainly has three types of components: Evidence Gathering, Evidence Repository and the Trust Calculation Engine. Evidence Gathering is responsible for the discovering and gathering the evidences from AliEn2 and MonALISA. Any gathered evidence will be transferred to the Evidence Repository. The Evidence Repository is the storage for all the gathered evidence. The Trust Calculation Engine aims at analyzing the evidences retrieved from the Evidence Repository and calculate the trust value for each trustee with different algorithms. There are many ways to calculate a trustee's reputation and the plug-in mechanisms are introduced to support as many algorithms as possible. The API is also provided to the applications or services for accessing the results of trust values.

The calculation of trust value in RBTM has three aspects: User Feedback, Basic Trust and Dynamic Trust. The trust value of the User Feedback is proposed by the users from their experiences with the grid system. The initial trust value is Basic Trust which is calculated from the resources contributed to the system). The trust value of Dynamic Trust is related with metrics which are collected during the running of the system.

In this paper, the Reputation-Based Trust Model for AliEn2 is introduced which can dynamically make trust decision in different situations. Finally we give the simulation results of our model and comparison with the related works.

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Poster Session / **237**

Hunting for hardware changes in data centers.

Author: Miguel Coelho Dos Santos[1]

Co-authors: Eric Bonfillou [1]; Iain Bradford Steers [1]; Imre Szebenyi ; Olof Barring [1]

[1] *CERN*

Corresponding Author: miguel.coelho.santos@cern.ch

With many servers and server parts the environment of warehouse sized data centers is increasingly complex. Server life-cycle management and hardware failures are responsible for frequent changes that need to be managed.
To manage these changes better a project codenamed “hardware hound” focusing on hardware failure trending and hardware inventory has been started at CERN.
By creating and using a hardware oriented data set - the inventory - with detailed information on servers and their parts, firmware levels, and other server related data, e.g. rack location, benchmarked processing performance and power consumption, warranty coverage, purchase order, deployment state (production, maintenance), etc; as well as tracking changes to this inventory, the project aims at, for example, being able to discover trends in hardware failure rates, e.g. lower mean time to failure of a given component in a given batch of servers. This contribution will describe the architecture of the project, the inventory data, and real life use cases.

Poster Session / **238**

Alignment Procedures for the CMS Silicon Tracker

Author: Joerg Behr[1]

Co-author: Gero Flucke [1]

[1] *Deutsches Elektronen-Synchrotron (DE)*

Corresponding Authors: joerg.behr@cern.ch, gero.flucke@cern.ch

The CMS all-silicon tracker consists of 16588 modules. Therefore its alignment procedures require sophisticated algorithms. Advanced tools of computing, tracking and data analysis have been deployed for reaching the targeted performance. Ultimate local precision is now achieved by the determination of sensor curvatures, challenging the algorithms to determine about 200k parameters simultaneously. Systematic biases in the geometry are controlled by adding further information into the alignment workflow, e.g. the mass of decaying resonances. The orientation of the tracker with respect to the magnetic field of CMS is determined with a stand-alone chi-square minimization procedure. The geometries are finally carefully validated. The monitored quantities include the basic track quantities for tracks from both collisions and cosmic muons and physics observables.

Poster Session / **239**

APEnet+: a 3-D Torus network optimized for GPU-based HPC Systems

Author: Piero Vicini[1]

Co-authors: Alessandro Lonardo [1]; Davide Rossetti [2]; Francesca Lo Cicero [1]; Francesco Simula [3]; Laura Tosoratto [4]; Pier S. Paolucci [1]; Roberto Ammendola [5]; andrea biagioni [1]; ottorino Frezza [1]

[1] *INFN Roma - Roma*

[2] *INFN - Rome*

[3] *Sapienza Universita' di Roma*

[4] *INFN*

[5] *INFN Tor Vergata - Roma*

Corresponding Authors: laura.tosoratto@roma1.infn.it, piero.vicini@roma1.infn.it

The emerging of hybrid GPU-accelerated clusters in the supercomputing landscape is a matter of fact.
In this framework we proposed a new INFN initiative, the QUonG project, aiming to deploy a high performance computing system dedicated to scientific computations leveraging on commodity multi-core processors coupled with last generation GPUs.
The multi-node interconnection system is based on a point-to-point, high performance, low latency 3-d torus network built in the framework of the APEnet+ project: it consists of an FPGA-based PCI Express board exposing six full bidirectional links running at 34 Gbps each, and implementing RDMA protocol.
In order to enable significant access latency reduction for inter-node data transfer a direct network-to-GPU interface was built. The specialized hardware blocks, integrated in the APEnet+ board, provide support for GPU-initiated communications using the so called PCI Express peer-to-peer (P2P) transactions. To this end we are strongly collaborating with NVidia GPU vendor.
The final shape of a complete QUonG deployment is an assembly of standard 42U racks, each one capable of ˜80 TFlops/rack of peak performance, at a cost of 5 KEuro/TFlops and for an estimated power consumption of 25 KW/rack.
A first reduced QUonG system prototype is expected to be delivered by the end of the year 2011.
In this talk we will report on the status of final rack deployment and on the 2012 R&D activities that will focus on performance enhancing of the APEnet+ hardware through the adoption of new generation 28nm FPGA allowing the implementation of PCI-e Gen3 host interface and the addition of new fault tolerance oriented capabilities.

Poster Session / 240

CMS reconstruction improvements for the tracking in large pile-up events

Authors: Domenico Giordano[1]; Giacomo Sguazzoni[2]

[1] *CERN*

[2] *Universita e INFN (IT)*

Corresponding Authors: giacomo.sguazzoni@cern.ch, domenico.giordano@cern.ch

The CMS tracking code is organized in several levels, known as 'iterative steps', each optimized to reconstruct a class of particle trajectories, as the ones of particles originating from the primary vertex or displaced tracks from particles resulting from secondary vertices. Each iterative step consists of seeding, pattern recognition and fitting by a kalman filter, and a final filtering and cleaning. Each subsequent step works on hits not yet associated to a reconstructed particle trajectory. The CMS tracking code underwent a major upgrade needed to make the reconstruction computing load compatible with the increasing instantaneous luminosity of LHC, resulting in a large number of primary vertices and tracks per bunch crossing. The iterative steps have been reorganized and optimized and an iterative step specialized for the reconstruction of photon conversion has been added. It is based on the innovative idea to use an existing track to build up a custom seed in the conversion hypothesis. For special event reconstruction applications, as the particle flow algorithm, it is necessary to test the possible association between a given reconstructed track and an energy deposit in calorimeters (cluster). The implementation of a k-dimensional tree in two dimensions allowed the combinatorics of links between tracks and clusters to be reduced from N*N to N*log(N), where N is the number of objects. The impact on reconstruction performances are promising and the prospects for future applications are discussed.

Poster Session / 241

An innovative seeding technique for photon conversion reconstruction at CMS

Authors: Domenico Giordano[1]; Giacomo Sguazzoni[2]

[1] *CERN*

[2] *Universita e INFN (IT)*

Corresponding Author: domenico.giordano@cern.ch

The conversion of photons into electron-positron pairs in the detector material is a nuisance in the event reconstruction of high energy physics experiments, since the measurement of the electromagnetic component of interaction products results degraded. Nonetheless this unavoidable detector effect can be also extremely useful. The reconstruction of photon conversions can be used to probe the detector material and to accurately measure soft photons that come from radiative decays in heavy flavor physics. In fact a converted photon can be measured with very high momentum resolution by exploiting the excellent reconstruction of charged tracks of a tracking detector as the one of CMS at LHC. The main issue is that photon conversion tracks are difficult to reconstruct for standard reconstruction algorithms. They are typically soft and very displaced from primary interaction vertex. An innovative seeding technique that exploits the peculiar photon conversion topology, successfully applied in the CMS track reconstruction sequence, is presented. The performances of this technique and the substantial enhancement of photon conversion reconstruction efficiency are discussed. Application examples are given.

Poster Session / 242

Major changes to the LHCb Grid computing model in year 2 of LHC data

Authors: Alexey Zhelezov[1]; Andrei Tsaregorodtsev[2]; Daniela Remenska[3]; David Bouvet[4]; Elisa Lanciotti[5]; Federico Stagni[5]; Joel Closier[5]; Marco Cattaneo[5]; Mario Ubeda Garcia[5]; Peter Clarke[6]; Philippe Charpentier[5]; Raja Nandakumar[7]; Ricardo Graciani Diaz[8]; Roberto Santinelli[5]; Stefan Roiser[5]; Victor Mendez Munoz[9]; Vincent Roger Yvan Bernardoff[10]; Vladimir Romanovskiy[11]

[1] *Ruprecht-Karls-Universitaet Heidelberg (DE)*

[2] *Universite d'Aix - Marseille II (FR)*

[3] *NIKHEF (NL)*

[4] *Universite Claude Bernard-Lyon I (FR)*

[5] *CERN*

[6] *University of Edinburgh (GB)*

[7] *Rutherford Appleton Laboratory*

[8] *University of Barcelona (ES)*

[9] *PIC*

[10] *Univ. P. et Marie Curie (Paris VI) (FR)*

[11] *Institute for High Energy Physics (RU)*

Corresponding Author: stefan.roiser@cern.ch

The increase of luminosity in the LHC during its second year of operation (2011) was achieved by delivering more protons per bunch and increasing the number of bunches. This change of running conditions required some changes in the LHCb Computing Model. The consequences of the higher pileup are a bigger event size and processing time but also the possibility for LHCb to propose and get approved a new physics program, implying an increase in the trigger rate by 50%. These changes led to shortages in the offline distributed data processing resources such an increased need of cpu

capacity by a factor 2 for reconstruction, higher storage needs at T1 sites by 70 % and subsequently problems with data throughput for file access from the storage elements. To accommodate these changes the online running conditions and the Computing Model for offline data processing had to be adapted accordingly.

This talk will describe in detail the changes implemented for the offline data processing on the Grid, relaxing the Monarc model in a first step and going beyond it subsequently. It will further describe other operational issues discovered and solved during 2011, present the performance of the system and conclude by lessons learned to further improve the data processing reliability and quality for the 2012 run. If available, first results on the computing performance from 2012 run will be presented.

Poster Session / **243**

Storage Element performance optimization for CMS analysis jobs

Author: Tomas Linden[1]

Co-authors: Gerd Behrmann [2]; Guldmyr Johan [3]; Jonas Dahlblom [3]; Kalle Happonen [4]

[1] *Helsinki Institute of Physics (FI)*

[2] *Nordic Data Grid Facility*

[3] *CSC —IT Center for Science Ltd*

[4] *Helsinki Institute of Physics HIP*

Corresponding Author: tomas.linden@helsinki.fi

Tier-2 computing sites in the Worldwide Large Hadron Collider Computing Grid (WLCG) host CPU-resources (Compute Element, CE) and storage resources (Storage Element, SE). The vast amount of data that needs to processed from the Large Hadron Collider (LHC) experiments requires good and efficient use of the available resources. Having a good CPU efficiency for the end users analysis jobs requires that the performance of the storage system is able to scale with I/O requests from hundreds or even thousands of simultaneous jobs.

In this presentation we report on the work on improving the SE performance at the Helsinki Institute of Physics (HIP) Tier-2 used for the Compact Muon Experiment (CMS) at the LHC. Statistics from CMS grid jobs are collected and stored in the CMS Dashboard for further analysis, which allows for easy performance monitoring by the sites and by the CMS collaboration. As part of the monitoring framework CMS uses the JobRobot which sends every four hours 100 analysis jobs to each site. CMS also uses the HammerCloud (HC) tool for site monitoring and stress testing and HC will replace soon replace the JobRobot. The performance of the analysis workflow submitted with JobRobot or HC can be used to track the performance due to site configuration changes, since the analysis workflow is kept the same for all sites and for months in time. The CPU efficiency of the JobRobot jobs at HIP was increased approximately by 50 % to more than 90 %, by tuning the SE and by improvements in the CMSSW and dCache software. The performance of the CMS analysis jobs improved significantly too. Similar work has been done on other CMS Tier-sites, since on average the CPU efficiency for CMSSW jobs has increased during 2011. Better monitoring of the SE allows faster detection of problems, so that the performance level can be kept high. The next storage upgrade at HIP will consist of SAS disk enclosures which can be stress tested on demand with HC workflows, to make sure that the I/O-performance is good.

Poster Session / **244**

Trying to Predict the Future - Resource Planning and Allocation in CMS

Author: Peter Kreuzer[1]

[1] *RWTH Aachen*

Corresponding Author: peter.kreuzer@cern.ch

In the large LHC experiments the majority of computing resources are provided by the participating countries. These resource pledges account for more than three quarters of the total available computing. The experiments are asked to give indications of their requests three years in advance and to evolve these as the details and constraints become clearer. In this presentation we will discuss the resource planning techniques used in CMS to predict the computing resources several years in advance. We will discuss how we attempt to implement the activities of the computing model in spreadsheets and formulas to calculate the needs. We will talk about how those needs are reflected in the 2012 running and how the planned long shutdown of the LHC in 2013 and 2014 impact the planning process and the outcome. In the end we will speculate on the computing needs in the second major run of LHC.

Poster Session / **245**

Service monitoring in the LHC experiments

Authors: Alessandro Di Girolamo[1]; Diego Da Silva Gomes[2]; Fernando Harald Barreiro Megino[3]; José Flix[None]; Peter Kreuzer[4]; Stefan Roiser[1]; Vincent Roger Yvan Bernardoff[5]

[1] *CERN*

[2] *Universidade do Estado do Rio de Janeiro (BR)*

[3] *CERN IT ES*

[4] *Rheinisch-Westfaelische Tech. Hoch. (DE)*

[5] *Univ. P. et Marie Curie (Paris VI) (FR)*

Corresponding Authors: fernando.harald.barreiro.megino@cern.ch, alessandro.di.girolamo@cern.ch

The LHC experiments' computing infrastructure is hosted in a distributed way across different computing centers in the Worldwide LHC Computing Grid and needs to run with high reliability. It is therefore crucial to offer a unified view to shifters, who generally are not experts in the services, and give them the ability to follow the status of resources and the health of critical systems in order to alert the experts whenever a system becomes unavailable.
Several experiments have chosen to build their service monitoring on top of the flexible Service Level Status (SLS) framework developed by CERN IT. Based on examples from ATLAS, CMS and LHCb, this contribution will describe the complete development process of a service monitoring instance and explain the deployment models that can be adopted. We will also describe the software package used in ATLAS Distributed Computing to send health reports through the MSG messaging system and publish them to SLS on a lightweight web server.

Poster Session / **246**

Data compression in ALICE by on-line track reconstruction and space point analysis

Author: Matthias Richter[1]

[1] *University of Oslo (NO)*

Corresponding Author: matthias.richter@cern.ch

High resolution detectors in high energy nuclear physics deliver a huge amount of data which is often a challenge for the data acquisition and mass storage. Lossless compression techniques on the level of the raw data can provide compression ratios up to a factor of 2. In ALICE, an effective compression factor of >5 for the Time Projection Chamber (TPC) is needed to reach an overall compression factor suited for data taking in Heavy Ion data-taking.

The ALICE High Level Trigger provides online calculation of the TPC clusters from the raw data, followed tracking, thus producing a fully reconstructed event. Storing the reconstructed cluster data in an appropriate compressed format for utilization in the off-line reconstruction allows to discard the original raw data of the TPC. In the presented solution, compression factors of 5 to 6 are achieved without any loss in the physics information of the event. By associating space points to reconstructed tracks, all relevant parameters can be transformed into a format suitable for huffman compression. In a first conservative approach, all reconstructed clusters are kept in the data. Enhanced data compression factors can be achieved by further analysis of the space point properties and discarding clusters which are irrelevant for the measured observables.

Data compression has been implemented for the ALICE TPC in 2011 for usage in the Heavy Ion data-taking. The generic implementation of the track model concept supports the application for other detectors. In this contribution the results for TPC data compression from the 2011 Heavy Ion run and studies for other detectors are presented.

Poster Session / **247**

Tape status and strategy at CERN

Author: German Cancio Melia[1]

Co-authors: Daniele Francesco Kruse [1]; Eric Cano [1]; Giuseppe Lo Re [1]; Steven Murray [1]; Vlado Bahyl [1]

[1] *CERN*

Corresponding Author: german.cancio.melia@cern.ch

With currently around 55PB of data stored on over 49000 cartridges, and around 2PB of fresh data coming every month, CERN's large tape infrastructure is continuing its growth. In this contribution, we will detail out the progress achieved and the ongoing steps towards our strategy of turning tape storage from a HSM environment into a sustainable long-term archiving solution. In particular, we report on the experiences gained in the production deployment of our new high-performance tape format, the optimization and reduction of random end-user access to tape-resident data, the deployment of a new media migration (repack) facility, and the review of our monitoring subsystem. We will also explain the recent infrastructure upgrades at CERN in terms of new-generation tape drives and testing/integration of new tape library models. An outlook on future plans in view of better integration with the EOS disk pool management suite will also be given.

Event Processing / **248**

Refactoring, reengineering and evolution: paths to Geant4 uncertainty quantification and performance improvement

Authors: Gabriela Hoff[1]; Maria Grazia Pia[2]; Matej Batic[3]; Steffen Hauf[4]

Co-authors: Andreas Zoglauer [5]; Chan Hyeong Kim [6]; Georg Weidenspointner [7]; Hee Seo [6]; Jason P. Hayward [8]; Marcia Begalli [9]; Markus Kuster [10]; Mincheol Han [6]; Paolo Giovanni Saracco ; Zane W. Bell [11]

[1] *CERN*

[2] *Universita e INFN (IT)*

[3] *Jozef Stefan Institute*

[4] *Technische Universitaet Darmstadt-Unknown-Unknown*

[5] *UC Berkeley*

[6] *Hanyang Univ.*

[7] *MPI Halbleiterlabor*

[8] *Univ. of Tennessee*

[9] *State Univ. Rio de Janeiro*

[10] *XFEL*

[11] *ORNL*

Corresponding Author: maria.grazia.pia@cern.ch

Quantitative results on Geant4 physics validation and computational performance are reported: they cover a wide spectrum of electromagnetic and hadronic processes, and are the product of a systematic, multi-disciplinary effort of collaborating physicists, nuclear engineers and statisticians. They involve comparisons with established experimental references in the literature and ad hoc measurements by collaborating experimental groups.

The results highlight concurrent effects of Geant4 software design and implementation on physics accuracy, computational speed and memory consumption. Prototype alternatives, which improve these three aspects, are presented: they span a variety of strategies - from refactoring and reengineering existing Geant4 code to new and significantly different approaches in physics modeling, software design and software development methods. Solutions that simultaneously contribute to both physics and computational performance improvements are highlighted.

In parallel, knowledge gaps embedded in Geant4 physics models are identified and discussed: they are due to lack of experimental data or conflicting measurements preventing the validation of the models themselves, and represent a potential source of systematic effects in detector observables.

Poster Session / 249

Characterisation of HEP database applications

Authors: Eric Grancher[1]; Mariusz Piorkowski[None]

Co-author: Anton Topurov [1]

[1] *CERN*

Corresponding Author: mariusz.piorkowski@cern.ch

Oracle-based database applications underpin many key aspects of operations for both the LHC accelerator and the LHC experiments. In addition to overall performance, predictability of response is a key requirement to ensure smooth operations—and delivering predictability requires understanding the applications from the ground up. Fortunately, the Oracle database management system provides several tools to check, measure, analyse and gather useful information. We present our experiences characterising the performance of several typical HEP database applications—performance characterisations that were used to deliver improved predictability and scalability as well as for optimising the hardware platform choice as we migrated to new hardware and Oracle 11g.

Computer Facilities, Production Grids and Networking / 250

High Performance Experiment Data Archiving with gStore

Author: Horst Göringer[1]

Co-authors: Matthias Feyerabend [1]; Serguei Sedykh [2]

[1] *GSI*

[2] *Gesellschaft fuer Schwerionenforschung mbH (GSI)*

Corresponding Author: h.goeringer@gsi.de

GSI in Darmstadt (Germany) is a center for heavy ion research. It hosts an Alice Tier2 center and is the home of the future FAIR facility. The planned data rates of the largest FAIR experiments, CBM and Panda, will be similar to those of the current LHC experiments at Cern.
gStore is a hierarchical storage system with unique name space and successfully in operation since more than fifteen years. Its core consists of several tape libraries and currently ˜20 data mover nodes connected within a SAN network. The gStore clients transfer data via fast socket connections from/to the disk cache of the data movers (˜240 TByte currently). Each data mover has also a high speed connection to the GSI lustre file system (˜3 PByte data capacity currently). The overall bandwidth between gStore (disk cache or tape) and lustre amounts to 6 GByte/s and will be duplicated in 2012. In the near future the lustre HSM functionality will be implemented with gStore.
Each tape drive is accessible from any data mover, fully transparent to the users. The tapes and libraries are managed by commercial software (IBM Tivoli Storage Manager TSM), whereas the disk cache management and the TSM and user interfaces are provided by GSI software. This provides the flexibility needed to tailor gStore according to the always developing requirements of the GSI and FAIR user communities. For Alice users all gStore data are worldwide accessible via Alice grid software.
Data streams from running experiments at GSI u(p to 500 MByte/s) are written via sockets from the event builders to gStore write cache for migration to tape. In parallel the data are also copied to lustre for online evaluation and monitoring.
As all features related to tapes and libraries are handled by TSM gStore is practically completely hardware independent. Additionally, according to the design principles gStore is fully scalable in data capacity and I/O bandwidth. Therefore we are optimistic to fulfill also the dramatically increased mass storage requirements of the FAIR experiments in 2018, which will be some orders of magnitude higher than those of today.

Poster Session / 251

Consistency between Grid Storage Elements and File Catalogs for the LHCb experiment's data

Author: Elisa Lanciotti[1]

Co-authors: Alexey Zhelezov [2]; Andrei Tsaregorodtsev [3]; Cedric Serfon [4]; Daniela Remenska [5]; David Bouvet [6]; Federico Stagni [1]; Joel Closier [1]; Marco Cattaneo [1]; Mario Ubeda Garcia [1]; Natalia Ratnikova [7]; Nicolo Magini [1]; Peter Clarke [8]; Philippe Charpentier [1]; Raja Nandakumar [9]; Ricardo Graciani Diaz [10]; Roberto Santinelli [1]; Stefan Roiser [1]; Victor Mendez Munoz [11]; Vincent Roger Yvan Bernardoff [12]; Vladimir Romanovskiy [13]

[1] *CERN*

[2] *Ruprecht-Karls-Universitaet Heidelberg (DE)*

[3] *Universite d'Aix - Marseille II (FR)*

[4] *Ludwig-Maximilians-Univ. Muenchen (DE)*

[5] *NIKHEF (NL)*

[6] *Universite Claude Bernard-Lyon I (FR)*

[7] *KIT - Karlsruhe Institute of Technology (DE)*

[8] *University of Edinburgh (GB)*

[9] *STFC - Science & Technology Facilities Council (GB)*

[10] *University of Barcelona (ES)*

[11] *Universitat de Barcelona*

[12] *Univ. P. et Marie Curie (Paris VI) (FR)*

[13] *Institute for High Energy Physics (RU)*

Corresponding Author: elisa.lanciotti@cern.ch

In the distributed computing model of WLCG Grid Storage Elements (SE) are by construction completely decoupled from the File Catalogs (FC) where the experiment's files are registered. On the basis of the experience of managing large volumes of data in such environment, inconsistencies have often happened either causing a waste of disk space, in case the data were deleted from the FC, but still physically on the SE, or serious operational problems in the opposite case, when some data registered in the FC was not found on the SE. Therefore, the LHCbDirac data management system has been equipped with a new dedicated system to ensure the consistency of the data stored on the SEs with the information reported in the FCs implementing systematic checks. Objective of the checks is to spot any inconsistency above a certain threshold, that cannot only be due to the expected latency between data upload and registration, and in such case try and identify the problematic data. The system relies on information provided by the sites who should make available to the experiment a full dump of their SEs on weekly or monthly basis.
In this talk we shall present the definition of a common format and procedure to produce the storage dumps that has been coordinated with the other LHC experiments in order to provide a solution as generic as possible that can suit all LHC experiments and will reduce the effort for the sites who are asked to provide such data. We will also present the LHCb specific implementation for checking the consistency between SEs and FC and discuss the results.

Poster Session / 252

SSD Scalability Performance for HEP data analysis using PROOF

Author: Giacinto Donvito[1]

Co-authors: Alexis Pompili [2]; Lucia Barbone [2]

[1] *INFN-Bari*

[2] *Universita e INFN (IT)*

Corresponding Author: giacinto.donvito@ba.infn.it

Nowadays the storage systems are evolving not only in size but also in terms of used technologies. SSD disks are currently introduced in storage facilities for HEP experiments and their performance is tested in comparison with standard magnetic disks.
The tests are performed by running a real CMS data analysis for a typical use case and exploiting the features provided by PROOF-Lite, that allows to distribute a huge number of events to be processed among different CPU cores in order to reduce the overall time needed to complete the analysis task. These tests are carried on comparing performances over a few computational devices typically hosted at a current Tier2/Tier3 facility.
The performance results are provided by focusing on scalability issues in terms of speed up factor and processing event rate, and can be assumed as guidelines for both the typical HEP analyst and the T2/T3 manager.
For the former in the configuration of his own analysis task
while dealing with increasing data sizes, for the latter in the implementation of interactive data analysis facility for HEP experiments while facing solutions that concern both technological and economical aspects.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Distributed Processing and Analysis on Grids and Clouds / 253

dCache, agile adoption of storage technology

Authors: Patrick Fuhrmann[1]; Paul Millar[2]

[1] *DESY*

[2] *Deutsches Elektronen-Synchrotron (DE)*

Corresponding Author: paul.millar@desy.de

For over a decade, dCache has been synonymous with large-capacity, fault-tolerant storage using commodity hardware that supports seamless data migration to and from tape. Over that time, it has satisfied the requirements of various demanding scientific user communities to store their data, transfer it between sites and fast, site-local access.

When the dCache project started, the focus was on managing a relatively small disk cache in front of large tape archives. Over the project's lifetime storage technology has changed. During this period, technology changes have driven down the cost-per-GiB of harddisks. This resulted in a shift towards systems where the majority of data is stored on disk. More recently, the availability of Solid State Disks, while not yet a replacement for magnetic disks, offers an intriguing opportunity for significant performance improvement if they can be used intelligently within an existing system.

New technologies provide new opportunities and dCache user communities' computing models are changing. The traditional data models, in which tape is used as an active storage, are being revised with tape adopting a more archival model. The symbiotic relationship between dCache and the end-users means that dCache is both driven by and facilitating these changes.

Recently, dCache introduced support for WebDAV and the NFS 4.1/pNFS protocols. This move away from bespoke protocols towards standards is the result of the availability of protocols that support large storage systems. dCache's adoption of standards allows end-users to use their favourite desktop data-transfer clients or unmodified analysis software. This keeps dCache competitive with industry solutions.

Hadoop FS (HDFS) provides an easy-to-maintain backend storage that is showing promise as an easy-to-maintain storage system. dCache is adopting HDFS as an alternative to local filesystem storage. Since HDFS doesn't offer file system semantics, integrating support into dCache provides some challenges. Once solved, this work will allow dCache integration with other storage technologies such as object stores and cloud storage.

We present a short summary of what dCache is providing in new long-term support release (the next "Golden Release") and offers a glimpse into the future of dCache with the emerging storage technology.

Poster Session / 254

Algorithms and parameters for improved accuracy in physics data libraries

Authors: Hee Seo[1]; Maria Grazia Pia[2]; Matej Batic[3]

Co-authors: Chan Hyeong Kim [1]; Lorenzo Moneta [4]; Mincheol Han [1]; Paolo Giovanni Saracco

[1] *Hanyang Univ.*

[2] *Universita e INFN (IT)*

[3] *Jozef Stefan Institute*

[4] *CERN*

Corresponding Authors: shee@hanyang.ac.kr, maria.grazia.pia@cern.ch

Physics data libraries play an important role in Monte Carlo simulation systems: they provide fundamental atomic and nuclear parameters, and tabulations of basic physics quantities (cross sections, correction factors, secondary particle spectra etc.) for particle transport.

This report summarizes recent efforts for the improvement of the accuracy of physics data libraries, concerning two complementary areas: the refinement of atomic parameters and the development of software tools for their effective management, and the investigation of interpolation algorithms used in association with physics data libraries.

Results are reported about a large scale validation analysis of atomic parameters used by major Monte Carlo systems (Geant4, EGS, MCNP, Penelope etc.); their contribution to the accuracy of simulation observables are quantitatively documented. A new atomic data management software package, which optimizes the provision of state-of-the-art atomic parameters to physics models, is illustrated. To the best of the authors' knowledge, this is the first comprehensive study in this domain.

A variety of interpolation algorithms have been developed and investigated to improve the accuracy of simulation models based on tabulated data libraries: quantitative results are reported, that illustrate the effects of interpolation algorithms on the physical and computational performance of the simulation.

Software Engineering, Data Stores and Databases / 255

Cling - The LLVM-based C++ Interpreter

Author: Vasil Georgiev Vasilev[1]

Co-authors: Axel Naumann [1]; Paul Russo [2]; Philippe Canal [2]

[1] *CERN*

[2] *FERMILAB*

Corresponding Author: vasil.georgiev.vasilev@cern.ch

Cling (http://cern.ch/cling) is a C++ interpreter, built on top of clang (http://clang.llvm.org) and LLVM (http://llvm.org). Like its predecessor CINT, cling offers an interactive, terminal-like prompt. It enables exploratory programming with rapid edit / run cycles.

The ROOT team has more than 15 years of experience with C++ interpreters, and this has been fully exploited in the design of cling. However, matching the concepts of an interpreter to a compiler library is a non-trivial task; we will explain how this is done for cling, and how we managed to implement cling as a small (10,000 lines of code) extension to the clang and llvm libraries.

The resulting features clearly show the advantages of basing an interpreter on a compiler. Cling uses clang's praised concise and easy to understand diagnostics. Building an interpreter on top of a compiler library makes the transition between interpreted and compiled code much easier and smoother. We will present the design, e.g. how cling treats the C++ extensions that used to be available in CINT. We will also present the new features, e.g. how C++11 will come to cling, and how dictionaries will be simplified due to cling. We describe the state of cling's integration in the ROOT Framework.

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Poster Session / 256

EGI Security Monitoring integration into the Operations Portal

Authors: Cyril L'Orphelin[1]; Daniel Kouril[2]; Mingchao Ma[3]

[1] *CNRS/IN2P3*

[2] *Unknown*

[3] *STFC - Rutherford Appleton Laboratory*

Corresponding Authors: kouril@ics.muni.cz, mingchao.ma@stfc.rl.ac.uk, cyril.lorphelin@cc.in2p3.fr

The Operations Portal is a central service being used to support operations in the European Grid Infrastructure: a collaboration of National Grid Initiatives (NGIs) and several European International Research Organizations (EIROs). The EGI Operation Portal is providing a single access point to operational information gathered from various sources such as site topology database, monitoring systems, user support helpdesk, grid information system, VO database and VOMS servers etc.

Significant development effort has been put in place to implement synoptic view. The single operations platform has been proved invaluable for those who involve EGI operations such as site administrators, NGI representatives, VO managers and NGI operators.

In parallel with this work, over the years, the EGI CSIRT (Computer Security Incident Response Team) has been developing security monitoring tools to monitor the infrastructure and to alert resource providers on any identified security problem. Due to the large and increasing number of resources joining the EGI e-Infrastructure it becomes more and more challenging for the EGI CSIRT to follow up all identified security issues.

In order to scale up the operation capability a security dashboard has been developed. The security dashboard integrates into the EGI Operations Portal as a module which allows resource providers' security officers and its NGI operation staff to access the monitoring results, and therefore to handle the issues directly. The dashboard aggregates the data produced by different security monitoring components and provides interfaces to its visualization. Access to the collected data is subject to strict access control so that sensitive information is accessed in a controlled manner. The integration will also allow operational security issue handling workflow to be easily incorporated into existing issue handling procedure, thus significantly reduces overall operational cost.

The paper will first briefly introduce current security monitoring framework
and its key components : Nagios and Pakiti, followed by the detail design and implementation of the security dashboard. we will also present some early experience gained with regular utilization of the security dashboard and results that have improved security of the whole environment recently.

Poster Session / 257

The Database on Demand service

Author: Ruben Domingo Gaspar Aparicio[1]

Co-authors: Daniel Gomez Blanco [1]; Dawid Wojcik [1]; Ignacio Coterillo Coz [2]

[1] *CERN*

[2] *Universidad de Cantabria (ES)*

Corresponding Author: ruben.gaspar.aparicio@cern.ch

At CERN, and probably elsewhere, centralised Oracle-database services deliver high levels of service performance and reliability but are sometimes perceived as overly rigid and inflexible for initial application development. As a consequence a number of key database applications are running on user-managed MySQL database services. This is all very well when things are going well, but the user-managed database infrastructure rarely delivers the same service levels, most notably in terms of backup and backup verification, as the centrally managed services. This weakness in backend infrastructure could have major adverse consequences in the event of, for example, a hardware failure. To address these issues, CERN has recently been exploring the possibility of supporting a "Database on Demand"service. Such a service would deliver a simple and intuitive web interface to empower users to create and exploit database instances without having to worry about the "back-end"aspects of database management.
The presentation will cover
• the rich web application, based on J2EE, that has been developed to allow users to request, start up, shutdown, reconfigure, backup and restore databases;
• the provision of database monitoring information to users;
• how we intend to handle database and operating system upgrades;
• details of how we exploit virtualisation and storage technologies to minimise our management costs; and
• possible future directions—although we have focussed on MySQL during the development phase, the architecture has been designed to be database system agnostic.

Poster Session / 258

A new development cycle of the Statistical Toolkit

Authors: Alberto Ribon[1]; Andreas Pfeiffer[1]; Maria Grazia Pia[2]; Matej Batic[3]

[1] *CERN*

[2] *Universita e INFN (IT)*

[3] *Jozef Stefan Institute*

Corresponding Authors: matej.batic@ijs.si, maria.grazia.pia@cern.ch

The Statistical Toolkit is an open source system specialized in the statistical comparison of distributions. It addresses requirements common to different experimental domains, such as simulation validation (e.g. comparison of experimental and simulated distributions), regression testing in software development and detector performance monitoring.

The first development cycles concerned the provision of a wide set of non-parametric goodness-of-fit tests for the so-called two sample problem, i.e. the comparison of two distributions. The active use of the Statistical Toolkit in real-life applications, documented in the literature, has highlighted new requirements, that are addressed by a new development cycle. The new product includes extensions of the functionality of the toolkit, refinements of existing algorithms and tools and improved usability of the system.

Various sets of statistical tests have been added to the existing collection to deal with the one sample problem (i.e. the comparison of a data distribution to a function, including tests for normality), the comparison of two-dimensional distributions, categorical analysis and the estimate of randomness. Improved algorithms and software design contribute to the robustness of the results. A simple user layer dealing with primitive data types and an improved ROOT user layer facilitate the use of the toolkit both in standalone analyses and in large scale experiments. Interface to the R package extends the native functionality of the toolkit.

An overview of the new developments is presented, along with applications to concrete experimental scenarios.

Poster Session / 259

Regression testing in the TOTEM DCS

Authors: Fernando Lucas Rodriguez[1]; Ivan Atanassov[2]; Jani Tapani Taskinen[3]; Oliver Frost[None]; Ville Tulimaki[4]

Co-author: Paul Burkimsher [1]

[1] *CERN*

[2] *Bulgarian Academy of Sciences (BG)*

[3] *Helsinki Institute of Physics (FI)*

[4] *Helsinki Institute of Physics (HIP)*

Corresponding Author: fernando.lucas.rodriguez@cern.ch

The Detector Control System of the TOTEM experiment at the LHC is built with the industrial product WinCC OA (PVSS). The TOTEM system is generated automatically through scripts using as input the detector PBS structure and pinout connectivity, archiving and alarm meta-information, and some other heuristics based on the naming conventions. When those initial parameters and code are modified to include new features, the resulting PVSS system can also included undesired side-effects.

In a daily basis, a custom developed regression testing tool takes the most recent code from a SVN repository, builds a new control system from scratch. This system is exported in a plain text format using the PVSS export tool, and compared with a system previously validated by a human. A report is sent to the developers with the differences observed, in view of validation.

This regression approach is not dependent on any development framework or methodology. It has been used successfully for several months proving to be very valuable as final validation before deploying a new production version.

Poster Session / 260

Towards higher reliability of CMS Computing Facilities

Authors: Andrea Sciaba[1]; José Flix[None]

Co-authors: Chris Brew [2]; Giuseppe Bagliesi [3]; Kenneth Bloom [4]; Peter Kreuzer [5]

[1] *CERN*

[2] *Particle Physics-Rutherford Appleton Laboratory-STFC - Science &*

[3] *Sezione di Pisa (IT)*

[4] *University of Nebraska (US)*

[5] *Rheinisch-Westfaelische Tech. Hoch. (DE)*

Corresponding Author: jose.flix.molina@cern.ch

The CMS experiment has adopted a computing system where resources are distributed worldwide in more than 50 sites. The operation of the system requires a stable and reliable behavior of the underlying infrastructure. CMS has established procedures to extensively test all relevant aspects of a site and their capability to sustain the various CMS computing workflows at the required scale. The Site Readiness monitoring infrastructure has been instrumental in understanding how the system as a whole was improving towards LHC operations, measuring the reliability of sites when running CMS activities, and providing sites with the information they need to solve eventual problems. This paper reviews the complete automation of the Site Readiness program, with the description of monitoring tools and their inclusion into the Site Status Board (SSB), the performance checks, the use of tools like HammerCloud, and the impact in improving the overall reliability of the Grid from the point

of view of the CMS computing system. Based on these results, CMS automatically excludes sites to conduct workflows, in order to maximize workflows efficiencies. The performance against these tests seen at the sites during the first years of LHC running will be as well reviewed.

Poster Session / **261**

Performance studies and improvements of CMS Distributed Data Transfers

Author: José Flix[None]

Co-authors: Andrea Sartirana [1]; Daniele Bonacorsi [2]; James Letts [3]; Nicolo Magini [4]

[1] *Ecole Polytechnique (FR)*

[2] *Universita e INFN (IT)*

[3] *Univ. of California San Diego (US)*

[4] *CERN*

Corresponding Author: jose.flix.molina@cern.ch

CMS computing needs reliable, stable and fast connections among multi-tiered computing infrastructures. CMS experiment relies on File Transfer Services (FTS) for data distribution, a low level data movement service responsible for moving sets of files from one site to another, while allowing participating sites to control the network resource usage. FTS servers are provided by Tier-0 and Tier-1 centers and used by all the computing sites in CMS, subject to established CMS and sites setup policies, including all the virtual organizations making use of the Grid resources at the site, and properly dimensioned to satisfy all the requirements for them. Managing the service efficiently needs good knowledge of the CMS needs for all kind of transfer routes, and the sharing and interference with other Virtual Organizations using the same FTS transfer managers. This contribution deals with a complete revision of all FTS servers used by CMS, customizing the topologies and improving their setup in order to keep CMS transferring data to the desired levels in a reliable and robust way, as well as complete performance studies for all kind of transfer routes, including overheads measurements introduced by SRM servers and storage systems, FTS server misconfigurations and identification of congested channels, historical transfer throughputs per stream for site-to-site data transfer comparisons, file-latency studies, among others... This information is retrieved directly from the FTS servers through the FTS Monitor webpages and conveniently archived for further analysis. The project provides a monitoring interface for all these values. Measurements, problems and improvements in CMS sites connected to LHCOPN are shown, where differences up to x100 are visible, constant performance measurements of data flowing from Tier-0 to Tier-1s, comparison to other existing monitoring tools (PerfSonar, LHCOPN dashboard), as well as the usage of the graphical interface to understand, among others, the effects for sites when connecting to LHCONE network. Given the multi-VO added value of this tool, this work is serving as a reference for building up the WLCG FTS monitoring tool, which will be based on the FTS messaging system.

Poster Session / **262**

Evolving ATLAS computing for today's networks

Author: Collaboration Atlas[1]

Co-authors: Cedric Serfon [2]; Fernando Harald Barreiro Megino [3]; I Ueda [4]; Simone Campana [5]; Stephane Jezequel [6]

[1] *Atlas*

[2] *Ludwig-Maximilians-Univ. Muenchen (DE)*

[3] *CERN IT ES*

[4] *University of Tokyo (JP)*

[5] *CERN*

[6] *Centre National de la Recherche Scientifique (FR)*

Corresponding Author: simone.campana@cern.ch

The ATLAS computing infrastructure was designed many years ago based on the assumption of rather limited network connectivity between computing centers. ATLAS sites have been organized in a hierarchical model, where only a static subset of all possible network links can be exploited and a static subset of well connected sites (CERN and the T1s) can cover important functional roles such as hosting master copies of the data.
The pragmatic adoption of such simplified approach, in respect of a more relaxed scenario interconnecting all sites, was very beneficial during the commissioning of the ATLAS distributed computing system and essential in reducing the operational cost during the first two years of LHC data taking.
In the mean time, networks evolved far beyond this initial scenario: while a few countries are still poorly connected with the rest of the WLCG infrastructure, most of the ATLAS computing centers are now efficiently interlinked. Our operational experience in running the computing infrastructure in the last years demonstrated many limitations of the current model: statically defined network paths are sometimes abused, while most of the network links are underutilized together with computing and storage resources at many sites, under the wrong assumption of limited connectivity with the rest of the infrastructure.
In this contribution we describe the various steps which ATLAS Distributed Computing went through in order to benefit from the network evolution and move from the current static model to a more relaxed scenario. This will include the development of monitoring and testing tools and the commissioning effort. We will finally describe the gains of the new model in terms of resource utilization at grid sites after many months of experience.

Poster Session / 263

New solutions for large scale functional tests in the WLCG infrastructure with SAM/Nagios: the experiments experience

Author: Andrea Sciaba[1]

Co-authors: Akshat Kakkar [2]; Alessandro Di Girolamo [1]; Amol Wakankar [2]; Biswajit Sarkar [3]; Guidone Negri [1]; Julia Andreeva [1]; Maarten Litmaath [1]; Maria Dolores Saiz Santos [4]; Nicolo Magini [1]; Pablo Saiz [1]; Partha Dhara [5]; Stefan Roiser [1]; Suja Ramachandran [6]

[1] *CERN*

[2] *Bhabha Atomic Research Centre (BARC)*

[3] *Department of Atomic Energy (DAE)*

[4] *Conseil Europeen Recherche Nucl. (CERN)*

[5] *Variable Energy Cyclotron Centre, Kolkata (India)*

[6] *Indira Gandhi Centre for Atomic Res*

Corresponding Authors: andrea.sciaba@cern.ch, alessandro.di.girolamo@cern.ch

Since several years the LHC experiments rely on the WLCG Service Availability Monitoring framework (SAM) to run functional tests on their distributed computing systems. The SAM tests have become an essential tool to measure the reliability of the Grid infrastructure and to ensure reliable computing operations, both for the sites and the experiments.
Recently the old SAM framework was replaced with a completely new system based on Nagios and ActiveMQ to better support the transition to EGI and to its more distributed infrastructure support model and to implement several scalability and functionality enhancements.
This required all LHC experiments and the WLCG support teams to migrate their tests, to acquire expertise on the new system, to validate the new availability and reliability computations and to adopt new visualisation tools.
In this contribution we describe in detail the current state of the art of functional testing in WLCG:

how the experiments use the new SAM/Nagios framework, the advanced functionality made available by the new framework and the future developments that are foreseen, with a strong focus on the improvements in terms of stability and flexibility brought by the new system.

Distributed Processing and Analysis on Grids and Clouds / 264

Exploiting Virtualization and Cloud Computing in ATLAS

Author: Collaboration Atlas[1]

Co-authors: Daniel Colin Van Der Ster [2]; Fernando Harald Barreiro Megino [3]; Kaushik De [4]; Rodney Walker [5]

[1] *Atlas*

[2] *CERN*

[3] *CERN IT ES*

[4] *University of Texas at Arlington (US)*

[5] *Ludwig-Maximilians-Univ. Muenchen (DE)*

Corresponding Author: fernando.harald.barreiro.megino@cern.ch

The ATLAS Computing Model was designed around the concepts of grid computing; since the start of data-taking, this model has proven very successful in the federated operation of more than one hundred Worldwide LHC Computing Grid (WLCG) sites for offline data distribution, storage, processing and analysis. However, new paradigms in computing, namely virtualization and cloud computing, present improved strategies for managing and provisioning IT resources that could allow ATLAS to more flexibly adapt and scale its storage and processing workloads on varied underlying resources. In particular, ATLAS is developing a "grid-of-clouds" infrastructure in order to utilize WLCG sites that make resources available via a cloud API.
This work will present the current status of the Virtualization and Cloud Computing R&D project in ATLAS Distributed Computing. First, strategies for deploying PanDA queues on cloud sites will be discussed, including the introduction of a "cloud factory" for managing cloud VM instances. Next, performance results when running on virtualized/cloud resources at CERN LxCloud, StratusLab, and elsewhere will be presented. Finally, we will present the ATLAS strategies for exploiting cloud-based storage, including remote XROOTD access to input data, management of EC2-based files, and the deployment of cloud-resident LCG storage elements.

Poster Session / 265

ATLAS R&D Towards Next-Generation Distributed Computing

Author: Collaboration Atlas[1]

[1] *Atlas*

The ATLAS Distributed Computing (ADC) project delivers production quality tools and services for ATLAS offline activities such as data placement and data processing on the Grid. The system has been capable of sustaining with large contingency the needed computing activities in the first years of LHC data taking, and has demonstrated flexibility in reacting promptly to new challenges. Development activities in this period have focused on consolidating existing services and increasing automation to be able to sustain existing loads. At the same time, an R&D program has evaluated new solutions and promising technologies capable of extending the operational scale, manageability and feature set of ATLAS distributed computing, several of which have selectively been brought to maturity as production-level tools and services. We will give an overview of R&D work in evaluating new tools and approaches and their integration into production services. A non exhaustive list

of items includes cloud computing and virtualization, non-relational databases, utilizing multicore processors, the CERNVM File System, end to end network monitoring, event and file level caching, and federated distributed storage systems. The R&D initiative, while focused on ATLAS needs, has aimed for a broad scope involving many other parties including other LHC experiments, ATLAS Grid sites, the CERN IT department, and WLCG and OSG programs.

Poster Session / **266**

Data analysis system for Super Charm-Tau Factory at BINP

Author: Ivan Logashenko[1]

Co-authors: Aleksandr Korol [2]; Alexandr Zaytsev [3]; Evgeny Baldin [1]

[1] *Budker Institute Of Nuclear Physics*

[2] *Budker Institute of Nuclear Physics (BINP)*

[3] *Budker Institute of Nuclear Physics (RU)*

Super Charm–Tau Factory (CTF) is a future electron-positron collider with center-of-mass energy range from 2 to 5 GeV and unprecedented for this energy range peak luminosity of about 10**35 cm−2s−1. The project of CTF is being developed in the Budker Institute of Nuclear Physics (Novosibirsk, Russia). The main
goal of experiments at Super Charm-Tau Factory is a study of the processes with charm quarks or tau leptons in the final state using data samples, which are by 3–4 orders of magnitude higher than collected by now in any other experiments.

The peak input data flow up to 10 GBytes/s and very large collected data volume, estimated to be 200 PBytes, require to design large scale data storage and data analysis system. We overview the requirements for the computer infrastructure of Super Charm-Tau Factory and discuss the main design solutions.

Poster Session / **267**

Distributed Data Analysis in the ATLAS Experiment: Challenges and Solutions

Author: Collaboration Atlas[1]

[1] *Atlas*

The ATLAS experiment at the LHC at CERN is recording and simulating several 10's of PetaBytes of data per year. To analyse these data the ATLAS experiment has developed and operates a mature and stable distributed analysis (DA) service on the Worldwide LHC Computing Grid. The service is actively used: more than 1400 users have submitted jobs in the year 2011 and a total of more 1 million jobs run every week. Users are provided with a suite of tools to submit Athena, ROOT or generic jobs to the grid, and the PanDA workload management system is responsible for their execution. The reliability of the DA service is high but steadily improving; grid sites are continually validated against a set of standard tests, and a dedicated team of expert shifters provides user support and communicates user problems to the sites. This talk will review the state of the DA tools and services, summarize the past year of distributed analysis activity, and present the directions for future improvements to the system.

Poster Session / 268

The evolving role of Tier2s in ATLAS with the new Computing and Data Model

Author: Collaboration Atlas[1]

Co-author: Santiago Gonzalez De La Hoz [2]

[1] *Atlas*

[2] *Universidad de Valencia (ES)*

Corresponding Author: santiago.gonzalezdelahoz@cern.ch

Originally the ATLAS computing model assumed that the Tier2s of each of the 10 clouds keep on disk collectively at least one copy of all "active" AOD and DPD datasets. Evolution of ATLAS computing and data models requires changes in ATLAS Tier2s policy for the data replication, dynamic data caching and remote data access.
Tier2 operations take place completely asynchronously with respect to data taking. Tier2s do simulation and user analysis. Large-scale reprocessing jobs on real data are at first taking place mostly at Tier1s but will progressively move to Tier2s as well. The availability of disk space at Tier2s is extremely important in the ATLAS computing model as it allows more data to be readily accessible for analysis jobs to all users, independently of their geographical location. The Tier2s disk space has been reserved for real, simulated, calibration and alignment, group, and user data. A buffer disk space is needed for input and output data for simulations jobs.
Tier2s are going to be used more efficiently. In this way Tier1s and Tier2s are becoming more equivalent for the network and the Hierarchy of Tier1, 2 is not longer so important. This talk will present the usage of Tier2s resources in different GRID activities, caching of data at Tier2s, and their role in the analysis in the new ATLAS computing model.

Poster Session / 269

ATLAS Distributed Computing Operations: Experience and improvements after 2 full years of data-taking

Author: Collaboration Atlas[1]

[1] *Atlas*

Corresponding Authors: stephane.jezequel@cern.ch, graeme.andrew.stewart@cern.ch

This paper will summarize operational experience and improvements in ATLAS computing infrastructure during 2010 and 2011.
ATLAS has had 2 periods of data taking, with many more events recorded in 2011 than in 2010. It ran 3 major reprocessing campaigns. The activity in 2011 was similar to that in 2010, but scalability issues had to be adressed due to the increase in luminosity and trigger rate. Based on improved monitoring of ATLAS Grid computing, the evolution of computing activities (data/group production, their distribution and grid analysis) over time will be presented.
The major bottlenecks and the implemented solutions will be described. The main changes in the implementation of the computing model that will be shown are: the optimisation of data distribution over the Grid, according to effective transfer rate and site readiness for analysis; the relaxation of the cloud model, for data distribution and data processing; software installation migration to cvmfs; changing database access to a Frontier/squid infrastructure.

Poster Session / 270

Enabling data analysis à la PROOF on the Italian ATLAS-Tier2's using PoD

Author: Collaboration Atlas[1]

Co-authors: Agnese Martini [2]; Alberto Annovi [3]; Alessandra Doria [4]; Alessandro De Salvo [5]; Anar Manafov [6]; Elisabetta Vilucchi [3]; Gerardo Ganis [7]; Giampaolo Carlino [8]; Manoj Kumar Jha [9]; Marianna Testa [3]; Mario Antonelli [3]; Roberto Di Nardo [3]

[1] *Atlas*

[2] *Istituto Nazionale Fisica Nucleare (INFN)*

[3] *Istituto Nazionale Fisica Nucleare (IT)*

[4] *Universita e INFN (IT)*

[5] *Universita e INFN, Roma I (IT)*

[6] *GSI - Helmholtzzentrum fur Schwerionenforschung GmbH (DE)*

[7] *CERN*

[8] *INFN, Sezione di Napoli-Universita & INFN, Napoli*

[9] *INFN Bologna*

Corresponding Authors: roberto.di.nardo@cern.ch, elisabetta.vilucchi@lnf.infn.it

In the ATLAS computing model, Tier2 resources are intended for MC productions and end-user analyses activities. These resources are usually exploited via the standard GRID resource management tools, which are de facto a high level interface to the underlying batch systems managing the contributing clusters. While this is working as expected, there are user-cases where a more dynamic usage of the resources may be more appropriate. For example, the design and optimization of an analysis on a large data sample available on the local storage of the Tier2, requires many iterations and fast turn around. In these cases a 'pull' model for work distribution, like the one implemented by PROOF, may be more effective.
This contribution describes our experience using PROOF for data analysis on the Italian ATLAS-Tier2: Frascati, Napoli and Roma1. To enable PROOF on the cluster we used PoD, PROOF on Demand. PoD is a set of tools designed to interact with any resource management system (RMS) to start the PROOF daemons. In this way any user can quickly setup its own PROOF cluster on the resources, with the RMS taking care of scheduling, priorities and accounting. Usage of PoD has steadily increased in the last years, and the product has now reached a production level quality.
PoD features an abstract interface to RMSs and provides several plugins for the most common RMSs. In our tests we used both the gLite and PBS plug-ins, the latter being the native RMS handling the resources under test. Data were accessed via xrootd, with file discovery provided by the standard ATLAS tools. The SRM is DPM (Disk Pool Manager) which has rfio as standard data access protocol; so we provided DPM of Xrootd protocol too.
We will describe the configuration and setup details and the results of some benchmark tests we run on the facility.

Poster Session / 271

CMS Tier-0: Preparing for the future

Author: Dirk Hufnagel[1]

[1] *Fermi National Accelerator Lab. (US)*

Corresponding Author: dirk.hufnagel@cern.ch

The Tier-0 processing system is the initial stage of the multi-tiered computing system of CMS. It is responsible for the first processing steps of data from the CMS Experiment at CERN. This talk covers the complete overhaul (rewrite) of the system for the 2012 run, to bring it into line with the

new CMS Workload Management system, improving scalability and maintainability for the next few years.

Summary:

In the last CHEP we presented the current CMS Tier0 system. It has worked very well for us, but due to the deployment of a new CMS workload management system this year, we were looking at changes to bring the Tier0 in sync with it. The changes were extensive enough to warrant a complete overhaul, a redesign based on lessons learned and rewrite from scratch.

Poster Session / **272**

The next generation ARC middleware and ATLAS computing model

Author: Collaboration Atlas[1]

Co-authors: Aleksandr Konstantinov [2]; Andrej Filipcic [3]; David Cameron [4]; Dmytro Karpenko [5]; Oxana Smirnova [6]

[1] *Atlas*

[2] *University of Helsinki (FI)*

[3] *Jozef Stefan Institute (SI)*

[4] *University of Oslo (NO)*

[5] *University of Oslo*

[6] *Lund University (SE)*

Corresponding Author: andrej.filipcic@ijs.si

The distributed NDGF Tier-1 and associated Nordugrid clusters are well integrated into the ATLAS computing model but follow a slightly different paradigm than other ATLAS resources. The current strategy does not divide the sites as in the commonly used hierarchical model, but rather treats them as a single storage endpoint and a pool of distributed computing nodes. The next generation ARC middleware with its several new technologies provides new possibilities in development of the ATLAS computing model, such as pilot jobs with pre-cached input files, automatic job migration between the sites, integration of remote sites without connected storage elements, and automatic brokering for jobs with non-standard resource requirements. ARC's data transfer model provides an automatic way for the computing sites to participate in ATLAS' global task management system without requiring centralised brokering or data transfer services. The powerful API combined with Python and Java bindings can easily be used to build new services for job control and data transfer. Integration of the ARC core into the EMI middleware provides a natural way to implement the new services using the ARC components.

Distributed Processing and Analysis on Grids and Clouds / **273**

Consolidation and development roadmap of the EMI middleware

Author: Balazs Konya[1]

Co-authors: John White White [2]; Jon Kerr Nilsen [3]; Laurence Field [4]; Marco Cecchi [5]; Patrick Fuhrmann [6]

[1] *Lund University (SE)*

[2] *Helsinki Institute of Physics (FI)*

[3] *University of Oslo (NO)*

[4] *CERN*

[5] *Istituto Nazionale Fisica Nucleare (IT)*

[6] *DESY*

Corresponding Author: balazs.konya@hep.lu.se

Scientific research communities have benefited recently from the increasing availability of computing and data infrastructures with unprecedented capabilities for large scale distributed initiatives. These infrastructures are largely defined and enabled by the middleware they deploy. One of the major issues in the current usage of research infrastructures is the need to use similar but often incompatible middleware solutions.

The European Middleware Initiative (EMI) is a collaboration of the major European middleware providers ARC, dCache, gLite and UNICORE. EMI aims to: deliver a consolidated set of middleware components for deployment in EGI, PRACE and other Distributed Computing Infrastructures; extend the interoperability between grids and other computing infrastructures; strengthen the reliability of the services; establish a sustainable model to maintain and evolve the middleware; fulfill the requirements of the user communities.

This paper presents the consolidation and development objectives of the EMI software stack covering the next two years. Details will be given concerning how the most important requirements of the key user groups, including the high energy physics community, were taken into account. The EMI development roadmap will be introduced along the four technical areas of compute, data, security and infrastructure.

The compute area plan focuses on consolidation of standards and agreements through an unified interface for job submission and management, a common format for accounting, the wide adoption of GLUE schema version 2.0 and the provision of a common framework for the execution of parallel jobs. The security area is working towards a unified security model and lowering the barriers to Grid usage by allowing users to gain access with their own credentials. The data area is focusing on implementing standards to ensure interoperability with other grids and industry components and to reuse already existing clients in operating systems and open source distributions. One of the highlights of the infrastructure area is the consolidation of the information system services via the creation of a common information backbone.

Wherever possible early results of the consolidation plan and the ongoing development will be covered by introducing EMI technical agreements and development prototypes.

Distributed Processing and Analysis on Grids and Clouds / 274

PD2P : PanDA Dynamic Data Placement for ATLAS

Author: Collaboration Atlas[1]

Co-authors: Kaushik De [2]; Sergey Panitkin [3]; Tadashi Maeno [3]

[1] *Atlas*

[2] *University of Texas at Arlington (US)*

[3] *Brookhaven National Laboratory (US)*

Corresponding Author: tmaeno@bnl.gov

The PanDA Production and Distributed Analysis System is the ATLAS workload management system for processing user analysis, group analysis and production jobs.
In 2011 more than 1400 users have submitted jobs through PanDA to the ATLAS grid infrastructure. The system processes more than 2 million analysis jobs per week. Analysis jobs are routed to sites based on the availability of relevant data and processing resources, taking account of the nonuniform distribution of CPU and storage resources in the ATLAS grid. The data distribution has to be optimized to fit the resource distribution, and also has to be dynamically changed to meet rapidly evolving requirements for analysis use cases.

The PanDA Dynamic Data Placement (PD2P) system has been developed to cope with difficulties of data placement for ATLAS. PD2P is an intelligent subsystem of PanDA to distribute data by taking the following factors into account: popularity, locality, the usage pattern of the data, the distribution of CPU and storage resources, network topology between sites, site operation downtime and reliability, and so on. We will describe the design of the new system, its performance during the past year of data taking, dramatic improvements it has brought about in the efficient use of storage and processing resources, associated reductions in average wait time for user analysis jobs, and plans for the future.

Poster Session / **275**

Evolution of ATLAS PanDA System

Authors: Alden Stradling[1]; Collaboration Atlas[2]

Co-authors: Kaushik De [1]; Paul Nilsson [1]; Rodney Walker [3]; Sergey Panitkin [4]; Tadashi Maeno [4]; Torre Wenaus [4]; Valeri Fayn [4]

[1] *University of Texas at Arlington (US)*

[2] *Atlas*

[3] *Ludwig-Maximilians-Univ. Muenchen (DE)*

[4] *Brookhaven National Laboratory (US)*

Corresponding Author: tmaeno@bnl.gov

The PanDA Production and Distributed Analysis System plays a key role in the ATLAS distributed computing infrastructure.
PanDA is the ATLAS workload management system for processing all Monte-Carlo simulation and data reprocessing jobs in addition to user and group analysis jobs. The system processes more than 5 million jobs in total per week, and more than 1400 users have submitted analysis jobs in 2011 through PanDA. PanDA has performed well with high reliability and robustness during the two years of LHC data-taking, while being actively evolved to meet the rapidly changing requirements for analysis use cases. We will present an overview of system evolution including PanDA's roles in data flow, automatic rebrokerage and reattempt for analysis jobs, adaptation for the CERNVM File System, support for the 'multi-cloud' model through which Tier 2s act as members of multiple clouds, pledged resource management, monitoring improvements, and so on. We will also describe results from the analysis of two years of PanDA usage statistics, current issues, and plans for the future.

Computer Facilities, Production Grids and Networking / **277**

Dimensioning storage and computing clusters for efficient High Throughput Computing

Author: Xavier Espinal Curull[1]

Co-authors: Arnau Bria [2]; Elena Planas [3]; Esther Accion Garcia [4]; Fernando Lopez Munoz [3]; Francisco Martinez Ramirez De Loaysa [4]; Gerard Bernabeu Altayó [5]; Manuel Delfino Reznicek [1]; Marc Caubet Serrabou [6]

[1] *Universitat Autònoma de Barcelona (ES)*

[2] *Port d'Informació Científica (PIC)*

[3] *PIC*

[4] *Unknown*

[5] *PIC (Tier-1)*

[6] *Universitat Autònoma de Barcelona*

Corresponding Author: espinal@pic.es

Scientific experiments are producing huge amounts of data, and they continue increasing the size of their datasets and the total volume of data. These data are then processed by researchers belonging to large scientific collaborations, with the Large Hadron Collider being a good example. The focal point of Scientific Data Centres has shifted from coping efficiently with PetaByte scale storage to deliver quality data processing throughput. The dimensioning of the internal components in High Throughput Computing (HTC) data centers is of crucial importance to cope with all the activities demanded by the experiments, both the online (data acceptance) and the offline (data processing, simulation and user analysis). This requires a precise setup involving disk and tape storage services, a computing cluster and the internal networking to prevent bottlenecks, overloads and undesired slowness that lead to losses cpu cycles and batch jobs failures. In this paper we point out relevant features for running a successful storage setup in an intensive HTC environment

Poster Session / **278**

Managing a site with Puppet

Author: Xavier Espinal Curull[1]

Co-authors: Arnau Bria [2]; Elena Planas [3]; Esther Accion Garcia [4]; Fernando Lopez Munoz [3]; Francisco Martinez Ramirez De Loaysa [4]; Gerard Bernabeu Altayó [5]; Manuel Delfino Reznicek [1]; Marc Caubet Serrabou [6]

[1] *Universitat Autònoma de Barcelona (ES)*

[2] *Port d'Informació Científica*

[3] *PIC*

[4] *Unknown*

[5] *PIC (Tier-1)*

[6] *Universitat Autònoma de Barcelona*

Corresponding Author: espinal@pic.es

Installation and post-installation mechanisms are critical points for the computing centres to streamline production services. Managing hundreds of nodes is a challenge for any computing centre and there are many tools able to cope with this problem. The desired features includes the ability to do incremental configuration (no need to bootstrap the service to make it manageable by the tool), simplicity in the description language for the configurations and in the system itself, ease of extension of the properties/capabilities of the system, a rich community for assistance and development, and open-source software. A possible choice to steer post-installations and dynamic post-configurations is Puppet. Puppet is a central point where profiles can be defined, those can easily be propagated around the cluster hence fulfilling the necessities of post-install configurations after the raw Operating System installation. Puppet also ensures the enforcement of the profile and the defined services once has been completely installed. We found in puppet a correct trade-off among simplicity and flexibility, and it was the most fitting to our requirements. Puppet approach to system management is simplistic, non-intrusive and incremental; puppet do not try to control every aspect of the configuration but only the ones you are interested in. Allows to manage a whole site from a central service, easing a lot potential reconfiguration or speeding up disaster recovery procedures.

Poster Session / **279**

Extra Dimensions: Creating 3D content in PDF

Author: Norman Anthony Graf[1]

[1] *SLAC National Accelerator Laboratory (US)*

Corresponding Author: norman.graf@slac.stanford.edu

Experimental science is replete with multi-dimensional information which is often poorly represented by the two dimensions of presentation slides and print media. Past efforts to disseminate such information to a wider audience have failed for a number of reasons, including a lack of standards which are easy to implement and have broad support. Adobe's Portable Document Format (PDF) has in recent years become the de facto standard for secure, dependable electronic information exchange. It has done so by creating an open format, providing support for multiple platforms and being reliable and extensible. By providing support for the ECMA standard Universal 3D (U3D) and the ISO PRC file format in its free Adobe Reader software, Adobe has made it easy to distribute and interact with 3D content. Until recently, Adobe's Acrobat software was also capable of incorporating 3D content into PDF files from a variety of 3D file formats, including proprietary CAD formats. However, this functionality is no longer available in Acrobat X, having been spun off to a separate company. Incorporating 3D content now requires the additional purchase of a separate plug-in. In this talk we present alternatives based on open source libraries which allow the programmatic creation of 3D content in PDF format. While not providing the same level of access to CAD files as the commercial software, it does provide physicists with an alternative path to incorporate 3D content into PDF files from such disparate applications as detector geometries from Geant4, 3D data sets, mathematical surfaces or tesselated volumes.

Poster Session / **280**

ATLAS Grid Data Processing: system evolution and scalability

Authors: Andrei Minaenko[1]; Collaboration Atlas[2]

Co-authors: Alexandre Vaniachine [3]; Alexei Klimentov [4]; Borut Kersevan [5]; Dmitri Golubkov [6]; Pavel Nevski [4]; Rodney Walker [7]

[1] *Institute for High Energy Physics (RU)*

[2] *Atlas*

[3] *ATLAS*

[4] *Brookhaven National Laboratory (US)*

[5] *Jozef Stefan Institute*

[6] *Institute for High Energy Physics (IHEP)-Unknown-Unknown*

[7] *Ludwig-Maximilians-Univ. Muenchen (DE)*

Corresponding Author: pavel.nevski@cern.ch

The production system for Grid Data Processing (GDP) handles petascale ATLAS data reprocessing and Monte Carlo activities. The production system empowered further data processing steps on the Grid performed by dozens of ATLAS physics groups with coordinated access to computing resources worldwide, including additional resources sponsored by regional facilities.
The system provides knowledge management of configuration parameters for massive data processing tasks, reproducibility of results, scalable database access, orchestrated workflow and performance monitoring, dynamic workload sharing, automated fault tolerance and petascale data integrity control. The system evolves to accommodate a growing number of users and new requirements from our contacts in ATLAS main areas: Trigger, Physics, Data Preparation and Software & Computing. To assure scalability, the next generation production system architecture development is in progress. We report on scaling up the GDP production system for a growing number of users providing data for physics analysis and other ATLAS main activities.

Poster Session / **281**

BOINC service for volunteer cloud computing

Author: Nils Hoimyr[1]

Co-authors: Alvaro Gonzalez Alvarez [1]; Anton Karneyeu [2]; Artem Harutyunyan [1]; Ben Segal [3]; Daniel Lombraña Gonzále [4]; Eric Mcintosh [1]; Francois Grey [5]; Igor Zacharov [6]; Jakob Blomer [1]; Massimo Giovannozzi [1]; Miguel Marquina [1]; Pete Jones [1]; Peter Skands [1]; Predrag Buncic [1]

[1] *CERN*

[2] *Russian Academy of Sciences (RU)*

[3] *Unknown*

[4] *Citizens Cyberscience Centre (CCC)*

[5] *University of London (GB)*

[6] *EPFL*

Corresponding Authors: alvaro.gonzalez.alvarez@cern.ch, nils.hoimyr@cern.ch, miguel.marquina@cern.ch, b.segal@cern.ch, predrag.buncic@cern.ch, jakob.blomer@cern.ch, artem.harutyunyan@cern.ch, anton.karneyeu@cern.ch, peter.skands@cern.ch, eric.mcintosh@cern.ch, francois.grey@cern.ch, massimo.giovannozzi@cern.ch

Since a couple of years, a team at CERN and partners from the Citizen Cyberscience Centre (CCC) have been working on a project that enables general physics simulation programs to run in a virtual machine on volunteer PCs around the world. The project uses the Berkeley Open Infrastructure for Network Computing (BOINC) framework. Based on CERNVM and the job management framework Co-Pilot, this project was made available for public beta-testing in August 2011 with Monte Carlo simulations of LHC physics under the name "LHC@home 2.0" and the BOINC project: "Test4Theory". At the same time, CERN's efforts on Volunteer Computing for LHC machine studies have been intensified; this project has previously been known as LHC@home, and has been running the "Sixtrack"beam dynamics application for the LHC accelerator, using a classic BOINC framework without virtual machines. CERN-IT has set up a BOINC server cluster, and has provided and supported the BOINC infrastructure for both projects. CERN intends to evolve the setup into a generic BOINC application service that will allow scientists and engineers at CERN to profit from volunteer computing. The authors describe the experience with the 2 different approaches to volunteer computing as well as the status and outlook of a general BOINC service.

Please see also the presentation of CernVM Co-Pilot by Artem Harutyunyan

https://indico.cern.ch/contributionDisplay.py?contribId=94&confId=149557

Online Computing / 282

Upgrade of the CMS Event Builder

Authors: Alexander Flossdorf[1]; Andre Georg Holzner[2]; Andrea Petrucci[3]; Andrei Cristian Spataru[3]; Attila Racz[3]; Aymeric Arnaud Dupont[3]; Christian Deldicque[3]; Christian Hartl[3]; Christoph Paus[4]; Christoph Schwick[3]; Dennis Shpakov[5]; Dominique Gigi[3]; Emilio Meschi[3]; Frank Glege[3]; Frans Meijers[3]; Gerry Bauer[4]; Giovanni Polese[3]; Hannes Sakulin[3]; James Branson[6]; Jeroen Hegeman[3]; Jose Antonio Coarasa Perez[3]; Konstanty Sumorok[4]; Lorenzo Masetti[3]; Luciano Orsini[3]; Marc Dobson[3]; Marco Pieri[2]; Matteo Sani[2]; Matthew Bowen[7]; Michal Simon[None]; Olivier Raginel[4]; Remi Mommsen[5]; Robert Gomez-Reino Garrido[3]; Samim Erhan[8]; Sebastian Bukowiec[3]; Sergio Cittolin[2]; Ulf Behrens[9]; Vivian O'Dell[10]; Yi Ling Hwong[3]

[1] *DESY*

[2] *Univ. of California San Diego (US)*

[3] *CERN*

[4] *Massachusetts Inst. of Technology (US)*

[5] *Fermi National Accelerator Lab. (US)*

[6] *UC San Diego*

[7] *University of the West of England*

[8] *Univ. of California Los Angeles (US)*

[9] *Deutsches Elektronen-Synchrotron (DE)*

[10] *Fermi National Accelerator Laboratory (FNAL)*

Corresponding Author: andrea.petrucci@cern.ch

The Data Acquisition (DAQ) system of the Compact Muon Solenoid (CMS) experiment at CERN assembles events at a rate of 100 kHz, transporting event data at an aggregate throughput of 100 GB/s. By the time the LHC restarts after the 2013/14 shut-down, the current compute nodes and networking infrastructure will have reached the end of their lifetime. We are presenting design studies for an upgrade of the CMS event builder based on advanced networking technologies such as 10 Gb/s Ethernet. We report on tests and performance measurements with small-scale test setups.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Distributed Processing and Analysis on Grids and Clouds / 283

Experience in Grid Site Testing for ATLAS, CMS and LHCb with HammerCloud

Author: Daniel Colin Van Der Ster[1]

Co-authors: Andrea Sciaba [1]; Federica Legger [2]; Johannes Elmsheuser [2]; Mario Ubeda Garcia [1]; Ramon Medrano Llamas [3]

[1] *CERN*

[2] *Ludwig-Maximilians-Univ. Muenchen (DE)*

[3] *Universidad de Oviedo (ES)*

Corresponding Author: daniel.vandersterr@cern.ch

Frequent validation and stress testing of the network, storage and CPU resources of a grid site is essential to achieve high performance and reliability. HammerCloud was previously introduced with the goals of enabling VO- and site-administrators to run such tests in an automated or on-demand manner. The ATLAS, CMS and LHCb experiments have all developed VO plugins for the service and have successfully integrated it into their grid operations infrastructures.

This work will present the experience in running HammerCloud at full scale for more than 3 years and present solutions to the scalability issues faced by the service. First, we will show the particular challenges faced when integrating with CMS and LHCb offline computing, including customized dashboards to show site validation reports for the VOs and a new API to tightly integrate with the LHCbDIRAC Resource Status System. Next, a study of the automatic site exclusion component used by ATLAS will be presented along with results for tuning the exclusion policies. A study of the historical test results for ATLAS, CMS and LHCb will be presented, including comparisons between the experiments' grid availabilities and a search for site-based or temporal failure correlations. Finally, we will look to future plans that will allow users to gain new insights into the test results; these include developments to allow increased testing concurrency, increased scale in the number of metrics recorded per test job (up to hundreds), and increased scale in the historical job information (up to many millions of jobs per VO).

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Poster Session / 284

Evolution of Version Control Services at CERN: Life-cycle of Services

Authors: Alvaro Gonzalez Alvarez[1]; David Asbury[1]; Georgios Koloventzos[2]

[1] *CERN*

[2] *University of Athens (GR)*

Corresponding Author: alvaro.gonzalez.alvarez@cern.ch

In 2002, the first central CERN service for version control based on CVS was set up. Since then, three different services based on CVS and SVN have been launched and run in parallel; there are user requests for another service based on git. In order to ensure that the most demanded services are of high quality in terms of performance and reliability, services in less demand had to be shut down. The support team has recently closed one flavour of the CVS services, and is working, together with user groups concerned, on closing the remaining CVS service; both closures have shown to be both technical and social challenges. Meanwhile work is going on in order to consolidate and improve the SVN service, ensuring proper scalability with user requirements. In parallel, the team is studying how to potentially set up a git service. The presentation will report on our experience with service management from a life-cycle point of view: creation, maintenance, evolution and closure, taking into account the user, technical and managerial perspectives using the version control services as a real-life example.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Event Processing / 285

Rethinking particle transport in the many-core era

Authors: Andrei Gheata[1]; Federico Carminati[1]; Rene Brun[1]

[1] *CERN*

Corresponding Author: federico.carminati@cern.ch

Detector simulation is one of the most CPU intensive tasks in modern High Energy Physics. While its importance for the design of the detector and the estimation of the efficiency is ever increasing, the amount of events that can be simulated is often constrained by the available computing resources. Various kind of "fast simulations" have been developed to alleviate this problem, however, while successful, these are mostly "ad hoc" solutions which do not replace completely the need for detailed simulations. One of the common features of both detailed and fast simulation is the inability of the codes to exploit fully the parallelism which is increasingly offered by the new generations of CPUs. In the next years it is reasonable to expect an increase on one side of the needs for detector simulation, and on the other in the parallelism of the hardware, widening the gap between the needs and the available means. In the past years, and indeed since the beginning of simulation programs, several unsuccessful efforts have been made to exploit the "embarrassing parallelism" of simulation programmes. After a careful study of the problem, and based on a long experience in simulation codes, the authors have concluded that an entirely new approach has to be adopted to exploit parallelism. The talk will review current prototyping work, encompassing both detailed and fast simulation use cases. Performance studies will be presented, together with a roadmap to develop a new full-fledged transport program efficiently exploiting parallelism for the physics and geometry computations, while adapting the steering mechanisms to accommodate detailed and fast simulation in a single framework.

Poster Session / 286

Virtualization of Grid Services

Author: Andreas Gellrich[1]

[1] *DESY*

Corresponding Author: andreas.gellrich@desy.de

Virtualization techniques have become a key topic in computing in the last years. In the Grid, discussions on the virtualization of worker nodes is most prominent. Currently, concepts for the provenience and sharing if images are under debate. The virtualization of Grid servers though is already a common and successful practice.

At DESY, one of the largest WLCG Tier-2 centres world-wide and home of a number of global VOs, already half of the Grid services run on virtual machines. This approach helped to improve the reliability, redundancy, and efficiency of the Grid infrastructure.

In the contribution to CHEP 2012 we will describe our set-up with regard to the usage of virtualization techniques for Grid servers. We will discuss the choice of products, free and commercial ones, and their implications, and present findings and experiences of day-to-day operations.

Event Processing / 287

Exploiting new CPU architectures in the SuperB software framework

Author: Marco Corvo[1]

Co-authors: Alberto Gianoli [2]; Alejandro Perez [3]; Andrea Di Simone [4]; Armando Fella [1]; Bruno Santeramo [5]; Domenico DelPrete [6]; Eleonora Luppi [7]; Fabrizio Bianchi [8]; Francesco Giacomini [9]; Giacinto Donvito [10]; Guido Russo [11]; Luca Tomassetti [7]; Matteo Manzali [2]; Matteo Rama [12]; Roberto Stroili [13]; Silvio Pardi [14]; Stefano Longo [13]; Steffen Luitz [15]; Vincenzo Ciaschini [9]

[1] *CNRS*

[2] *INFN Ferrara*

[3] *INFN Pisa*

[4] *Universita degli Studi di Roma Tor Vergata (IT)*

[5] *INFN Bari*

[6] *INFN Napoli*

[7] *Universita' di Ferrara and INFN Ferrara*

[8] *Universita' di Torino and INFN Torino*

[9] *INFN CNAF*

[10] *INFN-Bari*

[11] *Universita' di Napoli and INFN (IT)*

[12] *INFN LNF*

[13] *INFN Padova*

[14] *INFN*

[15] *SLAC*

Corresponding Authors: corvo@pd.infn.it, fabrizio.bianchi@to.infn.it

The SuperB asymmetric energy e+e- collider and detector to be built at the newly founded Nicola Cabibbo Lab will provide a uniquely sensitive probe of New Physics in the flavor sector of the Standard Model. Studying minute effects in the heavy quark and heavy lepton sectors requires a data sample of 75 ab-1 and a luminosity target of 10ˆ36 cm-2 s-1.

These parameters require a substantial growth in computing requirements and performances. The SuperB collaboration is thus investigating the advantages of new CPU architectures (multi and many cores) and how to exploit their capability of task parallelization in the framework for simulation and analysis software.
In this work we present the underlying architecture which we intend to use and some preliminary performance results of the first framework prototype.

Poster Session / **288**

Model of shared ATLAS Tier2 and Tier3 facilities in EGI/gLite Grid flavour

Author: Santiago Gonzalez De La Hoz[1]

Co-authors: Andres Pacheco Pages [2]; Collaboration Atlas [3]; Daniel Colin Van Der Ster [4]; Doug Benjamin [5]; Helmut Wolters [6]; Horst Severini [7]; Javier Sanchez [8]; Juan Jose Pardo Navarro [9]; Lorne Levinson [10]; Mario Solano Gonzales [11]; Martin Gasthuber [12]; Miguel Villaplana Perez [13]; Simone Campana [4]; Wahid Bhimji [14]; Xavier Espinal Curull [15]; Yves Kemp [12]

[1] *IFIC-Valencia*

[2] *IFAE*

[3] *Atlas*

[4] *CERN*

[5] *Duke University (US)*

[6] *Universidade de Coimbra (PT)*

[7] *University of Oklahoma (US)*

[8] *Consejo Superior de Investigaciones Cientificas (CSIC)-Universi*

[9] *Universidad Autonoma de Madrid (ES)*

[10] *Weizmann Institute of Science (IL)*

[11] *George Mason University (US)*

[12] *Deutsches Elektronen-Synchrotron (DE)*

[13] *Universidad de Valencia (ES)*

[14] *University of Edinburgh (GB)*

[15] *Universitat Autònoma de Barcelona (ES)*

Corresponding Author: santiago.gonzalez@ific.uv.es

The ATLAS computing and data models have moved/are moving away from the strict MONARC model (hierarchy) to a mesh model. Evolution of computing models also requires evolution of network infrastructure to enable any Tier2 and Tier3 to easily connect to any Tier1 or Tier2. In this way some changing of the data model are required:
a) Any site can replicate data from any other site.
b) Dynamic data caching. Analysis sites receive datasets from any other site “on demand”based on usage pattern, and possibly using a dynamic placement of datasets by centrally managed replication of whole datasets. Unused data is removed.
c) Remote data access. Local jobs could access data stored at remote sites using local caching on a file or sub-file level.
In this contribution, the model of shared ATLAS Tier2 and Tier3 facilities in the EGI/gLite flavour is explained. The Tier3s in the US and the Tier3s in Europe are rather different because in Europe we have facilities which are Tier2s with a Tier3 component (Tier3 with a co-located Tier2).
Data taking in ATLAS has been going on for more than one year. The Tier2 and Tier3 facility setup, how do we get the data, how do we enable at the same time grid and local data access, how Tier2

and Tier3 activities affect the cluster differently and process of hundreds of million of events, will be presented.
Finally, an example of how a real physics analysis is working at these sites will be shown, and this is a good occasion to see if we have developed all the Grid tools necessary for the ATLAS Distributed Computing community, and in case we do not, to try to fix it, in order to be ready for the foreseen increase in ATLAS activity in the next years.

Poster Session / **289**

Providing WLCG Global Transfer monitoring

Authors: David Kingsley Tuckett[None]; Julia Andreeva[1]

Co-authors: Alexander Uzhinskiy [2]; Daniel Dieguez Arias [3]; Gunnar Ro [1]; José Flix ; Michail Salichos [1]; Nicolo Magini [1]; Oliver Keeble [1]; Pablo Saiz [1]; Simone Campana [1]; Tony Wildish [4]; Zsolt Molnar [1]

[1] *CERN*

[2] *Joint Inst. for Nuclear Research (JINR)*

[3] *University of Vigo (ES)*

[4] *Princeton University (US)*

Corresponding Author: julia.andreeva@cern.ch

The WLCG Transfer Dashboard is a monitoring system which aims to provide a global view of the WLCG data transfers and to reduce redundancy of monitoring tasks performed by the LHC experiments. The system is designed to work transparently across LHC experiments and across various technologies used for data transfer. Currently every LHC experiment monitors data transfers via experiment-specific systems but the overall cross-experiment picture is missing. Even for data transfers handled by FTS, which is used by 3 LHC experiments, monitoring tasks such as aggregation of FTS transfer statistics or estimation of transfer latencies are performed by every experiment separately. These tasks could be performed once, centrally, and then served to all experiments via a well-defined set of APIs. In the design and development of the new system, experience accumulated by the LHC experiments in the data management monitoring area is taken into account and a considerable part of the code of the ATLAS DDM Dashboard is being re-used. The presentation will describe the architecture of the Global Transfer monitoring system, the implementation of its components and the first prototype.

Poster Session / **290**

Optimizing Resource Utilization in Grid Batch Systems

Author: Andreas Gellrich[1]

[1] *DESY*

Corresponding Author: andreas.gellrich@desy.de

DESY is one of the largest WLCG Tier-2 centres for ATLAS, CMS and LHCb world-wide and the home of a number of global VOs. At the DESY-HH Grid site more than 20 VOs are supported by one common Grid infrastructure to allow for the opportunistic usage of federated resources. The VOs share roughly 4800 job slots in 800 physical CPUs of 400 hosts operated by a TORQUE/MAUI batch system.

On Tier-2 sites, the utilization of computing, storage, and network requirements of the Grid jobs differ widely. For instance Monte Carlo production jobs are almost purely CPU bound, whereas physics analysis jobs demand high data rates.

In order to optimize the utilization of resources, jobs must be distributed intelligently over the slots, CPUs, and hosts. Although the jobs resource requirements cannot be deduced directly, jobs are mapped to POSIX user/group ID based on their VOMS-proxy. The user/group ID allows to distinguish jobs, assuming VOs make use of the VOMS group and role mechanism. This was implemented in the job scheduler (MAUI) configuration.

In the contribution to CHEP 2012 we will sketch our set-up, describe our configuration, and present experiences based on monitoring information.

Poster Session / 291

A new era for central processing and production in CMS

Authors: Edgar Mauricio Fajardo Hernandez[1]; Guillelmo Gomez Ceballos Retuerto[2]; Jacob Linacre[3]; Oliver Gutsche[4]

Co-authors: Ajit Kumar Mohapatra [5]; Dave Evans [6]; Markus Klute [7]; Matthew Norman [8]; Rapolas Kaselis [9]; Simon Metson [10]; Stephen Foulkes [11]; Valentina Dutta [2]; Vincenzo Spinoso [12]; Zdenek Maxa [13]

[1] *Universidad de los Andes (CO)*

[2] *Massachusetts Inst. of Technology (US)*

[3] *Oxford*

[4] *FERMILAB*

[5] *University of Wisconsin (US)*

[6] *Fermi National Accelerator Lab. (US)*

[7] *Massachusettes Institute of Technology*

[8] *University of California at San Diego*

[9] *Vilnius University (LT)*

[10] *University of Bristol (GB)*

[11] *Fermi National Accelerator Lab. (Fermilab)*

[12] *Universita e INFN (IT)*

[13] *California Institute of Technology (US)*

Corresponding Authors: rapolas.kaselis@cern.ch, edgar.mauricio.fajardo.hernandez@cern.ch, oliver.gutsche@cern.ch, linacre@fnal.gov, guillelmo.gomez.ceballos@cern.ch, markus.klute@cern.ch, ajit.kumar.mohapatra@cern.ch, valentina.dutta@cern.ch, evansde@fnal.gov, simon.metson@cern.ch, stephen.foulkes@cern.ch, mnorman@fnal.gov, zdenek.maxa@hep.caltech.edu

The goal for CMS computing is to maximise the throughput of simulated event generation while also processing the real data events as quickly and reliably as possible. To maintain this achievement as the quantity of events increases, since the beginning of 2011 CMS computing has migrated at the Tier 1 level from its old production framework, ProdAgent, to a new one, WMAgent. The WMAgent framework offers improved processing efficiency and increased resource usage as well as a reduction in manpower.

In addition to the challenges encountered during the design of the WMAgent framework, several operational issues have arisen during its commissioning. The largest operational challenges were in the usage and monitoring of resources, mainly a result of a change in the way work is allocated. Instead of work being assigned to operators, all work is centrally injected and managed in the Request Manager system and the task of the operators has changed from running individual workflows to monitoring the global workload.

In this report we present how we tackled some of the operational challenges, and how we benefitted from the lessons learned in the commissioning of the WMAgent framework at the Tier 2 level in late 2011. As case studies, we will show how the WMAgent system performed during some of the large data reprocessing and Monte Carlo simulation campaigns.

Poster Session / 292

DIRAC evaluation for the SuperB experiment

Author: Armando Fella[1]

Co-authors: Alberto Gianoli [2]; Alejandro Perez [3]; Andrea Di Simone [4]; Bruno Santeramo [5]; Domenico DelPrete [6]; Eleonora Luppi [7]; Fabrizio Bianchi [8]; Francesco Giacomini [9]; Giacinto Donvito [10]; Guido Russo [11]; Luca Tomassetti [7]; Marco Corvo [1]; Matteo Manzali [2]; Matteo Rama [12]; Roberto Stroili [13]; Silvio Pardi [14]; Stefano Longo [13]; Steffen Luitz [15]; Vincenzo Ciaschini [9]

[1] *CNRS*

[2] *INFN Ferrara*

[3] *INFN Pisa*

[4] *Universita degli Studi di Roma Tor Vergata (IT)*

[5] *INFN Bari*

[6] *INFN Napoli*

[7] *Universita' di Ferrara and INFN Ferrara*

[8] *Universita' di Torino and INFN Torino*

[9] *INFN CNAF*

[10] *INFN-Bari*

[11] *Universita di Napoli and INFN (IT)*

[12] *INFN LNF*

[13] *INFN Padova*

[14] *INFN*

[15] *SLAC*

Corresponding Authors: giacinto.donvito@ba.infn.it, armando.fella@pi.infn.it, fabrizio.bianchi@to.infn.it

The SuperB asymmetric energy e+e- collider and detector to be built at the newly founded Nicola Cabibbo Lab will provide a uniquely sensitive probe of New Physics in the flavor sector of the Standard Model. Studying minute effects in the heavy quark and heavy lepton sectors requires a data sample of 75 ab-1 and a luminosity target of 10^36 cm-2 s-1.

In this work we will present our evaluation of the DIRAC Distributed Infrastructure for use in the SuperB experiment based on the two use cases:
End User Analysis and Monte Carlo Production. We will present:
1) The test bed layout with DIRAC site and service configurations and the efforts to enable and manage OSG-EGI interoperability.
2) Our specific use cases ported to the DIRAC test bed with the computational and data management requirements and the DIRAC subsystem configuration.
3) The test results obtained from running both SuperB Monte Carlo and end user analysis with details about the performance achieved, the efficiency and the failures that occurred during the tests.
4) An evaluation and comparison of the two catalogue systems provided by the DIRAC framework, LFC (LHC File Catalogue) and DIRAC File Catalog in terms of features, performance and reliability.
5) Evaluation of capabilities and performance tests of the DIRAC Cloud capabilities as potentially applicable to SuperB computing.
6) A comparison of DIRAC with other submission systems available in the HEP community with pros and cons of each system.

Distributed Processing and Analysis on Grids and Clouds / 294

SuperB R&D computing program: HTTP direct access to distributed resources

Author: Giacinto Donvito[1]

Co-authors: Alberto Gianoli [2]; Alejandro Perez [3]; Andrea Di Simone [4]; Armando Fella [3]; Bruno Santeramo [5]; Domenico DelPrete [6]; Eleonora Luppi [7]; Fabrizio Bianchi [8]; Francesco Giacomini [9]; Guido Russo [10]; Luca Tomassetti [7]; Marco Corvo [11]; Matteo Manzali [2]; Matteo Rama [12]; Paolo Franchini [13]; Roberto Stroili [14]; Silvio Pardi [15]; Stefano Longo [14]; Steffen Luitz [16]; Vincenzo Ciaschini [9]

[1] *INFN-Bari*

[2] *INFN Ferrara*

[3] *INFN Pisa*

[4] *Universita degli Studi di Roma Tor Vergata (IT)*

[5] *INFN Bari*

[6] *INFN Napoli*

[7] *Universita' di Ferrara and INFN Ferrara*

[8] *Universita' di Torino and INFN Torino*

[9] *INFN CNAF*

[10] *Universita' di Napoli and INFN (IT)*

[11] *CNRS*

[12] *INFN LNF*

[13] *INFN CNAF, Bologna, Italy*

[14] *INFN Padova*

[15] *INFN*

[16] *SLAC*

Corresponding Authors: armando.fella@pi.infn.it, giacinto.donvito@ba.infn.it, fabrizio.bianchi@to.infn.it

The SuperB asymmetric energy e+e- collider and detector to be built at the newly founded Nicola Cabibbo Lab will provide a uniquely sensitive probe of New Physics in the flavor sector of the Standard Model. Studying minute effects in the heavy quark and heavy lepton sectors requires a data sample of 75 ab-1 and a luminosity target of 10ˆ36 cm-2 s-1.

The increasing network performance also in the Wide Area Network environment and the capability to read data remotely with good efficiency are providing new possibilities and opening new scenarios in the data access field.
Subjects like data access and data availability in a distributed environment are key points in the definition of the computing model for an HEP experiment like SuperB. R&D efforts in such a field have been brought on during the last year in order to release the Computing Technical Design Report within 2012.
Among the possible data access models resulting of interest for a mid-term future scenario we identify the WAN direct access via robust and reliable protocols such as HTTP/WebDAV and xrootd as a viable option.
In this work we present the R&D results obtained in the study of new data access technologies for typical HEP use cases, focusing on specific protocols such as HTTP and WebDAV in Wide Area Network scenarios. Reports on efficiency, performance and reliability tests have been included, using both Monte Carlo production and Analysis use cases. We also compare the results obtained with HTTP and xrootd protocols, in terms of performance, efficiency, security and features available.

Poster Session / 295

Configuration management and monitoring of the middleware at GridKa

Authors: Dimitri Nilsen[1]; Pavel Weber[1]

[1] *Karlsruhe Institute of Technology (KIT)*

GridKa is a computing centre located in Karlsruhe. It serves as Tier-1 centre for the four LHC experiments and also provides its computing and storage resources for other non-LHC HEP and astroparticle physics experiments as well as for several communities of the German Grid Initiative D-Grid.

The middleware layer at GridKa comprises three main flavours: Globus, gLite and UNICORE. This layer provides the access to the several clusters, according to the requirements of the corresponding communities. The heterogeneous structure of middleware resources and services requires their effective administration for stable and sustainable operation of the whole computing centre. In the presentation the overview of the middleware system at GridKa is given with focus on the configuration management and monitoring. These are the crucial components of the administration task for the system with high-availability setup. The various configuration tools used at GridKa, their benefits and limitations as well as developed automation procedures of the configuration management will be discussed. The overview of the monitoring system which evaluates the information delivered by central and local grid information services and provides status and detailed diagnostics for the middleware services is presented.

Poster Session / **296**

Grid Computing at GSI (ALICE and FAIR) - present and future

Author: Kilian Schwarz[1]

Co-authors: Dan Protopopescu [2]; Florian Uhlig [3]

[1] *GSI - Helmholtzzentrum fur Schwerionenforschung GmbH (DE)*

[2] *University of Glasgow*

[3] *GSI*

Corresponding Author: k.schwarz@gsi.de

The future FAIR experiments CBM and PANDA have computing requirements that fall in a category that could currently not be satisfied by one single computing centre. One needs a larger, distributed computing infrastructure to cope with the amount of data to be simulated and analysed.

Since 2002, GSI operates a Tier2 center for ALICE@CERN. The central component of the GSI computing facility and hence the core of the ALICE Tier2 centre is a LSF/SGE batch farm of 4200+ CPU cores shared by the participating experiments, and accessible both locally and via Grid. In terms of data storage, a 2.5 PB Lustre file system, directly accessible from all worker nodes is maintained, as well as a 400 TB xrootd-based Grid storage element.

Based on this existing expertise, and utilising ALICE's middleware 'AliEn', the Grid infrastructure for PANDA and CBM is being built. Besides a Tier0 centre at GSI, the computing Grids of the two FAIR collaborations encompass now more than 17 sites in 11 countries and are constantly expanding.

The operation of the distributed FAIR computing infrastructure benefits significantly from the experience gained with the ALICE Tier2 centre. A close collaboration between ALICE Offline and FAIR provides mutual advantages. The employment of a common Grid middleware as well as compatible simulation and analysis software frameworks ensure significant synergy effects.

However, there are certain distinctions in usage and deployment between ALICE, CBM and PANDA. Starting from the common attributes, this talk goes on to explore the particularities of the three Grids and the dynamics of knowledge transfer between them.

Summary:

GSI operates an ALICE T2 centre since 2002. Based on the corresponding experiences and in close collaboration with ALICE Offline the distributed computing infrastructure for the FAIR experiments is being set up.

Poster Session / **297**

ROOT: High Quality, Systematically

Author: Axel Naumann[1]

[1] *CERN*

Corresponding Author: axel.naumann@cern.ch

We will present new approaches to implementing quality control procedures in the development of the ROOT data processing framework. A multi-platform, cloud-based infrastructure is used for supporting the incremental build and test procedures employed in the ROOT software development process. Tests run continuously and a custom generic tool has been adopted for CPU and heap regression monitoring. We also use static analysis to check for coding errors. We will show how the adoption of these new procedures has influenced the way ROOT is developed.

Poster Session / **298**

Preparing for the new C++11 standard

Author: Axel Naumann[1]

[1] *CERN*

Corresponding Author: axel.naumann@cern.ch

C++11 is a new standard for the C++ language that includes several additions to the core language and that extends the C++ standard library. New features, such as move semantics, are expected to bring performance benefits and as soon as these benefits have been demonstrated, it will undoubtedly become widely adopted in the development of HEP code. However it will be shown that this may well be achieved only at the expense of an even more complex syntax, which may well impact on the readability of code (examples will be provided). One approach to addressing this issue can be to restrict the set of features C++ provides that are allowed to be used, e.g. in headers, and the best way of implementing restrictions of this sort is by an automated means. We argue that a compiler library, such as clang http://clang.llvm.org, can facilitate the implementation of such a code syntax checker, in particular by exploiting clang's already existing static code analysis functionality.

Poster Session / **299**

Testing and evaluating storage technology to build a distributed Tier1 for SuperB in Italy

Author: Silvio Pardi[1]

Co-authors: Alberto Gianoli [2]; Alejandro Perez [3]; Andrea Di Simone [4]; Armando Fella [5]; Bruno Santeramo [6]; Domenico DelPrete [7]; Eleonora Luppi [8]; Fabrizio Bianchi [9]; Francesco Giacomini [10]; Giacinto Donvito [11]; Guido Russo [12]; Luca Tomassetti [13]; Marco Corvo [5]; Matteo Manzali [2]; Matteo Rama [14]; Roberto Stroili [15]; Stefano Longo [15]; Steffen Luitz [16]; Vincenzo Ciaschini [10]

[1] *INFN*

[2] *INFN Ferrara*

[3] *INFN Pisa*

[4] *Universita degli Studi di Roma Tor Vergata (IT)*

[5] *CNRS*

[6] *INFN Bari*

[7] *INFN Napoli*

[8] *Universita' di Ferrara and INFN Ferrara*

[9] *Univerita' di Torino and INFN Torino*

[10] *INFN CNAF*

[11] *INFN-Bari*

[12] *Universita' di Napoli and INFN (IT)*

[13] *Univerita' di Ferrara and INFN Ferrara*

[14] *INFN LNF*

[15] *INFN Padova*

[16] *SLAC*

Corresponding Authors: spardi@na.infn.it, fabrizio.bianchi@to.infn.it

The SuperB asymmetric energy e+e- collider and detector to be built at the newly founded Nicola Cabibbo Lab will provide a uniquely sensitive probe of New Physics in the flavor sector of the Standard Model. Studying minute effects in the heavy quark and heavy lepton sectors requires a data sample of 75 ab-1 and a luminosity target of 10^36 cm-2 s-1.

This luminosity translate in the requirement of storing more than 50 PByte of additional data each year, making SuperB an interesting challenge to the data management infrastructure, both at site level as at Wide Area Network level.
A new Tier1, distributed among 3 or 4 sites in the south of Italy, is planned as part of the SuperB computing infrastructure.
In this paper we evaluate the advantages and draw backs in terms of service availability and reliability for a set of possible design for this distributed Tier1 site.
Different data and CPU resources access strategies will be described for a typical HEP experiment use cases.
We will report the activity of testing and evaluating several available technology that could be used in order to build a distributed sites for a
typical HEP experiment. In particular we will report about the test on the software like: Hadoop, EOS, Lustre, GPFS, and other similar product. We will also describe in details the algorithm used in order to guarantee the needed data resiliency.
All those software were tested both in a Local Area Network farm infrastructure as on a Wide Area Network test bed in which each site could exploit high performance network connection (ranging from 1 up to 10 Gbps).
A particular attention will be paid to the level of security provided by each solution and on the implication on the job management and scheduling coming from each design and software used.
At the end of the work we will show also few real cases test in order to show how each of the analyzed schema can or cannot fulfill the experiment requirements.

Software Engineering, Data Stores and Databases / 300

Designing and developing portable large-scale JavaScript web applications within the Experiment Dashboard framework

Author: David Tuckett[1]

Co-authors: Edward Karavakis [1]; Ivan Antoniev Dzhunov [2]; Julia Andreeva [1]; Lukasz Kokoszkiewicz [1]; Michal Maciej Nowotka [3]; Pablo Saiz [1]

[1] *CERN*

[2] *University of Sofia*

[3] *Warsaw University of Technology (PL)*

Corresponding Author: david.tuckett@cern.ch

Improvements in web browser performance and web standards compliance, as well as the availability of comprehensive JavaScript libraries, provides an opportunity to develop functionally rich yet intuitive web applications that allow users to access, render and analyse data in novel ways. However, the development of such large-scale JavaScript web applications presents new challenges, in particular with regard to code sustainability and team-based work.

We present an approach that meets the challenges of large-scale JavaScript web application design and development, including client-side model-view-controller architecture, design patterns, and JavaScript libraries. Furthermore, we show how the approach leads naturally to the encapsulation of the data source as a web API, allowing applications to be easily ported to new data sources.

The Experiment Dashboard framework is used for the development of applications for monitoring the distributed computing activities of virtual organisations on the Worldwide LHC Computing Grid. We demonstrate the benefits of the approach for large-scale JavaScript web applications in this context by examining the design of several Experiment Dashboard applications for data processing, data transfer and site status monitoring, and by showing how they have been ported for different virtual organisations and technologies.

Summary:

We present an approach to designing and developing large-scale JavaScript web applications that achieves the goals of rich functionality, code sustainability, and data source portability. We provide examples of the benefits of the approach by examining several Experiment Dashboard applications for monitoring distributed computing activities on the Worldwide LHC Computing Grid.

Poster Session / 301

SuperB Simulation Production System

Author: Luca Tomassetti[1]

Co-authors: Alberto Gianoli [2]; Alejandro Perez [3]; Alessandro Paolini [4]; Andrea Di Simone [5]; Armando Fella [6]; Bruno Santeramo [7]; Domenico DelPrete [8]; Eleonora Luppi [9]; Fabrizio Bianchi [10]; Francesco Giacomini [11]; Giacinto Donvito [12]; Guido Russo [13]; Marco Corvo [6]; Matteo Manzali [2]; Matteo Rama [14]; Roberto Stroili [15]; Silvio Pardi [16]; Stefano Longo [15]; Steffen Luitz [17]; Vincenzo Ciaschini [11]

[1] *University of Ferrara and INFN*

[2] *INFN Ferrara*

[3] *INFN Pisa*

[4] *INFN CNAF, Bologna, Italy*

[5] *Universita degli Studi di Roma Tor Vergata (IT)*

[6] *CNRS*

[7] *INFN Bari*

[8] *INFN Napoli*

[9] *Universita' di Ferrara and INFN Ferrara*

[10] *Universita' di Torino and INFN Torino*

[11] *INFN CNAF*

[12] *INFN-Bari*

[13] *Universita' di Napoli and INFN (IT)*

[14] *INFN LNF*

[15] *INFN Padova*

[16] *INFN*

[17] *SLAC*

Corresponding Authors: tomassetti@fe.infn.it, fabrizio.bianchi@to.infn.it

The SuperB asymmetric energy e+e- collider and detector to be built at the newly founded Nicola Cabibbo Lab will provide a uniquely sensitive probe of New Physics in the flavor sector of the Standard Model. Studying minute effects in the heavy quark and heavy lepton sectors requires a data sample of 75 ab-1 and a luminosity target of 10ˆ36 cm-2 s-1.

Since 2009 the SuperB Computing group is working on developing a simulation production framework capable to satisfy the experiment needs. It provides access to distributed resources in order to support both the detector design definition and the its performance evaluation studies.
During last year the framework has evolved from the point of view of job workflow, Grid services interfaces and technologies adoption.
A complete code refactoring and sub-component language porting now permits the framework to sustain distributed production involving resources from three continents and Grid Flavors.
In this paper we will report a complete description of the production system status of the art, its evolution and its integration with Grid
services; in particular, we will focus on the utilization of new Grid component features as in LB and WMS version 3.
The last official SuperB production cycle has been completed; results and digests will be reported.

Poster Session / **302**

PREP: Production and Reprocessing management tool for CMS

Author: Fabio Cossutti[1]

Co-authors: Dirk Samyn [2]; Fabian Stoeckli [3]; Piergiulio Lenzi [2]

[1] *Universita e INFN (IT)*

[2] *CERN*

[3] *Massachusetts Inst. of Technology (US)*

Corresponding Author: fabio.cossutti@ts.infn.it

The production of simulated samples for physics analysis at LHC represents a noticeable organization challenge, because it requires the management of several thousands different workflows. The submission of a workflow to the grid based computing infrastructure is just the arrival point of a long decision process: definition of the general characteristics of a given set of coherent samples, called campaign; definition of the physics settings to be used for each sample corresponding to a specific process to be simulated, both at hard event generation and detector simulation level. In order to have an organized control of the of the definition of the large number of MC samples needed by CMS, from the initial request to the acknowledgment of the completion of each sample, a dedicated management tool, called PREP, has been built. Its basic component is a databased storing all the relevant information about the sample and the actions implied by the workflow definition, approval and production. A web based interface allows the database to be used from experts involved in production to trigger all the different actions needed, as well as by normal physicists involved in analyses to retrieve the relevant information. The tool is integrated through a set of dedicated APIs with the production agent and information storage utilities of CMS.

Poster Session / **303**

Integrated cluster management at the Manchester Tier-2

Authors: Alessandra Forti[1]; Andrew Mcnab[2]

[1] *University of Manchester (GB)*

[2] *University of Manchester*

Corresponding Author: andrew.mcnab@cern.ch

We describe our experience of operating a large Tier-2 site since 2005 and how we have developed an integrated management system using third-party, open source components. This system tracks individual assets and records their attributes such as MAC and IP addresses; derives DNS and DHCP configurations from this database; creates each host's installation and re-configuration scripts; monitors the services on each host according to the records of what should be running; and cross references tickets with asset records and per-asset monitoring pages. In addition, scripts which detect problems and automatically remove hosts record these new states in the database which are available to operators immediately through the same interface as tickets and monitoring.

Poster Session / 305

Monitoring ARC services with GangliARC

Authors: David Cameron[1]; Dmytro Karpenko[2]

[1] *University of Oslo (NO)*

[2] *University of Oslo*

Corresponding Author: david.cameron@cern.ch

Monitoring of Grid services is essential to provide a smooth experience for users and provide fast and easy to understand diagnostics for administrators running the services. GangliARC makes use of the widely-used Ganglia monitoring tool to present web-based graphical metrics of the ARC computing element. These include statistics of running and finished jobs, data transfer metrics, as well as showing the availability of the computing element and hardware
information such as free disk space left in the ARC cache. Ganglia presents metrics as graphs of the value of the metric over time and shows an easily-digestable summary of how the system is performing, and enables quick and easy diagnosis of common problems. This paper describes how GangliARC works and shows numerous examples of how the generated data can quickly be used by an administrator to investigate problems. It also presents possibilities of combining GangliARC with
other commonly-used monitoring tools such as Nagios to easily integrate ARC monitoring into the regular monitoring infrastructure of any site or computing centre.

Poster Session / 306

Multi-platform Automated Software Building and Packaging

Author: Andres Abad Rodriguez[1]

Co-authors: Alberto Aimar [1]; Alberto Di Meglio [1]; Duarte Bacelar De Begonha De Meneses [2]; Fabio Capannini [3]; Lorenzo Dini [1]; Vitor Emanuel Gomes Gouveia [4]

[1] *CERN*

[2] *LIP Laboratorio de Instrumentacao e Fisica Experimental de Part*

[3] *INFN*

[4] *Universidade de Lisboa*

Corresponding Author: andres.abad.rodriguez@cern.ch

One of the major goals of the EMI (European Middleware Initiative) project is the integration of several components of the pre-existing middleware (ARC, gLite, UNICORE and dCache) into a single

consistent set of packages with uniform distributions and repositories. Those individual middleware projects have been developed in the last decade by tens of development teams and before EMI were all built and tested using different tools and dedicated services. The software, millions of lines of code, is written in several programming languages and supports multiple platforms. Therefore a viable solution ought to be able to build and test applications on multiple programming languages using common dependencies on all selected platforms. It should, in addition, package the resultant software in formats compatible with the popular Linux distributions, such as Fedora and Debian, and store them in repositories from which all EMI software can be accessed and installed in a uniform way.
Despite this highly heterogeneous initial situation, a single common solution, with the aim of quickly automating the integration of the middleware products, had to be selected and implemented in a few months after the beginning of the EMI project. Because of the previous knowledge and the short time available in order to provide this common solution, the ETICS service, where the gLite middleware was already built for years, was selected.
This contribution describes how the team in charge of providing a common EMI build and packaging infrastructure to the whole project has developed a homogeneous solution for releasing and packaging the EMI components from the initial set of tools used by the earlier middleware projects. An important element of the presentation is about the developers experience and feedback on converging on ETICS and on the ongoing work in order to integrate more widely used and supported build and packaging solutions of the Linux platforms.

Poster Session / 307

CMS resource utilization and limitations on the grid after the first two years of LHC collisions

Authors: Andrea Sciaba[1]; Chris Brew[2]; Daniele Bonacorsi[3]; Giuseppe Bagliesi[4]; Ian Fisk[5]; José Flix[None]; Kenneth Bloom[6]; Peter Kreuzer[7]

[1] *CERN*

[2] *STFC - Science & Technology Facilities Council (GB)*

[3] *Universita e INFN (IT)*

[4] *Sezione di Pisa (IT)*

[5] *Fermi National Accelerator Lab. (US)*

[6] *University of Nebraska (US)*

[7] *Rheinisch-Westfaelische Tech. Hoch. (DE)*

Corresponding Author: kenbloom@unl.edu

After years of development, the CMS distributed computing system is now in full operation. The LHC continues to set records for instantaneous luminosity, and CMS records data at 300 Hz. Because of the intensity of the beams, there are multiple proton-proton interactions per beam crossing, leading to larger and larger event sizes and processing times. The CMS computing system has responded admirably to these challenges, but some reoptimization of the computing model has been required to maximize the physics output of the collaboration in the face of increasingly constrained computing resources. We present the current status of the system, describe the recent performance, and discuss the challenges ahead and how we intend to meet them.

Poster Session / 308

JavaFIRE: A Replica and File System for Grids

Author: Marko Petek[1]

Co-authors: Alberto Santoro [2]; Claudio Fernando Resin Geyer [3]; Diego Da Silva Gomes [2]; Stephen Gowdy [4]

[1] *UERJ*
[2] *Universidade do Estado do Rio de Janeiro (BR)*
[3] *UFRGS*
[4] *CERN*

Corresponding Author: stephen.gowdy@cern.ch

The work is focused on the creation and validation tests of a replica and transfers system for Computational Grids

inspired on the needs of the High Energy Physics (HEP).

Due to the high volume of data created by the HEP experiments, an efficient file and dataset replica system may play

an important role on the computing model. Data replica systems allow the creation of copies, distributed between

the different storage elements on the Grid.

In the HEP context, the data files are basically immutable. This eases the task of the replica system, because given

sufficient local storage resources any given dataset only needs to be replicated to a particular site once.

Concurrent with the advent of computational Grids, another important theme in the distributed systems area that has

also seen some significant interest is that of peer-to-peer networks (p2p). P2p networks are an important and

evolving mechanism that facilitates the use of distributed computing and storage resources by end users.

One common technique to achieve faster file downloads from possibly overloaded storage elements over congested

networks is to split the files into smaller pieces. This way, each piece can be transferred from a different

replica, in parallel or not, optimizing the moments in that the network conditions are better suited to the

transfer.

The main tasks achieved by the system are: the creation of replicas, the development of a system for replicas

transfer (RFT) and for replicas location (RLS) with a different architecture that the one provided by Globus and the

development of a system for file transfer in pieces on computational grids with interfaces for several storage

elements.

The RLS uses a p2p overlay based on the Kademlia algorithm.

Poster Session / 309

INFN Tier1 test bed facility.

Authors: Alessandro Cavalli[1]; Andrea Prosperini[2]; Daniele Gregori[3]; Elisabetta Ronchieri[4]; Luca dell'Agnello[1]; Pier Paolo Ricci[2]; Stefano Dal Pra[5]; Vladimir Sapunenko[6]

[1] *INFN-CNAF*

[2] *INFN CNAF*

[3] *Istituto Nazionale di Fisica Nucleare (INFN)*

[4] *Universita e INFN (IT)*

[5] *Unknown*

[6] *INFN*

Corresponding Author: pierpaolo.ricci@cnaf.infn.it

The INFN Tier1 at CNAF is the first level Italian High Energy Physics computing center that shares resources to the scientific community using the grid infrastructure. The Tier1 is composed of a very complex infrastructure divided into different parts: the hardware layer, the storage services, the computing resources (i.e. worker nodes adopted for analysis and other activities) and finally the interconnection layer used for data transfers between different Tiers over the grid. Any update of the different parts of this infrastructure, in particular a software update or a change in the services software code, as the activity of adding new hardware, should be carefully tested and debugged before switching to production. For this reason a test bed facility has beed gradually built in order to reproduce the behaviour of the different layers of the Tier1 in a smaller but meaningful scale. Using this test bed system it is possible to perform extensive testing of both the software and hardware layers and certify them before the use at the Tier1.

Poster Session / 310

Geant4 Graphical User Interface OpenGL developments

Author: Laurent Garnier[1]

[1] *LAL-IN2P3-CNRS*

Corresponding Author: garnier@lal.in2p3.fr

New developments on visualization drivers in Geant4 software toolkit

Summary:

The Geant4 software toolkit simulates the passage of particles through matter. Visualization is a key part of it. Geant4 is used in many application domains including high energy, nuclear and accelerator physics, and in medical and space science. We have developed several visualization drivers, such as OpenInventor, HepRep, DAWN, VRML, RayTracer, ASCIITree, gMocren and OpenGL to fit the various requirements of each domain.

During the last 3 years, the OpenGL suite of visualization drivers has been significantly improved by adding a lot of functionalities, in particular a new OpenGL Qt driver. Qt is a free and well-known toolkit available on all platforms, including Windows, that has enabled us to offer Geant4 visualization that has the same look and feel on all systems. Geant4 release 9.5 integrates the latest improvements in the OpenGL and Qt viewer, including faster first time rendering, integration of multiple visualization frames and the user interface into same window, making posters (thanks to gl2ps), a new Qt viewer components help tree and volume tree, easy creation of videos, “free hand” rotation mode, etc. Thanks to the cmake build system, compiling Geant4 with the Qt viewer is simple. Also use and choice of user interface and visualization drivers has been simplified in all examples.

Poster Session / 311

Performance Tests of CMSSW on the CernVM

Author: Marko Petek[1]

Co-author: Stephen Gowdy [2]

[1] *Universidade do Estado do Rio de Janeiro (BR)*

[2] *CERN*

Corresponding Authors: stephen.gowdy@cern.ch, marko.petek@cern.ch

The CERN Virtual Machine (CernVM) Software Appliance is a project developed in CERN with the goal of allowing the execution of the experiment's software on different operating systems in an easy way for the users. To achieve this it makes use of Virtual Machine images consisting of a JEOS (Just Enough Operational System) Linux image, bundled with CVMFS, a distributed file system for software. This image can this be run with a proper virtualizer on most of the platforms available. It also aggressively caches data on the local user's machine so that it can operate disconnected from the network.

CMS wanted to compare the performance of the CMS Software running in the virtualized environment with the same software running on a native Linux box.

To answer this need a series of tests were made on a controlled environment during 2010-2011. This work presents the results of those tests.

Poster Session / 312

Proof of concept - CMS Computing Model into volunteer computing

Author: Marko Petek[1]

Co-authors: Alan Malta Rodrigues [1]; Samir Cury Siqueira [1]; Sandro Fonseca De Souza [1]

[1] *Universidade do Estado do Rio de Janeiro (BR)*

Corresponding Author: marko.petek@cern.ch

The motivation of this work is about the ongoing efforts to integrate the CMS Computing Model with a project of volunteer computing under development at CERN, the LHC@home, thus allowing the CMS Analysis jobs and Monte Carlo production activities to be executed on this paradigm that has a growing user base.

The LCH@home project allows the use of the CernVM (a virtual machine technology developed at CERN that enables complex simulation code to run easily on the diverse platforms) in an autonomous way on the volunteered machines. To do this it uses on the client side the BOINC (an open-source software platform for computing using volunteered resources) and on the server side the Co-Pilot, a framework developed by the CernVM team that allows to instantiate a distributed computing infrastructure on top of virtualized computing resources.

We developed the plugin which can be adapted in the CRAB submission system to use the Co-Pilot system to the CernVMs running into BOINC, and did a proof of concept of the performance of volunteer computing into CMS Computing Model.

A possible spin-off is to make it easier for CMS (and many other experiments), a submission system interface with the Co-Pilot system based on Globus, that would mimic a regular Grid Computing Element, but instead, would schedule jobs to the BOINC/CernVM Cloud.

Poster Session / 313

LET Estimation for Heavy Ion Particles based on a Timepix-based Si Detector

Author: SON HOANG[1]

Co-authors: Lawrence Pinsky [2]; Ricardo Vilalta [1]

[1] *University of Houston*

[2] *UNIVERSITY OF HOUSTON*

In the quest to develop a Space Radiation Dosimeter based on the Timepix chip from Medipix2 Collaboration, the fundamental issue is how Dose and Dose-equivalent can be extracted from the raw Timepix outputs. To calculate the Dose-equivalent, each type of potentially incident radiation is given a Quality Factor, also referred to as Relative Biological Effectiveness (RBE). As proposed in the National Council on Radiation Protection 153 (2008), the Quality Factor is the function of Linear Energy Transfer (LET) of the traversing particle. With the Timepix chip device, LET can be measured by the total energy deposited –which could be calibrated from the Timepix chip– divided by the path length of traversing when the particle passes through the Si layer. The Si layer is in this case 300 μm thick, which can be used to calculate the traversing length if there is an algorithm to estimate for angular resolution of incidence.

The raw Timepix-outputs are generated from the Medipix2 Timepix-based Si detector, which is a hybrid semiconductor pixel detector made of 256x256 pixels with 55 μm each readout CMOS-based integrated circuit. This device, developed through a collaboration based at CERN, is able to survive and perform for extended periods in strong radiation fields. When an incident heavy ion penetrates the Si layer, it diffuses and produces a core of charge carriers with track structures embedded within the pixel footprint. The structures are complicated due to the diffusion and the existence of δ-rays, which are recoil particles caused by secondary ionization. A carefully analysis of these track structures is needed to understand the characteristic of particles.

In this paper, we propose an algorithm for estimating angular resolution of incident heavy ion particles, which is an essential step toward calculating LET, Dose and Dose-equivalent based on a Timepix-based Si detector. Given the raw Timepix outputs, we use a segmentation operator to identify clusters –groups of contiguous pixels forming track structures. We then apply a morphology operator to get the primary and stable shape of a cluster. By doing this, noise and δ-rays are effectively removed. It also helps to recognize and separate simple overlapping particles that occur when the exposure time is not shortened enough. A linear regression has been found to determine the direction of incidence. We extract a skeleton of the track based on an analysis of pixels projected onto this line. The angular resolution of incidence is dependent on the length of this skeleton. Having the angle and energies calibrated, LET is estimated before getting Dose and Dose-equivalent.

We use data from HIMAC (Heavy Ion Medical ACcelerator) in Chiba, Japan, and NASA Space Radiation Laboratory at Brookhaven in USA for our experiments. The data frames were taken at different angles and particle charges. In particular, we show experimental results with H-100MeV (0, 30, 60, 75 degree), He-100MeV (0, 30, 60, 75 degree), C-400MeV (0, 30, 60, 85 degree), Fe-400MeV (0, 30, 60, 75, 85 degree). Results show the advantage of our algorithm for angular calculation and LET estimation.

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Poster Session / 314

A Grid storage accounting system based on DGAS and HLRmon

Authors: Andrea Cristofori[1]; Andrea Guarise[2]; Enrico Fattibene[1]; Luciano Gaido[2]; Paolo Veronesi[1]

[1] *INFN-CNAF, IGI*

[2] *INFN-TO, IGI*

Corresponding Author: andrea.cristofori@cern.ch

The accounting activity in a production computing Grid is of paramount importance in order to understand the utilization of the available resources. While several CPU accounting systems are deployed within the European Grid Infrastructure (EGI), storage accounting systems, that are stable enough to be adopted on a production environment, are not yet available. A growing interest is being put on the storage accounting and work is being carried out in the Open Grid Forum (OGF) to write a standard Usage Record (UR) definition suitable for this kind of resources.

In this paper we present a storage accounting system which is composed of three parts: a sensors layer, a data repository and transport layer (Distributed Grid Accounting System - DGAS) and a web portal that generates graphical and tabular reports (HLRmon). The sensors layer is responsible for the creation of URs according to the schema that will be presented in the paper and that is being discussed in OGF. DGAS is one of the CPU accounting systems used in EGI, by the Italian Grid Infrastructure (IGI) and other National Grid Initiatives (NGIs) and other projects that relies on the Grid . DGAS is evolving towards an architecture that allows the collection of URs for different resources. Those features allows DGAS to be used as data repository and transport layer of the accounting system we depicted. HLRmon is the web interface for DGAS. It has been further developed to retrieve storage accounting data from the repository and create reports in an easy to access fashion in order to be useful to the Grid stakeholders.

Distributed Processing and Analysis on Grids and Clouds / 315

Offline Processing in the Online Computer Farm

Author: Luis Granado Cardoso[1]

Co-authors: Beat Jost [1]; Clara Gaspar [1]; Guoming Liu [1]; Joel Closier [1]; Markus Frank [1]; Niko Neufeld [1]; Olivier Callot [2]; Philippe Charpentier [1]

[1] *CERN*

[2] *LAL-Orsay (FR)*

Corresponding Author: luis.granado@cern.ch

LHCb is one of the 4 experiments at the LHC accelerator at CERN. LHCb has approximately 1600 (8 cores) PCs for processing the High Level Trigger (HLT) during physics data acquisition. During periods when data acquisition is not required or the resources needed for data acquisition are reduced, like accelerator Machine Development (MD) periods or technical shutdowns, most of these PCs are idle or very little used. In these periods it is possible to profit from the unused processing capacity to reprocess earlier datasets with the newest applications (code and calibration constants), thus reducing the CPU capacity needed on the Grid.
The offline computing environment is based on LHCb-DIRAC (Distributed Infrastructure with Remote Agent Control) to process physics data on the Grid. In DIRAC, agents are started on Worker Nodes, pull available jobs from the DIRAC central WMS (Workload Management System) and process them on the available resources.
A Control System was developed which is able to launch, control and monitor the agents for the offline data processing on the HLT Farm. It can do so without overwhelming the offline resources (e.g. DBs) and in case of change of the accelerator planning it can easily return the used resources

for online purposes. This control system is based on the existing Online System Control infrastructure, the PVSS SCADA and the FSM toolkit. A web server was also developed to provide a highly available and easy view of the status of the offline data processing on the online HLT farm.

Student? Enter 'yes'. See http://goo.gl/MVv53:

No

Poster Session / **316**

Particle Tracking in a Solenoidal Field with an Adaptive Hough Transform

Author: Alan Dion[1]

[1] *Brookhaven National Laboratory*

An algorithm is presented which reconstructs helical tracks in a solenoidal magnetic field using a generalized Hough Transform. While the problem of reconstructing helical tracks from the primary vertex can be converted to the problem of reconstructing lines (with 3 parameters), reconstructing secondary tracks requires a full helix to be used (with 5 parameters). The Hough transform memory requirements typically grow exponentially with the number of parameters. To reduce the amount of memory used, this algorithm adapts the granularity of the accumulator array depending on the given distribution of detector hits. Furthermore, only a small portion of the accumulator array needs to be explicitly stored at a time. It will be shown that the time required for event reconstruction of the presented algorithm grows more slowly asymptotically as a function of the number of detector hits in the event than the time required for road-finding techniques. In addition, the algorithm is easliy implemented in a cache-oblivious manner. Thus, the presented adaptive Hough Transform is well-suited for reconstruction of the high-multiplicity events in heavy ion collisions.

Results of the algorithm will be shown for heavy ion collisions in various simulated detectors, as well as on data from the PHENIX Silicon Vertex Detector.

Poster Session / **317**

Improving ATLAS grid site reliability with functional tests using HammerCloud

Author: Computing Atlas[1]

Co-authors: Daniel Colin Van Der Ster [2]; Federica Legger [3]; Gianfranco Sciacca [4]; Johannes Elmsheuser [5]; Ramon Medrano Llamas [6]

[1] *Atlas*

[2] *CERN*

[3] *Ludwig-Maximilians-Univ. Muenchen*

[4] *Universitaet Bern (CH)*

[5] *Ludwig-Maximilians-Univ. Muenchen (DE)*

[6] *Universidad de Oviedo (ES)*

Corresponding Author: federica.legger@physik.uni-muenchen.de

With the exponential growth of LHC (Large Hadron Collider) data in 2011, and more to come in 2012, distributed computing has become the established way to analyse collider data. The ATLAS grid infrastructure includes more than 80 sites worldwide, ranging from large national computing centers to smaller university clusters. These facilities are used for data reconstruction and simulation, which are centrally managed by the ATLAS production system, and for distributed user analysis. To ensure the smooth operation of such a complex system, regular tests of all sites are necessary to validate the site capability of successfully executing user and production jobs. We report on the development, optimization and results of an automated functional testing suite using the HammerCloud framework. Functional tests are short light- weight applications covering typical user analysis and production schemes, which are periodically submitted to all ATLAS grid sites. Results from those tests are collected and used to evaluate site performances. Sites that fail or are unable to run the tests are automatically excluded from the PanDA brokerage system, therefore avoiding user or production jobs to be sent to problematic sites.
We show that stricter exclusion policies help to increase the grid reliability, and the percentage of user and production jobs aborted due to network or storage failures can be sensibly reduced using such a system.

Poster Session / 318

Management of virtualized infrastructure for databases in HEP

Author: Anton Topurov[1]

Co-author: Mariusz Piorkowski

[1] *CERN*

Corresponding Author: anton.topurov@cern.ch

As elsewhere in today's computing environment, virtualisation is becoming prevalent in the database management area where HEP laboratories, and industry more generally, seek to deliver improved services whilst simultaneously increasing efficiency. We present here our solutions for the effective management of virtualised databases, building on over five years of experience dating back to studies with the Xen hypervisor in early 2006.
After reviewing the evolving functionality of virtualisation solutions for database and middle tier and their associated virtualisation management applications, we will present CERN's solutions for managing our virtualized database infrastructure efficiently and explain how these solutions have enabled us to meet the rapidly increasing demands for database storage of physics metadata with an improved service availability.

Poster Session / 319

Key developments of the Ganga task-management framework.

Authors: Alexander John Richards[1]; Ivan Antoniev Dzhunov[2]; Jakub Moscicki[3]; Mark William Slater[4]; Michael John Kenyon[3]

Co-authors: Daniel Colin Van Der Ster [3]; Frederic Brochu [5]; Hurng-Chun Lee [6]; J Michael Williams [1]; Johannes Ebke [7]; Johannes Elmsheuser [7]; Manoj Jha [8]; Ulrik Egede [1]

[1] *Imperial College Sci., Tech. & Med. (GB)*

[2] *University of Sofia*

[3] *CERN*

[4] *University of Birmingham (GB)*

[5] *University of Cambridge (GB)*

[6] *NIKHEF (NL)*

[7] *Ludwig-Maximilians-Univ. Muenchen (DE)*

[8] *Universita e INFN (IT)*

Corresponding Author: mkenyon@cern.ch

Ganga is an easy-to-use frontend for the definition and management of analysis jobs, providing a uniform interface across multiple distributed computing systems. It is the main end-user distributed analysis tool for the ATLAS and LHCb experiments and provides the foundation layer for the HammerCloud sytem, used by the LHC experiments for validation and stress testing of their numerous distributed computing facilities.

This poster will illustrate recent developments aimed at improving both the efficiency with which computing resources are utilised, and the end-user experience. Notable highlights include a new web-based monitoring interface (WebGUI) that allows users to conveniently view the status of their submitted Ganga jobs and browse the local job repository. Improvements to the core Ganga package will also be outlined. Specifically we will highlight the development of procedures for automatic handling and resubmission of failed jobs, alongside a mechanism that stores an analysis application such that it can be repeated (optionally using different input data) at any point in the future.

We will demonstrate how tools that were initially developed for a specific user community have been migrated into the Ganga core, and so can be exploited by a wider user-base. Similarly, examples will be given where Ganga components have been adapted for use by communities in their custom analysis packages.

Student? Enter 'yes'. See http://goo.gl/MVv53:

No

Summary:

An overview of recent Ganga developments, stressing improvements to the user-experience and demonstrating how originally community-specific tools have been adapted for use by a wider user-base.

Poster Session / 320

CREAM Computing Element: a status update

Author: Massimo Sgaravatto[1]

Co-authors: Alessio Gianelle [2]; Alvise Dorigo [3]; Eric Frizziero [4]; Fabio Capannini [5]; Luigi Zangrando [6]; Marco Cecchi [7]; Paolo Andreetto [8]; Salvatore Monforte [8]; sara bertocco [9]

[1] *Universita e INFN (IT)*

[2] *Istituto Nazionale di Fisica Nucleare (INFN)*

[3] *INFN PADOVA*

[4] *Laboratori Nazionali di Legnaro*

[5] *Unknown-Unknown-Unknown*

[6] *INFN*

[7] *Istituto Nazionale Fisica Nucleare (IT)*

[8] *Unknown*

[9] *INFN-PD*

Corresponding Author: massimo.sgaravatto@pd.infn.it

The European Middleware Initiative (EMI) project aims to deliver a consolidated set of middleware products based on the four major middleware
providers in Europe - ARC, dCache, gLite and UNICORE.

The CREAM (Computing Resource Execution And Management) Service, a service for job management operation at the Computing Element (CE) level, is one of the software product part of the EMI middleware distribution.
In this paper we discuss about some new functionality in the CREAM CE
introduced with the first EMI major release (EMI-1, codename Kebnekaise).
The integration with the Argus authorization service is one of these implementations: the use of a unique authorization system, besides
simplying the overall management, allows also to avoid inconsistent authorization decisions.
An improved support for complex deployment scenarios (e.g. for sites having multiple CE head nodes and/or having heterogeneous resources)
is another new achievement.
The improved support for resource allocation in a multicore environments, and the initial support of version 2.0 of the Glue specification for resource
publication are other new functionality introduced with the first EMI release.

Poster Session / **321**

New developments in the CREAM Computing Element

Authors: Alessio Gianelle[1]; Alvise Dorigo[2]; Eric Frizziero[3]; Fabio Capannini[4]; Luigi Zangrando[5]; Marco Cecchi[6]; Massimo Sgaravatto[7]; Paolo Andreetto[8]; Salvatore Monforte[8]; sara bertocco[9]

[1] *Istituto Nazionale di Fisica Nucleare (INFN)*

[2] *INFN PADOVA*

[3] *Laboratori Nazionali di Legnaro*

[4] *Unknown-Unknown-Unknown*

[5] *INFN*

[6] *Istituto Nazionale Fisica Nucleare (IT)*

[7] *Universita e INFN (IT)*

[8] *Unknown*

[9] *INFN-PD*

Corresponding Author: massimo.sgaravatto@pd.infn.it

The EU-funded project EMI, now at its second year, aims at providing a unified, standardized, easy to install software for distributed computing infrastructures.
CREAM is one of the middleware product part of the EMI middleware distribution:
it implements a Grid job management service which allows the submission, management and monitoring of computational jobs to local resource management systems.
In this paper we discuss about some new features being implemented in the CREAM Computing Element.
The implementation of the EMI Execution Service (EMI-ES) specification (an agreement in the EMI consortium on interfaces and protocols to be
used in order to enable computational job submission and management
required across technologies) is one of the new functionality being implemented.
New developments are also focusing in the High Availability (HA) area, to
improve performance, scalability, availability and fault tolerance.

Poster Session / **322**

The Memory of MICE, the Configuration Database

Author: Antony Wilson[1]

Co-authors: David Colling [2]; Pierrick Hanlet [3]; on behalf of the MICE Collaboration [4]

[1] *STFC - Science & Technology Facilities Council (GB)*

[2] *Imperial College Sci., Tech. & Med. (GB)*

[3] *Illinois Institute of Technology*

[4] *MICE*

Corresponding Author: antony.wilson@stfc.ac.uk

The configuration database (CDB) is the memory of the Muon Ionisation Cooling Experiment (MICE). Its principle aim is to store temporal data associated with the running conditions of the experiment. These data can change on a per run basis (e.g. magnet currents, high voltages), or on long time scales (e.g. cabling, calibration, and geometry). These data are used throughout the life cycle of experiment, from running the experiment through to data analysis and reconstruction.
The CDB has expanded from its initial use to form an essential part of the MICE state machine as used by the controls and monitoring system. The state of MICE: off, testing, running, etc., dictates the possible states of a hierarchy of sub-systems. The CDB stores information about allowed combinations of states along with allowed settings for all controls for every state.
Master and slave CDBs have been set up in different parts of the site to increase resilience. Both machines have multiple mirrored pair raid arrays, with the data stored on one mirrored pair and the transaction logs stored on another mirrored pair of each machine. Off site backups of the data are also kept. Access to the CDB is via a Python API, which communicates with a WSDL interface provided by a web-service on the CDB.
The priority is to ensure availability of the CDB to the control room systems. The master CDB is located in the control area where it is only used by the running experiment. In the event of the failure of the master, the slave can be promoted and the control room services can be switched to use the use the new master. Read only access to the CDB for data analysis and reconstruction is provided by the slave which has an up to the minute copy of the data.
MICE is a precision experiment, it is imperative that we minimize our systematic errors; the CDB will ensure reproducible and documented running conditions in a highly resilient manner. This information is crucial to the running of the experiment and understanding the experimental data.

Poster Session / **323**

Hybrid C++/Python components for physics analysis and trigger

Author: Ivan BELYAEV[1]

Co-authors: Gerhard Raven [2]; Juan Palacios [3]; Patrick Koppenburg [4]

[1] *ITEP/MOSCOW*

[2] *Free University (NL)*

[3] *CERN*

[4] *NIKHEF (NL)*

Corresponding Author: ivan.belyaev@itep.ru

A hybrid C++/Python environment built from the standard components is being heavily and successfully used in LHCb, both for off-line physics analysis as well as for the High Level Trigger. The approach is based on the LoKi toolkit and the Bender analysis framework. A small set of highly configurable C++ components allows
to describe the most frequirent analysis tasks, e.g. combining and filtering of particles, in compact & coherent way.
The action of these components is defined at an initialization phase using a palette of C++/Python functors, provided by LoKi/Bender
framework, using the full power of the python language. The C++/Python binding and intercommunications have been performed
using Reflex dictionaries. The system is currently being exteded to cover all
steps of the High Level Trigger, thus providing a coherent solution for the whole trigger and analysis chain.

We shall describe the overal design and key features of the major C++/Python analysis&trigger components.

Poster Session / **324**

A PROOF Analysis Framework

Authors: Ana Rodriguez Marrero[1]; Isidro Gonzalez Caballero[2]

Co-authors: Alberto Cuesta Noriega [3]; Enol Fernandez Del Castillo [4]

[1] *Universidad de Cantabria (ES)*

[2] *Universidad de Oviedo (ES)*

[3] *Universidad de Oviedo*

[4] *CERN*

Corresponding Author: isidro.gonzalez.caballero@cern.ch

The analysis of the complex LHC data usually follows a standard path that aims at minimizing not only the amount of data but also the number of observables used. After a number of steps of slimming and skimming the data, the remaining few terabytes of ROOT files hold a selection of the events and a flat structure for the variables needed that can be more easily inspected and traversed in the final stages of the analysis. PROOF arises at this point as an efficient mechanism to distribute the analysis load by taking advantage of all the cores in modern CPUs through PROOF Lite, or by using PROOF Cluster or PROOF on Demand tools to build dynamic PROOF cluster on computing facilities with spare CPUs. However using PROOF at the level required for a serious analysis introduces some difficulties that may scare new adopters. We have developed the PROOF Analysis Framework (PAF) to facilitate the development of new analysis by uniformly exposing the PROOF related configurations across technologies and by taking care of the routine tasks as much as possible. We describe the details of the PAF implementation as well as how we succeeded in engaging a group of CMS physicists to use PAF as their daily analysis framework.

Poster Session / **325**

Atlas Analysis and Conference Notes

Author: Collaboration Atlas[1]

Co-authors: Bruno Lange Ramos [2]; Carmen Maidantchik [2]; Felipe Fink Grael [3]; Kaio Karam Galvao [4]; Kathy Pommes [5]; Laura De Oliveira Fernandes Moraes [2]; Luiz Fernando Cagiano Parodi De Frias [2]; Luiz Henrique Ramos De Azevedo Evora [2]

[1] *Atlas*

[2] *Univ. Federal do Rio de Janeiro (BR)*

[3] *Univ. Federal do Rio de Janeiro (UFRJ)*

[4] *Instituto de Física / Universidade Federal do Rio de Janeiro*

[5] *CERN*

Corresponding Author: luiz.fernando.cagiano.parodi.de.frias@cern.ch

In 2010, the LHC experiment produced 7 TeV and heavy-ions collisions continually, generating a huge amount of data, which was analyzed and reported throughout several performed studies. Since then, physicists are bringing out papers and conference notes announcing results and achievements. During 2010, 37 papers and 102 conference notes were published and until September 2011 there are already 131 papers and 189 conference notes in preparation.
This paper presents the ATLAS Analysis Papers and ATLAS Analysis Conference Notes systems,

developed to monitor the entire publication procedure up to the final submission and to promote the communication among the collaboration members. The software supports the paper elaboration process, tracking the analysis results status and improvement of the paper initial version, presenting a step-by-step procedure overview and promoting communication among collaborators.
Along with the increasing flow of papers and conference notes, one of the issues is the way to guarantee that all members who participate in the analysis studies are aware of not only the discussion deadlines but also of the publication process, which involves 17 steps, split in 3 different phases for papers and 10 steps in 1 phase for conference notes. By sending notifications based on predefined rules the systems inform members to approve each step and provide further information such as the approval conditions and the documents in which the publication is based on. Through the software it is also possible to manage dates and members of the editorial team. The data processing is performed by using the Glance System, the main data retrieval platform used for ATLAS information management.

Poster Session / **326**

Increasing performance in KVM virtualization within a Tier-1 environment

Author: Andrea Chierici[1]

Co-author: Davide Salomoni [1]

[1] *INFN-CNAF*

Corresponding Author: chierici@cnaf.infn.it

This work shows the optimizations we have been investigating and implementing at the KVM virtualization layer in the INFN Tier-1 at CNAF, based on more than a year of experience in running thousands of virtual machines in a production environment used by several international collaborations. These optimizations increase the adaptability of virtualization solutions to demanding applications like those run in our institute (High-Energy Physics).

We will show performance differences among different filesystems (like ext3 vs ext4 vs xfs) and caching options, when used as KVM host local storage. We will provide guidelines for solid state disks (SSD) adoption, for deployment of SR-IOV enabled hardware, for providing PCI-passthrough network cards to virtual machines and what is the best solution to distribute and instantiate read-only virtual machine images.

This work has been driven by the project called Worker Nodes on Demand Service (WNoDeS), a framework designed to offer local, grid or cloud-based access to computing and storage resources, preserving maximum compatibility with existing computing center policies and work-flows.

Summary:

We will show performance differences among different filesystems (like ext3 vs ext4 vs xfs) and caching options, when used as KVM host local storage. We will provide guidelines for solid state disks (SSD) adoption, for deployment of SR-IOV enabled hardware, for providing PCI-passthrough network cards to virtual machines and what is the best solution to distribute and instantiate read-only virtual machine images.

Poster Session / 328

Big data log mining: the key to efficiency

Author: Paul Rossman[1]

[1] *Fermi National Accelerator Laboratory (FNAL)*

Corresponding Author: rossman@fnal.gov

In addition to the physics data generated each day from the CMS detector, the experiment also generates vast quantities of supplementary log data. From reprocessing logs to transfer logs this data could shed light on operational issues and assist with reducing inefficiencies and eliminating errors if properly stored, aggregated and analyzed. The term "big data" has recently taken the spotlight with organizations worldwide using tools such as CouchDB, Hadoop and Hive. In this paper we present a way of evaluating the capture and storage of log data from various experiment components to provide analytics and visualization in near real time.

Poster Session / 329

AutoPyFactory: A Scalable Flexible Pilot Factory Implementation

Author: Collaboration Atlas[1]

Co-authors: Graeme Andrew Stewart [2]; John Hover [3]; Jose Caballero Bejar [4]; Peter Love [5]

[1] *Atlas*

[2] *CERN*

[3] *Brookhaven National Laboratory (BNL)-Unknown-Unknown*

[4] *Brookhaven National Laboratory (US)*

[5] *LANCASTER UNIVERSITY*

Corresponding Author: jose.caballero@cern.ch

The ATLAS experiment at the CERN LHC is one of the largest users of grid computing infrastructure, which is a central part of the experiment's computing operations. Considerable efforts have been made to use grid technology in the most efficient
and effective way, including the use of a pilot job based workload management framework.
In this model the experiment submits 'pilot' jobs to sites without payload. When these
jobs begin to run they contact a central service to pick-up a real payload to execute.
The first generation of pilot factories were usually specific to a single VO, and were bound to the particular architecture of that VO's distributed processing. A second
generation provides factories which are more flexible, not tied to any particular VO,
and provide new or improved features such as monitoring, logging, profiling, etc.
In this paper we describe this key part of the ATLAS pilot architecture, a second generation pilot factory, AutoPyFactory.
AutoPyFactory has a modular design and is highly configurable. It is able to send different types of pilots to sites and exploit different submission mechanisms and queue characteristics. It is tightly integrated with the PanDA job submission framework, coupling pilot flow to the amount of work the site has to run. It gathers information from many sources in order to correctly configure itself for a site, and its decision logic can easily be updated.
Integrated into AutoPyFactory is a flexible system for delivering both generic and specific job wrappers which can perform many useful actions before starting to run end-user scientific applications, e.g. validation of the middleware, node profiling and diagnostics, and monitoring.
AutoPyFactory now also has a robust monitoring system and we show how this has helped establish a reliable pilot factory service for ATLAS.

Poster Session / 330

Popularity framework for monitoring user workload

Author: Collaboration Atlas[1]

Co-authors: Angelos Molfetas [2]; Cedric Serfon [3]; Graeme Andrew Stewart [2]; Mario Lassnig [2]; Martin Barisits [4]; Vincent Garonne [2]

[1] *Atlas*

[2] *CERN*

[3] *Ludwig-Maximilians-Univ. Muenchen (DE)*

[4] *Vienna University of Technology (AT)*

Corresponding Author: vincent.garonne@cern.ch

This paper describes a user monitoring framework for very large data management systems that maintain high numbers of data movement transactions. The proposed framework prescribes a method for generating meaningful information from collected tracing data that allows the data management system to be queried on demand for specific user usage patterns in respect to source and destination locations, period intervals, and other searchable parameters.
The feasibility of such a system at the petabyte scale is demonstrated by describing
the implementation and operational experience of an enterprise information system employing the proposed framework that uses data movement traces collected by the ATLAS data management system for operations occurring on the Worldwide LHC Computing Grid (WLCG). Our observations suggest that the proposed user
monitoring framework is capable of scaling to meet the needs of very large data management systems.

Poster Session / 331

ATLAS job monitoring in the Dashboard Framework

Author: Collaboration Atlas[1]

Co-authors: David Tuckett [2]; Edward Karavakis [2]; Jaroslava Schovancova [3]; Julia Andreeva [2]; Laura Sargsyan [4]; Lukasz Kokoszkiewicz [2]; Simone Campana [2]

[1] *Atlas*

[2] *CERN*

[3] *Acad. of Sciences of the Czech Rep. (CZ)*

[4] *A.I. Alikhanyan National Scientific Laboratory (AM)*

Corresponding Author: laura.sargsyan@cern.ch

Monitoring of the large-scale data processing of the ATLAS experiment includes monitoring of production and user analysis jobs.
Experiment Dashboard provides a common job monitoring solution, which is shared by ATLAS and CMS experiments. This includes an accounting portal as well as real-time monitoring.
Dashboard job monitoring for ATLAS combines information from the Panda job processing DB, Production system DB and monitoring information from jobs submitted through Ganga to WMS or local batch systems. Usage of Dashboard-based job monitoring applications will decrease load on the PanDA DB and overcome scale limitations in PanDA monitoring caused by the short job rotation cycle in the PanDA DB. Aggregation of the task/job metrics from different sources will provide complete view of job processing in scope of ATLAS.
The presentation will describe the architecture, functionality and the future plans of the new monitoring applications, including the accounting portal and task monitoring for production and analysis users.

Poster Session / 332

ATLAS Distributed Computing Monitoring tools after full 2 years of LHC data taking

Author: Collaboration Atlas[1]

[1] *Atlas*

Corresponding Author: jaroslava.schovancova@cern.ch

This talk details variety of Monitoring tools used within the ATLAS Distributed Computing during the first 2 years of LHC data taking. We discuss tools used to monitor data processing from the very first steps performed at the Tier-0 facility at CERN after data is read out of the ATLAS detector, through data transfers to the ATLAS computing centers distributed world-wide. We present an overview of monitoring tools used daily to track ATLAS Distributed Computing activities ranging from network performance and data transfers throughput, through data processing and readiness of the computing services at the ATLAS computing centers, to the reliability and usability of the ATLAS computing centers. Described tools provide monitoring for issues of different level of criticality: from spotting issues with the instant online monitoring to the long-term accounting information.

Poster Session / 333

Automating ATLAS Computing Operations using the Site Status Board

Author: Collaboration Atlas[1]

Co-authors: Alessandro Di Girolamo [2]; Carlos Borrego Iglesias [3]; Erekle Magradze [4]; Graeme Andrew Stewart [2]; Jaroslava Schovancova [5]; Julia Andreeva [2]; Lorenzo Rinaldi [6]; Michael Wright [7]; Michal Maciej Nowotka [8]; Pablo Saiz [2]; Simone Campana [2]; Stavro Gayazov [9]; Xavier Espinal Curull [10]

[1] *Atlas*

[2] *CERN*

[3] *IFAE*

[4] *Georg-August-Universitaet Goettingen (DE)*

[5] *Acad. of Sciences of the Czech Rep. (CZ)*

[6] *INFN CNAF*

[7] *Department of Physics and Astronomy-University of Glasgow*

[8] *Warsaw University of Technology (PL)*

[9] *Budker Institute of Nuclear Physics (RU)*

[10] *Universitat Autònoma de Barcelona (ES)*

Corresponding Author: erekle.magradze@cern.ch

The automation of operations is essential to reduce manpower costs and improve the reliability of the system. The Site Status Board (SSB) is a framework which allows Virtual Organizations to monitor their computing activities at distributed sites and to evaluate site performance.
The ATLAS experiment intensively uses SSB for the distributed computing shifts, for estimating data processing and data transfer efficiencies at a particular site, and for implementing automatic exclusion of sites from computing activities, in case of potential problems. ATLAS SSB provides a real-time aggregated monitoring view and keeps the history of the monitoring metrics. Based on this history, usability of a site from the perspective of ATLAS is calculated.
The presentation will describe how SSB is integrated in the ATLAS operations and computing infrastructure and will cover implementation details of the ATLAS SSB sensors and alarm system, based

on the information in SSB. It will demonstrate the positive impact of the use of SSB on the overall performance of ATLAS computing activities and will overview future plans.

Poster Session / **334**

CMS CSC Expert System: towards the detector control automation

Authors: Evaldas Juska[1]; Valdas Rapsevicius[1]

Co-author: Karoly Banicz [1]

[1] *Fermi National Accelerator Lab. (US)*

Corresponding Author: evaldas.juska@cern.ch

Cathode strip chambers (CSC) compose the endcap muon system of the CMS experiment at the LHC. Two years of data taking have proven that various online systems like Detector Control System (DCS), Data Quality Monitoring (DQM), Trigger, Data Acquisition (DAQ) and other specialized applications are doing their task very well. But the need for better integration between these systems is starting to emerge. Automatic and fast problem identification and resolution, tracking detector performance trend, maintenance of known problems, current and past detector status and alike tasks are still hard to handle and require a lot of efforts from many experts. Moreover, this valuable expert knowledge is not always well documented.

CSC Expert System prototype is aiming to fill in these gaps and provides a solution for online systems integration and automation. Its design is based on solid industry standards –Service Bus and Application Integration, Data Warehouse and Online analytical processing (OLAP), Complex Event Processing (CEP, i.e. Rule Engine) and ontology based Knowledge Base. CSC Expert system receives and accumulates Facts (i.e. detector status, conditions, shifter/expert actions), derives and manages Conclusions (i.e. hot device, masked chamber, weak HV segment, high radiation background), stores detector inventory –Assets (i.e. hardware, software, links) and outputs Conclusions, Facts and Assets for other applications and users. CEP engine allows experts to describe their valuable knowledge in SQL-like language and to execute it taking subsequent action in real time (e.g. sends emails, SMS'es, commands and fact requests to other applications, raise alarms).

A year of running the CSC Expert System has proven the correctness of the solution and displays its applicability in detector control automation.

Summary:

Cathode strip chambers (CSC) compose the endcap muon system of the CMS experiment at the LHC. Two years of data taking have proven that various online systems like Detector Control System (DCS), Data Quality Monitoring (DQM), Trigger, Data Acquisition (DAQ) and other specialized applications are doing their task very well. But the need for better integration between these systems is starting to emerge. Automatic and fast problem identification and resolution, tracking detector performance trend, maintenance of known problems, current and past detector status and alike tasks are still hard to handle and require a lot of efforts from many experts. Moreover, this valuable expert knowledge is not always well documented.

CSC Expert System prototype, which is based on solid industry standards, is aiming to fill in these gaps and provides a solution for online systems integration and automation. A year of running the CSC Expert System has proven the correctness of the solution and displays its applicability to this task.

Poster Session / **335**

Application of rule based data mining techniques to real time ATLAS Grid job monitoring data

Author: Collaboration Atlas[1]

Co-authors: Frank Volkmer [2]; Marisa Sandhoff [3]; Raphael Ahrens ; Sergey Kalinin [3]; Tim Dos Santos [3]; Torsten Harenberg [4]

[1] *Atlas*

[2] *Fachbereich C / Physik-Bergische Universitaet Wuppertal-Unknown*

[3] *Bergische Universitaet Wuppertal (DE)*

[4] *UNIVERSITY OF WUPPERTAL*

Corresponding Author: sergey.kalinin@cern.ch

The Job Execution Monitor (JEM), a job-centric grid job monitoring software, is actively developed at the University of Wuppertal. It leverages Grid-based physics analysis and Monte Carlo event production for the ATLAS experiment by monitoring job progress and grid worker node health. Using message passing techniques, the gathered data can be supervised in real time by users, site admins and shift personnel.
Imminent error conditions can be detected early and countermeasures taken by the Job's owner. Grid site admins can access aggregated data of all monitored jobs to infer the site status and to detect job and Grid worker node misbehavior. Shifters can use the same aggregated data to quickly react to site error conditions and broken production jobs. JEM is integrated into ATLAS' Pilot-based "PanDA" job brokerage system.
In this work, the application of novel data-centric rule based methods and data-mining techniques to the real time monitoring data is discussed. The usage of such automatic inference techniques on monitoring data to provide job- and site-health summary information to users and admins is presented. Finally, the provision of a secure real-time control- and steering channel to the job as extension of the presented monitoring software is considered and a possible architecture is shown

Distributed Processing and Analysis on Grids and Clouds / 336

The ATLAS Distributed Data Management project: Past and Future

Author: Collaboration Atlas[1]

Co-authors: Angelos Molfetas [2]; Graeme Andrew Stewart [2]; Mario Lassnig [2]; Martin Barisits [3]; Thomas Beermann [4]; Vincent Garonne [2]

[1] *Atlas*

[2] *CERN*

[3] *Vienna University of Technology (AT)*

[4] *Bergische Universitaet Wuppertal (DE)*

Corresponding Author: vincent.garonne@cern.ch

The ATLAS collaboration has recorded almost 5PB of RAW data since the LHC started running at the end of 2009. Together with experimental data generated from RAW and complimentary simulation data, and accounting for data replicas on the grid, a total of 74TB is currently stored in the Worldwide LHC Computing Grid by ATLAS. All of this data is managed by the ATLAS Distributed Data Management system, called Don Quixote 2 (DQ2).
The DQ2 system has over time rapidly evolved to assist the ATLAS collaboration management to properly manage the data, as well as provide an effective interface allowing physicists easy access to this data. Numerous
new requirements and operational experience of ATLAS' use cases have necessitated the need for a next generation data management system, called Rucio, which will re-engineer the current system to cover new high-level use cases and workflows such as the management of data for physics groups.

In this talk, we will describe the state of the current of DQ2, and present an overview of the upcoming Rucio system, covering it's architecture, new innovative features, and preliminary benchmarks.

Poster Session / **337**

The ATLAS DDM Tracer monitoring framework

Author: Collaboration Atlas[1]

Co-authors: Angelos Molfetas [2]; Donal Zang [3]; Graeme Andrew Stewart [2]; Mario Lassnig [2]; Martin Barisits [4]; Thomas Beermann [5]; Vincent Garonne [2]

[1] *Atlas*

[2] *CERN*

[3] *Chinese Academy of Sciences (CN)*

[4] *Vienna University of Technology (AT)*

[5] *Bergische Universitaet Wuppertal (DE)*

Corresponding Author: vincent.garonne@cern.ch

The DDM Tracer Service is aimed to trace and monitor the atlas file operations on the Worldwide LHC Computing Grid. The volume of traces has increased significantly since the service started in 2009. Now there are about ˜5 million trace messages every day and peaks of greater than 250Hz, with peak rates continuing to climb, which gives the current service structure a big challenge.
Analysis of large datasets based on on-demand queries to the relational database management system (RDBMS), i.e. Oracle, can be problematic, and have a significant effect on the database's performance. Consequently, We
have investigated some new high availability technologies like messaging infrastructure, specifically ActiveMQ, and key-value stores. The advantages of key value store technology are that they are distributed and have high scalability; also their write performances are usually much better than RDBMS, all of which are very useful for the Tracer service.
Indexes and distributed counters have been also tested to improve query performance and provided almost real time results.
In this talk, the design principles, architecture and main characteristics of Tracer monitoring framework will be described and examples of its usage will be presented.

Poster Session / **338**

Executor framework for DIRAC

Author: Adrian Casajus Ramo[1]

Co-author: Ricardo Graciani Diaz [1]

[1] *University of Barcelona (ES)*

Corresponding Author: adria@ecm.ub.es

DIRAC framework for distributed computing has been designed as a group of collaborating components, agents and servers, with persistent database back-end. Components communicate with each other using DISET, an in-house protocol that provides Remote Procedure Call (RPC) and file transfer capabilities. This approach has provided DIRAC with a modular and stable design by enforcing stable interfaces across releases. But it made complicated to scale further with commodity hardware.

To further scale DIRAC, components needed to send more queries between them. Using RPC to do so requires a lot of processing power just to handle the secure handshake required to stablish the connection. DISET now provides a way to keep stable connections and send and receive queries between components. Only one handshake is required to send and receive any number of queries. Using this new communication mechanism DIRAC now provides a new type of component called executor. Executors process any task (such as resolving the input data of a job) sent to them by a task dispatcher. This task dispatcher takes care of persisting the state of the tasks to the storage backend and distributing them amongst all the executors based on the requirements of each task.

In case of a high load, several executors can be started to process the extra load and stop them once the tasks have been processed. This new approach of handling tasks in DIRAC makes executors easy to replace and replicate, thus enabling DIRAC to further scale beyond the current approach based on polling agents.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Poster Session / **339**

AGIS: The ATLAS Grid Information System

Author: Collaboration Atlas[1]

Co-authors: Alessandro Di Girolamo [2]; Alexander Senchenko [3]; Alexei Klimentov [4]; Alexey Anisenkov [3]

[1] *Atlas*

[2] *CERN*

[3] *Budker Institute of Nuclear Physics (RU)*

[4] *Brookhaven National Laboratory (US)*

Corresponding Author: alexey.anisenkov@cern.ch

The ATLAS Grid Information System (AGIS) centrally stores and exposes static, dynamic and configuration parameters required to configure and to operate ATLAS distributed computing systems and services. AGIS is designed to integrate information about resources, services and topology of the ATLAS grid infrastructure from various independent sources including BDII, GOCDB, the ATLAS data management system and the ATLAS PanDA workload management system.
Being an intermediate middleware system between a client and external information sources, AGIS automatically collects and keeps data up to date, caching information required by and specific for ATLAS, removing the source as a direct dependency for clients but without duplicating the source information system itself. All interactions with various information providers are hidden. Synchronization of AGIS content with external sources is performed by agents which periodically communicate with sources via standard interfaces and update database content. For some types of information AGIS is itself the primary repository. AGIS stores data objects in a way convenient for ATLAS, introduces additional object relations required by ATLAS applications, exposes the data via API and web front end services.
Through the API clients are able to update information stored in AGIS. A python API and command line tools further help end users and developers use the system conveniently. Web interfaces such as a site downtime calendar and ATLAS topology viewers are widely used by shifters and data distribution experts.

Poster Session / **340**

Integration of Globus Online with the ATLAS PanDA Workload Management System

Author: Collaboration Atlas[1]

Co-authors: Carlos Contreras [2]; Maxim Potekhin [3]; Paul Nilsson [4]; Tadashi Maeno [3]

[1] *Atlas*

[2] *Departamento de Fisica-Univ. Tecnica Federico Santa Maria (UTFSM*

[3] *Brookhaven National Laboratory (US)*

[4] *University of Texas at Arlington (US)*

Corresponding Author: maxim.potekhin@cern.ch

The PanDA Workload Management System is the basis for distributed production and analysis for the ATLAS experiment at the LHC. In this role, it relies on sophisticated dynamic data movement facilities developed in ATLAS.
In certain scenarios, such as small research teams in ATLAS Tier-3 sites and non-ATLAS Virtual Organizations supported by the Open Science Grid consortium (OSG), the overhead of installation and operation of this component makes its use not cost effective. Globus Online is an emerging new tool from the Globus Alliance, which already proved popular within the OSG community. It provides the users with fast and robust file transfer capabilities that can also be managed from a Web interface,
and in addition to grid sites, can have individual workstations and laptops serving as data transmission endpoints. We will describe the integration of the Globus Online functionality into the PanDA suite of software, in order to give more flexibility in choosing the method of data transfer to ATLAS Tier-3 and OSG users.

Computer Facilities, Production Grids and Networking / 341

CMS Data Transfer operations after the first years of LHC collisions

Author: Rapolas Kaselis[1]

Co-authors: Andrea Sartirana [2]; José Flix ; Markus Klute [3]; Nicolo Magini [4]; Oliver Gutsche [5]; Peter Kreuzer [6]; Stefan Piperov [7]

[1] *Vilnius University (LT)*

[2] *Ecole Polytechnique (FR)*

[3] *Massachusettes Institute of Technology*

[4] *CERN*

[5] *FERMILAB*

[6] *Rheinisch-Westfaelische Tech. Hoch. (DE)*

[7] *INRNE/FermiLab*

Corresponding Author: rapolas.kaselis@cern.ch

CMS experiment possesses distributed computing infrastructure and its performance heavily depends on the fast and smooth distribution of data between different CMS sites. Data must be transferred from the Tier-0 (CERN) to the Tier-1 for storing and archiving, and time and good quality are vital to avoid overflowing CERN storage buffers. At the same time, processed data has to be distributed from Tier-1 sites to all Tier-2 sites for physics analysis while MonteCarlo simulations synchronized back to Tier-1 sites for further archival. At the core of all transferring machinery is PhEDEx (Physics Experiment Data Export) data transfer system. It is very important to ensure reliable operation of the system, and the operational tasks comprise monitoring and debugging all transfer issues. Based on transfer quality information Site Readiness tool is used to create plans for resources utilization in the future. We review the operational procedures created to enforce reliable

data delivery to CMS distributed sites all over the world. Additionally, we need to keep data consistent at all sites and both on disk and on tape. In this presentation, we describe the principles and actions taken to keep data consistent on sites storage systems and central CMS Data Replication Database (TMDB/DBS) while ensuring fast and reliable data samples delivery of hundreds of terabytes to the entire CMS physics community.

Poster Session / **342**

ATLAS Distributed Computing Shift Operation in the first 2 full years of LHC data taking

Author: Collaboration Atlas[1]

Co-authors: Alessandro Di Girolamo [2]; Daniel Colin Van Der Ster [2]; Guidone Negri [2]; Hiroshi Sakamoto [3]; I Ueda [3]; Jaroslava Schovancova [4]; Johannes Elmsheuser [5]; Mark Slater [6]; Nurcan Ozturk [7]; Stephane Jezequel [8]; Yuri Smirnov [9]

[1] *Atlas*

[2] *CERN*

[3] *University of Tokyo (JP)*

[4] *Acad. of Sciences of the Czech Rep. (CZ)*

[5] *Ludwig-Maximilians-Univ. Muenchen (DE)*

[6] *Birmingham University*

[7] *University of Texas at Arlington (US)*

[8] *Centre National de la Recherche Scientifique (FR)*

[9] *Brookhaven National Laboratory (US)*

Corresponding Author: jaroslava.schovancova@cern.ch

ATLAS Distributed Computing organized 3 teams to support data processing at Tier-0 facility at CERN, data reprocessing, data management operations, Monte Carlo simulation production, and physics analysis at the ATLAS computing centers located world-wide. In this talk we describe how these teams ensure that the ATLAS experiment data is delivered to the ATLAS physicists in a timely manner in the glamorous era of the LHC data taking. We describe experience with ways how to improve degraded service performance, we detail on the Distributed Analysis support over the exciting period of the computing model evolution.

Poster Session / **343**

ATLAS DQ2 Deletion Service

Author: Collaboration Atlas[1]

Co-authors: Artem Petrosyan [2]; Danila Oleynik [2]; Simone Campana [3]; Vincent Garonne [3]

[1] *Atlas*

[2] *Joint Inst. for Nuclear Research (RU)*

[3] *CERN*

Corresponding Author: danila.oleynik@cern.ch

The ATLAS Distributed Data Management project DQ2 is responsible for the replication, access and bookkeeping of ATLAS data across more than 100 distributed grid sites. It also enforces data management policies decided on by the collaboration and defined in the ATLAS computing model.

The DQ2 deletion service is one of the most important DDM services. This distributed service interacts with 3rd party grid middleware and the DQ2 catalogs to serve data deletion requests on the grid. Furthermore, it also takes care of retry strategies, check-pointing transactions, load management and fault tolerance.
In this paper special attention is paid to the technical details which are used to achieve the high performance of service (peaking at more than 4 millions files deleted per day), accomplished without overloading either site storage, catalogs or other DQ2 components.
Special attention is also paid to the deletion monitoring service that allows operators a detailed view of the working system.

Poster Session / **344**

Recent Improvements in the ATLAS PanDA Pilot

Author: Collaboration Atlas[1]

Co-authors: Alden Stradling [2]; Carlos Contreras [3]; Jose Caballero Bejar [4]; Kaushik De [2]; Maxim Potekhin [4]; Paul Nilsson [2]; Tadashi Maeno [4]; Tim Dos Santos [5]; Torre Wenaus [4]

[1] *Atlas*

[2] *University of Texas at Arlington (US)*

[3] *Departamento de Fisica-Univ. Tecnica Federico Santa Maria (UTFSM*

[4] *Brookhaven National Laboratory (US)*

[5] *Bergische Universitaet Wuppertal (DE)*

Corresponding Author: paul.nilsson@cern.ch

The Production and Distributed Analysis system
(PanDA) in the ATLAS experiment uses pilots to execute submitted jobs on the worker nodes.
The pilots are designed to deal with different runtime conditions and failure scenarios, and support many storage systems.
This talk will give a brief overview of the PanDA pilot system and will present major features and recent improvements including CERNVM File System integration, file transfers with Globus Online, the job retry mechanism,
advanced job monitoring including JEM technology, and validation of new pilot code using the HammerCloud stress--testing system.
PanDA is used for all ATLAS distributed production and is the primary system for distributed analysis. It is currently used at over 100 sites world--wide.
We analyze the performance of the pilot system in processing LHC data on the OSG, LCG and Nordugrid infrastructures used by ATLAS, and describe plans for its further evolution.

Distributed Processing and Analysis on Grids and Clouds / **345**

Multi-core job submission and grid resource scheduling for ATLAS AthenaMP

Author: Collaboration Atlas[1]

Co-authors: Andrew John Washbrook [2]; Armin Nairz [3]; David Crooks [4]; David Lesny [5]; Douglas Smith [6]; Horst Severini [7]; Paolo Calafiura [8]; Robert Duane Harrington Jr [9]; Sam Skipsey [10]; Stuart Purdie [11]; Vakhtang Tsulaia [8]

[1] *Atlas*

[2] *University of Edinburgh (GB)*

[3] *CERN*

[4] *University of Glasgow (GB)*

[5] *Univ. Illinois at Urbana-Champaign (US)*

[6] *SLAC National Accelerator Laboratory (US)*

[7] *University of Oklahoma (US)*

[8] *Lawrence Berkeley National Lab. (US)*

[9] *University of Edinburgh*

[10] *NeSC/Edinburgh University*

[11] *University of Glasgow-Unknown-Unknown*

Corresponding Author: andrew.washbrook@cern.ch

AthenaMP is the multi-core implementation of the ATLAS software framework and allows the efficient sharing of memory pages between multiple threads of execution. This has now been validated for production and delivers a significant reduction on overall memory footprint with negligible CPU overhead.
Before AthenaMP can be routinely run on the LHC Computing Grid, it must be determined how the computing resources available to ATLAS can best exploit the notable improvements delivered by switching to this multi-process model. In particular, there is a need to identify and assess the potential impact of scheduling issues where single core and multi-core job queues have access to the same underlying resources.
A study into the effectiveness and scalability of AthenaMP in a production environment will be presented. Submitting AthenaMP tasks to the Tier-0 and candidate Tier-2 sites will allow detailed measurement of worker node performance and also highlight the relative performance of local resource management systems (LRMS) in handling large volumes of multi-core jobs.
Best practices for configuring the main LRMS implementations currently used by Tier-2 sites will be identified in the context of multi-core job optimisation. There will also be a discussion on how existing Grid middleware and the ATLAS job submission pilot model could use scheduling information to increase the overall efficiency of multi-core job throughput.

Poster Session / 346

GoCxx: a tool to easily leverage C++ legacy code for multicore-friendly Go libraries and frameworks

Author: Sebastien Binet[1]

[1] *LAL/IN2P3*

Corresponding Author: sebastien.binet@cern.ch

Current HENP libraries and frameworks were written before multicore systems became widely deployed and used.
From this environment, a ‘single-thread’ processing model naturally emerged but the implicit assumptions it encouraged are greatly impairing our abilities to scale in a multicore/manycore world.

Writing scalable code in C++ for multicore architectures, while doable, is no panacea. Sure, C++11 will improve on the current situation (by standardizing on std::thread, introducing lambda functions and defining a memory model) but it will do so at the price of complicating further an already quite sophisticated language.
This level of sophistication has probably already strongly motivated analysis groups to migrate to CPython, hoping for its current limitations with respect to multicore scalability to be either lifted (Grand Interpreter Lock removal) or for the advent of a new Python VM better tailored for this kind of environment (PyPy, Jython,...)

Could HENP migrate to a language with none of the deficiencies of C++ (build time, deployment, low level tools for concurrency) and with the fast turn-around time, simplicity and ease of coding of Python ?

This paper will try to make the case for Go - a young open source language with built-in facilities to easily express and expose concurrency - being such a language.

We will first present a status update on go-gaudi, a framework written in Go loosely modeled after the C++ framework Gaudi used by two LHC experiments, and how its event loop was modified to expose more concurrency.
Then, benchmarks fed with data flows extracted from current C++ frameworks and with different toy-components (thread-safe/non-thread-safe, I/O bound, CPU bound, ...) will be discussed.

Finally, we will introduce GoCxx, a tool leveraging gcc-xml's output to automatize the tedious work of creating Go wrappers for foreign languages, a critical task for any language wishing to leverage legacy and field-tested code.
We will conclude with the first results of applying GoCxx to real C++ Gaudi components, effectively enabling go-gaudi with LHC know-how.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Summary:

We present GoCxx, a tool to automatize the wrapping of C++ libraries, and its impact on next-generation parallel event processing frameworks.

Poster Session / 347

A Study of ATLAS Grid Performance for Distributed Analysis

Author: Collaboration Atlas[1]

Co-authors: Sergey Panitkin [2]; Torre Wenaus [2]; Valeri Fayn [2]

[1] *Atlas*

[2] *Brookhaven National Laboratory (US)*

Corresponding Author: panitkin@bnl.gov

In the past two years the ATLAS Collaboration at the LHC has collected a large volume of data and published a number of ground breaking papers. The Grid-based ATLAS distributed computing infrastructure played a crucial role in enabling timely analysis of the data. We will present a study of the performance and usage of the ATLAS Grid as platform for physics analysis and discuss changes that analysis usage patterns underwent in 2011. This includes studies of timing properties of user jobs (wait time, run time, etc) and analysis of data format popularity evolution that significantly affected ATLAS data distribution policies. These studies are based on mining of data archived by the PanDA workload management system.

Poster Session / 348

DCS Data Viewer, a Application that Access ATLAS DCS Historical Data.

Author: Charilaos Tsarouchas[1]

Co-authors: Dirk Hoffmann [2]; Mirjam Lena Fehling [3]; Saverio D'Auria [4]; Shaun Roe [5]; Stefan Schlenker [5]; Stefan Winkelmann [3]

[1] *National Technical Univ. of Athens (GR)*

[2] *Universite d'Aix - Marseille II (FR)*

[3] *Albert-Ludwigs-Universitaet Freiburg (DE)*

[4] *University of Glasgow (GB)*

[5] *CERN*

Corresponding Author: charilaos.tsarouchas@cern.ch

The ATLAS experiment at CERN is one of the four Large Hadron Collider ex- periments. The Detector Control System (DCS) of ATLAS is responsible for the supervision of the detector equipment, the reading of operational parame- ters, the propagation of the alarms and the archiving of important operational data in a relational database. DCS Data Viewer (DDV) is an application that provides access to the ATLAS DCS historical data through a web interface. Its design is structured using a client-server architecture. The pythonic server connects to the DB and fetches the data by using optimized SQL requests. It communicates with the outside world, by accepting HTTP requests and it can be used stand alone. The client is an AJAX interactive web application devel- oped under the Google Web Toolkit (GWT) framework. Its web interface is user friendly, platform and browser independent. The selection of metadata is done via a column-tree view or with a powerful search engine. The final visualization of the data is done using java applets or java script applications as plugins. The default output is a value-over-time chart, but other types of outputs like tables, ascii or ROOT files are supported too. Excessive access or malicious use of the database is prevented by a dedicated protection mechanism, allowing the expo- sure of the tool to hundreds of inexperienced users. The current configuration of the client and of the outputs can be saved in an XML file. Protection against web security attacks is foreseen and authentication constrains have been taken into account, allowing the exposure of the tool to hundreds of users world wide. Due to its flexible interface and its generic and modular approach, DDV could be easily used for other experiment control systems.

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Summary:

A web application for the visualization of ATLAS data.

Poster Session / 349

Software installation and condition data distribution via CernVM FileSystem in ATLAS

Author: Collaboration Atlas[1]

Co-authors: Alessandro De Salvo [2]; Alexander Undrus [3]; Artem Harutyunyan [4]; Asoka De Silva [5]; Doug Benjamin [6]; Jakob Blomer [7]; Predrag Buncic [4]; Yushu Yao [8]

[1] *Atlas*

[2] *Universita e INFN, Roma I (IT)*

[3] *Brookhaven National Laboratory (US)*

[4] *CERN*

[5] *TRIUMF (CA)*

[6] *Duke University (US)*

[7] *Ludwig-Maximilians-Univ. Muenchen (DE)*

[8] *LBNL*

Corresponding Author: alessandro.de.salvo@cern.ch

The ATLAS Collaboration is managing one of the largest collections of software among the High Energy Physics Experiments. Traditionally this
software has been distributed via rpm or pacman packages, and has been installed in every site and user's machine, using more space than needed since the releases could not always share common binaries. As soon as the software has grown in size and number of releases this approach showed its limits, both in terms of manageability, used disk space and
performance. The adopted solution is based on the CernVM FileSystem, a fuse-based http, read-only filesystem which guarantees file
de-duplication, scalability and performance. Here we describe the ATLAS experience in setting up the CVMFS facility and putting it into production, for different type of use-cases, ranging from single users' machines up to large Data Centers, for both Software and Conditions Data. The performance of CernVMFS, both with software and condition data
access, will be shown, comparing with other filesystems currently in use by the Collaboration.

Software Engineering, Data Stores and Databases / 350

dCache: implementing a high-end NFSv4.1 service using a Java NIO framework

Author: Tigran Mkrtchyan[1]

[1] *DESY/dCache.ORG*

Corresponding Author: kofemann@gmail.com

dCache is a high performance scalable storage system widely used by HEP community. In addition to set of home grown protocols we also provide industry standard access mechanisms like WebDAV and NFSv4.1. This support places dCache as a direct competitor to commercial solutions. Nevertheless conforming to a protocol is not enough; our implementations must perform comparably or even better than commercial systems. To achieve this, dCache uses two high-end IO frameworks from well know application servers: GlassFish and JBoss.

This presentation describes how we implemented an rfc1831 and rfc2203 compliant ONC RPC (Sun RPC) service based on the Grizzly NIO framework, part of the GlassFish application server. This ONC RPC service is the key component of dCache's NFSv4.1 implementation, but is independent of dCache and available for other projects. We will also show some details of dCache NFS v4.1 implementations, describe some of the Java NIO techniques used and, finally, present details of our performance evaluation.

Software Engineering, Data Stores and Databases / 351

Handling of time-critical Conditions Data in the CMS experiment - Experience of the first year of data taking

Author: Giacomo Govi[1]

Co-authors: Andreas Pfeiffer [2]; Francesca Cavallari [3]; Salvatore Di Guida [2]; Vincenzo Innocente [2]

[1] *Fermi National Accelerator Lab. (US)*

[2] *CERN*

[3] *Universita e INFN, Roma I (IT)*

Corresponding Author: giacomo.govi@cern.ch

Data management for a wide category of non-event data plays a critical role in the operation of the CMS experiment. The processing chain (data taking-reconstruction-analysis) relies in the prompt availability of specific, time dependent data describing the state of the various detectors and their calibration parameters, which are treated separately from event data. The Condition Database system is the infrastructure established to handle these data and to make sure that they are available to both offline and online workflows. The Condition Data layout is designed such that the payload data (the Condition) is associated to an Interval Of Validity (IOV). The IOV allows accessing selectively the sets corresponding to specific intervals of time, run number or luminosity section. Both payloads and IOVs are stored in a cluster of relational database servers (Oracle) using an object-relational access approach. The strict requirements of security and isolation of the CMS online systems are imposing a redundant architecture to the database system. The master database is located in the experiment area within the online network, while a read-only replica is kept in sync via Oracle streaming in the CERN computing center and this is the one which is accessible by worldwide computing jobs. The synchronization of the condition data is performed with specific jobs deployed within the online networks, and with dedicated "drop-box"services. We will discuss the overall architecture of the system, the implementation choices and the experience gained in the first year of operation.

Poster Session / **352**

Monitoring of services with non-relational databases and map-reduce framework

Author: Marian Babik[1]

Co-authors: David Collados Polidura [1]; Fabio Souto Moure [2]

[1] *CERN*

[2] *University of Vigo (ES)*

Corresponding Author: marian.babik@cern.ch

Service Availability Monitoring (SAM) is a well-established monitoring framework that performs regular measurements of the core services and reports the corresponding availability and reliability of the Worldwide LHC Computing Grid (WLCG) infrastructure. One of the existing extensions of SAM is a Site Wide Area Testing (SWAT), which gathers monitoring information from the worker nodes via instrumented jobs. This generates quite a lot of monitoring data to digest, as there are several data points for every job and several million jobs are executed every day. The recent uptake of non-relational databases opens a new paradigm in the large-scale storage and distributed processing of systems with heavy read-write workloads. For SAM this brings new possibilities to improve its model from performing aggregation of measurements to storing raw data and subsequent re-processing. Both SAM and SWAT are currently tuned to run at top performance reaching some of the limits in storage and processing power of their existing Oracle relational database. We investigated the usability and performance of non-relational storage together with its distributed data processing capabilities. For this, several popular systems have been compared.

In this contribution we describe our investigation of the existing non-relational databases suited for monitoring systems covering Cassandra, HBase, OpenTSDB and MongoDB. Further, we present our experiences in data modeling and prototyping map-reduce algorithms focusing on the extension of the already existing availability and reliability computations. Finally, possible future directions in this area are discussed, analyzing the current deficiencies of the existing Grid monitoring systems and proposing solutions how to leverage the benefits of the non-relational databases to get more scalable and flexible frameworks.

Event Processing / 353

Track finding and fitting on GPUs, first steps toward a software trigger

Author: Mohammad Al-Turany[1]

[1] *GSI*

Corresponding Author: mohammad.al-turany@cern.ch

The high data rates expected from the planned detectors at FAIR (CBM, PANDA) call for dedicated attention with respect to the computing power needed in online (e.g. High level event selection) and offline analysis. The graphics processor units (GPUs) have evolved into high performance co-processors that can be easily programmed with common high-level language such as C, Fortran and C++. Todays GPUs greatly
outpace CPUs in arithmetic performance and memory bandwidth, making them the ideal co-processor to accelerate a variety of data parallel applications. For the online processing (i.e: Software triggers and online event selections) GPUs are an attractive solution, online applications include high level processing which require floating point operations. However, the widely used FPGA (Field Programmable Gate Array) does not have such capabilities. The users have to program in the low-level hardware description language (HDL). GPUs, on the contrary, are programmable with high-level languages and meanwhile provide support even for double precision. An algorithm based upon conformal mapping and Hough transform was implemented on GPUs. The algorithm is tested with the PANDA central tracker simulated data. The results of the same algorithm are compared with CPU and FPGA implementations.

Event Processing / 354

The art framework

Author: Christopher Green[1]

Co-authors: Jim Kowalkowski [2]; Marc Paterno [3]; Walter E Brown [3]

[1] *Department of Physics*

[2] *Fermi National Accelerator Laboratory (FNAL)*

[3] *Fermilab*

Corresponding Author: paterno@fnal.gov

Future "Intensity Frontier" experiments at Fermilab are likely to be conducted by smaller collaborations, with fewer scientists, than is the case for recent "Energy Frontier" experiments. *art* is an event-processing framework designed with the needs of such experiments
in mind.

The authors have been involved with the design and implementation of frameworks for several experiments, including D0, BTeV, and CMS. Although many of these experiments' requirements were the same, they shared little effort, and even less code. This resulted in significant duplication of development effort. The *art* framework project is intended to avoid such duplication of effort for the experiments planned, and under consideration, at Fermilab.

The *art* framework began as an evolution of the framework of the CMS experiment, and has since been heavily adapted for the needs of the intensity frontier experiments. Trade-offs have been made to simplify the
code in order for it to be maintainable and usable by much smaller groups. The current users of *art* include mu2e, NOvA, g-2, and LArSoft (ArgoNeuT,
MicroBooNE, LBNE-LAr).

The *art* framework relies upon a number of external products (e.g., the Boost C++ library and Root); these products are built by the *art* team and deployed through a simplified UPS package deployment system. The *art*
framework is itself deployed via the same mechanism, and is treated by the experiments using it as just another external product upon which their code relies.

Because of the increasing importance of multi-core and many-core architectures, current development plans center around the migration of *art* to support parallel processing of independent events as well as to permit parallel processing within events.

Summary:

We describe the art framework, current used by several planned and proposed Intensity Frontier experiments at Fermilab, and plans for its future development.

Poster Session / **356**

Certified Grid Job Submission in the ALICE Grid Services

Author: Steffen Schreiner[1]

Co-authors: Costin Grigoras [2]; Latchezar Betev [2]; Maarten Litmaath [2]

[1] *CERN, CASED/TU Darmstadt*

[2] *CERN*

Corresponding Author: steffen.schreiner@cern.ch

Grid computing infrastructures need to provide traceability and accounting of their users'activity and protection against misuse and privilege escalation, where the delegation of privileges in the course of a job submission is a key concern. This work describes an improved handling of multi-user Grid jobs in the ALICE Grid Services.
A security analysis of the ALICE Grid job model is presented with derived security objectives, followed by a discussion of existing approaches of unrestricted delegation based on X.509 proxy certificates and the Grid middleware gLExec. Unrestricted delegation has severe security consequences and limitations, most importantly allowing for identity theft and forgery of jobs and data. These limitations are discussed and formulated, both in general and with respect to an adoption in line with multi-user Grid jobs. A new general model of mediated definite delegation is developed and formulated, allowing a broker to assign context-sensitive user privileges to agents while providing strong accountability and long-term traceability. A prototype implementation allowing for certified Grid jobs is presented including a potential interaction with gLExec. The achieved improvements regarding system security, malicious job exploitation, identity protection, and accountability are emphasized, followed by a discussion of non-repudiation in the face of malicious Grid jobs.

Summary:

This contribution will demonstrate an in-depth security analysis and discussion of proxy certificate based authentication and authorization of multi-user pilot jobs and present a new model of delegation as a proposed solution. The model's implementation in a prototype and its performance testing will be shown as a proof of concept.

Poster Session / **358**

Ksplice: Update without rebooting

Author: Waseem Daher[1]

[1] *Oracle*

Corresponding Author: waseem.daher@oracle.com

Today, every OS in the world requires regular reboots in order to be up to date and secure. Since reboots cause downtime and disruption, sysadmins are forced to choose between security and convenience.

Until Ksplice. Ksplice is new technology that can patch a kernel while the system is running, with no disruption whatsoever. We use this technology to provide Ksplice Uptrack, a service that delivers important security and bugfix updates to your systems. (It's free for Ubuntu Desktop and Fedora, and is also a free feature of Oracle Linux Premier support.)

In this talk, we'll very briefly provide an overview of how Ksplice can dramatically reduce the pain associated with large-scale installation maintenance, which is why it is used at 700+ customer sites, on 100,000+ customer systems, including at Brookhaven National Lab.

More importantly, we'll provide a detailed look into how the Ksplice technology works and how the Ksplice Uptrack service works, at a technical level primarily targeted at system administrators and developers, but largely accessible to the average Linux user as well.

Summary:

Today, every OS in the world requires regular reboots in order to be up to date and secure. Since reboots cause downtime and disruption, sysadmins are forced to choose between security and convenience.

Until Ksplice. Ksplice is new technology that can patch a kernel while the system is running, with no disruption whatsoever. We use this technology to provide Ksplice Uptrack, a service that delivers important security and bugfix updates to your systems. (It's free for Ubuntu Desktop and Fedora, and is also a free feature of Oracle Linux Premier support.)

In this talk, we'll very briefly provide an overview of how Ksplice can dramatically reduce the pain associated with large-scale installation maintenance, which is why it is used at 700+ customer sites, on 100,000+ customer systems, including at Brookhaven National Lab.

More importantly, we'll provide a detailed look into how the Ksplice technology works and how the Ksplice Uptrack service works, at a technical level primarily targeted at system administrators and developers, but largely accessible to the average Linux user as well.

Poster Session / **359**

Development of noSQL data storage for the ATLAS PanDA Monitoring System

Author: Collaboration Atlas[1]

Co-authors: Hironori Ito [2]; Maxim Potekhin [2]; Torre Wenaus [2]

[1] *Atlas*

[2] *Brookhaven National Laboratory (US)*

Corresponding Author: maxim.potekhin@cern.ch

For several years the PanDA Workload Management System has been the basis for distributed production and analysis for the ATLAS experiment at the LHC. Since the start of data taking PanDA usage has ramped up steadily, typically exceeding 500k completed jobs/day by June 2011. The associated monitoring data volume has been rising as well, to levels that present a new set of challenges in the areas of database scalability and monitoring system performance
and efficiency. These challenges have being met with a R&D and development effort aimed at

implementing a scalable and efficient monitoring data storage based on a noSQL solution (Cassandra).
We present the data design and indexing strategies for efficient queries, as well
as our experience of operating a Cassandra cluster and interfacing it with a Web service.

Poster Session / **360**

Software Validation in ATLAS

Author: Collaboration Atlas[1]

Co-authors: Brinick Simmons [2]; David Rousseau [3]; Mark Hodgkinson [4]; Peter Sherwood [5]; Rolf Seuster [6]

[1] *Atlas*

[2] *Department of Physics and Astronomy - University College London*

[3] *Laboratoire de l'Accelerateur Lineaire (LAL)-Universite de Paris*

[4] *University of Sheffield*

[5] *University College London (UK)*

[6] *Max-Planck-Institut fuer Physik (Werner-Heisenberg-Institut) (D*

Corresponding Authors: mark.hodgkinson@cern.ch, rolf.seuster@cern.ch

The ATLAS collaboration operates an extensive set of protocols to validate the quality of the offline software in a timely manner. This is essential in order to process the large amounts of data being collected by the ATLAS detector in 2011 without complications on the offline software side. We will discuss a number of different strategies used to validate the ATLAS offline software; running the Athena software in a variety of configurations daily on each nightly build via the ATN and RTT systems; the monitoring of these tests and checking the compilation of the software via distributed teams of rotating shifters; monitoring of and follow up on bug reports by the shifter teams and periodic software cleaning weeks to improve the quality of the offline software further.

Poster Session / **361**

Rebootless Linux Kernel Patching with Ksplice Uptrack at BNL

Author: Christopher Hollowell[1]

Co-authors: James Pryor [1]; Jason Alexander Smith [2]

[1] *Brookhaven National Laboratory*

[2] *Brookhaven National Laboratory (US)*

Corresponding Author: hollowec@bnl.gov

Ksplice/Oracle Uptrack is a software tool and update subscription service which allows system administrators to apply security and bug fix patches to the Linux kernel running on servers/workstations without rebooting them. The RHIC/ATLAS Computing Facility at Brookhaven National Laboratory (BNL) has deployed Uptrack on nearly 2000 hosts running Scientific Linux and Red Hat Enterprise Linux. The use of this software has minimized downtime, and increased our security posture. In this presentation, we provide an overview of Ksplice's rebootless kernel patch creation/insertion mechanism, and our experiences with Uptrack.

Poster Session / **362**

WHALE, a management tool for Tier-2 LCG sites

Author: Ivano Giuseppe Talamo[1]

Co-authors: Giovanni Organtini [1]; Luciano Barone [1]

[1] *Universita e INFN, Roma I (IT)*

Corresponding Author: ivano.giuseppe.talamo@cern.ch

The LCG (Worldwide LHC Computing Grid) is a grid-based hyerarchical computing distributed facility, composed of more than 140 computing centers, organized in 4 tiers, by size and offer of services. Every site, although indipendent for many technical choices, has to provide services with a well-defined set of interfaces. For this reason, different LCG sites need frequently to manage very similar situations, like jobs behaviour on the batch system, dataset transfers between sites, operating system and experiment software installation and configuration, monitoring of services.

In this context we created WHALE (WHALE Handles Administration in an LCG Environment), a software actually used at the T2_IT_Rome site, an LCG Tier-2 for the CMS experiment.

WHALE is a generic, site indipendent tool written in python: it allows administrator to interact in a uniform and coherent way with several subsystems using a high level syntax which hides specific commands.

The architecture of WHALE is based on the plugin concept and on the possibility of connecting the output of a plugin to the input of the next one, in a pipe-like system, giving the administrator the possibility of making complex functions by combining the simpler ones. The core of WHALE just handles the plugin orchestrations, while even the basic functions (eg. the WHALE activity logging) are performed by plugins, giving the capability to tune and possibly modify every component of the system. WHALE already provides many plugins useful for a LCG site and some more for a Tier-2 of the CMS experiment, expecially in the field of job management, dataset transfer and analysis of performance results and availability tests (eg. Nagios tests, SAM tests). Thanks to its architecture and the provided plugins WHALE makes easy to perform tasks that, even if logically simple, are technically complex or tedious, like eg. closing all the worker nodes with a job-failure rate greater than a given threshold. Finally, thanks to the centralization of the activities on a single point and to its logging functionalities, WHALES acts as a knowledge-base of the site and a handful tool to keep track of the activities at a given site. For this reason it also provides a tailored plugin to perform advanced searches in the activity log.

Poster Session / **363**

Evolution of the ATLAS Nightly Build System

Author: Collaboration Atlas[1]

Co-author: Alexander Undrus [2]

[1] *Atlas*

[2] *Brookhaven National Laboratory (US)*

Corresponding Author: undrus@bnl.gov

The ATLAS Nightly Build System is a major component in the ATLAS collaborative software organization, validation, and code approval scheme. For over 10 years of development it has evolved into a factory for automatic release production and grid distribution. The 50 multi-platform branches of ATLAS releases provide vast opportunities for testing new packages, verification of patches to existing software, and migration to new platforms and compilers for ATLAS code that currently contains 2000 packages with 4 million C++ and 1.4 million python scripting lines written by 1000

developers. Recent development was focused on the integration of ATLAS Nightly Builds and Installation systems. The nightly releases are distributed and validated and some are transformed into stable releases used for data processing worldwide. The ATLAS Nightly System is managed by the NICOS control tool on a computing farm with 50 powerful multiprocessor nodes. NICOS provides the fully automated framework for the release builds, testing, and creation of distribution kits. The ATN testing framework of the Nightly System runs unit and integration tests in parallel suites, fully utilizing the resources of multi-processor machines, and provides the first results even before compilations complete. The NICOS error detection system is based on several techniques and classifies the compilation and test errors according to their severity. It is periodically tuned to place greater emphasis on certain software defects by highlighting the problems on NICOS web pages and sending automatic e-mail notifications to responsible developers. These and other recent developments will be presented and future plans will be described.

Poster Session / **365**

The Monitoring and Calibration Web Systems for the ATLAS Tile Calorimeter Data Quality Analysis

Author: Andressa Sivolella Gomes[1]

Co-authors: Carmen Maidantchik [1]; Collaboration Atlas [2]; Fernando Guimaraes Ferreira [1]

[1] *Univ. Federal do Rio de Janeiro (BR)*

[2] *Atlas*

Corresponding Authors: andressa@cern.ch, fernando.guimaraes.ferreira@cern.ch

The Tile Calorimeter (TileCal), one of the ATLAS detectors. has four partitions, where each one contains 64 modules and each module has up to 48 PhotoMulTipliers (PMTs), totalizing more than 10,000 electronic channels. The Monitoring and Calibration Web System (MCWS) supports data quality analyses at channels level. This application was developed to assess the detector status and verify its performance, presenting the problematic known channels list from the official database that stores the detector conditions data (COOL). The bad channels list guides the data quality validator during analyses in order to identify new problematic channels. Through the system, it is also possible to update the channels list directly in the COOL database. MCWS generates results, as etaphi plots and comparative tables with masked channels percentage, which concerns TileCal status, and it is accessible by all ATLAS collaboration. Annually, there is an intervention on LHC (Large Hadronic Collider) when the detector equipments (PMTs, motherboards, voltages and cables, for example) are fixed or replaced by new ones. When a channel needs to be repaired, the calibration constants stored into COOL database must be updated, otherwise they may negatively interfere in the data quality analyses. A MCWS functionality manages the calibration constants by updating their values in COOL database. The development team foresees an integration with the Tile detector control Web system (DCS) in order to automatically identify voltage problems, since the channels are fed by high voltage sources. The MCWS has been used by the Tile community since 2008, during the commissioning phase, and was upgraded to respect the ATLAS operation specifications.

Poster Session / **366**

File and Dataset Metadata Collection and Use in Atlas

Author: Collaboration Atlas[1]

Co-authors: Elizabeth Gallas [2]; Fabian Lambert [3]; Jerome Fulachier [3]; Solveig Albrand [3]

[1] *Atlas*

[2] *University of Oxford (GB)*

[3] *Universite Joseph Fourier (FR)*

Corresponding Author: elizabeth.gallas@physics.ox.ac.uk

The ATLAS Metadata Interface ("AMI") was designed as a generic cataloguing system, and as such it has found many uses in the experiment including software release management, tracking of reconstructed event sizes and control of dataset nomenclature. In this paper we will discuss the primary use of AMI which is to provide a catalogue of datasets (file collections) which is searchable using physics criteria.
The AMI dataset catalogues are filled from several sources:
- The Tier 0 database for raw data and first pass reconstruction.
- The Production System database for Monte Carlo and reprocessed data.
- The Distributed Data Management system.
- Direct input from the physicist community.
We will summarize the information taken from each source, and discuss the different mechanisms used to obtain it.
By correlating information from different sources we can derive aggregate information which is important for physics analysis; for example the total number of events contained in dataset, and possible reasons for missing events such as a lost file.
Finally we will describe some specialized interfaces which were developed for the Data Preparation and reprocessing coordinators. These interfaces manipulate information from both the dataset domain held in AMI, and the run-indexed information held in the ATLAS COMA application (Conditions and Configuration Metadata).

Poster Session / **367**

The Geant4 Virtual Monte Carlo

Author: Ivana Hrivnacova[1]

[1] *IPN Orsay, CNRS/IN2P3*

Corresponding Author: ivana@ipno.in2p3.fr

The Virtual Monte Carlo (VMC) provides the abstract interface into the Monte Carlo transport codes: GEANT3, Geant4 and FLUKA. The user VMC based application, independent from the specific Monte Carlo codes, can be then run with all three simulation programs. The VMC has been developed by the ALICE Offline Project and since then it draw attention in more experimental frameworks.

Since its first release in 2002, the implementation of the VMC for Geant4 (Geant4 VMC) is in continuous maintenance and development, driven by the evolution of Geant4 on one side and the requirements from users on the other side. In this presentation we will give an overview and the present status of this interface and report on new features. The performance evaluation and the time comparisons for equivalent Geant4 native and VMC test applications will be presented. The Geant4 multi-threading prototype (Geant4 MT) represents a new challenge. The study of a feasibility of the migration of Geant4 VMC to Geant4 MT will be also presented.

Poster Session / **368**

IPv6 testing and deployment at Prague Tier 2

Author: Tomas Kouba[1]

Co-authors: Jan Kundrat [2]; Jan Svec [1]; Jiri Chudoba [1]; Jiri Horky [1]; Lukas Fiala [1]; Marek Elias [3]

[1] *Acad. of Sciences of the Czech Rep. (CZ)*

[2] *Unknown-Unknown-Unknown*

[3] *FZU ASCR*

Corresponding Author: koubat@fzu.cz

Computing Centre of the Institute of Physics in Prague provides computing and storage resources for various HEP experiments (D0, Atlas, Alice, Auger) and currently operates more than 300 worker nodes with more than 2500 cores and provides more than 2PB of disk space. Our site is limited to one C-sized block of IPv4 addresses, and hence we had to move most of our worker nodes behind the NAT. However this solution demands more difficult routing setup. We see the IPv6 deployment as a solution that provides less routing, more switching and therefore promises higher network throughput.

The administrators of the Computing Centre strive to configure and install all provided services automatically. For installation tasks we use PXE and kickstart, for network configuration we use DHCP and for software configuration we use CFengine. Many hardware boxes are configured via specific web pages or telnet/ssh protocol provided by the box itself. All our services are monitored with several tools e.g. Nagios, Munin, Ganglia. We rely heavily on the SNMP protocol for hardware health monitoring.

All these installation, configuration and monitoring tools must be tested before we can switch completely to IPv6 network stack. In this contribution we present the tests we have made, limitations we have faced and configuration decisions that we have made during IPv6 testing. We also present testbed built on virtual machines that was used for all the testing and evaluation.

Software Engineering, Data Stores and Databases / 369

A Programmatic View of Metadata, Metadata Services, and Metadata Flow in ATLAS

Author: Collaboration Atlas[1]

Co-authors: David Malon [2]; Elizabeth Gallas [3]; Graeme Andrew Stewart [4]; Solveig Albrand [5]

[1] *Atlas*

[2] *Argonne National Laboratory (US)*

[3] *University of Oxford (GB)*

[4] *CERN*

[5] *Universite Joseph Fourier (FR)*

Corresponding Author: malon@anl.gov

The volume and diversity of metadata in an experiment of the size and scope of ATLAS is considerable. Even the definition of metadata may seem context-dependent: data that are primary for one purpose may be metadata for another. Trigger information and data from the Large Hadron Collider itself provide cases in point, but examples abound.
Metadata about logical or physics constructs, such as data-taking periods and runs and luminosity blocks and events and algorithms,
often need to be mapped to deployment and production constructs, such as datasets and jobs and files and software versions, and vice versa.
Metadata at one level of granularity may have implications at another.
ATLAS metadata services must integrate and federate information from inhomogeneous sources and repositories, map metadata about logical or physics constructs to deployment and production constructs, provide a means to associate metadata at one level of granularity with processing or decision-making at another, offer a coherent and integrated view to physicists, and support both human use and programmatic access.
In this paper we consider ATLAS metadata, metadata services, and metadata flow principally from the illustrative perspective
of how disparate metadata are made available to executing jobs and, conversely, how metadata generated by such jobs are returned.
We describe how metadata are read, how metadata are cached, and how metadata generated by

jobs and the tasks of which they are a part are communicated, associated with data products, and preserved. We also discuss the principles that guide decision-making about metadata storage, replication, and access.

Poster Session / 370

An Extensible Infrastructure for Querying and Mining Event-level Metadata in ATLAS

Author: Collaboration Atlas[1]

Co-authors: David Malon [2]; Elisabeth Vinek [3]; Jack Cranshaw [2]; Qizhi Zhang [2]

[1] *Atlas*

[2] *Argonne National Laboratory (US)*

[3] *University of Vienna (AT)*

Corresponding Author: cranshaw@anl.gov

The ATLAS event-level metadata infrastructure supports applications that range from data quality monitoring, anomaly detection, and fast physics monitoring to event-level selection and navigation to file-resident event data at any processing stage, from raw through analysis object data, in globally distributed analysis. A central component of the infrastructure is a distributed TAG database, which contains event-level metadata records for all ATLAS events, real and simulated.
This resource offers a unique global view of ATLAS data, and provides an opportunity, not only for stream-style mining of event data,
but also for an examination of data across streams, across runs, and across (re)processings.
The TAG database serves as a natural locus for run-level and processing-level integrity checks, for investigations of event duplication and other issues in the trigger and offline systems, for questions about stream overlap, for queries about interesting but out-of-stream events, for statistics, and more.
In early ATLAS running, such database queries were largely ad hoc, and were handled manually.
In this paper, we describe an extensible infrastructure for addressing these and other use cases during upload and post-upload processing, and discuss some of the uses to which this infrastructure has been applied.

Software Engineering, Data Stores and Databases / 371

RooStats: Statistical Tools for the LHC

Authors: Lorenzo Moneta[1]; Sven Kreiss[2]

Co-authors: Alfio Lazzaro [3]; Gennadiy Kukartsev [4]; Giovanni Petrucciani [5]; Gregory Alfred Schott [6]; Kyle Stuart Cranmer [2]; Wouter Verkerke [7]

[1] *CERN*

[2] *New York University (US)*

[3] *Universita degli Studi di Milano-Universita e INFN*

[4] *Brown University (US)*

[5] *Univ. of California San Diego (US)*

[6] *KIT - Karlsruhe Institute of Technology (DE)*

[7] *NIKHEF (NL)*

Corresponding Authors: sven.kreiss@cern.ch, lorenzo.moneta@cern.ch

RooStats is a project providing advanced statistical tools required for the analysis of
LHC data, with emphasis on discoveries, confidence intervals, and combined measurements in the

both the Bayesian and Frequentist approaches. The tools are built on top of the RooFit data modeling language and core ROOT mathematics libraries and persistence technology.
These tools have been developed in collaboration with the LHC experiments and used by them to produce numerous physics results, such as the combination of ATLAS and CMS Higgs searches that resulted in a model with more than 200 parameters. We will review new developments which have been included in RooStats and the performance optimizations, required to cope with such complex models used by the LHC experiments. We will show as well the parallelization capability of these statistical tools using multiple-processors via PROOF.

Poster Session / **372**

TAG Base Skimming In ATLAS

Author: Collaboration Atlas[1]

Co-authors: Donnchadha Quilty [2]; Jack Cranshaw [3]; Julius Hrivnac [4]; Marcin Nowak [5]; Mark Slater [6]; Qizhi Zhang [3]; Thomas Doherty [7]

[1] *Atlas*

[2] *University College Dublin School of Physics (UCD)*

[3] *Argonne National Laboratory (US)*

[4] *Universite de Paris-Sud 11 (FR)*

[5] *Brookhaven National Laboratory (US)*

[6] *Birmingham University*

[7] *Department of Physics and Astronomy-University of Glasgow*

Corresponding Author: cranshaw@anl.gov

TAGs are event-level metadata allowing a quick search for interesting events for further analysis, based on selection criteria defined by the user. They are stored in a file-based format as well as in relational databases. The overall TAG system architecture encompasses a range of interconnected services that provide functionality for the required use cases such as event level selection, display, extraction and skimming. Skimming can be used to produce any of the pre-TAG data products by pure copy or any post-TAG data products if these can be made from a pre-TAG data product. The implemented use cases for the skimming service scale from a physicist wishing to select a handful of interesting events for an analysis specific study to the creation of physics working group samples on the ATLAS production system.
This paper will focus on the workflow aspects involved in creating pre and post TAG data products from a TAG selection using the Grid in the context of the overall TAG system architecture. The emphasis will be on the range of demands that the implemented use cases place on these workflows and on the infrastructure. The tradeoffs of various workflow strategies will be discussed including scalability issues and other concerns that occur when integrating with data management and production systems.

Poster Session / **373**

Conditions and Configuration Metadata for the ATLAS experiment

Author: Elizabeth Gallas[1]

Co-authors: Collaboration Atlas [2]; Fabian Lambert [3]; Jeffrey Tseng [1]; Jerome Fulachier [3]; Katherine Pachal [1]; Qizhi Zhang [4]; Solveig Albrand [3]

[1] *University of Oxford (GB)*

[2] *Atlas*

[3] *Centre National de la Recherche Scientifique (FR)*

[4] *High Energy Physics Division*

Corresponding Authors: elizabeth.gallas@physics.ox.ac.uk, solveig.albrand@lpsc.in2p3.fr, jerome.fulachier@lpsc.in2p3.fr, fabian.lambert@lpsc.in2p3.fr, katherine.pachal@cern.ch, jeffrey.tseng@cern.ch, qzhang@anl.gov

In the ATLAS experiment, database systems generally store the bulk of conditions and configuration data needed by event-wise reconstruction and analysis jobs. These systems can be relatively large stores of information, organized and indexed primarily to store all information required for system-specific use cases and efficiently deliver
the required information to event-based jobs.
Metadata in these systems may include the indexes themselves, but frequently important metadata for forming, for example, collections of events for analysis or for the management of that system may not be readily accessible
for more global purposes.
Moreover, the systems may have been developed before important metadata quantities were recognized.
A system, called COMA (Conditions/Configuration Metadata for ATLAS),
has been developed to make globally important metadata more readily accessible.
It is based on a relational database storing directly extracted, refined, reduced, and derived information from these system-specific data sources as well as information from non-database sources. A variety of unique interfaces have emerged and additional interfaces are in development.
This presentation will give an overview of the components of the system and describe the unique interfaces which it facilitates.
We summarize the challenges in defining and loading the requisite data and specify how consistency is maintained between COMA and the primary data sources.

Poster Session / **374**

Monitoring of computing resource utilization of the ATLAS experiment

Author: Collaboration Atlas[1]

Co-authors: David Rousseau [2]; Gancho Dimitrov [3]; Ilija Vukotic [4]; Osman Aidel [5]; Solveig Albrand [6]

[1] *Atlas*

[2] *Laboratoire de l'Accelerateur Lineaire (LAL)-Universite de Paris*

[3] *Brookhaven National Laboratory (US)*

[4] *Universite de Paris-Sud 11 (FR)*

[5] *Unknown*

[6] *Universite Joseph Fourier (FR)*

Corresponding Author: ilija.vukotic@cern.ch

Due to the good performance of the LHC accelerator, the ATLAS experiment has seen higher than anticipated levels for both the event rate and the average number of interactions per bunch crossing. In order to respond to these changing requirements, the current and future usage of CPU, memory and disk resources has to be monitored, understood and acted upon. This requires data collection at a fairly fine level of granularity: the performance of each object written and each algorithm run, as well as a dozen per-job variables, are gathered for the different processing steps of Monte Carlo generation and simulation and the reconstruction of both data and Monte Carlo. We present a system to collect and visualize the data from both the online Tier-0 system and distributed grid production jobs. Around 40 GB of performance data are expected from up to 200k jobs per day, thus making performance optimization of the underlying Oracle database of utmost importance.

Poster Session / **375**

Applicability of modern, scale-out file services in dedicated LHC data analysis environments.

Author: Martin Gasthuber[1]

Co-author: Yves Kemp [2]

[1] *Deutsches Elektronen-Synchrotron (DE)*

[2] *DESY/IT*

Corresponding Author: martin.gasthuber@cern.ch

DESY has started to deploy modern, state of the art, industry based, scale out file services together with certain extension as a key component in dedicated LHC analysis environments like the National Analysis Facility (NAF) @DESY. In a technical cooperation with IBM, we will add identified critical features to the standard SONAS product line of IBM to make the system best suited for the already high and increasing demands of the NAF@DESY. Initially we will give a short introduction of the core system and their basic mode of operations - followed by a detailed description of the identified additional components/services addressed within the DESY/IBM cooperation and largely worked out by talking to the physicists doing analysis on the NAF today. Already known areas are for example: interface to tertiary storage (archive), system federation through industry standard protocols, X509 integration and far more aggressive caching of physics data (immutable data). Finally we will show in detail the first results of the newly implemented features including lectures learned regarding the basic suitability in our community.

Poster Session / **376**

New features in the ROOT mathematical and statistical libraries

Author: Lorenzo Moneta[1]

Co-author: Christian Gumpert [2]

[1] *CERN*

[2] *Technische Universitaet Dresden (DE)*

Corresponding Author: lorenzo.moneta@cern.ch

ROOT, a data analysis framework, provides advanced numerical and statistical methods via the ROOT Math work package.
Now that the LHC experiments have started to analyze their data and produce physics results, we have acquired experience in the way these numerical methods are used and the libraries have been consolidated taking into account also the received feedback. At the same time, new features have been introduced as required by the experiments. One of these new features is a better support for dealing with multi-dimensional data structure. A new class based on a binary kd-tree has been introduced for dealing with multi-dimensional data. We will show examples on how this class can be used for efficient binning of multidimensional data and for constructing non-parametric density estimation, which can be used for fitting or data classification.
We will show as well the improvements added in the mathematical libraries for analyzing and fitting weighted data sets. In particular we will show examples in fitting Poisson (histograms) and binomial data.
We will present as well some of the improvements in the core numerical algorithms and the optimization and performance studies which have been performed.

Event Processing / **377**

I/O Strategies for Multicore Processing in ATLAS

Author: Collaboration Atlas[1]

Co-authors: David Malon [2]; Paolo Calafiura [3]; Peter Van Gemmeren [2]; Sebastien Binet [4]; Vakhtang Tsulaia [3]; Wim Lavrijsen [3]

[1] *Atlas*

[2] *Argonne National Laboratory (US)*

[3] *Lawrence Berkeley National Lab. (US)*

[4] *Universite de Paris-Sud 11 (FR)*

Corresponding Author: peter.van.gemmeren@cern.ch

A critical component of any multicore/manycore application architecture is the handling of input and output.
Even in the simplest of models, design decisions interact both in obvious and in subtle ways with persistence strategies.
When multiple workers handle I/O independently using distinct instances of a serial I/O framework, for example, it may happen that because of the way data from consecutive events are compressed together, there may be serious inefficiencies, with workers redundantly reading the same buffers, or multiple instances thereof. With shared reader strategies, caching and buffer management by the persistence infrastructure and by the control framework may have decisive performance implications for a variety of design choices. Providing the next event may seem straightforward when all event data are contiguously stored in a block, but there
may be performance penalties to such strategies when only a subset of a given event's data are needed; conversely, when event data are partitioned by type in persistent storage, providing the next event becomes more complicated, requiring marshaling of data from many I/O buffers.
Output strategies pose similarly subtle problems, with complications that may lead to significant serialization and the possibility of serial bottlenecks, either during writing or in post-processing, e.g., during data stream merging.
In this paper we describe the I/O components of AthenaMP, the multicore implementation of the ATLAS control framework, and the considerations that have led to the current design, with attention to how these I/O components interact with ATLAS persistent data organization and infrastructure.

Event Processing / 378

The ATLAS ROOT-based data formats: recent improvements and performance measurements

Author: Collaboration Atlas[1]

Co-authors: David Malon [2]; Ilija Vukotic [3]; Jack Cranshaw [2]; Peter Van Gemmeren [2]; R D Schaffer [3]; Wahid Bhimji [4]

[1] *Atlas*

[2] *Argonne National Laboratory (US)*

[3] *Universite de Paris-Sud 11 (FR)*

[4] *University of Edinburgh (GB)*

Corresponding Author: wahid.bhimji@cern.ch

We detail recent changes to ROOT-based I/O within the ATLAS experiment. The ATLAS persistent event data model continues to make considerable use of a ROOT I/O backend through POOL persistency. Also ROOT is used directly in later stages of analysis that make use of a flat-ntuple based "D3PD" data-type. For POOL/ROOT persistent data, several improvements have been made including implementation of automatic basket optimisation, memberwise streaming, and changes to split and compression levels. Optimisations are also planned for the D3PD format. We present

a full evaluation of the resulting performance improvements from these, including in the case of selected retrieval of events. We also evaluate ongoing changes internal to ROOT, in the ATLAS context, for both POOL and D3PD data. We report results not only from test systems, but also utilising new automated tests on real ATLAS production resources which employ a wide range of storage technologies.

Poster Session / 379

A browser-based event display for the CMS experiment at the LHC

Author: Thomas Mc Cauley[1]

Co-authors: Mihael Hategan [2]; Phong Nguyen [3]

[1] *Fermi National Accelerator Lab. (US)*

[2] *University of Chicago*

[3] *Fermilab*

Corresponding Author: thomas.mccauley@cern.ch

The line between native and web applications is becoming increasingly blurred as modern web browsers are becoming powerful platforms on which applications can be run. Such applications are trivial to install and are readily extensible and easy to use. In an educational setting, web applications permit a way to rapidly deploy tools in a highly-restrictive computing environment.

The I2U2 collaboration has developed a browser-based event display for viewing events in data collected and released to the public by the CMS experiment at the LHC. The application itself reads a JSON event format and uses the JavaScript 3D rendering engine pre3d. The only requirement is a modern browser using HTML5 canvas. The event display has been used by thousands of high school students in the context of programs organized by I2U2, Quarknet, and IPPOG. This browser-based approach to display of events can have broader usage and impact for experts and public alike.

Computer Facilities, Production Grids and Networking / 380

Evaluation of 40 Gigabit Ethernet technology for data servers

Author: Artur Jerzy Barczyk[1]

Co-authors: Azher Mughal [2]; Harvey Newman [1]; Iosif Legrand [1]; Ramiro Voicu [1]; sandor Rozsa [3]

[1] *California Institute of Technology (US)*

[2] *California Institute of Technology*

[3] *California Institute of Technology (CALTECH)*

Corresponding Authors: artur.barczyk@cern.ch, azher.mughal@cern.ch, sandor.gyula.rozsa@cern.ch

40Gb/s network technology is increasingly available today in the data centers as well as in the network backbones. We have built and evaluated storage systems equipped with the last generation of 40GbE Network Interface Cards. The recently available motherboards with the PCIe v3 bus provide the possibility to reach the full 40Gb/s rate per network interface.
A fast caching system was built using 16 SSD drives in a single server. The single-node system has been designed for disk data throughput at full 40Gb/s. We have evaluated data transfer performance in the data center environment using 40GbE switches. The last step in the evaluation was the demonstration, during SuperComputing 2011, of 40Gb/s disk-to-disk data throughput between a pair of servers over close to 4000 km WAN circuit.

We review our experience with 40GbE technology in the LAN and WAN environment. We describe the system design, tuning performed, and the performance achieved.
The system described has potential application as a caching or front-end system to a large, conventional, storage system, allowing fast data movement over high-capacity network channels. Such a system is of particular interest in combination with dynamic bandwidth reservation systems, as it allows efficient use of network resources available during the reservation period.

Student? Enter 'yes'. See http://goo.gl/MVv53:

No

Computer Facilities, Production Grids and Networking / 381

Using Xrootd to Federate Regional Storage

Authors: Brian Paul Bockelman[1]; Robert GARDNER[2]

Co-authors: Andrew Hanushevsky [3]; Avi Yagil [4]; Daniel Charles Bradley [5]; David Lesny [6]; Doug Benjamin [7]; Frank Wurthwein [8]; Giacinto Donvito [9]; Hironori Ito [10]; Horst Severini [11]; Igor Sfiligoi [12]; Kenneth Bloom [1]; Lothar Bauerdick [13]; Matevz Tadel [4]; Michael Ernst [14]; Ofer Rind [15]; Patrick Mcguigan [16]; Sarah Williams [17]; Shawn Mc Kee [18]; Sridhara Dasu [19]; Wei Yang [20]

[1] *University of Nebraska (US)*

[2] *UNIVERSITY OF CHICAGO*

[3] *STANFORD LINEAR ACCELERATOR CENTER*

[4] *Univ. of California San Diego (US)*

[5] *High Energy Physics*

[6] *Univ. Illinois at Urbana-Champaign (US)*

[7] *Duke University (US)*

[8] *UCSD*

[9] *Universita e INFN (IT)*

[10] *Brookhaven National Laboratory (US)*

[11] *University of Oklahoma (US)*

[12] *University of California San Diego*

[13] *FERMILAB*

[14] *Unknown*

[15] *BROOKHAVEN NATIONAL LABORATORY*

[16] *University of Texas at Arlington (US)*

[17] *Indiana University (US)*

[18] *University of Michigan (US)*

[19] *University of Wisconsin (US)*

[20] *SLAC National Accelerator Laboratory (US)*

Corresponding Authors: brian.bockelman@cern.ch, robert.w.gardner@cern.ch

While the LHC data movement systems have demonstrated the ability to move data at the necessary throughput, we have identified two weaknesses: the latency for physicists to access data and the complexity of the tools involved. To address these, both ATLAS and CMS have begun to federate regional storage systems using Xrootd. Xrootd, referring to a protocol and implementation, allows us to provide data access to all disk-resident data from a single virtual endpoint. This "redirector" endpoint (which may actually be multiple physical hosts) discovers the actual location of the data and redirects the client to the appropriate site. The approach is particularly advantageous since typically the redirection requires much less than 500 milliseconds (bounded by network round trip time) and the Xrootd client is conveniently built into LHC physicist's analysis tools.

Currently, there are three regional storage federations - a US ATLAS region, a European CMS region, and a US CMS region. The US ATLAS and US CMS regions include their respective Tier 1 and Tier 2 facilities, meaning a large percentage of experimental data is available via the federation. There are plans for federating storage globally and so studies of the peering between the regional federations is of particular interest.

From the base idea of federating storage behind an endpoint, the implementations and use cases diverge. For example, the CMS software framework is capable of efficiently processing data over high-latency data, so using the remote site directly is comparable to accessing local data. ATLAS's processing model is currently less resilient to latency, and they are particularly focused on the physics n-tuple analysis use case; accordingly, the US ATLAS region relies more heavily on caching in the Xrootd server to provide data locality.

Both VOs use GSI security. ATLAS has developed a mapping of VOMS roles to specific filesystem authorizations, while CMS has developed callouts to the site's mapping service. Each federation presents a global namespace to users. For ATLAS, the global-to-local mapping is based on a heuristic-based lookup from the site's local file catalog, while CMS does the mapping based on translations given in a configuration file.

We will also cover the latest usage statistics and interesting use cases that have developed over the previous 18 months.

Online Computing / 382

DZERO Level 3 DAQ/Trigger Closeout

Authors: Andrew Haas[1]; Aran Garcia-Bellido[2]; David Cutts[3]; Douglas Chapin[4]; Gordon Watts[5]; John Alexander Backus Mayes[1]; Lidija Zivkovic[6]; Thomas Gadfort[7]; Yun-Tse Tsai[2]; Yunhe Xie[8]

[1] *SLAC National Accelerator Laboratory (US)*

[2] *University of Rochester*

[3] *Brown University (US)*

[4] *BROWN UNIVERSITY*

[5] *University of Washington (US)*

[6] *Brown University*

[7] *Brookhaven National Laboratory (US)*

[8] *Fermilab*

Corresponding Author: gwatts@uw.edu

The Tevatron Collider, located at the Fermi National Accelerator Laboratory, delivered its last 1.96 TeV proton-antiproton collisions on September 30th, 2011. The DZERO experiment continues to take cosmic data for final alignment for several more months . Since Run 2 started, in March 2001, all DZERO data has been collected by the DZERO Level 3 Trigger/DAQ System. The system is a modern, networked, commodity hardware trigger and data acquisition system based around a large central switch with about 60 front ends and 200 trigger computers. DZERO front end crates are VME based. Single Board Computer interfaces between detector data on VME and the network transport for the DAQ system. Event flow is controlled by the Routing Master which can steer events to clusters of farm nodes based on the low level trigger bits that fired. The farm nodes are multi-core commodity computer boxes, without special hardware, that run isolated software to make the final Level 3 trigger decision. Passed events are transferred to the DZERO online system. We will report on the final status and state of the system, along with some of the more interesting milestones throughout its history.

Event Processing / 383

Using Functional Languages and Declarative Programming to Analyze Large Datasets: LINQToROOT

Author: Gordon Watts[1]

[1] *University of Washington (US)*

Corresponding Author: gwatts@uw.edu

Modern HEP analysis requires multiple passes over large datasets. For example, one has to first reweight the jet energy spectrum in Monte Carlo to match data before you can make plots of any other jet related variable. This requires a pass over the Monte Carlo and the Data to derive the reweighting, and then another pass over the Monte Carlo to plot the variables you are really interested in. With most modern ROOT based tools this requires separate analysis loops for each pass, and script files to glue to the two analysis loops together. A prototype framework has been developed that uses the functional and declarative features of C# and LINQ to specify the analysis. The framework uses language tools to convert the analysis into C++ and runs ROOT or PROOF as a backend to get the results. This gives the analyzer the full power of a object-oriented programming language to put together the analysis and at the same time the speed of C++ for the analysis loop. The tool allows one to incorporate C++ algorithms written for ROOT by others. The code is mature enough to have been used in ATLAS analyses. The package is open source and available on the open source site Codeplex.

Summary:

A new approach to end-user analysis that tries to take advantage of modern (i.e. computer language research from the '70s) and the already exiting infrastructure in HEP (i.e. ROOT).

Poster Session / 384

ROOT.NET: Using ROOT from .NET languages like C# and F#

Author: Gordon Watts[1]

[1] *University of Washington (US)*

Corresponding Author: gwatts@uw.edu

ROOT.NET provides an interface between Microsoft's Common Language Runtime (CLR) and .NET technology and the ubiquitous particle physics analysis tool, ROOT. ROOT.NET automatically generates a series of efficient wrappers around the ROOT API. Unlike pyROOT, these wrappers are statically typed and so are highly efficient as compared to the Python wrappers. The connection to .NET means that one gains access to the full series of languages developed for the CLR including functional languages like F# (based on OCaml). Many features that make ROOT objects work well in the .NET world are added (properties, IEnumerable interface, LINQ compatibility, etc.). Dynamic languages based on the CLR can be used as well, of course (Python, for example). Additionally it is now possible to access ROOT objects that are unknown to the translation tool. This poster will describe the techniques used to effect this translation, along with performance comparisons, and examples. All described source code is posted on the open source site Codeplex.

Online Computing / 385

Operational experience with the CMS Data Acquisition System

Author: Hannes Sakulin[1]

Co-authors: Alexander Flossdorf [2]; Andre Georg Holzner [3]; Andrea Petrucci [1]; Andrei Cristian Spataru [1]; Attila Racz [1]; Aymeric Arnaud Dupont [1]; Christian Deldicque [1]; Christian Hartl [1]; Christoph Paus [4]; Christoph Schwick [1]; Dennis Shpakov [5]; Dominique Gigi [1]; Emilio Meschi [1]; Frank Glege [1]; Frans Meijers [1]; Gerry Bauer [4]; Giovanni Polese [1]; James Branson [3]; Jeroen Hegeman [1]; Jose Antonio Coarasa Perez [1]; Konstanty Sumorok [4]; Lorenzo Masetti [1]; Luciano Orsini [1]; Marc Dobson [1]; Marco Pieri [3]; Matteo Sani [3]; Matthew Bowen [6]; Michal Simon ; Olivier Raginel [4]; Remi Mommsen [5]; Robert Gomez-Reino Garrido [1]; Samim Erhan [7]; Sebastian Bukowiec [1]; Sergio Cittolin [3]; Ulf Behrens [2]; Vivian O'Dell [8]; Yi Ling Hwong [1]

[1] *CERN*

[2] *Deutsches Elektronen-Synchrotron (DE)*

[3] *Univ. of California San Diego (US)*

[4] *Massachusetts Inst. of Technology (US)*

[5] *Fermi National Accelerator Lab. (US)*

[6] *University of the West of England*

[7] *Univ. of California Los Angeles (US)*

[8] *Fermi National Accelerator Laboratory (FNAL)*

Corresponding Authors: hannes.sakulin@cern.ch, emilio.meschi@cern.ch

The data-acquisition (DAQ) system of the CMS experiment at the LHC performs the read-out and assembly of events accepted by the first level hardware trigger. Assembled events are made available to the high-level trigger (HLT), which selects interesting events for offline storage and analysis. The system is designed to handle a maximum input rate of 100 kHz and an aggregated throughput of 100 GB/s originating from approximately 500 sources and 10^8 electronic channels. An overview of the architecture and design of the hardware and software of the DAQ system is given. We report on the performance and operational experience of the DAQ and its Run Control System in the first two years of collider run of the LHC, both in proton-proton and Pb-Pb collisions. We present an analysis of the current performance, its limitations, and the most common failure modes and discuss the ongoing evolution of the HLT capability needed to match the luminosity ramp-up of the LHC.

Summary:

The data-acquisition (DAQ) system of the CMS experiment at the LHC performs the read-out and assembly of events accepted by the first level hardware trigger. Assembled events are made available to the high-level trigger (HLT), which selects interesting events for offline storage and analysis. The system is designed to handle a maximum input rate of 100 kHz and an aggregated throughput of 100 GB/s originating from approximately 500 sources and 10^8 electronic channels. An overview of the architecture and design of the hardware and software of the DAQ system is given. We report on the performance and operational experience of the DAQ and its Run Control System in the first two years of collider run of the LHC, both in proton-proton and Pb-Pb collisions. We present an analysis of the current performance, its limitations, and the most common failure modes and discuss the ongoing evolution of the HLT capability needed to match the luminosity ramp-up of the LHC.

Poster Session / 386

Using Zoom Technologies To Display HEP Plots and Talks

Author: Gordon Watts[1]

[1] *University of Washington (US)*

Corresponding Author: gwatts@uw.edu

Particle physics conferences and experiments generate a huge number of plots and presentations. It is impossible to keep up. A typical conference (like CHEP) will have 100's of plots. A single analysis result from a major experiment will have almost 50 plots. Scanning a conference or sorting out what

plots are new is almost a full time job. The advent of multi-core computing and advanced video cards means that we have more processor power available for visualization than any time in the past. This poster describes two related projects that take advantage of this to solve the viewing problem. The first, Collider Plots, has a backend that looks fro new plots released by ATLAS, CMS, CDF, and DZERO and organizes them by date, by experiment, and by subgroup for easy viewing and sorting. It maintains links back to associated conference notes and web pages with full result information. The second project, Deep Conference, renders all the slides as a single large zoomable picture. In both cases, much like a web mapping program, details are revealed as you zoom in. In the case of Collider Plots the plots are stacked as histograms to give visual clues for the most recent updates and activity have occurred. Standard plug-in software for a browser allows a user to zoom in on a portion of the conference that looks interesting. As the user zooms further more and more details become visible, allowing the user to make a quick and cheep decision on whether to spend more time on a particular talk or series of plots. Both projects are available at http://deeptalk.phys.washington.edu. The poster discusses the implementation and use as well as cross platform performance and possible future directions. A demo will be shown.

Software Engineering, Data Stores and Databases / **387**

Tiered Storage For LHC

Authors: Andrew Hanushevsky[1]; Wei Yang[2]

[1] *STANFORD LINEAR ACCELERATOR CENTER*

[2] *SLAC National Accelerator Laboratory (US)*

Corresponding Authors: yangw@slac.stanford.edu, abh@stanford.edu

For more than a year, the ATLAS Western Tier 2 (WT2) at SLAC National Accelerator has been successfully operating a two tiered storage system based on Xrootd's flexible cross-cluster data placement framework, the File Residency Manager. The architecture allows WT2 to provide both, high performance storage at the higher tier to ATLAS analysis jobs, as well as large, low cost disk capacity at the lower tier. Data automatically moves between the two storage tiers based on the needs of analysis jobs and is completely transparent to the jobs.

Poster Session / **388**

Application of Control System Studio for the NOvA Detector Control System.

Authors: Gennadiy Lukhanin[1]; Martin Frank[2]

Co-authors: Athanasios Hatzikoutelis [3]; James Kowalkowski [4]; Kurt Biery [5]; Ronald Rechenmacher [6]

[1] *Fermi National Accelerator Lab. (US)*

[2] *UVA*

[3] *UTK*

[4] *Fermi National Accelerator Laboratory (FNAL)*

[5] *CMS/Fermilab*

[6] *FNAL*

Corresponding Authors: lukhanin@fnal.gov, mfrank@fnal.gov

In the NOvA experiment, the Detector Controls System (DCS) provides a method for controlling and monitoring important detector hardware and environmental parameters. It is essential for operating the detector and is required to have access to roughly 370,000 independent programmable channels

via more than 11,600 physical devices.
In this paper, we demonstrate an application of Control System Studio (CSS), developed by Oak Ridge National Laboratory, for the NOvA experiment. The application of CSS for the DCS of the NOvA experiment has been divided into three phases: (1) user requirements and concept prototype on a test-stand, (2) small scale deployment at the prototype Near Detector on the Surface, and (3) a potential for a larger scale deployment at the Far Detector. We also give an outline of the CSS integration with the NOvA online software and the alarm handling logic for the Front-End electronics.

Poster Session / **389**

Eurogrid: a new glideinWMS based portal for CDF data analysis.

Authors: Donatella Lucchesi[1]; Doug Benjamin[2]; Gabriele Compostella[3]; Silvia Amerio[4]

[1] *INFN Padova*

[2] *Duke University (US)*

[3] *Max-Planck-Institut fuer Physik-Max-Planck-Gesellschaft (MPG)*

[4] *University of Padova & INFN*

Corresponding Author: silvia.amerio@pd.infn.it

The CDF experiment at Fermilab ended its Run-II phase on September 2011 after 11 years of operations and 10 fb-1 of collected data.
CDF computing model is based on a Central Analysis Farm (CAF) consisting of local computing and storage resources, supported by
OSG and LCG resources accessed through dedicated portals.
Recently a new portal, Eurogrid, has been developed to effectively exploit computing and
disk resources in Europe: a dedicated farm and storage area at the TIER-1 CNAF computing
center in Italy, and additional LCG computing resources at different TIER-2 sites in Italy, Spain, Germany and France, are accessed through a common interface.
The goal of this project was to develop a portal 1) easy to integrate in the existing CDF computing model, 2) completely transparent to the user and 3) requiring a minimum amount of maintenance support by the CDF collaboration.
In this talk we will review the implementation of this new portal, and the performance in the first months of usage.
Eurogrid is based on the glideinWMS[1] software, a Glidein Based WMS that
works on top of Condor [2]. As CDF CAF is based on
Condor, the choice of the glideinWMS software was natural and the implementation seamless.
Thanks to the pilot jobs, user needs and site resources are matched in a very efficient way, completely transparent to the users.
Official since June 2011, Eurogrid effectively complements and supports CDF computing resources and is the best solution for the future in terms of required manpower for administration, support and development.

Poster Session / **390**

Belle II High Level Trigger at SuperKEKB

Authors: Ryosuke ITOH[1]; Soohyung Lee[2]

Co-authors: Eunil Won [2]; Mikihiko Nakao [1]; Soh Suzuki [1]; Takeo Higuchi [1]

[1] *KEK*

[2] *Korea University*

Corresponding Author: shlee@hep.korea.ac.kr

A next generation B-factory experiment, Belle II, is now being constructed at KEK in Japan. The upgraded accelerator SuperKEKB is designed to have the maximum luminosity of 8 × 10ˆ35 cmˆ−2sˆ−1 that is a factor of 40 higher than the current world record. As a consequence, the Belle II detector yields a data stream of the event size ˜1 MB at a Level 1 rate of 30 kHz.
The Belle II High Level Trigger (HLT) is designed to reduce the Level 1 rate to 1/5 by performing the real time full event reconstruction and by applying the physics level event selection as the software trigger. The results of the processing are also fed back to the readout system of the pixel detector for the further data size reduction.
The event processing framework for HLT is intended to be the same as that used in offline so that the same reconstruction codes can be shared. The HLT framework is desired to be based on "basf2" , which is the unified software framework for the Belle II data processing. The basf2 framework is designed to be used in a single node although it has the parallel processing capability utilizing multicores. For the HLT purpose, we need to extend the parallel processing to make use of multiple PC servers connected over the network, and therefore, a super-framework called hbasf2 is developed.
The Belle II High Level Trigger system takes full advantages of multicore and network connected PC servers for the parallel processing with the hbasf2 framework. It also provides the control function of HLT such as the configuration management, the real time monitoring, etc.
In this contribution, the details of the design and implementation of hbasf2 are presented. The processing performance of hbasf2 measured using the prototype HLT test bench is also reported.

Student? Enter 'yes'. See http://goo.gl/MVv53:

Yes

Poster Session / 391

ATLAS Data Caching based on the Probability of Data Popularity

Authors: Alexei Klimentov[1]; Gergely Zaruba[2]; Kaushik De[2]; Mikhail Titov[2]

[1] *Brookhaven National Laboratory (US)*

[2] *University of Texas at Arlington (US)*

Corresponding Author: mikhail.titov@cern.ch

Efficient distribution of physics data over ATLAS grid sites is one of the most important tasks for user data processing. ATLAS' initial static data distribution model over-replicated some unpopular data and under-replicated popular data, creating heavy disk space loads while under-utilizing some processing resources due to low data availability. Thus, a new data distribution mechanism was implemented, PD2P (PanDA Dynamic Data Placement) within the production and distributed analysis system PanDA that dynamically reacts to user data needs [1], basing dataset distribution principally on user demand. Data deletion is also demand driven, reducing replica counts for unpopular data [2]. This dynamic model has led to substantial improvements in efficient utilization of storage and processing resources.
Based on this experience, in this work we seek to further improve data placement policy by investigating in detail how data popularity is calculated. For this it is necessary to precisely define what data popularity means, what types of data popularity exist, how it can be measured, and most importantly, how the history of the data can help to predict the popularity of derived data. We introduce locality of the popularity: a dataset may be only of local interest to a subset of clouds/sites or may have a wide (global) interest. We also extend the idea of the "data temperature scale"model [3] and a popularity measure.
Using the ATLAS data replication history, we devise data distribution algorithms based on popularity measures and past history. Based on this work we will describe how to explicitly identify why and how datasets become popular and how such information can be used to predict future popularity.

[1] Kaushik De, Tadashi Maeno, Torre Wenaus, Alexei Klimentov, Rodney Walker, Graeme Stewart, "PD2P –PanDA Dynamic Data Placement", ATLAS Notes, CERN
[2] Angelos Molfetas, Fernando Barreiro Megino, Andrii Tykhonov, Vincent Garonne, Simone Campana, Mario Lassnig, Martin Barisits, Gancho Dimitrov, Florbela Tique Aires Viegas, "Popularity framework to process dataset tracers and its application on dynamic replica reduction in the ATLAS experiment", CHEP, Taipei, Taiwan, October 18-22, 2010
[3] Alexei Klimentov, "ATLAS data over Grid (data replication, placement and deletion policy)", ATLAS Notes, CERN, March 17, 2009

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Poster Session / 392

Data acquisition and online monitoring software for CBM test-beams

Author: Jorn Adamczewski-Musch[1]

Co-authors: Nikolaus Kurz [2]; Peter Zumbruch [2]; Sergey Linev [3]

[1] *GSI - Helmholtzzentrum fur Schwerionenforschung GmbH (DE)*

[2] *GSI*

[3] *GSI DARMSTADT*

Corresponding Authors: jorn.adamczewski-musch@cern.ch, s.linev@gsi.de, p.zumbruch@gsi.de

The Compressed Baryonic Matter (CBM) experiment is intended to run at the FAIR facility that is currently being build at GSI in Darmstadt, Germany. For testing of future CBM detector and readout electronics prototypes, several test beamtimes have been performed at different locations, such as GSI, COSY, and CERN PS.
The DAQ software has to treat various data inputs, e.g. standard VME modules on the MBS system, or different kinds of the FPGA boards, read via USB, Ethernet or optical links.
The Data Acquisition Backbone Core framework (DABC) is able to combine such different data sources with event builder processes running on regular Linux PCs.
DABC can also retrieve the instrumental set up data from EPICS slow control systems and insert it into the event data stream for later analysis. Vice versa, the DIM based DABC control protocol has been integrated to the general CBM EPICS ioc by means of an EPICS-DIM interface. Hence the DAQ can be monitored and steered with an CSS based operator GUI.
The CBM online monitoring analysis is based on the GSI Go4 framework which can directly connect to DABC online data via sockets, or process previous data from listmode files. A Go4 subframework was implemented to provide possibility of parallel development of analysis code for different subdetectors groups. This allows to divide up the Go4 components into independent software packages that can run either standalone, or together at the beamtime in a full set up.

Poster Session / 393

The WLCG Messaging Service and its Future

Authors: Lionel Cons[1]; Massimo Paladin[2]

[1] *CERN*

[2] *Universita degli Studi di Udine*

Corresponding Authors: lionel.cons@cern.ch, massimo.paladin@cern.ch

Messaging is seen as an attractive mechanism to simplify and extend several portions of the Grid middleware, from low level monitoring to experiments dashboards. The messaging service currently used by WLCG is operated by EGI and consists of four tightly coupled brokers running ActiveMQ and designed to host the Grid operational tools such as SAM.

This service is successfully being used by several Grid operational tools. To improve these services and widen the use of the technology we identified three core aspects that have to evolve: security, scalability and availability/reliability.

In this paper we describe the WLCG messaging service, it's future and the technical solutions being put in place to address the anticipated needs while preserving backward compatibility for its current applications.

Event Processing / **394**

Event Reconstruction in the PandaRoot framework

Author: Stefano Spataro[1]

[1] *University of Turin*

Corresponding Author: spataro@to.infn.it

The PANDA experiment will study the collisions of beams of anti-protons, with momenta ranging from 2-15 GeV/c, with fixed proton and nuclear targets in the charm energy range, and will be built at the FAIR facility. In preparation for the experiment, the PandaRoot software framework is under development for detector simulation, reconstruction and data analysis, running on an Alien2-based grid. The basic features are handled by the FairRoot framework, based on ROOT and Virtual Monte Carlo, while the PANDA detector specifics and reconstruction code are implemented inside PandaRoot. The realization of Technical Design Reports for the tracking detectors has pushed the finalization of the tracking reconstruction code, which is complete for the Target Spectrometer, and of the analysis tools. Particle Identification algorithms are implemented using Bayesian approach and compared to Multivariate Analysis methods. Moreover, the PANDA data acquisition foresees a triggerless operation in which events are not defined by a 1st level trigger decision, but all the signals are stored with time stamps requiring a deconvolution by the software. This has led to a redesign of the software from an event basis to a time-ordered structure. In this contribution, the reconstruction capabilities of the Panda spectrometer will be reported, focusing on the performances of the tracking system and the results for the analysis of physics benchmark channels, as well as the new (and challenging) concept of time-based simulation and its implementation.

Poster Session / **395**

GFAL 2.0 Evolutions & GFAL-File system introduction

Author: Adrien Devresse[1]

[1] *University of Nancy I (FR)*

Corresponding Author: adrien.devresse@cern.ch

The Grid File Access Library (GFAL) is a library designed for a universal and simple access to grid storage systems. Re-designed and re-written completely, the 2.0 version of GFAL provides a complete abstraction of the complexity and heterogeneity of the grid storage systems (DPM, LFC, Dcache, Storm, arc, ...) and of the data management protocols (RFIO, gsidcap, LFN, dcap, SRM, Http/webdav, gridFTP) by a simpler, faster, more reliable and more consistent POSIX API.

GFAL 2.0 is not only an improvement of the GFAL 1.0's reliability, several new functionalities have been developed like the extended attributes management, the runtime configuration setter/getter, a new scalable plugin system, new operations and new protocol (http/webdav) support and the GFAL FUSE module.

GFAL 2.0 is delivered with gfalFS (GFAL 2.0 FUSE module), a new tool that provides a Virtual File System common to all the grid storage systems (Dcache, DPM, , WebDAV server), allowing a user to mount these resources.

In this paper I analyse in detail the new functionality and the new possibilities brought by GFAL 2.0 and gfalFS, like the new plugin system for the support of the new protocols , the new error report system, the old issues corrected, the new development-kit provided. A comparison of the performance benefit/loss of the GFAL 2.0/gfalFS vs the other existing tools on the different storage systems is explained. More details are presented as well on the GFAL 2.X future improvements and possibilities.

Poster Session / **396**

LCIO2.0: Event Data Model and Persistency for HEP

Authors: Frank Gaede[1]; Jan Dominik Engels[2]; Jeremy McCormick[3]; Norman Anthony Graf[4]; Steven Aplin[5]; Tony Johnson[6]

[1] *DESY IT*

[2] *Deutsches Elektronen-Synchrotron (DESY)*

[3] *Unknown*

[4] *SLAC National Accelerator Laboratory (US)*

[5] *DESY*

[6] *Nuclear Physics Laboratory*

Corresponding Author: norman.graf@slac.stanford.edu

LCIO is a persistency framework and event data model which, as originally presented at CHEP 2003, was developed for the next linear collider physics and detector response simulation studies. Since then, the data model has been extended to also incorporate raw data formats as well as reconstructed object classes. LCIO defines a common abstract user interface (API) and is designed to be lightweight and flexible without introducing additional dependencies on other software packages. Concrete implementations are provided in several programming languages, providing end users the flexibility of using multiple simulation, reconstruction and analysis frameworks. Persistence is provided by a simple binary format that supports data compression and random event access.
LCIO is being used by the ILC and CLiC physics and detector communities to conduct performance benchmarking studies such as the recently completed CLiC CDR and the ILC Detector Baseline Design study to be completed in 2012. Detector studies for the Muon Collider are also being conducted using LCIO as the event data model and persistency. Multiple test-beam collaborations have used LCIO to store and process tens of millions of events, providing experience with real data. Recently the Heavy Photon Search collaboration adopted LCIO as its event data model and offline persistency format.
In this talk we present details of its use in these various applications, and discuss the successful cooperation and collaboration LCIO has enabled. We will also present the design and implementation of new features introduced in LCIO2.0.

Poster Session / **397**

mesh2gdml: from CAD to Geant4

Author: Norman Anthony Graf[1]

[1] *SLAC National Accelerator Laboratory (US)*

Corresponding Author: norman.graf@slac.stanford.edu

The ability to directly import CAD geometries into Geant4 is an often requested feature, despite the recognized limitations of the difficulty in accessing proprietary formats, the mismatch between level of detail in producing a part and simulating it, the often disparate approaches to parent-child relationships and the difficulty in maintaining or assigning material definitions to parts.

Geant4 provides a very rich library of basic geometrical shapes, often referred to as "primitives" , plus the ability to define compound geometries via boolean operations. It is therefore capable of supporting extremely complex physical geometries composed of simple primitives. Most CAD systems also incorporate primitive volumes, but their definitions differ between programs and often do not map onto the Geant4 primitives, making the conversion difficult at best. However, one can also define a solid in Geant4 as a volume composed of surface facets. This G4TessellatedSolid can be composed of either triangular or quadrangular facets and therefore provides a mechanism for the programmatic importation of shapes and volumes defined in many CAD systems. In addition to CAD programs, there are very many 3D modeling programs which provide the user with convenient graphical user interfaces to create solid models. Usually aimed at gaming or rendering engines, these could be useful as a front end for a graphical geometry editor. Many output formats are supported, including tesselations. Furthermore, this approach provides a useful solution in cases where the objects are intrinsically irregular, such as biological phantoms.

The main impediment to the importation of CAD files into Geant4 has been their proprietary formats. Some existing solutions target recognized interchange formats such as STEP, but even these formats provide challenges, such as complicated file formats, possible loss of hierarchy or material association and little or no mapping to primitives. Thanks to the proliferation of rapid prototyping and additive manufacturing processes, the surface tesselation language (STL) format is the industrial standard for handling triangulated meshes and is ubiquitous as an export format for both CAD and other 3D modelling software. The format consists of a plain list of three dimensional corner point coordinates (vertex) and flat triangles (facet) with an associated normal vector, making it an ideal candidate for importation into Geant4.

In this talk, we present mesh2gdml, a solution which converts an STL file into a GDML file which can be imported directly into Geant4. The STL facets are translated directly into G4TriangularFacets which are used to create G4TessellatedSolids. Since there is no other structure in an STL file, one has to also solve the problem of creating "topology from a bucket of facets", which we have done. The one area requiring manual intervention is the assignment of material to the newly created solid or solids. Finally, one can either create a world volume from the bounding box of the volume(s) found in the STL file to use standalone within Geant4, or leave the resulting gdml file as individual volumes to aggregate or incorporate into a common world volume later.

In order to benchmark the performance of Geant4 using tesselated volumes, we have written code which allows the exportation of Geant4 geometries in STL format. This geometry is then reimported into Geant4 after being processed with mesh2gdml, allowing us to directly compare the CPU time difference between a geometry composed of primitives and tesselated solids. A side effect of this STL export is the ability to create a real 3D model of the Geant4 geometry on 3D printers. This enables rapid prototyping of parts, direct comparison of the modeled geometry to CAD geometry, communication with colleagues and outreach to the public.

Despite the inherent performance issues related to navigating through geometries composed of many individual facets and the requirement that material be assigned manually to volumes during the translation process, we believe the approach outlined in this talk provides access to a wider range of geometry inputs and will prove to be useful to a number of user communities.

Poster Session / 398

A General Purpose Grid Portal for simplified access to Distributed Computing Infrastructures

Authors: Andrea Ceccanti[1]; Diego Michelotto[2]; Giacinto Donvito[3]; Giuseppe Misurelli[4]; Luciano Gaido[5]; Marco Bencivenni[6]; Paolo Veronesi[4]; Riccardo Brunetti[4]; Valerio Venturi[6]; Vincenzo Ciaschini[7]

[1] *Istituto Nazionale Fisica Nucleare (IT)*

[2] *INFN Ferrara & IGI*

[3] *INFN-Bari*

[4] *Unknown*

[5] *Universita e INFN (IT)*

[6] *INFN*

[7] *INFN CNAF*

Corresponding Author: marco.bencivenni@cnaf.infn.it

One of the main barriers against Grid widespread adoption in scientific communities stems from the intrinsic complexity of handling X.509 certificates, which represent the foundation of the Grid security stack.
To hide this complexity, in recent years, several Grid portals have been proposed which, however, do not completely solve the problem, either requiring that users manage their own certificates or proposing solutions that weaken the Grid middleware authorization and accounting mechanisms by obfuscating the user identity.
General purpose Grid portals aim at providing a powerful and easy to use gateway to distributed computing resources. They act as incubators where users can securely run their applications without facing the complexity of the authentication infrastructure (e.g., handling X.509 certificates and VO membership requests, accessing resources through dedicated shell-based UIs).
In this paper, we discuss a general purpose Grid portal framework, based on Liferay, which provides several important services such as job submission, workflow definition, data management and accounting services. It is also interfaced with external Infrastructure-as-a-Service (IaaS) frameworks for the dynamic provisioning of computing resources.
In our model, authentication is demanded to a Shibboleth 2.0 federation while the generation and management of Grid credentials is handled securely integrating an On-Line CA with the MyProxy server. Consequently, the portal gives users full access to Grid functionality without exposing the complexity of X.509 certificates and proxy management.
Unlike other existing solutions, our portal does not leverage robot certificates for the user credentials. This approach offers twofold benefits. On the one hand, user identity is not obfuscated across the middleware stack thus preserving the functionality and effectiveness of existing distributed accounting and authorization mechanisms. On the other hand, users are not constrained to a predefined set of applications but can freely take advantage of Grid facilities for any computational or data-intensive activity.
The portal also provides simplified access to common Grid data-management operations. Our solution manages the staging of input and output data for Grid jobs to an external WebDAV storage service. The staged data is then transferred to or from Grid SE and registered in data catalogs on behalf of the user. This approach has two main benefits. Firstly, by delegating the file transfer handling to an external service, the portal is relieved from the potential load caused by many concurrent large file transfers operations that would severely impact its scalability. Secondly, the use of standard protocols like WebDAV enables any client machine to upload and download files to the Grid without requiring installation of custom software on the client side.

Poster Session / 399

Electron reconstruction and identification capabilities of the CBM Experiment at FAIR

Author: Semen Lebedev[1]

Co-authors: Andrey Lebedev [2]; Claudia Höhne [3]; Gennady Ososkov [4]

[1] *GSI - Helmholtzzentrum fur Schwerionenforschung GmbH (DE)*

[2] *GSI, Darmstadt / LIT JINR, Dubna*

[3] *JLU Giessen, Giessen*

[4] *LIT JINR, Dubna*

Corresponding Author: s.lebedev@gsi.de

The Compressed Baryonic Matter (CBM) experiment at the future FAIR facility at Darmstadt will measure dileptons emitted from the hot and dense phase in heavy-ion collisions. In case of an electron measurement, a high purity of identified electrons is required in order to suppress the background. Electron identification in CBM will be performed by a Ring Imaging Cherenkov (RICH) detector and Transition Radiation Detectors (TRD).
In this contribution, algorithms which were developed for the electron reconstruction and identification in RICH and TRD detectors are presented. A fast RICH ring recognition algorithm based on the Hough Transform was implemented. An ellipse fitting algorithm was elaborated because most of the CBM RICH rings have elliptic shapes. An efficient algorithm based on the Artificial Neural Network is implemented for electron identification in RICH. In TRD track reconstruction algorithm which is based on track following and Kalman Filter methods was implemented. Several algorithms for electron identification in TRD were developed and investigated. The best-performed algorithm is based on the special transformation of energy losses measured in TRD and usage of the Boosted Decision Tree (BDT) as classifier. Results and comparison of different methods of electron identification and pion suppression are presented.

Software Engineering, Data Stores and Databases / **400**

Evolution of grid-wide access to database resident information in ATLAS using Frontier

Author: Collaboration Atlas[1]

Co-authors: Alastair Dewhurst [2]; Carlos Fernando Gamboa [3]; Dario Barberis [4]; Dave Dykstra [5]; David Front [6]; Elizabeth Gallas [7]; Florentin Bujor [8]; Fred Luehring [9]; John Steven De Stefano Jr [10]; Rodney Walker [11]

[1] *Atlas*

[2] *STFC - Science & Technology Facilities Council (GB)*

[3] *Department of Physics-Brookhaven National Laboratory (BNL)-Unkno*

[4] *Universita e INFN (IT)*

[5] *Fermi National Accelerator Lab. (US)*

[6] *SW engineer*

[7] *University of Oxford (GB)*

[8] *University of Wisconsin (US)*

[9] *Indiana University (US)*

[10] *Brookhaven National Laboratory (US)*

[11] *Ludwig-Maximilians-Univ. Muenchen (DE)*

Corresponding Author: alastair.dewhurst@cern.ch

The ATLAS experiment deployed Frontier technology world-wide during the the initial year of LHC collision data taking to enable user analysis jobs running on the World-wide LHC Computing Grid to access database resident data. Since that time, the deployment model has evolved to optimize resources, improve performance, and streamline maintenance of Frontier and related infrastructure. In this presentation we focus on the specific changes in the deployment and improvements undertaken such as the optimization of cache and launchpad location, the use of RPMs for more uniform deployment of underlying Frontier related components, improvements in monitoring, optimization of fail-over,
and an increasing use of a centrally managed database containing site specific information
(for configuration of services and monitoring).
In addition, analysis of Frontier logs has allowed us a deeper understanding of problematic queries and understanding of use cases. Use of the system has grown beyond just user analysis and subsystem specific tasks such as calibration and alignment, extending into production processing areas such as initial reconstruction and trigger reprocessing. With a more robust and tuned system, we are

better equiped to satisfy the still growing number of diverse clients and the demands of increasingly sofisticated processing and analysis.

Software Engineering, Data Stores and Databases / 401

New software library of geometrical primitives for modelling of solids used in Monte Carlo detector simulations

Author: Marek Gayer[1]

Co-authors: Andrei Gheata [1]; Gabriele Cosmo [1]; Jean-Marie Guyader [1]; John Apostolakis [1]; Tatiana Nikitina [2]

[1] *CERN*

[2] *Universite de Franche-Comte*

Corresponding Author: marek.gayer@cern.ch

We present our effort for the creation of a new software library of geometrical primitives, which are used for solid modelling in Monte Carlo detector simulations. We plan to replace and unify current geometrical primitive classes in the CERN software projects Geant4 and ROOT with this library. Each solid is represented by a C++ class with methods suited for measuring distances of particles from the surface of a solid and for determination as to whether the particles are located inside, outside or on the surface of the solid. We use numerical tolerance for determining whether the particles are located on the surface. The class methods also contain basic support for visualization.

We use dedicated test suites for validation of the shape codes. These include also special performance and numerical value comparison tests for help with analysis of possible candidates of class methods as well as to verify that our new implementation proposals were designed and implemented properly.

Currently, bridge classes are used for simple integration of the library to existing versions of Geant4 and ROOT software. New versions of Geant4 and ROOT are planned to be modified in the way that our new solids library can be used there directly.

Summary:

We present our effort for the creation of a new software library of geometrical primitives, which will unify current geometrical primitive classes in the CERN software projects Geant4 and ROOT. The solids are represented by C++ classes with several methods, namely those determining distance and location of the particle with relation to the surface of solids. We use several test suites for validation. A simple integration method for easy integration to the current versions of Geant4 and ROOT is proposed, that can be used in the interim period before both ROOT and GEANT4 are adapted to use our library directly.

Poster Session / 403

Evaluation of a new data staging framework for the ARC middleware

Authors: Aleksandr Konstantinov[1]; Andrej Filipcic[2]; David Cameron[3]; Dmytro Karpenko[4]

[1] *VILNIUS UNIVERSITY*

[2] *Jozef Stefan Institute (SI)*

[3] *University of Oslo (NO)*

[4] *University of Oslo*

Corresponding Author: david.cameron@cern.ch

Staging data to and from remote storage services on the Grid for users' jobs is a vital component of the ARC computing element. A new data staging framework for the computing element has recently been developed to address issues with the present framework, which has essentially remained unchanged since its original implementation 10
years ago. This new framework consists of an intelligent data transfer scheduler which handles priorities and fair-share, a rapid caching system, and the ability to delegate data transfer over multiple nodes to increase network throughput. This paper uses data from real user
jobs running on production ARC sites to present an evaluation of the new framework. It is shown to make more efficient use of the available
resources, reduce the overall time to run jobs, and avoid the problems seen with the previous simplistic scheduling system. In addition, its
simple design coupled with intelligent logic provides greatly increased flexibility for site administrators, end users and future development.

Poster Session / **404**

Service Availability Monitoring framework based on commodity software

Author: Pedro Manuel Rodrigues De Sousa Andrade[1]

Co-authors: Christos Triantafyllidis [2]; David Collados Polidura [1]; Emir Imamagic [3]; Kislay Bhatt [4]; Kumar Vaibhav [1]; Marian Babik [1]; Paloma Fuente Fernandez [1]; Phool Chand [4]; Pradyumna Joshi [4]; Rajesh Kalmady [4]; Urvashi Karnani [4]; Wojciech Lapka [1]; vibhuti duggal [4]

[1] *CERN*

[2] *AUTH*

[3] *SRCE*

[4] *BARC*

Corresponding Authors: pedro.andrade@cern.ch, wojciech.lapka@cern.ch

The Worldwide LHC Computing Grid (WLCG) infrastructure continuously operates thousands of grid services scattered around hundreds of sites. Participating sites are organized in regions and support several virtual organizations, thus creating a very complex and heterogeneous environment. The Service Availability Monitoring (SAM) framework is responsible for the monitoring of this infrastructure.

SAM is a complete monitoring framework for grid services and grid operational tools. Its current implementation tailored for a decentralized operation replaces the old SAM system which is now being decommissioned from production. SAM provides functionality for submission of monitoring probes, gathering of probes results, processing of monitoring data, and retrieval of monitoring data in terms of service status, availability, and reliability.

In this paper we present the SAM framework. We motivate the need from moving from the old SAM to a new monitoring infrastructure deployed and managed in a distributed environment and explain how SAM exploits and builds on top of commodity software, such as Nagios and Apache ActiveMQ, to provide a reliable and scalable system. We also present the SAM architecture by highlighting the adopted technologies and how the different SAM components deliver a complete monitoring framework.

Poster Session / **405**

VISPA@Web: A Server-Client-Based Graphical Development Environment for Physics Analyses

Authors: Andreas Hinzmann[1]; Dennis Klingebiel[1]; Gero Müller[1]; Hans-Peter Bretz[1]; Jan Steggemann[1]; Joschka Lingemann[1]; Marcel Rieger[1]; Martin Erdmann[1]; Matthias Komm[1]; Robert Fischer[1]; Tobias Winchen[1]

[1] *Rheinisch-Westfaelische Tech. Hoch. (DE)*

Corresponding Author: martin.erdmann@cern.ch

The Visual Physics Analysis (VISPA) project addresses the typical development cycle of (re-)designing, executing, and verifying an analysis.
It presents an integrated graphical development environment for physics analyses, using the Physics eXtension Library (PXL) as underlying C++ analysis toolkit.
Basic guidance to the project is given by the paradigms of object oriented programming, data flow management, and graphical representation.
In this contribution we present extension of the project to make the physics analysis functionality accessible via a standard internet browser.
Utilizing the server-client based approach avoids common requirements such as package installations or specialized computing resources on the client side.
With the web browser being the only needed software, mobile devices like tablet computers or smart phones can now be used for physics analysis.
New use cases like selected access to experiment data and analyses by the public are possible.

Poster Session / 407

An Exhibition Booth for demonstrating recent developments in data processing software used at the LHC

Author: John Harvey[1]

[1] *CERN*

Corresponding Author: john.harvey@cern.ch

The PH/SFT group at CERN is responsible for developing, releasing and deploying some of the software packages used in the data processing systems of CERN experiments, in particular those at the LHC. They include ROOT, GEANT4, CernVM, Generator Services, and Multi-core R&D (http://sftweb.cern.ch/). We have already submitted a number of abstracts for oral presentations at the conference. Here we request access to a booth so that we can continue a dialogue with interested delegates in front of practical demos that we have prepared and that illustrate new developments in detail. We would undertake to keep the booth manned during the breaks and poster sessions.

We would plan to show a number of demos covering a variety of different software domains, namely:
- our latest development for Apple's mobile devices; firstly a "RootBrowser" for iPad and iPhone and secondly ROOT's event visualization framework ported to OpenGL for embedded systems/iOS. ("EVE for iPad")
- a prototype of ROOT that highlights a new C++ interpreter (called cling) that has been developed using the new Low Level Virtual Machine compiler technology (LLVM)
- use of the MCPLOTS tool, which is dedicated to the tuning and the validation of MonteCarlo event generators, such as Pythia, and its connection to the LHC@home project; we will show the physics content of MCPLOTS, with different comparisons to the LHC data, as well as the underlying computing technology used to produce these results.
- an automated solution for validation and testing of the Geant4 toolkit through the integration of a variety of tools and technologies, including the configuration and submission of jobs on grid-based resources, as well as the analysis and recording of results.
- the procedures and tools used in the CernVM project to manage the virtual machine (VM) lifecycle for developing, testing and running the software frameworks of the LHC experiments; we will present how one can manage the full CernVM lifecycle using a Web-based user interface as well as a lightweight application for portable devices such as mobile phones and tablets.

Summary:

The PH/SFT group at CERN is responsible for developing, releasing and deploying some of the software packages used in the data processing systems of CERN experiments, in particular those at the LHC. They include ROOT, GEANT4, CernVM, Generator Services, and Multi-core R&D (http://sftweb.cern.ch/). We have already submitted a number of abstracts for oral presentations at the conference. Here we request access to an exhibition booth so that we can continue a dialogue with interested delegates in front of practical demos that we have prepared and that illustrate new developments in detail. Topics to be covered include our latest developments for running ROOT components on mobile devices (iPad & iPhone), a prototype of the new ROOT C++ interpreter, demos of the MCPLOTS tool and of the Geant4 test and validation suite, and finally the procedures and tools used in the CernVM project to manage the virtual machine (VM) lifecycle.

Computer Facilities, Production Grids and Networking / **408**

Review of CERN Computer Centre Infrastructure

Author: Tim Bell[1]

Co-author: Bernd Panzer-Steindel [1]

[1] *CERN*

Corresponding Author: tim.bell@cern.ch

The CERN Computer Centre is reviewing strategies for optimizing the use of the existing infrastructure in the future, and in the likely scenario that any extension will be remote from CERN, and in the light of the way other large facilities are today being operated. Over the past six months, CERN has been investigating modern and widely-used tools and procedures used for virtualisation, clouds and fabric management in order to reduce operational effort, increase agility and support unattended remote computer centres. This presentation will give the details on the project's motivations, current status and areas for future investigation.

Online Computing / **409**

ALICE HLT TPC Tracking of Heavy-Ion Events on GPUs

Author: David Michael Rohr[1]

[1] *Johann-Wolfgang-Goethe Univ. (DE)*

Corresponding Author: drohr@jwdt.org

The ALICE High Level Trigger (HLT) is capable of performing an online reconstruction of heavy-ion collisions.
The reconstruction of particle trajectories in the Time Projection Chamber (TPC) is the most compute intensive step.
The TPC online tracker implementation combines the principle of the cellular automaton and the Kalman filter.
It has been accelerated by the usage of graphics cards (GPUs).
A pipelined processing allows to perform the tracking on the GPU, the data transfer, and the preprocessing on the CPU in parallel.

In order to use data locality, the tracking is split in multiple phases.
At first, track segments are searched in local sectors of the detector, independently and in parallel.
These segments are then merged at a global level.

A shortcoming of this approach is that if a track contains only a very short segment in one particular sector, the local search possibly does not find this short part.
The fast GPU processing allowed to add an additional step:
all found tracks are extrapolated to neighboring sectors and the unassigned clusters which constitute the missing track segment are collected.

For running the QA on computers without a GPU, it is important that the output of the CPU and the GPU tracker is as consistent as possible.
One major challenge was to implement the tracker such that the output is not affected by concurrency, while maintaining peak performance and efficiency.
For instance, a naive implementation depended on the order of the tracks which is nondeterministic when they are created in parallel.
Still, due to non-associative floating point arithmetic a direct binary comparison of the CPU and the GPU tracker output is impossible.
Thus, the approach chosen for evaluating the GPU tracker efficiency is to compare the cluster to track assignment of the CPU and the GPU tracker cluster by cluster.

With the above comparison scheme, the output of the CPU and the GPU tracker differ by 0.00024%.
The GPU tracker outperforms its CPU analog by a factor of three.
Recently, the ALICE HLT cluster was upgraded with new GPUs and will be able to process central heavy ion events at a rate of approximately 200 Hz.
The tracking algorithm together with the necessary modifications, a performance comparison of the CPU and the GPU version, and QA plots will be presented.

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Poster Session / 410

Distributed monitoring infrastructure for Worldwide LHC Computing Grid

Author: Wojciech Lapka[1]

Co-authors: Christos Triantafyllidis [2]; David Collados Polidura [1]; Emir Imamagic [3]; Kislay Bhatt [4]; Marian Babik [1]; Paloma Fuente Fernandez [1]; Pedro Manuel Rodrigues De Sousa Andrade [1]; Phool Chand [4]; Pradyumna Joshi [4]; Rajesh Kalmady [4]; Robert Quick [5]; Scott Werner Teige [5]; Soichi Hayashi [5]; Urvashi Karnani [4]; Vaibhav Kumar [4]; vibhuti duggal [4]

[1] *CERN*
[2] *AUTH*
[3] *SRCE*
[4] *BARC*
[5] *Indiana University*

Corresponding Authors: wojciech.lapka@cern.ch, david.collados@cern.ch

The journey of a monitoring probe from its development phase to the moment its execution result is presented in an availability report is a complex process. It goes through multiple phases such as development, testing, integration, release, deployment, execution, data aggregation, computation, and reporting. Further, it involves people with different roles (developers, site managers, VO managers, service managers, management), from different middleware providers (ARC, dCache, gLite, UNICORE and VDT), consortiums (WLCG, EMI, EGI, OSG), and operational teams (GOC, OMB, OTAG, CSIRT). The seamless harmonization of these distributed actors is in daily use for monitoring of the WLCG infrastructure.

In this paper we describe the monitoring of the WLCG infrastructure from the operational perspective. We explain the complexity of the journey of a monitoring probe from its execution on a grid

node to the visualization on the MyWLCG portal where it is exposed to other clients. This monitoring workflow profits from the interoperability established between the SAM and RSV frameworks. We show how these two distributed structures are capable of uniting technologies and hiding the complexity around them, making them easy to be used by the community. Finally, the different supported deployment strategies, tailored not only for monitoring the entire infrastructure but also for monitoring sites and virtual organizations, are presented and the associated operational benefits highlighted.

Online Computing / 411

A Final Review of the Performance of the CDF Run II Data Acquisition System

Author: William Badgett[1]

[1] *Fermilab*

Corresponding Author: william.badgett@cern.ch

The CDF Collider Detector at Fermilab ceased data collection on September 30, 2011 after over twenty five years of operation. We review the performance of the CDF Run II data acquisition systems over the last ten of these years while recording nearly 10 fb-1 of proton-antiproton collisions with a high degree of efficiency. Technology choices in the online control and configuration systems and front-end embedded processing have impacted the efficiency and quality of the data accumulated by CDF, and have had to perform over a large range of instantaneous luminosity values and trigger rates. We identify significant sources of problems and successes. In particular, we present our experience computing and acquiring data in a radiation environment, and attempt to correlate system technical faults with radiation dose rate and technology choices.

Software Engineering, Data Stores and Databases / 412

Experiences with Software Quality Metrics in the EMI Middleware

Author: Maria Alandes Pradillo[1]

Co-authors: Duarte Bacelar De Begonha De Meneses [2]; Eamonn Kenny [3]; Gianni Pucciani [1]

[1] *CERN*

[2] *LIP Laboratorio de Instrumentacao e Fisica Experimental de Part*

[3] *Unknown*

Corresponding Author: maria.alandes.pradillo@cern.ch

The EMI Quality Model has been created to define, and later review, the EMI (European Middleware Initiative) software product and process quality. A quality model is based on a set of software quality metrics and helps to set clear and measurable quality goals for software products and processes. The EMI Quality Model follows the ISO/IEC 9126 Software Engineering –Product Quality to identify a set of characteristics that need to be present in the EMI software. For each software characteristic, such as portability, maintainability, compliance, etc, a set of associated metrics and KPIs (Key Performance Indicators) are identified.

This article presents how the EMI Quality Model and the EMI Metrics have been defined in the context of the software quality assurance activities carried out in EMI. It also describes the measurement plan and presents some of the metrics reports that have been produced for the EMI releases and updates. It also covers which tools and techniques can be used by any software project to extract "code

metrics"on the status of the software products and "process metrics"related to the quality of the development and support process such as reaction time to critical bugs, requirements tracking and delays in product releases.

Poster Session / 413

Why Are Common Quality and Development Policies Needed?

Author: Maria Alandes Pradillo[1]

Co-author: Lorenzo Dini [1]

[1] *CERN*

Corresponding Author: maria.alandes.pradillo@cern.ch

The EMI project is based on the collaboration of four major middleware projects in Europe, all already developing middleware products and having their pre-existing strategies for developing, releasing and controlling their software artefacts. In total, the EMI project is made up of about thirty development individual teams, called "Product Teams"in EMI. A Product Team is responsible for the entire lifecycle of specific products or small groups of tightly coupled products, including the development of test-suites to be peer reviewed within the overall certification process.

The Quality Assurance in EMI (European Middleware Initiative), as requested by the grid infrastructures and the EU funding agency, must support the teams in providing uniform releases and interoperable middleware distributions, with a common degree of verification and validation of the software and with metrics and objective criteria to compare product quality and evolution over time. In order to achieve these goals the QA team in EMI has defined and now it monitors the development work and release with a set of comprehensive policies covering all aspects of a software project such as packaging, configuration, documentation, certification, release management and testing.

This contribution will present with practical and useful examples the achievements, problems encountered and lessons learned in the definition, implementation and review of Quality Assurance and Development policies. It also describes how these policies have been implemented in the EMI project including the benefits and difficulties encountered by the developers in the project. The main value of this contribution is that all the policies explained are not depending on EMI or grid environments and can be used by any software project.

Poster Session / 414

The Detector Control System of the ATLAS experiment

Author: Sebastien Franz[1]

Co-author: Kerstin Lantzsch [2]

[1] *CERN*

[2] *Bergische Universitaet Wuppertal (DE)*

Corresponding Authors: kerstin.lantzsch@cern.ch, sebastien.franz@cern.ch

The ATLAS experiment is one of the multi-purpose experiments at the Large Hadron Collider (LHC), constructed to study elementary particle interactions in collisions of high-energy proton beams. Twelve different sub-detectors as well as the common experimental infrastructure are supervised by the Detector Control System (DCS). The DCS enables equipment supervision of all ATLAS sub-detectors by using a system of 140 server machines running the industrial SCADA product PVSS.

This highly distributed system reads, processes and archives of the order of 10^6 operational parameters. Higher level control system layers based on the CERN JCOP framework allow for automatic control procedures, efficient error recognition and handling, manage the communication with external control systems such as the LHC controls, and provide a synchronization mechanism with the ATLAS physics data acquisition system. A condition database is used to store the online parameters of the experiment and a subset of the parameters is replicated to an offline database for off-site access and as part of the physics data reconstruction. A configuration database is used to ease the mass parameterization of the detector. This contribution first describes the computing architecture which has been build and the software tools which are used to handle this complex and highly interconnected control system. Secondly, the experience gained during the first operation period of the LHC is given. And finally, the ongoing studies for future upgrades and the usage of new technology standards are presented.

Poster Session / **415**

Tape write efficiency improvements in CASTOR

Author: Steven Murray[1]

Co-authors: Eric Cano [1]; German Cancio Melia [1]; Giuseppe Lo Presti [1]; Giuseppe Lo Re [1]; Sebastien Ponce [1]; Victor Kotlyar [2]; Vlado Bahyl [1]

[1] *CERN*

[2] *Institute for High Energy Physics (RU)*

Corresponding Author: steven.murray@cern.ch

The CERN Advanced STORage manager (CASTOR) is used to archive to tape the physics data of past and present physics experiments. Data is migrated (repacked) from older, lower density tapes to newer, high-density tapes approximately every two years to follow the evolution of tape technologies and to keep the volume occupied by the tape cartridges relatively stable. Improving the performance of writing files smaller than 2G to tape is essential in order to keep the time needed to repack all of the tape resident data within a period of no more than 1 year. Until now CASTOR has flushed the write buffers of the underlying tape-system 3 times per user-file, using up to 7 seconds. With current drive-writing speeds reaching over 240MB/s per second, 7 seconds of flush-time equates to an approximate loss of 1.5 GB of data transfer time per user-file. This paper reports on the solution to writing efficiently to tape that is currently in its early deployment phases at CERN. Write speeds have been increased whilst preserving the existing tape-format by using immediate (non-flushing) tapemarks to write multiple user-files before flushing the tape-system write-buffers. The solution has been realized as a set of incremental upgrades to minimize risk, maximize backwards compatibility and work safely with the legacy modules of CASTOR. Unit testing has been used to help reduce the risk of working with legacy code. This solution will enable CASTOR to continue to be a long-term and performant tool for archiving past and present experiment data to tape.

Summary:

The CERN Advanced STORage manager (CASTOR) is used to archive to tape the physics data of past and present physics experiments. For reasons of physical storage space, all of the tape resident data in CASTOR are repacked onto higher density tapes approximately every two years. Improving the performance of writing files smaller than 2G to tape is essential in order to keep the time needed to repack all of the tape resident data within a period of no more than 1 year. This paper reports on the solution to writing efficiently to tape that is currently in its early deployment phases at CERN.

Poster Session / **416**

Elastic Testbed at CERN for the Integration of the EMI Middleware

Author: Tomasz Wolak[1]

Co-author: Andrew Elwell [1]

[1] *CERN*

Corresponding Author: tomasz.wolak@cern.ch

The development and distribution of Grid middleware software projects, as large, complex, distributed systems require a sizeable computing infrastructure for each stage of the software process: for instance pools of machines for building, and testing on several platforms. Software testing and the possibility of implementing realistic scenarios for the verification of grid middleware are a crucial part of the testing process. System integration testing of a large number of components requires a large dedicated testing infrastructure installed and ready to host such tests. In the grid community such testing environment is described as a "grid integration testbed". It is a dedicated grid infrastructure having similar organization, in smaller scale, as production installations where inter-component tests can be executed on different versions and platforms.

This contribution presents the implementation, based on elastic virtualized resources, of the grid testbed provided by the Grid Technologies group at CERN in order to support the developers of the DPM, FTS, LFC teams and the part of the EMI integration testbed hosted at CERN. The implementation of the EMI testbed also provides the integration of the Nagios monitoring probes of the installed services and supports several platforms such as Scientific Linux and Debian for 32 and 64 bits architectures. We will also present the lessons learned and the experience gained during the migration from the Linux Xen virtualization platform, used for gLite, to Microsoft Hyper V currently used for the EMI testbed.

Poster Session / 418

A distributed agent based framework for high-performance data transfers

Author: Ramiro Voicu[1]

Co-authors: Artur Jerzy Barczyk [1]; Azher Mughal [2]; Harvey Newman [1]; Iosif Legrand [1]; sandor Rozsa [2]

[1] *California Institute of Technology (US)*

[2] *California Institute of Technology (CALTECH)*

Corresponding Author: ramiro.voicu@cern.ch

Current network technologies like dynamic network circuits and emerging protocols like OpenFlow, enable the network as an active component in the context of data transfers.

We present framework which provides a simple interface for scientists to move data between sites over Wide Area Network with bandwidth guarantees. Although the system hides the complexity from the end users, it was designed to include all the security, redundancy and fail-over aspects of large distributed systems. The agents collaborate between them over secure channels to advertise their presence, request and allocate network resources like Dynamic Network Circuits, end-host routing tables and IP addresses, when a user requests a data transfer and release the resources when a transfer finishes. The data transfer tool used by the framework is Fast Data Transfer (http://fdt.cern.ch/). It provides the dynamic bandwidth adjustments capabilities also at application level, so bandwidth scheduling can be used where network circuits are not available.

The framework is currently being deployed and tested between a set of US HEP sites, part of the NSF-funded DYNES project: Dynamic Network System (http://internet2.edu/dynes). The DYNES "cyber-instrument"interconnects ˜40 institutes participating in both US-Atlas and US-CMS collaborations.

Student? Enter 'yes'. See http://goo.gl/MVv53:

No

Poster Session / 419

SYNCAT - Storage Catalogue Consistency

Authors: Fabrizio Furano[1]; Michele Dibenedetto[None]; Paul Millar[2]; Riccardo Zappi[3]

[1] *CERN*

[2] *Deutsches Elektronen-Synchrotron (DE)*

[3] *INFN*

Corresponding Author: fabrizio.furano@cern.ch

Born in the context of EMI (European Middleware Initiative), the SYNCAT project considers as its main purpose the incremental reduction of the divergence of the content of remote file catalogues, like the ones represented by LFC, the Grid Storage Elements and the experiments' private databases. Aiming at giving ways for these remote systems to interact transparently in order to keep their file metadata synchronized, the SYNCAT project is a step towards improved coherence by coupling heterogeneous catalogues and Storage Elements in the Grid infrastructure.
Using standard messaging tools, a set of core libraries called SEMsg has been produced and integrated in a working prototype.
SEMsg has been integrated with LFC and DPM, in addition it was connected to the LHCb framework. Currently we are working on the integration with other EMI storage elements. Deployment and packaging issues have been addressed and the components are now available for evaluation.

Poster Session / 420

The ATLAS LFC consolidation

Author: Fabrizio Furano[1]

Co-authors: Cedric Serfon [2]; Luca Canali [1]; Marcin Blaszczyk [1]; Simone Campana [1]; Stewart Graeme ; Vincent Garonne [1]

[1] *CERN*

[2] *Ludwig-Maximilians-Univ. Muenchen (DE)*

Corresponding Author: fabrizio.furano@cern.ch

ATLAS decided to move from a globally distributed file catalogue to a central instance at CERN.
This talk describes the ATLAS LFC merge exercise from the analysis phase over the prototyping and stress testing to the final execution phase.
We demonstrate that with careful preparation even major architectural changes could be implemented while minimizing the impact on the experiments production and analysis operations.
Merging these large catalogues by processing partially inconsistent metadata for many tens of millions of files and replicas posed several challenges.
We show how the new LFC instance was stress tested to ensure it met ATLAS's requirements and how the LFC schema evolved to support this. We also describe the main reasons why the merging process had to be done with a specialized multithreaded application. This was developed in order to accommodate the peculiarities that make this process much more challenging than a mere movement of data between SQL database instances. The process has to take into account a number of situations where the metadata records clash, or contain errors that have to be fixed on the fly, while still guaranteeing a high level of performance.

Poster Session / 421

Preparing for long-term data preservation and access in CMS

Author: Kati Lassila-Perini[1]

Co-authors: David Colling [2]; Elizabeth Sexton-Kennedy [3]; Jesus Marco [4]; Lucas Taylor [3]; Roberto Tenchini [5]; Sudhir Malik [6]

[1] *Helsinki Institute of Physics (FI)*

[2] *Imperial College Sci., Tech. & Med. (GB)*

[3] *Fermi National Accelerator Lab. (US)*

[4] *Universidad de Cantabria (ES)*

[5] *Sezione di Pisa (IT)*

[6] *University of Nebraska-Lincoln*

Corresponding Author: katri.lassila-perini@cern.ch

The data collected by the LHC experiments are unique and present an opportunity and a challenge for a long-term preservation and re-use. The CMS experiment is defining a policy for the data preservation and access to its data and is starting the implementation of the policy. This note describes the driving principles of the policy and summarises the actions and activities which are planned for its implementation.

Software Engineering, Data Stores and Databases / 422

ATLAS DDM/DQ2 & NoSQL databases: Use cases and experiences

Author: Collaboration Atlas[1]

Co-authors: Angelos Molfetas [2]; Gancho Dimitrov [3]; Graeme Andrew Stewart [2]; Luca Canali [2]; Mario Lassnig [2]; Martin Barisits [4]; Vincent Garonne [2]

[1] *Atlas*

[2] *CERN*

[3] *Brookhaven National Laboratory (US)*

[4] *Vienna University of Technology (AT)*

Corresponding Author: mario.lassnig@cern.ch

The Distributed Data Management System DQ2 is responsible for the global management of petabytes of ATLAS physics data. DQ2 has a critical dependency on Relational Database Management Systems (RDBMS), like Oracle, as RDBMS are well suited to enforce data integrity in online transaction processing application. Despite these advantages, concerns have been raised recently on the scalability of data warehouse-like workload against the relational schema, in particular for the analysis of archived data or the
aggregation of data for summary purposes. Therefore, we have considered new approaches of handling very large amount of data. More specifically, we investigated a new class of database technologies commonly referred to as
NoSQL databases. This includes distributed file system like HDFS that support parallel execution of computational tasks on distributed data, as well as schema-less approaches via key-value/document stores, like HBase, Cassandra or MongoDB. These databases provide solutions to particular types of problems: for example, NoSQL databases have demonstrated horizontal scalability, high throughput, automatic fail-over mechanisms, and provide easy replication support over LAN and WAN.
In this talk, we will describe our use cases in ATLAS, and share our experiences with NoSQL databases in a comparative study with Oracle.

Poster Session / 423

ATLAS software packaging

Author: Collaboration Atlas[1]

Co-author: Grigori Rybkin [2]

[1] *Atlas*

[2] *Universite de Paris-Sud 11 (FR)*

Corresponding Author: grigori.rybkine@cern.ch

Software packaging is indispensable part of build and prerequisite for deployment processes. Full ATLAS software stack consists of TDAQ, HLT, and Offline software. These software groups depend on some 80 external software packages. We present tools, package PackDist, developed and used to package all this software except for TDAQ project. PackDist is based on and driven by CMT, ATLAS software configuration and build tool, and consists of shell and Python scripts. The packaging unit used is CMT project. Each CMT project is packaged as several packages - platform dependent (one per platform available), source code excluding header files, other platform independent files, documentation, and debug information packages (the last two being built optionally). Packaging can be done recursively to package all the dependencies. The whole set of packages for one software release, distribution kit, also includes configuration packages and contains some 120 packages for one
platform. Also packaged are physics analysis projects (currently 6) used by particular physics groups on top of the full release. The tools provide an installation test for the full distribution kit. Packaging is done in two formats for use with the Pacman and RPM package managers. The tools are functional on the platforms supported by ATLAS - GNU/Linux and Mac OS X. The packaged software is used for software deployment on all ATLAS computing resources from the detector and trigger computing farms, collaboration laboratories computing centres, grid sites, to physicist laptops, and CERN VMFS and covers the use cases of running all applications as well as of software development.

Poster Session / 424

Simulating the ATLAS Distributed Data Management System

Author: Collaboration Atlas[1]

Co-authors: Angelos Molfetas [2]; Mario Lassnig [2]; Martin Barisits [3]; Vincent Garonne [2]

[1] *Atlas*

[2] *CERN*

[3] *Vienna University of Technology (AT)*

Corresponding Author: martin.barisits@cern.ch

The ATLAS Distributed Data Management system stores more than 75PB of physics data across 100 sites globally. Over 8 million files are transferred daily with strongly varying usage patterns. For performance and scalability reasons it is imperative to adapt and improve the data management system continuously. Therefore future system modifications in hardware, software as well as policy, need to be evaluated to accomplish good results and avoid unwanted side effects. Due to the complexity of large-scale distributed systems this evaluation process is primarily based on expert-knowledge, as conventional evaluation
methods are inadequate. However, this error-prone process lacks quantitative estimations and leads to inaccuracy as well as incorrect evaluations.
In this work we present a novel, full-scale simulation framework. This flow-level based simulator is able to accurately model the ATLAS Distributed Data Management system. The design and architecture of the component-based software is presented and discussed. The evaluation concentrates on the accuracy and scalability of the simulation framework. Finally, selected use-cases where simulation could be hugely beneficial to distributed data management systems are presented and discussed.

Poster Session / **425**

Accounting the ATLAS DDM system – A case study with Oracle, MongoDB and HBase

Author: Collaboration Atlas[1]

Co-authors: Gancho Dimitrov [2]; Lisa Azzurra Chinzer [3]; Luca Canali [4]; Mario Lassnig [4]; Vincent Garonne [4]

[1] *Atlas*

[2] *Brookhaven National Laboratory (US)*

[3] *Universita e INFN (IT)*

[4] *CERN*

Corresponding Author: mario.lassnig@cern.ch

The ATLAS Distributed Data Management system requires accounting of its contents at the metadata layer. This presents a hard problem
due to the large scale of the system and the high rate of concurrent modifications of data. The system must efficiently account more than 80PB of disk and tape that store upwards of
500 million files across 100 sites globally.
In this work a generic accounting system is presented, which is able to scale to the requirements of ATLAS. The design and architecture is presented, and three implementations are discussed, the reference
implementation in Oracle RAC, and two alternative implementations in MongoDB and HBase. A strong emphasis is placed on the necessary design choices such that the underlying data models are generally applicable to many kinds of accounting, reporting and monitoring. The evaluation then focuses on principal architectural differences,
read-insert-update-delete performance, support for concurrent operations, deployment and operational effort, and possible means to
calculate the actual accounting values based on metadata critera. Finally, a recommendation is presented for the applicability of each
implementation under different accounting use cases, as well as an overall recommendation for useful and required data models.

Poster Session / **426**

ATLAS off-Grid sites (Tier 3) monitoring. From local fabric monitoring to global overview of the VO computing activities

Author: Collaboration Atlas[1]

Co-authors: Artem Petrosyan [2]; Danila Oleynik [2]; Ivan Kadochnikov [2]; Julia Andreeva [3]; Sergey Belov [4]

[1] *Atlas*

[2] *Joint Inst. for Nuclear Research (RU)*

[3] *CERN*

[4] *Joint Inst. for Nuclear Research (JINR)*

Corresponding Author: danila.oleynik@cern.ch

The ATLAS Distributed Computing activities have so far concentrated in the “central” part of the experiment computing system, namely the first 3 tiers (the CERN Tier0, 10 Tier1 centers and over 60 Tier2 sites). Many ATLAS Institutes and National Communities have deployed (or intend to) deploy Tier-3 facilities. Tier-3 centers consist of non-pledged resources, which are usually dedicated to data analysis tasks by the geographically close or local scientific groups, and which usually comprise a range of architectures without Grid middleware. Therefore a substantial part of the ATLAS monitoring tools which make use of Grid middleware, cannot be used for a large fraction of Tier3 sites.

The presentation will describe the T3mon project, which aims to develop a software suite for monitoring the Tier3 sites, both from the perspective of the local site administrator and that of the ATLAS VO, thereby enabling the global view of the contribution from Tier3 sites to the ATLAS computing activities.
Special attention in presentation will be paid generic monitoring solutions for PROOF and xrootd, covering monitoring components which collect, store and visualise monitoring data. One of the popular solutions for local data analysis is the PROOF-based computing facility with a simple storage system based on xrootd protocol. Monitoring of user activities at the PROOF-based computing facility as well as data access and data movement with xrootd is useful, both on the local and global VO level.
The proposed PROOF and xrootd monitoring systems can be deployed as a part of the T3mon monitoring suite or separately as standalone components and can easily be integrated in the global VO tools for monitoring data movement, data access or job processing.

Poster Session / **427**

Dynamic federations: storage aggregation using open tools and protocols

Authors: Fabrizio Furano[1]; Patrick Fuhrmann[2]; Ricardo Brito Da Rocha[1]

[1] *CERN*

[2] *DESY*

Corresponding Author: fabrizio.furano@cern.ch

A number of storage elements now offer standard protocol interfaces like NFS 4.1/pNFS and WebDAV, for access to their data repositories, in line with the standardization effort of the European Middleware Initiative (EMI). Here we report on work which seeks to exploit the federation potential of these protocols and build a system which offers a unique view of the storage ensemble and the possibility of integration of other compatible resources such as those from cloud providers.
The challenge, here undertaken by the providers of dCache and DPM, but pragmatically open to other Grid and Cloud storage solutions, is to build such a system while being able to accommodate name translations from existing catalogues (e.g. LFCs), experiment-based metadata catalogues, or stateless algorithmic name translations, also known as "trivial file catalogues".
Such so-called storage federations of standard protocols-based storage elements will give a unique view of their content, thus promoting simplicity in accessing the data they contain and offering new possibilities for resilience and data placement strategies.
The goal is to consider HTTP and NFS4.1-based storage elements and make them able to cooperate through an architecture that properly feeds the redirection mechanisms that they are based upon, thus giving the functionalities of a "loosely coupled"storage federation. One of the key requirements is to use standard clients (provided by OS'es or open source distributions, e.g. Web browsers) to access an already aggregated system; this approach is quite different from aggregating the repositories at the client side through some wrapper API, like for instance GFAL, or by developing new custom clients.
Other technical challenges that will determine the success of this initiative include performance, latency and scalability, and the ability to create worldwide storage federations that are able to redirect clients to repositories that they can efficiently access, for instance trying to choose the endpoints that are closer or applying other criteria.
We believe that the features of a loosely coupled federation of open-protocols-based storage elements will open many possibilities of evolving the current computing models without disrupting them, and, at the same time, will be able to operate with the existing infrastructures, follow their evolution path and add storage centers that can be acquired as a third-party service.

Poster Session / **428**

Refurbishing the CERN fabric management system

Authors: Gavin Mccance[1]; Steve Traylen[1]

[1] *CERN*

Corresponding Author: gavin.mccance@cern.ch

The CERN Computer Centre is reviewing strategies for optimizing the use of the existing infrastructure in the future. There have been significant developments in the area of computer centre and configuration management tools over the last few years. CERN is examining how these modern, widely-used tools can improve the way in which we manage the centre, with a view to reducing the overall operational effort, increasing agility and automating as much as possible. This presentation will focus on the current status of deployment of the new configuration toolset, the reasons why specific tools were chosen, and give an outlook on how we plan to manage the change to the new system.

Event Processing / **429**

Medical imaging inspired vertex reconstruction at the large hadron collider

Authors: Eckhard Von Toerne[1]; Harris Kagan[2]; Stephan G. Hageboeck[3]

[1] *Universitaet Bonn (DE)*

[2] *Ohio State University*

[3] *University of Bonn*

Corresponding Author: hageboeck@physik.uni-bonn.de

Three dimensional image reconstruction in medical imaging applies sophisticated filter algorithms to linear trajectories of coincident photon pairs in PET. The goal is to reconstruct an image of a source density distribution.
In a similar manner, tracks in particle physics originate from vertices that need to be distinguished from background track combinations.
We investigate if methods from medical imaging can be applied to vertex reconstruction in high energy proton collisions at the large hadron collider.
Therefore, a new method of vertex finding has been developed based on a three dimensional filtered backprojection algorithm. It has been compared to standard vertex reconstruction methods of the RAVE package.
We tested the performance of the vertex finding algorithms as a function of instantaneous luminosity using simulated LHC collisions. Tracks in these collisions resemble a simplified detector model that simulates the tracking performance of e.g. ATLAS or CMS.
Our talk discusses the similarities between medical image reconstruction and vertexing and presents results with both standard and medical imaging inspired finders. We discuss the similarities and differences of our method compared to David J. Jackson's ZV-Top algorithm.

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Poster Session / **430**

IFIC-Valencia Analysis Facility

Author: Miguel Villaplana Perez[1]

Co-authors: Alvaro Fernandez Casani [1]; Collaboration Atlas [2]; Elena Oliver Garcia [1]; Javier Sanchez [3]; Jose Salt [4]; Mohammed KACI [5]; Sanchez Martinez Victoria [6]; Santiago Gonzalez De La Hoz [6]

[1] *Universidad de Valencia (ES)*

[2] *Atlas*

[3] *Consejo Superior de Investigaciones Cientificas (CSIC)-Universi*

[4] *Instituto de Fisica Corpuscular (IFIC) - Universidad de Valencia*

[5] *IFIC - Valencia - Spain*

[6] *IFIC-Valencia*

Corresponding Author: miguel.villaplana.perez@cern.ch

The ATLAS Tier3 at IFIC-Valencia is attached to a Tier2 that has 50% of the Spanish Federated Tier2 resources. In its design, the Tier3 includes a GRID-aware part that shares some of the features of Valencia's Tier2 such
as using Lustre as a file system. ATLAS users, 70% of IFIC's users, also have the possibility of analysing data with a PROOF farm and storing them locally.
In this contribution we discuss the design of the analysis facility as well as the monitoring tools we use to control and improve its performance. We also comment on how the recent changes in the ATLAS computing GRID model affect IFIC. Finally, how this complex system can coexist with the other science
applications running at IFIC (non-ATLAS users) is presented.

Event Processing / 431

The FairRoot framework

Authors: Denis Bertini[1]; Dmytro Kresan[2]; Mohammad Al-Turany[2]; Peter Malzacher[3]; Radek Karabowicz[2]; florian Uhlig[2]

[1] *GSI Darmstadt*

[2] *GSI*

[3] *GSI - Helmholtzzentrum fur Schwerionenforschung GmbH (DE)*

Corresponding Authors: f.uhlig@gsi.de, mohammad.al-turany@cern.ch

The FairRoot framework is an object oriented simulation, reconstruction and data analysis framework based on ROOT. It includes core services for detector simulation and offline analysis. The project started as a software framework for the CBM experiment at GSI, and later became the standard software for simulation, reconstruction and analysis for CBM, PANDA, R3B and ASYEOS at GSI/FAIR, as well as the MPD (NICA) at JINR, Russia. Technical design reports, detector studies and physics performance studies are carried out for FAIR experiments based on the FairRoot services. The framework delivers base classes which enable the users to construct their detectors and /or analysis tasks in a simple way, it also delivers some general functionality like track visualization. Parameter handling and data base connections are also handled by the framework. Beside of the the traditional services of an event processing framework, FairRoot deliver also the possibility to run some Tasks on GPU through FairCuda interface. A CMake-CDash building and monitoring system is also part of the FairRoot services. Time ordered simulations are meanwhile possible with FairRoot. In this contribution, the capabilities of the framework and usage of the different services by the experiments will be presented.

Poster Session / 433

Experience of using the Chirp distributed file system in ATLAS

Author: Collaboration Atlas[1]

Co-author: Rodney Walker [2]

[1] *Atlas*

[2] *Ludwig-Maximilians-Univ. Muenchen (DE)*

Chirp is a distributed file system specifically designed for the wide area network, and developed by the University of Notre Dame CCL group. We describe the design features making it particularly suited to the Grid environment,
and to ATLAS use cases. The deployment and usage within ATLAS distributed computing are discussed, together with scaling tests and evaluation for the various use cases.

Online Computing / 434

Prototyping a 10Gigabit-Ethernet Event-Builder for a Cherenkov Telescope Array

Authors: Dirk Hoffmann[1]; Julien Houles[2]

Co-author: for the CTA consortium [3]

[1] *Universite d'Aix - Marseille II (FR)*

[2] *Centre de Physique des Particules de Marseille*

[3] *CTA*

Corresponding Author: dirk.hoffmann@cern.ch

We present the prototyping of a 10Gigabit-Ethernet based UDP data acquisition (DAQ) system that has been conceived in the context of the Array and Control group of CTA (Cherenkov Telescope Array). The CTA consortium plans to build the next generation ground-based gamma-ray instrument, with approximately 100 telescopes of at least three different sizes installed on two sites. The genuine camera dataflow amounts to 1.2 GByte/s per camera. We have conceived and built a prototype of a front-end event builder DAQ able to receive and compute such a data rate, allowing a more sustainable level for the central data logging of the site by data reduction. We took into account characteristics and constraints of several camera electronics projects in CTA, thus keeping a generic approach to all front-end types. The big number of telescopes and the remoteness of the array sites imply that any front-end element must be robust and self-healing to a large extent. The main difficulty is to combine very high performances with a good reliability and rude environmental conditions. We will present the iterations we made to maximize the performances and the results we obtained with hundreds of IP nodes connected to a switch simulating the camera elements.

Summary:

We present the prototyping of a 10Gigabit-Ethernet based UDP data acquisition (DAQ) system that has been conceived in the context of the Array and Control group of CTA (Cherenkov Telescope Array). The CTA consortium plans to build the next generation ground-based gamma-ray instrument, with approximately 100 telescopes of at least three different sizes installed on two sites. The genuine camera dataflow amounts to 1.2 GByte/s per camera. We have conceived and built a prototype of a front-end event builder DAQ able to receive and compute such a data rate, allowing a more sustainable level for the central data logging of the site by data reduction. We took into account characteristics and constraints of several camera electronics projects in CTA, thus keeping a generic approach to all front-end types. The big number of telescopes and the remoteness of the array sites imply that any front-end element must be robust and self-healing to a large extent. The main difficulty is to combine very high performances with a good reliability and rude environmental conditions. We will present the iterations we made to maximize the performances and the results we obtained with hundreds of IP nodes connected to a switch simulating the camera elements.

Poster Session / 435

Grid Information Systems Revisited

Author: Laurence Field[1]

Co-author: Lorenzo Dini [1]

[1] *CERN*

Corresponding Author: laurence.field@cern.ch

The primary goal of a Grid information system is to display the current composition and state of a Grid infrastructure. It's purpose is to provide the information required for workload and data management. As these models evolve, the information system requirements need to be revisited and revised. This paper first documents the results from a recent survey of LHC VOs on the information system requirements. An evaluation of how well these requirements are met by the current system is conducted and directions for future improvements are suggested. It is shown that due to the changing computing models, predominately the adoption of the pilot job paradigm, the main focus for the information system has shifted from scheduling towards service discovery and service monitoring. Six use cases are identified and directions for improved support for these are presented. The paper concludes by suggesting changes to existing system that will provide improved support while maintaining continuity of the service.

Distributed Processing and Analysis on Grids and Clouds / 436

Next generation WLCG File Transfer Service (FTS)

Author: Zsolt Molnár[1]

Co-authors: Jean-Philippe Baud [1]; Michail Salichos [1]; Oliver Keeble [1]

[1] *CERN*

Corresponding Author: zsolt.molnar@cern.ch

LHC experiments at CERN and worldwide utilize WLCG resources and middleware components to perform distributed computing tasks. One of the most important tasks is reliable file replication. It is a complex problem, suffering from transfer failures, disconnections, transfer duplication, server and network overload, differences in storage systems, etc. To address these problems, EMI and gLite have provided the independent File Transfer Service (FTS) and Grid File Access Library (GFAL) tools. Their development started almost a decade ago, in the meantime, requirements in data management have changed - the old architecture of FTS and GFAL cannot keep support easily these changes. Technology has also been progressing: FTS and GFAL do not fit into the new paradigms (cloud, messaging, for example).

To be able to serve the next stage of LHC data collecting (from 2013), we need a new generation of these tools: FTS 3 and GFAL 2. We envision a service requiring minimal configuration, which can dynamically adapt to the state of its resources (endpoints (SE-s) and network) and which offers an SE-centric configuration model for administrators to use where necessary.

Main problems that we solve: scalability problems of the static channel model, resource and network states are not taken into account when scheduling transfer jobs, supporting multiple database backends (Oracle, MySQL), multiple transfer and control protocols. The new system will be much easier to configure and manage. FTS 3 will provide transparent job submission and monitoring features based on messaging.
We report on design and prototype experience. We provide an overview about how FTS fits into the more general suite of data management utilities provided by the EMI project.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Summary:

We present our plans and prototyping experience about the next generation WLCG FTS and GFAL tools. We address the problems that the actual tools face with: static channel model, configuration and scalability problems, etc. We present the solution we proposed and the design of the new tools as well. We also overview how FTS fits into the more general suite of data management utilities provided by the EMI project.

Poster Session / **437**

Deployment and Operational Experiences with CernVM-FS at the GridKa Tier-1 Center

Author: Andreas Petzold[1]

Co-authors: Axel Jäger [1]; Manfred Alef [1]

[1] *KIT*

Corresponding Author: andreas.petzold@cern.ch

In 2012 the GridKa Tier-1 computing center hosts 130kHEPSPEC06 computing resources and 11PB disk and 17.7PB tape space. These resources are shared between the four LHC VOs and a number of national and international VOs from high energy physics and other sciences. CernVM-FS has been deployed at GridKa to supplement the existing NFS-based system to access VO software on the worker nodes. It provides a solution tailored to the requirement of the LHC VOs.
We will focus on the first operational experiences and the monitoring of CernVM-FS on the worker nodes and the squid caches.

Poster Session / **438**

Performance of Standards-based transfers in WLCG SEs

Author: Sam Skipsey[1]

Co-authors: Christopher John Walker [2]; Graeme Andrew Stewart [3]; Matthew Doidge [4]; Ricardo Brito Da Rocha [3]

[1] *University of Glasgow / GridPP*

[2] *University of London (GB)*

[3] *CERN*

[4] *Lancaster University*

Corresponding Author: samuel.cadellin.skipsey@cern.ch

While, historically, Grid Storage Elements have relied on semi-proprietary protocols for data transfer (gridftp for site-to-site, and (rfio/dcap/other) for local transfers)), the rest of the world has not stood still in providing its own solutions to data access.
dCache, DPM and StoRM all now support access via the widely implemented HTTP/WebDAV standard, and dCache and DPM both support NFS4.1/pNFS, which is partly implemented in newer releases of the linux kernel.
We present results of comparing the performance of these new protocols against the older, more parochial protocols, both on DPM and StoRM systems.
The next-iteration ATLAS data movement tool, Rucio, is used for some of these tests, which include

examination of interoperability between the DPM and StoRM sites, as well as internal transfer performance within each site.

Poster Session / **439**

Coping with the Data Rates and Volumes of the PHENIX Experiment

Author: Martin Purschke[1]

[1] *BROOKHAVEN NATIONAL LABORATORY*

Corresponding Author: purschke@bnl.gov

The PHENIX detector system at the Relativistic Heavy Ion Collider (RHIC) was one of the first experiments getting to "LHC-era" data rates in excess of 500 MB/s of compressed data in 2004. In step with new detectors and increasing event sizes and rates, the data logging capability has grown to about 1500MB/s since then.

We will explain the strategies we employ to cope with the data volumes in the online system and at the analysis stage. We will detail and discuss in the role of our data format, and the success of our concept of managed access to the DST data by way of "analysis taxis".

We will outline the upgrades currently in progress and envisioned on our way to the "sPHENIX" project, expected to begin in 2015, and their impact on our current computing paradigms.

Summary:

The PHENIX Experiment has started to exceed its originally designed data rates in 2004, when we switched to a compressed data format on disk and upgraded to a maximum data rate of 600 MB/s. This was made possible with a distributed compression system, which uses a large number of CPUs in the event builder to compress the data, rather than the logger machines. The switch, which has allowed us to run our Level-2 trigger is "tag-only", rather than in filter mode, has given us access to physics signals which one can not normally trigger on due to the high multiplicity in heavy-ion collisions.

With the increased data rate, we had to implement a managed access to the data. Compared to a traditional staging model, we have achieved an throughput increase for the data analysis by an estimated factor 30.

We will explain the technologies involved in the DAQ and analysis procedures, and give an overview of the strategies we will use to maintain our event rate in spite of the increasing event sizes after a future upgrade.

Poster Session / **440**

EMI_datalib - joining the best of ARC and gLite data libraries

Author: Jon Kerr Nilsen[1]

Co-authors: Adrien Devresse [2]; David Cameron [1]; Michail Salichos [3]; Zsolt Molnar [3]; Zsombor Nagy [4]

[1] *University of Oslo (NO)*

[2] *University of Nancy I (FR)*

[3] *CERN*

[4] *Nemzeti Informacios Infrastruktura Fejlesztesi Intezet*

Corresponding Author: j.k.nilsen@fys.uio.no

To manage data in the grid, with its jungle of protocols and enormous amount of data in different storage solutions, it is important to have a strong, versatile and reliable data management library. While there are several data management tools and libraries available, they all have different strengths and weaknesses, and it can be hard to decide which tool to use for which purpose.

EMI is a collaboration between the European middleware providers aiming to take the best out of each middleware to create one consolidated, all-purpose grid middleware. When EMI started there were two main tools for managing data - gLite had lcg_util and the GFAL library, ARC had the ARC data tools and libarcdata2. While different in design and purpose, they both have the same goal; to manage data in the grid.

The design of the new EMI_datalib was ready by the end of 2011, and a first prototype is now implemented and going through a thorough testing phase. This presentation will give the latest results of the consolidated library together with an overview of the design, test plan and roadmap of EMI_datalib.

Software Engineering, Data Stores and Databases / 441

Parallelization of the AliRoot event reconstruction by performing a semi- automatic source-code transformation

Author: Stefan Lohn[1]

Co-authors: Federico Carminati [2]; Xin Dong [3]

[1] *Universitaet Bonn (DE)*

[2] *CERN*

[3] *Northeastern University*

Corresponding Author: stefan.lohn@cern.ch

Chip multiprocessors are going to support massive parallelism to provide further processing capacities by adding more and more physical and logical cores. Unfortunately the growing number of cores come along with slower advances in speed and size of the main memory, the cache hierarchy, the front-side bus or processor interconnections. Parallelism can only result in performance gain, if the memory usage is optimized, memory locality improved and the communication between threads is minimized. But the domain of concurrent programming has become a field for highly skilled experts, as the implementation of multithreading is difficult, error prone and labor intensive. A full re-implementation for parallel execution of existing offline frameworks, like AliRoot in ALICE, is thus unaffordable. An alternative method, is to use a semi-automatic source-to-source transformation for getting a simple parallel design, with almost no interference between threads. This reduces the need of rewriting the developed software and avoids excessive communication between threads.
This paper evaluates the adaption of the AliRoot event reconstruction, by taking the following steps: introduction of thread-safety to bring the original sequential and thread unaware source-code into the position of using multithreading; identification of classes that can be shared between threads as a method to reduce the memory footprint; transformation of the source code to share these classes; verification of the resulting code to localize and eliminate any remaining interferences between threads.

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Summary:

This work is about the content of my ongoing PhD thesis at the ALICE experiment at CERN to provide methods for a better exploitation of the available and prospective CPUs. It is following the idea to introduce scalable multithreading by using a source-to-source transformation. First results and upcoming problems will be presented.

Computer Facilities, Production Grids and Networking / 442

Monitoring the US ATLAS Network Infrastructure with perfSONAR-PS

Author: Shawn Mc Kee[1]

Co-authors: Andrew Lake [2]; Collaboration Atlas [3]; Horst Severini [4]; Jason Zurawski [5]; Philippe Laurens [6]; Stephen Wolff [5]; Tomasz Wlodek [7]

[1] *University of Michigan (US)*

[2] *ESnet*

[3] *Atlas*

[4] *University of Oklahoma (US)*

[5] *Internet2*

[6] *Michigan State University*

[7] *Brookhaven National Laboratory*

Corresponding Author: shawn.mckee@cern.ch

Global scientific collaborations, such as ATLAS, continue to push the network requirements envelope. Data movement in this collaboration is projected to include the regular exchange of petabytes of datasets between the collection and analysis facilities in the coming years. These requirements place a high emphasis on networks functioning at peak efficiency and availability; the lack thereof could mean critical delays in the overall scientific progress of distributed data-intensive experiments like ATLAS.

Network operations staff routinely must deal with problems deep in the infrastructure; this may be as benign as replacing a failing piece of equipment, or as complex as dealing with a multi-domain path that is experiencing data loss. In either case, it is crucial that effective monitoring and performance analysis tools are available to ease the burden of management.

We will report on our experiences deploying and using the perfSONAR-PS Performance Toolkit at ATLAS sites in the United States. This software creates a dedicated monitoring server, capable of collecting and performing a wide range of passive and active network measurements. Each independent instance is managed locally, but able to federate on a global scale; enabling a full view of the network infrastructure that spans domain boundaries. This information, available through web service interfaces, can easily be retrieved to create customized applications. USATLAS has developed a centralized "dashboard" offering network administrators, users, and decisions makers the ability to see the performance of the network at a glance. The dashboard framework includes the ability to notify users (alarm) when problems are found, thus allowing rapid response to potential problems and making perfSONAR-PS crucial to the operation of our distributed computing infrastructure.

Poster Session / 443

Status of the DIRAC Project

Authors: Adrian Casajus Ramo[1]; Andrei Tsaregorodtsev[2]; Federico Stagni[3]; Matvey Sapunov[2]; Ricardo Graciani Diaz[1]; Vanessa Hamar[4]

[1] *University of Barcelona (ES)*

[2] *Universite d'Aix - Marseille II (FR)*

[3] *CERN*

[4] *CPPM-IN2P3-CNRS, Marseille*

Corresponding Author: atsareg@in2p3.fr

The DIRAC Project was initiated to provide a data processing system for the LHCb Experiment at CERN. It provides all the necessary functionality and performance to satisfy the current and projected future requirements of the LHCb Computing Model. A considerable restructuring of the DIRAC software was undertaken in order to turn it into a general purpose framework for building distributed computing systems that can be used by various user communities in High Energy Physics and other application domains. The ILC Collaboration started to use DIRAC for their data production system. The Belle Collaboration at KEK, Japan, has adopted the Computing Model based on the DIRAC system for its second phase starting in 2015. The CTA Collaboration uses DIRAC for the data analysis tasks. A large number of other experiments are starting to use DIRAC or are evaluating this solution for their data processing tasks. DIRAC services are included as part of the production infrastructure of the GISELA Latin America grid. Similar services are provided for the users of the French segment of the EGI Grid. The new communities using DIRAC started to provide important contributions to its functionality. Among recent additions can be mentioned the support of the Amazon EC2 computing resources; a versatile File Replica Catalog with the File Metadata capabilities; support for running MPI jobs in the pilot based Workload Management System. Integration with existing application Web Portals, like WS-PGRADE, is demonstrated.

In this paper we will describe the current status of the DIRAC Project, recent developments of its framework and functionality as well as the status of the rapidly evolving community of the DIRAC users.

Poster Session / 444

Prototype of a cloud-based Computing Service for ATLAS at PIC Tier1

Author: Collaboration Atlas[1]

Co-authors: Alexey Sedov [2]; Gonzalo Merino Arevalo [3]; Pau Tallada Crespi [4]; Xavier Espinal Curull [2]

[1] *Atlas*

[2] *Universitat Autònoma de Barcelona (ES)*

[3] *Centro de Investigaciones Energ. Medioambientales y Tecn. - (ES*

[4] *Unknown*

We present the prototype deployment of a private cloud at PIC and the tests performed in the context of providing a computing service for ATLAS. The prototype is based on the OpenNebula open source cloud computing solution. The possibility of using CernVM virtual machines as the standard for ATLAS cloud computing is evaluated by deploying a
Panda pilot agent as part of the VM contextualization. Different mechanisms to do this are compared (EC2, OCCI, cvm-tools) on the basis of their suitability for implementing a VM-pilot factory service. Different possibilities to access the Tier1 Storage Service (based in dCache) from the VMs are also tested: dcap, NFS4.1, http. As a
conclusion, the viability of private clouds to be considered as a candidate implementations for Tier1 computing services in the future
is discussed.

Poster Session / 445

VM-based infrastructure for simulating different cluster and storage solutions used on ATLAS Tier-3 sites

Author: Mikalai Kutouski[1]

Co-authors: Artem Petrosyan [2]; Danila Oleynik [2]; Ivan Kadochnikov [2]; Sergey Belov [1]; Vladimir Korenkov [1]

[1] *Joint Inst. for Nuclear Research (JINR)*

[2] *Joint Inst. for Nuclear Research (RU)*

Corresponding Author: mikalai.kutouski@cern.ch

The current ATLAS Tier3 infrastructure consists of a variety of sites of different sizes and with a mix of local resource management systems (LRMS) and mass storage system (MSS) implementations. The Tier3 monitoring suite, having been developed in order to satisfy the needs of Tier3 site administrators and to aggregate Tier3 monitoring information on the global VO level, needs to be validated for various combinations of LRMS and MSS solutions along with the corresponding Ganglia and/or Nagios plugins. For this purpose the Testbed infrastructure, which allows simulation of various computational cluster and storage solutions, had been set up at JINR (Dubna). This infrastructure provides the ability to run testbeds with various LRMS and MSS implementations, and with the capability to quickly redeploy particular testbeds or their components. Performance of specific components is not a critical issue for development and validation, whereas easy management and deployment are crucial. Therefore virtual machines were chosen for implementation of the validation infrastructure which, though initially developed for Tier3 monitoring project, can be exploited for other purposes. Load generators for simulation of the computing activities at the farm were developed as a part of this task. The poster will cover concrete implementation, including deployment scenarios, hypervisor details and load simulators.

Poster Session / 446

Optimising the read-write performance of mass storage systems through the introduction of a fast write cache

Authors: David Colling[1]; Simon William Fayer[1]; Stuart Wakefield[1]

[1] *Imperial College Sci., Tech. & Med. (GB)*

Corresponding Authors: simon.fayer@cern.ch, stuart.wakefield@gmail.com

Reading and writing data onto a disk based high capacity storage system has long been a troublesome task. While disks handle sequential reads and writes well, when they are interleaved performance drops off rapidly due to the time required to move the disk's read-write head(s) to a different position. An obvious solution to this problem is to replace the disks with an alternative storage technology such as solid-state devices which have no such mechanical limitations, however in most applications this is prohibitively expensive. This problem is commonly seen at computer facilities where new data need to be stored while old data are being processed from the same storage, such as WLCG grid sites. In the WLCG case this problem is only going to become more prominent as the LHC luminosity increases, creating larger data-sets.

In this paper we explore the possibility of introducing a fast write cache in-front of the storage system to buffer inbound data. This cache allows writes to be coalesced into larger, more efficient blocks before being committed to the primary storage, while also allowing this action to be postponed until the primary storage is sufficiently quiescent. We demonstrate that this is a viable solution to the problem using a real WLCG site as an example of deployment. Finally we also discuss the steps required to tune the surrounding infrastructure, such as the computer network and storage metadata server in order to sustain high write rates to the cache and allow for the data to be flushed to the bulk storage successfully.

Poster Session / **447**

The ATLAS Computing activities and developments of the Italian Cloud

Author: Collaboration Atlas[1]

Co-authors: Agnese Martini [2]; Alberto Annovi [3]; Alessandra Doria [4]; Alessandro Brunengo [5]; Alessandro De Salvo [6]; Alessandro Di Girolamo [7]; Claudia Ciocca [8]; Dario Barberis [9]; David Rebatto [9]; Elisabetta Vilucchi [3]; Francesco Prelz [10]; Giampaolo Carlino [11]; Guido Russo [9]; Lamberto Luminari [6]; Laura Perini [12]; Leonardo Carminati [13]; Leonardo Merola [14]; Lorenzo Rinaldi [15]; Luca Vaccarossa ; Manoj Jha [9]; Mario Antonelli [3]; Mirko Corosu [16]; Rosario Esposito [17]; Simone Campana [7]; Stefano Barberis [18]; Vincenzo Capone [9]

[1] *Atlas*

[2] *Istituto Nazionale Fisica Nucleare (INFN)*

[3] *Istituto Nazionale Fisica Nucleare (IT)*

[4] *Univ. + INFN*

[5] *Sezione di Genova (INFN)-Universita e INFN*

[6] *Universita e INFN, Roma I (IT)*

[7] *CERN*

[8] *Dipartimento di Fisica-Universita degli Studi di Bologna-Univers*

[9] *Universita e INFN (IT)*

[10] *Sezione di Milano (INFN)-Universita e INFN*

[11] *INFN, Sezione di Napoli-Universita & INFN, Napoli*

[12] *Università degli Studi e INFN Milano (IT)*

[13] *INFN Sezione di Milano (INFN)*

[14] *Unknown*

[15] *INFN CNAF*

[16] *INFN*

[17] *INFN - Napoli*

[18] *INFN Milano*

The large amount of data produced by the ATLAS experiment needs new computing paradigms for data processing and analysis, involving many
Computing Centres spread around the world. The computing workload is managed by regional federations, called Clouds.
The Italian Cloud consists of a main (Tier-1) centre, located in Bologna, four secondary (Tier-2) centres, and a few smaller (Tier-3) sites.
In this contribution we describe the Italian Cloud site facilities and the activities of Data Processing, Analysis, Simulation and Software Development performed within the Cloud, and we discuss the tests of the new Computing Technologies contributing to the ATLAS Computing Model evolution, namely Network Monitoring, Software Installation and use of CVMFS, Virtualization and Multi-Core processing.

Poster Session / **448**

New Developments in the GENFIT track fitting framework

Author: Felix Valentin Böhmer[1]

Co-authors: Christian Höppner [1]; Johannes Rauch [1]; Sebastian Neubert [2]

[1] *Technische Universität München*

[2] *Technical University Munich*

GENFIT is a framework for track fitting in nuclear and particle physics. Its defining feature is the conceptual independence of the specific detector and field geometry, achieved by modular design of the software.

A track in genfit is a collection of detector hits and a collection of track representations.It can contain hits from different detector types (planar hits, space points, isochrones from wire detectors) in their natural coordinates. The track representations define the extrapolation through the detector material and field configuration. Several track representations are available and can be easily exchanged or complemented with new models. Different representations (e.g. for different particle hypotheses) can be fitted simultaneously.

Its application in different collaborations (e.g. PANDA, Belle II, COMPASS) has sparked new developments and improvements. The fitting routines have been upgraded with algorithms for hit smoothing. A Deterministic Annealing Filter (DAF) has been implemented, tested and validated. Due to a new interfacing with the RAVE vertexing framework developed for CMS, GENFIT is now capable of vertex reconstruction and fitting. Furthermore, GENFIT has been complemented with built-in capabilities for event display, allowing direct visual validation of fit results.

Results from simulation tests and physics data will be presented.

Poster Session / **449**

Investigating the performance of CMSSW on the AMD Bulldozer micro-architecture

Authors: David Colling[1]; Simon William Fayer[1]; Stuart Wakefield[1]

[1] *Imperial College Sci., Tech. & Med. (GB)*

Corresponding Authors: simon.fayer@cern.ch, stuart.wakefield@gmail.com

The density of rack-mount computers is continually increasing, allowing for higher performance processing in smaller and smaller spaces. With the introduction of its new Bulldozer micro-architecture, AMD have made it feasible to run up to 128 cores within a 2U rack-mount space. CPUs based on Bulldozer contain a series of modules, each module containing two processing cores which share some resources, while also having dedicated versions of other resources. As the LHC luminosity increases, in turn increasing pile-up, more and more computing power will be needed to reconstruct and analyse each event. In-order to provide this increase in computing power without a large increase in space, higher density computing must be used.

In this paper we explore the possibilities of running the CMS software stack, CMSSW, on one implementation of this new architecture. Initially we look at running traditional single core jobs within the architecture in such a way that it is directly comparable to jobs run on older architectures. We then go on to explore the possibility of multi-core and whole-node processing within CMSSW, which would allow for much better memory utilisation. Using less memory in any large job has the advantage of allowing the CPU to churn less data through the memory caches while analysing data, resulting in better overall performance.

Poster Session / **450**

CMS integrated central monitoring and validation system

Author: Kaori Maeshima[1]

[1] *Fermi National Accelerator Lab. (US)*

Corresponding Author: kaori.maeshima@cern.ch

In operating a complex high energy physics experiment such as CMS, two of the important issues are to record high quality data as efficiently as possible and, correspondingly, to have well validated and certified data in a timely manner for physics analyses. Integrated and user-friendly monitoring systems and coherent information flow play an important role to accomplish this. The CMS integrated central monitoring and validation system (CICMS) is often described separately as two parts: Web Based Monitoring (WBM) and Data Quality Monitoring (DQM). Both are monitoring systems, but information for WBM is typically from non-event sources such as online databases and real-time messaging, while the primary DQM monitoring source is the event data. We discuss here both systems together, focusing on how we track the online operation run time (Run Time Logger), how we propagate the input information necessary for data certification, how we do the book-keeping (Run Registry), and how we visualize the data certification statistics (Data Quality Logger).

Poster Session / 451

DIRAC File Replica and Metadata Catalog

Authors: Andrei Tsaregorodtsev[1]; Stephane Poss[2]

[1] *Universite d'Aix - Marseille II (FR)*

[2] *Unknown*

Corresponding Author: atsareg@in2p3.fr

File replica and metadata catalogs are essential parts of any distributed data management system, which are largely determining its functionality and performance. A new File Catalog (DFC) was developed in the framework of the DIRAC Project that combines both replica and metadata catalog functionality. The DFC design is based on the practical experience with the data management system of the LHCb Collaboration. It is optimized for the most common patterns of the catalog usage in order to achieve maximum performance from the user perspective. The DFC supports bulk operations for replica queries and allows quick analysis of the storage usage globally and for each Storage Element separately. It supports flexible ACL rules with plug-ins for various policies that can be adopted by a particular community. The DFC catalog allows to store various types of metadata associated with the files and directories and to perform efficient queries for the data based on complex metadata combinations. Definition of file ancestor-descendent chains is also possible. It is implemented in the DIRAC distributed computing framework following the standard grid security architecture.
In this contribution we describe the design of the DFC and its implementation details. The performance measurements are compared with other grid file catalog implementations. The experience of the DFC Catalog usage in the ILC Collaboration is discussed.

Poster Session / 452

A hybrid Monte Carlo Generator for Ultra High Energy Cosmic Rays from their Sources to the Observer

Authors: David Walz[1]; Gero Müller[1]; Hans-Peter Bretz[1]; Klaus Dolag[2]; Martin Erdmann[1]; Tobias Winchen[1]

[1] *III. Physikalisches Institut A, RWTH Aachen University, Germany*

[2] *MPI für Astrophysik, Garching, Germany*

Corresponding Author: gero.mueller@physik.rwth-aachen.de

To understand in detail cosmic magnetic fields and sources of Ultra High Energy Cosmic Rays (UHECRs) we have developed a Monte Carlo simulation for galactic and extragalactic propagation.

In our approach we identify three different propagation regimes for UHECRs, the Milky Way, the local universe out to 110 Mpc, and the distant universe.
For deflections caused by the Galactic magnetic field a lensing technique based on matrices is applied which are created from backtracking of antiparticles through Galactic field models.
Propagation in the local universe uses forward tracking through structured magnetic fields extracted from simulations of the large scale structure of the universe.
UHECRs from distant sources are simulated using parameterized models.
Interactions with backround photons are taken into account per simulation step or as continuous energy loss.
In this contribution we present the combination of all three simulation techniques by means of propability maps.
The combined propability maps are used to generate UHECRs for large scale mass production, and to create distributions with realistic arrival directions and energies.
Comparisons with physics analyses of UHECR measurements enables the development of new analysis techniques and constrain parameters of the underlying physics models like the source density and magnetic field strength.

Poster Session / 455

Integration of WS-PGRADE/gUSE portal and DIRAC

Authors: Albert Puig Navarro[1]; Damia Viana Casals[2]

Co-authors: Adrian Casajus Ramo [1]; Ricardo Graciani Diaz [1]

[1] *University of Barcelona (ES)*

[2] *Universitat de Barcelona*

Corresponding Author: albert.puig@epfl.ch

The gUSE (Grid User Support Environment) framework allows to create, store and distribute application workflows. This workflow architecture includes a wide variety of payload execution operations, such as loops, conditional execution of jobs and combination of output. These complex multi-job workflows can easily be created and modified by application developers through the WS-PGRADE portal. The portal also allows end users to download and use existing workflows, as well as executing them.
The DIRAC framework for distributed computing, a complete Grid solution for a community of users needing access to distributed computing resources, has been integrated into the WS-PGRADE/gUSE system. This integration allows the execution of gUSE workflows in a distributed computing environment, thus greatly expanding the capability of the portal to several Grids and Cloud Computing facilities.
The main features and possibilities of the WS-PGRADE/gUSE-DIRAC system, as well as the benefits for users, will be outlined and discussed.

Student? Enter 'yes'. See http://goo.gl/MVv53:

No

Poster Session / 456

iSpy: a powerful and lightweight event display

Author: Thomas Mc Cauley[1]

Co-authors: George Alverson [2]; Giulio Eulisse [1]; Lucas Taylor [1]

[1] *Fermi National Accelerator Lab. (US)*

[2] *Northeastern University (US)*

Corresponding Author: thomas.mccauley@cern.ch

iSpy is a general-purpose event data and detector visualization program that was developed as an event display for the CMS experiment at the LHC and has seen use by the general public and teachers and students in the context of education and outreach.

Central to the iSpy design philosophy is ease of installation, use, and extensibility. The application itself uses the open-access packages Qt4 and Open Inventor and is distributed either as a fully-bound executable or a standard installer package: one can simply download and double-click to begin. Mac OSX, Linux, and Windows are supported. iSpy renders the standard 2D, 3D, and tabular views, and the architecture allows for a generic approach to production of new views and projections.

iSpy reads and displays data in the ig format: event information is written in compressed JSON format files designed for distribution over a network. This format is easily extensible and makes the iSpy client indifferent to the original input data source. The ig format is the one used for release of approved CMS data to the public.

Poster Session / 457

Precision measurements of cosmic shear fields using weak gravitational lensing for dark energy search

Author: Nobu Katayama[1]

[1] *HIGH ENERGY ACCELERATOR RESEARCH ORGANIZATION*

Corresponding Author: nobu.katayama@kek.jp

Dark Energy is one of the most intriguing questions in the field of particle physics and cosmology. We expect the first lignt of Hyper Suprime Cam (HSC) at the Subaru Telescope on top of Mauna Kea in Hawaii island in 2012. HSC will measure the shapes of billions of galaxies precisely to construct the 3D map of the dark matter in the universe, characterizing the properties of dark energy. We will discuss many aspects of the latest data processing pipeline built for HSC and other experiments in the field of observational cosmology.

Computer Facilities, Production Grids and Networking / 458

The Grid Enabled Mass Storage System (GEMMS): the Storage and Data management system used at the INFN Tier1 at CNAF.

Authors: Alessandro Cavalli[1]; Andrea Prosperini[2]; Daniele Gregori[3]; Elisabetta Ronchieri[4]; Luca dell'Agnello[1]; Pier Paolo Ricci[2]; Vincenzo Vagnoni[4]; Vladimir Sapunenko[5]

[1] *INFN-CNAF*

[2] *INFN CNAF*

[3] *Istituto Nazionale di Fisica Nucleare (INFN)*

[4] *Universita e INFN (IT)*

[5] *INFN*

Corresponding Author: pierpaolo.ricci@cnaf.infn.it

The storage solution currently used in production at the INFN Tier-1 at CNAF, is the result of several years of case studies, software development and tests. This solution, called the Grid Enabled

Mass Storage System (GEMSS), is based on a custom integration between a fast and reliable parallel filesystem (IBM GPFS), with a complete integrated tape backend based on TIVOLI TSM Hierarchical storage management (HSM) and the Storage Resource Manager (StoRM), providing access to grid users through a standard SRM interface. Since the start of the operations of the Large Hadron Collider (LHC), all the LHC experiments have been using GEMMS at CNAF for both the fast access to data on disk and the long-term tape archive. Moreover, during the last year, GEMSS has become the standard solution for all the other experiments hosted at CNAF, allowing the definitive consolidation of the data storage layer. Our choice has proved to be successful in the last two years of production with constant enhancements in the software releases, accurate monitoring of the data throughput and effective customizations to the end-user requests.
In this paper a brief description of the system is reported with a particular focus on the new improvements of the code and with detailed overview of the administration and monitoring tools. We also report all the solutions adopted in order to grant the maximum avaliability of the service in case of software and hardware failures and the latest optimization features within the data access process. Finally we include an overall report of the results obtained during the last years of activity from the experiment user perspective, which clearly shows the reliability and the high performance throughput that can be obtained using GEMMS.

Summary:

SUMMARY:
The presentation will focus on the last years improvements in the implementation of the complete Hierarchical Storage Management GEMSS system with a particular attention on the monitoring and optimization methods in the disk/tape access layers. The GEMSS primary components are described in a clear and understandable way with a particular attention in providing information on the interaction between the disk and tape layers. Detailed information will be provided on what we have learned during the last years of activity and all the system fixes and optimization that has been introduced in order to provide a stable, redundant and efficient service.
Also the reports of the last years of LHC and non-LHC experiment activity will be summarized for a clear view of the thoughtput and avaliability level obtained with the definitive choice of GEMSS as the storage solution at the INFN Tier-1 at CNAF.

Collaborative tools / 459

Preparing experiments' software for long term analysis and data preservation (DESY-IT)

Authors: Dmitry Ozerov[1]; Yves Kemp[1]

[1] *Deutsches Elektronen-Synchrotron (DE)*

Corresponding Author: yves.kemp@cern.ch

Preserving data from past experiments and preserving the ability to perform analysis with old data is of growing importance in many domains of science, including High Energy Physics (HEP). A study group on this issue, DPHEP, has been established in this field to provide guidelines and a structure for international collaboration on data preservation projects in HEP. This contribution aims at preparing experiments' software for long term analysis and data preservation.
In a first part, we discuss the use of modern techniques like virtualization or Cloud for this purpose. In a second part, we detail the constraints of a supporting IT center for future legacy experiments. In a third part, we present a framework that allows

experimentalists to validate their software against a previously defined set of tests in an automated way. We show first usage of the system, and present results gained from the experience with early-bird-users, and future adaptations to the system.

Poster Session / **461**

An XML generic detector description system and geometry editor for the ATLAS detector at the LHC

Author: Collaboration Atlas[1]

Co-authors: Andrea Dell'Acqua [2]; Jochen Meyer [3]; Laurent Chevalier [4]

[1] *Atlas*

[2] *CERN*

[3] *Bayerische Julius Max. Universitaet Wuerzburg (DE)*

[4] *CEA - Centre d'Etudes de Saclay (FR)*

Corresponding Author: jochen.meyer@cern.ch

Accurate and detailed descriptions of the HEP detectors are turning out to be crucial elements of the software chains used for simulation, visualization and reconstruction programs: for this reason, it is of paramount importance to dispose of and to deploy generic detector description tools which allow for precise modeling, visualization, visual debugging and interactivity and which can be used to feed information in e.g. Geant4 based simulation programs and in reconstruction-oriented geometry models: at the same time, these tools must allow for different levels of descriptions, ranging from very accurate
geometries aimed at very precise Geant simulation to more generic descriptions of scattering centers in a track reconstruction program.
In this paper we describe a system which was developed to describe the ATLAS muon spectrometer, which is based on a generic XML detector description system (AGDD, ATLAS Generic Detector Description), on the Persint visualization program and on a series of parsers/converters which build a generic, transient geometry model and translate it into the commonly used geometry descriptions (Geant4, the ATLAS GeoModel, ROOT TGeo etc.). These tools permit an easy, self descriptive approach to the detector description problem, intuitive visualization and rapid turn-around, since the results of the description process can be immediately fed into e.g. a Geant4 simulation for rapid prototyping. Examples of the current usage for the ATLAS detector description as well as prototyping for upgrade elements will be given and further developments needed to meet future requirements will be discussed

Poster Session / **462**

The ZEUS data preservation project (ZEUS Collaboration)

Authors: Janusz Malka[1]; Katarzyna Wichmann[1]

[1] *DESY*

A project to allow long term access and physics analysis of ZEUS data (ZEUS data preservation) has been established in collaboration with the DESY-IT group. In the ZEUS approach the analysis model is based on the Common Ntuple

project, under development since 2006. The real data and all presently available Monte Carlo samples are being preserved in a flat ROOT ntuple format. There is ongoing work to provide the ability to simulate new, additional Monte Carlo samples also in the future. The validation framework of such a scheme using virtualisation techniques is being explored. The goal is to validate the frozen ZEUS software against future changes in hardware and operating system. A cooperation between ZEUS, DESY-IT and the library was established for document digitisation and long-term preservation of collaboration web pages. Part of the ZEUS internal documentation has already been stored within the HEP documentation system INSPIRE. Existing digital documentation, needed to perform physics analysis also in the future, is being centralised and completed.

Poster Session / **464**

The H1 data preservation project (H1 Collaboration)

Authors: David South[1]; Michael Steder[1]

[1] *DESY*

Corresponding Authors: michael.steder@desy.de, david.south@cern.ch

The H1 data preservation project was started in 2009 as part of the global data preservation in high-energy physics (DPHEP) initiative. In order to retain the full potential for future improvements, the H1 collaboration aims for level 4 of the DPHEP recommendations, requiring the full simulation and reconstruction chain to be available for analysis. A major goal of the H1 project is therefore to provide secure, long-lived and validated access to the H1 data and analysis software, which is realised in collaboration with DESY-IT using virtualisation techniques. By implementing such a system, it is hoped that the lifetime of the unique HERA data will be extended, providing the possibility for novel analysis in the near future. Improvements may come from for example the development of new experimental techniques or the implementation of currently unavailable higher order Monte Carlo corrections. The preservation of data and software is performed alongside a consolidation programme of all (non-)digital documentation since the beginning of the H1 collaboration in the early 1980s. This presentation will introduce idea, structure and status of the H1 data preservation program.

Poster Session / **465**

Prompt data reconstruction of the ATLAS experiment

Author: Collaboration Atlas[1]

[1] *Atlas*

Corresponding Author: graeme.andrew.stewart@cern.ch

Abstract: The ATLAS experiment at the LHC collider recorded more than 3 fb-1 data of pp collisions at the center of mass energy
of 7 TeV by September 2011. The recorded data are promptly reconstructed in two steps at a large computing farm at CERN to provide fast access to high quality data for physics analysis. In the first step a subset of the collision data corresponding to 10 Hz is processed in parallel with data taking. Data quality, detector calibration constants and beam spot position are determined using the reconstructed data in 36 hours. In the
second step the whole recorded data are processed with the updated parameters. The LHC largely increased the instantaneous
luminosity and the number of interactions per bunch crossing in 2011 and the data recording rate by ATLAS exceeds 400 Hz. To cope with these challenges the performance and reliability of the ATLAS reconstruction software have been improved.
In this presentation we describe how the prompt data reconstruction system quickly and stably provides high quality data to analyzers

Poster Session / **466**

Taking the C out of CVMFS: providing repositories for country-local VOs.

Author: Sam Skipsey[1]

Co-authors: Andy Turner [2]; Thomas Doherty [3]

[1] *University of Glasgow / GridPP*

[2] *University of Leeds*

[3] *University of Glasgow (GB)*

Corresponding Author: samuel.cadellin.skipsey@cern.ch

The caching, http-mediated filesystem "CVMFS", while first developed for use with the Cern Virtual Machines project, has quickly become a significant part of several VOs software distribution policy, with ATLAS being particularly interested.
The benefits of CVMFS do not simply extend to large VOs, however; small virtual organisations can find software distribution to be problematic, as they don't have any real effort to manage multiple sites.
We explore a case study of a local CVMFS repository, installed at Glasgow, for the UK local neiss.org.uk VO, comparing it to the less structured software management used previously.

Poster Session / **467**

H1 Monte Carlo Production on the Grid (H1 Collaboration)

Author: Bogdan Lobodzinski[1]

[1] *DESY*

Corresponding Author: bogdan@mail.desy.de

The H1 Collaboration at HERA is now in the era of high precision analyses based on the final and complete data sample. A natural consequence of this is the huge increase in requirement for simulated Monte Carlo (MC) events. As a response to this increase, a framework for large scale MC production using the LCG Grid Infrastructure was developed. After 3 years, the H1 MC Computing Framework has become a high performance, reliable and robust platform operating on the top of gLite infrastructure. The original framework has been expanded into a tool which can handle 600 million simulated MC events per month and 20,000 simultaneously supported jobs on the LHC Grid, decreasing operator effort to the minimum. An annual MC event production rate of over 2.5 billion events has been achieved, and the project is integral to the data analysis performed by H1. Tools have also been developed to allow modifications of H1 detector details, for different levels of MC production steps and for full monitoring of the jobs on the Grid sites. The H1 MC Framework will be described, based on the experience gained during the successful MC simulation for the H1 Experiment, focussing on the solutions which can be implemented for other types of experiments - not only those devoted to HEP. Failure states, deficiencies, bottlenecks and scaling boundaries observed during this full scale physics analysis endeavour are also addressed.

Poster Session / **468**

Track finding in ATLAS using GPUs

Author: Collaboration Atlas[1]

Co-author: Christian Schmitt [2]

[1] *Atlas*

[2] *CERN*

Corresponding Author: johannes.mattmann@cern.ch

The reconstruction and simulation of collision events is a major task in modern HEP experiments involving several ten thousands of
standard CPUs. On the other hand the graphics processors (GPUs) have become much more powerful and are by far outperforming the standard CPUs in terms of floating point operations due to their massive parallel approach. The usage of these GPUs could therefore significantly reduce the overall reconstruction time per event or allow for the
usage of more sophisticated algorithms.
In this contribution the track finding in the ATLAS experiment will be used as an example on how the GPUs can be used in this context: both the seed finding as well as the Kalman filter show already a speed increase of one order of
magnitude compared to the same implementation on a standard CPU. On the other hand the implementation on the GPU requires a change in the algorithmic flow to allow the code to work in the rather limited environment on the GPU in terms of memory, cache, and transfer speed
from and to the GPU.

Poster Session / 469

ATLAS Offline Data Quality System Upgrade

Author: Collaboration Atlas[1]

Co-authors: Helen Hayward [2]; Peter Onyisi [3]; Peter Waller [4]; Tobias Golling [5]

[1] *Atlas*

[2] *University of Liverpool (GB)*

[3] *University of Chicago (US)*

[4] *University of Liverpool-Unknown-Unknown*

[5] *Yale University (US)*

Corresponding Author: steven.farrell@cern.ch

The ATLAS data quality software infrastructure provides tools for prompt investigation of and feedback on collected data and propagation of these results to analysis users. Both manual and automatic inputs are used in this system. In 2011, we upgraded our framework to record
all issues affecting the quality of the data in a manner which allows users to extract as much information (of the data) for their particular analyses as possible. By improved recording of issues, we are allowed the
ability to reassess the impact of the quality of the data on different physics measurements and adapt accordingly. We have gained
significant experience with collision data operations and analysis; we have used this experience to improve the data quality system,
particularly in areas of scaling and user interface. This talk describes the experience gained in assessing and recording of the data
quality of ATLAS and subsequent benefits to the analysis users.

Computer Facilities, Production Grids and Networking / 470

Centralized Fabric Management Using Puppet, Git, and GLPI

Author: Jason Alexander Smith[1]

Co-authors: Christopher Hollowell [2]; Hironori Ito [1]; James Pryor [2]; John Peter Fetzko [3]; John Steven De Stefano Jr [1]; Mizuki Karasawa [4]; William Strecker-Kellogg [5]

[1] *Brookhaven National Laboratory (US)*

[2] *Brookhaven National Laboratory*

[3] *Brookhaven National Laboratory (BNL)-Unknown-Unknown*

[4] *BNL*

[5] *Brookhaven National Lab*

Corresponding Authors: smithj4@bnl.gov, jd@bnl.gov

Managing the infrastructure of a large and complex data center can be extremely difficult without taking advantage of automated services. Puppet is a seasoned, open-source tool designed for enterprise-class centralized configuration management. At the RHIC/ATLAS Computing Facility at Brookhaven National Laboratory, we have adopted Puppet as part of a suite of tools, including Git, GLPI, and some custom scripts, that comprise our centralized configuration management system. In this paper, we discuss the use of these tools for centralized configuration management of our servers and services; change management, which requires authorized approval of production changes; a complete, version-controlled history of all changes made; separation of production, testing, and development systems using Puppet environments; semi-automated server inventory using GLPI; and configuration change monitoring and reporting via the Puppet dashboard. We will also

discuss scalability and performance results from using these tools on a 2,000+ node cluster and a pool of over 400 infrastructure servers with an administrative staff of only about 20 full-time employees.

In addition to managing our data center servers, we've also used this Puppet infrastructure successfully to satisfy recent security mandates from our funding agency (the U.S. Department of Energy) to centrally manage all staff Linux and UNIX desktops; in doing so, we've extended several core Puppet modules to not only support both RHEL 5 and 6, but also to include support for other operating systems, including Fedora, Gentoo, OS X, and Ubuntu.

Poster Session / **471**

Toolkit for data reduction to tuples for the ATLAS experiment

Author: Scott Snyder[1]

Co-authors: Attila Krasznahorkay [2]; Collaboration Atlas [3]

[1] *Brookhaven National Laboratory (US)*

[2] *New York University (US)*

[3] *Atlas*

Corresponding Author: scott.snyder@cern.ch

The final step in a HEP data-processing chain is usually to reduce the data to a 'tuple' form which can be efficiently read by interactive analysis tools such as ROOT. Often, this is implemented independently by each group analyzing the data, leading to duplicated effort and needless divergence in the format of the reduced data. ATLAS has implemented a common toolkit for performing this processing step. By using tools from this package, physics analysis groups can produce
tuples customized for a particular analysis but which are still consistent in format and vocabulary with those produced by other physics groups.
The package is designed so that almost all the code is independent of the specific form used to store the tuple. The code that does depend on this is grouped into a set of small backend packages. While the ROOT backend is the most used, backends also exist for HDF5 and for specialized databases.
By now, the majority of ATLAS analyses rely on this package, and it is an important contributor to the ability of ATLAS to rapidly analyze physics data.

Poster Session / **472**

The ATLAS physics analysis model and production of derived datasets

Author: Collaboration Atlas[1]

[1] *Atlas*

Corresponding Author: amir.farbin@cern.ch

The ATLAS experiment has collected vast amounts of data with the arrival of the inverse-femtobarn era at the LHC. ATLAS has developed an intricate analysis model with several types of derived datasets, including
their grid storage strategies, in order to make data from O(109) recorded events readily available to physicists for analysis. Several use cases have been considered in the ATLAS analysis model with a few distinct classes
of analyses that need to look at various parts of the overall data. A rst class of analysis needs very detailed information in order to

study detector and reconstruction performance, as well as performing physics analyses for non-standard scenarios. For this case, specialized Derived Event Summary Data (DESD) are produced at the Tier 0, right after the main reconstruction was performed. These DESDs contain only specic events and sometimes also very specic per-event content in order to keep their total size manageable. They are distributed on the grid in order to allow easy access for physicists. All other types of analysis could in principle be performed on the
Analysis Object Data (AOD) which has a size of ‘150 kByte/event. It is generally considered the main data format for physics analysis in ATLAS. However, its still very large total size makes it unpractical in most cases to frequently process it. Thus, further size reduction is necessary. Derived AODs (DAOD) are produced from the AOD and distributed
on the grid with only events and content of interest to a specic class of physics analysis. These DAODs contain the full object structure of the AOD. The most commonly used data format for physics analysis is today
a type of ROOT les known as D3PDs which contain only simple types and vectors of simple types for selected events, i.e. no ATLAS-specic software is needed to process them. Both DAODs and D3PDs can be
produced centrally for individual analyses or whole analysis groups. They are stored on the grid using dedicated group-specic space quota, but fully available to the whole collaboration.
In all cases, the selection criteria, luminosity bookkeeping, and other relevant information is stored inside the resulting les as meta-data.

Poster Session / **473**

Performance of the ATLAS Reconstruction Software with high level of Pileup

Author: Collaboration Atlas[1]

[1] *Atlas*

Corresponding Author: rolf.seuster@cern.ch

In 2011 the LHC provided excellent data, the integrated luminosity of about 5fb-1 was more than what was expected. The price for this
huge data set is the in and out of time pileup, additional soft events overlaid on top of the interesting event. The reconstruction software is very sensitive to these additional particles in the event, as the reconstruction time increases due to increased combinatorics. During the running of the experiment in 2011, several successful changes to the software were made that sped up the reconstruction. Pileup has different effects on the various detector technologies used in ATLAS and a general recipe is not applicable.

Poster Session / **474**

The ”NetBoard”: Network Monitoring Tools Integration for INFN Tier-1 Data Center

Author: Donato De Girolamo[1]

Co-author: Stefano Zani [1]

[1] *INFN-CNAF*

Corresponding Author: donato.degirolamo@cnaf.infn.it

The monitoring and alert system is fundamental for the management and the operation of the network in a large data center such as an LHC Tier-1.

The network of the INFN Tier-1 at CNAF is a multi-vendor environment: for its management and monitoring several tools have been adopted and different sensors have been developed.

In this paper, after an overview on the different aspects to be monitored and the tools used for this (i.e. MRTG, Nagios, Arpwatch, NetFlow, Syslog, etc), we will describe the "NetBoard", a monitoring toolkit developed at the INFN Tier-1.

NetBoard, developed for a multi-vendor network, is able to install and auto-configure all tools needed for its monitoring, either via network devices discovery mechanism or via configuration file or via wizard. In this way, we are also able to activate different types of sensors and Nagios checks according to the equipment vendor specifications. Moreover, when a new devices is connected in the LAN, NetBoard can detect where it is plugged.

Finally the NetBoard web interface allows to have the overall status of the entire network "at a glance", both the local and the geographical (including the LHCOPN and the LHCONE) link utilization, health status of network devices (with active alerts) and flow analysis.

Distributed Processing and Analysis on Grids and Clouds / 475

The Open Science Grid –Support for Multi-Disciplinary Team Science –the Adolescent Years

Author: Ruth Pordes[1]

Co-author: Miron Livny [2]

[1] *Fermi National Accelerator Lab. (US)*

[2] *University of Wisconsin Madison*

Corresponding Author: ruth@fnal.gov

As it enters adolescence the Open Science Grid (OSG) is bringing a maturing fabric of Distributed High Throughput Computing (DHTC) services that supports an expanding HEP community to an increasingly diverse spectrum of domain scientists. Working closely with researchers on campuses throughout the US and in collaboration with national cyberinfrastructure initiatives, we transform their computing environment through new concepts, advanced tools and deep experience. We discuss examples of these including: the pilot-job overlay concepts and technologies now in use throughout OSG and delivering 1.4 Million CPU hours/day; the role of campus infrastructures- built out from concepts of sharing across multiple local faculty clusters (made good use of already by many of the HEP Tier-2 sites in the US); the work towards the use of clouds and access to high throughput parallel (multi-core and GPU) compute resources; and the progress we are making towards meeting the data management and access needs of non-HEP communities with general tools derived from the experience of the pariochial tools in HEP (integration of Globus Online, prototyping with IRODS, investigations into Wide Area Lustre).

We will also review our activities and experiences as HTC Service Provider to the recently awarded NSF XD XSEDE project, the evolution of the US NSF TeraGrid project, and how we are extending the reach of HTC through this activity to the increasingly broad national cyberinfrastructure. We believe that a coordinated view of the HPC and HTC resources in the US will further expand their impact on scientific discovery.

Poster Session / 477

The "NetBoard": Network Monitoring Tools Integration for INFN Tier-1 Data Center

Author: Donato De Girolamo[1]

Co-author: Stefano Zani

[1] *INFN*

Corresponding Authors: donato.degirolamo@cnaf.infn.it, stefano.zani@cnaf.infn.it

The monitoring and alert system is fundamental for the management and the operation of the network in a large data center such as an LHC Tier-1.

The network of the INFN Tier-1 at CNAF is a multi-vendor environment: for its management and monitoring several tools have been adopted and different sensors have been developed.

In this paper, after an overview on the different aspects to be monitored and the tools used for this (i.e. MRTG, Nagios, Arpwatch, NetFlow, Syslog, etc), we will describe the "NetBoard", a monitoring toolkit developed at the INFN Tier-1.

NetBoard, developed for a multi-vendor network, is able to install and auto-configure all tools needed for its monitoring, either via network devices discovery mechanism or via configuration file or via wizard. In this way, we are also able to activate different types of sensors and Nagios checks according to the equipment vendor specifications. Moreover, when a new devices is connected in the LAN, NetBoard can detect where it is plugged.

Finally the NetBoard web interface allows to have the overall status of the entire network "at a glance", both the local and the geographical (including the LHCOPN and the LHCONE) link utilization, health status of network devices (with active alerts) and flow analysis.

Poster Session / **478**

Fast simulation for ATLAS: Atlfast-II and ISF

Author: Wolfgang Lukas[1]

Co-authors: Daniel Froidevaux [2]; Philip Clark [3]

[1] *University of Innsbruck (AT)*

[2] *CERN*

[3] *University of Edinburgh (GB)*

Corresponding Author: wolfgang.lukas@cern.ch

We present the ATLAS simulation packages ATLFAST-II and ISF.
Atlfast-II is a sophisticated fast parametrized simulation in the Calorimeter system in combination with full Geant4 simulation precision in the Inner Detector and Muon Systems. This combination offers a relative increase in speed of around a factor of ten compared to the standard ATLAS detector simulation and is being used to supplement the ATLAS MC needs for 2011.
The design of the parametrized simulation components allows for a flexible tuning of the simulated detector response to direct
measurements with the ATLAS detector in order to increase the overall simulation quality. Hence the tuned fast simulation approach promises to be usable for the production of large MC samples needed for new physics searches as well as precision measurements.
The newly developed highly configurable ATLAS Integrated Simulation Framework (ISF) extends Atlfast-II by fast inner detector and muon system simulation. Furthermore, depending on the required accuracy, differently flavored full and fast simulation approaches can be mixed within a single event to allow an optimal balance between precision and execution time. The ISF is built into the ATLAS event data processing framework Athena and is designed to allow for future
extension and the application of parallel computing techniques.

Poster Session / 479

Parallel algorithms for track reconstruction in the CBM experiment

Author: Igor Kulakov[1]

Co-authors: Ivan Kisel [2]; Maksym Zyzak [1]

[1] *Goethe Universitaet Frankfurt*

[2] *GSI Helmholtzzentrum fuer Schwerionenforschung GmbH*

Corresponding Author: i.kulakov@gsi.de

The CBM experiment is a future fixed-target experiment at FAIR/GSI (Darmstadt, Germany). It is being designed to study heavy-ion collisions at extremely high interaction rates. The main tracking detectors are the Micro-Vertex Detector (MVD) and the Silicon Tracking System (STS). Track reconstruction in these detectors is very complicated task because of several factors. Up to 1000 tracks per central Au+Au collision intersect 5x5 cm2 region of the first MVD detector plane. Double-sided strip detector modules are used in STS, that leads to about 85% additional combinatorial space points. The detectors are placed in the non-homogeneous magnetic field. The full event reconstruction is required for online event selection. Therefore, both the speed of the reconstruction algorithms and their efficiency are crucial.
The Cellular Automaton (CA) algorithm is used for the track reconstruction. It is based on a local reconstruction and therefore is robust, fast and easily parallelizable. The algorithm is optimized for the very complicated and realistic simulation of the detectors. Reconstruction of the central collisions shows 95% efficiency for most of signal particles, 5% incorrectly reconstructed tracks and speed of 200 ms per event per core. The algorithm is stable against detector inefficiency. The CA algorithm is suitable for complicated conditions and high interaction rates of the CBM experiment.
The Kalman filter (KF) based package is used for precise estimation of track parameters. It includes track fitter, track smoother and deterministic annealing filter (DAF). Initial approximate estimation of the track parameters with the least square method is used in order to increase stability of the KF algorithms. Several approaches of the Kalman filter track fit are implemented: conventional Kalman filter, U-D filtering and two approaches for the square root Kalman filter. The square root approach, which based on the Potter's measurement-update equations, is robust with respect to computational round-off errors, gives 1.1% momentum resolution and takes less then 2.5 us per track. It appears to be the most suitable for track fitting in CBM. Two procedures for the track propagation in non-homogeneous magnetic field are implemented: a standard fourth-order Runge-Kutta method and a method based on the analytic formula, specially developed for the CBM experiment. The DAF based procedure rejects with 99% efficiency the noise hits placed in 300 um from the true track position.
The track finder and track fitter procedures are implemented in single precision, use the SIMD instruction set and multithreading for parallel computations. The algorithms show a strong scalability with respect to number of cores. Results for the newest AMD Opteron CPU with 48 cores and Intel Westmere CPU with 40 hardware (80 logical) cores are presented and discussed.
Future plans include investigation of the parallel algorithms for track reconstruction on non-homogeneous many-core CPU/GPU systems.

Event Processing / 480

Parallel implementation of the KFParticle vertexing package for the CBM and ALICE experiments

Author: Maksym Zyzak[1]

Co-authors: Igor Kulakov [1]; Ivan Kisel [2]

[1] *Goethe Universitaet Frankfurt*

[2] *GSI, Gesellschaft fuer Schwerionenforschung mbH*

Corresponding Authors: i.kulakov@gsi.de, m.zyzak@gsi.de

Modern heavy-ion experiments operate with very high data rates and track multiplicities. Because of time constraints the speed of the reconstruction algorithms is crucial both for the online and offline data analysis. Parallel programming is considered nowadays as one of the most efficient ways to increase the speed of event reconstruction.

Reconstruction of short-lived particles is one of the most important tasks in data analysis of high energy physics experiments. The KFParticle package for short-lived particles reconstruction, based on the Kalman filter, is presented and described with mathematical apparatus. The package is actively used both in the CBM experiment at FAIR/GSI (Darmstadt, Germany) and the ALICE experiment at CERN (Geneva, Switzerland). The high computational speed of the KFParticle package in the CBM experiment is of the particular importance, because the full event reconstruction is required for the online event selection. Also in the ALICE experiment it is important for the analysis of the already collected data. The KFParticle package is geometry independent and can be used in other experiments too.

The package has rich functionality: the complete particle reconstruction with momentum and covariance matrix calculation; reconstruction of decay chains; daughter particles can be added one by one; simple access to parameters of the particle, such as mass, lifetime, decay length, rapidity, and their errors; transport of the particle; estimation of the distance between particles etc.

KFParticle has been vectorized using the SIMD instructions set. Since modern processors have SIMD units, vectoization is a simple and efficient way to increase the computational speed of the algorithms running on the same CPU. The package has been implemented in single precision for more efficient vectorization. The additional speedup factor of 3-5 has been achieved for the CBM and ALICE experiments. The Intel TBB library is used for parallelization between cores. The quality analysis of parameters, their errors and covariance matrix, which are obtained with KFParticle, has been performed using Monte Carlo simulated data. Results of the analysis are presented and discussed.

More sophisticated statistical methods for the particle analysis are under implementation within the KFParticle package. Implementation of the package using the parallel Intel ArBB library, as well as parallelization on GPU architectures are foreseen.

Poster Session / 481

Methods and the computing challenges of the realistic simulation of physics events in the presence of pile-up in the ATLAS experiment

Author: Collaboration Atlas[1]

Co-authors: Daniel Froidevaux [2]; Philip Clark [3]

[1] *Atlas*

[2] *CERN*

[3] *University of Edinburgh (GB)*

Corresponding Author: drandyhaas@gmail.com

We are now in a regime where we observe substantial multiple proton-proton collisions within each filled LHC bunch-crossing and also multiple filled bunch-crossings within the sensitive time window of the ATLAS detector. This will increase with increased luminosity in the near future.
Including these effects in Monte Carlo simulation poses significant computing challenges. We present a description of the standard approach used by the ATLAS experiment and details of how we manage the conflicting demands of keeping the background dataset size as small as possible while minimizing the effect of background event re-use. We also present details of the methods used to minimize the memory footprint of these digitization jobs, to keep them within the grid limit, despite combining the information from thousands of simulated events at once.
We also describe an alternative approach, known as Overlay. Here, the actual detector conditions

are sampled from raw data using a special zero-bias trigger, and the simulated physics events are overlaid on top of this zero-bias data. This gives a realistic simulation of the detector response to physics events. The overlay simulation runs in time linear in the number of events and consumes memory proportional to the size of a single event, with small overhead. We explain the computational issues and challenges that will arise in running overlay in production mode on the grid. Finally we discuss the computational issues that may arise in the future in generating large amount of luminosity weighted zero-bias data and making it available on the grid

Poster Session / **482**

The HERMES data preservation project (HERMES Collaboration)

Author: Eduard Avetisyan[1]

[1] *DESY*

Corresponding Author: dich@mail.desy.de

We discuss the steps and efforts required to secure the continued analysis and data access for the HERMES experiment after the end of the
active collaboration period. The model for such an activity has been developed within the framework of the DPHEP initiative in a close
collaboration of HERA experiments and the DESY IT. For HERMES the preservation scheme foresees a possibility of full data production chain
starting from the raw data, as well as MC productions using existing and future generators. In that scheme, the main analysis data format to preserve are microDSTs based on ADAMO tables wrapped in a special portability layer DAD. The necessary software packages are preserved and validated in a special virtual environment developed by the IT, to allow a
flowless porting of the software to future OS and compiler libraries if need be. In parallel, reliable storage and access to relevant
documentation is pursued. The lessons we learned from the past may help currently active collaborations to avoid the penalties that come
for starting late.

Distributed Processing and Analysis on Grids and Clouds / **484**

Dynamic Extension of a Virtualized Cluster by using Cloud Resources

Authors: Oliver Oberst[1]; Thomas Hauth[1]

Co-authors: David Kernert [1]; Gunter Quast [1]; Stephan Riedel [1]

[1] *KIT - Karlsruhe Institute of Technology (DE)*

Corresponding Author: oliver.oberst@cern.ch

The specific requirements concerning the software environment within the HEP community constrain the choice of resource providers for the outsourcing of computing infrastructure. The use of virtualization in HPC clusters and in the context of cloud resources is therefore a subject of recent developments in scientific computing.
The dynamic virtualization of worker nodes in common batch systems provided by ViBatch serves each user with a dynamically virtualized subset of worker nodes on a local cluster. Now it can be transparently extended by the use of common open source cloud interfaces like OpenNebula or Eucalyptus, launching a subset of the virtual worker nodes within the cloud.
It is demonstrated how a dynamically virtualized computing cluster is combined with cloud resources by attaching remotely started virtual worker nodes to the local batch system.

Summary:

The IEKP institute at the Karlsruhe Institute of Technology (KIT) is sharing a cluster with nine different departments. The cluster, maintained by the central computing department of KIT is installed with a SuSE Enterprise Linux. To be able to use CERN specific setups and software (e.g. AFS, CMSSW) the IEKP relies on a Scientific Linux OS environment. Therefore, we developed the dynamic virtualization of worker nodes within common batch systems (ViBatch) as presented already at CHEP09. In order to cope peak load times of the cluster we now extended this local "by-job" virtualization system by using worker nodes which are automatically spawned within cloud resources through the "Responsive On-demand Cloud Enabled Deployment" (ROCED). ROCED hereby manages the monitoring of the local job queues and spawns automatically new cloud worker nodes if a certain queue length threshold is reached. The locally used virtual machines are setup with SLC5 and use the CernVMFS to provide the CMS software transparently to our users.

The presentation will give a summary on both tools, ViBatch and ROCED, their new features and the experiences of using the combination of both in our local analysis production system.

Poster Session / **485**

Many-core experience with HEP software at CERN openlab

Authors: Alfio Lazzaro[1]; Andrzej Nowak[1]; Julien Leduc[1]; Sverre Jarp[1]

[1] *CERN openlab*

Corresponding Author: andrzej.nowak@cern.ch

The continued progression of Moore's law has led to many-core platforms becoming easily accessible commodity equipment. New opportunities that arose from this change have also brought new challenges: harnessing the raw potential of computation of such a platform is not always a straightforward task. This paper describes practical experience coming out of the work with many-core systems at CERN openlab and the observed differences with respect to their predecessors. We provide the latest results for a set of parallelized HEP benchmarks running on several classes of many-core platforms.

Software Engineering, Data Stores and Databases / **486**

The future of commodity computing and many-core versus the interests of HEP software

Authors: Alfio Lazzaro[1]; Andrzej Nowak[1]; Julien Leduc[1]; Sverre Jarp[1]

[1] *CERN openlab*

Corresponding Authors: andrzej.nowak@cern.ch, sverre.jarp@cern.ch

As the mainstream computing world has shifted from multi-core to many-core platforms, the situation for software developers has changed as well. With the numerous hardware and software options available, choices balancing programmability and performance are becoming a significant challenge. The expanding multiplicative dimensions of performance offer a growing number of possibilities that need to be assessed and addressed on several levels of abstraction. This paper reviews the major tradeoffs forced upon the software domain by the changing landscape of parallel technologies –hardware and software alike. Recent developments, paradigms and techniques are considered with respect to their impact on the rather traditional HEP programming models. Other considerations addressed include aspects of efficiency and reasonably achievable targets for the parallelization of large scale HEP workloads.

Poster Session / 487

WMSMonitor advancements in the EMI era

Authors: Daniele Cesini[1]; Danilo Dongiovanni[1]; Enrico Fattibene[1]

[1] *INFN-CNAF, IGI*

Corresponding Author: danilo.dongiovanni@cnaf.infn.it

In production Grid infrastructures deploying EMI (European Middleware Initiative) middleware release, the Workload Management System (WMS) is the service responsible for the distribution of user tasks to the remote computing resources. Monitoring the reliability of this service, the job lifecycle and the workflow pattern generated by different user communities is an important and challenging activity.
Initially designed to monitor and manage a distributed cluster of gLite WMS/LB (Logging and Bookeeping) services, WMSMonitor has proved to be a useful and flexible tool for a variety of user categories. In fact, after asynchronously extracting information from all monitored instances, WMSMonitor reaggregates it by different keys (WMS instance, Virtual Organization, User, etc.) providing insight both on services status and on their usage to service administrators, developers, advanced Grid users and performance testers. The positive feedback on WMSMonitor utilization from various production Grid sites pushed us to improve the tool to enhance its flexibility and scalability exploiting a new architecture. Moreover the tool has been made compliant to recent evolutions in the monitored services. We therefore present the new version of WMSMonitor which can monitor EMI WMS/LB services and shows an improved user interface allowing better report capabilities. Among main novelties, we mention the collection of Job Submission Service (JSS) error type statistics and the adoption of ActiveMQ messaging system which now allows multiple data consumers to exploit collected information.
Finally, it is worth to mention that the implemented architecture and the exploitation of a messaging layer commonly adopted in EMI Grid applications make WMSMonitor a flexible tool that can be easily extended to monitor other Grid services.

Poster Session / 488

Numerical accuracy and auto-vectorization of probability density functions used in high energy physics

Authors: Alfio Lazzaro[1]; Andrzej Nowak[2]; Felice Pantaleo[3]; Julien Leduc[None]; Sverre Jarp[2]; Yngve Sneen Lindal[4]

Co-author: Vincenzo Innocente [2]

[1] *CERN openlab*

[2] *CERN*

[3] *University of Pisa (IT)*

[4] *Norwegian University of Science and Technology (NO)*

Corresponding Author: felice.pantaleo@cern.ch

Data analyses based on evaluation of likelihood functions are commonly used in the high-energy physics community for fitting statistical models to data samples. The likelihood functions require the evaluation of several probability density functions on the data. This is accomplished using loops. For the evaluation operations, the standard accuracy is double precision floating point. The probability density functions require the evaluation of several transcendental functions (mainly exponential and square roots). Therefore, fast evaluation of the likelihood functions can be achieved either by a faster execution of the transcendental expressions or using vectorization for the loops. The former can be achieved reducing the numerical accuracy, i.e. using single precision floating or in general less accurate functions. The latter requires special techniques to vectorize the transcendental functions. However, the impact of this optimization can be significant, and in particular in the future when the

vectors units will become larger and larger. Several compilers gives the possibility to apply auto-vectorization and several floating point optimizations. We will show results when using different compilers on different hardware systems for several probability distribution functions.

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Computer Facilities, Production Grids and Networking / 489

Overview of storage operations at CERN

Authors: Jan Iven[1]; Massimo Lamanna[1]

Co-authors: Andreas Joachim Peters [1]; Giuseppe Lo Presti [1]; Ignacio Reguero [1]; John Hefferman [1]; Luca Mascetti [1]; Sebastien Ponce [1]

[1] *CERN*

Corresponding Authors: jan.iven@cern.ch, massimo.lamanna@cern.ch

Large-volume physics data storage at CERN is based on two services, CASTOR and EOS:
* CASTOR - in production for many years - now handles the Tier0 activities (including WAN data distribution), as well as all tape-backed data;
* EOS - in production since 2011 - supports the fast-growing need for high-performance low-latency (i.e. diskonly) data access for user analysis.

In 2011, a large part of the original CASTOR storage has been migrated into EOS, which grew from the original testbed installation (1 PB usable capacity for ATLAS) to over 6 PB for ATLAS and CMS. EOS has
has been validated for several month under production conditions and has already replaced several CASTOR service classes.

CASTOR also evolved during this time with major improvements in critical areas - notably the internal scheduling of requests, the simplifications of the database structure and a complete overhaul of
the tape subsystem.

The talk will compare the two systems from an operation's perspective (setup, day-by-day user support, upgrades, resilience to common
failures) while taking into account their different scope. In the case of CASTOR we will analyse the impact of the 2011 improvements on delivering Tier0 services, while for EOS we will focus on the steps to achieve to a production-quality service.
For both systems, the upcoming changes will be discussed in relation with the evolution of the LHC programme and computing models (data volumes, access patterns, relations among computing sites).

Poster Session / 490

Parallel Likelihood Function Fits on Heterogeneous Many-core Systems with OpenMP, CUDA, and MPI technologies

Authors: Alfio Lazzaro[1]; Andrzej Nowak[2]; Felice Pantaleo[2]; Julien Leduc[None]; Ruggero Caravita[3]; Sverre Jarp[2]; Yngve Sneen Lindal[4]

[1] *CERN openlab*

[2] *CERN*

[3] *Universita e INFN (IT)*

[4] *Norwegian University of Science and Technology (NO)*

Corresponding Authors: julien.leduc@cern.ch, felice.pantaleo@cern.ch, alfio.lazzaro@cern.ch

Data analyses based on evaluation of likelihood functions are commonly used in the high energy physics community for fitting statistical models to data samples. These procedures require several evaluations of these functions and they can be very time consuming. Therefore, it becomes particularly important to have fast evaluations. This paper describes a parallel implementation that allows to run cooperatively the evaluations of the negative log-likelihood function for data analysis methods on heterogeneous computational devices (i.e. CPU and GPU) belonging to a single computational node or on several homogeneous nodes connected by a network. The implementation is able to split and balance the workload needed for the evaluation of the function in corresponding sub-workloads to be executed in parallel on each computational device. The CPU parallelization is implemented using OpenMP, while the GPU implementation is based on CUDA. The parallelization over several nodes is based on MPI. The comparison of the performance of these implementations for different configurations and different hardware systems is reported. Tests are based on real data analyses carried out by the high energy physics community taken from RooFit and RooStats packages.

Computer Facilities, Production Grids and Networking / 491

Status and trends in networking at LHC Tier1 facilities

Author: Andrey Bobyshev[1]

Co-authors: Aurelie Reymund [2]; Bruno Heinrich Hoeft [3]; Philip DeMar [1]; Vytautas Grigaliunas [4]; john bigrow

[1] *FERMILAB*

[2] *Forschungszentrum Karlsruhe*

[3] *KIT - Karlsruhe Institute of Technology (DE)*

[4] *Fermilab*

Corresponding Authors: bobyshev@fnal.gov, demar@fnal.gov, big@bnl.gov, bruno.hoeft@kit.edu, aurelie.reymund@kit.edu

The LHC is entering its fourth year of production operation. Many Tier1 facilities can count up to a decade of existence when development and ramp-up efforts are included. LHC computing has always been heavily dependent on high capacity, high performance network facilities for both the LAN and WAN data movement, particularly within the Tier1 centers. As a result, the Tier1 centers tend to be on the leading edge of data center networking technology. In this paper, we conduct an analysis of past and current developments in Tier1 networking, as well as extrapolating where we anticipate things are heading. A large part of our analysis will be based on the US-CMS Tier1 at Fermilab as a use case. Our analysis will include examination into the following areas:

• Evolution of the US-CMS Tier1 center to its current state⋯
• The changing environment of data center networking approaches & practices and how they may apply to Tier-1 centers
• Likely impact of emerging network technologies (10GE-connected hosts, 40GE/100GE links, IPv6) on Tier-1 centers
• Trends in WAN data movement and emergence of software-defined WAN network capabilities (end-to-end circuits, OpenFlow, etc)
• PerfSONAR framework's current and potential use within Tier-1 centers for performance measurement and analysis

Summary:

Analysis of state and trends in networking at LHC Tier1 facilities. A large part of our analysis will be based on the US-CMS Tier1 facility but we also intend to analyze experience, both LAN and WAN of

other Tier1 facilities.

Event Processing / 492

Acceleration of multivariate analysis techniques in TMVA using GPUs

Author: Andrew John Washbrook[1]

Co-authors: Andreas Hoecker [2]; Jan Therhaag [3]; Robert Duane Harrington Jr [4]

[1] *University of Edinburgh (GB)*

[2] *CERN*

[3] *Universitaet Bonn (DE)*

[4] *University of Edinburgh*

Corresponding Author: andrew.washbrook@cern.ch

Multivariate classification methods based on machine learning techniques are commonly used for data analysis at the LHC in order to look for signatures of new physics beyond the standard model. A large variety of these classification techniques are contained in the Toolkit for Multivariate Analysis (TMVA) which enables training, testing, performance evaluation and application of the chosen methods.

As data continues to be successfully collected at the LHC at record rates the sample size processed by TMVA is expected to grow by orders of magnitude. However, it is known that some classification techniques are likely to be process bound as the sample size is significantly increased. Other input factors - such as the number of classifier variables defined for a given method - can also lead to an appreciable increase in overall execution time.

A feasibility study into the acceleration of multivariate analysis techniques using Graphics Processing Units (GPUs) will be presented. The MLP-based Artificial Neural Network method has been chosen as a focus for investigation. The challenges faced when refactoring the existing codebase to the CUDA programming language will be considered as well as determining how possible performance improvements can be integrated and extended to other classification techniques in the TMVA framework.

Poster Session / 493

Lxcloud: A Prototype for an Internal Cloud in HEP. Experiences and Lessons Learned

Author: Ulrich Schwickerath[1]

Co-author: Belmiro Moreira [1]

[1] *CERN*

Corresponding Author: ulrich.schwickerath@cern.ch

In 2008 CERN launched a project aiming at virtualising the batch farm. It strictly distinguishes between infrastructure and guests, and is thus able to serve, along with its initial batch farm target, as an IaaS infrastructure, which can be exposed to users. The system was put into production at small scale at Christmas 2010, and has since grown to almost 500 virtual machine slots in spring 2011. It was opened to test case users deploying CERNVM images on it, which opened new possibilities for

delivering IT resources to users in a cloud-like way. This presentation gives an overview over the project, its evolution and growth, as well as the different real-life use cases. Operational experiences and issues will reported as well.

Poster Session / **494**

New Developments in Web Based Monitoring at the CMS Experiment

Author: Irakli Chakaberia[1]

Co-author: Aron Soha [2]

[1] *Kansas State University*

[2] *Fermi National Accelerator Lab. (US)*

Corresponding Authors: irakli@phys.ksu.edu, soha@fnal.gov

The rate of performance improvements of the LHC at CERN has had strong influence on the characteristics of the monitoring tools developed for the experiments. We present some of the latest additions to the suite of Web Based Monitoring services for the CMS experiment, and explore the aspects that address the roughly 20-fold increase in peak instantaneous luminosity over the course of 2011. One of these user-friendly tools allows collaborators to easily view, and make correlations among, accelerator configuration information such as bunch patterns, measured quantities such as intensities, vacuum pressures, and background conditions, as well as derived quantities such as luminosity and the number of simultaneous interactions per beam crossing. An additional tool summarizes the daily, weekly, and yearly luminosity and efficiency. Finally, we discuss a trigger cross section and rate fitting service, that uses data from previous runs to validate current running conditions as well as to serve as a predictive extrapolation tool for developing triggers for higher luminosity running.

Poster Session / **495**

The new CERN Controls Middleware

Author: Andrzej Dworak[1]

Co-authors: Joel Lauener ; Pierre Charrue [1]; Wojtek Sliwinski [1]

[1] *CERN*

Corresponding Author: andrzej.dworak@cern.ch

The Controls Middleware (CMW) project was launched over ten years ago. Its main goal was to unify middleware solutions used to operate CERN accelerator complex. A key part of the project, the equipment access library RDA, was based on CORBA, an unquestionable middleware standard at the time. RDA became an operational and critical part of the infrastructure, yet the demanding run-time environment revealed shortcomings of the system. Accumulation of fixes and workarounds led to unnecessary complexity. RDA became difficult to maintain and to extend. CORBA proved to be rather a cumbersome product than a panacea. Fortunately, many new transport frameworks appeared since then. They boasted a better design and supported concepts that made them easier to use. Willing to profit from the coming long LHC shutdown which will make it possible to update the operational software, the CMW team reviewed user requirements and in their terms investigated eventual CORBA substitutes. Evaluation of several market recognized products helped to identify

three most-suitable middleware solutions: ZeroMQ, Ice and YAMI. This paper presents the prototyping process using the three libraries, its outcome and the influence of the chosen product on the internal implementation of the new RDA. Also, the new generic API and its strengths are presented. The article ends with an outline of the planned deployment process and explains how backward compatibility problems are addressed.

Poster Session / **496**

The Data Operation CEntre Tool. Architecture and population strategies

Author: Stefano Dal Pra[1]

Co-author: Alberto Crescente [1]

[1] *INFN*

Corresponding Author: stefano.dalpra@cnaf.infn.it

Keeping track of the layout of the informatic resources in a big datacenter is a complex task.

DOCET is a database-based webtool designed and implemented at INFN. It aims at providing a uniform interface to manage and retrieve needed information about one or more datacentre, such as available hardware, software and their status.

Having a suitable application is however useless until most of the information about the centre are not inserted in the DOCET's database. Manually inserting all the information from scratch is an unfeasible task.

After describing DOCET's high level architecture, its main features and current development track, we present and discuss the work done to populate the DOCET database for the INFN-T1 site by retrieving information from a heterogenous variety of authoritative sources, such as DNS, DHCP, Quattor host profiles, etc. We then describe the work being done to integrate DOCET with some common management operation, such as adding a newly installed host to DHCP and DNS, or creating a suitable Quattor profile template for it.

Poster Session / **497**

Web enabled data management with DPM & LFC

Authors: Alejandro Alvarez Ayllon[1]; Ricardo Brito Da Rocha[2]

[1] *University of Cadiz*

[2] *CERN*

Corresponding Authors: alejandro.alvarez.ayllon@cern.ch, ricardo.rocha@cern.ch

The Disk Pool Manager (DPM) and LCG File Catalog (LFC) are two grid data management components currently used in production at more than 240 sites. Together with a set of grid client tools they give the users a unified view of their data, hiding most details concerning data location and access.

Recently we've put a lot of effort in developing a reliable and high performance HTTP/WebDAV frontend to both our grid catalog and storage components, exposing the existing functionality to users accessing the services via standard clients - e.g. web browsers, curl - present in all operating systems, giving users a simple and straigh-forward way of interaction. In addition, as other relevant grid storage components (like dCache) expose their data using the same protocol, for the first time we had the opportunity of attempting a unified view of all grid storage using HTTP.

We describe the mechanism used to integrate the grid catalog(s) with the multiple storage components - HTTP redirection -, including details on some assumptions made to allow integration with other implementations. We describe the way we hide the
details regarding site availability or catalog inconsistencies, by switching the standard HTTP client automatically between multiple replicas. We also present measurements of access performance, and the relevant factors regarding replica selection - current throughput and load, geographic proximity, etc.

Finally, we report on some additional work done to have this system as a viable alternative to GridFTP, providing multi-stream transfers and exploiting some additional features of WebDAV to enable third party copies - essential for managing data movements between storage systems - with equivalent performance.

Poster Session / 498

Planning for Obsolescence in a Production Environment: Migration from a Legacy Geometry Code to an Abstract Geometry Modeling Language in STAR

Author: Jason Webb[1]

Co-authors: Jerome LAURET [2]; Victor Perevoztchikov [1]

[1] *Brookhaven National Lab*

[2] *BROOKHAVEN NATIONAL LABORATORY*

Corresponding Author: jwebb@bnl.gov

Faced with the abundance of geometry models available within the HENP community, long running experiments face a daunting challenge: how to migrate legacy GEANT3 based detector geometries to new technologies, such as the ROOT/TGeo framework [1]. One approach, entertained by the community for some time, is to introduce a level of abstraction: implementing the geometry in a higher order language independent of the concrete implementation of the geometry model. This approach faces many practical challenges and, until now, has remained at the conceptual design level. The STAR experiment has successfully stepped back from its legacy Geant 3 model (AGI [2]) and implemented a front-end abstract geometry modeling language (AgML), based on an XML syntax enriched with mathematical expressions. AgML allows STAR to leverage recent developments in simulation and detector description, provides a clear path for the seamless integration of future technologies, and enables us to support both the past and ongoing experimental program of STAR by maintaining a consistent single-source description of the detector geometry across its decade-long lifespan. It is complemented by parsers and a C++ class library which enables the automated conversion of the original source code to AgML, supports export back to the STAR original format (regression testing), and creates the concrete ROOT/TGeo geometry model used in our reconstruction framework. In this talk we present our approach, design and experience and will demonstrate physical consistency between the original AGI and new AgML geometry models and discuss its integration within the STAR framework.

[1] R. Brun et al, "The ROOT Geometry Package", Nucl. Instrum. Meth. A502:676-680,2003.
[2] A. Artamonov et al, "DICE-95", internal note ATLAS-SOFT/95-14, CERN, 1995.

Distributed Processing and Analysis on Grids and Clouds / 499

Employing peer-to-peer software distribution in ALICE Grid Services to enable opportunistic use of OSG resources

Authors: Iwona Sakrejda[None]; Jeff Porter[1]

Co-authors: Costin Grigoras [2]; Federico Carminati [2]; Latchezar Betev [2]; Pablo Saiz [2]

[1] *Lawrence Berkeley National Lab. (US)*

[2] *CERN*

Corresponding Authors: rjporter@lbl.gov, isakrejda@lbl.gov

The ALICE Grid infrastructure is based on AliEn, a lightweight open source framework built on Web Services and a Distributed Agent Model in which job agents are submitted onto a grid site to prepare the environment and pull work from a central task queue located at CERN. In the standard configuration, each ALICE grid site supports an ALICE-specific VO box as a single point of contact between the site and the ALICE central services. VO box processes monitor site utilization and job requests (ClusterMonitor), monitor dynamic job and site properties (MonaLisa), perform job agent submission (CE) and deploy job-specific software (PackMan). In particular, requiring a VO box at each site simplifies deployment of job software, done onto a shared file system at the site, and adds redundancy to the overall Grid system. ALICE offline computing, however, has also implemented a peer-to-peer method (based on BitTorrent) for downloading job software directly onto each worker node as needed. By utilizing both this peer-to-peer deployment model and job agent submission onto remote Open Science Grid (OSG) Compute Elements, we are able relax the site VO box requirement and run jobs opportunistically on independent OSG resources from a single VO box. In this paper, we will describe the implementation of the peer-to-peer method and the full configuration of the setup. We will cover the deployment of this configuration at NERSC utilizing a VO box at PDSF and an OSG gatekeeper on the NERSC Carver system from which we can directly compare the performance to that of a standard ALICE Grid installation. We will also describe our experience with wider deployments.

Poster Session / 500

The WNoDeS Cache Manager, an efficient method to self-allocate virtual resources

Authors: Claudio Grandi[1]; Daniele Andreotti[2]; Davide Salomoni[3]; Francesco Pepe[4]; Gianni Dalla Torre[None]

[1] *INFN - Bologna*

[2] *Universita e INFN (IT)*

[3] *Istituto Nazionale Fisica Nucleare (IT)*

[4] *Sezione di Bologna (INFN)-Universita e INFN*

Corresponding Authors: daniele.andreotti@cern.ch, gianni.dalla.torre@cern.ch, davide.salomoni@cnaf.infn.it

The WNoDeS software framework (http://web.infn.it/wnodes) uses virtualization technologies to provide access to a common pool of dynamically allocated computing resources. WNoDeS can process batch and interactive requests, in local, Grid and Cloud environments.

A problem of resource allocation in Cloud environments is the time it takes to actually allocate the resource and make it available to customers. WNoDeS, for its resource scheduling and allocation tasks, uses an underlying batch system. The time to allocate resources is therefore dictated by this batch system, by its configuration, and by site-specific peculiarities.

Interactive access to resources is supplied by WNoDeS in two ways: a Web-based application, and a command line interface, called the Virtual Interactive Pool (VIP). Both of them interact with a central component, the WNoDeS Cache Manager (CM), providing the actual resource allocation.

The CM, the topic of this poster, has been designed to speed up the allocation of virtual machines, be they requested via the Web-based application of via the command-line interface. The CM keeps a cache of ready-to-use virtual machines, matches them to user requirements and makes them readily available for consumption.

We will show how the adoption of the WNoDeS CM speeds up considerably resource allocation, thereby significantly improving user experience in the self-allocation of virtual nodes used for Cloud computing, or for the self-instantiation of machine pools used, for example, for physics analysis. We will then show how the CM is being used in the WNoDeS installation at the INFN Tier-1 located in Bologna.

Collaborative tools / 501

Code and papers: computing publication patterns in the LHC era

Authors: Maria Grazia Pia[1]; Tullio Basaglia[2]

[1] *Universita e INFN (IT)*

[2] *CERN*

Corresponding Author: maria.grazia.pia@cern.ch

Publications in scholarly journals establish the body of knowledge deriving from scientific research; they also play a fundamental role in the career path of scientists and in the evaluation criteria of funding agencies.

This presentation reviews the evolution of computing-oriented publications in HEP following the start of operation of LHC. Quantitative analyses are illustrated, which document the production of scholarly papers on computing-related topics by HEP experiments and core tools projects (including distributed computing R&D), and the citations they receive. Several scientometric indicators are analyzed to characterize the role of computing in HEP literature. Distinctive features of scholarly publication production in the software-oriented and hardware-oriented experimental HEP communities are highlighted. Current patterns and trends are compared to the situation in previous generations' HEP experiments at LEP, Tevatron and B-factories.

The results of this scientometric analysis document objectively the contribution of computing to HEP scientific production and technology transfer to other fields. They also provide elements for discussion about how to more effectively promote the role played by computing-oriented research in high energy physics.

Poster Session / 502

DPM: Future-proof storage

Author: Ricardo Brito Da Rocha[1]

Co-authors: Alejandro Alvarez Ayllon [2]; Alexandre Beche [3]; Fabrizio Furano [1]; Jean-Philippe Baud [1]; Oliver Keeble [1]

[1] *CERN*

[2] *University of Cadiz*

[3] *SUPINFO International University (FR)*

Corresponding Author: ricardo.rocha@cern.ch

The Disk Pool Manager (DPM) is a lightweight solution for grid enabled disk storage management. Operated at more than 240 sites it has the widest distribution of all grid storage solutions in the WLCG infrastructure.

It provides an easy way to manage and configure disk pools, and exposes multiple interfaces for data access (rfio, xroot, nfs, gridftp and http/dav) and control (srm). During the last year we have been working on providing stable, high performant data access to our storage system using standard protocols, while extending the storage management functionality and adapting both configuration and deployment procedures to reuse commonly used building blocks.

In this contribution we cover in detail the extensive evaluation we have performed of our new HTTP/WebDAV and NFS 4.1 frontends, in terms of functionality and performance. We summarize the issues we faced and the solutions we developed to turn them into valid alternatives to the existing grid protocols - namely the additional work required to provide multi-stream transfers for high performance wide area access, support for third party copies, credential delegation or the required changes in the experiment and fabric management frameworks and tools.

We describe new functionality that has been added to ease system administration, such as different filesystem weights and a faster disk drain, and new configuration and monitoring solutions based on the industry standards Puppet and Nagios. Finally, we explain some of the internal changes we had to do in the DPM architecture to better handle the additional load from the analysis use cases.

Poster Session / 503

The DESY Grid Lab in action

Authors: Dmitry Ozerov[1]; Patrick Fuhrmann[2]; Yves Kemp[1]

[1] *Deutsches Elektronen-Synchrotron (DE)*

[2] *DESY*

Corresponding Authors: yves.kemp@cern.ch, dmitry.ozerov@cern.ch

Since mid of 2010, the Scientific Computing department at DESY is operating a storage and data access evaluation laboratory, DESY Grid Lab, equipped with 256 CPU cores, and about 80 Tbytes of data distributed among 5 servers and interconnected via up to 10-GiGE links.

The system has been dimensioned to be equivalent to the size of a medium WLCG Tier 2 center to provide commonly exploitable results.

It is integrated in the WLCG Grid infrastructure and as such can execute standard LHC experiment jobs including the hammercloud framework.

During its 18 month of operation, results of data access performance evaluations, especially in the context of NFS 4.1/pNFS but not limited to that, have been presented at various conference and workshops.

The goal of this poster is to give a comprehensive summary the collected findings and to attract the attention of the storage expert community, as the DESY Grid Lab is open to everyone to evaluate the performance of their application(s) against various protocols provided by the Grid Lab environment.

Distributed Processing and Analysis on Grids and Clouds / 504

Connecting multiple clouds and mixing real and virtual resources via the open source WNoDeS framework

Authors: Alessandro Italiano[1]; Andrea Chierici[1]; Daniele Andreotti[2]; Davide Salomoni[3]; Elisabetta Ronchieri[2]; Giacinto Donvito[4]; Gianni Dalla Torre[None]; Vincenzo Spinoso[2]

[1] *INFN-CNAF*

[2] *Universita e INFN (IT)*

[3] *Istituto Nazionale Fisica Nucleare (IT)*

[4] *INFN-Bari*

Corresponding Authors: giacinto.donvito@ba.infn.it, alessandro.italiano@cnaf.infn.it, davide.salomoni@cnaf.infn.it

In this paper we present the latest developments introduced in the WNoDeS framework (http://web.infn.it/wnodes); we will in particular describe inter-cloud connectivity, support for multiple batch systems, and co-existence of virtual and real environments on a single hardware.

Specific effort has been dedicated to the work needed to deploy a "multi-sites" WNoDeS installation. The goal is to give end users the possibility to submit requests for resources using cloud interfaces on several sites in a transparent way. To this extent, we will show how we have exploited already existing and deployed middleware within the framework of the IGI (Italian Grid Initiative) and EGI (European Grid Infrastructure) services. In this context, we will also describe the developments that have taken place in order to have the possibility to dynamically exploit public cloud services like Amazon EC2. The latter gives WNoDeS the capability to serve, for example, part of the user requests through external computing resources when needed, so that peak requests can be honored.

We will then describe the work done to add support for open source batch systems like Torque/Maui to WNoDeS. This makes WNoDeS a fully open source product and gives the possibility to smaller sites as well (where often there is no possibility to run commercial batch systems) to install it and exploit its features. We will also describe recent WNoDeS I/O optimizations, showing results of performance tests executed using Torque as batch system and Lustre as a distributed file system.

Finally, starting from the consideration that not all tasks are equally suited to run on virtual environments, we will describe a novel feature added to WNoDeS, allowing the possibility to use the same hardware to run both virtual machines and real jobs (i.e., jobs running on the bare metal and not in a virtualized environment). This increases flexibility and may optimize the usage of available resources. In particular, we will describe tests performed in order to show how this feature can help in fulfilling requests for "whole-node jobs" (which are becoming increasingly popular in the HEP community), for efficient analysis support, and for GPU-based resources (which are typically not easily amenable to be virtualized).

Poster Session / 505

Campus Grids Bring Additional Computational Resources to HEP Researchers

Author: Derek John Weitzel[1]

Co-authors: Brian Bockelman [2]; Dan Fraser [3]; David Swanson [2]

[1] *University of Nebraska (US)*

[2] *University of Nebraska*

[3] *Argonne National Laboratory*

Corresponding Author: derek.weitzel@cern.ch

It is common at research institutions to maintain multiple clusters that represent different owners or generations of hardware, or that fulfill different needs and policies. Many of these clusters are consistently under utilized while researchers on campus could greatly benefit from these unused capabilities. By leveraging principles from the Open Science Grid it is now possible to utilize these resources by forming a lightweight Campus Grids. The Campus Grids framework enables jobs that are submitted to one cluster to overflow, when necessary, to other clusters within the campus using

whatever authentication mechanisms are available on campus. This framework is currently being used on several campuses to run HEP and other science jobs. Further, the framework has in some cases been expanded beyond the campus boundary by bridging campus grids into a regional grid, and can even be used to integrate resources from a national cyberinfrastructure such as the Open Science Grid. This poster will highlight 18 months of operational experiences creating campus grids in the US, and the different campus configurations that have successfully utilized the campus grid infrastructure.

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Poster Session / **506**

MPI support in the DIRAC Pilot Job Workload Management System

Authors: Andrei Tsaregorodtsev[1]; Vanessa Hamar[2]

[1] *Universite d'Aix - Marseille II (FR)*

[2] *CPPM-IN2P3-CNRS*

Corresponding Author: atsareg@in2p3.fr

Parallel job execution in the grid environment using MPI technology presents a number of challenges for the sites providing this support. Multiple flavors of the MPI libraries, shared working directories required by certain applications, special settings for the batch systems make the MPI support difficult for the site managers. On the other hand the workload management systems with pilot jobs became ubiquitous although the support for the MPI applications in the pilot frameworks was not available. This support was recently added in the DIRAC Project in the context of the GISELA Latin American Grid. Special services for dynamic allocation of virtual computer pools on the grid sites were developed in order to deploy MPI rings corresponding to the requirements of the jobs in the central task queue of the DIRAC Workload Management systems. The required MPI software is installed automatically by the pilot agents using user space file system techniques. The same technique is used to emulate shared working directories for the parallel MPI processes. This makes it possible to execute MPI jobs even on the sites not supporting them officially. Reusing so constructed MPI rings for execution of a series of parallel jobs increases dramatically their efficiency and turnaround.

In this contribution we will describe the design and implementation of the DIRAC MPI Service as well as its support for various types of the MPI libraries. Advantages of coupling the MPI support with the pilot frameworks will be outlined and examples of usage with real applications will be presented.

Poster Session / **507**

An automated virtual testing environment for StoRM

Authors: Elisabetta Ronchieri[1]; Michele Dibenedetto[2]; Riccardo Zappi[3]

[1] *Universita e INFN (IT)*

[2] *INFN CNAF*

[3] *INFN*

Corresponding Authors: luca.dellagnello@cnaf.infn.it, elisabetta.ronchieri@cnaf.infn.it, michele.dibenedetto@cnaf.infn.it, riccardo.zappi@cnaf.infn.it

An automated virtual test environment is a way to improve testing, validation and verification activities when several deployment scenarios must be considered. Such solution has been designed and developed at INFN CNAF to improve software development life cycle and to optimize the deployment of a new software release (sometimes delayed for the difficulties
met during the installation and configuration of a testing environment). Its main characteristic is the set-up of a virtual environment where the downloading and installation of the packages, the configuration of the services and the tests execution are orchestrated by a proper deployment and test engine fed with a pre-built configuration file. Running automated tests by using virtual environment follows the same process as running automated tests with physical environment, allowing much more testing flexibility, dynamic on-demand resources provisioning, greatly simplifying the use of the test-bed, and optimizing the usage of test-bed machines. Virtual images, with the required Operating System version, including host certificate when necessary, are provisioned automatically before running tests. This virtual test environment is being used by the StoRM team for testing, validation and verification activities: however, it is not peculiar for StoRM and can be easily customized for other software team who just needs to provide configuration file and virtual images for the deployment and test engine. In this paper, we describe the design and development of an automated virtual test environment, and we present its usage during the StoRM development life ycle.

Summary:

StoRM, one of the SRM implementation, is a multi-service software subject to intense testing, validation and verification activities in order to guarantee high-quality services. Its characteristics of being usable on different file systems (such as IBM GPFS, Lustre and POSIX), and of supporting several transfer protocols (like gsiFTP, file and HTTPS) raise the need of StoRM to be validated on a variety of deployment scenarios. Moreover, the StoRM distributed nature requires that services are tested with multiple machines.
With this in mind, manual testing is extremely time consuming, inconsistent to be effective, error prone and inaccurate to cover all cases. While automating manual testing can, however, be very expensive in order to maintain a set of scripts that describes a given set of tests. The usage of virtualization technology can contribute to making automating testing accurate, efficient, reliable and cost effective.
Here lies the need of an automated virtual test environment in order to improve testing, validation and verification activities when several deployment scenarios must be considered. The running tests belong to several categories (like system, functionality and stress) and are all automitized. The StoRM sofware has been considered to validate this solution.

Poster Session / 508

Creating Dynamic Virtual Networks for network isolation to support Cloud computing and virtualization in large computing centers

Authors: Davide Salomoni[1]; Marco Caberletti[1]

[1] *Istituto Nazionale Fisica Nucleare (IT)*

Corresponding Author: davide.salomoni@cnaf.infn.it

The extensive use of virtualization technologies in cloud environments has created the need for a new network access layer residing on hosts and connecting the various Virtual Machines (VMs). In fact, massive deployment of virtualized environments imposes requirements on networking for which traditional models are not well suited. For example, hundreds of users issuing cloud requests for which full access (i.e., including root privileges) to VMs are requested, typically requires the definition of network separation at layer 2 through the use of virtual LANs (VLANs). However, in large computing centers, due for example to the number of installed network switches, to their characteristics, or to their heterogeneity, the dynamic (or even static) definition of many VLANs is often impractical or simply not possible.

In this paper, we present a solution to the problem of creating dynamic virtual networks based on the use of the Generic Routing Protocol (GRE). GRE is used to encapsulate VM traffic so that the configuration of the physical network switches doesn't have to change. In particular, we describe how this solution can be used to tackle problems such as dynamic network isolation and mobility of VMs across hosts or even sites. We will then show how this solution has been integrated in the WNoDeS framework (http://web.infn.it/wnodes) and tested in the WNoDeS installation at the INFN Tier-1, presenting performance metrics and an analysis of the scalability of the system.

Poster Session / **509**

Improving Geant4 multi-core's performance and usability

Author: Xin Dong[1]

Co-authors: Andrzej Nowak [2]; Gene Cooperman [3]; John Apostolakis [2]; Makoto Asai [4]; Sverre Jarp [2]

[1] *Northeastern University*

[2] *CERN*

[3] *Unknown*

[4] *SLAC National Accelerator Laboratory (US)*

Corresponding Authors: xin.dong@cern.ch, john.apostolakis@cern.ch

We report on the progress of the multi-core versions of Geant4, including multi-process and multi-threaded Geant4.

The performance of the multi-threaded version of Geant4 has been measured, identifying an overhead compared with the sequential version of 20-30%. We explain the reasons, and the improvements introduced to reduce this overhead.

In addition we have improved the design of a few key classes of Geant4 were revised in order to simplify the design and improve the implementation of multi-threaded and reduce the memory footprint of multi-process Geant4.

The process for adapting user applications to Geant4 multi-threaded has been documented and streamlined. Most applications can be adapted within 1-2 working days. Tools to verify that the results of a multi-threaded application are exactly equal to the sequential version are under development.

In addition we present an overview of the test coverage undertaken to ensure that the Geant4 multi-threaded are fully compatible with the sequential version.

Student? Enter 'yes'. See http://goo.gl/MVv53:

No

Poster Session / **511**

Tier2 procurements experiences in the UK

Author: Alessandra Forti[1]

[1] *University of Manchester (GB)*

Corresponding Author: alessandra.forti@cern.ch

In this paper we will describe primarily the experience of going through an EU procurement. We will describe what a PQQ (Pre-Qualification Questionaire) is and some of the requirments for vendors such as ITIL and PRINCE2 project management qualifications. We will describe how the technical part was written including requirements from the main users and the university logistic requirements to the importance of including the acceptance tests and what are considered valid acceptance tests.

Poster Session / **514**

Collaborative development. Case study of the development of flexible monitoring applications

Authors: Alessandro Di Girolamo[1]; Andrea Sciaba[1]; David Tuckett[1]; Ivan Antoniev Dzhunov[2]; Jaroslava Schovancova[3]; Jose Flix Molina[4]; Julia Andreeva[1]; Lukasz Kokoszkiewicz[1]; Michal Maciej Nowotka[5]; Pablo Saiz[1]; Pekka Karhula[1]; Peter Kreuzer[6]

[1] *CERN*

[2] *University of Sofia*

[3] *Acad. of Sciences of the Czech Rep. (CZ)*

[4] *Centro de Investigaciones Energ. Medioambientales y Tecn. - (ES*

[5] *Warsaw University of Technology (PL)*

[6] *Rheinisch-Westfaelische Tech. Hoch. (DE)*

Corresponding Author: pablo.saiz@cern.ch

Collaborative development proved to be a key of the success of the Dashboard Site Status Board (SSB) which is heavily used by ATLAS and CMS for the computing shifts and site commissioning activities.
The Dashboard Site Status Board (SSB) is an application that enables Virtual Organisation (VO) administrators to monitor the status of distributed sites. The selection, significance and combination of monitoring metrics falls clearly in the domain of the administrators, depending not only on the VO but also on the role of the administrator. Therefore, the key requirement for SSB is that it be highly customisable, providing an intuitive yet powerful interface to define and visualise various monitoring metrics.
We present SSB as an example of a development process typified by very close collaboration between developers and the user community. The collaboration extends beyond the customisation of metrics and views to the development of new functionality and visualisations. SSB Developers and VO administrators cooperate closely to ensure that requirements are met and, wherever possible, new functionality is pushed upstream to benefit all users and VOs.
The contribution covers the evolution of SSB over recent years to satisfy diverse use cases through this collaborative development process.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Poster Session / **515**

Disk to Disk network transfers at 100 Gb/s using a handful of servers

Authors: Artur Jerzy Barczyk[1]; Azher Mughal[2]; Colin Leavett-Brown[3]; Dianne Lee[3]; Don McWilliam[4]; Harvey Newman[1]; Ian Gable[5]; Iosif Legrand[1]; Kim Lewall[3]; Marilyn Hay[4]; Ramiro Voicu[1]; Randy Sobie[5]; Thomas Tam[6]; Yvan Savard[3]; sandor Rozsa[2]

[1] *California Institute of Technology (US)*

[2] *California Institute of Technology (CALTECH)*

[3] *University of Victoria*

[4] *BCNet*

[5] *University of Victoria (CA)*

[6] *CANARIE*

Corresponding Authors: artur.barczyk@cern.ch, igable@uvic.ca

For the Super Computing 2011 conference in Seattle, Washington, a 100 Gb/s connection was established between the California Institute of Technology conference booth and the University of Victoria.

A small team performed disk to disk data transfers between the two sites nearing 100 Gb/s, using only a small set of properly
configured transfer servers equipped with SSD drives.The circuit was established over the BCnet, CANARIE and SCinet (the SuperComputing conference network) using network equipment dedicated to the demonstration.The end-sites' setups involved a mix of 10GE and 40GE technologies. Three servers were equipped with PCIe v3, with a theoretical throughput per network interface of 40Gb/s.

We examine the design of the circuit and the work necessary to establish it. The technical hardware design of each end system is described. We discuss the transfer tools, disk configurations, and monitoring tools used in the test with particular emphasis on disk to disk throughput.

We review the final test results in addition to discussing the practical problems encountered and overcome during the demonstration.

Finally, we evaluate the performance obtained, both with regard to the 100Gb/s WAN circuit as well as end-system and LAN setups, and discuss potential application as a high-rate data access system, and/or caching front-end to a large conventional storage system.

Distributed Processing and Analysis on Grids and Clouds / 516

AliEn: ALICE Environment on the GRID

Author: Pablo Saiz[1]

Co-authors: Alina Gabriela Grigoras [1]; Almudena Del Rocio Montiel Gonzalez [2]; Armenuhi Abramyan [3]; Costin Grigoras [1]; Dushyant Goyal [4]; Federico Carminati [1]; Jeff Porter [5]; Jianlin Zhu [6]; Latchezar Betev [1]; Narine Manukyan [7]; Stefano Bagnasco [8]; Steffen Schreiner [9]; Subho Sankar Banerjee [10]

[1] *CERN*

[2] *GSI - Gesellschaft fuer Schwerionen forschung (GSI)*

[3] *Yerevan Physics Institute*

[4] *LNM Institute of Information Technology (IN)*

[5] *Lawrence Berkeley National Lab. (US)*

[6] *Central China Normal University (CN)*

[7] *A.I. Alikhanyan National Scientific Laboratory (AM)*

[8] *I.N.F.N. TORINO*

[9] *Technische Universitaet Darmstadt (DE)*

[10] *LNM Institute of Information Technology*

Corresponding Author: pablo.saiz@cern.ch

AliEn is the GRID middleware used by the ALICE collaboration. It provides all the components that are needed to manage the distributed resources. AliEn is used for all the computing workflows

of the experiment: Montecarlo production, data replication and reconstruction and organixed or chaotic user analysis. Moreover, AliEn is also being used by other experiments like PANDA and CBM.

The main components of AliEn are a centralized file and metadata catalogue, a job execution model and file replication model. These three components have been evolving over the last 10 years to make sure that the satisfy the computing requirements of the experiment, which keep increasing every year.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Summary:

This contribution will present the current status of the AliEn components, with special emphasis on the latest development, in particular data handling. We will also outline the future development plans.

Poster Session / 517

Sysematic analysis of job failures at a Tier-2, and mitgation of the causes.

Author: Stuart Purdie[1]

Co-authors: David Crooks [1]; Mark Mitchell [1]; Sam Skipsey [2]

[1] *University of Glasgow*

[2] *University of Glasgow / GridPP*

Corresponding Author: stuart.purdie@glasgow.ac.uk

Failure is endemic in the Grid world - as with any large, distributed computer system, at some point things will go wrong. Wether it is down to a problem with hardware, network or software, the shear size of a production Grid requires operation under the assumption that some of the jobs will fail. Some of those are anavoidable (e.g. network loss during data staging), some are preventable but only within engineering tradeoffs (e.g. uninteruptable power supplies on worker nodes), and some are fully preventable (e.g. software problems).

It is not clear that the current level of job failures is at the minimum level. We have been logging all failed jobs, and classifying then according to how and why they failed. Some work have been invested into automated systems to collated job failure information from various sources, and these will be presented.

This work reports on that data, and supplies an analysis, as quantitative as possible, on what would need to have been different to have prevented those jobs from failing.

Some subtlty lies in the definition of a 'failed job'. From the perspective of an end user, any job that does not do what the user wants can be considered failed, but that is not the most useful definition for an infrastructure provider. However, it is useful to track such cases in order to provide a comparison point to the infrastrucutre caused job failures. Not all problems in the infrastructure result in user visibale job failures; for example a problem in a batch system schedular can result in no jobs being started at for some point, which is only visible as reduced throughput. These were tracked, but can't be quantified in the same scale as user visible job failure modes.

Clearly, there are some cases of job failure that it is not within the capability of a site administrator to resolve - If user code divides by zero, no aspect of site administration can resolve that. However, there are many other sources of problems other than that. Of particular interest are jobs that report a failure the first time, but succeded on a re-submission. Jobs falling into that description include (but are not limited to) all the jobs where something transient went wrong at a site. This is the class

of failures which it is within the capability of a site manager to reduce to zero, which is the long term goal of this work.

Deep analysis of these probalemtic cases is required, in order to determine the underlying causes, and further work is needed in order to prevent the problems from re-occuring. One early observation was that, in general, the slower the rate of failures detected, the more likely a root fix was to be found. A detailed analysis of this will be reported, but if this observation holds then it suggests that there might be net super-linear improvements in reliability from this sort of work.

In cases where it appeared that the root cause was located in some component, although no precise reason could be found, an alternative for that component was sourced, and compared with the original. Where possible, such as with a computing element, these were run in parralell. One such case, were no alternative could be found, we wrote an alternative implementation of the BLAH parser for CREAM, and compared it's stability to the supplied one.

Many problems that have been identified come down to either a hardware problem, or some interaction between multiple components. A survey of these will be presented. For hardware problems, an estimated cost to prevent the problem from being user visible will be given, and for hte problems with interacting components a series of reccomendations can be given.

Overall, this work is of importance in ensuring the imporved user experience on the Grid, and also a reduction in the manpower required to operate a site. Although the analysis might not be feasable at smaller sites, the lessons learned from this work will be directly applicable to many sites, and should contribute to the smooth running of the Grid in the future.

Summary:

Some degree of job failure is inevitable, but this work assumes, and demonstrates, that the level of failure is not at the minimum level. Sources of job failures are found, analysed and mitigations and solutions given. In some cases this is as simple as configuration, but includes cases where new software components have been written, in addition to monitoring and analysis tools.

Some discussion is included over the concept of a failed job, and to what extent these can be eliminated - it clearly being infeasble to prevent any jobs fromfailing.

Overall, this work is of importance in ensuring the imporved user experience on the Grid, and also a reduction in the manpower required to operate a site. Although the analysis might not be feasable at smaller sites, the lessons learned from this work will be directly applicable to many sites, and should contribute to the smooth running of the Grid in the future.

Poster Session / 518

AliEn Extreme JobBrokering

Authors: Alina Gabriela Grigoras[1]; Almudena Del Rocio Montiel Gonzalez[2]; Armenuhi Abramyan[3]; Costin Grigoras[1]; Dushyant Goyal[4]; Federico Carminati[1]; Jianlin Zhu[5]; Latchezar Betev[1]; Narine Manukyan[6]; Pablo Saiz[1]; Stefano Bagnasco[7]; Steffen Schreiner[8]; Subho Sankar Banerjee[9]

[1] *CERN*

[2] *GSI - Gesellschaft fuer Schwerionen forschung (GSI)*

[3] *Yerevan Physics Institute*

[4] *LNM Institute of Information Technology (IN)*

[5] *Central China Normal University (CN)*

[6] *A.I. Alikhanyan National Scientific Laboratory (AM)*

[7] *I.N.F.N. TORINO*

[8] *Technische Universitaet Darmstadt (DE)*

[9] *LNM Institute of Information Technology*

Corresponding Author: pablo.saiz@cern.ch

The AliEn workload management system is based on a central job queue wich holds all tasks that have to be executed. The job brokering model itself is based on pilot jobs: the system submits generic pilots to the compuiting centres batch gateways, and the assignment of a real job is done only when the pilot wakes up on the worker node. The model facilitates
a flexible fair share user job distribution. This job model has proven stable and reliable over the past years and has surviced well the first two years of LHC operation with very little changes.
Nonetheless there are several areas where the model can be pushed to the next level: most notably in the area of 'just in time' job and data assignment, where the decisions will be based on data closeness (relaxed locality) and the data which already has been processed. This methods
will have significant efficiency enhancement effect for end user analysis tasks.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Summary:

This contribution will describe the new mechanism implemented in AliEn for the job splitting and matchmaking. It will also show how the users
benefit from such a model.

Poster Session / 519

Investigation of many-core scalability of the track reconstruction in the CBM experiment

Author: Pavel Kisel[1]

Co-authors: Igor Kulakov [2]; Sergey Baginyan [1]; Victor Ivanov [1]

[1] *JINR*

[2] *GSI*

Corresponding Author: i.kulakov@gsi.de

Search for particle trajectories is a basis of the on-line event reconstruction in the heavy-ion CBM experiment (FAIR/GSI, Darmstadt, Germany). The experimental requirements are very high, namely: up to 10^7 collisions per second, up to 1000 charged particles produced in a central collision, a non-homogeneous magnetic field, about 85% of the additional background combinatorial measurements in the detector, full on-line event reconstruction and selection. This requires use of the full potential of modern many-core CPU/GPU architectures.

The Cellular Automaton (CA) method is one of the most efficient methods of searching for charged particles trajectories. The implementation of the CA algorithm in the CBM experiment is well optimized with respect to time consumptions, the calculations are carried out in parallel with use of parallelism at the level of data, as well as at the level of cores. We present a detailed description of the algorithm realization and results of the many-core scalability tests on a server at the Laboratory of Information Technologies (JINR, Dubna, Russia) with 2 Intel Xeon E5640 CPUs (in total 8 physical or 16 logical cores). The track reconstruction efficiency, the speed of the algorithm and its scalability with respect to number of cores are presented in detail. Using the Nvidia GPU card as an accelerator is also discussed.

Poster Session / 520

Experiment Dashboard - a generic, scalable solution for monitoring of the LHC computing activities, distributed sites and services

Authors: Daniel Dieguez Arias[1]; David Tuckett[2]; Edward Karavakis[2]; Gunnar Ro[2]; Ivan Antoniev Dzhunov[3]; Julia Andreeva[2]; Laura Sargsyan[4]; Lukasz Kokoszkiewicz[2]; Mattia Cinquilli[5]; Michael John Kenyon[6]; Michal Maciej Nowotka[7]; Pablo Saiz[2]; Pekka Karhula[2]

[1] *University of Vigo (ES)*

[2] *CERN*

[3] *University of Sofia*

[4] *A.I. Alikhanyan National Scientific Laboratory (AM)*

[5] *Univ. of California San Diego (US)*

[6] *University of Glasgow*

[7] *Warsaw University of Technology (PL)*

Corresponding Author: pablo.saiz@cern.ch

The Experiment Dashboard system provides common solutions for monitoring job processing, data transfers and site/service usability. Over the last seven years, it proved to play a crucial role in the monitoring of the LHC computing activities, distributed sites and services.
It has been one of the key elements during the commissioning of the distributed computing systems of the LHC experiments.

The first years of data taking represented a serious test for Experiment Dashboard in terms of functionality, scalability and performance. And given that the usage of the Experiment Dashboard applications has been steadily increasing over time, it can be asserted that all the objectives were fully accomplished.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Summary:

The presentation will describe the different applications within the Experiment Dashboard, putting special emphasis in the recent evolutions and improvements regarding performance and scalability.

It will cover as well the usage of the system by the experiments during data taking.

Poster Session / 521

New developments on visualization drivers in Geant4 software toolkit

Author: Laurent Garnier[1]

[1] *LAL-IN2P3-CNRS*

Corresponding Author: garnier@lal.in2p3.fr

New developments on visualization drivers in Geant4 software toolkit

Summary:

The Geant4 software toolkit simulates the passage of particles through matter. Visualization is a key part of it. Geant4 is used in many application domains including high energy, nuclear and accelerator

physics, and in medical and space science. We have developed several visualization drivers, such as OpenInventor, HepRep, DAWN, VRML, RayTracer, ASCIITree, gMocren and OpenGL to fit the various requirements of each domain.

During the last 3 years, the OpenGL suite of visualization drivers has been significantly improved by adding a lot of functionalities, in particular a new OpenGL Qt driver. Qt is a free and well-known toolkit available on all platforms, including Windows, that has enabled us to offer Geant4 visualization that has the same look and feel on all systems. Geant4 release 9.5 integrates the latest improvements in the OpenGL and Qt viewer, including faster first time rendering, integration of multiple visualization frames and the user interface into same window, making posters (thanks to gl2ps), a new Qt viewer components help tree and volume tree, easy creation of videos, "free hand" rotation mode, etc. Thanks to the cmake build system, compiling Geant4 with the Qt viewer is simple. Also use and choice of user interface and visualization drivers has been simplified in all examples.

Poster Session / **522**

Fermi Offline Software: The Pros and Cons of Beg, Borrow, and Steal

Author: Heather Kelly[1]

[1] *SLAC National Accelerator Laboratory*

Corresponding Author: heather625@gmail.com

The Fermi Gamma-ray Observatory, including the Large Area Telescope (LAT), was launched June 11, 2008. We are a relatively small collaboration, with a maximum of 25 software developers in our heyday. Within the LAT collaboration we support Redhat Linux, Windows, and are moving towards Mac OS as well for offline simulation, reconstruction and analysis tools. Early on it was decided to use one software system to run our simulations as well as ultimately handle the event processing for real data. We leveraged many existing HEP external libraries (Geant4, Gaudi Framework, ROOT, CLHEP, CMT) to ease the burden on our developers. This strategy of re-using existing software helped us pull together our system quickly and test during our beam tests and data challenges. Now, after launch, we are in a new phase of the project, where we must move forward to support modern operating systems and compilers to get us through the life of the mission. This means upgrading our external libraries as well, which are not under our direct control. Meanwhile, it is crucial to our production system that we carefully orchestrate all upgrades to insure stability. An additional hurtle is that our number of active developers has dwindled dramatically. Many of those left are Windows developers reliant on the Visual Studio development environment, while our user base and production system depend on our Linux distributions. There have been a number of lessons learned, with undoubtedly more to come.

Poster Session / **524**

The CC1 project - Cloud Computing for Science

Authors: Mariusz Witek[1]; Milosz Zdybal[2]

Co-authors: Andrzej Olszewski [1]; Bartlomiej Henryk Zabinski [1]; Danielowski Krzysztof [3]; Grzymkowski Rafal [3]; Janusz Chwastowski [1]; Maciej Kruk [2]; Maciej Piotr Nabozny [4]; Piotr Wojcik [2]; Przemyslaw Syktus [3]; Tomasz Sosnicki [2]; Tomasz Wojton [3]

[1] *Polish Academy of Sciences (PL)*

[2] *Institute of Nuclear Physics*

[3] *Institute of Nuclear Physics PAN*

[4] *Cracow University of Technology*

Corresponding Author: mariusz.witek@cern.ch

Providing computer infrastructure to end-users in an efficient and user-friendly way was always a big challenge in the IT market. "Cloud computing"is an approach that addresses these issues and recently it has been gaining more and more popularity. A well designed Cloud Computing system gives elasticity in resources allocation and allows for efficient usage of computing infrastructure. The underlying virtualization technology and the self-service type of access are the two key features that make the software independent of the specific hardware and enable a significant decrease in system administration effort.
The growing popularity of cloud computing led to the appearance of many open source systems offering cloud computing environments, such as Eucalyptus, OpenNebula, Nimbus or OpenStack. These solutions make it possible to construct a computing cloud in a relatively short time and do not require a deep understanding of virtualization techniques and network administration. The main drawback of using this type of toolkits is a difficulty in customization to special needs. A significant effort is needed to implement some non standard features.
The CC1 Project started in 2009. The proposed solution is based on Libvirt, a lower level virtualization toolkit. It provides a full set of VM management actions on a single node. For the top layer the PYTHON programming language was chosen as it ensures fast development environment (interpreter) and offers a number of useful modules. At present most of the required features are being implemented:

- custom web-based user interface,
- automatic creation of virtual clusters ("farms") with preconfigured batch system,
- groups of users with the ability to share resources,
- permanent virtual storage volumes that can be mounted to a VM,
- distributed structure –federation of clusters running as a uniform cloud,
- quota for user resources,
- monitoring and accounting system.

The CC1 system consists of two main layers. The top element of the system is called Cloud Manager (CLM). It receives calls from user interfaces (web browser based interface or EC2 interface) and passes commands to Cluster Managers (CMs). Cluster Manager, running on each individual cluster, handles all low-level operations required to control virtual machines.
The project is close to reach its first milestone. The production quality system (Private Cloud) will be made available to researchers of IFJ PAN at the beginning of 2012. The next step is to build federated systems with universities that expressed their interest in the project.
The CC1 system will be described and the experience from the firsts months of its usage will be reported.

Computer Facilities, Production Grids and Networking / 525

Experience with HEP analysis on mounted filesystems

Authors: Dmitry Ozerov[1]; Martin Gasthuber[1]; Patrick Fuhrmann[2]; Yves Kemp[1]

[1] *Deutsches Elektronen-Synchrotron (DE)*

[2] *DESY*

Corresponding Authors: martin.gasthuber@cern.ch, patrick.fuhrmann@desy.de, yves.kemp@cern.ch, dmitry.ozerov@cern.ch

We present results on different approaches on mounted filesystems in use or under investigation at DESY.

dCache, established since long as a storage system for physics data has implemented the NFS v4.1/pNFS protocol. New performance results will be shown with the most current version of the dCache server. In addition to the native usage of the mounted filesystem in a LAN environment, the results are given for the performance of the dCache NFS v4.1/pNFS in WAN case.

Several commercial vendors are currently in alpha or beta phase of adding the NFS v4.1/pNFS protocol to their storage appliances. We will test some of these vendor solutions for their readiness for HEP analysis.

DESY has recently purchased an IBM Sonas system. We will present the result of a thourough performance evaluation using the native protocols NFS (v3 or v4) and GPFS.

As the emphasis is on the usability for end user analysis, we will use latest ROOT versions and current end user analysis code for benchmark scenarios.

Poster Session / **526**

Fermilab Multicore and GPU-Accelerated Clusters for Lattice QCD

Author: Don Holmgren[1]

Co-authors: Amitoj Singh [1]; James Simone [1]; Nirmal Seenu [1]

[1] *Fermilab*

Corresponding Author: djholm@fnal.gov

As part of the DOE LQCD-ext project, Fermilab designs, deploys, and operates dedicated high performance clusters for parallel lattice QCD (LQCD) computations. Multicore processors benefit LQCD simulations and have contributed to the steady decrease in price/performance for these calculations over the last decade. We currently operate two large conventional clusters, the older with over 6,800 AMD Barcelona cores distributed across 8-core systems interconnected with DDR Infiniband, and the newer with over 13,400 AMD Magny-Cours cores distributed across 32-core systems interconnected with QDR Infiniband. We will describe the design and operations of these clusters, as well as their performance and the benchmarking data that were used to select the hardware and the techniques used to handle their NUMA architecture.

We will also discuss the design, operations, and performance of a GPU-accelerated cluster that Fermilab will deploy in late November 2011. This cluster will have 152 nVidia Fermi GPUs distributed across 76 servers coupled with QDR Infiniband. In the last several years GPUs have been used to increase the throughput of some LQCD simulations by over tenfold compared with conventional hardware of the same cost. These LQCD codes have evolved from using single GPUs to using multiple GPUs within a server, and now to multiple GPUs distributed across a cluster. The primary goal of this cluster's design is the optimization of large GPU-count LQCD simulations.

Poster Session / **527**

Application of Bayesian inference with usage of Markov Chain Monte Carlo to a many-parameter fit of ep-collider HERA data to extract the proton structure functions.

Author: Julia Grebenyuk[1]

Co-authors: Allen Caldwell [2]; Daniel Kollar [2]; Frederik Beaujean [3]; Kevin Alexander Kroeninger [4]; Shabnaz Pashapouralamdari [4]

[1] *DESY*

[2] *Max Planck Institute*

[3] *Max Planck Institute for Physics*

[4] *Georg-August-Universitaet Goettingen (DE)*

Corresponding Author: julia.grebenyuk@desy.de

A many-parameter fit to extract the the proton structure functions from the Neutral Current deep-inelastic scattering cross sections, measured from the data collected at HERA ep-collider with the ZEUS detector, will be presented. The structure functions F_2 and F_L are extracted as a function of Bjorken-x in bins of virtuality Q2. The fit is performed with the Bayesian Analysis Toolkit (BAT) which allows the investigation of complex statistical problems encountered in Bayesian inference. It is realised with the use of Markov Chain Monte Carlo and gives access to the full posterior probability distribution, which enables straightforward parameter estimation and uncertainty propagation. 78 parameters are fit in total, 54 central F_2 and F_L values, 3 normalisations, and 21 systematic uncertainties included as nuisance parameters. The resulting posterior distributions showed correlations between parameters. The experience gained from this analysis will be discussed.

Poster Session / **528**

Evolution of Data Acquisition in the PHENIX Experiment

Author: John Haggerty[1]

[1] *Brookhaven National Laboratory*

Corresponding Author: haggerty@bnl.gov

The architecture of the PHENIX data acquisition system will be reviewed, and how it has evolved in 12 years of operation. Custom data acquisition hardware front end modules embedded in the detector operated in a largely inaccessible experimental hall have been controlled and monitored, and a large software infrastructure has been developed around remote objects which are controlled from a relatively small number of applications. A number of different networking technologies are used to control, acquire, and record data from a dozen different detectors. The challenges of adapting new detectors and increasing performance while continuing to operate the experiment will be discussed.

Online Computing / **529**

The NOvA Data Acquistion System: A highly distributed, synchronized, continuous readout system for a long baseline neutrino experiment

Author: Andrew Norman[1]

[1] *Fermilab*

Corresponding Author: anorman@fnal.gov

The NOvA experiment at Fermi National Accelerator Lab, has been designed and optimized to perform a suite of measurements critical to our understanding of the neutrino's properties, their oscillations and their interactions. NOvA presents a unique set of data acquisition and computing challenges due to the immense size of the detectors, the data volumes that are generated through the continuous digitization of the frontend systems, and the need to buffer the full data stream to allow for highly asynchronous triggering and extraction of physics events. These challenges are compounded by the stringent timing and synchronization requirements that are placed on the acquisition systems by the need to precisely correlate information between the accelerator complex and the remote detector locations.

The NOvA Data Acquisition system has been designed and built to meets these challenges. The system utilizes a highly modular, novel acquisition and event building scheme, which has been deployed on a large hierarchical organization of both custom and commodity computing. This system

is coupled with highly optimized software and firmware to aggregate over 350,000 continuously sampled, readout channels into arbitrary length time windows, which are buffered in large compute farms for analysis. These windows allow the experiment to perform not only standard event-trigger based data analysis, but also permit non-traditional searches for macroscopic phenomena, such as core collapse supernova, whose time scales and event signatures are uncharacteristic of the ranges that are addressable by most high energy physics experiments.

In this paper we cover the overall design of the NOvA DAQ system and its capabilities. We present results from its initial deployment with our Near Detector in a surface configuration, and from its deployment on the first blocks of our far detector. We also discuss the planned upgrades to this system that expand its capabilities and allow the experiment to address other topics in high energy physics.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Summary:

The NOvA collaboration describes the designs, capabilities and performance of their data acquisition and timing system. Emphasis is placed on the roll of distributed acquisition systems for event building and the roll of large CPU farms for data driven triggering. The challenges of performing absolute time synchronization across massive detectors and between remote sites are addressed.

Poster Session / **530**

NOvA Event Building, Buffering, and Filtering within the DAQ System

Authors: Andrew Norman[1]; Marc Paterno[1]; Ronald Rechenmacher[2]

Co-author: Jim Kowalkowski [3]

[1] *Fermilab*

[2] *Fermi National Accelerator Lab. (Fermilab)*

[3] *Fermi National Accelerator Laboratory (FNAL)*

Corresponding Authors: anorman@fnal.gov, jbk@fnal.gov

The NOvA experiment at Fermi National Accelerator Lab features a free running, continuous readout system without dead time, which collects and buffers time-continuous data from over 350,000 readout channels. The raw data must be searched to correlate it with beam spill events from the NuMI beam facility. They are also analyzed in real-time to identify event topologies of interest. The analysis results then are fed back into the experiment's triggering systems to form data-driven decisions.

The NOvA event building layer is designed to continuously process data at full sampling rate from the NOvA detectors using commodity networking and computing equipment. For the far detector, custom designed upstream hardware delivers fragments of data in 5ms time slices to more than 180 multicore commodity buffering nodes using standard gigabit ethernet switches. The fragments are assembled into full time-synchronized windows and indexed to allow for efficient search and delivery to downstream applications upon receipt of positive trigger broadcasts. The system can sustain a raw data input rate of greater than 2GB/s and buffer in excess of 20 seconds worth of data. The buffer management software feeds all raw time slices into an event processing framework that is common with the offline production. This framework runs analysis modules that examine each slice and generates positive trigger decisions in real-time causing windows of raw data to be transferred to downstream subsystems using global trigger messages.

This paper will describe the system architecture and software processes constructed to perform the buffering and filtering operations within the NOvA DAQ system. The paper will also describe the advantages of the data driven triggering model and the physics potential that it provides.

Summary:

The NOvA DAQ event building, buffering, and filtering system are described in this paper.

Collaborative tools / 531

Electronic Collaboration Logbook

Author: Igor Mandrichenko[1]

Co-authors: Federica Moscato [2]; Margherita Vittone-Wiersma [2]; Suzanne Gysin [2]; Vladimir Podstavkov [2]

[1] *Fermilab*

[2] *FNAL*

Corresponding Author: ivm@fnal.gov

In HEP, scientific research is performed by large collaborations of organizations and individuals. Log book of a scientific collaboration is important part of the collaboration record. Often, it contains experimental data.
At FNAL, we developed an Electronic Collaboration Logbook (ECL) application which is used by about 20 different collaborations, experiments and groups at FNAL. ECL is the latest iteration of the project formerly known as Control Room Logbook (CRL).

We have been working on mobile (IOS and Android) clients for ECL.

We will present history, current status and future plans of the project, as well as design, implementation and support solutions made by the project.

532

Building, distributing and running big software projects on MacOSX... There is an app for that!

Author: Giulio Eulisse[1]

Co-authors: Andreas Pfeiffer [2]; Lassi Tuura [1]; Peter Elmer [3]; Shahzad Muzaffar [4]

[1] *Fermi National Accelerator Lab. (US)*

[2] *CERN*

[3] *Princeton University (US)*

[4] *NORTHEASTERN UNIVERSITY*

Corresponding Author: giulio.eulisse@cern.ch

We present CMS' experience in porting its full offline software stack to MacOSX. In the first part we will focus on the system level issues encountered while doing the port, in particular with respect to the different behavior of the compiler and linker in handling common symbols. In the second part we present our progress with an alternative approach of distributing large software projects which is in line with the click and run installation common to most of the mac applications.

Poster Session / 533

The NOvA Timing System: A system for synchronizing a Long Baseline Neutrino Experiment.

Author: Andrew Norman[1]

Co-authors: Gregory Deuerling [1]; Neal Wilcer [1]; Rick Kwarciany [1]

[1] *Fermilab*

Corresponding Author: anorman@fnal.gov

The NOvA experiment at Fermi National Accelerator Lab, uses a sophisticated timing distribution system to perform synchronization of more than 12,000 front end readout and data acquisition systems at both the near detector and accelerator complex located at Fermilab and at the far detector located 810km away at Ash River, MN. This global synchronization is performed to an absolute clock time with a system wide variation of less than 16ns, which allows for the direct comparisons of detector data with the accelerator beam spills. The system accomplishes this through the use of high precision GPS receivers, which are decoded by custom hardware to both determine the absolute wall clock times and propagate them to the readout systems. This custom hardware is able to perform detector wide calibrations for the paths to each frontend readout system that take into account the signal propagation and retransmission delays. The resulting system ensure that the electronics clock registers tick in perfect unison regardless of their position on the faces of the 220ft long, five story tall far detector. The paper will cover the details of the timing system, its characteristic and performance as demonstrated on the NOvA detectors. The paper will also describe the prospects for performing specific measurements, that rely on high precision timing, related to the properties of the neutrino.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Summary:

The NOvA data acquisition groups presents the designs, capabilities and performance of a new system for performing timing and synchronization with the next generation of massive, distributed readouts for neutrino detectors. Results from the initial deployment of the system will be presented.

Distributed Processing and Analysis on Grids and Clouds / 534

Evaluation of benefits of a three tier data model for WLCG analysis

Authors: Dmitry Ozerov[1]; Patrick Fuhrmann[2]

[1] *Deutsches Elektronen-Synchrotron (DE)*

[2] *DESY*

Corresponding Authors: patrick.fuhrmann@desy.de, dmitry.ozerov@cern.ch

One of the most crucial requirement for online storage is the fast and efficient access to data.

Although smart client side caching often compensates for discomforts like latencies and server disk congestion, spinning disks, with their limited ability to serve multi stream random access patterns, seem to be the cause of most of the observed inefficiencies.

With the appearance of the different variants of solid state disks (SSD), this deficiency could be overcome, however, replacing the entire experiment data repositories by SSDs is not feasible in the foreseeable future.

Moreover, spinning disks are still appropriate media for controlled streaming applications.

Assuming a deployment of a mixture of media, like spinning disks, SSDs and tape, at a site, the authors argue for the introduction of a three tier media structure within a single storage system with automatic transitions, based on usage patterns, in contrast to interlinking and maintaining different mediatypes in different systems with external procedures taking care of proper data placement.

The feasibility of the suggested approach is studied, using the analysis of access logs of the DESY WLCG Tier II storage elements, hosting the largest part of the data
to be analyzed by the CMS and ATLAS Collaborations.

Finally we will report on a prototype implementing of the three tier media structure into dCache, a storage technology widely used in WLCG.

Poster Session / 535

Applying formal verification methods to experiment triggers

Author: Swain John[1]

Co-authors: Gene Cooperman [2]; Pete Manolios [3]; Thomas Paul [2]

[1] *Noreastern University*

[2] *Northeastern University*

[3] *Northeatern University*

Modern particle physics experiments use short pieces of
code called “triggers” in order to make rapid decisions about whether incoming
data represents potentially interesting physics or not. Such decisions are
irreversible and while it is extremely important that they are
made correctly, little use has been made in the community of formal verification
methodology.

The goal of this research is to determine both a restricted language for writing software triggers and a formal verification methodology that can be learned by non-experts in time similar to what they would invest to learn a new programming language. That methodology will also include a more formal specification for the software triggers.

We describe domain-specific languages for preparing software triggers and their properties. These languages will be specified in ACL2, a theorem proving system that was awarded the ACM Software System Award, and has been used in industry to prove some of the most complex theorems ever proved about commercial systems. We develop libraries of definitions and theorems to significantly automate the testing, validation, and verification of software triggers.

This work will provide a bridge technology to allow physicists to
produce reliable software triggers. The burden of using a restricted
language is not large, since the software trigger programs are short.
Hence, the formal verification methodology and the need for a restricted
programming language represent a modest burden to the physicists.
This burden is considered a bargain in exchange for the greater software
reliability in triggers that will be the outcome of this work.

Summary:

We apply methods of formal verifications to trigger software as a means to ensure higher software quality for short, critical pieces of code.

Poster Session / **536**

Double Chooz Physical Environment Monitoring System

Author: Chang Pi-Jung[1]

Co-authors: David McKee [1]; Glenn Horton-Smith [2]; Janet Conrad [3]; Lindley Winslow [3]

[1] *Kansas University*

[2] *Kansas Univeristy*

[3] *MIT*

The Double Chooz experiment will measure reactor antineutrino flux from two detectors with a relative normalization uncertainty less than 0.6%. The Double Chooz physical environment monitoring system records conditions of the experiment's environment to ensure the stability of the active volume and readout electronics. The system monitors temperatures in the detector liquids, temperatures and voltages in electronics, experimental hall environmental conditions, magnetic field, radon concentrations in the air, and phototube high voltages. The system scans all channels automatically, stores data in a common database, and warns of changes in the detector's physical environment. The design and performance of the Double Chooz physical environment monitoring system is presented.

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Poster Session / **537**

The Double Chooz Online Monitor Framework

Author: Tomoyuki Konno[1]

Co-author: Arthur Franke [2]

[1] *Tokyo Tech*

[2] *Columbia University*

Corresponding Author: ajf2140@columbia.edu

The Double Chooz reactor antineutrino experiment employs a network-distributed DAQ divided among a number of computing nodes on a Local Area Network. The Double Chooz Online Monitor Framework has been developed to provide short-timescale, real-time monitoring of multiple distributed DAQ subsystems and serve diagnostic information to multiple clients. Monitor information can be accessed via a Java GUI or a web-based interface implemented in HTML5 with Google Web Toolkit. An automatic email notification system has been developed. The Online Monitor Framework has been designed to be scalable to other experiments with similar network-distributed DAQ systems, with DAQ components implemented in multiple programming languages."

Student? Enter 'yes'. See http://goo.gl/MVv53:

yes

Poster Session / 538

The Double Chooz Data Streaming

Author: Alberto Remoto[1]

Co-author: Kazuhiro Terao [2]

[1] *APC/in2p3*

[2] *MIT*

Corresponding Author: kazuhiro@mit.edu

The Double Chooz reactor anti-neutrino experiment have developed a automatised system for data streaming from the detector site to the different nodes of data analysis in Europe, Japan and USA. The system both propagates and triggers the processing of data as it goes through low level data analysis. All operations (propagation and processing) are tracked file-wise in real time using DB (MySQL based technology). Web interfaces have been also developed to allow any member of the experiment (like data-taking shifters) to follow in real time the status and location of data as it streams to their final location where higher level data analysis starts.

Poster Session / 539

Automating Linux Deployment with Cobbler

Author: James Pryor[1]

Co-author: Jason Alexander Smith [2]

[1] *Brookhaven National Laboratory*

[2] *Brookhaven National Laboratory (US)*

Cobbler is a network-based Linux installation server, which, via a choice of web or CLI tools, glues together PXE/DHCP/TFTP and automates many associated deployment tasks. It empowers a facility's systems administrators to write scriptable and modular code, which can pilot the OS installation routine to proceed unattended and automatically, even across heterogeneous hardware. These tools make it so system administrators do not have to move between various commands and applications and then and tweak machine specific configuration files when deploying the OS. Network deployments can be configured for new and re-installations via PXE, media-based over-the-network installations, and virtualized installations that support Xen, qemu, KVM, and some variants of VMware. Cobbler supports most large Linux distributions, including Red Hat Enterprise Linux, Scientific Linux, Centos, SuSE Enterprise Linux, Fedora, Debian, and Ubuntu.

Here at the RACF at Brookhaven National Laboratory, we had been deploying network PXE installs for many years, and needed a centralized and scalable solution for Linux deployments. This paper will discuss the ways in which we now use Cobbler for nearly all Linux OS deployments, both physical and virtualized. We will discuss our existing Cobbler setup, and the details of how we use Cobbler to deploy variants of the RHEL OS to our 250+ infrastructure servers.

Online Computing / 540

The DoubleChooz DAQ systems.

Author: camillo mariani[1]

Co-authors: Arthur Franke [2]; Matt Toups [2]

[1] *Columbia university*

[2] *Columbia University*

Corresponding Authors: mht2114@columbia.edu, mariani@nevis.columbia.edu

The Double Chooz (DC) reactor anti-neutrino experiment consists of a neutrino detector and a large area Outer Veto detector. A custom data-acquisition (DAQ) system written in Ada language for all the sub-detector in the neutrino detector systems and a generic object oriented data acquisition system for the Outer Veto detector were developed. Generic object-oriented programming was also used to support several electronic systems to be readout providing a simple interface for any new electronics to be added given its dedicated driver. The core electronics of the experiment is based on FADC electronics (500MHz sampling rate), therefore a data-reduction scheme has been implemented to reduce the data volume per trigger. A dynamic data-format was created to allow dynamic reduction of each trigger before data is written to disk. The decision is based on low level information that determines the relevance of each trigger. The DAQ is structured internally into two types of processors: several read-out processors reading and processing data at crate level and one event-builder processor collecting data from all crates and further processing data before writing into disk. An average rate of 40MB/s data output can be handled without dead-time. The Outer Veto DAQ uses a token-passing scheme to read out five daisy chains of multi-anode PMTs via five USB interfaces. The maximum rate that this system can handle is up to 40MB/s limited only by the USB2.0 throughput. A dynamic data reducer is implemented to reduce the amount of data written to disk. An object-oriented event builder process was developed to collect the data from the multiple USB streams and merge them into a single data stream ordered in time. A separate object oriented code was developed to merge the information coming from the neutrino and Outer Veto DAQ in a single event based on time information.
The internal architecture and functioning of the Double Chooz DAQs as well as examples of performance and other capabilities will be described.

Poster Session / **541**

Comparative Investigation of Shared Filesystems for the LHCb Online Cluster

Author: Vijay Kartik Subbiah[1]

Co-author: Niko Neufeld [1]

[1] *CERN*

Corresponding Authors: vijay.kartik@cern.ch, niko.neufeld@cern.ch

This paper describes the investigative study undertaken to evaluate shared filesystem performance and suitability in the LHCb Online environment. Particular focus is given to the measurements and field tests designed and performed on an in-house AFS setup, and related comparisons with NFSv3 and pNFS are presented. The motivation for the investigation and the test setup arises from the need to serve common user-space like home directories, experiment software and control areas, and clustered log areas. Since the operational requirements on such user-space are stringent in terms of read-write operations (in frequency and access speed) and unobtrusive data relocation, test results are presented with emphasis on file-level performance, stability and "high-availability"of the shared filesystems. Use-cases specific to the experiment operation in LHCb, including the specific handling of shared filesystems served to a cluster of 1500 diskless nodes, are described. Issues of authentication token expiry are explicitly addressed, keeping in mind long-running analysis jobs on the Online cluster. In addition, quantitative test results are also presented with alternatives including pNFS, which is now being seen as an increasingly viable option for shared filesystems in many medium to large networks. Comparative measurements of filesystem performance benchmarks are presented, which are seen to be used as reference for decisions on potential migration of the current storage solution deployed in the LHCb online cluster.

Poster Session / 542

Shibboleth Federation in BNL

Author: Mizuki Karasawa[1]

Co-author: John Steven De Stefano Jr [2]

[1] *BNL*

[2] *Brookhaven National Laboratory (US)*

Corresponding Author: mizuki@bnl.gov

In BNL, we are planning to establish a federation with different organizations by using a SSO technology - Shibboleth. It provides the underlying mechanism for leveraging institutional authentication and exchanging of user attributes for authorization. This framework will allow us to collaborate not only with organizations inside of BNL but institutions/organizations outside of BNL to be able to access RACF resources (and vice versa) with ease of user account management, reduce the need for per-service account provisioning. Meanwhile reduce the opportunities for account to be compromised from security's point of view and provides users convenience to access any number of resources while singing on only once. We currently replaced our existing SSO with Shibboleth successfully in RACF, we also collaborated with Scifed and CERN and tested the framework. We foresee the federation will happen in a real world in near future.

Poster Session / 543

RooFit - a data modeling language for physics analysis

Author: Wouter Verkerke[1]

[1] *NIKHEF (NL)*

Corresponding Author: verkerke@nikhef.nl

RooFit is a library of C++ classes that facilitate data modeling in the ROOT environment. Mathematical concepts such as variables, (probability density) functions and integrals are represented as C++ objects. The package provides a flexible framework for building complex fit models through classes that mimic math operators. For all constructed models RooFit provides a concise yet powerful interface for fitting, plotting and toy Monte Carlo generation as well as sophisticated tools to manage large scale projects. RooFit has been used in countless published B-factory results and more recently also at the LHC. We will review recent developments such as the ability to persist models in ROOT files in container classes, which provides the basis for several new concepts and techniques. This enables the concept of digital publishing of analytical likelihood functions with an arbitrary number of parameters, which in turn is the basis of the RooStats statistical tools that combine Higgs analysis channels of ATLAS and CMS. Combined models can be technically trivially constructed exploiting the editing and introspection methods provided by RooFit modeling classes. Persistability also enable streaming of tasks to other computers which facilitates parallelized calculation of computing intensive problems.

Poster Session / 544

The Double Chooz Online System

Author: Matthew Toups[1]

Co-author: Anatael Cabrera [2]

[1] *Columbia University*

[2] *APC/in2p3*

Corresponding Author: mht2114@columbia.edu

The Double Chooz experiment searches for reactor neutrino oscillations at the Chooz nuclear power plant. A client/server model is used to coordinate actions among several online systems over TCP/IP sockets. A central run control server synchronizes data-taking among two independent data acquisition (DAQ) systems via a common communication protocol and state machine definition. Calibration subsystems are controlled by a calibration server which establishes a connection to one the DAQs. The data are written to buffer disks in the experimental hall and diagnostic information is generated using fast reconstructions. An automatic data transfer system tracks and manages the relocation of the data files to permanent, offsite storage. Various hardware-level and environmental information are monitored by a slow-control system. The DAQ, slow control, and data transfer systems send information to a centralized monitoring server from which diagnostic information can be visualized via a java-based graphical user interface. Since access to the experimental site is restricted, all systems have been designed to operate remotely and employ robust exception-handling techniques.

Poster Session / **547**

the INFN Tier-1

Author: luca dell'agnello[1]

[1] *infn*

Corresponding Author: luca.dellagnello@cnaf.infn.it

INFN-CNAF is the central computing facility of INFN: it is the Italian Tier-1 for the experiments at LHC, but also one of the main Italian computing facilities of several other experiments such as BABAR, CDF, SuperB, Virgo, Argo, AMS, Pamela, MAGIC, Auger etc..
Currently there is an installed CPU capacity of 100,000 HS06, a net disk capacity of 9 PB and an equivalent amount of tape storage (these figures are going to be increased in the first half of 2012 respectively to 125,000 HS06, 12 PB and 18 PB).
More than 50,000 computing jobs are executed daily on the farm, managed by LSF, accessing the storage, managed by GPFS, with an aggregate bandwidth up to several GB/s. The access to the storage system from the farm is direct through the file protocol. The interconnection of the computing resources and the data storage
is based on 10 Gbps technology.
The disk-servers and the storage systems are connected through a Storage Area Network allowing a complete flexibility and easiness of management; dedicated disk-servers are connected, also via teh SAN, to the tape library.
The INFN Tier-1 is connected to the other centers via 3x10 Gbps links (to be upgraded next year), including the LHCOPN and to the LHCONE.
In this paper we show the main results of our center after 2 full years of run of LHC.

Poster Session / **548**

NUMA memory hierarchies experience with multithreaded HEP software at CERN openlab

Authors: Alfio Lazzaro[1]; Andrzej Nowak[2]; Julien Leduc[None]; Sverre Jarp[2]

[1] *CERN openlab*

[2] *CERN*

Corresponding Author: julien.leduc@cern.ch

Newer generations of processors come with no increase in their clock frequency, and the same is true for memory chips. In order to achieve more performance, the core count is getting higher, and to feed all the cores on a chip with instructions and data, the number of memory channels must follow the same trend.
Non Uniform Memory Access (NUMA) architecture allowed the CPU manufacturers to reduce nicely the impact of memory subsystem bottlenecks, but, in turn, this solution introduces a cost at the application level. This paper describes our practical experience with the typical CPU servers currently available to the HEP community, based on work with NUMA systems at CERN openlab. We provide the latest measurements of the different NUMA implementations from AMD and Intel, as well as NUMA consequences on some parallelized HEP codes.

Poster Session / 549

UK efforts to improve networking rates on WAN transfers

Authors: Alessandra Forti[1]; Brian Davies[2]; Sam Skipsey[3]

[1] *University of Manchester (GB)*

[2] *STFC RALLCG2 Tier1*

[3] *University of Glasgow / GridPP*

Corresponding Author: alessandra.forti@cern.ch

In this paper we will present the efforts carried out in the UK to fix the WAN transfers problem highlighted by the ATLAS sonar tests. We will present the work done at site level, the monitoring tools at local level on the machines (ifstat, tcpdump, netstat...), between sites (iperf) and at FTS level monitoring. We will describe the effort to setup a mini-mesh to simplify the sonar tests setup separating the FTS layer from the transport layer. We will present the improvements and optimizations carried out by sites on kernel parameters and bonding setup at machine and switch level the attempt to understand differences between sites rates despite similar setup and most of all the asymmetric traffic obesrved at some sites. We will also describe work on the FTS channel organisation and configuration that has been carried out in parallel and talk about the opportunity and plans to upgrade the site insfrastructure to 10Gbps.

Poster Session / 550

Improving the quality of EMI Releases by leveraging the EMI Testing Infrastructure

Authors: Danilo Dongiovanni[1]; Doina Cristina Aiftimiei[2]

Co-authors: Alberto Di Meglio [3]; Andrea Ceccanti [2]; Francesco Giacomini [4]

[1] *INFN*

[2] *Istituto Nazionale Fisica Nucleare (IT)*

[3] *CERN*

[4] *INFN CNAF*

Corresponding Authors: aiftim@pd.infn.it, danilo.dongiovanni@cnaf.infn.it

What is an EMI Release? What is its life-cycle? How is its quality assured through a continuous integration and large scale acceptance testing? These are the main questions that this article will

answer, by presenting the EMI release management process with emphasis on the role played by the Testing Infrastructure in improving the quality of the middleware provided by the project.

The European Middleware Initiative (EMI) is a close collaboration of four major European technology providers: ARC, gLite, UNICORE and dCache. Its main objective is to deliver a consolidated set of components for deployment in EGI (as part of the Unified Middleware Distribution, UMD), PRACE and other DCIs. The harmonized set of EMI components thus enables the interoperability and integration between Grids. EMI aims at creating an effective environment that satisfies the requirements of the scientific communities relying on it.

The EMI distribution is organized in periodic major releases whose development and maintenance follow a 5-phase yearly cycle: i) requirements collection and analysis; ii) development and test planning; iii) software development, testing and certification; iv) release certification and validation and v) release and maintenance.

In this article we present in detail the implementation of operational and infrastructural resources supporting the certification and validation phase of the release. The main goal of this phase is to harmonize into a single release the strongly inter-dependent products coming from various development teams through parallel certification paths. To achieve this goal the continuous integration and large scale acceptance testing performed on the EMI Testing Infrastructure plays a key role. The purpose of this infrastructure is to provide a system where both the production and the release candidate product versions are deployed. On this system inter-component testing by different product team testers can concurrently take place. The Testing Infrastructure is also continuosly monitored through Nagios and exposed both to automatic testing and to usage by volunteer end-users. Furthermore the infrastructure size is increased with resources made available by volunteer end-users that are interested in implementing production-like deployments or specific test scenarios.

Computer Facilities, Production Grids and Networking / 551

The DYNES Instrument: A Description and Overview

Authors: Eric Boyd[1]; Harvey Newman[2]; Jason Zurawski[1]; Paul Sheldon[3]; Shawn Mc Kee[4]

Co-authors: Aaron Brown ; Alan Tackett [5]; Artur Jerzy Barczyk [2]; Azher Mughal [6]; Ben Meekhof [7]; Bobby Brown [8]; Jeff Boote [1]; Mathew Binkley [8]; Ramiro Voicu [2]; Robert Ball [4]; Stephen Wolff [1]; Tom Lehman [9]; Xi Yang [9]; sandor Rozsa [6]

[1] *Internet2*

[2] *California Institute of Technology (US)*

[3] *Vanderbilt University (US)*

[4] *University of Michigan (US)*

[5] *VANDERBILT UNIVERSITY*

[6] *California Institute of Technology (CALTECH)*

[7] *University of Michigan*

[8] *Vanderbilt*

[9] *ISI*

Corresponding Authors: zurawski@internet2.edu, shawn.mckee@cern.ch

Scientific innovation continues to increase requirements for the computing and networking infrastructures of the world. Collaborative partners, instrumentation, storage, and processing facilities are often geographically and topologically separated, as is the case with LHC virtual organizations. These separations challenge the technology used to interconnect available resources, often delivered by Research and Education (R&E) networking providers, and leads to complications in the overall process of end-to-end data management.

Capacity and traffic management are key concerns of R&E network operators; a delicate balance is required to serve both long-lived, high capacity network flows, as well as more traditional end-user activities. The advent of dynamic circuit services, a technology that enables the creation of variable duration, guaranteed bandwidth networking channels, allows for the efficient use of common network infrastructures. These gains are seen particularly in locations where overall capacity is scarce compared to the (sustained peak) needs of user communities. Related efforts, including those of the LHCOPN operations group and the emerging LHCONE project, may take advantage of available resources by designating specific network activities as a "high priority", allowing reservation of dedicated bandwidth or optimizing for deadline scheduling and predicable delivery patterns.

This paper presents the DYNES instrument, an NSF funded cyberinfrastructure project designed to facilitate end-to-end dynamic circuit services. This combination of hardware and software innovation is being deployed across R&E networks in the United States at selected end-sites located on University Campuses. DYNES is peering with international efforts in other countries using similar solutions, and is increasing the reach of this emerging technology. This global data movement solution could be integrated into computing paradigms such as cloud and grid computing platforms, and through the use of APIs can be integrated into existing data movement software.

Summary:

A description and overview of a distributed virtual instrument for the dynamic creation of end-to-end circuits to support distributed scientific collaborations.

Poster Session / **552**

Lessons Learned from Migrating Open Science Grid to a Native Packaging Software Distribution

Author: Alain Roy[1]

[1] *University of Wisconsin-Madison*

Corresponding Author: roy@cs.wisc.edu

We recently completed a significant transition in the Open Science Grid in which we moved our software distribution mechanism from the useful but niche system called Pacman to a community-standard native packaged system (RPM). Despite the challenges, this migration was both useful and necessary. In this paper we explore some of the lessons learned during this transition, lessons which we believe are valuable not only for software distribution and packaging, but for software engineering in a distributed computing environment where reliability is critical. We discuss the benefits found in moving to a community standard, including the abilities to reuse existing packaging, to donate existing packaging back to the community, and to leverage existing skills in the community. We describe our approach to testing in which we test our software against multiple versions of the OS, including pre-releases of the OS, in order to find surprises before our users do. We also discuss our large-scale evaluation testing and community testing, which are essential for both quality and community acceptance. Finally, we discuss how we can share our expertise, tools, and perhaps even testing infrastructure to benefit other communities building distributed software.

Poster Session / **553**

Using CernVM and EDGI to transparently use desktop resources for LHC related computation in a traditional data grid context

Authors: Anders Waananen[1]; Chrulle Soettrup[2]

[1] *Niels Bohr Institute*

[2] *University of Copenhagen (DK)*

Corresponding Authors: waananen@nbi.dk, christian.soettrup@cern.ch

Modern HEP related calculations have traditionally been beyond the capabilities of donated desktop machines, particularly because of complex deployment of the needed software.
The popularization of efficient virtual machine technology and in particular the CernVM appliance, that allows for only the needed subset of the ATLAS software environment to be dynamically downloaded, has made such computation feasible.
We report on the results of integrating the ARC Grid Middleware and the EDGI infrastructure with Virtual Machine enabled BOINC for running ATLAS related computations on publicly donated desktop machines. The approach allows the user to transparently benefit from both private and public desktop grid resources as well as standard ARC based resources.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Poster Session / 554

A Fully Software-based Online Test-bench for LHCb

Author: Vijay Kartik Subbiah[1]

Co-authors: Beat Jost [1]; Clara Gaspar [1]; Eric Van Herwijnen [1]; Jean-Christophe Garnier [1]; Markus Frank [1]; Niko Neufeld [1]

[1] *CERN*

Corresponding Authors: niko.neufeld@cern.ch, vijay.kartik@cern.ch

This contribution describes the design and development of a fully software-based Online test-bench for LHCb. The current "Full Experiment System Test"(FEST) is a programmable data injector with a test setup that runs using a simulated data acquisition (DAQ) chain. FEST is heavily used in LHCb by different groups, and thus the motivation for complete software emulation of the test-bench is to enable running parallel tests by sharing resources and removing all dependency on detector-related hardware. The Timing and Fast Control (TFC) used in FEST, originally in hardware, is now completely replaced with a software module that emulates the behaviour of sending trigger decisions to the test-bench. The design of a monolithic structure encompassing the former data injector and the developed TFC emulator is described in detail, and the advantages of disconnecting the test-bench from the hardware are discussed. In particular, design details for emulating (user-defined) trigger decisions and multiple event data flags are shown, and the advantages of having complete control over every stage of the DAQ chain are demonstrated using measurements made on different configurations of the test-bench done through software. The integration of the full software emulator in the run-control of FEST completes the switch. The installation of a "development"computing farm is also shown in brief, which allows the allocation of resources to different groups so that instances of FEST may be run concurrently. Additionally, the setup allows the High Level Trigger algorithms to be benchmarked on different hardware with controlled input, due to the complete software emulation of all data. Results of performance tests on the independent test setup are presented to underline the data throughput levels in the DAQ chain and the utility of this modified design and implementation.

Poster Session / 555

Present and future of Identity Management in Open Science Grid

Authors: Jim Basney[1]; Mine Altunay[2]; Von Welch[3]

[1] *University of Illinois*

[2] *Fermi National Accelerator Laboratory*

[3] *Indiana University*

Corresponding Author: maltunay@fnal.gov

Identity management infrastructure has been a key work area for the Open Science Grid (OSG) security team for the past year. The progress of web-based authentication protocols such as openID, SAML, and scientific federations such as InCommon, prompted OSG to evaluate its current identity management infrastructure and propose ways to incorporate new protocols and methods.

For the couple of years we have been working on documenting and then improving the user experience. Our identity roadmap has evolved. In one next step we are working closely with the ESNET DOE Grids CA group on the future for the main US x509 CA. We are now starting a pilot project using a commercial CA, DigiCert CA, which is currently undergoing IGTF accreditation for user and host certificates. We then plan to investigate multiple back end services from a new OSG front-end service to enable integration and support of the new technologies and mechanisms needed by our users. We are participating in the cross-agency MAGIC forum to look at a high level at some of these futures.

In this talk, we will present our ideas and activities and speculate on the future.

Student? Enter 'yes'. See http://goo.gl/MVv53:

No

Poster Session / 556

The future Tier1, sharing a dedicated computing environment

Author: Jos Van Wezel[1]

[1] *KIT - Karlsruhe Institute of Technology (DE)*

Corresponding Author: jos.van.wezel@cern.ch

Resources of large computer centers used in physics computing today. are optimised for the WLCG framework and reflect the typical data access footprint of reconstruction and analysis. A traditional Tier 1 centre like GridKa at KIT hosts thousands of hosts and many PetaBytes of disk and tape storage that is used mostly by a single community. The required size as well as the intrinsic difficulties that came with the deployment of a new infrastructure over the last ten years made it necessary to build a dedicated environment which has been optimised for a small number of middleware stacks. Although computing demands will grow during the lifetime of the LHC, the relative requirements, compared to the growth of hardware capabilities are diminishing. The hardware for computing in high energy physics will soon fit in a few racks, 4 TB disks and 128 core systems are at the horizon as is 100 Gbit networking and the economy of scale is no longer working for LHC computing.

The next generation of dedicated clusters at Tier 1 sites will be of moderate size and could be integrated with environments that are build for fast rising computing consumers such as biology or climatology. This raises the question what characteristics of LHC computing and the attached middleware must be preserved or looked after in these hosted environments. GridKa at KIT, the German Tier1 WLCG centre is actively investigating the possibility to expand the use of its infrastructure beyond physics computing, currently its main task. Alternatively physics jobs may run at different environments at KIT or beyond. This is possible with the use of virtual machines and cloud computing, techniques that allow common environments as well as transparent job migration. Either way is a promising direction and may well lead to more efficient use of ever limited computing, storage and networking resources for example because it allows temporary grow-as-you need expansion of the compute and disk farms which can than be planned with less spare capacity.
The presentation discusses the advantages and disadvantages of a shared computing environment at an LHC Tier 1 in technical as well as organisational sense, the changes to specific hard and software required to enable sharing and the steps to be taken at KIT to enable sharing of or with T1 resources.

Poster Session / **557**

Data transfer test with 100 Gb network

Author: haifeng pi[1]

[1] *CMS*

Corresponding Author: hpi@physics.ucsd.edu

As part of the Advanced Networking Initiative (ANI) of ESnet, we exercise
a prototype 100Gb network infrastructure for data transfer and processing for
OSG HEP applications.
We present results of these tests.

Poster Session / **558**

lcsim: An integrated detector simulation, reconstruction and analysis environment

Authors: Jeremy McCormick[1]; Norman Anthony Graf[2]

[1] *Unknown*

[2] *SLAC National Accelerator Laboratory (US)*

Corresponding Author: norman.graf@slac.stanford.edu

slic: Geant4 simulation program

As the complexity and resolution of particle detectors increases,
the need for detailed simulation of the experimental setup also
increases. Designing experiments requires efficient tools to
simulate detector response and optimize the cost-benefit ratio
for design options. We have developed efficient and flexible
tools for detailed physics and detector response simulation which
builds on the power of the Geant4 toolkit but frees the end user
from any C++ coding. The primary goal has been to develop a software
toolkit and computing infrastructure to allow physicists from
universities and labs to quickly and easily contribute to detector

design without requiring either coding expertise or experience with Geant4.

Geant4 is the de facto high-energy physics standard for simulating the interaction of particles with fields and materials. However, the end user is required to write their own C++ program, and the learning curve for setting up the detector geometry and defining sensitive elements and readout can be quite daunting. We have developed the Geant4-based detector simulation program, slic, which employs generic IO formats as well as a textual detector description. Extending the pure geometric capabilities of GDML, LCDD enables fields, regions, sensitive detector readout elements, etc. to be fully described at runtime using an xml file. We also describe how more complex geometries, such as those from CAD programs, can be seamlessly incorporated into the xml files. We provide executable programs for Windows, Mac OSX and Linux, allowing physicists to design detectors within minutes.

We present the architecture as well as the implementation for several candidate ILC, CLIC and Muon Collider detector designs. We also describe the implementation of a fixed target experiment (HPS at JLab) and a proton computed tomography (pCT) implementation, demonstrating both the flexibility and the power of the system.

org.lcsim: event reconstruction and analysis

Maximizing the physics performance of detectors being designed for the ILC, while remaining sensitive to cost constraints, requires a powerful, efficient, and flexible simulation, reconstruction and analysis environment to study the capabilities of a large number of different detector designs. The preparation of Letters Of Intent for the ILC involved the detailed study of dozens of detector options, layouts and readout technologies; the final physics benchmarking studies required the reconstruction and analysis of hundreds of millions of events.

We describe the Java-based software toolkit (org.lcsim) which was used for full event reconstruction and analysis. The components are fully modular and are available for tasks from digitization of tracking detector signals through to cluster finding, pattern recognition, track-fitting, calorimeter clustering, individual particle reconstruction, jet-finding, and analysis. The detector is defined by the same xml input files used for the detector response simulation, ensuring the simulation and reconstruction geometries are always commensurate by construction. We discuss the architecture as well as the performance.

In addition to the ILC LOI studies, we describe the use of the org.lcsim software at CERN for CLiC physics and detector studies which culmintaed in the successful completion of their CDR, its application for dual-readout crystal calorimeter detector R&D at Fermilab, and detector design and event reconstruction, including an online trigger, for the proposed “heavy photon” experiment HPS at JLAB.

Poster Session / 559

Software For the Mu2e Experiment at Fermilab

Author: Robert Kutschke[1]

[1] *Femilab*

Corresponding Author: kutschke@fnal.gov

The Mu2e experiment at Fermilab is in proceeding through its R&D and approval processes. Two critical elements of R&D towards a design that will achieve the physics goals are an end-to-end simulation package and reconstruction code that has reached the stage of an advanced prototype. These codes live within the environment of the experiment's intrastructure software. Mu2e uses art as the infrastructure software, Geant4 as the simulation engine, a port of Kalman Filter from the Super-B FastSim package as the final track fitter, and Mu2e-developed code for event generators, creation of digis and the remaining reconstruction algorithms. A ROOT-based event display runs within the art based framework. This talk will present the Mu2e software with emphasis on two topics: a) defining the right boundaries between the component parts in order to ease the job of making a coherent whole and b) interacting with the art development team in order to influence the directions of art.

Poster Session / **560**

Implementation and use of BaBar Long Term Data Access.

Author: Douglas Smith[1]

Co-authors: Concetta Cartaro [1]; Homer A. Neal [1]; Igor Gaponenko [1]; Kyle Fransham [2]; Marcus Ebert [1]; Steffen Luitz [1]; Wilko Kroeger [1]

[1] *SLAC National Accelerator Lab.*

[2] *University of Victoria*

Corresponding Author: douglas@slac.stanford.edu

The BaBar high energy physics experiment acquired data from 1999 until 2008. Soon after the end of data taking, the effort to produce the final dataset started. This final dataset contains over 11x10ˆ9 events, in 1.6x10ˆ6 files, over a petabyte of storage. The Long Term Data Access (LTDA) project aims at the preservation of the BaBar data, analysis tools and documentation to ensure the capability to perform physics analyses and publish new physics results. It also foresees to use the data for education and outreach, and for the combination of BaBar results with other experiments. The central element of the BaBar LTDA is an integrated cluster of computation and storage resources which will become the primary facility for the analysis of BaBar data in the coming years. The cluster uses virtualization technologies to ensure continued operation with future hardware and software platforms, and utilizes distributed computation and storage methods for scalability. The design has been developed with particular attention to computer security and portability of the model. This presentation will focus on the details of the implementation and use of the LTDA within the BaBar Collaboration, and the first user experience with BaBar data analyses.

Poster Session / **561**

MAUS Online Data Quality

Authors: Christopher Tunnell[1]; Michael Jackson[2]

[1] *Oxford*

[2] *EPCC*

Corresponding Author: ctunnell@nikhef.nl

Within the Muon Ionization Cooling Experiment (MICE), the MICE Analysis User Software (MAUS) framework performs both online analysis of live data and detailed offline data analysis, simulation, and accelerator design. The MAUS Map-Reduce API parallelizes computing in the control room, ensures that code can be run both offline and online, and displays plots for users in an easily extendable manner. The original Map-Reduce design can be advantageous for offline computing but cannot be used in online settings. It expects all map operations to terminate before running the reduction;

however, the data flow for online analysis requires the continuous updating of live plots as data arrives. For online running, the 'map'and 'reduce'steps must happen concurrently; therefore, new parallelization routines were developed specifically for this use. The 'map'step is parallelized using a Python-based distributed task queue called Celery, and output from these tasks is then written into a NoSQL database called CouchDB. As the 'mapper'writes output, the plotting 'reducers'query the database, request data from a user-specified window in time, and make plots using Matplotlib or PyRoot. The 'reducers'serialize the plots into the data stream after which all the data is written to the database by the output routines. Finally, plots are displayed on the web using the Django platform, which queries the database and displays the plots to the control room and the world. By maintaining the API and modifying the data flow, MICE is able to use identical analysis software in both offline and online scenarios, thus avoiding a common issue in experimental particle physics.

Student? Enter 'yes'. See http://goo.gl/MVv53:

no

Software Engineering, Data Stores and Databases / 563

MAUS: MICE Analysis User Software

Author: Christopher Tunnell[None]

Co-author: Durga Rajaram [1]

[1] *IIT, Chicago*

Corresponding Authors: durga@fnal.gov, ctunnell@nikhef.nl

The Muon Ionization Cooling Experiment (MICE) has developed the MICE Analysis User Software (MAUS) to simulate and analyse experimental data. It serves as the primary codebase for the experiment, providing for online data quality checks and offline batch simulation and reconstruction. The code is structured in a Map-Reduce framework to allow parallelization whether on a personal machine or in the control room. Various software engineering practices from industry are also used to ensure correct and maintainable physics code, which include unit, functional and integration tests, continuous integration and load testing, code reviews, and distributed version control systems. Lastly, there are various small design decisions like using JSON as the data structure, using SWIG to allow developers to write components in either Python or C++, or using the SCons python-based build system that may be of interest to other experiments.

Summary:

Lessons learned from adopting numerous software engineering practices from industry when developing the MICE experiment software framework. Experiences will be shared.

Poster Session / 564

Improving Phenix search experience with Solr/Lucene and Nutch

Authors: Dave Morrison[1]; Irina Sourikova[1]

[1] *Brookhaven National Laboratory*

Corresponding Author: irina@bnl.gov

During its 20 years of R&D, construction and operation the Phenix experiment at RHIC has accumulated large amounts of proprietary collaboration data that is hosted
on many servers around the world and is not open for commercial search engines for indexing and searching.The legacy search infrastructure did not
scale well with the fast growing Phenix document
base and produced results inadequate in both precision
and recall.

After considering the possible alternatives that would provide an aggregated, fast, full text sear

formats (text, pdf, ppt, etc) we decided to use Nutch as a web crawler and Solr/Lucene as a search engine.
Nutch support of crawling multiple domains helps Phenix aggregate collaboration data from many participating institutions. The ability of Nutch to parse large variety
of file formats greatly increases the search domain.
Phenix search got a substantial boost in precision by using the Solr support for faceted navigation - indexing data under custom categories and then combining text search with a progressive narrowing of choices in available categories. Most users in Phenix know how the data is structured fairly well and can limit the search to a specific area right away, for example searching only in the mail archives or published papers.

To present XML-based Solr search results in a user-friendly manner we decided to use Drupal as a we

We will report on Phenix experience searching with Solr/Lucene, Nutch and Drupal.

Online Computing / 565

The MICE Online Systems

Author: Linda Coney[1]

[1] *University of California, Riverside*

Corresponding Author: lconey@fnal.gov

The Muon Ionization Cooling Experiment (MICE) is designed to test transverse cooling of a muon beam, demonstrating an important step along the path toward creating future high intensity muon beam facilities. Protons in the ISIS synchrotron impact a titanium target, producing pions which decay into muons that propagate through the beam line to the MICE cooling channel. Along the beam line, particle identification (PID) detectors, scintillating fiber tracking detectors, and beam diagnostic tools identify and measure individual muons moving through the cooling channel.

The MICE Online Systems encompass all tools; including hardware, software, and documentation, within the MLCR (MICE Local Control Room) that allow the experiment to efficiently record high quality data. Controls and Monitoring (C&M), Data Acquisition (DAQ), Online Monitoring and Reconstruction, Data Transfer, and Networking all fall under the Online Systems umbrella.

C&M controls all MICE systems including the target, conventional and superconducting magnets, detectors, and cooling channel components. Monitoring of environment and equipment function during data-taking is provided by the Alarm Handler, and the Archiver saves a record of all run conditions. C&M also provides an interface with the Configuration Database to retrieve pre-selected run configurations and to save new configurations.

The DAQ controls the taking and recording of all data in MICE, and must allow the collection of data for up to 600 muons in MICE during a 3 ms data acquisition gate. Equipment readout, event building, and the DAQ user interface software has been developed from the DATE package, originally from the ALICE experiment. Within the DAQ, the trigger system initiates the digitization of detector signals and controls the timing of the subsequent readout and local storage of data.

Online Monitoring provides an immediate, low-level diagnostic monitoring capability for all DAQ hardware. It displays DAQ performance and allows for individual channel-by-channel assessment

of detector component behavior. Online Reconstruction and Data Quality provide real-time physics information during data-taking, immediate feedback to experimenters, and a first look at analysis quantities. It includes histograms filled during data-taking for checks of data quality, beam dynamics, and detector function and necessarily interfaces with the MICE DAQ and offline software. After each period of MICE running, all data and related histograms are transferred to remote storage on the GRID for later analysis.

Collaborative tools / 566

Project Management Web Tools at the MICE experiment

Author: Linda Coney[1]

[1] *University of California, Riverside*

Corresponding Author: lconey@fnal.gov

Project management tools like Trac are commonly used within the open-source community to coordinate projects. The Muon Ionization Cooling Experiment (MICE) uses the project management web application Redmine to host mice.rl.ac.uk. Many groups within the experiment have a Redmine project: analysis, computing and software (including offline, online, controls and monitoring, and database subgroups), executive board, and operations. All of these groups use the website to communicate, track effort, develop schedules, and maintain documentation. The issue tracker is a rich tool that is used to identify tasks and monitor progress within groups on timescales ranging from immediate and unexpected problems to milestones that cover the life of the experiment. It allows the prioritization of tasks according to time-sensitivity, while providing a searchable record of work that has been done. This record of work can be used to measure both individual and overall group activity, identify areas lacking sufficient personnel or effort, and as a measure of progress against the schedule. Given that MICE, like many particle physics experiments, is an international community, such a system is required to allow easy communication within a global collaboration. Unlike systems that are purely wiki-based, the structure of a project management tool like Redmine allows information to be maintained in a more structured and logical fashion.

Poster Session / 567

The ATLAS database application enhancements using Oracle 11g

Author: Gancho Dimitrov[1]

Co-authors: Luca Canali [2]; Marcin Blaszczyk [2]; Roman Sorokoletov [3]

[1] *Brookhaven National Laboratory (US)*

[2] *CERN*

[3] *University of Texas at Arlington (US)*

Corresponding Author: gancho.dimitrov@cern.ch

The ATLAS experiment at LHC relies on databases for detector online data-taking, storage and retrieval of configurations, calibrations and alignments, post data-taking analysis, file management over the grid, job submission and management, data replications to other computing centers, etc. The Oracle Relational Database Management System has been addressing the ATLAS database requirements to a great extent for many years. Several database clusters were deployed for the needs of the different applications. The data volume, complexity and demands from the users are increasing steadily with time. Nowadays about 20 TB of data are stored in the ATLAS

Oracle databases at CERN (not including the index overhead), but the most impressive number is the hosted 260 database schemas (in the common case each schema is related to a dedicated client application with its own requirements). At the beginning of 2012 all ATLAS databases at CERN are upgraded to the newest Oracle version 11g Release 2. In order to make the ATLAS DB applications more reliable and performant we explored and evaluated the new 11g database features. In this work we present some of the Oracle 11g enhancements and typical ATLAS application use cases which suit best and the gain from the implemented changes.

Poster Session / **568**

Architecture and evolution of the CMS High Level Trigger

Author: Andrea Bocci[1]

[1] *CERN*

Corresponding Author: andrea.bocci@cern.ch

The CMS experiment has been designed with a 2-level trigger system: the Level 1 Trigger, implemented using FPGA and custom ASIC technology, and the High Level Trigger (HLT), implemented running a streamlined version of the CMS offline reconstruction software on a cluster of commercial rack-mounted computers, comprising thousands of CPUs.
The design of a software trigger system requires a tradeoff between the complexity of the algorithms running online, the output rate, and the selection efficiency. The complexity is limited by the available computing power, while the rate is constrained by the offline storage and processing capabilities. The main challenge faced during 2011 was the fine-tuning and optimisation of the algorithms, in order to cope with the increasing LHC luminosity without impacting the physics performance.

The flexibility of a single all-software trigger running on the full L1 output rate also allowed the introduction of different data "streams": in order to monitor the performance of the detector and the HLT itself, to collect dedicated data for the detector calibrations, and for special physics analysis.
Here we will present the architecture of the High Level Trigger, its operation and evolution.
We will outline the improvements introduced during 2011, such as particle-flow techniques, pile-up subtraction and rejection, and optimisation of the tracking algorithms, including their impact on the CPU-time of the HLT process. We will then discuss the improvements planned for the 2012 data taking.

Poster Session / **569**

Performance of the CMS High Level Trigger

Author: Andrea Bocci[1]

[1] *CERN*

Corresponding Author: andrea.bocci@cern.ch

The CMS experiment has been designed with a 2-level trigger system: the Level 1 Trigger, implemented using FPGA and custom ASIC technology, and the High Level Trigger (HLT), implemented running a streamlined version of the CMS offline reconstruction software on a cluster of commercial rack-mounted computers, comprising thousands of CPUs.
The design of a software trigger system requires a tradeoff between the complexity of the algorithms running online, the output rate, and the selection efficiency. The complexity is limited by the available computing power, while the rate is constrained by the offline storage and processing capabilities.

The main challenge faced during 2011 was the fine-tuning and optimisation of the algorithms, in order to cope with the increasing LHC luminosity without impacting the physics performance.

Here we will present a review of the performance of the main triggers used during the 2011 data taking, ranging from simpler single-object selections to more complex algorithms combining different objects, and applying analysis-level reconstruction and selection.
We will discuss how the increasing LHC luminosity and pile-up have affected their performance, and how these effects have been mitigated.

Poster Session / 570

ConfDB: a database backend and GUI program for the management and development of CMS High Level Trigger

Author: Andrea Bocci[1]

[1] *CERN*

Corresponding Author: andrea.bocci@cern.ch

The CMS experiment has been designed with a 2-level trigger system: the Level 1 Trigger, implemented using FPGA and custom ASIC technology, and the High Level Trigger (HLT), implemented running a streamlined version of the CMS offline reconstruction software on a cluster of commercial rack-mounted computers, comprising thousands of CPUs.

The CMS software is written mostly in C++, using Python as its configuration language through an embedded CPython interpreter. The configuration of each process is made up of hundreds of "modules", organised in "sequences" and "paths". As an example, the latest HLT configurations used for 2011 data taking comprised over 2200 different modules, organized in more than 400 independent trigger paths.
To manage the complexity of each HLT configuration, and the number of different configurations used to cope with the changing LHC luminosity and specific detector conditions, all configurations used for data taking are stored in a database ("ConfDB") and developed with a dedicated GUI program.
A configuration can be converted back to python format through the GUI itself, or with command-line tools which interact with the database backend through a web server.
In addition, an experimental Jython interface is under development, to allow loading a python configuration directly into the GUI and in the database.

We will describe how the CMS python configuration language is used to steer the High Level Trigger, and detail the ConfDB GUI used to edit such configurations, with special emphasis on the features introduced specifically for trigger development.

Collaborative tools / 571

Next Generation High Quality Videoconferencing Service for the LHC

Authors: Joao Correia Fernandes[1]; Marek Domaracky[1]

Co-author: Thomas Baron [1]

[1] *CERN*

Corresponding Author: marek.domaracky@cern.ch

In recent times, we have witnessed an explosion of video initiatives in the industry worldwide. Several advancements in video technology are currently improving the way we interact and collaborate. These advancements are forcing tendencies and overall experiences: any device in any network can be used to collaborate, in most cases with an overall high quality. To cope with this technology progresses, CERN IT Department has taken the leading role to establish strategies and directions to improve the user experience in remote dispersed meetings and remote collaboration at large in the worldwide LHC communities. Due to the high rate of dispersion in the LHC user communities, these are critically dependent of videoconferencing technology, with a need of robustness and high quality for the best possible user experience. We will present an analysis of the factors that influenced the technical and strategic choices to improve the reliability, efficiency and overall quality of the LHC remote sessions. In particular, we are going to describe how the new videoconferencing service offered by CERN IT, based on Vidyo technology suits these requirements. During a vidyoconference, Vidyo's core technology continuously monitors the performance of the underlying network and the capabilities of each endpoint device, in order to adapt video streams in real time and optimize video communication. This results in offering telepresence-quality videoconferencing over the commercial Internet and at the same time, in providing a robust platform to make video communications universally available on any device ranging from traditional videoconferencing room systems, to multiplatform PCs, the latest smartphones and tablets PCs, over any network. The infrastructure deployed to offer this new service, its integration in the specific CERN environment will be presented as well as recent use cases.

Summary:

A status update about the current videoconferencing technology and services offered by CERN IT to the LHC community, in particular the new Vidyo service.

Poster Session / **572**

CERN Lecture archiving and Video Delivery to any screen

Author: Marek Domaracky[1]

Co-author: Thomas Baron [1]

[1] *CERN*

Corresponding Author: marek.domaracky@cern.ch

Over the last few years, we have seen the broadcast industry moving to mobile devices and to the broadband Internet delivering HD quality. To keep up with the trends, we deployed a new streaming infrastructure. We are now delivering live and on-demand video to all major platforms like Windows, Linux, Mac, iOS and Android running on PC, Smart Phone, Tablet or TV.
To optimize the viewing quality and pleasure on any device we improved the process of publishing recorded lectures with the new release of the Lecture Archiving system - Micala. As a result of our effort we developed a new lecture viewer, which gives our users the best possible experience for watching recorded lectures, with both video of the speaker and slides in high resolution and a fully customizable experience. For the mobile devices we improve quality and usability to watch any Video from CDS even on low bandwidth conditions. The various tools and processes involved will be described as well as the integration of these services in the wider CERN collaboration services offer.

Summary:

A review of the Webcast and Recording activities at CERN and showing the new improvements in the delivery of the Video to any screen.

Collaborative tools / **573**

Indico: CERN Collaboration Hub

Authors: Jose Benito Gonzalez Lopez[1]; Pedro Ferreira[1]

[1] *CERN*

Corresponding Author: pedro.ferreira@cern.ch

Since 2009, the development of Indico has focused on usability, performance and new features, especially the ones related to meeting collaboration. Usability studies have resulted in the biggest change Indico has experienced up to now, a new web layout that makes the user experience better. Performance improvements were also a key goal since 2010; the main features of Indico have been optimized remarkably. Along with usability and performance, new features have been added to Indico such as webchat integration, video services bookings, webcast and recording requests, designed to really reinforce Indico position as the main hub for all CERN collaboration services, and many others which aim is to complete the conference lifecycle management.
Indico development is also moving towards a broader collaboration where other institutes, hosting their own Indico instance, can contribute to the project in order make it a better and more complete tool.

Summary:

A review of all the enhancements done in the recent past to Indico, and especially a view of Indico as CERN collaboration hub.

Collaborative tools / 574

The Workflow of LHC Papers

Author: Jean-Yves Le Meur[1]

[1] *CERN*

Corresponding Authors: ludmila.marian@cern.ch, jean-yves.le.meur@cern.ch

In this talk, we will explain how CERN digital library services have evolved to deal with the publication of the first results of the LHC.
We will describe the work-flow of the documents on CERN Document Server and the diverse constraints relative to this work-flow.
We will also give an overview on how the underlying software, Invenio, has been enriched to cope with special needs.
In a second part, the impact in terms of user access to the publication of the experiments and to the multimedia material will be detailed.
Finally, the talk will focus on how the institutional repository (CDS) is being linked to the HEP disciplinary archive (INSPIRE) in order to provide users with a central access point to reach LHC results.

Poster Session / 575

Recent Developments in the Geant4 Precompound and Deexcitation Models

Authors: Jose Manuel Quesada Molina[1]; Jose Manuel Quesada Molina[1]

Co-authors: Anton Ivantchenko [2]; Vladimir Ivantchenko [3]

[1] *Universidad de Sevilla (ES)*

[2] *Geant 4 Associates International Experts in Radiation Simulatio*

[3] *M.V. Lomonosov Moscow State University (RU)*

Corresponding Author: jose.manuel.quesada.molina@cern.ch

The final stages of a number of generators of inelastic hadron/ion interactions with nuclei in Geant4 are described by native pre-equilibrium and de-excitation models. The pre-compound model is responsible for pre-equilibrium emission of protons, neutrons and light ions. The de-excitation model provides sampling of evaporation of neutrons, protons and light fragments up to magnesium. Fermi break-up model is invocated for decay of light fragments (Z<9, A<17) whereas statistical multifragmentation and fission are invoked for heavier ones. Photon emission has a chance for all excited fragments. Recently, model improvements in Fermi break-up and photon evaporation as well as changes in the logic of the de- excitation handler have been made. New sets of experimental data have been included in the validation, which are presented.

Submitted on behalf of Hadronic Physics Working Group of the Geant4 Collaboration

Poster Session / **576**

Automating MICE Controls and Monitoring

Author: Pierrick Hanlet[1]

[1] *Illinois Institute of Technology*

Corresponding Author: hanlet@fnal.gov

The Muon Ionization Cooling Experiment (MICE) is a demonstration experiment to prove the feasibility of cooling a beam of muons for use in a Neutrino Factory and/or Muon Collider. The MICE cooling channel is a section of a modified Study II cooling channel which will provide a 10% reduction in beam emittance. In order to ensure a reliable measurement, MICE will measure the beam emittance before and after the cooling channel at the level of 1%, or an absolute measurement of 0.001. This renders MICE a precision experiment which requires strict controls and monitoring of all experimental parameters in order to control systematic errors. The MICE Controls and Monitoring system is based on EPICS and integrates with the DAQ, Data monitoring systems, and a configuration database. A new paradigm is being developed for MICE to ensure proper sequencing of equipment and use of system resources to protect data quality. A description of this system, its implementation, and performance during recent muon beam data collection will be discussed.

Poster Session / **577**

BAT - The Bayesian Analysis Toolkit

Authors: Allen Caldwell[1]; Daniel Kollar[2]; Frederik Beaujean[3]; Kevin Alexander Kroeninger[4]; Shabnaz Pashapouralamdari[4]

[1] *Max Planck Institute*

[2] *Max-Planck-Institut fuer Physik, Munich*

[3] *Max Planck Institute for Physics*

[4] *Georg-August-Universitaet Goettingen (DE)*

Corresponding Author: dkollar@mppmu.mpg.de

The main goals of data analysis are to infer the parameters of models from data, to draw conclusions on the validity of models,

and to compare their predictions allowing to select the most appropriate model.

The Bayesian Analysis Toolkit, BAT, is a tool developed to evaluate the posterior probability distribution for models and their parameters. It is centered around Bayes' Theorem and is realized with the use of Markov Chain Monte Carlo giving access to the full posterior probability distribution. This enables straightforward parameter estimation, limit setting and uncertainty propagation.

BAT is implemented in C++ and allows a flexible definition of models. It is interfaced to other software packaged commonly used in high-energy physics: ROOT, Minuit, RooStats and CUBA. A set of predefined models exists to cover standard statistical cases.

We will present an overview of the software and the algorithms implemented. Recent updates and future plans will be summarized.

Event Processing / **578**

Recent Developments and Validation of Geant4 Hadronic Physics

Authors: D.H. Wright[1]; Dennis Herbert Wright[2]; Hans-Joachim Wenzel[3]; Julia Yarba[3]; Michael Kelsey[2]; Sunanda Banerjee[4]; Vladimir Uzhinsky[5]

Co-authors: Andreas Schaelicke [6]; Jennifer Karkoska [7]

[1] *SLAC*

[2] *SLAC National Accelerator Laboratory (US)*

[3] *Fermi National Accelerator Lab. (US)*

[4] *Saha Institute of Nuclear Physics (IN)*

[5] *CERN and JINR*

[6] *University of Edinburgh (GB)*

[7] *University of Rochster*

Corresponding Author: julia.yarba@cern.ch

In the past year several improvements in Geant4 hadronic physics code have been made, both for
HEP and nuclear physics applications. We discuss the implications of these changes for physics
simulation performance and user code. In this context several of the most-used codes will be covered
briefly. These include the Fritiof (FTF) parton string model which has been extended to include
antinucleon and antinucleus interactions with nuclei, the Bertini-style cascade with its improved CPU
performance and extension to include photon interactions, and the precompound and deexcitation
models. We have recently released new models and databases for low energy neutrons, and the
radioactive decay process has been improved with the addition of forbidden beta decays and
better gamma spectra following internal conversion.

As new and improved models become available, the number of tests and comparisons to data has
increased. One of these is a validation of the parton string models

against data from the MIPP
experiment, which covers the largely untested range of 50 to 100 GeV.
At the other extreme, a
new stopped hadron validation will cover pions, kaons and antiprotons.
These, and the ongoing
simplified calorimeter studies, will be discussed briefly. We also
discuss the increasing number of
regularly performed validations, the demands they place on both software
and users, and the
automated validation system being developed to address them.

Distributed Processing and Analysis on Grids and Clouds / **579**

The CMS workload management system

Author: Stuart Wakefield[1]

[1] *Imperial College London*

Corresponding Author: stuart.wakefield@imperial.ac.uk

CMS has started the process of rolling out a new workload management system. This system is currently used for reprocessing and monte carlo production with tests under way using it for user analysis.

It was decided to combine, as much as possible, the production/processing, analysis and T0 codebases so as to reduce duplicated functionality and make best use of limited developer and testing resources.

This system now includes central request submission and management (Request Manager); a task queue for parcelling up and distributing work (WorkQueue) and agents which process requests by interfacing with disparate batch and storage resources (WMAgent).

Plenary / **580**

Welcome to CHEP 2012

Corresponding Author: mernst@bnl.gov

Plenary / **581**

Welcome to NYU

Plenary / **582**

Keynote Address: High Energy Physics and Computing –Perspectives from DOE

Plenary / **583**

LHC experience so far, prospects for the future

Corresponding Author: incandel@fnal.gov

Plenary / **584**

HEP Computing

Corresponding Author: rene.brun@cern.ch

Plenary / **585**

Upgrade of the LHC Experiment Online Systems

Corresponding Author: wsmith@hep.wisc.edu

Plenary / **586**

Perspective Across The Technology Landscape: Existing Standards Driving Emerging Innovations

Plenary / **587**

ROOT

Corresponding Author: fons.rademakers@cern.ch

Plenary / **588**

GEANT4 Roadmap

Corresponding Author: asai@slac.stanford.edu

Plenary / **589**

A reflection on Software Engineering in HEP

Author: Federico Carminati[1]

[1] *CERN*

Corresponding Author: federico.carminati@cern.ch

Plenary / **590**

A review of analysis in different experiments

Corresponding Author: markus.klute@cern.ch

Plenary / **591**

New computing models and LHCONE

Corresponding Author: ian.fisk@cern.ch

Plenary / **592**

Current operations and future role of the Grid

Author: Oxana Smirnova[1]

[1] *LUND UNIVERSITY*

Corresponding Author: oxana.smirnova@hep.lu.se

Plenary / **593**

Middleware Evolution

Author: Sebastien Goasguen[1]

[1] *Clemson University*

Corresponding Author: sebgoa@clemson.edu

From Grid to Cloud: A Perspective

Plenary / **594**

New SW trends and technologies in the Internet industry

Plenary / **595**

Computing Technology Future

Corresponding Author: johnsson@cs.uh.edu

Plenary / **596**

Track Summary: Online Computing

Plenary / **597**

Track Summary: Distributed Processing & Analysis

Plenary / **598**

Track Summary: Event Processing

Corresponding Author: lyon@fnal.gov

Plenary / **599**

Track Summary: Fabrics & Networking

Corresponding Author: andreas.heiss@kit.edu

Plenary / **600**

Track Summary: Software Engineering

Author: David Lange[1]

[1] *Lawrence Livermore Nat. Laboratory (US)*

Corresponding Author: david.lange@cern.ch

Plenary / **601**

Introduction of CHEP2013 in Amsterdam

Author: David Groep[1]

Co-author: NIKHEF Team

[1] *NIKHEF*

Corresponding Author: david.groep@cern.ch

Plenary / **602**

Analysis with Extremely Large Datasets

Author: Jacek Becla[1]

[1] *SLAC*

Corresponding Author: becla@slac.stanford.edu

Plenary / **603**

Large Storage Systems: Present and Future

Corresponding Author: andreas.joachim.peters@cern.ch

Plenary / **604**

Networking: 100G Across Oceans: Where and When?

Corresponding Author: artur.barczyk@cern.ch

Plenary / **605**

Computing the Universe

Corresponding Author: apope@anl.gov

Plenary / **606**

Future Experiments and Impact on Computing

Corresponding Author: messchendorp@kvi.nl

Plenary / **607**

Data Preservation and Long Term Analysis in High Energy Physics

Author: David South[1]

[1] *DESY*

Corresponding Author: david.south@cern.ch

Plenary / **608**

VC in HEP: Status and Perpective

Corresponding Author: philippe.galvez@cern.ch

Plenary / **609**

Lightning Talks (Session 1)

Plenary / **610**

Track Summary: Collaborative Tools

Corresponding Author: tony.johnson@slac.stanford.edu

Plenary / **611**

Lightning Talks (Session 2)

Poster Session / **612**

Linear photodiode array for tracking and video recording of a human speaker

Author: Daniel DeTone[1]

Co-authors: Bob Lougheed [1]; Homer A. Neal [1]

[1] *University of Michigan*

Communication and collaboration using stored digital media has recently garnered increasing interest in many facets of business, government and education. This is primarily due to improvements in the quality of cameras and the speed of computers. Digital media serves as an effective alternative in the absence of physical interaction between multiple individuals. Video recordings that allow for intimate interaction—the viewer's ability to discern a presenter's facial features, lips and hand motions —have been shown to be more effective than videos that do not. To achieve this, a video capture must ensure that the speaker occupies a significant portion of the captured pixels. But camera operators are costly and often do an imperfect job of tracking presenters in unrehearsed situations. This

creates the need for a robust, automated system that directs a video camera to follow a presenter as he or she walks anywhere in the front of a lecture hall or large conference room. We present such a system.

The system consists of a commercial, off-the-shelf pan/tilt/zoom (PTZ) color video camera, a necklace of infrared LEDs and a linear photodiode array detector. Electronic output from the photodiode array is processed to generate the location of the LED necklace, which is worn by a human speaker. The computer controls the video camera movements to record video of the speaker. The speaker's vertical position and depth are assumed to remain relatively constant –the video camera is sent only panning (horizontal) movement commands. The LED necklace is flashed at 70Hz at 50% duty cycle to provide noise-filtering capability. The benefit to using a photodiode array versus a standard video camera is its higher frame rate (4kHz vs. 60Hz). The higher frame rate allows for the filtering of noise infrared such as sunlight and indoor lighting –a capability absent from other tracking technologies. The system has been tested in a large lecture hall, and is shown to be effective.

Poster Session / **613**

PLUME –FEATHER

Author: Dirk Hoffmann[1]

[1] *CPPM, Aix-Marseille Université, CNRS/IN2P3, Marseille, France*

Corresponding Author: dirk.hoffmann@cern.ch

on behalf of the PLUME Technical Committee http://projet-plume.org“
for the PLUME abstract.

PLUME - FEATHER is a non-profit project created to Promote economicaL, Useful and Maintained softwarE For the Higher Education And THE Research communities. The site references software, mainly Free/Libre Open Source Software (FLOSS) from French universities and national research organisations, (CNRS, INRA...), laboratories or departments. Plume means feather in French.
The main goals of PLUME –FEATHER are:
• promote the community's own developments,
• contribute to the developement and sharing FLOSS (Free/Libre Open Source Software) information, experiences and expertise in the community,
• bring together FLOSS experts and knowledgeable people to create a community,
• foster and facilitate FLOSS use, deployment and contribution in the higher education and the research communities.
PLUME - FEATHER was initiated by the CNRS unit UREC. The UREC unit has been integrated to the CNRS computing division DSI in 2011. The different resources are provided by the main partners involved in the project.
The French PLUME server contains more than 1000 software reference cards, edited and peer- reviewed by members of the research and education community. It is online since November 2007, and the first English pages have been published in April 2009. Currently there are 84 software products referenced in the PLUME-FEATHER area. Therefore time has come to announce the availability and potential of the PLUME project on the international level to find not only users, but also contributors: editors and reviewers of frequently used software in our domain.

BoF / **614**

Report on International Data Exchange Requirements (RIDER) – What are the international data flow requirements for 2020?

Corresponding Author: gemmill@clemson.edu

RIDER is an NSF-funded study (Award #1223688) of the current and future 2020 international data requirements of the science and engineering community,

specifically flow of data into the US. Results will assist NSF in predicting future capacity requirements and planning funding for the International Research Network Connections (IRNC) programs.

This BoF is an opportunity to provide your input to this NSF study. Discussion topics will include:
* Are you thinking about moving data from other countries into the US now and in the future? What data? From where?
* What do you view as the current top three data sources important to your research?
* What would you expect the top three data drivers to be in 2020?
* How do you access large data sets originating overseas (What are the steps?)
* What worries/concerns you about current and future international data movement?
* Are there gaps in current network infrastructure ? Do you forsee possible future gaps ?
* How do you let your funding agencies know about your future international network requirements? (by email? Phone? Face to face?other?)
* How do you let your network / infrastructure support people know of your future international network requirements? ? (by email? Phone? Face to face? Other?)
* What else should be paid attention to besides moving the data? Are storage and access (search/locate) adequately addressed? What will be the challenges in 2020?
* Will there be large amounts of data coming from new international sources? What data? What countries?

BoF / **615**

Report on International Data Exchange Requirements (RIDER) – What are the international data flow requirements for 2020?

Corresponding Author: gemmill@clemson.edu

RIDER is an NSF-funded study (Award #1223688) of the current and future 2020 international data requirements of the science and engineering community, specifically flow of data into the US. Results will assist NSF in predicting future capacity requirements and planning funding for the International Research Network Connections (IRNC) programs.

This BoF is an opportunity to provide your input to this NSF study. Discussion topics will include:
* Are you thinking about moving data from other countries into the US now and in the future? What data? From where?
* What do you view as the current top three data sources important to your research?
* What would you expect the top three data drivers to be in 2020?
* How do you access large data sets originating overseas (What are the steps?)
* What worries/concerns you about current and future international data movement?
* Are there gaps in current network infrastructure ? Do you forsee possible future gaps ?
* How do you let your funding agencies know about your future international network requirements? (by email? Phone? Face to face?other?)
* How do you let your network / infrastructure support people know of your future international network requirements? ? (by email? Phone? Face to face? Other?)
* What else should be paid attention to besides moving the data? Are storage and access (search/locate) adequately addressed? What will be the challenges in 2020?
* Will there be large amounts of data coming from new international sources? What data? What countries?

DPHEP / 616

DPHEP