ACAT 2008

Monday 03 November 2008 - Friday 07 November 2008

Ettore Majorana Foundation and Centre for Scientific Culture

Book of abstracts

Table of contents

Two approaches to Combining Significances 1

Multi-threaded event processing with JANA 1

TMVA- the toolkit for multivariate data analysis 1

Porting Reconstruction Algorithms to the Cell Broadband Engine 2

Higgs bbbar decay at NNLO and beyond : the uncertainties of QCD predictions 2

Development, validation and maintenance of Monte Carlo event generators and generator services in the LHC era 3

Forget multicore! The future is many-core: An outlook to the explosion of parallelism likely to occur in the LHC era 4

Throughput Computing in C++ 4

Code quality from the programmer's perspective 5

PDE-FOAM - a probability-density estimation method based on self-adapting phase-space binning 5

Goodness of fit tests for weighted histograms 6

Application of the rule-growing algorithm RIPPER to particle physics analysis 6

Recent Progress of Geant4 Electromagnetic Physics and Readiness for the LHC Start 7

The Role of Interpreters in High Performance Computing 7

C++ and Data 8

g-Eclipse - user and developer friendly access to Grids and Clouds 8

Dedicated Services to Support Data Replication over GRID using ATLAS Distributed Data Management System. 9

Petaminer: Efficient Navigation to Petascale Data Using Event-Level Metadata 10

Strategies for btagging calibration using data at CMS 11

The CMS Framework for Alignment and Calibration 11

Track Reconstruction and Muon Identification in the Muon Detector of the CBM Experiment at FAIR 12

Building Grid client applications with the g-Eclipse framework 13

MINUIT Package Parallelization and applications using the RooFit Package 14

EVE - Event Visualization Environment of the ROOT framework 14

An overview of the b-Tagging algorithms in the CMS Offline software 15

Radiative corrections to Drell-Yan like processes in SANC 15

The PanDA System in the ATLAS Experiment 15

Standard SANC Modules 16
Distributed analysis in CMS using CRAB: the client-server architecture evolution and commissioning 16
VISPA: a Novel Concept for Visual Physics Analysis 17
Mass Storage System for Disk and Tape resources at the Tier1. 17
Hadronic Physics in Geant4: Improvements and Status for LHC Start 18
Feynman Diagrams, Differential Reduction and Hypergeometric Functions 18
Current status of FORM parallelization 18
ATLAS Handling Problematic Events in Quasi Real-Time 19
The Monte Carlo generators in CMS 19
Enhanced Gene Expression Programming for signal-background discrimination in particle physics 20
Large Scale Job Management and Experience in Recent Data Challenges within the LHC CMS experiment. ... 21
CMS Dataset Bookkeeping, Discovery and Query System 21
Data transfer infrastructure for CMS data taking 22
Software validation and performance optimisation in the CMS experiment 23
PARADIGM, a Decision Making Framework for Variable Selection and Reduction in High Energy Physics ... 23
An artificial neural network based algorithm for calorimetric energy measurements in CMS 24
WebDat: Bridging the Gap Between Unstructured and Structured Data 24
Profiling Post-GRID analysis 25
Distributed Computing in ATLAS 25
Alignment of the ATLAS Inner Detector tracking system 26
PERFORMANCE ASSESSMENT OF THE 2008 CONFIGURATION OF THE CICC JINR CLUSTER 27
Constraint Vertex Fitting in the ATLAS Experiment at LHC 27
Computational aspects for three-loop DIS calculations 28
The CMS Tier 0 28
Improving replica placement strategies using information from existing communication infrastructures 29
Mathematical model of magnetically interacting rigid bodies 29
Software development, release integration and distribution tools for the CMS experiment 30
LCG MCDB and HepML, next step to unified interfaces of Monte-Carlo Simulation 31
Data Analysis with PROOF 31
Optimized Calorimeter Signal Compaction for an Independent Component based ATLAS Electron/Jet Second-level Trigger 32
The ATLAS Conditions Database Model for the Muon Spectrometer 33
MDT data quality assessment at the Calibration centre for the ATLAS experiment at LHC 34
Early Experience with the CMS Computing Model 35
Computing Technology for Physics Research - Summary 35

Data Analysis - Algorithms and Tools - Summary 35
Methodology of Computations in Theoretical Physics - Summary 35
ACAT2008 Summary 35
High-Precision Arithmetic and Mathematical Physics 36
Using constraint programing to resolve the multi-source / multi-site data movement paradigm on the Grid ... 36
Evolution of the STAR Framework OO model for the Multi-Core era 37
Recent Improvements of the ROOT Fitting and Minimization Classes 37
New results for loop integrals 38
XCFS - an analysis disk pool & filesystem based on FUSE and xroot protocol 38
FROG : The Fast & Realistic OpenGl Event Displayer 39
Combining Hough Transform and Optimal Filtering for Efficient Cosmic Ray Detection with a Hadronic Calorimeter 39
ALICE Analysis Framework 40
The commissioning of CMS computing centres in the WLCG Grid 40
Efficient Level 2 Trigger System Based on Artificial Neural Networks 41
Sophisticated algorithms of analysis of spectroscopic data 42
Visualization of multidimensional histograms using hypervolume techniques 42
RDIG Alice computing at the time of the first LHC data 43
From moments to functions in higher order QCD 43
LHC phenomenology at next-to-leading order QCD: theoretical progress and new results 44
Many-core Round Table - How to prepare for the future 44
Aspects of Intellectual Property Law for HEP Software Developers 44
Tools for systematic event generator tuning and validation 45
A Numeric Comparison of Feature Selection Algorithms for Supervised Learning 45
The ATLAS Muon Trigger vertical slice at LHC startup 46
Job Centric Monitoring for ATLAS jobs in the LHC Computing Grid 47
Tau identification using multivariate techniques in ATLAS 48
CernVM - a virtual appliance for LHC applications 48
CVMFS - a file system for the CernVM virtual appliance 49
Numerical Evaluation of Feynman Integrals by a Direct Computation Method 49
Concurrent Collections: Parallel Programming Without Worrying About Parallelism 50
New implementation of the sector decomposition on FORM 50
The DAQ/HLT system of the ATLAS experiment 51
VARIOUS RUNTIME ENVIRONMENTS IN GRID BY MEANS OF VIRTUALIZATION OF WORKING NODES 51
The Advanced Resource Connector for Distributed LHC Computing 52

ATLAS Muon Spectrometer Monitoring using cosmic rays and First Beam Data 52
Getting ready for next generation computing 53
FairRoot Framework 53
Resonance analysis in ALICE 53
MonALISA : A Distributed Service System for Monitoring, Control and Global Optimization 54
Unitarity Methods For 1-Loop Amplitudes 54
Numerical calculations of Multiple Polylog functions 55
Advanced online system for data analysis and Monte-Carlo simulations in High Energy Physics 55
CompHEP status report (version 4.5) 55
Fireworks: A Physics Event Display for CMS 56
The ALICE Global Redirector. A step towards real storage robustness. 56
A prototype of a dinamically expandable Virtual Analysis Facility 57
ATLAS trigger status and results from commissioning operations 57
Interactive Data Analysis with PROOF, Experience at GSI 58
FormCalc 6 58
GENIUS/EnginFrame Grid Portal: new features and enhancements 59
Software Validation Infrastructure for the Atlas High-Level Trigger Validation Infrastructure for the ATLAS High-Level Trigger 60
Java based software for High-Energy and Astro-physics 61
The Blue Brain Project - Simulation-based Research in Neuroscience 61
Recent developments of GRACE 62
Are SE Architectures Ready For LHC 62
Introduction to The LLVM Compiler System 62
Multivariate Methods in Particle Physics: Today and Tomorrow 63
Computing in HEP at the time of the grid 63
Introduction to ACAT 2008 63
Introduction to Morning Session 63

Methodology of Computations in Theoretical Physics - Session 2 / 0

Two approaches to Combining Significances

Dr. BITYUKOV, Sergey [1]; Prof. KRASNIKOV, Nikolai [2]; Dr. NIKITENKO, Alexander [3]

[1] *INSTITUTE FOR HIGH ENERGY PHYSICS, PROTVINO*

[2] *Institute for nuclear research RAS, Moscow, Russia*

[3] *Imperial College, London, United Kingdom*

Corresponding Author: serguei.bitioukov@cern.ch

We compare two approaches to the combining of signal significances: the approach, in which the signal significances are considered as corresponding random variables, and the approach with the using of confidence distributions. Several signal significances, which are often used in analysis of data in experimental physics as a measure of excess of the observed or expected number of signal events above the predicted number of background events, are considered.

Data Analysis - Algorithms and Tools / 1

Multi-threaded event processing with JANA

Dr. LAWRENCE, David [1]

[1] *Jefferson Lab*

Corresponding Author: davidl@jlab.org

The C++ reconstruction framework JANA has been written to support the next generation of Nuclear Physics experiments at Jefferson Lab in anticipation of the 12GeV upgrade. This includes the GlueX experiment in the planned 4th experimental hall "Hall-D". The JANA framework was designed to allow multi-threaded event processing with a minimal impact on developers of reconstruction software. As we enter the multi-core era, thread-enabled code will become essential to utilizing the full processor power available without invoking the logistical overhead of managing many individual processes. Event-based reconstruction lends itself naturally to mutli-threaded processing. Emphasis will be placed on the multi-threading features of the framework. Test results of the scaling of event processing rates with number of threads will be shown.

Data Analysis - Algorithms and Tools / 2

TMVA- the toolkit for multivariate data analysis

VOSS (FOR THE TMVA-TEAM), Helge [1]

[1] *MPI-Heidelberg*

Corresponding Author: helge.voss@cern.ch

Multivariate data analysis techniques are becoming increasingly important for high energy physics experiments. TMVA is a tool, integrated in the ROOT environment, which allows easy access to sophisticated multivariate classifiers allowing for a widespread use of these very effective data selection techniques. It furthermore provides a number of pre-processing capabilities and numerous additional evaluation features that help in getting the most out of the data. The talk gives an overview of the TMVA package, and presents new features and developments.

Thursday, 06 November 2008 / 3

Porting Reconstruction Algorithms to the Cell Broadband Engine

Dr. KISEL, Ivan [1]; Mr. GORBUNOV, Sergey [2]; Prof. KEBSCHULL, Udo [2]; Prof. LINDENSTRUTH, Volker [2]; Dr. MUELLER, Walter [1]

[1] *Gesellschaft fuer Schwerionenforschung mbH (GSI), Darmstadt, Germany*

[2] *Kirchhoff Institute for Physics, University of Heidelberg, Germany*

Corresponding Author: ikisel@kip.uni-heidelberg.de

On-line processing of large data volumes produced in modern HEP experiments requires using maximum capabilities of the computer architecture. One of such powerful feature is a SIMD instruction set, which allows packing several data items in one register and to operate on all of them, thus achieving more operations per clock cycle. The novel Cell processor extends the parallelization further by combining a general-purpose PowerPC processor core with eight streamlined coprocessing elements which greatly accelerate vector processing applications.

In order to investigate a possible speed up of the reconstruction stage of data processing, we have ported a track fitting package based on the Kalman filter to the Cell processor. The overall speed up in 120000 times has been obtained on a Cell Blade computer compared to the initial scalar implementation on a Pentium 4 machine. Major steps of the porting procedure (memory optimization, numerical analysis, vectorization with inline operator overloading, and optimization using the Cell simulator) are presented and discussed.

Monday, 03 November 2008 - Morning session 2 / 6

Higgs bbbar decay at NNLO and beyond : the uncertainties of QCD predictions

Dr. KATAEV, Andrei [1]; Dr. KIM, Victor [2]

[1] *Institute for Nucleaer Research , Moscow, Russia*

[2] *St, Petersburg Institute for Nuclear Physics, Gatchina, Russia*

Corresponding Author: andrei.kataev@cern.ch

Different methods for treating the results of higher-order perturbative QCD calculations of the decay width of the Standard Model Higgs boson into bottom quarks are discusssed. Special attention is paid to the analysis of the M_H dependence of the decay width $\Gamma(H\to \bar{b}b})$ in the cases when the mass of b-quark is defined as the running parameter in the $\bar{MS}$-scheme and as the quark pole mass. The relation between running and pole masses is taken into account in the order α_s^4-approximation. The way of fixation of theoretical uncertainties of the f QCD predictions are fproposed. Thev results may be of interest in the process of matching with the results ofexisting NNLO Monte Carlo program to compute Higgs boson production at hadron colliders at NNLO

Methodology of Computations in Theoretical Physics - Session 1 / 8

Development, validation and maintenance of Monte Carlo event generators and generator services in the LHC era

Dr. KIRSANOV, Mikhail [1]; Dr. RIBON, Alberto [2]; Dr. ZENIN, Oleg [3]

[1] *Institute for Nuclear Research (INR), Moscow*

[2] *CERN*

[3] *Institute for High Energy Physics (IHEP), Protvino*

Corresponding Author: mikhail.kirsanov@cern.ch

The Generator Services project collaborates with the Monte Carlo generators authors and with the LHC experiments in order to prepare validated LCG compliant code for both the theoretical and the experimental communities at the LHC. On the one side it provides the technical support as far as the installation and the maintenance of the generators packages on the supported platforms is concerned and on the other side it participates in the physics validation of the generators.

The libraries of the Monte Carlo generators maintained within this project are currently widely adopted by the LHC collaborations and are used in large scale productions. The existing testing and validation tools are regularly used and the additional ones are being developed, in particular for the new object-oriented generators. The aim of the validation activity is also to participate in the tuning of the generators in order to provide appropriate settings for the proton-proton collisions at the LHC energy level.

This talk presents the current status and the future plans of the Generator Services project. The approach used in order to provide tested Monte Carlo generators for the LHC experiments is discussed and some of the testing and validation tools are presented.

Thursday, 06 November 2008 / 11

Forget multicore! The future is many-core: An outlook to the explosion of parallelism likely to occur in the LHC era

Mr. JARP, Sverre [1]

[1] *CERN*

Corresponding Author: sverre.jarp@cern.ch

This talk will start by reminding the audience that Moore's law is very much alive (even after 40+ years of existence).
Transistors will continue to double for every new silicon generation every other year.
Chip designers are therefore trying every possible "trick" for putting the transistors to good use.
The most notable one is to push more parallelism into each CPU: More and longer vectors, more parallel execution units,
more cores and more hyperthreading inside each core. In addition highly parallel grphics processing units (GPUs) are
also entering the game and compete efficiently with CPUs in several computing fields.
The speaker will try to predict the CPU dimensions we will reach during the LHC era, based on what we have seen in the recent past and the projected roadmap for silicon.
He will also discuss the impact on HEP software.
Can we continue to rely on event-level parallelism at the process levels or do we need to move to a new software paradigm?

Thursday, 06 November 2008 / 12

Throughput Computing in C++

Dr. GHULOUM, Anwar [1]

[1] *Intel Corporation*

Corresponding Author: anwar.ghuloum@intel.com

Power consumption is the ultimate limiter to current and future processor design, leading us to focus on more power efficient architectural features such as multiple cores, more powerful vector units, and use of hardware multi-threading (in place of relatively expensive out-of-order techniques). It is (increasingly) well understood that developers face new challenges with multi-core software development. The first of these challenges is a significant productivity burden particular to parallel programming. A big contributor to this burden is the relative difficulty of tracking down data races, which manifest non-deterministically. The second challenge is parallelizing applications so that they effectively scale with new core counts and the inevitable enhancement and evolution of the instruction set. This is a new and subtle qualifier to the benefits of backwards compatibility inherent in Intel® Architecture (IA): performance may not scale forward with new micro-architectures and, in some cases, actually regress. I assert that forward-scaling is an essential requirement for new programming models, tools, and methodologies intended for multi-core software development.
We are implementing a programming model called Ct (C for Throughput Computing) that leverages the strengths of data parallel programming to help address these challenges. Ct is a C++-hosted deterministic parallel programming model integrating the nested data parallelism of Blelloch and bulk synchronous processing of Valiant (with a dash of SISAL for good measure). Ct uses meta-programming and dynamic compilation to essentially embed a pure functional programming language in impure and unsafe C++. A key objective of the Ct project is to create both high-level and low-level abstractions that forward-scale across IA. I will describe the surface API and runtime architecture that we've built to achieve this, as well as some performance results

Wednesday, 05 November 2008 / 13

Code quality from the programmer's perspective

TONELLA, Paolo [1]

[1] *FBK-IRST*

Corresponding Author: tonella@fbk.eu

Code quality has traditionally been decomposed into internal and external quality. In this talk, I will discuss the differences between these two views and I will consider the contexts in which either of the two becomes the main quality goal. I will argue that for physics software the programmer's perspective, focused on the internal quality, is the most important one. Then, I will revise the available tools and techniques for the verification and improvement of the internal code quality, having in mind the programmer's perspective. I will conclude with a list of challenges for research in software engineering about aspects of the internal code quality that are largely neglected, but affect deeply the programmer's ability to carry out code modification and bug fixing tasks.
Such aspects revolve around the way in which the natural language is embedded into the code as a form of domain modeling.

Data Analysis - Algorithms and Tools / 14

PDE-FOAM - a probability-density estimation method based on self-adapting phase-space binning

Dr. CARLI, Tancredi [1]; Dr. DANNHEIM, Dominik [1]; Mr. VOIGT, Alexander [1]; Dr. SPECKMAYER, Peter [2]

[1] *CERN*

[2] *Technische Universität Wien*

Corresponding Author: dominik.dannheim@cern.ch

Probability-Density Estimation (PDE) is a multivariate discrimination technique based on sampling signal and background densities in a multi-dimensional phase space. The signal and background densities are defined by event samples (from data or monte carlo) and are evaluated using a binary search tree (range searching). This method is a powerful classification tool for problems with highly non-linearly correlated observables. In this paper, we present an innovative improvement of the PDE method that uses a self-adapting binning method to divide the multi-dimensional phase space in a finite number of hyper-rectangles (boxes). For a given number of boxes, the binning algorithm adjusts the size and position of the boxes inside the multidimensional phase space, minimizing the variance of the signal and background densities inside the boxes. The binned density information is stored in binary trees, allowing for a very fast and memory-efficient classification of events. The implementation of the binning algorithm (PDE-FOAM) is based on the monte-carlo integration package TFOAM included in the analysis package ROOT and has been developed within the framework of the Toolkit for Multivariate Data Analysis with ROOT (TMVA). We present performance results for representative examples (toy models) and discuss the dependence of the obtained results on the choice of parameters. The new PDE-FOAM is compared to the original PDE method based on range-searching.

15

Goodness of fit tests for weighted histograms

Prof. GAGUNASHVILI, Nikolai [1]

[1] *University of Akureyri, Iceland*

Corresponding Author: nikolai@unak.is

Weighted histograms in Monte-Carlo simulations are often used for the estimation of
a probability density functions. They are obtained as a result of random experiment with random events that have weights. In this paper the bin contents of weighted histogram are considered as a sum of random variables with random number of terms. Goodness of fit tests for weighted histograms and for weighted histograms with unknown normalization are proposed. Sizes and powers of the tests are investigated numerically.

Data Analysis - Algorithms and Tools / 16

Application of the rule-growing algorithm RIPPER to particle physics analysis

Dr. BRITSCH, Markward [1]; Prof. GAGUNASHVILI, Nikolai [2]; Prof. SCHMELLING, Michael [3]

[1] *Max-Planck-Institut fuer Kernphysik (MPI)-Unknown-Unknown*

[2] *University of Akureyri*

[3] *Max-Planck-Institut fuer Kernphysik (MPI)*

Corresponding Author: markward@mpi-hd.mpg.de

A large hadron machine like the LHC with its high track multiplicities always asks for powerful tools that drastically reduce the large background while selecting signal events efficiently. Actually such tools are widely needed and used in all parts of particle physics. Regarding the huge amount of data that will be produced at the LHC, the process of training as well as the process of applying these tools to data, must be time efficient. Such tools can be multi-variate analysis -- also called data mining -- tools. In this talk we present the results for the application of the multi-variate analysis, rule growing algorithm RIPPER on a problem of particle selection. Minimum-bias Monte-Carlo data for the LHCb-Experiment are used. It turns out that the meta-methods bagging and cost-sensitivity are essential for the quality of the outcome. The results are compared to other multi-variate analysis techniques as well as to the traditional cuts based method.

Methodology of Computations in Theoretical Physics - Session 2 / 17

Recent Progress of Geant4 Electromagnetic Physics and Readiness for the LHC Start

Prof. IVANTCHENKO, Vladimir [1]; Dr. MAIRE, Michel [2]

[1] *CERN, ESA*

[2] *LAPP*

Corresponding Author: vladimir.ivantchenko@cern.ch

The status of Geant4 electromagnetic (EM) physics models is presented, focusing on the models most relevant for collider HEP experiments, at LHC in particular. Recently improvements were undertaken in models for the transport of electrons and positrons, and for hadrons. Models revised included those for single and multiple scattering, ionization at low and high energies, bremsstrahlung, annihilation, scintillation and Cerenkov. Validation has been performed against experimental data. Typical results of comparisons are shown.

There was a significant update of the bremsstrahlung models. This introduced a new description of the relativistic regime for electrons and positrons, which describes precisely the recent LPM experiment at CERN. New models for bremsstrahlung and electron-positron pair production by hadrons were introduced. A significant effect is observed due to the bremsstrahlung of pions is observed, affecting the signal in EM calorimeters of LHC detectors.

With a focus on the LHC start-up, we discuss performance versus precision of different configurations of EM physics.

Data Analysis - Algorithms and Tools / 18

The Role of Interpreters in High Performance Computing

NAUMANN, Axel [1]; Mr. CANAL, Philippe [2]

[1] *CERN*

[2] *FERMILAB*

Corresponding Author: axel.naumann@cern.ch

Compiled code is fast, interpreted code is slow. There is not much we can do
about it, and it's the reason why interpreters use in high performance computing is usually restricted to job submission. I will show where interpreters make sense even in the context of analysis code, and what aspects have to be taken into account to make this combination a success.

Data Analysis - Algorithms and Tools / 19

C++ and Data

NAUMANN, Axel [1]; Mr. CANAL, Philippe [2]

[1] *CERN*

[2] *FERMILAB*

Corresponding Author: axel.naumann@cern.ch

High performance computing with a large code base and C++ has proved to be a
good combination. But when it comes to storing data, C++ is a really bad choice: it offers no support for serialization, type definitions are amazingly complex to parse, and the dependency analysis (what does object A need to be stored?) is incredibly difficult. Nevertheless, the LHC data consists of C++ objects that are serialized with help from ROOT's interpreter CINT. The fact that we can do it on that scale, and the performance with which we do it makes this approach unique and stirs interest even outside HEP. I will show how CINT collects and stores information about C++ types, what the current major challenges are (dictionary size!), and what CINT and ROOT have done and plan to do about it.

Computing Technology for Physics Research / 20

g-Eclipse - user and developer friendly access to Grids and Clouds

Dr. GARCIA, Ariel [1]; WOLNIEWICZ, Pawel [2]; KORNMAYER, Harald [3]; STUEMPERT, Mathias [1]

[1] *FORSCHUNGSZENTRUM KARLSRUHE, GERMANY*

[2] *Poznań Supercomputing and Networking Center, Poland*

[3] *NEC Laboratories Europe*

Corresponding Author: garcia@iwr.fzk.de

g-Eclipse is both a user friendly graphical user interface and a programming framework for accessing Grid and Cloud infrastructures. Based on the extension mechanism of the well known Eclipse platform, it provides a middleware independent core implementation including standardized user interface components. Based on these components, implementations for any available Grid and Cloud middleware can be added by using the Eclipse plug-in mechanism.

Currently g-Eclipse provides support for the gLite and GRIA Grid middlewares, as well as for the Amazon WebServices cloud computing (EC2) and storage (S3) offers. The data management component enables a seamless and interworking storage access across different middlewares. I.e. data from Amazon S3 can be moved to gLite resources just by the drag-and-drop mechanism. Furthermore, the tool provides Grid job management functionality including a JSDL standard conformant job description editor, automatic update of submitted jobs' status, parametric job creation and direct access to the job input and output files. Many other elements contribute to the g-Eclipse eco system to make up for a fully integrated user, operator and developer environment, like a graphical workflow editor, a batch system management component, a resources test framework, scientific visualization support and others. g-Eclipse provides a variety of integrated tools for the future e-Scientist working on the Grid and other emerging e-Infrastructures.

In this presentation we will introduce the framework and demonstrate the tool online by accessing the different supported infrastructures. The presentation will focus on the user and the developer's perspectives.

Computing Technology for Physics Research / 21

Dedicated Services to Support Data Replication over GRID using ATLAS Distributed Data Management System.

TITOV, Mikhail [1]; KLIMENTOV, Alexei [2]

[1] *Moscow Physical Engineering Inst. (MePhI)*

[2] *Brookhaven National Laboratory (BNL)*

Corresponding Author: mikhail.titov@cern.ch

There is ATLAS wide policy how different types of data is distributed between centers of different level (T0/T1/Tn) it is well defined and centrally operated activity (uses Atlas Central Services which include Catalogue services, Sites services, T0 services, Panda Services and etc). At the same ATLAS Operations Group designed user oriented services to allow ATLAS physicists to place data replication request and using Distributed Data Management (DDM) as a low level distribute data between more than 70 sites.

The DDM System consists of a bookkeeping system (dataset-based) and a set of local site services to handle data transfers, building upon Grid technologies. The software stack is called DQ2 [1].

So replication methods for physicist's requests include DQ2 service, DQ2 End-User tools [2], Web-based interface for user's requests (which is using internally DQ2 service) [3] and for replication control uses DDM/DQ2 data transfer monitoring. These methods supplement each other, because all of them are having some restrictions and policies, which are defined by ATLAS Operations Group or corresponding cloud coordinators.

[1] https://twiki.cern.ch/twiki/bin/view/Atlas/DistributedDataManagement
[2] https://twiki.cern.ch/twiki/bin/view/Atlas/DDMEndUserTutorial
[3] http://panda.atlascomp.org?mode=reqsubs0

Data Analysis - Algorithms and Tools / 22

Petaminer: Efficient Navigation to Petascale Data Using Event-Level Metadata

HAMILL, Paul [1]; CRANSHAW, Jack [2]; MALON, David [2]; VANIACHINE, Alexandre [2]

[1] *Tech-X Corporation*

[2] *Argonne National Laboratory*

Corresponding Author: sacha.vaniachine@cern.ch

HEP experiments at the LHC store petabytes of data in ROOT files described with TAG metadata. The LHC experiments have challenging goals for efficient access to this data. Physicists need to be able to compose a metadata query and rapidly retrieve the set of matching events. Such skimming operations will be the first step in the analysis of LHC data, and improved efficiency will facilitate the discovery process by permitting rapid iterations of data evaluation and retrieval. Furthermore, efficient selection of LHC data helps enable the tiered data distribution system adopted by LHC experiments, in which massive raw data resides at a few central sites, while higher quality, smaller scale skimmed data is replicated at many lower tier sites with more modest computational resources.

To address this problem, we are developing a custom MySQL storage engine to enable the MySQL query processor to directly access TAG data stored in ROOT TTrees. As ROOT TTrees are column-oriented, reading them directly will provide improved performance over traditional row-oriented TAG databases. In addition, to the efficient SQL query interface to the data stored in ROOT TTrees, the Petaminer technology will enable rich MySQL index-building capabilities to add indices to the data in ROOT TTrees, providing further optimization to TAG query performance.

Column-oriented databases are an emerging technique for achieving higher performance than traditional row-oriented databases, especially in large scale data-mining scenarios.

We will present first results of our feasibility studies of creating a column-oriented MySQL storage engine that enables MySQL to access TAG metadata directly from ROOT files.

23

Strategies for btagging calibration using data at CMS

Dr. BAZTERRA, Victor Eduardo [1]

[1] *Univ Illinois at Chicago*

Corresponding Author: tommaso.boccali@cern.ch

The CMS Collaboration is studying several algorithms to discriminate jets coming from the hadronization of b quarks from the lighter background. These will be used to identify top quarks and in searches of the Higgs boson and non-Standard Model processes. A reliable estimate of the performance of these algorithms is therefore crucial, and methods to estimate efficiencies and mistag rates directly on data are needed. While on simulated data it was shown the btagging algorithms reach adequate performance for 'standard model and beyond' analyses, when searching for the better b efficiency / light rejection, it is definitely not trivial to extract and validate these figures on real data. The CMS Monte Carlo simulation, even if tuned for more than 10 years, is in fact not expected to be reliable on the first data, and large discrepancies can also come from the experimental inputs on the production of heavy flavours. The CMS b-tagging group has prepared several strategies to extract efficiencies and rejection rates from data, which should work even on the first data expected during the 2008 running period (10 pb-1). Three methods are described that take into account startup conditions expected from both the LHC machine (luminosity) and the CMS detector (alignment/calibration).

Computing Technology for Physics Research / 24

The CMS Framework for Alignment and Calibration

MANKEL, Rainer [1]; FUTYAN, David [2]; FLUCKE, Gero [3]

[1] *DESY/CERN*

[2] *Imperial College London*

[3] *Universität Hamburg*

Corresponding Author: rainer.mankel@cern.ch

The ultimate performance of the CMS detector relies crucially on precise and prompt alignment and calibration of its components. A sizable number of workflows need to be coordinated and performed with minimal delay through the use of a computing infrastructure which is able to provide the constants for a timely reconstruction of the data for subsequent physics analysis. The framework supporting these processes and results from testing it in recent commissioning campaigns are presented.

Data Analysis - Algorithms and Tools / 26

Track Reconstruction and Muon Identification in the Muon Detector of the CBM Experiment at FAIR

Mr. LEBEDEV, Andrey [1]; Dr. HOEHNE, Claudia [2]; Dr. KISEL, Ivan [3]; Dr. KISELEVA, Anna [2]; Prof. OSOSKOV, Gennady [4]

[1] *GSI, Darmstadt / JINR, Dubna*

[2] *GSI, Darmstadt*

[3] *GSI, Darmstadt / KIP, University of Heidelberg*

[4] *JINR, Dubna*

Corresponding Author: andrey.lebedev@gsi.de

The Compressed Baryonic Matter (CBM) experiment at the future FAIR accelerator at Darmstadt is being designed for a comprehensive measurement of hadron and lepton production in heavy-ion collisions from 8-45 AGeV beam energy, producing events with large track multiplicity and high hit density. The setup consists of several detectors, including the silicon tracking system (STS) placed in a dipole magnet close to the target region, and the MUCH (MUon CHamber) detector. The MUCH detector is aimed for muon identification down to momenta of 1.5 GeV/c. It consists of a sequence of absorber layers and detector stations. The concept for the MUCH detector and the status of the track reconstruction software are presented in this contribution.
The reconstruction software is organized to be flexible with respect to feasibility studies of different physics channels and to optimization of the detector itself. The main blocks of the reconstruction package include track finding, fitting and propagation in the material of the detector. The track finding algorithm is based on the track following method with branches, using tracks reconstructed in the STS as initial seeds. The magnetic field is taken into account during extrapolation through the detector. Track propagation in the material includes accurate calculation of multiple scattering and energy losses. The performance of the track propagation is shown to be similar to the results of GEANE. Track parameters and covariance matrices are estimated using the Kalman filter method. At the final competition, tracks with larger number of hits and with better chi-square value are more preferable.
The track reconstruction efficiency for muons embedded in central Au+Au collisions at 25 AGeV beam energy from the UrQMD model is at the level of 95%. In these collisions, feasibility studies of low mass vector meson measurements in the dimuon channel result in 1.7% total reconstruction efficiency of the omega meson and a signal to background ratio of 0.15. Currently, ongoing work focuses on detector layout studies in order to optimize the detector setup while keeping a high performance.

27

Building Grid client applications with the g-Eclipse framework

WOLNIEWICZ, Pawel [1]; GARCIA, Ariel [2]

[1] *PSNC*

[2] *FZK*

Corresponding Author: pawelw@man.poznan.pl

The g-Eclipse is an integrated workbench framework to access the power of existing Grid infrastructures. g-Eclipse can be used on user level or application level. On user level g-Eclipse is just a rich client application with user friendly interface which allows users to access Grid resources, operators to manage Grid resources and developers to speed up the development cycle of new Grid applications. On application level the g-Eclipse components can be used as a base for writing customised grid application using g-Eclipse model and g-Eclipse common grid library.

There are three ways to prepare and manage grid applications job with g-Eclipse:
- provide XML application description,
- enhance g-Eclipse with specific application editor and submission service,
- prepare Rich Client Platform (RCP) application using g-Eclipse as a common grid library.

g-Eclipse supports the first approach by offering XML-based application descriptions and corresponding tools to specify application parameters in much more friendly way than just filling "Executable" and "Arguments" fields in JDL or JSDL. g-Eclipse offers tools to share such job description documents within application communities by using repositories like CVS or SVN. The same is valid for workflow description files. With the help of the g-Eclipse framework application group manager or VO manager can prepare applications, both simple or workflow, and share them with the rest of the group.

The second approach enhances the g-Eclipse framework with new plugins for application specific job description languages and submission methods. Many applications use job description files in the specific format (e.g. running Gaussian jobs requires ".inp" file with all data). The new plugins registers such files as grid job descriptions and supporting editors (textual or graphical ones) which can handle this file format. Using the existing g-Eclipse infrastructure, such job description are submitted to the Grid by transforming the application specific description in a middleware specific one.

The last method is the most advanced and can be used to prepare custom application. By extensively using the Eclipse extension mechanism in combination with object-oriented design patterns, the framework can be easily extended by application-specific implementations. The application is not plugged into g-Elipse, but rather built on top of it. It follows the approach of the Rich Client Platform (RCP) of Eclipse. Dedicated applications can be build by using parts of g-Eclipse as a common library to handle resources and grid access. RCP application can use the g-Eclipse core model only, or can use some GUI components, too. With such an approach existing applications can be gridyfied. The g-Eclipse core model is grid middleware independent. By electing the a appropriate set of plug-ins, the same application can access resources from different grid middleware. The g-Eclipse model defines interfaces for accessing virtual organisations, information system, job submission services, job monitoring services and grid file systems. There is also support for managing authentication tokens, storing submitted jobs locally, automatic notifications about job statuses and many others. If the results of computation needs to be visualised then another plug-in can be added that will handle the output file format and will visualise it, possibily using g-Eclipse visualisation.

g-Eclipse also include a Developer Perspective which contain tools that can be used while developing grid application for seamless integration of the remote development process into the existing Eclipse development environments. This include remote debugging support, application monitoring support, visualisation support, and application deployment.

Data Analysis - Algorithms and Tools / 28

MINUIT Package Parallelization and applications using the RooFit Package

Dr. LAZZARO, Alfio [1]; Dr. MONETA, Lorenzo [2]

[1] *Universita' degli Studi and INFN, Milano*

[2] *CERN*

Corresponding Author: alfio.lazzaro@mi.infn.it

MINUIT is the most common package used in high energy physics for numerical minimization of multi-dimensional functions. The major algorithm of this package, MIGRAD, searches for the minimum by using the function gradient. For each minimization iteration, MIGRAD requires the calculation of the first derivatives for each parameter of the function to be minimized.
In this presentation we will show how the algorithm can be easily parallelized using MPI techniques to scale over multiple nodes or multi-threads for multi-cores in a single node.
We will present the speed-up improvements obtained in typical physics applications such as complex maximum likelihood fits using the RooFit package. Furthermore, we will also show results of hybrid parallelization between MPI and multi-threads, to take full advantage of multi-core architectures.

29

EVE - Event Visualization Environment of the ROOT framework

Dr. TADEL, Matevz [1]

[1] *CERN*

Corresponding Author: matevz.tadel@cern.ch

EVE is a high-level environment using ROOT's data-processing, GUI and OpenGL interfaces. It can serve as a framework for object management offering hierarhical data organization, object interaction and visualization via GUI and OpenGL representations and automatic creation of 2D projected views. On the other hand, it can serve as a toolkit satisfying most HEP requirements, allowing visualization of geometry, simulated and reconstructed data such as hits, clusters, tracks and calorimeter information. Special classes are available for visualization of raw-data. EVE is used in the ALICE experiment as the standard visualization tool, AliEVE, using the full feature set of the environment. In the CMS experiment, EVE is used as the underlying toolkit of the cmsShow physics-analysis oriented event-display. Both AliEVE and cmsShow are also used for the online data-quality monitoring.

Data Analysis - Algorithms and Tools / 30

An overview of the b-Tagging algorithms in the CMS Offline software

Mr. SAOUT, Christophe [1]

[1] *CMS, CERN IEKP, University of Karlsruhe*

Corresponding Author: christophe.saout@cern.ch

The CMS Offline software contains a widespread set of algorithms to identify jets originating from the weak decay of b-quarks. Different physical properties of b-hadron decays like lifetime information, secondary vertices and soft leptons are exploited. The variety of selection algorithms range from simple and robust ones, suitable for early data-taking and online environments as the trigger system, to highly discriminating ones, exploiting all the information available. For the latter, a generic discriminator computing framework has been developed that allows to exploit the full power of multi-variate analysis techniques in an flexible way.

Methodology of Computations in Theoretical Physics - Session 2 / 31

Radiative corrections to Drell-Yan like processes in SANC

Dr. ARBUZOV, Andrej [1]; KOLESNIKOV, Vladimir [1]; SADYKOV, Renat [1]; ANDONOV, Anton [2]; Prof. BARDIN, Dmitry [1]; Dr. BONDARENKO, Serge [1]; Dr. CHISTOVA, Pena [1]; Dr. KALINOVSKAJA, Lidija [1]; Dr. NANAVA, Gizo [3]

[1] *Joint Institute for Nuclear Research (JINR)*

[2] *Shoumen State university, Bulgaria*

[3] *University of Bonn*

Corresponding Author: vladimir.kolesnikov@cern.ch

Radiative corrections to processes of single Z and W boson production are obtained within the SANC computer system. Interplay of one-loop QCD and electroweak corrections is studied. Higher order QED final state radiation is taken into account. Monte Carlo event generators at the hadronic level are constructed. Matching with general purpose programs like HERWIG and PYTHIA is performed to include the effect of partonic showers. Numerical results for LHC conditions are demonstrated. The resulting theoretical uncertainty in the description of these processes is discussed.

Computing Technology for Physics Research / 32

The PanDA System in the ATLAS Experiment

NILSSON, Paul [1]

[1] *University of Texas at Arlington*

Corresponding Author: paul.nilsson@cern.ch

The PanDA system was developed by US ATLAS to meet the requirements for full scale production and distributed analysis processing for the ATLAS Experiment at CERN. The system provides an integrated service architecture with late binding of job, maximal automation through layered services, tight binding with the ATLAS Distributed Data Management system, advanced job recovery and error discovery functions, among other features. This talk will present the components and performance of the PanDA system which has been in production in the US since early 2006 and ATLAS wide since the end of 2007.

Methodology of Computations in Theoretical Physics - Session 2 / 33

Standard SANC Modules

Dr. BONDARENKO, Serge [1]; Prof. BARDIN, Dmitry [1]; ANDONOV, Anton [2]; Dr. ARBUZOV, Andrey [1]; Dr. KALINOVSKAJA, Lidija [1]; SADYKOV, Renat [1]; KOLESNIKOV, Vladimir [1]; Dr. CHISTOVA, Pena [1]

[1] *Joint Institute for Nuclear Research (JINR)*

[2] *Shoumen State university, Bulgaria*

Corresponding Author: vladimir.kolesnikov@cern.ch

Two types of SANC system output are presented.
At first the status of stand-alone packages for calculations of the EW and
QCD NLO RC at the parton level (Standard SANC FORM and/or FORTRAN Modules)
are done. Short overview of these packages in sector of the Neutral Current:
(uu, dd) -> (mu,mu, ee) and ee(uu, dd) -> HZ;
and in the sector of the Charge Current:
ee(uu, dd) -> (mu nu_mu, e nu_e) are described.
In addition second type of SANC output -- MC event generator for
production of event distributions at the hadronic level, based on the FOAM
algorithm, are demonstrated.

Computing Technology for Physics Research / 34

Distributed analysis in CMS using CRAB: the client-server architecture evolution and commissioning

CODISPOTI, Giuseppe [1]

[1] *Dipartimento di Fisica*

Corresponding Author: giuseppe.codispoti@cern.ch

CRAB (CMS Remote Analysis Builder) is the tool used by CMS to enable running physics analysis in a transparent manner over data distributed across many sites. It abstracts out the interaction with the underlying batch farms, grid infrastructure and CMS workload management tools, such that it is easily usable by non-experts.

CRAB can be used as a direct interface to the computing system or can delegate the user task to a server. Major efforts have been dedicated to the client-server system development, allowing the user to deal only with a simple and intuitive interface and to delegate all the work to a server.

The server takes care of handling the users jobs during the whole lifetime of the users task. In particular, it takes care of the data and resources discovery, process tracking and output handling. It also provides services such as automatic resubmission in case of failures, notification to the user of the task status, and automatic blacklisting of sites showing evident problems beyond what is provided by existing grid infrastructure.

The CRAB server architecture and its deployment will be presented, as well as the current status and future development. In addition the experience in using the system for initial detector commissioning activities and data analysis will be summarized.

Data Analysis - Algorithms and Tools / 35

VISPA: a Novel Concept for Visual Physics Analysis

ERDMANN, Martin [1]; FISCHER, Robert [1]; HINZMANN, Andreas [1]; KIRSCH, Matthias [1]; KLIMKOVICH, Tatsiana [1]; MUELLER, Gero [1]; PLUM, Matthias [1]; STEGGEMANN, Jan [1]; ACTIS, Oxana [1]

[1] *RWTH-Aachen*

Corresponding Author: klimkovich@physik.rwth-aachen.de

VISPA is a novel graphical development environment for physics analysis, following an experiment-independent approach. It introduces a new way of steering a physics data analysis, combining graphical and textual programming. The purpose is to speed up the design of an analysis, and to facilitate its control.

As the software basis for VISPA the C++ toolkit Physics eXtension Library (PXL) is used which is a successor project of the Physics Analysis eXpert (PAX) package. The most prominent features of this toolkit are the management of relations, a copyable container holding different aspects of physics events, the ability to store arbitrary user data, and a fast I/O.

In order to support modular physics analysis, VISPA provides a module handling system using the above mentioned event container as the interface. Several analysis modules are provided, e.g. a module for automated reconstruction of particle cascades. All modules can be steered through Python scripts. Physicists can easily write their own modules to the module handling system or extend the existing ones.

The concept of VISPA will be presented. Some application examples for different physics analyses will be shown.

Computing Technology for Physics Research / 36

Mass Storage System for Disk and Tape resources at the Tier1.

RICCI, Pier Paolo [1]

[1] *INFN CNAF*

Corresponding Author: pierpaolo.ricci@cnaf.infn.it

The activities in the last 5 years for the storage access at the INFN CNAF Tier1 can be enlisted under two different solutions efficiently used in production: the CASTOR software, developed by CERN, for Hierarchical Storage Manager (HSM), and the General Parallel File System (GPFS), by IBM, for the disk resource management.
In addition, since last year, a promising alternative solution for the HSM, using Tivoli Storage Manager (TSM) and GPFS, has been under intensive test.
This paper reports the description of the current hardware and software installation with an outlook on the last GPFS and TSM tests results.

Methodology of Computations in Theoretical Physics - Session 2 / 37

Hadronic Physics in Geant4: Improvements and Status for LHC Start

Prof. IVANTCHENKO, Vladimir [1]; Dr. APOSTOLAKIS, JOhn [2]; Dr. FOLGER, Gunter [2]; Dr. RIBON, Alberto [2]

[1] *CERN, ESA*

[2] *CERN*

Corresponding Author: vladimir.ivantchenko@cern.ch

An overview of recent developments for the Geant4 hadronic modeling is provided with a focus on the start of the LHC experiments. Improvements in Pre-Compound model, Binary and Bertini cascades, models of elastic scattering, quark-gluon string and Fritiof high energy models, and low-energy neutron transport were introduced using validation versus data from thin target experiments. Many of these developments were directed to improve simulation of hadronic showers for LHC. As a result, starting from Geant4 8.3, the Physics List QGSP_BERT describes reasonably well all the main observables that have been measured in different test-beam setups for ATLAS and CMS experiments.

Methodology of Computations in Theoretical Physics - Session 2 / 38

Feynman Diagrams, Differential Reduction and Hypergeometric Functions

Dr. KALMYKOV, mikhail [1]; Prof. KNIEHL, Bernd [2]

[1] *Hamburg U./JINR*

[2] *II. Institut für Theoretische Physik, Hamburg U.*

Corresponding Author: kalmykov@theor.jinr.ru

Recent results related with manipulation of hypergeometric functions: reduction and construction of higher-order terms in epsilon-expansion is revised. The application of given technique to the analytical evaluation of Feynman diagrams is considered.

Methodology of Computations in Theoretical Physics - Session 1 / 39

Current status of FORM parallelization

TENTYUKOV, Mikhail [1]; Dr. VERMASEREN, Jos [2]

[1] *Karlsruhe University*

[2] *NIKHEF*

Corresponding Author: tentukov@particle.uni-karlsruhe.de

We report on the status of the current development in parallelization of the symbolic manipulation system FORM. Most existing FORM programs will be able to take advantage of the parallel execution, without the need for modifications.

Computing Technology for Physics Research / 40

ATLAS Handling Problematic Events in Quasi Real-Time

GARITAONANDIA, Hegoi [1]; KLOUS, Sander [1]; PETERSEN, Brian [2]; SFYRLA, Anna [3]

[1] *NIKHEF*

[2] *CERN*

[3] *UCI*

Corresponding Author: hegoi.garitaonandia@cern.ch

The ATLAS experiment at CERN will require about 4000 CPUs for the online data acquisition system (DAQ). When the DAQ system experiences software errors, such as event selection algorithm problems, crashes or timeouts, the fault tolerance mechanism routes the corresponding event data to the so called debug stream. During first beam commissioning and early data taking, a large fraction of events is expected to end up in this stream. In order to identify problems with the DAQ as soon as possible and reduce the turn-around time for fixing these problems, it is of prime importance to treat the debug stream. We have adopted a quasi real-time approach. We have developed an automated system that analyzes the contents of the debug stream and provides fine grained error classification. A high percentage of error events is related to online transient problems. Many of those events are recovered by feeding them to an independent system that reruns the trigger software. To be flexible in terms of computing power requirements, we added a layer of abstraction over the computing backend. This gives the possibility of using the Grid as well as dedicated resources. Using cosmic ray runs, we validated the automatic error analysis and recovery procedure.

Methodology of Computations in Theoretical Physics - Session 1 / 41

The Monte Carlo generators in CMS

Dr. BARTALINI, Paolo [1]

[1] *CERN*

Corresponding Author: paolo.bartalini@cern.ch

The CMS collaboration supports a wide spectrum of Monte Carlo generator packages in its official production, each of them requiring a dedicated software integration and physics validation effort. We report on the progress of the usage of these external programs with particular emphasis on the handling and tuning of the Matrix Element tools. The first integration tests in a large scale production for a new family of Object Oriented Monte Carlo generators are also reported.

Data Analysis - Algorithms and Tools / 42

Enhanced Gene Expression Programming for signal-background discrimination in particle physics

TEODORESCU, Liliana [1]; HUANG, Zhengwen [1]

[1] *Brunel University*

Corresponding Author: liliana.teodorescu@cern.ch

In order to address the data analysis challenges imposed by the complexity of the data generated by the current and future particle physics experiments, new techniques for performing various analysis tasks need to be investigated. In 2006 we introduced to the particle physics field one such new technique, based on Gene Expression Programming (GEP), and successfully applied it to an event selection problem. While GEP, as initially proposed, was proven to be more flexible and more efficient than other evolutionary algorithms, it does not incorporate many of the advanced developments in the field of evolutionary computation. This paper will present our developments of the algorithm and will discuss results obtained with alternative mapping mechanisms between the solution space and the representation space, the effect of a more controlled selection process of the candidate solutions, and of adaptable discrimination thresholds for supervised classification problems. The enhanced version of the algorithm was applied to a signal-background discrimination problem in a particle physics data analysis. Comparative studies of the initial and the enhanced version of GEP were performed and the results will be presented and discussed.

Computing Technology for Physics Research / 43

Large Scale Job Management and Experience in Recent Data Challenges within the LHC CMS experiment.

Dr. EVANS, David [1]; Dr. METSON, Simon [2]; Dr. WAKEFIELD, Stuart [3]; Dr. HUFNAGEL, Dirk [4]; Mr. HASSAN, Ahmad [4]; Dr. MOHAPATRA, Ajit [5]; Dr. MASON, David [1]; Dr. GUTSCHE, Oliver [1]; Dr. MILLER, Mike [6]; Mr. VAN LINGEN, Frank [7]

[1] *Fermilab, Batavia, IL, USA*

[2] *H.H. Wills Physics Laboratory, Bristol University*

[3] *Imperial College London*

[4] *CERN, Geneva, Switzerland*

[5] *Wisconsin*

[6] *MIT*

[7] *Caltech*

Corresponding Author: simon.metson@cern.ch

From its conception the job management system has been distributed to increase scalability and robustness. The system consists of several applications (called prodagents) which each manage Monte Carlo, reconstruction and skimming jobs on collections of sites within different Grid environments (OSG, NorduGrid?, LCG) and submission systems (GlideIn?, local batch, etc..).

Production of simulated data in CMS will take place mainly on so called Tier2s (small to medium size computing centers) resources. Approximately ~50% of the CMS Tier2 resources are allocated to running simulation jobs. While the so called Tier1s (medium to large size computing centers with high capacity tape storage systems) will be mainly used for skimming and reconstructing detector data. During the last one and a half years the system has also been adapted such that it can be configured for converting Data Acquisition (DAQ)/ High Level Trigger (HLT) output from the CMS detector to the CMS data format and manage the real time data stream from the experiment. Simultaneously the system has been upgraded to facilitate the increasing scale of the CMS production and adapting to the procedures used by its operators.

In this paper we discuss the current (high level) architecture of ProdAgent, the experience in using this system in computing challenges, feedback from these challenges, and future work including migration to a set of core libraries to facilitate convergence between the different data management projects within CMS that deal with analysis, simulation, and initial reconstruction of real data. This migration is important as it will decrease the code footprint used by these projects and increase maintainability of the code base.

44

CMS Dataset Bookkeeping, Discovery and Query System

LUEKING, Lee [1]; AFAQ, Anzar [1]; SEKHRI, Vijay [1]; KUZNETSOV, Valentin [2]

[1] *Fermilab, Batavia, IL, USA*

[2] *Cornell University, Ithaca, NY, USA*

Corresponding Author: simon.metson@cern.ch

The CMS experiment has implemented a flexible and powerful approach to enable users to find data within the CMS physics data catalog. The Dataset Bookkeeping Service (DBS) comprises a database and the services used to store and access metadata related to its physics data. In addition to the existing WEB based and programatic API, a generalized query system has been designed and built. This query system has a query language that hides the complexity of the underlying database structure. This provides a way of querring the system that is straightforward for CMS data managers and physicists. We will describe the design of the query system and provide details of the language components. We will also provide an overview of how this component fits into the overall data discovery, and bookkeeping systems for CMS.

Computing Technology for Physics Research - Session 1 / 45

Data transfer infrastructure for CMS data taking

Mr. EGELAND, Ricky [1]; Dr. METSON, Simon [2]; Dr. WILDISH, Tony [3]

[1] *University of Minnesota – Twin Cities, Minneapolis, MN, USA*

[2] *Bristol University, Bristol, UK*

[3] *Princeton University*

Corresponding Author: simon.metson@cern.ch

The CMS PhEDEx (Physics Experiment Data Export) project is responsible for facilitating large-scale data transfers across the grid ensuring transfer reliability, enforcing data placement policy, and accurately reporting results and performance statistics. The system has evolved considerably since its creation in 2004, and has been used daily by CMS since then. Currently CMS tracks over 2 PB of data in PhEDEx, and it has been tested well beyond the requirements of CMS.

Over the past year PhEDEx has evolved considerably, making use of new technologies (chiefly POE, an asynchronous, event-driven, cooperative-multitasking framework) and to consolidate the various components such that it is easy to reuse existing techniques and components in new features. This has resulted in changes to nearly every piece of the PhEDEx code base, creating a flexible modular framework. We are able to evolve the implementation to match changes in the requirements of the experiment, without changing the fundamental design.

Two major new features have recently been added to the PhEDEx system; an extensible data service and an improved transfer back-end module. The extensible data service provides machine-readable data over HTTP as the primary means of integration with other CMS services. An authenticated command line interface is also provided, making it possible to provide new utilities quickly with minimal development effort. The new transfer back-end module now integrates closely with FTS, the glite provided transfer tool, to provide accurate status information while keeping as much data in flight as possible. The new transfer back-end is transfer technology independent, and we expect to be able to support new transfer tools as they become available.

We describe the CMS PhEDEx system that is in place for CMS "first data taking" in 2008, provide details on the benefits and implementations of the new features, and describe other new tools that are now available.

46

Software validation and performance optimisation in the CMS experiment

HEGNER, Benedikt [1]; BENELLI, Gabriele [1]; LANGE, David [2]; PFEIFFER, Andreas [1]; GUTSCHE, Oliver [3]

[1] *CERN*

[2] *LLNL*

[3] *FNAL*

Corresponding Author: benedikt.hegner@cern.ch

The CMS experiment at LHC has a very large body of software of its own and uses extensively software from outside the experiment. Ensuring the software quality of such a large project requires checking and testing at every level of complexity. The aim is to give the developers very quick feedback on all the relevant CMS offline workflows during the (twice daily) Integration Builds. In addition the computing requirements constrain performance in terms of cpu time, memory footprint and event size on disk. A complex set of validation steps is used to verify the software at various stages from the regular Integration Builds to running a full s/w and physics validation suite on the grid for major releases.

In this talk, we briefly describe the software validation tools used and present the status of the performance of the CMS software at the start of data taking.

Data Analysis - Algorithms and Tools / 47

PARADIGM, a Decision Making Framework for Variable Selection and Reduction in High Energy Physics

GLEYZER, Sergei V. [1]

[1] *Florida State University*

Corresponding Author: sergei.v.gleyzer@cern.ch

In high energy physics, variable selection and reduction are key to a high quality multivariate analysis. Initial variable selection often leads to a variable set cardinality greater than the underlying degrees of freedom of the model, which motivates the needs for variable reduction and more fundamentally, a consistent decision making framework. Such a framework called PARADIGM, based on a global reduction criterion called the gloss function and relevant to searches for new phenomena in physics, is described. We illustrate the common pitfalls of variable selection and reduction associated with variable interactions and show that PARADIGM gives consistent results in their presence. We discuss PARADIGM's application to several HEP searches for new phenomena and compare the performance of different measures of relative variable importance, in particular with those based on binary regression. Finally we describe a technique called variable amplification that shows how PARADIGM's results lead to improved classification performance.

48

An artificial neural network based algorithm for calorimetric energy measurements in CMS

GLEYZER, Sergei V. [1]

[1] *Florida State University*

Corresponding Author: sergei.v.gleyzer@cern.ch

The Compact Muon Solenoid (CMS) experiment features an electromagnetic calorimeter (ECAL) composed of lead tungstate crystals and a sampling hadronic calorimeter (HCAL) made of brass and scintillator, along with other detectors. For hadrons, the response of the electromagnetic and hadronic calorimeters is inherently different. Because sampling calorimeters measure a fraction of the energy spread over several measuring towers, the energy resolution as well as the linearity are not easily preserved, especially at low energies. Several sophisticated algorithms have been developed to improve the resolution of the CMS calorimeter system for single particles. One such algorithm, based on artificial neural network application to a combined electromagnetic and hadronic calorimeter system, was developed and applied to test beam data using particles in the momentum range of 3-300 GeV/c. The method uses multivariate machinery to improve the energy measurement and linearity, especially at low energies below 10 GeV/c. Details of the algorithm are presented and a comprehensive comparison is made to other available methods.

Data Analysis - Algorithms and Tools / 49

WebDat: Bridging the Gap Between Unstructured and Structured Data

Dr. NOGIEC, Jerzy [1]

[1] *FERMI NATIONAL ACCELERATOR LABORATORY*

Corresponding Author: nogiec@fnal.gov

Accelerator R environments produce data characterized by different levels of organization. Whereas some systems produce repetitively predictable and standardized structured data, others may produce data of unknown or changing structure. In addition, structured data, typically sets of numeric values, are frequently logically connected with unstructured content (e.g., images, graphs, comments). Despite these different characteristics, a coherent, organized and integrated view of all information is sought out. WebDat is a system conceived as a result of efforts in this direction. It provides a uniform and searchable view of structured and unstructured data via common metadata, regardless of the repository used (DBMS or file system). It also allows for processing data and creating interactive reports. WebDat supports metadata management, administration, data and content access, application integration via Web services, and Web-based collaborative analysis.

Computing Technology for Physics Research - Session 1 / 51

Profiling Post-GRID analysis

Dr. SHIBATA, Akira [1]

[1] *New York University*

Corresponding Author: akira.shibata@nyu.edu

An impressive amount of effort has been put in to realize a set of frameworks to support analysis in this new paradigm of GRID computing. However, much more than half of a physicist's time is typically spent after the GRID processing of the data. Due to the private nature of this level of analysis, there has been little common framework or methodology.

While most physicists agree to use ROOT as the basis of their analysis, a number of approaches are possible for the implementation of the analysis using ROOT: conventional methods using CINT/ACLiC, development using g++, alternative interface through python, and parallel processing methods such as PROOF are some of the choices currently available on the market. Furthermore, in the ATLAS collaboration an additional layer of technology adds to the complexity because the data format is based on the POOL technology, which tends to be less portable.

In this study, various modes of ROOT analysis are profiled for comparison with the main focus on the processing speed. Input data is or derived from the ATLAS Full-Dress-Rehearsal, which was meant to stress test the whole computing system of ATLAS.

Computing Technology for Physics Research - Session 1 / 52

Distributed Computing in ATLAS

JIM, Shank [1]; BOS, Kors [2]; LAMANNA, Massimo [3]; KLIMENTOV, Alexei [4]; BARBERIS, Dario [3]

[1] *Boston University*

[2] *NIKHEF/CERN*

[3] *CERN*

[4] *BNL*

Corresponding Author: shank@bu.edu

The LHC machine has just started operations. Very soon, Petabytes of data from the ATLAS detector will need to be processed, distributed worldwide, re-processed and finally analyzed. This data-intensive physics analysis chain relies on a fabric of computer centers on three different sub-grids: the Open Science Grid, the LHC Computing Grid and the Nordugrid Data Facility--all part of the Worldwide LHC Computing Grid (wLCG). This fabric is arranged in a hierarchy of computing centers form Tier0 to Tier3. The role of the Tier-0 center is to perform prompt reconstruction of the raw data coming from the on-line data acquisition system, and to distribute raw and reconstructed data to the associated Tier-1 centers. The Tier1 centers mainly do raw data reprocessing after updated software releases and calibration constants are ready. The Tier2 centers have two major roles: simulation and physics analysis. This talk will describe the software components of the ATLAS data chain and the flow of data from the Tier0 center at CERN to the distributed Tier1, Tier2 and Tier3 centers.
There are five major components which will be discussed. The ATLAS Distributed Data Management system, that is responsible for all data movement and registration in ATLAS. The Storage Resource Management system for dealing with heterogeneous local storage systems. The PanDA pilot based system used to run managed production for both simulated data and real data re-processing. The detailed monitoring system (ARDA dashboard monitoring system) which allows us to debug problems. Finally, the systems which allow distributed physics analysis called GANGA and pAthena.

Data Analysis - Algorithms and Tools / 53

Alignment of the ATLAS Inner Detector tracking system

SCHIECK, Jochen [1]; Mr. ALISON, John [2]

[1] *Max-Planck-Institut für Physik*

[2] *Department of Physics and Astronomy, University of Pennsylvania*

Corresponding Author: jochen.schieck@cern.ch

The CERN's Large Hadron Collider (LHC) is the world largest particle accelerator. It will collide two proton beams at an unprecedented center of mass energy of 14 TeV and first colliding beams are expected during summer 2008. ATLAS is one of the two general purpose experiments that will record the decay products of the proton-proton collisions. ATLAS is equipped with a charge particle tracking system built on two technologies: silicon and drift tube based detectors, composing the ATLAS Inner Detector (ID). The performance of the Inner Detectors has to be optimized in order to achieve the ATLAS physics goals. The alignment of the tracking system poses a challenge as one should solve a linear equation with almost 36000 degrees of freedom. The required precision for the alignment of the most sensitive coordinates of the silicon sensors is just few microns. This limit comes from the requirement that the misalignment should not worsen the resolution of the track parameter measurements by more than 10%. Therefore the alignment of the ATLAS ID requires complex algorithms with extensive CPU and memory usage.

So far the proposed alignment algorithms are exercised on several applications. We will present the outline of the alignment approach and results from a Combined Test Beam, Cosmic Ray runs and large scale computing simulation of physics samples mimicking the ATLAS operation during real data taking. For the later application the trigger of the experiment is simulated and the event filter is applied in order to produce an alignment input data stream. The full alignment chain is tested using that stream and alignment constants are produced and validated within 24 hours. Cosmic ray data serves to produce an early alignment of the real ATLAS Inner Detector even before the LHC start up. Beyond all tracking information, the assembly survey data base contains essential information in order to determine the relative position of one module with respect to its neighbors. Finally a hardware system measuring an array of grid lines in the modules support structure with a Frequency Scan Interferometer monitors short time system deformations.

54

PERFORMANCE ASSESSMENT OF THE 2008 CONFIGURATION OF THE CICC JINR CLUSTER

Mr. AYRIYAN, Alexander [1]; Prof. ADAM, Gheorghe [2]; Dr. ADAM, Sanda [2]; Dr. DUSHANOV, Ermuhammad [1]; Prof. KORENKOV, Vladimir [1]; Mr. LUTSENKO, Alexey [1]; Mr. MITSYN, Valery [1]

[1] *JINR*

[2] *JINR, Dubna/IFIN-HH, Bucharest*

Corresponding Author: ayriyan@jinr.ru

The CICC JINR cluster has been installed in 2007-2008 years increasing computational power and disk space memory. It is generally used for distributed computing as part of Russian Data Intensive Grid (EGEE-RDIG) to work in LHC Computing Grid (LCG).
With the just installed superblade modules at mid-May 2008, the CICC JINR cluster reached a heterogeneous 560-core structure. The system consists three different homogeneous parts now. Here we discuss an independent assessment of its performance by means of the High Performance LINPACK (HPL) Benchmark and compare it with the results reported for the homogeneous 240-core configuration installed in the first half of 2007. In addition we discuss the superposition of the performances of each homogeneous part.
To define the consistency of the measured computing time dependence on the system dimension entering the HPL benchmark, we implemented a least squares fit procedure the output of which showed that, under the assumption of independent time measurements, the optimal polynomial fitting degree equals three, in agreement with the problem complexity following from the count of the number of involved elementary operations in the LU-decomposition solution of the algebraic system.
A comparison with the reported statistics on the TOP500 data show that the home made free software installation is consistent with the similar works worldwide.

55

Constraint Vertex Fitting in the ATLAS Experiment at LHC

Mrs. LIMPER, Maaike [1]; Dr. LIEBIG, Wolfgang [1]; Dr. PROKOFIEV, Kirill [2]

[1] *NIKHEF*

[2] *CERN*

Corresponding Author: wolfgang.liebig@cern.ch

On behalf of the ATLAS Collaboration.

The ATLAS collaboration at the Large Hadron Collider at CERN intends to study a variety of final states produced in proton-proton collisions at the energy of 14 TeV. The precise reconstruction of trajectories of charged and neutral particles including those which underwent decays is crucial for many phycics analyses. In addition, a study of tracking performance on the well-known benchmark channels will help understanding the properties of ATLAS detectors during the initial phase of the LHC. For instance, the investigation of correlations of parameters of the final state tracks having the same mother particle can help to correct for systematic effects in the alignment of the detector.

Described in this talk is the ATLAS framework for vertex reconstruction with simultaneous application of kinematic constraints. The implementation follows a very modular design based on object oriented C++.
A special emphasis is given on the use of the current approach in the alignment of the ATLAS detector and the reconstruction of exclusive decays.

Methodology of Computations in Theoretical Physics - Session 1 / 56

Computational aspects for three-loop DIS calculations

Dr. ROGAL, Mikhail [1]

[1] *DESY*

Corresponding Author: mrogal@ifh.de

will be sent later

Computing Technology for Physics Research - Session 1 / 57

The CMS Tier 0

Dr. EVANS, David [1]; Dr. HUFNAGEL, Dirk [2]; Dr. METSON, Simon [3]; Dr. MASON, David [1]; Dr. GOWDY, Stephen [2]; FOULKES, Stephen [1]; SEANGCHAN, Ryu [1]

[1] *Fermilab, Batavia, IL, USA*

[2] *CERN, Geneva, Switzerland*

[3] *Bristol University, Bristol, UK*

Corresponding Author: simon.metson@cern.ch

The CMS Tier 0 is responsible for handling the data in the first period of it's life, from being written to a disk buffer at the CMS experiment site in Cessy by the DAQ system, to the time transfer completes from CERN to one of the Tier1 computing centres. It contains all automatic data movement, archival and processing tasks run at CERN.

This includes the bulk transfers of data from Cessy to a Castor disk pool at CERN, repacking the data into Primary Datasets, storage to tape of and export to the Tier 1 centres. It also includes a first reconstruction pass over all data and and the tape archival and export to the Tier1 centres of the reconstructed data. While performing these tasks, the Tier 0 has to maintain redundant copies of the data and flush it through the system within a narrow time window to avoid data loss.

With data taking being imminent, this aspect of the CMS computing effort becomes of the upmost importance. We discuss and explain here the work developing and commissioning the CMS Tier0 undertaken over the last year.

58

Improving replica placement strategies using information from existing communication infrastructures

Mr. LASSNIG, Mario [1]; Mr. HALL, Mark Michael [2]

[1] *CERN University of Innsbruck, Austria*

[2] *Cardiff University, Wales, UK*

Corresponding Author: mario.lassnig@cern.ch

In highly data-driven environments such as the LHC experiments a reliable and high-performance distributed data management system is a primary requirement. Existing work shows that intelligent data replication is the key to achieving such a system, but current distributed middleware replication strategies rely mostly on computing, network and storage properties when deciding how to replicate data-sets across a global set of data centres.

While the distributed nature of such data management systems reduces the requirement for co-location of data and users interested in specific data, reliability and performance considerations mean that where possible co- or close-location are preferred.
We present an approach for improving existing replication strategies based on geographical data available in existing communication infrastructures. Information on the geographical distribution of interested users is extracted from an existing communication infrastructure using automated analysis of locational expressions in research documentation, operational logbooks, e-mail correspondence or web presences. Combined with the linking of data-sets to interested users this allows for an intelligent, anticipatory data replication strategy for data placement at locations close to the interested users.

Methodology of Computations in Theoretical Physics - Session 2 / 59

Mathematical model of magnetically interacting rigid bodies

Dr. ZUB, Stanislav [1]

[1] *National Science Center, Kharkov Institute of Physics and Techn*

Corresponding Author: stanislav.zub@cern.ch

Dynamics of two bodies, which interacts by magnetic forces, is considered. Model of interaction builds on quasi-stationary approach for electromagnetic field, and symmetric rotors with different moments of inertia of the bodies are considered. Interaction energy general form is discovered for the case of coincidence of mass and magnetic symmetries. Since the energy of interaction depends only from relative position of the bodies, then the consideration is too much simplified in c.m. system, notwithstanding that force is noncentral. The task requires development of the classic Hamilton formalism for the systems of magnetically interactive bodies, including the systems of the magnets and/or superconductive magnets (mixed systems). Hamilton motion equations are obtained on the basis of Poisson structure in the dynamic variables area. Such approach allows represent the equations in galilei-invariant vector form in contrast to default definition in Euler's angles. Invariance laws follow from system symmetry is considered. This variant of Hamilton formalism easily spreads in the case of arbitrary number of magnetically interactive symmetric symmetric rotors. All equations with Poisson brackets are tested with symbolic features of Maple system. For the numeral modelling of magnetic rigid bodies dynamics Maple and MATLAB packages are used. The obtained mathematical model allows investigate the possibility of orbital motion in the system of magnetically interactive bodies.

Computing Technology for Physics Research / 60

Software development, release integration and distribution tools for the CMS experiment

LANGE, David [1]; ELMER, Peter [2]; PFEIFFER, Andreas [3]; EULISSE, Giulio [4]; MUZAFFAR, Shahzad [4]; RATNIKOVA, Natalia [5]

[1] *LLNL*

[2] *Princeton University*

[3] *CERN*

[4] *Northwestern University*

[5] *FNAL*

Corresponding Author: david.lange@cern.ch

The offline software suite of the CMS experiment must support the production and analysis activities across a distributed computing environment. This system relies on over 100 external software packages and includes the developments of more than 250 active developers. This system requires consistent and rapid deployment of code releases, a stable code development platform, and efficient tools to enable code development and production work across the facilities utilized by the experiment. Recent developments have resulted in significant improvements in these areas. We report the concept, status, recent improvements and future plans for these aspects of the CMS offline software environment.

Methodology of Computations in Theoretical Physics - Session 1 / 61

LCG MCDB and HepML, next step to unified interfaces of Monte-Carlo Simulation

Mr. BELOV, Sergey [1]; DUDKO, Lev [2]; SHERSTNEV, Alexander [3]

[1] *JINR, Dubna*

[2] *Faculty of Physics*

[3] *University of Cambridge*

Corresponding Author: sergey.belov@cern.ch

In this talk we present a way of making Monte-Carlo simulation chain fully automated.automation
Last years there was a need for common place to store sophisticated MC event samples prepared by experienced theorists. Also such samples should be accessible in some standard manner to be easyly imported and used in experiments' software.
The main motivation behind the LCG MCDB project is to make sophisticated MC event samples and their structured descriptions available for various groups of physicists working on LHC. All the data from MCDB is accessible for end-users in several convenient ways from Grid, on the Web and via application program interface.
Developed in collaboration of LCG MCDB and CEDAR teams and several MC generator authors, HepML (High Energy Markup Language) is aimed be a unified XML description of event samples simulated by Matrix Element (ME) generators. The other main purpose of HepML is to keep MC generation parameters for further MC generators tuning.
It is possible to extend HepML as an XML standard to keep necessary information for the different levels of simulation in HEP, from theoretical model to a simulation of detector responds. HepML provides the possibility to use and develop many standard tools for the comparison, validation, graphical representation of the results and create transparent unified interfaces for the different software in HEP on the modern level of Computer science.
Using MCDB and HepML together gives a possibility of automation of such significant part of MC simulation chain as correct transfer physics events from Matrix Element generators to Shower generators and then to detector simulation. Such machine-driven manner allows to avoid errors coming from human factor (physical data are storing with complete unified description directly from MC generator), save a lot of time and efforts for end users of trusted and verified shared MC samples.

LCG MCDB is developing within CERN LCG Application Area Simulation Project.
This talk is given on behalf of the Generator Services subproject.

Tuesday, 04 November 2008 - Morning session 2 / 62

Data Analysis with PROOF

Dr. RADEMAKERS, Fons [1]; Dr. GANIS, Gerardo [1]; Mr. IWASZKIEWICZ, Jan [1]

[1] *CERN*

Corresponding Author: fons.rademakers@cern.ch

In this talk we describe the latest developments in the PROOF system. PROOF is the parallel extension of ROOT and allows large datasets to be processed in parallel on large clusters and/or multi-core machines. The recent developments have focused on readying PROOF for the imminent data analysis tasks of the LHC experiments. Main improvements have been made in the areas of overall robustness and fault tolerance, multi-user scheduling, generic processing, e.g. Monte Carlo's, and optimizations for many-core architectures. ALICE is deploying PROOF for prompt reconstruction and analysis on the CERN Analysis Facility (CAF) and ATLAS if focusing on Tier-3 deployment.

63

Optimized Calorimeter Signal Compaction for an Independent Component based ATLAS Electron/Jet Second-level Trigger

Mr. SIMAS, Eduardo [1]; Dr. SEIXAS, Jose [1]; Dr. LUIZ, Caloba [1]

[1] *Federal University of Rio de Janeiro*

Corresponding Author: esimas@lps.ufrj.br

The ATLAS online trigger system has three filtering levels and accesses information from calorimeters, muon chambers and the tracking system. The electron/jet channel is very important for triggering system performance as Higgs signatures may be found efficiently through decays that produce electrons as final-state particles.
Electron/jet separation relies very much on calorimeter information. ATLAS calorimeter system is segmented into seven layers and due to differences both in depth and cell granularity of these layers, trigger algorithms may benefit from performing feature extraction at the layer level.
This work addresses the second level (LVL2) filtering restricted to calorimeter data. The LVL2 operation is split into two phases: feature extraction, in which relevant information is extracted from the measured signals, and hypothesis testing, where particle discrimination is performed over relevant variables.
Here, Region of Interest (RoI) data are formatted into concentric ring sums. At each calorimeter layer, the most energetic cell is defined as the first ring, and the following rings are formed around this point, so that all cells belonging to a ring have their sampled energies added together. For each event to be processed, a RoI with a size of 0.4 x 0.4 in the eta x phi plane is defined and gets described by a total of 100 ring sums. After this pre-processing step, segmented independent component analysis (SICA) is applied in order to extract from ring sums relevant features for particle identification. Independent Component Analysis (ICA) aims at extracting the statistically independent sources that produced the observed calorimeter signals and here ICA is applied in a segmented way. The number of independent components to be extracted from a ROI is estimated through different signal compaction strategies, such as Principal Component Analysis, Nonlinear Principal Component Analysis and Principal Components for Discrimination. These compaction techniques are evaluated with respect to dimensionality reduction (and processing speed) and classification efficiency. The hypothesis testing is performed by a Multi-Layer Perceptron classifier fed from the segmented independent components. The classifier is designed using simulated level one trigger accepted signals.

Data Analysis - Algorithms and Tools / 64

The ATLAS Conditions Database Model for the Muon Spectrometer

Dr. VERDUCCI, Monica [1]

[1] *INFN RomaI*

Corresponding Author: monica.verducci@cern.ch

The ATLAS Muon System has extensively started to use the LCG conditions database project 'COOL' as the basis for all its conditions data storage both at CERN and throughout the worlwide collaboration as decided by the ATLAS Collaboration. The management of the Muon COOL conditions database will be one of the most challenging applications for Muon System, both in terms of data volumes and rates, but also in terms of the variety of data stored. The Muon Conditions database is responsible for almost of all the 'non-event' data and detector quality flags storage needed for debugging of the detector operations and for performing reconstruction and analysis. The COOL database allows database applications to be written independently of the underlying database technology and ensures long-term compatibility with entire ATLAS Software. COOL implements an interval of validity database, i.e. objects stored or referenced in COOL have an associated start and end time between which they are valid, the data is stored in folders, which are themselves arranged in a hierarchical structure of foldersets. The structure is simple and mainly optimsed to store and retrieve object(s) associated to a particular time. In this work, an overview of the entire Muon Conditions Database architecture is given, including the different sources of the data and the storage model used. In addiction the software interfaces used to access to the Conditions Data are described, more emphasis is given to the Offline Reconstruction framework ATHENA and the services developed to provide the Conditions data to the reconstruction.

Computing Technology for Physics Research / 66

MDT data quality assessment at the Calibration centre for the ATLAS experiment at LHC

Dr. VERDUCCI, Monica [1]; Dr. SOLFAROLI, Elena [2]

[1] *INFN RomaI*

[2] *INFN RomaI Universita' di Roma La Sapienza*

Corresponding Author: monica.verducci@cern.ch

ATLAS is a large multipurpose detector, presently in the final phase of construction at LHC, the CERN Large Hadron Collider accelerator. In ATLAS the muon detection is performed by a huge magnetic spectrometer, built with the Monitored Drift Tube (MDT) technology. It consists of more than 1,000 chambers and 350,000 drift tubes, which have to be controlled to a spatial accuracy better than 10 micrometers and an efficiency close to 100%.

Therefore, the detector automated monitor is an essential aspect of the operation of the spectrometer. The quality procedure collects data from online and offline sources and from the "calibration stream" at the calibration centres, situated in Ann Arbor (Michigan), MPI (Munich) and INFN Rome. The assessment at the Calibration Centres is performed using the DQHistogramAnalyzer utility of the Athena package. This application checks the histograms in an automated way and, after a further inspection with a human interface, reports results and summaries.
In this study a complete description of the entire chain, from the calibration stream up to the database storage is presented. Special algorithms have been implemented in the DQHistogramAnalyzer for the Monitored Drift Tube chambers. A detailed web display is provided for easy data quality consultation. The analysis flag is stored inside an Oracle Database using the COOL LCG library, through a C++ object-oriented interface. This quality flag is compared with the online and offline results, produced in a similar way, and the final decision is stored in a DB using a standalone C++ tool. The final DB, which uses the same COOL technology, is accessed by the reconstruction and analysis programs.

Computing Technology for Physics Research - Session 2 / 67

Early Experience with the CMS Computing Model

Dr. FISK, Ian [1]

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

Corresponding Author: ian.fisk@cern.ch

In this presentation we will discuss the early experience with the CMS computing model from the last large scale challenge activities to the first days of data taking. The current version of the CMS computing model was developed in 2004 with a focus on steady state running. In 2008 a revision of the model was made to concentrate on the unique challenges associated with the commission period. The types of changes needed for commissioning will be presented. In addition we will discuss the challenges in commissioning the major processing workflows for analysis and production and the workflows for data management and data consistency across the distributed computing infrastructure. We will present experiences and results from the final scalability tests of services and techniques. We will also discuss the initial experience with active users and real data from cosmic and collision running. We will address the issues that worked well in addition to identifying areas where future development and refinement is needed.

Friday, 07 November 2008 / 160

Computing Technology for Physics Research - Summary

Friday, 07 November 2008 / 161

Data Analysis - Algorithms and Tools - Summary

Dr. SPEER, Thomas [1]

[1] *Brown University*

Friday, 07 November 2008 / 162

Methodology of Computations in Theoretical Physics - Summary

Corresponding Author: kato@cc.kogakuin.ac.jp

Friday, 07 November 2008 / 163

ACAT2008 Summary

Monday, 03 November 2008 - Morning session 2 / 166

High-Precision Arithmetic and Mathematical Physics

BAILEY, David [1]

[1] *Lawrence Berkeley Laboratory*

Corresponding Author: dhbailey@lbl.gov

For the vast majority of computations done both in pure and applied physics, ordinary 64-bit floating-point arithmetic (about 16 decimal digits) is sufficient. But for a growing body of applications, this level is not sufficient. For applications such as supernova simulations, climate modeling, n-body atomic structure calculations, "double-double" (approx. 32 digits) or even "quad-double" (approx. 64 digits) is required. For yet other applications, notably arising in quantum field theory and statistical mechanics, much higher precision (hundreds or even thousands of digits) is required. Armed with software for performing computation to these high levels of precision, the tools of "experimental mathematics" can be brought to bear, such as in recognizing the values of definite integrals that arise in the theory via their decimal values. Numerous recent studies, including some rather remarkable results in quantum field theory, will be mentioned.

Computing Technology for Physics Research - Session 2 / 167

Using constraint programing to resolve the multi-source / multi-site data movement paradigm on the Grid

Mr. ZEROLA, Michal [1]; Dr. LAURET, Jerome [2]; Dr. SUMBERA, Michal [1]; Dr. BARTAK, Roman [3]

[1] *Nuclear Physics Inst., Academy of Sciences, Praha*

[2] *BROOKHAVEN NATIONAL LABORATORY*

[3] *Faculty of Mathematics and Physics, Charles University, Praha*

Corresponding Author: jlauret@bnl.gov

In order to achieve both fast and coordinated data transfer to collaborative sites as well as to create a distribution of data over multiple sites, efficient data movement is one of the most essential aspects in distributed environment. With such capabilities at hand, truly distributed task scheduling with minimal latencies would be reachable by internationally distributed collaborations (such as ones in HENP) seeking for scavenging or maximizing on geographically spread computational resources. But it is often not all clear (a) how to move data when available from multiple sources or (b) how to move data to multiple compute resources to achieve an optimal usage of available resources.

Constraint programming (CP) is a technique from artificial intelligence and operations research allowing to find solutions in a multi-dimensional space of variables. We present a method of creating a CP model consisting of sites, links and their attributes such as bandwidth for grid network data transfer also considering user tasks as part of the objective function for an optimal solution. We will explore and explain trade-off between schedule generation time and divergence from the optimal solution and show how to improve and render viable the solution's finding time by using search tree time limit, approximations, restrictions such as symmetry breaking or grouping similar tasks together, or generating sequence of optimal schedules by splitting the input problem.

Results of data transfer simulation for each case will also include a well known Peer-2-Peer model, and time taken to generate a schedule as well as time needed for a schedule execution will be compared to a CP optimal solution. We will additionally present a possible implementation aimed to bring a distributed datasets (multiple sources) to a given site in a minimal time.

Computing Technology for Physics Research - Session 2 / 168

Evolution of the STAR Framework OO model for the Multi-Core era

Dr. FINE, Valeri [1]; Dr. LAURET, Jerome [1]; Mr. PEREVOZTCHIKOV, Victor [1]

[1] *BROOKHAVEN NATIONAL LABORATORY*

Corresponding Author: jlauret@bnl.gov

With the era of multi-core CPUs, software parallelism is becoming both affordable as well as a practical need. Especially interesting is to re-evaluate the adaptability of the high energy and nuclear physics sophisticated, but time-consuming, event reconstruction frameworks to the reality of the multi-threaded environment.

The STAR offline OO ROOT-based framework implements a well known "standard model" composed of chained modules, where input for each module is the output of the other modules. At its basic principle, modules do not communicate with each other directly and act as consumers and providers of data structures. They use the framework via the special “query” / “publish” API to query the presence of the input data and publish the output results the modules produce.

We will show that by complementing the base framework with the ability to start several modules in parallel and synchronize the global data access between the "consumer" modules and "producer", one can transparently enhance the existent packages to leverage the multi-core hardware capability. Such approach allows re-using the existing offline software designed for the single thread batch applications in the multi-threaded environment if needed. However, we realize that the "query"/"publish" paradigm is not sufficient to run the multi-threaded application effectively. It should be complemented with an API to “register” the module output to notify the framework members about an output dataset “to be produced soon”. With such addition, the receiving module thread can be automatically suspended if the data it has requested is not ready yet.

We will explain how the STAR Offline Framework was modified to test the present approach by building the sophisticated interactive real time applications.

Data Analysis - Algorithms and Tools / 169

Recent Improvements of the ROOT Fitting and Minimization Classes

Dr. MONETA, Lorenzo [1]; Mr. GONZALEZ MALINE, David [1]

[1] *CERN*

Corresponding Author: lorenzo.moneta@cern.ch

Advanced mathematical and statistical computational methods are required by the LHC experiments for analyzing their data. Some of these methods are provided by the Math work package of the ROOT project, a C++ Object Oriented framework for large scale data handling applications.
We present in detail the recent developments of this work package, in particular the recent improvements in the fitting and minimization classes, which have been re-designed and re-implemented with an object-oriented approach. New minimization algorithms have been added recently in ROOT and they can be used consistently for fitting via a common interface.
These algorithms include Minuit2, the new OO version of Minuit,
various minimization methods from the GNU Scientific libraries, stochastic and genetic algorithms.
Furthermore, a new graphics user interface have been also developed for performing and monitoring fits on ROOT data objects such as histograms, graphs and threes in both one or multi-dimensions.
We will describe in detail the new capabilities provided by the new fitting and minimization classes and the functionality of the new user interface.

Methodology of Computations in Theoretical Physics - Session 2 / 170

New results for loop integrals

RIEMANN, Tord [1]; YUNDIN, Valery [1]

[1] *DESY*

Corresponding Author: tord.riemann@desy.de

We present some recent results on the evaluation of massive one-loop multileg Feynman integrals, which are of relevance for LHC processes.
An efficient complete analytical tensor reduction was derived and implemented in a Mathematica package hexagon.m.
Alternatively, one may use Mellin-Barnes techniques in order to avoid the tensor reduction.
We shortly report on a new version of the Matheamtica package AMBRE.m.

Computing Technology for Physics Research - Session 2 / 171

XCFS - an analysis disk pool & filesystem based on FUSE and xroot protocol

Mr. PETERS, Andreas Joachim [1]

[1] *CERN*

Corresponding Author: andreas.joachim.peters@cern.ch

One of the biggest challenges in LHC experiments at CERN is data management for data analysis. Event tags and iterative looping over datasets for physics analysis require many file opens per second and (mainly forward) seeking access. Analyses will typically access large datasets reading terabytes in a single iteration.
A large user community requires policies for space management and a highly performant, scalable, fault-tolerant and higly available system to store user data. While batch job access for analysis can be done using remote protocols experiment users expressed a need for a direct filesystem integration of their analysis (output) data to support file handling via standard unix tools, browsers,scripts etc.

XCFS - the xroot based catalog file system is an attempt to implement the above ideas based on xroot protocol. The implementation is done via a filesystem plugin for FUSE using the xroot posix client library which has been tested on LINUX and MAC OSX platform.

Filesystem meta data is stored on the head node in a XFS filesystem using sparse files and extended attributes. XCFS provides synchronous replica creation during write operations, a distributed unix quota system, krb5/gsi and voms authentication with support for secondary groups (via xroot remote protocol and through the mounted filesystem).

High availability of the headnode is achieved using a heartbeat setup and filesystem mirroring using DRBD.
The first 80 TB test setup allowing to store a maximum number of 800 million files has shown promising results with thousands of file open and meta data operations per second and saturation of gigabit ethernet executing single 'cp' commands on the mounted file system. The average latency for meta data commands is in the order of ~1ms, for file open operations it is <4ms.
The talk will discuss results of typical LHC analysis applications using remote or mounted filesystem access. A comparison will be made between XCFS and other filesystem implementations like AFS or Lustre. Strength and weaknesses of the approach and its possible usage in CASTOR - the CERN mass storage system - will be discussed.

172

FROG : The Fast & Realistic OpenGl Event Displayer

Mr. QUERTENMONT, Loic [1]

[1] *Universite Catholique de Louvain*

Corresponding Author: loic.quertenmont@cern.ch

FROG is a generic framework dedicated to visualize events in a given geometry. \newline
It has been written in C++ and use OpenGL cross-platform libraries. It can be used to any particular physics experiment or detector design. The code is very light and very fast and can run on various Operating System. Moreover, FROG is self consistent and does not require installation of ROOT or Experiment software (e.g. CMSSW) libraries on user's computer.\newline
The slides will describe the principle of the algorithm and its many functionalities such as : 3D and 2D visualization, graphical user interface, mouse interface, configuration files, production of pictures in various format, integration of personal objects... Finally the application of FROG for physic experiment, such as CMS experiment, will be described.

http://projects.hepforge.org/frog/
https://twiki.cern.ch/twiki/bin/view/CMS/FROG

173

Combining Hough Transform and Optimal Filtering for Efficient Cosmic Ray Detection with a Hadronic Calorimeter

Mr. MANHAES DE ANDRADE FILHO, Luciano [1]; Dr. MANOEL DE SEIXAS, Jose [1]

[1] *Universidade Federal do Rio de Janeiro*

Corresponding Author: luciano.manhaes.de.andrade.filho@cern.ch

The hadronic calorimeter of ATLAS, TileCal, provides a large amount of readout channels (about 10,000). Therefore, track detection may be performed by TileCal when cosmic muons cross the detector. The muon track detection has extensively been used in the TileCal commissioning phase, for both energy and timing calibrations, and it will also be important for background noise removal during nominal LHC operation.
This work presents a cosmic ray detection algorithm based on TileCal information. The algorithm employs the Hough Transform to map the data from activated calorimeter cells into a parameter (straight-line track) space, in which detection is effectively performed. Due to intrinsic low signal-to-noise ratio for cosmic ray detection with TileCal, a preprocessing algorithm based on a matched filter operating over TileCal time sampled signals is implemented to improve detection efficiency. Experimental data shows that the proposed method possesses extensive efficiency superiority over straight-line fitting produced by least-square methods.

Computing Technology for Physics Research / 174

ALICE Analysis Framework

Mr. GHEATA, Andrei [1]; FOR, ALICE Offline collaboration [2]

[1] *ISS/CERN*

[2] *CERN*

Corresponding Author: andrei.gheata@cern.ch

The talk will describe the current status of the offline analysis framework used in ALICE. The software was designed and optimized to take advantage of distributed computing resources and be compatible with ALICE computing model. The framwork's main features: possibility to use parallelism in PROOF or GRID environments, transparency of the computing infrastructure and data model, scalability and data access performance. The framework provides a common "language" for all ALICE analysis users and is being heavily tested in view of the data to arrive soon.

Computing Technology for Physics Research / 175

The commissioning of CMS computing centres in the WLCG Grid

Dr. FLIX MOLINA, José [1]; Dr. SCIABA', Andrea [2]; Dr. WHÜRTHWEIN, Frank [3]; Dr. LETTS, James [3]; Dr. KLEM, Jukka [4]; Dr. SAIZ, Pablo [2]; Dr. HERNANDEZ, José [1]; Dr. BELFORTE, Stefano [5]; Dr. FANFANI, Alessandra [6]; Dr. FISK, Ian [7]; Dr. MAGINI, Nicolo [2]; Dr. MICCIO, Vincenzo [2]

[1] *Cent. Invest. Energ. Medioamb. Tec. (CIEMAT) - Consejo Sup. de I*

[2] *CERN, Geneva, Switzerland*

[3] *University of California at San Diego, San Diego, United States*

[4] *Helsinki Institute of Physics, Helsinki, Findland*

[5] *INFN, Sezione di Trieste, Italy*

[6] *Universita degli Studi di Bologna*

[7] *Fermi National Accelerator Laboratory, Batavia, United States*

Corresponding Author: jose.flix.molina@cern.ch

The computing system of the CMS experiment works using distributed resources from more than 80 computing centres worldwide. These centres, located in Europe, America and Asia are interconnected by the Worldwide LHC Computing Grid. The operation of the system requires a stable and reliable behaviour of the underlying infrastructure.

CMS has established a procedure to extensively test all relevant aspects of a Grid site, such as the ability to efficiently use their network to transfer data, the functionality of all the site services relevant for CMS and the capability to sustain the various CMS computing workflows (Monte Carlo simulation, event reprocessing and skimming, data analysis) at the required scale. This contribution describes in detail the procedure to rate CMS sites depending on their performance, including the complete automation of the program, the description of monitoring tools, and its impact in improving the overall reliability of the Grid from the point of view of the CMS computing system.

Data Analysis - Algorithms and Tools / 176

Efficient Level 2 Trigger System Based on Artificial Neural Networks

Ms. KHATCHADOURIAN, Sonia [1]; Mr. PRÉVOTET, Jean-Christophe [2]; Mr. KESSAL, Lounis [1]

[1] *ETIS - UMR CNRS 8051*

[2] *IETR - UMR CNRS 6164*

Corresponding Author: sonia.khatchadourian@ensea.fr

The HESS project is a major international experiment currently performed in gamma astronomy. This project relies on a system of four Cherenkov telescopes enabling the observation of cosmic gamma rays. The outstanding performance obtained so far in the HESS experiment has led the research labs involved in this project to improve the existing system: an additional telescope is currently being built and will soon take place within the previous telescope system. This telescope is designed to be more sensitive to the detection of low energy particles than the others, leading to an increase of the number of collected particle images. In this context which is tightly constrained in terms of latency, physicists have been compelled to design an additional L2 Trigger in order to deal with a huge amount of data. This trigger aims at selecting images of interest (ie. gamma particles) and rejecting all other events that are associated to noise. Contrary to classical methods that consist of strong cuts based on Hillas parameters, we propose an original approach based on artificial neural networks.

In this approach, collected events are first handled by a pre-processing level whose purpose consists in applying transformations on incoming images, thus reducing the dimensionality of the problem. It is based on Zernike moments computation that aims to extract the main features of the images and guarantee image invariance in translation and rotation. Zernike moments have also proved to be reliable in terms of their feature representation capability and low noise sensitivity.

In a second step, an artificial neural networks ensures the classification of events within two classes (gammas and hadrons), indicating whether to keep the image for future processing or to reject it.

In this presentation, we will describe the entire L2-Trigger system and provide some results in terms of classification performances. We will discuss the contribution of neural networks in this type of experiments compare to classical solutions.

Data Analysis - Algorithms and Tools / 177

Sophisticated algorithms of analysis of spectroscopic data

MORHAC, Miroslav [1]

[1] *Institute of Physics, Slovak Academy of Sciences*

Corresponding Author: miroslav.morhac@savba.sk

The accuracy and reliability of the analysis of spectroscopic data depend critically on the treatment in order to resolve strong peak overlaps, to account for continuum background contributions, and to distinguish artifacts to the responses of some detector types. Analysis of spectroscopic data can be divided to
1. estimation of peaks positions (peak searching)
2. fitting of peak regions.
One of the most delicate problems of any spectrometric method is that related to the extraction of the correct information out of the spectra sections, where due to the limited resolution of the equipment, the peaks as the main carrier of spectrometric information are overlapping. Conventional methods of peak searching based usually on spectrum convolution are inefficient and fail to separate overlapping peaks. The deconvolution methods can be successfully applied for the determination of positions and intensities of peaks and for the decomposition of multiplets. Several deconvolution algorithms are studied and and their efficiences compared in the contribution.
However before the application of deconvolution operation we need to remove the background from spectroscopic data. One of the basic problems in the analysis of the spectra is the separation of useful information contained in peaks from the useless information (background, noise). In order to process data from numerous analyses efficiently and reproducibly, the background approximation must be, as much as possible, free of user-adjustable parameters. Baseline removal, as the first preprocessing step of spectrometric data, critically influences subsequent analysis steps. The more accurately the background is estimated the more presicely we can estimate the existence of peaks. In the contribution we present a new algorithm to determine peak regions and separate them from peak-free regions. Subsequently it allows to propose a new baseline estimation method based on sensitive non-linear iterative peak clipping with automatic local adjusting of width of clipping window. Moreover automatic setting of peak regions can be used to confine intervals of fitting and to fit each region separately.

178

Visualization of multidimensional histograms using hypervolume techniques

MORHAC, Miroslav [1]; Dr. MATOUSEK, Vladislav [1]; Dr. TURZO, Ivan [1]

[1] *Institute of Physics, Slovak Academy of Sciences*

Corresponding Author: miroslav.morhac@savba.sk

Visualization is one of the most powerful and direct ways how the huge amount of information contained in multidimensional histograms can be conveyed in a form comprehensible to a human eye. With increasing dimensionality of histograms (nuclear spectra) the requirements in developing of multidimensional scalar visualization techniques become striking. In the contribution we present a hypervolume visualisation techniques that provide simple and fully explanatory images that give comprehensive insights into the global structure of scalar fields of any dimension. The presented method provides a dimension independent viewing system which scales nicely with the geometric dimension of the dataset. On the other hand the algorithm allows one to localize and scan interesting parts (peaks) in multidimensional histograms. It also permits to find correlations in the data, mainly among neighboring points in all dimensions, and thus to discover prevailing trends around multidimensional peaks using classical approaches like slicing and animation of slices of multidimensional data.

179

RDIG Alice computing at the time of the first LHC data

EYGENE, Ryabinkin [1]; GALINA, Shabratova [2]; VLADIMIR, Trofimov [2]; WASIL, Urazmetov [3]; ALEXEY, Bogdanov [4]; ANDREY, Zarochentsev [5]; SERGEY, Zotkin [6]; VICTOR, Kotlyar [3]; ANDREY, Kiryanov [7]; NIKOLAY, Kutouski [2]; VALERY, Mitsyn [2]; YEVGENY, Lyublev [8]; LIUDMILA, Stepanova [9]

[1] *Russian Research Centre "Kurchatov Institute"*

[2] *Joint Institute for Nuclear Research*

[3] *State Research Center of Russian Federation, Institute for High Energy Physics*

[4] *Moscow Engineering Physics Institute*

[5] *Saint-Petersburg State University*

[6] *Skobelitsyn Institute for Nuclear Physics, Moscow State University*

[7] *St'Petersburg Nuclear Physics Institute, Russian Academy of Sciences*

[8] *State Scientific Center of Russian Federation Institute for Theoretical and Experimental Physics*

[9] *Institute for Nuclear Research, Russian Academy of Sciences*

Corresponding Author: rea@mbslab.kiae.ru

The major subject of this talk is the presentation of the distributed computing status report for the ALICE experiment at Russian sites just before and at the time of the data taking at the Large Hadron Collider in CERN. We present the usage of the ALICE application software, AliEn[1], at the top of the modern EGEE middleware called gLite for the simulation and data analysis in the experiment at the Russian Tier2 in accordance with the ALICE computing model [2]. We outline the results of CPU and disk space usage at RDIG sites for the data simulation and analysis of first LHC data from the exposition of ALICE detector.

[1] P. Saiz et al., Nucl. Instrum. Methods, A502, 437-440 (2003); http://alien.cern.ch/.
[2] ALICE Collaboration, Technical Design Report of Computing, CERN-LHCC-2005-018.

This activity is supported by the INTAS+CERN grant 05-103-7484

Methodology of Computations in Theoretical Physics - Session 1 / 180

From moments to functions in higher order QCD

BLUEMLEIN, Johannes [1]

[1] *DESY*

Corresponding Author: johannes.bluemlein@desy.de

We present a method to unfold the complete functional dependence of single-scale quantities as QCD splitting functions and Wilson coefficients from a finite number of moments. These quantities obey recursion relations which can be found in an automated way. The exact functional form is obtained solving the corresponding difference equations. We apply the algorithm to the QCD Wilson coefficients for deep-inelastic scattering and splitting functions to 3-loop order which are associated with difference equations of rather high order and degree.

Wednesday, 05 November 2008 / 181

LHC phenomenology at next-to-leading order QCD: theoretical progress and new results

Dr. BINOTH, Thomas [1]

[1] *University of Edinburgh*

Corresponding Author: thomas.binoth@ed.ac.uk

In this talk I will motivate that a succesful descripton of LHC physics needs the inclusion of higher order corrections for all kinds of signal and background processes. In the case of multi-particle production the combinatorial complexity of standard approaches triggered many new developments which allow for the efficient evaluation of one-loop amplitudes for LHC phenomenology. I will discuss the basic new ideas for one-loop multi-leg computations including comments on computational issues and will review recent results relevant for LHC phenomenology.

Thursday, 06 November 2008 / 184

Many-core Round Table - How to prepare for the future

Monday, 03 November 2008 - Morning session 1 / 185

Aspects of Intellectual Property Law for HEP Software Developers

PINSKY, Lawrence [1]

[1] *University of Houston-Unknown-Unknown*

Corresponding Author: lawrence.pinsky@cern.ch

Intellectual Property, which includes the following areas of the law: Copyrights, Patents, Trademarks, Trade Secrets, and most recently Database Protection and Internet Law, might seem to be an issue for lawyers only. However, increasingly the impact of the laws governing these areas and the International reach of the effects of their implementation makes it important for all software developers and especially software development team leaders to be aware of the implications that might exist. For example, under copyright law, if their are multiple contributors to a software development project, the contribution of even one single line of code to million line program, will vest an equal copyright ownership share in the person making that one-line contribution. While patenting of software is being debated in Europe, it is allowed in the US, and with the Grid and the Internet, the effects of these patents can reach far beyond the US border. Likewise, the recent enactment of database protection in Europe, which grants property rights to the facts contained in databases can still have impacts on users in the US where such legislation has not been put in force. A brief primer on the basics of Intellectual Property Law relevant for software developers will be provided.

Methodology of Computations in Theoretical Physics - Session 1 / 187

Tools for systematic event generator tuning and validation

Dr. BUCKLEY, Andy [1]

[1] *Durham University*

Corresponding Author: andy.buckley@durham.ac.uk

Event generator programs are a ubiquitous feature of modern particle physics, since the ability to produce exclusive, unweighted simulations of high-energy events is necessary for design of detectors, analysis methods and understanding of SM backgrounds. However --- particularly in the non-perturbative areas of physics simulated by shower+hadronisation event generators --- there are many parameters which must be tuned to experimental data for useful predictions to be obtained. Attempting to globally tune these parameters to a wide range of experimental results is a task much better suited to systematic, computer-based optimisation than the traditional "tweaking by eye" approach. I will present the current status of the Rivet+Professor tuning/validation system, with emphasis on recent tunes of Pythia 6 to event shape, hadron multiplicity and underlying event data from LEP to the Tevatron.

Data Analysis - Algorithms and Tools / 188

A Numeric Comparison of Feature Selection Algorithms for Supervised Learning

Dr. PALOMBO, Giulio [1]; Dr. NARSKY, Ilya [2]

[1] *University of Milan - Bicocca*

[2] *California Institute of Technology*

Corresponding Author: palombo@cacr.caltech.edu

Datasets in modern High Energy Physics (HEP) experiments are often described by dozens or even hundreds of input variables (features). Reducing a full feature set to a subset that most completely represents information about data is therefore an important task in analysis of HEP data. We compare various feature selection algorithms for supervised learning using several datasets such as, for instance, imaging gamma-ray Cherenkov telescope (MAGIC) data found at the UCI repository.

We use classifiers and feature selection methods implemented in the statistical package StatPatternRecognition (SPR), a free open-source C++ package developed in the HEP community (http://sourceforge.net/projects/statpatrec/). For each dataset, we select a powerful classifier and estimate its learning accuracy on feature subsets obtained by various feature selection algorithms. When possible, we also estimate the CPU time needed for the feature subset selection. The results of this analysis are compared with those published previously for these datasets using other statistical packages such as R and Weka. We show that the most accurate, yet slowest, method is a wrapper algorithm known as generalized sequential forward selection ("Add N Remove R") implemented in SPR.

189

The ATLAS Muon Trigger vertical slice at LHC startup

GRANCAGNOLO, Sergio [1]

[1] *INFN University Lecce*

Corresponding Author: sergio.grancagnolo@le.infn.it

The ATLAS trigger system has a three-levels structure, implemented to retain interesting physics events, here described for the muon case ("Muon Vertical Slice"). The first level, implemented in a custom hardware, uses measurements from the trigger chambers of the Muon Spectrometer to select muons with high transverse momentum and defines a Region of Interest (RoI) in the detector. RoIs are then processed by a second trigger level, in which fast algorithms run on an online software architecture. Full granularity information from precision chambers is accessed inside RoIs. A third trigger level (Event Filter), using offline-like algorithms and accessing the full event, provide the best possible muon reconstruction/identification and finally confirm or discard the trigger hypothesis formed at earlier levels. Implementation and performance of the full muon trigger slice, together with first events triggered with LHC beams on, are presented.

Computing Technology for Physics Research - Session 2 / 190

Job Centric Monitoring for ATLAS jobs in the LHC Computing Grid

Prof. MAETTIG, Peter [1]; Prof. WULFF, Nikolaus [2]; Mr. MECHTEL, Markus [1]; Mr. RAU, Martin [3]; Prof. EHSES, Erich [3]; Mr. UEBERHOLZ, Peer [4]; Mr. MUENCHEN, Tim [5]; Dr. HARENBERG, Torsten [1]

[1] *University of Wuppertal*

[2] *University of applied sciences, Muenster*

[3] *University of applied sciences, Koeln*

[4] *Hochschule Niederrhein*

[5] *Bergische Universitaet Wuppertal*

Corresponding Author: muenchen@physik.uni-wuppertal.de

As the Large Hadron Collider (LHC) at CERN, Geneva, has begun operation in september, the large scale computing grid LCG (LHC Computing Grid) is meant to process and store the large amount of data created in simulating, measuring and analyzing of particle physic experimental data. Data acquired by ATLAS, one of the four big experiments at the LHC, are analyzed using compute jobs running on the grid and utilizing the ATLAS software framework 'Athena'. The analysis algorithms themselves are written in C++ by the physicists using Athena and the ROOT toolkit.

Identifying the reason for a job failure (or even the occurance of the failure itself) in this context is a tedious, repetitive and - more often than not - unsuccessful task. Often, to deal with failures in the RUNNING stage (as opposed to job submission failures or compilation errors in the user algorithms), the job is just being resubmitted. The debugging of such problems is made even more difficult by the fact that the output-sandbox, which contains the jobs' output and error logs, is discarded by the grid middleware if the job failed. So, valuable information that could aid in finding the failure reason is lost. These issues result in high job failure rates and less than optimal resource usage.

As part of the High Energy Particle Physics Community Grid project (HEPCG) of the German D-Grid Initiave, the University of Wuppertal has developed the Job Execution Monitor (JEM). JEM helps finding job failure reasons by two means: It periodically provides vital worker node system data and collects job run-time monitoring data. To gather this data, a supervised line-by-line execution of the user job is performed. JEM is providing new possibilities to find problems in largely distributed computing grids and to analyze these problems in nearly real-time.

All monitored information is presented to the user almost instantaneously and additionally stored in the jobs' output sandbox for further analysis. As a first step, JEM has been seamlessly integrated into ATLAS' and LHCb's grid user interface 'ganga'. In this way, submitted jobs are monitored transparently, requiring no additional effort by the user.

In this work, the functionality of and the concepts behind JEM are presented together with examples of typical problems that are easily discovered. Furthermore, we present an ongoing work of classifying problems automatically using expert systems.

Data Analysis - Algorithms and Tools / 191

Tau identification using multivariate techniques in ATLAS

Dr. WOLTER, Marcin [1]

[1] *Henryk Niewodniczanski Institute of Nuclear Physics PAN*

Corresponding Author: marcin.wolter@cern.ch

Tau leptons will play an important role in the physics program at the LHC. They will not only be used in electroweak measurements and in detector related studies like the determination of the E_T^miss scale, but also in searches for new phenomena like the Higgs boson or Supersymmetry.

Due to the overwhelming background from QCD processes, highly efficient algorithms are essential to identify hadronically decaying tau leptons. This can be achieved using modern multivariate techniques which make optimal use of all the information available. They are particularly useful in case the discriminating variables are not independent and no single variable provides good signal and background separation.

In ATLAS four algorithms based on multivariate techniques have been applied to identify hadronically decaying tau leptons: projective likelihood estimator (LL), Probability Density Estimator with Range Searches (PDE-RS), Neural Network (NN) and Boosted Decision Trees (BDT). All four multivariate methods applied to the ATLAS simulated data have similar performance, which is significantly better than the baseline cut analysis.

Thursday, 06 November 2008 / 192

CernVM - a virtual appliance for LHC applications

MATO, Pere [1]; BLOOMER, Jakob [2]; BUNCIC, Predrag [1]; AGUADO SANCHEZ, Carlos [1]; FRANCO, Leandro [1]; KLEMER, Steffen [3]

[1] *CERN*

[2] *Univ. of Karlsruhe*

[3] *Georg-August-University Goettingen*

Corresponding Author: predrag.buncic@cern.ch

CernVM is a Virtual Software Appliance to run physics applications from the LHC experiments at CERN. The virtual appliance provides a complete, portable and easy to install and configure user environment for developing and running LHC data analysis on any end-user computer (laptop, desktop) and on the Grid independently of operating system software and hardware platform (Linux, Windows, MacOS). The aim is to facilitate the installation of the experiment software on an user computer and minimize the number of platforms (compiler-OS combinations) on which experiment software needs to be supported and tested thus reducing the overall cost of LHC software maintenance. Two ingredients are necessary for CernVM. The first one is a thin virtual machine that contains 'just enough Operating System' to run any application framework of the four LHC experiments. The second is a file system (cvmfs) specifically designed for an efficient 'just in time' software distribution and installation. The CernVM project, which has started at the beginning of this year is funded for period of four years under the recently approved R program at CERN.

193

CVMFS - a file system for the CernVM virtual appliance

MATO, Pere [1]; BLOOMER, Jakob [2]; BUNCIC, Predrag [1]; AGUADO SANCHEZ, Carlos [1]; FRANCO, Leandro [1]; KLEMER, Steffen [3]

[1] *CERN*

[2] *Univ. of Karlsruhe*

[3] *Georg-August-University Goettingen*

Corresponding Author: predrag.buncic@cern.ch

The CernVM Virtual Software Appliance contains a minimal operating system sufficient to host the application frameworks developed by the LHC experiments. In CernVM model the experiment application software and it dependencies are built independently from CernVM Virtual Machine. The procedures for building, installing and validating each software release remains in the hands and under responsibility of each user community. We provide the tools to synchronize pre-built and configured experiment software releases with our central distribution point that we plan to evolve into decentralized content delivery network.

We will describe the architecture of the CVMFS file system and the implementation of the highly available service infrastructure, which is required in order to provide the reliable central component of this file system. In addition, we will outline how we plan to evolve this into a decentralized content delivery network.

Methodology of Computations in Theoretical Physics - Session 1 / 194

Numerical Evaluation of Feynman Integrals by a Direct Computation Method

Dr. YUASA, Fukuko [1]; Dr. ISHIKAWA, Tadashi [1]; Dr. FUJIMOTO, Junpei [1]; Mr. HAMAGUCHI, Nobuyuki [2]; Prof. DE DONCKER, Elise [3]; Prof. SHIMIZU, Yoshimitsu [4]

[1] *KEK*

[2] *Hitachi, Ltd.*

[3] *Western Michigan University*

[4] *The Graduate University for Advanced Studies, Sokendai*

Corresponding Author: fukuko.yuasa@kek.jp

We apply a 'Direct Computation Method', which is purely numerical, to evaluate Feynman integrals. This method is based on the combination of an efficient numerical integration and an efficient extrapolation strategy. In addition, high-precision arithmetic and parallelization techniques can be used if required. We present our recent progress in the development of this method and show test results such as for one-loop 5-point and two-loop 3-point integrals.

Thursday, 06 November 2008 / 195

Concurrent Collections: Parallel Programming Without Worrying About Parallelism

KNOBE, Kathleen [1]

[1] *Intel*

Corresponding Author: kath.knobe@intel.com

Concurrent Collections is a different way of writing parallel applications. Its major contribution is to isolate the task of specifying the application semantics from any consideration of its parallel execution. This isolation makes it much easier for the domain-expert, the physicist for example, to specify the application. It also makes the task of the tuning-expert, mapping the application to a parallel platform, more effective. These experts may be the same person at different times, distinct people, or the tuning-expert may be automated.

The application is written in terms of high-level operators, the chunks of computation in the application. The specification indicates neither an explicitly serial ordering nor an explicitly parallel one. It simply specifies constraints on the order among these high-level operators. These constraints are based on the flow of data and control in the application.

Concurrent Collections supports (and we have built) a wide range of different mapping approaches. The grain, the distribution among processors and the scheduling within a processor are done statically or dynamically and by a person or automatically.

Methodology of Computations in Theoretical Physics - Session 1 / 196

New implementation of the sector decomposition on FORM

Dr. UEDA, Takahiro [1]; Dr. FUJIMOTO, Junpei [1]

[1] *KEK*

Corresponding Author: uedat@post.kek.jp

Nowadays the sector decomposition technique, which can isolate divergences from parametric representations of integrals, becomes quite useful tool for numerical evaluations of the Feynman loop integrals. It is used to verify the analytical results of multi-loop integrals in the Euclidean region, or in some cases practically used in the physical region by combining with other methods handling the threshold. In an intermediate stage of the sector decomposition for the multi-loop integrals, one often have to handle enormously large expressions containing tons of terms. The symbolic manipulation system FORM is originally designed to treat such a huge expressions and has strong advantage for it. In this talk, the implementation of the sector decomposition algorithm on FORM is discussed. A number of concrete examples including cases of multi-loop diagrams are also shown.

Computing Technology for Physics Research / 197

The DAQ/HLT system of the ATLAS experiment

Dr. DOS ANJOS, André [1]

[1] *University of Wisconsin, Madison, USA*

Corresponding Author: andre.dos.anjos@cern.ch

The DAQ/HLT system of the ATLAS experiment at CERN, Switzerland, is being commissioned for first collisions in 2009. Presently, the system is composed of an already very large farm of computers that accounts for about one-third of its event processing capacity. Event selection is conducted in two steps after the hardware-based Level-1 Trigger: a Level-2 Trigger processes detector data based on regions of interest (RoI) and an Event Filter operates on the full event data assembled by the Event Building system. The detector readout is fully commissioned and can be operated at its full design capacity. This places on the High-Level Triggers system the responsibility to maximize the quality of data that will finally reach the offline reconstruction farms.

This paper brings an overview of the current ATLAS DAQ/HLT implementation and performance based on studies originated from its operation with simulated, cosmic particles and first-beam data. Its built-in event processing parallelism is discussed for both HLT levels as well as an outlook of options to improve it.

Computing Technology for Physics Research / 198

VARIOUS RUNTIME ENVIRONMENTS IN GRID BY MEANS OF VIRTUALIZATION OF WORKING NODES

KRYUKOV, Alexander [1]; DEMICHEV, Andrey [1]; GORBUNOV, Ilya [1]; ILYIN, Slava [1]; SHAMARDIN, Lev [1]

[1] *Skobeltsyn Institute for Nuclear Physics Moscow State University*

Corresponding Author: alexander.kryukov@cern.ch

Grid systems are used for calculations and data processing in various applied areas such as biomedicine, nanotechnology and materials science, cosmophysics and high energy physics as well as in a number of industrial and commercial areas. However, one of the basic problems costing on a way to wide use of grid systems is related to the fact that applied jobs, as a rule, are developed for execution in a definite runtime environment specified by type and version of operating systems, auxiliary software (libraries), type of file system, presence or absence of facilities for parallel computing, etc. On the other hand, working nodes in the resource grid centers (where the jobs are executed) operate under control of a certain OS and offer a fixed runtime environment. Therefore if applied jobs were not developed initially for the particular runtime environment of the WNs, they cannot be directly processed in the grid.

In the framework of this work an approach [1] for batch processing of computer jobs prepared for various runtime environments in grid is proposed. This method is based on the virtualization of working nodes of grid resource centers and enables executing applied jobs irrespective of runtime environment it has been initially developed for. In particular, jobs developed for execution in the environment of the widespread OS Windows, can be processed in resource centers of the grid system whose working nodes operate under OS Linux.

The realization of the proposed approach was made under gLite MW in the EGEE/WLCG project and was successfully tested in SINP MSU resource center.

[1] V.A.Ilyin, A.P.Kryukov, L.V.Shamardin, A.P.Demichev, I.N.Gorbunov, A method for submitting and processing jobss prepared for various runtime environments in grid, Numerical Methods and Programming, v. 9, pp. 41-47, 2008 (in Russian)

Computing Technology for Physics Research / 199

The Advanced Resource Connector for Distributed LHC Computing

CAMERON, David [1]; KONSTANTINOV, Aleksandr [1]; OULD-SAADA, Farid [1]; PAJCHEL, Katarina [1]; READ, Alexander [1]; SAMSET, Bjørn [1]; TAGA, Adrian [1]

[1] *University of Oslo*

Corresponding Author: david.cameron@cern.ch

The NorduGrid collaboration and its middleware product, ARC (the Advanced Resource Connector), span institutions in Scandinavia and several other countries in Europe and the rest of the world. The innovative nature of the ARC design and flexible, lightweight distribution make it an ideal choice to connect heterogeneous distributed resources for use by HEP and non-HEP applications alike. ARC has been used by scientific projects for many years and through experience it has been hardened and refined to a reliable, efficient software product. In this paper we present the architecture and design of ARC and show how ARC's simplicity eases application integration and facilitates taking advantage of distributed resources. Example applications are shown along with some results from one particular application, simulation production and analysis for the ATLAS experiment, as an illustration of ARC's common usage today. These results demonstrate ARC's ability to handle significant fractions of the computing needs of the LHC experiments today and well into the future.

200

ATLAS Muon Spectrometer Monitoring using cosmic rays and First Beam Data

Dr. BENEKOS, Nectarios [1]; Prof. LISS, Anthony Michael [1]; CORTES GONZALEZ, Arely [1]; COGGESHALL, James Christopher [1]

[1] *University of Illinois*

Corresponding Author: nectarios.benekos@cern.ch

ATLAS is a large multipurpose detector, presently in the final phase
of construction at LHC, the CERN Large Hadron Collider accelerator.
In ATLAS the Muon Spectrometer (MS) is optimized to measure final state muons of 14 TeV proton-proton interactions with a good momentum resolution of 2-3% at 10-100 GeV/c and 10% at 1 TeV, and an efficiency close to 100%, taking into account the high level background environment, the inhomogeneous magnetic field, and the large size of the apparatus (24 m diameter by 44 m length).

The complexity of the experiment, the number of electronics channels and the high rate of acquired events requires a detailed commissioning of the detector, during the installation phase of the experiment and in the early life of ATLAS, to verify the correct behaviour of the hardware and software systems.
This is done through the acquisition, monitoring, reconstruction and validation of calibration signals as well as processing data obtained with cosmic ray muons.
Tools to monitor the status of the ATLAS MS and determine the quality of the data are being developed.
The goal is to spot problems during data taking and flag the data. Careful monitoring of data is especially important at the beginning of an experiment where the environment is new and requires some experience to fully comprehend.
Therefore developing such tools can help the experiment to quickly determine problems and then proceed to solve them efficiently.

Thursday, 06 November 2008 / 202

Getting ready for next generation computing

Prof. LINDENSTRUTH, Volker [1]

[1] *Kirchhoff Institute for Physics*

Corresponding Author: voli@kip.uni-heidelberg.de

The ALICE High Level Trigger is a high performance computer, setup to process the ALICE on-line data, exceeding 25GB/sec in real time. The most demanding detector for the event reconstruction is the ALICE TPC. The HLT implements different kinds of processing elements, including AMD, Intel processors, FPGAs and GPUs. The FPGAs perform an on the fly cluster reconstruction and the tracks are planned to be computed on GPUs for speed. The ALICE event reconstruction software is designed from scratch to support multi core architectures. The status of the system- and analysis code architecture, which are optimised for speed are presented.

Several design and programming features of the ALICE HLT have been integrated into the plan to build a T2 "Landesrechner" in Frankfurt with up to 3000 processing nodes. The architecture and status of the project will be outlined. The planned system is founding member of the German Gauss Alliance.

Computing Technology for Physics Research / 203

FairRoot Framework

Dr. AL-TURANY, Mohammad [1]; Dr. UHLIG, Florian [1]

[1] *GSI DARMSTADT*

Corresponding Author: m.al-turany@gsi.de

The new development in the FairRoot framework will be presented. FairRoot is the simulation and anaysis framework used by CBM and PANDA at FAIR/GSI experiments. The CMake based building and testing system will be described. A new event display based on EVE-package from ROOT and Geane will be shown, also the new developments for using GPUs and multi-core systems will be discussed.

204

Resonance analysis in ALICE

PULVIRENTI, Alberto [1]

[1] *University of Catania - INFN Catania*

Corresponding Author: alberto.pulvirenti@ct.infn.it

ALICE is the LHC experiment most specifically aimed at studying the hot and dense nuclear matter produced in Pb-Pb collisions at 5.5 TeV, in order to investigate the properties of the Quark-Gluon Plasma, whose formation is expected in such conditions.
Among the physics topics of interest within this experiment, resonances play a fundamental role, since they allow one to probe the chiral symmetry restoration and to estimate the lifetime of the fireball.
In the ALICE official analysis and simulation framework, a complete package has been developed which is devoted to this topic and has been optimized for an efficient management of a huge amount of data. Besides, it has been optimized to deal with all the distributed analysis environments available for the ALICE collaboration and has been integrated in the general correction framework, under development within the collaboration itself.
Results and details of this package will be illustrated.

Tuesday, 04 November 2008 - Morning session 2 / 205

MonALISA : A Distributed Service System for Monitoring, Control and Global Optimization

LEGRAND, Iosif [1]; NEWMAN, Harvey [1]; VOICU, Ramiro [1]; CIRSTOIU, Catalin [2]; GRIGORAS, Costin [3]; DOBRE, Ciprian [4]

[1] *CALTECH*

[2] *UPB*

[3] *CERN*

[4] *UBP*

Corresponding Author: iosif.legrand@cern.ch

The MonALISA (Monitoring Agents in A Large Integrated Services Architecture) framework provides a set of distributed services for monitoring, control, management and global optimization for large scale distributed systems. It is based on an ensemble of autonomous, multi-threaded, agent-based subsystems which are registered as dynamic services. They can be automatically discovered and used by other services or clients. The distributed agents can collaborate and cooperate in performing a wide range of management, control and global optimization tasks using real time monitoring information.

An essential part of managing global-scale systems is a monitoring system that is able to monitor and track in real time many site facilities, networks, and tasks in progress. The monitoring information gathered is essential for developing the required higher level services, the components that provide decision support and some degree of automated decisions and for maintaining and optimizing workflow in large scale distributed systems. These management and global optimization functions are performed by higher level agent-based services. Current applications of MonALISA's higher level services include optimized dynamic routing, control and optimization for large scale data transfers on dedicated circuits, data transfers scheduling, distributed job scheduling and automated management of remote services among a large set of grid facilities. MonALISA is currently used around the clock in several major projects and has proven to be both highly scalable and reliable. More than 320 services are running at sites around the world, collecting information about computing facilities, local and wide area network traffic, and the state and progress of the many thousands of concurrently running jobs.

Methodology of Computations in Theoretical Physics - Session 1 / 206

Unitarity Methods For 1-Loop Amplitudes

PERKINS, Warren [1]

[1] *Swansea University UK*

Corresponding Author: w.perkins@swan.ac.uk

Unitarity methods provide an efficient way of calculating 1-loop amplitudes for which Feynman diagram techniques are impracticable. Recently several approaches have been developed that apply these techniques to systematically generate amplitudes. The 'canonical basis' implementation of the unitarity method will be discussed in detail and illustrated using seven point QCD processes.

Methodology of Computations in Theoretical Physics - Session 2 / 207

Numerical calculations of Multiple Polylog functions

Dr. KURIHARA, Yoshimasa [1]

[1] *KEK*

Corresponding Author: yoshimasa.kurihara@kek.jp

Multiple Polylog functions (MPL) often appear as a result of the Feynman parameter integrals in higher order correction in quantum field theory. Numerical evaluation of the MPL with higher depth and weight is necessary for multi-loop calculations. We propose a purely numerical method to evaluate MPL using numerical contour integral in multi-parameter complex-plane. We can obtain values of MPL for any complex variables.

208

Advanced online system for data analysis and Monte-Carlo simulations in High Energy Physics

Mr. NICULESCU, Mihai [1]; Mr. SEVCENCO, Adrian [1]; Mr. MITU, Ciprian Mihai [1]; Dr. ZGURA, Sorin Ion [1]; Dr. JIPA, Alexandru [2]

[1] *Institute of Space Sciences*

[2] *University of Bucharest - Faculty of Physics*

Corresponding Author: mihai.niculescu@cern.ch

In this paper we present an integrated system for online Monte Carlo simulations in High Energy Physics. Several Monte Carlo simulations codes will be implemented: GEANT, PYTHIA, FLUKA, HIJING. This system will be structured in several basic modules. First module will ensure the system's web interface, the access to the other modules and will allow the logging of many users at the same time. Another module will be made of several Monte Carlo simulation codes for studying phenomena in High Energy Physics, these phenomena being treated differently according to their specific simulation codes. This module will be a scalable one, that will allow the further addition of other codes. A third module will contain analysis programs in order to obtain scientific results from simulated data or from experimental data. The last module will have the necessary Application Program Interface (API) for the management of the connections and of data in databases: experimental data and simulated data for comparative studies. One of the this system's applications (as a consequence of its accessibility and functionality) will be the development and implementation of educational modules: laboratories, classes and seminars.

Wednesday, 05 November 2008 / 209

CompHEP status report (version 4.5)

Dr. SHERSTNEV, Alexander [1]

[1] *University of Oxford*

Corresponding Author: sherstnv@theory.sinp.msu.ru

We present a new version of the CompHEP program package, version 4.5. We describe shortly new techniques and options implemented: interfaces to ROOT and HERWIG, generation of the XML-based header in event files (HepML), full implementation of Les Houches agreements (LHA I, SUSY LHA, LHA PDF, Les Houches events), realisation of the improved von Neumann procedure for the event generation, etc. We also mention a few concrete and physically motivated examples of CompHEP-based event generators, which are important for the LHC experiments.

Data Analysis - Algorithms and Tools / 211

Fireworks: A Physics Event Display for CMS

Dr. JONES, Christopher [1]

[1] *CORNELL UNIVERSITY*

Corresponding Author: cdj@mail.lns.cornell.edu

Event displays in HEP are used for many different purposes, e.g. algorithm debugging, commissioning, geometry checking and physics studies. The physics studies case is unique since few user are likely to become experts on the event display, the breadth of information all such users will want to see is quite large although any one user may only want a small subset of information and the best way to display physics information sometimes requires a stylized rather than 3D accurate representation. Fireworks is a CMS event display which is specialized for the physics studies case. Fireworks provides an easy to use interface which allows a physicist to concentrate only on the data to which they are interested. Data is presented via graphical and textual views. Cross view data interpretation is easy since the same object is shown using the same color in all views and if the object is selected it is highlighted in all views. Objects which have been selected can be further studied by displaying a detailed view of just that object. Physicists can select which events (e.g. require a high energy muon), what data (e.g. which track list) and which items in a collection (e.g. only high-pt tracks) to show. Once the physicist has configured Fireworks to their liking they can save the configuration. Fireworks is built using the Eve subsystem of the CERN ROOT project and CMS's FWLite project. The FWLite project was part of CMS's new Event Data Model and recent code redesign which separates data classes into libraries separate from algorithms producing the data and uses ROOT directly for C++ object storage thereby allowing the data classes to be used directly in ROOT. The Fireworks project released its first Linux and mac version this summer and has received much positive feedback.

Data Analysis - Algorithms and Tools / 212

The ALICE Global Redirector. A step towards real storage robustness.

Dr. FURANO, Fabrizio [1]

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

Corresponding Author: fabrizio.furano@cern.ch

In this talk we address the way the ALICE Offline Computing is starting to exploit the possibilities given by the Scalla/Xrootd repository globalization tools. These tools are quite general and can be adapted to many situations, without disrupting existing designs, but adding a level of coordination among xrootd-based storage clusters, and the ability to interact between them.

Computing Technology for Physics Research / 213

A prototype of a dinamically expandable Virtual Analysis Facility

LUSSO, Stefano [1]; MASERA, Massimo [2]; BAGNASCO, Stefano [1]; BERZANO, Dario [2]

[1] *Istituto Nazionale di Fisica Nucleare (INFN)*

[2] *Istituto Nazionale di Fisica Nucleare (INFN) and University of Torino*

Corresponding Author: stefano.bagnasco@cern.ch

Current Grid deployments for LHC computing (namely the WLCG infrastructure) do not allow efficient parallel interactive processing of data. In order to allow physicists to interactively access subsets of data (e.g. for algorithm tuning and debugging before running over a full dataset) parallel Analysis Facilities based on PROOF have been deployed by the ALICE experiment at CERN and elsewhere. Whereas large Tier-1 centres can afford to build such facilities at the expense of their Grid farms, this is likely not to be true for smaller Tier-2s centres. Leveraging on the virtualisation of highly performant multi-core machines, it is possible to build a fully virtual Analysis Facility on the same Worker Nodes that compose an existing LCG Grid Farm. Using the Xen paravirtualisation hypervisor, it is then possible to dynamically move resources from the batch instance to the interactive one when needed. We present the status of the prototype being developed.

Data Analysis - Algorithms and Tools / 214

ATLAS trigger status and results from commissioning operations

Dr. MICHELA, Biglietti [1]

[1] *UNIVERSITY OF NAPOLI and INFN*

Corresponding Author: michela.biglietti@cern.ch

The ATLAS trigger system is designed to select rare physics processes of interest from an extremely high rate of proton-proton collisions, reducing the LHC incoming rate of about 10^7. The short LHC bunch crossing period of 25 ns and the large background of soft-scattering events overlapped in each bunch crossing pose serious challenges, both on hardware and software, that the ATLAS trigger must overcome in order to efficiently select interesting events. The ATLAS trigger consists of hardware based Level-1, and a two-level software based High-Level Trigger (HLT). Data bandwidth and processing times in the higher level triggers are reduced by region of interest guidance in the HLT reconstruction steps. High flexibility is critical in order to adapt to the changing luminosity, backgrounds and physics goals. This is achieved by inclusive trigger menus and modular software design. Selection algorithms have been developed which provide the required elasticity to detect different physics signatures and to control the trigger rates. In this talk an overview of the ATLAS trigger design, status and expected performance, as well as the results from the on-going commissioning with cosmic rays and first LHC beams, is presented.

Data Analysis - Algorithms and Tools / 215

Interactive Data Analysis with PROOF, Experience at GSI

MALZACHER, Peter [1]; KRESHUK, Anna [2]; MANAFOV, Anar [2]; PENSO, Victor [2]; PREUSS, Carsten [2]; SCHWARZ, Kilian [2]; ZYNOVYEV, Mykhaylo [2]

[1] *Gesellschaft fuer Schwerionenforschung mbH (GSI)*

[2] *GSI*

Corresponding Author: p.malzacher@gsi.de

This presentation discusses activities at GSI to support interactive data analysis for the LHC experiment ALICE. GSI is a tier-2 centre for ALICE. One focus is a setup where it is possible to dynamically switch the resources between jobs from the Grid, jobs from the local batch system and the GSI Analysis Facility (GSIAF), a PROOF farm for fast interactive analysis. The second emphasis is on creating PROOF clusters on demand - on a batchfarm or on the Grid. Our experience with PROOF has also allowed us to develop some additional features, that simplify PROOF analysis for users and make it even more interactive.

Methodology of Computations in Theoretical Physics - Session 1 / 216

FormCalc 6

HAHN, Thomas [1]

[1] *MPI Munich*

Corresponding Author: hahn@mppmu.mpg.de

The talk will cover the latest version of the Feynman-diagram calculator FormCalc. The most significant improvement is the communication of intermediate expressions from FORM to Mathematica and back, for the primary purpose of introducing abbreviations at an early stage. Thus, longer expressions can be treated and a severe bottleneck in particular for processes with high multiplicities removed.

217

GENIUS/EnginFrame Grid Portal: new features and enhancements

ARDIZZONE, Valeria [1]; BARBERA, Roberto [2]; FALZONE, Alberto [3]; LA ROCCA, Giuseppe [1]; VENUTI, Nicola [3]

[1] *Istituto Nazionale di Fisica Nucleare (INFN) Sez. Catania – Italy*

[2] *University of Catania and INFN – Italy*

[3] *NICE srl*

Corresponding Author: alberto.falzone@nice-italy.com

In order to address new challenges in modern e-Science and technological developments, the needs to have a transparent access to the distributed computational and storage resources within the grid paradigm is becoming of particular importance for different applications and communities.
So far, the basic know-how requested to access the grid infrastructures is not so easy, especially for not ITC expert users. Scientists and researchers have to learn many instructions on the command line interface, and the learning curve might be not easy.
Using the Web technology and its new recent developments, Grid Middlewares details can be hidden to the end users, giving access to the infrastructures in a very easy way as such as the common sense in web usage. This can be done with a Computing Web Portal.
GENIUS Grid portal, powered by EnginFrame, is an increasingly popular mechanism for creating customizable, Web-based interfaces to Grid services and resources.
This work describes GENIUS Grid Portal capabilities such as portal login, VOMS Proxy creation, access control, display management, access to remote catalogs referring storage data, new approach to building reusable portal components as plugins, better performance results as consequence of EnginFrame core functions improvement, simple tools and wizards to compose and submit jobs to the in an easy way.
Finally, new features about integration of new kind of certificates, the Robots one, are described in this work as a facility that helps different communities to use the required VOMS Proxy authentication without use of Personal certificates.

Data Analysis - Algorithms and Tools / 218

Software Validation Infrastructure for the Atlas High-Level Trigger Validation Infrastructure for the ATLAS High-Level Trigger

Mr. GONÇALO, Ricardo [1]; Mr. FERREIRA DE LIMA, Danilo Enoque [2]; FAULKNER, Peter John William; IGONKINA, Olga [3]; GEORGE, Simon [4]; HAMILTON, Andrew [5]; KIRK, Julie Hart [6]; KWEE, Regina [7]; MASIK, Jiri [8]; Prof. MINCER, Allen [9]; MONTICELLI, Fernando [10]; OMACHI, Chihiro [11]; ADORISIO, Cristina [12]; PANIKASHVILI, Natalia [6]; PEREZ REALE, Valeria [13]; POTTER, Chris [14]; QUINONEZ GRANADOS, Fernando; REINSCH, Andreas [15]; RODRIGUEZ PATARROYO, Diego Julian [16]; DE SEIXAS, Jose [17]; SINEV, Nikolai [18]; STROM, David [18]; Dr. SUTTON, Mark [19]; ASK, Stefan [20]; VENTURA, Andrea [21]; WINKLMEIER, Frank [22]; ZHAO, Long [9]; BEAUCHEMIN, Pierre-Hugues [23]; BELL, Paul [8]; BIGLIETTI, Michela [24]; COCCARO, Andrea [25]; OLIVEIRA DAMAZIO, Denis [26]; EHRENFELD, Wolfgang [27]

[1] *Royal Holloway College University of London*

[2] *Federal University of Rio de Janeiro (UFRJ) - COPPE/Poli*

[3] *Nikhef*

[4] *Royal Holloway*

[5] *Universite de Geneve*

[6] *Department of Physics*

[7] *CERN/HUB*

[8] *University of Manchester*

[9] *New York University*

[10] *Universidad Nacional de La Plata (UNLP)*

[11] *Kobe University*

[12] *Dipartimento di Fisica dell' Universita della Calabria e I.N.F.N., Cosenza*

[13] *Columbia University*

[14] *McGill University*

[15] *University of Oregon*

[16] *Univ. Antonio Narino, Theor. Phys.*

[17] *UFRJ*

[18] *University of Oregon, Eugene, Oregon*

[19] *UNIVERSITY COLLEGE LONDON*

[20] *European Organization for Nuclear Research (CERN)*

[21] *Univ. + INFN*

[22] *CERN*

[23] *Oxford University*

[24] *Dipartimento di Scienze Fisiche, Universita di Napoli 'Federico II' e I.N.F.N., Napoli*

[25] *Istituto Nazionale di Fisica Nucleare (INFN)*

[26] *Brookhaven National Laboratory*

[27] *DESY*

Corresponding Author: alina.gabriela.grigoras@cern.ch

The ATLAS trigger system is responsible for selecting the interesting collision events delivered by the Large Hadron Collider(LHC). The ATLAS trigger will need to achieve a ~10-7 rejection factor against random proton-proton collisions, and still be able to efficiently select interesting events. After a first processing level based on FPGAs and ASICS, the final event selection is based on custom software running on two CPU farms, containing around two thousand multi-core machines. This is known as the high-level trigger(HLT).

With more than 100 contributors and around 250 different packages, a thorough validation of the HLT software is essential. This paper describes the existing infrastructure used for validating the HLT software, as well as future plans.

Tuesday, 04 November 2008 - Morning session 1 / 223

Java based software for High-Energy and Astro-physics

JOHNSON, Tony [1]

[1] *SLAC*

Corresponding Author: alina.gabriela.grigoras@cern.ch

This talk will give a brief overview of the features of Java which make it well suited for use in High-Energy and Astro-physics, including recent enhancements such as the addition of parameterized types and advanced concurrency utilities, and its release as an open-source (GPL) product.

I will discuss the current status of a number of Java based tools for High-Energy and Astro-physics including JAS (GUI based analysis tool), WIRED (event display), AIDA (analysis toolkit). I will give examples of their use for building web-based and GUI based applications citing examples from GLAST (recently renamed the Fermi Gamma-Ray Space Telescope) and linear-collider detector R

I will also discuss the methodologies employed in developing such toolkits, challenges involved in supporting them, and lessons that can be learned for the future.

Monday, 03 November 2008 - Morning session 1 / 224

The Blue Brain Project - Simulation-based Research in Neuroscience

Mr. SCHUERMANN, Felix

Corresponding Author: alina.gabriela.grigoras@cern.ch

The initial phase of the Blue Brain Project aims to reconstruct the detailed cellular structure and function of the neocortical column (NCC) of the young rat. As a collaboration between the Brain Mind Institute of the Ecole Polytechnique Federale de Lausanne (EPFL) and IBM the project is based on the many years of experimental data from an electrophysiology lab and a dedicated massively parallel computing resource (4-rack BlueGene/L). Over the last 3 years an interdisciplinary team of 35 researchers has cast the reverse-engineering of the biological pieces and the forward construction of detailed mathematical models in an iterative process that allows continuous refinement. Particular efforts go into the preparation of 10,000 unique morphologically-complex electrical models representing all morpho-electrical classes as well as establishing their structural and functional connectivity. Once a multi-compartmental description for each neuron is generated and the exact locations of the synapses (~30 million) are determined, the simulation is supposed to reproduce emergent properties found in slice experiments. The refinement is directed by a bottom-up calibration that aligns the model across all levels - from the ion channels to the emergent network phenomena - with the experimental data. In order to put the expert in the loop, extensive use of visualization and interactive analysis is made, which is powered by another dedicated supercomputer in order to realize short turn-around times.

Methodology of Computations in Theoretical Physics - Session 1 / 226

Recent developments of GRACE

Dr. KURIHARA, Yoshimasa [1]

[1] *KEK*

Corresponding Author: alina.gabriela.grigoras@cern.ch

Automatic Feynman-amplitude calculation system, GRACE, has been extended to treat next-to-leading order (NLO) QCD calculations. Matrix elements of loop diagrams as well as those of tree level ones can be generated using the GRACE system. A soft/collinear singularity is treated using a leading-log subtraction method.
Higher order re-summation of the soft/collinear correction by the parton shower method is combined with the NLO matrix-element without any double-counting in this method. Event-generators created using a GRACE system are implemented in the GR@PPA, which is an event-generator framework for high energy hadron-collision interactions.

Monday, 03 November 2008 - Morning session 2 / 227

Are SE Architectures Ready For LHC

Mr. HANUSHEVSKY, Andrew [1]

[1] *Stanford Linear Accelerator Center (SLAC)*

Corresponding Author: alina.gabriela.grigoras@cern.ch

There are many ways to build a Storage Element. This talk surveys the common and popular architectures used to construct today's Storage Elements and presents points for consideration. The presentation then asks, "Are these architectures ready for LHC era experiments?". The answer may be surprising and certainly shows that the context in which they are used matters.

Tuesday, 04 November 2008 - Morning session 1 / 228

Introduction to The LLVM Compiler System

Mr. LATTNER, Chris

Corresponding Author: alina.gabriela.grigoras@cern.ch

This talk gives a high level introduction to the LLVM Compiler System (http://llvm.org/), which supports high performance compilation of C and C++ code, as well as adaptive runtime optimization and code generation. Using LLVM as a drop-in replacement for GCC offers several advantages, such as being able to optimize across files in your application, producing better generated code performance, and doing so with reduced compile times.

Wednesday, 05 November 2008 / 229

Multivariate Methods in Particle Physics: Today and Tomorrow

Dr. PROSPER, Harrison [1]

[1] *Department of Physics, Florida State University*

Corresponding Author: alina.gabriela.grigoras@cern.ch

ultivariate methods are used routinely in particle physics research to classify objects or to discriminate signal from background.
They have also been used successfully to approximate multivariate functions. Moreover, as is evident from this conference, excellent easy-to-use implementations of these methods exist, making it possible for everyone to deploy these sophisticated methods. From time to time, however, it is helpful to step back and reflect a little on what is being done. That is the aim of this talk. I begin with a brief introduction to the kind of problems such methods address and follow with a survey of a few of the most promising recent developments. The talk ends with a discussion of what I consider to be the outstanding issues and the prospects for future developments.

230

Computing in HEP at the time of the grid

Dr. BRUN, Rene [1]; Dr. CARMINATI, Federico [1]; Dr. GALLI CARMINATI, Giuliana

[1] *CERN*

Corresponding Author: alina.gabriela.grigoras@cern.ch

This poster presents a book which is due to be published in 2009 about HEP computing. HEP research has been constantly limited by technology, both in the accelerator and detector domains as well as that of computing. At the same time High Energy physicists have greatly contributed to the development of Information Technology. Several developments conceived for HEP have found applications well beyond, the World Wide Web being the most notable example. This book describes the evolution of HEP computing, and in particular those aspects that have been most innovative. The different aspects are described by contributions from different authors in a coherent ensemble. The subject has been treated in the past mostly in conference proceedings or in specialised books on physics computing, but, to our knowledge, no other book has given a complete account of the evolution of HEP computing.

Monday, 03 November 2008 - Morning session 1 / 231

Introduction to ACAT 2008

Mr. CARMINATI, Federico [1]

[1] *CERN*

Monday, 03 November 2008 - Morning session 1 / 232

Introduction to Morning Session

PERRET-GALLIX, Denis [1]

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