

CHEP 07

Sunday 02 September 2007 - Sunday 09 September 2007

Victoria, Canada



Book of Abstracts

Contents

.....	1
A DNS-based load-balancing mechanism for the gLite Workload Management System . . .	1
A Data Skimming Service for Locally Resident Analysis Data	1
A Distributed Storage System with dCache	2
A Distriuted Tier-1 for WLCG	2
A GEANT4 based simulation for proton therapy	3
A Job Monitoring System for the LCG Computing Grid	4
A Light-weight Intrusion Detection System for clusters	5
A Login Shell interface for INFN-GRID	5
A Quantitative Comparison Test of Workload Management Systems	6
A Scientific Overview of Network Connectivity and Grid Infrastructure in South Asian Countries	6
A Search Engine for the Engineering and Equipment Data Management System (EDMS) at CERN	7
A distributed approach for a regional grid operation centre	7
A multi-dimensional view on information retrieval of CMS data	8
A new inclusive secondary vertex algorithm for b-jet tagging in ATLAS	8
A novel design approach and new Geant4 physics developments for microdosimetry simulation	9
A pre-identification for electron reconstruction in the CMS particle-flow algorithm. . . .	10
A software and computing prototype for CMS Muon System alignment	10
A software framework for Data Quality Monitoring in ATLAS	11
A study of the accuracy of Network Time Protocol client synchronization in large computing clusters	11
ALICE - ARC integration	12

ALICE DAQ Online Transient Data Storage	12
ASAP is a system for enabling distributed analysis for the CMS Experiment	13
ATLAS Analysis Model	13
ATLAS Computing System Commissioning - Simulation Production Experience	14
ATLAS Conditions Database Experience with the LCG COOL Conditions Database Project	14
ATLAS DDM Integration in ARC	15
ATLAS Detector Maintenance And Operation Management Tool	15
ATLAS Distributed Data Management Operations. Experience and Projection	16
ATLAS Liquid Argon Calorimeter Reconstruction Software and Commissioning	17
ATLAS Muon Spectrometer Simulation and its Validation Algorithms	17
ATLAS Production Experience on OSG Infrastructure	18
ATLAS Tile Calorimeter Cesium calibration control and analysis software	18
ATLAS Tile Calorimeter Data Quality assessment with commissioning data	19
ATLAS Tile Calorimeter on-line monitoring system based on the Event Filter	19
ATLAS's EventView Analysis Framework	20
Addressing Future HPC Demand with Multi-core Processors	21
Addressing the Pilot Security Problem With gLExec	21
Advances in Integrated Storage,Transfer and Network Management	22
Advances in Monitoring of Grid Services in WLCG	23
AliEn2: the ALICE grid Environment	23
Alignment data streams for the ATLAS Inner Detector.	24
Alignment of the CMS Silicon Tracker using MillePede II	25
Alignment strategy for the CMS tracker	25
An Inconvenient Truth: file-level metadata and in-file metadata caching in the (file-agnostic) ATLAS distributed event store	26
An SSH Key management system: easing the pain of managing key/user association	26
An effective XML based name mapping mechanism within StoRM	27
An interface for GEANT4 simulation using ROOT geometry navigation.	28
An interface to tape for disk pool managers	28

An original model for the simulation of the stopping power of negative hadrons	29
Analysing CMS software performance using IgProf, OProfile and callgrind	29
Analysis Environments for CMS	30
Analysis tools for the LHC experiments	31
Assessment of Data Quality in ATLAS	31
Automatic Model Selection Using Machine Learning Techniques for Event Selection in Particle Physics	31
Automatic processing of CERN video, audio and photo archives	32
BNL dCache Status and Plan	32
BaBar MC Production on the Canadian Grid using a Web Services Approach	33
Benchmarks of medical dosimetry on the grid	34
Benefits of Grid Technology to Running Experiments	34
Beyond Grid Security	35
Booting ROOT with BOOT	35
Building a Scalable Event-Level Metadata System for ATLAS	36
Building a production-grade Grid infrastructure at DESY	37
Building a robust distributed system: some lessons from R-GMA	37
Building the WLCG file transfer service	38
Bulk file transfer and storage with GridSite and SlashGrid	38
CABS3 - CLEO Analysis by Script @ Belle	39
CASTOR2: design and development of a scalable architecture for a hierarchical storage system at CERN	39
CDF Monte Carlo data transfer and storage with Grid tools	40
CDF experience with Monte Carlo production using LCG Grid	40
CDF offline computing'07: computing of a HEP experiment in a mature stage	41
CERN Database Services for the LHC Computing Grid	42
CERN Single Sign On solution	42
CMS CSA06 experience at INFN	43
CMS Centers for Control, Monitoring, Offline Operations and Analysis	44
CMS Conditions Data Access using FroNTier	44

CMS Event Display and Data Quality Monitoring for LHC Startup	45
CMS Experiences with Computing Software and Analysis Challenges	45
CMS MC Production System Development & Design	46
CMS Monte Carlo production in the WLCG Computing Grid	47
CMS Offline Web Tools	48
CMS Online Web Based Monitoring	49
CMS Software Deployment on the Open Science Grid (OSG)	49
CMS Tier Structure and Operation of the Experiment-specific Tasks in Germany	50
CMS Tier0 - design, implementation and first experiences	51
CMS packaging system or: how I learned stop worrying and love RPM spec files	51
COOL Performance Tests and Optimization	52
COOL Software Development and Service Deployment Status	52
CRAB (CMS Remote Analysis Builder)	53
CRONUS: A Condor Glide-in Based ATLAS Production Executor	53
Calibration workflow and dataflow in CMS	54
Canadian Cyberinfrastructure	54
Cathode Strip Chamber (CSC) Raw Data Unpacking and Packing using bit field data classes	54
Circuit Oriented End-to-End Services for HEP Data Transfers	55
Collaborative Tools and the LHC: An Update	56
Commissioning of the ALICE Data-Acquisition System	56
Commissioning of the ATLAS Inner Detector with cosmic rays	57
Commissioning of the ATLAS offline software with cosmic rays	58
Commissioning with cosmic rays of the Muon Spectrometer of the ATLAS experiment at the Large Hadron Collider	58
Complete Distributed Computing Environment for a HEP Experiment: Experience with ARC-Connected Infrastructure for ATLAS	59
Computationally efficient algorithms for the two-dimensional Kolmogorov-Smirnov test	60
Computing Operations at CMS Facilities	61
Computing and Ground Data Handling for AMS-02 Mission	61
Computing in IceCube	62

Concepts, Design and Implementation of the New ATLAS Track Reconstruction (NEWT)	62
Conference Closeout	63
Conference Summary	63
Construct an LHC scale national analysis facility	63
Control and monitoring of alignment data for the ATLAS endcap Muon Spectrometer at the LHC	64
Creating End-to-End Guaranteed Bandwidth Network Paths Across Multiple Domains with TeraPaths	65
Cross middlewares Virtual Organization Authorization	65
DIRAC Agents and Services	66
DIRAC Data Management: consistency, integrity and coherence of data	66
DIRAC Framework for Distributed Computing	67
DIRAC Job Prioritization and Fair Share in LHCb	67
DIRAC Optimized Workload Management	68
DIRAC: A community grid solution	68
DIRAC: Data Production Management	69
DIRAC: Reliable Data Management for LHCb	69
DPM Status and Next Steps	70
DZero Data-Intensive Computing on the Open Science Grid	70
Data Acquisition Backbone Core DABC	71
Data Acquisition at the LHC experiments	71
Data Analysis through the DIANA Meta-Scheduling Approach	72
Data Quality Monitoring and Visualization for the CMS Silicon Strip Tracker	72
Data Quality Monitoring for the CMS Electromagnetic Calorimeter	73
Data Stream handling in the LHCb experiment	74
Data access performance through parallelization and vectored access. Some results.	74
Data management in BaBar	75
Data storage layout and architecture of the German T1	75
Database architecture for the calibration of ATLAS Monitored Drift Tube Chambers	76
Dataharvester - a library for reading and writing "hierarchic tuples" from/to various file formats	77

Deploying HEP Applications Using Xen and Globus Virtual Workspaces	77
Deploying perfSONAR-based End-to-End Monitoring for Production US-CMS Networking	78
Design Principles of a Web Interface for Monitoring Tools	78
Development Status and Plans for the LCG Common Database Access Layer (CORAL)	79
Development of the Tier-1 Facility at Fermilab	80
Development, Deployment and Operations of ATLAS Databases	80
Developments in BaBar simulation - life without a database	81
Distributed Analysis using GANGA on the EGEE/LCG infrastructure	81
Distributed Cluster dynamic storage: A comparison of dcache, xrootd and slashgrid storage systems running on batch nodes.	82
Distributed Data Analysis in LHCb	83
Distributed Database Access in the LHC Computing Grid with CORAL	83
Distributed Interactive Access to Large Amount of Relational Data	84
Distributed database services in PHENIX - what it takes to support a Petabyte experiment.	84
ETICS Meta-Data Software Editing - From Check Out To Commit Operations	85
ETICS: the international software engineering service for the grid	86
EVIO - A Lightweight Object-Oriented Event I/O Package	86
EVO (Enabling Virtual Organizations), the Next Generation Grid-enable Collaborative	87
Efficient Access to Remote Data in High Energy Physics	88
Efficient, Large Scale Data Transfers in Wide Area Network	88
Electron reconstruction in CMS	89
Electronic Shift and handover log and much more, STAR Elog	89
Embedding Python- a new approach to specifying program options	90
Enabling a priority-based FairShare in the EGEE Infrastructure	90
End-to-End Network/Application Performance Troubleshooting Methodology	91
Enriched namespace to support content-aware authorization policies	91
Ensuring GRID resource availability with the SAM framework in LHCb	92
Evaluating Disk Hardware for Small Deployments of PostgreSQL	92
Evaluation of Goodness-of-Fit tests in physical use cases	93

Event reconstruction algorithms for the ATLAS trigger	93
Event reconstruction in the Forward and Backward Silicon detectors of HERA experiment H1	94
Exercising CMS dataflows and workflows in computing challenges at the Spanish Tier-1 and Tier-2 sites	94
Experience Running a Distributed Tier-2 for the ATLAS Experiment	95
Experience and Lessons learnt from running high availability databases on Network At- tached Storage	96
Experience from a Pilot based system for ATLAS	97
Experience with monitoring of Prague T2 Site	97
Experience with the gLite Workload Management System in ATLAS Monte Carlo Produc- tion on LCG	97
Experience with validating GEANT4 v7 and v8 against v6 in BaBar	98
Experience with xrootd/PROOF solution as a data access model for a Computing Center	99
Experiences with gStore, a scalable Mass Storage System with Tape Backend	99
Experiences with the GLUE information schema in the LCG/EGEE production Grid	100
Experimental Evaluation of Job Provenance in ATLAS environment	100
Explicit state representation and the ATLAS event data model: theory and practice	101
Exploring Capable Systems for Fast Data Distribution over End-to-End 10Gb/s Paths	102
Extension of the DIRAC workload-management system to allow use of distributed Win- dows resources	103
Extra Dimensions	103
FPGA based Compute Notes for High Level Triggering in PANDA	104
FairRoot : the FAIR simulation and analysis framework	105
Farm Logbook, a tool to keep/trace all operations on a big computing farm	105
Fast Simulations for the PANDA Experiment at FAIR	106
Fast shower simulation in ATLAS Calorimeter	106
Feicim: A browser for data and algorithms.	107
FermiGrid - Experience and Future Plans	107
Final results of the precision validation of Geant4 models in the pre-equilibrium and nuclear de-excitation phase	108
Fluka and Geant4 simulations using common geometry source and digitization algorithms	109

Fluka and Geant4 simulations using common geometry source and digitization algorithms	109
Full-scale CMS Tracker application of the Kalman Alignment Algorithm	110
Future of Grid Computing	111
GFAL and LCG-Util	111
GRATIA, a resource accounting system for OSG	111
GSIMF: A Grid Software Installation Management Framework	111
Ganga - a job management and optimising tool	112
Geographical failover for the EGEE-WLCG Grid collaboration tools	113
Glance Project: a database retrieval mechanism for the ATLAS detector	114
Global Grid User Support - Building a worldwide distributed user support infrastructure	114
Globus BOF	115
Globus BOF Sessions	115
Grid Information system for EGEE, scalability and performance assessment and plans	115
Grid Interoperability: Joining Grid Information Systems	116
Grid Interoperability: The Interoperations Cookbook	117
Grid Monitoring from the VO/User perspective. Dashboard for the LHC experiments.	117
Grid data storage on widely distributed worker nodes using Scalla and SRM	118
Grid reliability	118
GridKa Tier1 Site Management	119
HEP System Management Working Group and WEB site	119
HEP and Non-HEP Computing at a Laboratory in Transition	120
HEP grids face IPv6: A readiness study	120
HEPTrails: An Analysis Workflow and Provenance Tracking Application	121
HLRmon: a Role-based Grid Accounting Reporting Web Tool	121
HLT Online Calibration framework in ALICE	122
HPC at the Petascale and Beyond	123
HPC at the Petascale and Beyond	123
HepData: restructuring the archive for LHC data	124
High Level Trigger Configuration and Handling of Trigger Tables in the CMS Filter Farm	124

High Performance Data Analysis for Particle Physics using the Gfarm File System	125
High Performance Storage Tests At The INFN Pisa Computing Centre	126
High precision physics simulation: experimental validation of Geant4 Atomic Relaxation	126
High-Performance Stream Computing for Particle Beam Transport Simulations	127
How good is the match between LHC software and current/future processors?	127
HyperNews use in HEP - bigger and better	128
INFN Tier-1 status report	128
ISSEG: Welcome	129
ISSEG: Feedback/questions	129
ISSEG: How to sell Security - and get resources for it	129
ISSEG: Recommendations	129
ISSEG: Risk assessments	130
Identity Management	130
Identity Management	130
Implementation and Performance of the ATLAS Second Level Jet Trigger	130
Implementation of INCL4 cascade with ABLA evaporation in Geant4	131
Implementation of chamber mis-alignments and deformations in the ATLAS Muon Spec- trometer description and estimate of the muon reconstruction performance reconstruc- tion performance	131
Implementing SRM V2.2 Functionality in dCache	132
Implementing SRM V2.2 Functionality in dCache	133
Implementing a Modular Framework in a Conditions Database Explorer for ATLAS	134
Improvements in ROOT I/O's Functionality and Performance	134
Individual Particle Reconstruction	135
Initial results from the MAGIC-II Datacenter at PIC	136
Integrated RT-AT-Nagios system at BNL USATLAS Tier1 computing center	136
Integrating Xgrid technology into HENP distributed computing model	137
Integration of the ATLAS Tag Database with Data Management and Analysis Components	138
Integration of the ATLAS VOMS system with the ATLAS Metadata Interface	138
Integration of the Trigger and Data Acquisition Systems in ATLAS	139

Intelligent Design	140
Interfacing with Sun Utility Computing, experience with on demand physics simulations on SunGrid	140
JAIDA, JAS3, WIRED4 and the AIDA tag library - experience and new developments . . .	141
JetWeb: A library and server for Monte Carlo validation and data comparisons	141
Job Submission and Management Through Web Services: the Experience with the CREAM Service	142
LCGO - geometry description for ILC detectors	142
LHC Expt Computing	143
LHCb Distributed Conditions Database	143
LHCb Online Interface to a Conditions Database	144
LHCb Online event processing and filtering	144
LHCb experience with LFC database replication	145
Lambda Station: Alternate Network Path Forwarding for Production SciDAC Applications	145
Large Scale Access Tests and Online Interfaces to ATLAS Conditions Databases	146
Large-scale ATLAS Production on EGEE	147
Latest Developments in the PROOF System	147
Lessons Learnt From WLCG Service Deployment	148
Level-1 RPC Trigger in the CMS experiment - software for emulation and commissioning	148
Lorentz Angle Calibration for the CMS Pixel Detector	149
Low-level reconstruction software and regional unpacking for the CMS strip tracker . . .	149
Maintenance, validation and tuning of Monte Carlo event generators for the LHC experi- ments in the Generator Services project	150
Management System of Event Processing and Data Files Based on XML Software Tools at Belle	151
Managing ATLAS data on a petabyte-scale with DQ2	152
Managing an Institutional Repository with CDS Invenio	152
Mapping GRID to Site Credentials using GUMS	153
Medical Data Management Status and Plans	153
Metropolitan Area Network Support at Fermilab	154
Modular grid middleware configuration system	154

MonALISA: An Agent Based, Dynamic Service System to Monitor, Control and Optimize Distributed Systems	155
Monitoring a WLCG Tier-1 computing facility aiming at a reliable 24/7 service	156
Monitoring of Grid Behaviour by LHCb	156
Monitoring the ATLAS Production System	157
Monitoring the Atlas Distributed Data Management System	157
Monitoring the EGEE/WLCG Grid Services	158
Monitoring the user/application activities on the grid	158
Monitoring with MonAMI: a case study.	159
Multi-Agent Framework for Experiment Control Systems (AF ECS)	160
Multi-threaded Event Reconstruction with JANA	160
Networks for High Energy Physics and Data Intensive Science, and the Digital Divide . .	160
Next Steps in the Evolution of GridICE: a Monitoring Tool for Grid Systems	161
Next generation of OpenGL support in ROOT	161
Nightly builds and software distribution in the LCG / AA / SPI project	162
On using a generic framework for integrating advanced batch systems to production-level grid infrastructures	163
Online Data Monitoring in the LHCb experiment	163
Online and offline software for the CMS strip tracker data acquisition system	164
Opening Session	165
Optimising LAN access to grid enabled storage elements	165
Optimization of dCache MSS tape efficiency through Virtual Volumes	165
Optimizations in Python-based HEP Analysis	166
Optimized tertiary storage access in the dCache SE	166
Oracle RAC (Real Application Cluster) application scalability, experience with PVSS and methodology	167
Overview of the Inner Silicon detector alignment procedure and techniques in the RHIC/STAR experiment	168
PanDA: Distributed production and distributed analysis system for ATLAS	168
Parameterization of the LHCb magnetic field map	169
Perfmon2 - A leap forward in performance monitoring	169

Performance Measurement and Monitoring for HENP Applications	170
Performance study of multicore systems with the ATLAS software Athena	170
PetaCache: Data Access Unleashed	171
Physics Analysis Tools for Beauty Physics in ATLAS	171
Physics Analysis Tools for CMS experiment at LHC	172
Portable Monitoring for Castor2	172
Power and Air Conditioning Challenges in Computer Centres	173
Power, Density, Reliability & Performance: customer-driven evolution of cluster computing and storage, in high energy physics and other scientific fields.	173
Preparation for WLCG production from a Tier-1 viewpoint	174
Primary Vertex Reconstruction in the ATLAS Experiment at the LHC	174
Printing at CERN	175
Production Experience Using Resilient dCache	176
Production Experience with Distributed Deployment of Databases for the LHC Computing Grid	176
Production of BaBar skimmed analysis datasets using the Grid	177
Providing a Single View of Heterogeneous Clusters using Platform LSF	177
Publication in scientific journals: the role of HEP computing	178
Quattor and QWG Templates : efficient management of (complex) grid sites	178
RGLite, an interface between ROOT and gLite	179
ROOT Graphics: status and future.	180
ROOTlets and Pythia: Grid enabling HEP applications using the Clarendon Toolkit	180
RSS based CERN Alerter	181
Rapid-response Adaptive Computing Environment	181
Raw-data display and visual reconstruction validation in ALICE	183
Readiness of CMS Simulation towards LHC Startup	183
Real-time Data Access Monitoring in Distributed, Multi-Petabyte Systems	184
Real-time analysis of the operational state of the CMS strip tracker readout system	184
Real-time dataflow and workflow with the CMS Tracker data	185
Recent Developments in LFC	187

Recent Developments of the ROOT Mathematical and Statistical Software	187
Reconstruction and identification of Tau decays at CMS	188
Reconstruction of Converted Photons in CMS	188
Relational databases for conditions data and event selection in ATLAS	189
Release Management - the EGEE project's approach	189
Remote Management of nodes in the ATLAS Online Processing Farms	190
Replication and load Balancing strategy of STAR's RDBM	190
Resource sharing and deployment using virtualization at CERN openlab	191
Rivet: a system for event generator validation and development	191
Role of Advanced Computation in the Design of the International Linear Collider.	192
Role of Digital Forensics in Service Oriented Architectures	192
Running BaBar simulation on "public" Grid resources in Italy	193
Running CE and SE in a Xen-virtualized environment	193
Runtime memory and data monitoring using Reflex for memory usage evaluation and data quality control.	194
Scaling CMS data transfer system for LHC start-up	194
Searching for CMS data: Rapid web development using python and AJAX	195
Security Incidents management in a Grid environment	196
Service Level Status - a new real-time status display for IT services	196
Shaping Collaboration 2006: Action Items for the LHC	197
Sharing LCG files across different platforms	198
Simulation and event reconstruction inside the PandaRoot frameworks.	198
Simulation readiness for the first data at LHC	199
Software for CMS Reconstruction	199
Software for the Commissioning of the CMS Micro Strip Silicon Tracker	200
Software representation of the ATLAS magnetic field	200
StatPatternRecognition: A C++ Package for Multivariate Classification	201
Steering of GRID production in ATLAS experiment	201
Storage Resource Manager version 2.2: design, implementation, and testing experience	202
Storage Technology Evaluations at BNL	203

Storage management solutions and performance tests at INFN Tier-1	203
Streamlining and Scaling Castor2 Operations	204
Summary of Collaborative Tools and Initiatives	204
Summary of Distributed data analysis and information management track	205
Summary of Online Computing Track	205
Summary of WLCG Collaboration Workshop 1-2 September 2007	205
Summary of computing facilities, production grids and networking track	206
Summary of event processing track	206
Summary of grid middleware and tools track	206
Summary of software components, tools and databases track	206
Summary of the WLCG Meeting	206
Survey of the main HEP software components (1970 -> 2010)	206
TMVA - Toolkit for Multivariate Data Analysis	207
Testing TMVA software in b-tagging for the search of MSSM Higgs bosons at the LHC	207
Testing and integrating the WLCG/EGEE middleware in the LHC computing	208
Testing gLite for releases	209
Testing suite for validation of Geant4 hadronic generators	209
The ALICE Offline Environment	209
The ALICE-LHC Online Data Quality Monitoring Framework	210
The ATLAS Canada Network	210
The ATLAS Computing Model	211
The ATLAS DAQ System Online Configurations Database Service Challenge	211
The ATLAS High Level Trigger Steering	212
The ATLAS METADATA INTERFACE	213
The ATLAS Software Installation System for LCG/EGEE	213
The ATLAS T0 Software Suite	214
The ATLAS Trigger: Commissioning with cosmic-rays	215
The Aragats Data Acquisition System for Highly Distributed Particle Detecting Networks	216
The BaBar Online Detector Control System Upgrade	216
The CEDAR event generator validation project	217

The CERN Analysis Facility - A PROOF Cluster for Day-One Physics Analysis	217
The CMS Dataset Bookkeeping Service	218
The CMS LoadTest 2007: an infrastructure to exercise CMS transfer routes among WLCG Tiers	219
The CMS Storage Manager and Data Flow in the High Level DAQ	219
The CMS Tracker Control System	220
The DZERO Run 2 L3/DAQ System Performance	220
The Design and Implementation of BES Grid Computing System	221
The Distributed Parallel Multi-Platform ATLAS Release Nightly Build System	221
The EELA Grid Infrastructure and HEP Applications in Latin America	222
The GLAST Data Handling Pipeline	222
The Geant4 Virtual Monte Carlo	223
The GridPP Collaboration Web Site, 2000-2007	223
The GridSite security architecture	224
The Health-e-Child Project: A Grid enabled Platform for European Paediatrics	224
The High-Level Trigger at the CMS experiment	225
The IceCube Data Acquisition Software: Lessons Learned during Distributed, Collaborative, Multi-Disciplined Software Development.	226
The LDC Software Framework for the ILC detector	226
The LHC Machine and Experiments: Status and Prospects	227
The LHCb Computing Data Challenge DC06	227
The LHCb High Level Trigger Software Framework	228
The LiC Detector Toy Program	228
The Open Science Grid - Its Status and Implementation Architecture	229
The PHENIX Experiment in the RHIC Run 7	230
The Pierre Auger Observatory Offline Software	230
The Pre-Production Service in WLCG/EGEE	231
The RAVE/VERTIGO Vertex Reconstruction Toolkit and Framework	232
The Reconstruction and Calibration of the BESIII Drift Chamber	232
The Run Control and Monitoring System of the CMS Experiment	233

The Simulation of the CMS Electromagnetic Calorimeter	234
The Status of Grid for Belle Experiment	234
The Swiss ATLAS Grid	234
The Tier0 road to LHC Data Taking	235
The US LHCNet Network for HEP	235
The Virtual Geometry Model	236
The configuration system of the ATLAS Trigger	236
The gLite Workload Management System	237
The global chi2 track fitter in ATLAS	238
The life cycle of HEP offline software	238
The performance of the Geant4 Standard EM package for LHC and other applications	239
Tools for the management of stored data and transfer of data: DPM and FTS	239
Towards GLUE 2: Evolution of the Computing Element Information Model	240
Towards a full implementation of a robust solution of a Domain Specific Visual Query Language DSVQL for HEP analysis	241
Towards the Integration of StoRM on Amazon Simple Storage Service (S3)	241
Track Reconstruction with the CMS Tracking Detector	242
Track based alignment of the ATLAS inner detector	242
Track based software package for measurement of the energy deposited in the calorimeters of the ATLAS detector	243
Track reconstruction of real cosmic muon events with CMS tracker detector	244
Trigger Selection Software for Beauty physics in ATLAS	244
UK Grid Computing for High-Energy Physics	245
US CMS Tier-2 Computing	245
Unified Storage Systems for Distributed Tier-2 Centres	246
University of Michigan Lecture Archiving	246
Usage and Extension of Maven to build, release and distribute Java and Native Code Projects	247
Usage of LSF for the batch farms at CERN	247
Use of Alternate Path WAN Circuits for CMS	248
Use of Cfengine for deploying LCG/gLite middle-ware	248

Use of Flow Data for Traffic Analysis and Network Performance Characterization	249
Use of GEANE for tracking in Virtual Monte Carlo	249
User Centric Monitoring (UCM) information service for the next generation of Grid-enabled scientists	250
Using Parrot to access the CDF software on LCG Grid sites	251
Using ROOT with Modern Programming Languages	252
Using the Grid for Large Scale and Nightly Testing of the ATLAS Trigger & Data Acquisition System	252
VINCI : Virtual Intelligent Networks for Computing Infrastructures	253
Virtual Organization Trustworthiness in the Grid world	253
Virtualization applications at the Brookhaven Computing Facility	254
WEB Based Online Event Displays for KASCADE-Grande	254
WLCG - Where we are now and the Real Data Challenges ahead	255
WLCG scale testing during CMS data challenges	255
Web System to support analysis for experimental equipment commissioning	256
Xrootd in BaBar	256
ZEUS Grid Usage: Monte Carlo Production and Data Analysis	257
'Anvil' a case study of the Unified Software Development Process	258
cMsg - A general purpose, publish-subscribe, interprocess communication implementation and framework	258
dCache NFSv4.1 - true nfs server with SRM interface and HSM support	259
dCache, the Challenge.	259
gPlazma in dCache	260
glExec - gluing grid computing jobs to the Unix world	260
glideinWMS - A generic pilot-based Workload Management System	261

495

Poster 2 - Board: 100 / 396

A DNS-based load-balancing mechanism for the gLite Workload Management System

Authors: Alessandro Maraschini¹ ; Daniele Cesini² ; Davide Salomoni² ; Fabrizio Pacini¹ ; Francesco Giacomini² ; Marco Cecchi²

¹ DATAMAT

² INFN-CNAF

Corresponding Authors: marco.cecchi@cnafe.infn.it, davide.salomoni@cnafe.infn.it

Since the beginning, one of the design guidelines for the Workload Management System currently included in the gLite middleware was flexibility with respect to the deployment scenario: the WMS has to work correctly and efficiently in any configuration: centralized, decentralized, and in perspective even peer-to-peer.

Yet the preferred deployment solution is to concentrate the workload management functionality on a small number of certified hosts. This is certainly favored by system and virtual organization administrators, because it limits the amount of system management needed to provide the service. But it also helps the users of the system because it simplifies the configuration of the user interface. On the negative side, it raises some scalability problems, because the overall system is requested to manage millions of jobs in the not-too-distant future.

In this paper we show how a well-known technique, hiding a number of machines under a DNS-based alias mechanism that takes into account suitable load parameters typical of and specific to Workload Management Systems, can be easily applied to the gLite WMS, addressing both the scalability and the usability issues mentioned above and paving the way to more advanced abstraction mechanisms.

Poster 2 - Board: 9 / 34

A Data Skimming Service for Locally Resident Analysis Data

Author: Marco Mambelli¹

Co-authors: David Malon² ; Jack Cranshaw² ; Jerry Gieraltowsky² ; May Edward² ; Robert Gardner³

¹ University of Chicago

² Argonne National Laboratory

³ UNIVERSITY OF CHICAGO

Corresponding Author: marco@hep.uchicago.edu

A Data Skimming Service (DSS) is a site-level service for rapid event filtering and selection from locally resident datasets based on metadata queries to associated “tag” databases. In US ATLAS, we expect most if not all of the AOD-based datasets to be replicated to each of the five Tier 2 regional facilities in the US Tier 1 “cloud” coordinated by Brookhaven National Laboratory. Entire datasets will consist

of on the order of several terabytes of data, and providing easy, quick access to skimmed subsets of these data will be vital to physics working groups. Typically, physicists will be interested in portions of the complete datasets, selected according to event-level attributes (number of jets, missing E_t , etc) and content (specific analysis objects for subsequent processing).

In this paper we describe methods used to classify data (metadata tag generation) and to store these results in a local database. Next we discuss a general framework which includes methods for accessing this information, defining skims, specifying event output content, accessing locally available storage through a variety of interfaces (SRM, dCache/dccp, gridftp), accessing remote storage elements as specified, and user job submission tools through local or grid schedulers.

The advantages of the DSS are the ability to quickly “browse” datasets and design skims, for example, pre-adjusting cuts to get to a desired skim level with minimal use of compute resources, and to encode these analysis operations in a database for re-analysis and archival purposes. Additionally the framework has provisions to operate autonomously in the event that external, central resources are not available, and to provide, as a reduced package, a minimal skimming service tailored to the needs of small Tier 3 centers or individual users.

Grid middleware and tools / 148

A Distributed Storage System with dCache

Authors: Gerd Behrmann¹ ; Josva Kleist¹ ; Patrick Fuhrmann²

¹ *Nordic Data Grid Facility*

² *Deutsches Elektronen-Synchrotron DESY*

Corresponding Author: kleist@ndgf.org

The LCG collaboration is encompassed by a number of Tier 1 centers. The nordic LCG Tier 1 is in contrast to other Tier 1 centers distributed over most of Scandinavia. A distributed setup was chosen for both political and technical reasons, but also provides a number of unique challenges. dCache is well known and respected as a powerful distributed storage resource manager, and was chosen for implementing the storage aspects of the nordic Tier 1. In contrast to classic dCache deployments, we deploy dCache over a WAN with limited bandwidth, high latency, frequent network failures, and spanning many administrative domains. These properties provide unique challenges, covering topics such as security, administration, maintenance, upgradability, reliability, and performance. Our initial focus has been on implementing the GridFTP 2 OGF recommendation in dCache and the Globus Toolkit. Compared to GridFTP 1, GridFTP 2 allows for more intelligent data flow between clients and storage pools, thus enabling more efficient use of our limited bandwidth. Future efforts will address other issues, such as reliability in case of network separation.

Computer facilities, production grids and networking / 147

A Distributed Tier-1 for WLCG

Authors: Josva Kleist¹ ; Lars Fischer¹ ; Michael Grønager¹ ; Oxana Smirnova¹

¹ *Nordic Data Grid Facility*

Corresponding Authors: lars@nordu.net, kleist@ndgf.org

The Tier-1 facility operated by the Nordic DataGrid Facility (NDGF) differs significantly from other Tier-1s in several aspects: It is not located one or a few locations but instead distributed throughout the Nordic, it is not under the governance of a single organization but instead a “virtual” Tier-1 build out of resources under the control of a number of different national organizations.

We present the technical implications of these aspects as well as the high-level design of this distributed Tier-1. The focus will be on computations, storage and monitoring.

In order to look like a single site some services requires one single entry point. Most notably is the srm entry point to storage. This is archived using dCache, but with a number of additional features developed by NDGF and DESY/FermiLab.

Computations are controlled by the NorduGrid ARC middleware, which has proven to be a reliable and easy to install at the sites and on top of that supports the distribution of the compute elements very well. Integration happens at the HEP experiment level, where one or more VO-boxes provide the interface between the experiments and the ARC middleware.

Finally, diagnosing problems at sites where NDGF staff does not have administrative access to the machines requires a well developed set of monitoring services. This includes SAM tests especially tailored to the NDGF setup and use of other standard monitoring software packages.

Event processing / 267

A GEANT4 based simulation for proton therapy

Author: Tsukasa Aso¹

Co-authors: Akinori Kimura²; Inder Daftari³; Koichi Murakami⁴; Satoru Kameoka⁴; Takashi Akagi⁵; Takashi Sasaki⁴; Teiji Nishio⁶

¹ Toyama National College of Maritime Technology, JST CREST

² Ashikaga Institute of Technology, JST CREST

³ UCSF

⁴ KEK, JST CREST

⁵ Hyogo Ion Beam Medical Center, JST CREST

⁶ National Cancer Center, JST CREST

Corresponding Author: aso@toyama-cmt.ac.jp

The GEANT4 Monte Carlo code provides many powerful functions for conducting particle transport simulations with great reliability and flexibility. GEANT4 has been extending the application fields for not only the high energy physics but also medical physics. Using the reliable simulation for the radiation therapy, it will become possible to validate treatment planning and select the most effective one. For the use of a simulation in the clinical application, the simulation has to reproduce the dose distributions in three-dimensions with the best accuracy for ensuring the patient safety.

As a generalized simulator, the GEANT4 based simulation framework has been developed and used for the verification of the simulated dose distribution to the measurements. Three types of irradiation systems for proton therapy were successfully implemented on the top of this framework; those are the gantry treatment nozzle at the Hyogo Ion Beam Medical Center (HIBMC), the gantry treatment nozzle at the National Cancer Center (NCC), and the eye treatment

facility of UC San Francisco at the Crocker Nuclear Laboratory cyclotron, UC Davis (CNL).

The validation of the simulation was performed for the proton ranges in important materials at beam line and the size of radiation field, respectively. Then dose distributions in simulation were verified with measurements for Bragg peak and spread out Bragg peak, respectively. We will report a brief description of the developed simulation, and the comparisons of simulated dose distributions with measurements as well as the validation of the beam irradiation system.

Poster 2 - Board: 76 / 163

A Job Monitoring System for the LCG Computing Grid

Author: Torsten Harenberg¹

Co-authors: David Meder-Marouelli¹ ; Markus Mechtel¹ ; Peer Ueberholz² ; Peter Mättig¹ ; Stefan Borovac¹

¹ *University of Wuppertal*

² *Niederrhein University of Applied Sciences*

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

Today, one of the major challenges in science is the processing of large datasets. The LHC experiments will produce an enormous amount of results that are stored in databases or files. These data are processed by a large number of small jobs that read only chunks.

Existing job monitoring tools inside the LHC Computing Grid (LCG) provide just limited functionality to the user.

These are either command line tools delivering simple text strings for every job or the provided information is very limited.

Other tools like GridIce focus on the monitoring of the infrastructure rather than the user application/job.

In contrast to these concept, we developed the Python-based “Job execution Monitor”.

Typically, the first thing to be executed on a worker node is not a binary executable, but a script file which sets up the environment (including environment variables and loading of data from a storage element, a tasks known to be critical). It is the goal of the Job Execution Monitor to monitor the execution of such critical commands and report their success or failure to the user.

The core module of the Job Execution Monitor is the script wrapper.

To gain detailed information about the job execution, a given script file (bash or python) is executed command by command. After each command, the complete environment is checked and logged.

Together with the other components of this system, an expert system tries to classify the reason for a failure. An integration into the Global Grid User Support is planned.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

D-Grid

Poster 1 - Board: 76 / 155

A Light-weight Intrusion Detection System for clusters

Author: Alasdair Earl¹

¹ CERN

Corresponding Author: alasdair.earl@cern.ch

The RPMVerify package is a light weight intrusion detection system (IDS) which is used at CERN as part of the wider security infrastructure. The package provides information about potentially nefarious changes to software which has been deployed using the RedHat Package Management system (RPM).

The purpose of the RPMVerify project has been to produce a system which makes use of the existing CERN infrastructure and tackles the scalability limitations of existing IDSs.

In this paper we discuss its design, implementation, limitations, and our experiences in using it. We will specifically comment from the system administration and service management perspective.

Summary:

IDS, Quattor, security, cluster security

Poster 2 - Board: 108 / 454

A Login Shell interface for INFN-GRID

Author: Silvio Pardi¹

Co-authors: Enrico Calloni ² ; Fabio Garufi ² ; Guido Russo ² ; Leopoldo Milano ² ; Rosario De Rosa ³

¹ University of Naples "Federico II" - C.S.I. and INFN

² University of Naples "Federico II"

³ University of Naples "Federico II" and INFN

Corresponding Author: spardi@na.infn.it

The user interface is a crucial service to guarantee the Grid accessibility. The goal to achieve, is the implementation of an environment able to hide the grid complexity and offer a familiar interface to the final user.

Currently many graphical interfaces have been proposed to simplify the grid access, but the GUI approach appears not very congenial to UNIX developers and users accustomed to work with command line interface.

In 2004 the GridShell project proposed an extension of popular UNIX shells such as TCSH and BASH with features supporting Grid computing.

Starting from the ideas included in GridShell, we propose IGSH (INFN-GRID SHELL) a new login shell for the INFN-GRID middleware, that interact with the Resource Broker services and integrates in a "naturally way" the grid functionality with a familiar interface.

The architecture of IGSH is very simple, it consist of a software layer on the top of the INFN-GRID middleware layer. When some operation is performed by the user, IGSH takes in charge to parse the syntax and translate it in the correspondents INFN-GRID commands according to some semantic rules specified in the next sections.

The final user interacts with the underlying distributed infrastructure by using IGSH instead of his default login shell, with the sensation to work on a local machine.

Moreover IGSH shows interesting potentialities, by allowing the user to create complex workflow by using the standard shell language.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

Virgo **Summary:**

The paper is structured as follows.

- the architecture of INFN-GRID shell
- jobs submission and data management facilities with IGSH.
- a first implementation with some use case.
- workflow

Grid middleware and tools / 216

A Quantitative Comparison Test of Workload Management Systems

Authors: Burt Holzman¹ ; Igor Sfligoi¹

¹ *FNAL*

Corresponding Author: burt@fnal.gov

The advent of the Grids have made it possible for any user to run hundreds of thousands of jobs in a matter of days.

However, the batch slots are not organized in a common pool, but are instead grouped in independent pools at hundreds of Grid sites distributed among the five continents.

A higher level Workload Management System (WMS) that aggregates resources from many sites is thus necessary.

There are several ways to design and implement a WMS; the purpose of this project it to show how some of the most commonly used WMS-es (including gLite WMS, ReSS, and glideinWMS) behave under realistic load conditions.

The results presented have been measured using the same tools for all the tested WMS-es, comparing those results against a baseline obtained by using plain Condor-G submissions. Tests were performed at various load levels and with different payloads to test for scalability and reliability issues.

Poster 1 - Board: 104 / 375

A Scientific Overview of Network Connectivity and Grid Infrastructure in South Asian Countries

Author: Shahryar Khan¹

Co-authors: Arshad Ali² ; Les Cottrell¹ ; Umar Kalim²

¹ *Stanford Linear Accelerator Center*

² *NUST Institute of Information Technology*

Corresponding Author: shahryar@slac.stanford.edu

The future of Computing in High Energy Physics (HEP) applications depends on both the Network and Grid infrastructure. Some South Asian countries such as India and Pakistan are making progress in this direction by not only building Grid clusters, but also by improving their network infrastructure. However to facilitate the use of these resources, they need to overcome the issues of network connectivity to be among the leading participants in Computing for HEP

experiments. In this paper we classify the connectivity for academic and research institutions of South Asia. The quantitative measurements are carried out using the PingER methodology; an approach that induces minimal ICMP traffic to gather end-to-end network statistics. The PingER project has been measuring the Internet performance for the last decade. Currently the measurement infrastructure comprises of over 700 hosts in more than 130 countries which collectively represents approximately 99% of the world's Internet-connected population. Thus, we are well positioned to characterize the world's connectivity. Here we present the current state of the National Research and Educational Networks (NRENs) and Grid Infrastructure in the South Asian countries and identify the areas of concern. We also present comparisons between South Asia and other developing as well as developed regions. We show that there is a strong correlation between the Network performance and several Human Development indices.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

SLAC Networking Group

Poster 1 - Board: 37 / 162

A Search Engine for the Engineering and Equipment Data Management System (EDMS) at CERN

Author: Andrey Tsyganov¹

Co-authors: Anna SUWALSKA²; Sonia MALLON AMERIGO²; Stephan PETIT²; Thomas PETERSSON²

¹ *Moscow Physical Engineering Inst. (MePhI)*

² *CERN*

Corresponding Author: andrey.tsyganov@cern.ch

CERN, the European Laboratory for Particle Physics, located in Geneva - Switzerland, is currently building the LHC, a 27 km particle accelerator. The equipment life-cycle management of this project is provided by the Engineering and Equipment Data Management System (EDMS) Service. Using Oracle, it supports the management and follow-up of different kinds of documentation through the whole life cycle of the LHC project: design, manufacturing, installation, commissioning data etc...

The equipment data collection phase is now slowing down and the project is getting closer to the "As-Built" phase; the phase of the project consuming and exploring the large volumes of data stored since 1996.

Searching through millions of pieces of information (documents, pieces of equipment, operations...) multiplied by dozens of points of view (operators, maintainers...)

require an efficient and flexible search engine. This paper describes the process followed by the team to implement the search engine for LHC As-built project in the EDMS Service. The emphasis is put on the design decision to decouple the search engine from any user interface, potentially enabling other systems to also use it.

Projections, algorithms, and the planned implementation are described in this paper. The implementation of the first version started in early 2007.

Poster 1 - Board: 75 / 154

A distributed approach for a regional grid operation centre

Author: Sven Hermann¹

Co-authors: Clemens Koerdt¹; Holger Marten¹; Jos van Wezel¹

¹ *Forschungszentrum Karlsruhe*

Corresponding Author: sven.hermann@iwr.fzk.de

Forschungszentrum Karlsruhe is one of the largest science and engineering research institutions in Europe. The resource centre GridKa as part of this science centre is building up a Tier 1 centre for the LHC project. Embedded in the European grid initiative EGEE, GridKa also manages the ROC (regional operation centre) for the German Swiss region. A ROC is responsible for regional coordination, organisation and support of the sustainable operations grid infrastructure in EGEE. A particularity of this specific ROC is its distributed organisational structure. The German Swiss operations community, currently consisting of 14 resource centres, is supported and organised by a team of 6 partners. On the one hand a consequence of this decentral approach is a larger effort for organisation and management. On the other hand grid knowledge and expertise is naturally spread among partners with various backgrounds. As a consequence grid technology is being developed and deployed with a wider usability over more use cases and science communities, then it would be in a central approach. Different aspects of this regional specific organisation structures as part of the global EGEE grid operations as well as roadmap for upcoming ROC services and tasks will be highlighted. The successful integration of grid tools for operations and support leads to a sustainable structure of the German Swiss region within a world wide grid community.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

EGEE

Poster 2 - Board: 16 / 224

A multi-dimensional view on information retrieval of CMS data

Author: Valentin Kuznetsov¹

Co-authors: Andrew Dolgert¹ ; Chris Jones¹ ; Dan Riley¹

¹ *Cornell University*

Corresponding Author: valentin.kuznetsov@cern.ch

The CMS Dataset Bookkeeping System (DBS) search page is a web-based application used by physicists and production managers to find data from the CMS experiment. The main challenge in the design of the system was to map the complex, distributed data model embodied in the DBS and the Data Location Service (DLS) to a simple, intuitive interface consistent with the mental model of physicists analysis the data. We used focus groups and user interviews to establish the required features. The resulting interface addresses the physicist and production manager roles separately, offering both a guided search structured for the common physics use cases as well as a dynamic advanced query interface.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS

Poster 2 - Board: 47 / 175

A new inclusive secondary vertex algorithm for b-jet tagging in ATLAS

Authors: Christian Weiser¹ ; Giacinto Piacquadio¹

¹ *Physikalisches Institut - Albert-Ludwigs-Universität Freiburg*

Corresponding Author: giacinto.piacquadio@cern.ch

A new inclusive secondary vertexing algorithm which exploits the topological structure of weak b- and c-hadron decays inside jets is presented. The primary goal is the application to b-jet tagging. The fragmentation of a b-quark results in a decay chain composed of a secondary vertex from the weakly decaying b-hadron and typically one or more tertiary vertices from c-hadron decays. The decay lengths and charged particle multiplicities involved in these decays, as well as the instrumental resolution, do not allow to separately reconstruct and resolve these vertices efficiently using conventional secondary vertexing algorithms based on the assumption of a common geometrical vertex. These difficulties are partially overcome in the algorithm presented in this paper, that is based on the hypothesis that the primary event vertex and the vertices of the weak b- and c-hadron decays lie on the same line, the flight direction of the b-hadron. The algorithm provides detailed information on the topology of the decay cascade, also allowing the reconstruction of topologies with only one charged particle from a b- and c-hadron decay, respectively, which are difficult to access for conventional algorithms. The algorithm based on this hypothesis is implemented mathematically as an extension of the Kalman Filter formalism for vertex reconstruction and technically as a set of flexible software modules integrated in the ATLAS software framework Athena, which make use of the existing Event Data Model for vertexing and B-Tagging. The application of the algorithm to b-jet tagging and the impact on its performance is shown.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Poster 2 - Board: 43 / 132

A novel design approach and new Geant4 physics developments for microdosimetry simulation

Authors: Maria Grazia Pia¹ ; Petteri Nieminen² ; Sebastien Incerti³

¹ *INFN GENOVA*

² *ESA*

³ *CENBG-IN2P3*

Corresponding Authors: incerti@cenbg.in2p3.fr, maria.grazia.pia@cern.ch

Detailed knowledge of the microscopic pattern of energy deposition related to the particle track structure is required to study radiation effects in various domains, like electronics, gaseous detectors or biological systems.

The extension of Geant4 physics down to the electronvolt scale requires not only new physics models, but also adequate design technology. For this purpose a novel approach, based on a policy-based class design, has been explored: the usage of this design technique represents an innovative design within Geant4, and more in general for Monte Carlo simulation for particle physics.

A policy-based design assembles classes with complex functionality out of simpler ones responsible of a single behavioural or structural aspect. Policies define a class interface or a class template interface; they are more loosely defined than conventional abstract interfaces, as they are syntax oriented rather than signature oriented.

A policy-based design is highly customisable; this feature is relevant to cases where a variety of physics models is desirable. It also provides advantages in performance with respect to more conventional techniques for interchangeable algorithms, such as a Strategy pattern: in fact, policies are compile-time bound and are exempt from the

drawbacks related to the virtual method table.

The novel design and a set of new physics models introduced in Geant4 are presented; this prototype demonstrates the feasibility of the new approach. The new design technique is suitable to be extended to other Geant4 physics domains to improve the flexibility of modelling configuration and the execution performance.

Poster 2 - Board: 56 / 273

A pre-identification for electron reconstruction in the CMS particle-flow algorithm.

Author: Michele Pioppi¹

¹ CERN

Corresponding Author: michele.pioppi@cern.ch

In the CMS software, a dedicated electron track reconstruction algorithm, based on a Gaussian Sum Filter (GSF), is used. This algorithm is able to follow an electron along its complete path up to the electromagnetic calorimeter, even in the case of a large amount of Bremsstrahlung emission. Because of the significant CPU consumption of this algorithm, however, it can be run only on a limited number of electron candidates. The standard GSF electron track reconstruction is triggered by the presence of high energy isolated electromagnetic clusters. Instead, a pre-identification algorithm based on both the tracker and the calorimeter has recently been developed. It allows electron tracks to be reconstructed within jets with a good efficiency even for small electron transverse momentum. This algorithm as well as its performance in terms of efficiency, mis-identification probability and timing are presented. Its implementation within the particle-flow algorithm is also described.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS

Poster 2 - Board: 29 / 393

A software and computing prototype for CMS Muon System alignment

Author: Francisco Matorras¹

Co-authors: Andre Sznajder²; Ibán Cabrillo¹; Isidro Gonzalez¹; Javier Fernandez¹; Pablo Martinez¹; Rafael Marco¹; Rebeca Gonzalez³

¹ Insitituto de Física de Cantabria

² Instituto de Fisica Univ. do Estado do Rio de Janeiro

³ Instituto de Fisica de Cantabria

Corresponding Author: francisco.matorras@cern.ch

A precise alignment of Muon System is one of the requirements to fulfill the CMS expected performance to cover its physics program. A first prototype of the software and computing tools to achieve this goal has been successfully tested during the CSA06, Computing, Software and Analysis Challenge in 2006. Data was exported from Tier-0 to Tier-1 and Tier-2, where the alignment software was run . Re-reconstruction with new geometry files was also performed at remote sites.

Performance and validation of the software has also been tested on cosmic data, taken during the MTCC in 2006, Magnet Test Cosmic Challenge.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS

Online computing / 112

A software framework for Data Quality Monitoring in ATLAS

Authors: Alina Corso-Radu¹ ; Haleh Hadavand² ; Michael Hauschild³ ; Robert Kehoe² ; Serguei Kolos¹

¹ *University of California Irvine*

² *Southern Methodist University*

³ *CERN*

Corresponding Author: serguei.kolos@cern.ch

Data Quality Monitoring (DQM) is an important and integral part of the data taking and data reconstruction of HEP experiments. In an online environment, DQM provides the shift crew with live information beyond basic monitoring. This is used to overcome problems promptly and help avoid taking faulty data. During the off-line reconstruction DQM is used for more complex analysis of physics quantities and its results are used to assess the quality of the reconstructed data.

The Data Quality Monitoring software Framework (DQMF) which has been provided for the ATLAS experiment performs analysis of monitoring data through user defined algorithms and relays the summary of the analysis results to the configurable Data Quality output stream. From this stream the results can be stored to a database, displayed on a GUI, or used to make some other relevant actions with respect to the operational environment ie sending alarms, stopping the run. This paper describes the implementation of the DQMF and discusses experience from usage and performance of the DQMF during ATLAS commissioning.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Poster 1 - Board: 77 / 160

A study of the accuracy of Network Time Protocol client synchronization in large computing clusters

Author: Nick Garfield¹

Co-author: Vlado Bahyl¹

¹ *CERN*

Corresponding Author: nicholas.garfield@cern.ch

As computing systems become more distributed and as networks increase in throughput and resources become ever increasingly dispersed over multiple administrative domains, even continents, there is a greater need to know the performance limits of the underlying protocols which make the foundations of complex computing and networking architectures. One such protocol is the Network Time Protocol (NTP) which is often overlooked as an important part of

any large scale computing system. With the adoption of new highly distributed technologies, such as those employed in grid computing, the increasing number of users and resources will test not only the synchronization of these resources but also the transaction logging and event correlation in any problem resolution/diagnostic systems. In essence, good quality and reliable time synchronization is a key component to the actual operation of any large scale production system incorporating many components. In this paper we present the CERN NTP server and client architecture and discuss the statistical quality of time synchronization of 4 computing clusters of increasing size from approximately 50 to 3000 nodes and inter-connected via a high-performance 10Gbit/s symmetrically routed network backbone infrastructure. Each cluster is dedicated to a specific task or application resulting in various IO load profiles, some more deterministic than others. The relationship between the reliability of time synchronization, system load and network IO is analysed and optimization suggestions are presented.

Poster 2 - Board: 75 / 151

ALICE - ARC integration

Authors: Csaba Anderlik¹ ; Josva Kleist¹

Co-authors: Andreas Peters² ; Pablo Siaz²

¹ *Nordic Data Grid Facility*

² *CERN*

Corresponding Author: kleist@ndgf.org

AliEn or Alice Environment is the Gridware developed and used within the ALICE collaboration for storing and processing data in a distributed manner. ARC (Advanced Resource Connector) is the Grid middleware deployed across the Nordic countries and gluing together the resources within the Nordic Data Grid Facility (NDGF). In this paper we will present our approach to integrate AliEn and ARC, in the sense that ALICE data management and job processing can be carried out on the NDGF infrastructure, using the client tools available in AliEn. The interoperation has two aspects, one is the data management part and the second the job management aspect. The first aspect was solved by using dCache across NDGF to handle data. dCache provides support for several data management tools (among them for xrootd the tools used by AliEn) using the so called “doors”. Therefore, we will concentrate on the second part. Solving it, was somewhat cumbersome, mainly due to the different computing models employed by AliEn and ARC. AliEn uses an Agent based pull model while ARC handles jobs through the more “traditional” push model. The solution comes as a module implementing the functionalities necessary to achieve AliEn job submission and management to ARC enabled sites.

Poster 1 - Board: 1 / 53

ALICE DAQ Online Transient Data Storage

Author: Ulrich Fuchs¹

Co-authors: Csaba Soós² ; Franco Carena² ; Irina Makhlyueva² ; Klaus Schossmair² ; Pierre Vande Vyvre² ; Roberto Divià² ; Sylvain Chapeland² ; Wisla Carena²

¹ *CERN & Ludwig-Maximilians-Universitat Munchen*

² *CERN*

Corresponding Author: ulrich.fuchs@cern.ch

ALICE is a dedicated heavy-ion detector to exploit the physics potential of nucleus-nucleus (lead-lead) interactions at LHC energies. The aim is to study the physics of strongly interacting matter at extreme energy densities, where the formation of a new phase of matter, the quark-gluon plasma, is expected.

Running in heavy-ion mode the data rate from event building to permanent storage is expected to be around 1.25 GB/s. To continue data recording even in the event of hardware failure or connection problems, a large disk pool has been installed at the experiment's site as buffering layer between the DAQ and the remote (~5km) tape facility in the CERN Computing Centre. This Transient Data Storage (TDS) disk pool has to provide the bandwidth to be able to simultaneously absorb data from the event building machines and to move data to the tape facility. The aggregated bandwidth of the TDS is expected to exceed 3 GB/s in mixed I/O traffic.

Extensive tests have been carried out on various hardware and software solutions with the goal to build a common file space shared by ~60 clients, whilst still providing maximum bandwidth per client (~400MB/s, 4Gbps Fibre Channel), fail-over safety and redundancy.

This talk will present the chosen hardware and software solution, the configuration of the TDS pool and the various modes of operation in the ALICE DAQ framework. It will also present the results of the performance tests carried out during the last ALICE Data Challenge.

Distributed data analysis and information management / 51

ASAP is a system for enabling distributed analysis for the CMS Experiment

Author: Akram Khan¹

Co-author: Craig Munro¹

¹ *Brunel University*

Corresponding Author: akram@slac.stanford.edu

ASAP is a system for enabling distributed analysis for CMS physicists. It was created with the aim of simplifying the transition from a locally running application to one that is distributed across the Grid. The experience gained in operating the system for the past 2 years has been used to redevelop a more robust, performant and scalable version. ASAP consists of a client for job creation, control and monitoring and an optional server side component. Once jobs are delegated to the server it will submit, update, fetch or resubmit the job on behalf of the user. ASAP is able to make decisions on the success of the users job and will resubmit if either a grid or application failure is detected. An advanced mode allows running jobs to communicate directly with the server in order to request additional jobs and to set the status of the job directly. These features reduce the turnaround time experienced by the user and increase the likelihood of success.

Software components, tools and databases / 83

ATLAS Analysis Model

Author: Amir Farbin¹

Co-authors: Ketevi Assamagan²; Kyle Cranmer²

¹ *European Organization for Nuclear Research (CERN)*

² *Brookhaven National Lab*

Corresponding Author: amir.farbin@cern.ch

As we near the collection of the first data from the Large Hadron Collider, the ATLAS collaboration is preparing the software and computing infrastructure to allow quick analysis of the first data and support of the long-term steady-state ATLAS physics program. As part of this effort considerable attention has been paid to the "Analysis Model", a vision of the interplay of the software design, computing constraints, and various physics requirements. An important input to this activity has been the experience of Tevatron and B-Factory experiments, one topic which was explored discussed in the ATLAS October 2006 Analysis Model workshop. Recently, much of the Analysis Model has focused on ensuring the ATLAS software framework supports the required manipulations of event data; the event data design and content is consistent with foreseen calibration and physics analysis tasks; the event data is optimized in size, access speed, and is accessible both inside and outside the software framework; and that the analysis software may be developed collaboratively.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS Offline Computing **Summary:**

In this talk, I will overview the lessons learned from various High Energy Experiments and the resulting actions within the ATLAS software.

Poster 1 - Board: 78 / 164

ATLAS Computing System Commissioning - Simulation Production Experience

Author: Kaushik De¹

¹ *UT-Arlington*

Corresponding Author: kaushik@uta.edu

During 2006-07, the ATLAS experiment at the Large Hadron Collider launched a massive Monte Carlo simulation production exercise to commission software and computing systems in preparation for data in 2007. In this talk, we will describe the goals and objectives of this exercise, the software systems used, and the tiered computing infrastructure deployed worldwide. More than half a petabyte of data was generated at more than 50 different sites. The results of this year-long exercise will be summarized, with special emphasis on the lessons learned from an international distributed computing exercise of unprecedented size and scope.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Poster 1 - Board: 28 / 90

ATLAS Conditions Database Experience with the LCG COOL Conditions Database Project

Author: Monica Verducci¹

¹ *European Organization for Nuclear Research (CERN)*

Corresponding Author: monica.verducci@cern.ch

One of the most challenging task faced by the LHC experiments will be the storage of “non-event data” produced by calibration and alignment stream processes into the Conditions Database. For the handling of this complex experiment conditions data the LCG Conditions Database Project has implemented COOL, a new software product designed to minimise the duplication of effort by developing a single implementation to support persistency for several relational technologies (Oracle,MySQL and SQLite). After several production releases of the COOL software, the project has be moved into the deployment phase in Atlas and LHCb, the two experiments that are developing the software in collaboration with CERN IT. In particular, the ATLAS Conditions Database, accessed by the offline reconstruction framework (ATHENA), is implemented using this COOL technology. The objects, stored or referenced in the COOL tables, have an associated start and end time (run or event number or absolute time-stamp) between they are valid (Interval of Validity). The storage and retrieving of data during a recontruction job inside ATHENA is guaranteed by the IOVService, a software interface between the COOL DB and the reconstruction algorithms. This work describes some practical examples already successfully tested in ATLAS and the further extensively tests of the entire chain foreseen during the next Computing and Detector Commissioning in 2007.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Poster 2 - Board: 74 / 149

ATLAS DDM Integration in ARC

Authors: Adrian Taga¹ ; David Cameron² ; Gerd Behrmann² ; Josva Kleist² ; Mattias Ellert²

¹ *Oslo University*

² *Nordic Data Grid Facility*

Corresponding Author: kleist@ndgf.org

The Nordic Data Grid Facility (NDGF) consists of Grid resources running ARC middleware in Scandinavia and other countries. These resources serve many virtual organisations and contribute a large fraction of total worldwide resources for the ATLAS experiment, whose data is distributed and managed by the DQ2 software. Managing ATLAS data within NDGF and between NDGF and other Grids used by ATLAS (the LHC Computing Grid and the Open Science Grid) presents a unique challenge for several reasons. Firstly, the entry point for data, the Tier 1 centre, is physically distributed among heterogeneous resources in several countries and yet must present a single access point for all data stored within the centre. The middleware framework used in NDGF differs significantly from other Grids, specifically in the way that all data movement and registration is performed by services outside the worker node environment. Also, the service used for cataloging the location of data files is different from other Grids but must still be useable by DQ2 and ATLAS users to locate data within NDGF. This paper presents in detail how we solve these issues to allow seamless access worldwide to data within NDGF.

Poster 1 - Board: 24 / 44

ATLAS Detector Maintenance And Operation Management Tool

Author: Brice Copy¹

Co-author: Maksim Tsikanin¹

¹ CERN

Corresponding Author: brice.copy@cern.ch

The maintenance and operation of the ATLAS detector will involve thousands of contributors from 170 physics institutes. Planning and coordinating the action of ATLAS members, ensuring their expertise is properly leveraged and that no parts of the detector are under or overstuffed will be a challenging task.

The ATLAS Maintenance and Operation (ATLAS M&O) application offers a fluent web based interface that combines the flexibility and comfort of a desktop application, intuitive data visualization and navigation techniques, with a lightweight service oriented architecture. We will review the application, its usage within the ATLAS experiment, its underlying design and the software project management techniques employed to complete the project.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Computer facilities, production grids and networking / 84

ATLAS Distributed Data Management Operations. Experience and Projection

Author: Alexei Klimentov¹

Co-authors: Pavel Nevski¹; Stephane Jezequel²

¹ BNL

² LAPP

Corresponding Author: alexei.klimentov@cern.ch

ATLAS Distributed Data Management Operations Team unites experts from Tier-1s and Tier-2s computer centers. The group is responsible for all day by day ATLAS data distribution between different sites and centers.

In our paper we describe ATLAS DDM operation model and address the data management and operation issues. A serie of Functional Tests have been conducted in the past and is in progress now. We will decrbe the problems we met and improvements we made during this work.

The system performance, reliability and robustness in a large scale operation are described.

We also address the issue of DDM and ATLAS production system integration and how ATLAS computing requirements for data sharing between sites are implemented by DDM oprations

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Poster 2 - Board: 40 / 98

ATLAS Liquid Argon Calorimeter Reconstruction Software and Commissioning

Author: Hong Ma¹

¹ Brookhaven National Laboratory (BNL)

Corresponding Authors: rolf.seuster@cern.ch, hong.ma@cern.ch

The ATLAS Liquid Argon Calorimeter consists of precision electromagnetic accordion calorimeters in the barrel and endcaps, hadronic calorimeters in the endcaps, and calorimeters in the forward region.

The initial high energy collision data at the LHC experiments is expected in the spring of 2008. While tools for the reconstruction of the calorimeter data are quite developed through years of Monte Carlo simulation and test beam studies, the processing, storing and archiving of all peripheral meta-data such as calibration constants, detailed conditions of the sub-detector systems are actively being worked on. The LAr calorimeter consists of over 180,000 electronic channels. The reconstruction of all the signal channels with proper calibration and other conditions is challenging.

The current status of these efforts for the ATLAS Liquid Argon Calorimeter is presented. The interfaces to access, and techniques to store, the conditions data are introduced. The current effort in commissioning the detector gives invaluable experience in larger scale application of these methods and is discussed in detail.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Poster 2 - Board: 36 / 33

ATLAS Muon Spectrometer Simulation and its Validation Algorithms

Authors: Daniela Rebuffi¹ ; Nectarios Benekos²

Co-authors: Georgios Dedes² ; Lashkar Kashif³ ; Matthias Schott⁴

¹ INFN, Sezione di Pavia

² Max-Planck-Institut für Physik

³ Harvard University

⁴ Ludwig-Maximilians-Univ. Muenchen

Corresponding Authors: nectarios.benekos@cern.ch, daniela.rebuffi@pv.infn.it

The ATLAS detector, currently being installed at CERN, is designed to make precise measurements of 14 TeV proton-proton collisions at the LHC, starting in 2007. Arguably the clearest signatures for new physics, including the Higgs Boson and supersymmetry, will involve the production of isolated final-state muons. The identification and precise reconstruction of muons are performed using a combination of detector components, including an inner detector, comprising a silicon tracker, pixel detector, and transition radiation tracker, housed in a uniform solenoidal field, and a precision standalone Muon Spectrometer, comprising monitored drift tubes and cathode strip chambers,

triggered by resistive plate chambers and thin-gap chambers, and housed in a toroidal field.

In order to manage the complexity and to understand the performance of the ATLAS MuonSpectrometer, a detailed full detector simulation is required and it should be kept under control by means of automatic validation procedures.

We describe the implementation and the functionalities of the recently developed MuonValidation package, which has been developed as a dedicated tool to monitor and validate the performance of the Full Simulation and Digitization of the Muon System. Its flexible design allows comparisons between different Muon geometrical layouts and different software releases.

Validation results based on fully simulated GEANT4 events, using the complete detailed geometrical description of the detector are shown.

Computer facilities, production grids and networking / 184

ATLAS Production Experience on OSG Infrastructure

Author: Smirnov Yuri¹

¹ *Brookhaven National Laboratory*

Corresponding Author: smirnov@hep.uchicago.edu

The Open Science Grid infrastructure provides one of the largest distributed computing systems deployed in the ATLAS experiment at the LHC. During the CSC exercise in 2006-2007, OSG resources provided about one third of the worldwide distributed computing resources available in ATLAS. About half a petabyte of ATLAS MC data is stored on OSG sites. About 2000k SpecInt2000 CPU's is available. In this talk, we will describe the software systems used, and the operational experience gained during one and a half years of continuous production for ATLAS on OSG resources.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Online computing / 209

ATLAS Tile Calorimeter Cesium calibration control and analysis software

Authors: A Isaev¹ ; A Karyukhin¹ ; A Solodkov¹ ; E Starchenko¹ ; N Shalanda¹ ; O Solovyanov¹

¹ *IHEP, Protvino, Russia*

Corresponding Author: oleg.solovianov@cern.ch

An online control system to calibrate and monitor ATLAS Barrel hadronic calorimeter (TileCal) with a movable radioactive source, driven by liquid flow, is described.

To read out and control the system an online software has been developed, using ATLAS TDAQ components like DVS (Diagnostic and Verification System) to verify the HW before running, IS (Information Server) for data and status exchange between networked computers, and other components like DDC, to connect to PVSS-based slow control systems of Tile Calorimeter, like high voltage and low voltage.

A system of scripting facilities, based on Python language, is used to handle all the calibration and monitoring processes from hardware perspective to final data storage, including various abnormal situations.

A QT based graphical user interface to display the status of the calibration system during the cesium source scan is described.

The software for analysis of the detector response, using online data, is discussed.

Performance of the system and first results from the pit are presented.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

Tile Collaboration of ATLAS experiment

Event processing / 418

ATLAS Tile Calorimeter Data Quality assessment with commissioning data

Authors: Alexander Solodkov¹ ; Andrea Dotti² ; Carlos Solans Sanchez³ ; Fernando Guimaraes Ferreira⁴ ; Jalal Mohamad Abdallah³ ; Luca Fiorini⁵ ; Tomas Davidek⁶ ; Ulrike Blumenschein⁵

¹ *Institute for High Energy Physics (IHEP)*

² *Università and INFN Pisa*

³ *Instituto de Fisica Corpuscular (IFIC) Universidad de Valencia*

⁴ *Instituto de Fisica Univ. Federal do Rio de Janeiro (UFRJ)*

⁵ *Inst. de Fisica de Altas Energias (IFAE) Universidad Autonoma de Barcellona*

⁶ *Institute of Particle & Nucl. Phys. Faculty of Mathematics and Physics Charles University Prague*

Corresponding Author: andrea.dotti@pi.infn.it

The Tile Calorimeter (TileCal) is the central hadronic calorimeter of the ATLAS experiment presently in an advanced state of installation and commissioning at the LHC accelerator.

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 cosmic muon data.

To assess the detector status and verify its performance a set of tools have been developed spanning from the hardware detector verification tests and the online monitoring to the offline reconstruction. Tools allowing for a fast and partly automated analysis of the result have been also developed. The system is completed with web interfaces to allow for remote monitoring and data quality assessment.

This set of tools is the prototype of the final TileCal data quality system that is under development and it is highly integrated with all ATLAS online and offline frameworks.

A review of the TileCal data quality system, current developments and the future foreseen upgrades will be presented together with a selection of results.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS Tile Calorimeter

Poster 1 - Board: 35 / 142

ATLAS Tile Calorimeter on-line monitoring system based on the Event Filter

Authors: Luca Fiorini¹ ; Nils Gollub²

¹ *IFAE Barcelona*

² *University of Uppsala*

Corresponding Authors: nils.gollub@cern.ch, nils.gollub@tsl.uu.se, luca.fiorini@cern.ch

ATLAS Tile Calorimeter detector (TileCal) is presently involved in an intense phase of commissioning with cosmic rays and subsystems integration. Various monitoring programs have been developed at different level of the data flow to tune the set-up of the detector running conditions and to provide a fast and reliable assessment of the data quality.

The presentation will focus on the on-line monitoring tools employed during TileCal detector commissioning and integration tests with cosmic rays and in particular on the monitoring system integrated in the highest level of the Trigger, the Event Filter (EF).

The key feature of EF monitoring is the capability of performing detector and data quality control on the complete physics event at the trigger level, hence before it is stored on disk. In the on-line data flow, this is the only monitoring system in ATLAS capable of giving a comprehensive event quality feedback.

The presentation will also show some monitoring results of the integration tests with other sub-detectors and performances and future upgrades of the current implementation.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

Atlas TileCal project

Poster 1 - Board: 27 / 82

ATLAS's EventView Analysis Framework

Author: Amir Farbin¹

Co-authors: Akira Shibata² ; Kyle Cranmer³

¹ *European Organization for Nuclear Research (CERN)*

² *Queen Mary University*

³ *Brookhaven National Lab*

Corresponding Author: amir.farbin@cern.ch

The EventView Analysis Framework is currently the basis for much of the analysis software employed by various ATLAS physics groups (for example the Top, SUSY, Higgs, and Exotics working groups). In ATLAS's central data preparation, this framework provides an assessment of data quality and the first analysis of physics data for the whole collaboration. An EventView is a self-consistent interpretation of a physics event or equivalently the state of a specific analysis. Analyses are constructed at runtime by chaining and configuring modular components consisting of tools, C++ implementation of specific analysis algorithms, and modules, python grouping and configuration of various tool. A large common library of general tools and modules serve as the building blocks of nearly all of the

steps of any analysis. The output is multiple simultaneous EventViews of every event, typically reflecting different choices of selections, reconstruction algorithms, combinatoric assignments, or input data (eg full or fast reconstruction or truth).

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS Offline Computing **Summary:**

In this talk I will motivate the EventView concept through various physics examples and then detail the design and implementation of this now widely-used analysis framework.

Plenary / 489

Addressing Future HPC Demand with Multi-core Processors

Dozens of cores will not be a dream. Multiple processor cores drive energy efficient performance for highly parallel applications. However, looking beyond cores, achieving balanced high performance throughput has many challenges. Intel Senior Fellow and CTO of Digital Enterprise Group Steve Pawlowski will provide his technology vision to address bandwidth, capacity and power needs on memory, I/O, intra-chip and inter-chip interconnections and outlook on future reliability challenges.

Summary:

Stephen S. Pawlowski is an Intel Senior Fellow. He is the Digital Enterprise Group chief technology officer and general manager for Architecture and Planning for Intel Corporation.

Pawlowski joined Intel in 1982. He led the design of the first Multibus I Single Board Computer based on the 386 processor. He was a lead architect and designer for Intel's early desktop PC and high performance server products and was the co-architect for Intel's first P6 based server chipsets. He helped define the system bus interfaces for Intel's P6 family processors, the Pentium® 4 processor and Itanium™ processor. He also created and led the research for Intel's agile radio architecture for a future generation of wireless products and prior to his current assignment was the director of Corporate Technology Group's Microprocessor Technology Lab.

Pawlowski graduated from the Oregon Institute of Technology in 1982 with bachelor's degrees in electrical engineering technology and computer systems engineering technology, and received a master's degree in computer science and engineering from the Oregon Graduate Institute in 1993.

Pawlowski holds 56 patents in the area of system, and microprocessor technologies. He has received three Intel Achievement Awards.

Computer facilities, production grids and networking / 92

Addressing the Pilot Security Problem With gLExec

Authors: Dan Yocum¹ ; Gerben Venekamp² ; Igor Sfiligoi¹ ; Oscar Koeroo²

Co-authors: David Groep² ; Don Petravick¹

¹ *Fermilab*

² *NIKHEF*

Corresponding Author: yocum@fnal.gov

Pilot jobs are becoming increasingly popular in the Grid world. Experiments like ATLAS and CDF are using them in production, while others, like CMS, are actively evaluating them.

Pilot jobs enter Grid sites using a generic pilot credential, and once on a worker node, call home to fetch the job of an actual user.

However, this operation mode poses several new security problems when used in the traditional Grid environment:

- Executing the code of another user without authenticating and authorizing the end user violates the security policies of any site that requires full knowledge and control of all users of its resources.
- All processes run under the same UID, allowing a malicious user to steal the credentials of both the pilot and potentially any other user handled by the same pilot infrastructure.

To solve this problem, a site-trusted, and necessarily setuid utility is needed to authorize the end user and switch to the correct local UID.

gLExec is a Grid-aware suexec derivative, developed for EGEE by the NIKHEF group. Recently it has been integrated with the distributed OSG security infrastructure making it easy to deploy on OSG worker nodes.

The initial OSG deployment of gLExec on worker nodes has been completed at Fermilab and the CDF and CMS experiments have been actively using it for several months.

An architectural overview and the experience gathered will be presented.

Computer facilities, production grids and networking / 72

Advances in Integrated Storage, Transfer and Network Management

Authors: Alan Tacket¹ ; Conrad Steenberg² ; Daniel Engh¹ ; Dimitri Bourilkov³ ; Frank van Lingen² ; Harvey Newman² ; Ilya Narsky² ; Iosif LeGrand² ; Julian Bunn² ; Michael Thomas² ; Paul Avery³ ; Paul Sheldon¹ ; Rick Cavanaugh³ ; Shawn McKee⁴

Co-authors: Alberto Santoro⁵ ; Andy Adamson⁴ ; Azher Amin⁶ ; Chip Chapman² ; Dan Nae² ; Don Petravick⁷ ; Dongchul Son⁸ ; Edoardo Martelli⁹ ; Ernesto Rubi¹⁰ ; Frank Burstein¹¹ ; John Bigrow¹¹ ; Julio Ibarra¹⁰ ; Matt Crawford⁷ ; Minsun Lee¹² ; Ramiro Voicu⁹ ; Scott Bradley¹¹ ; Sergio Novaes¹³ ; Steven Low² ; Yang Xia²

¹ *Vanderbilt University*

² *CALIFORNIA INSTITUTE OF TECHNOLOGY*

³ *University of Florida*

⁴ *University of Michigan*

⁵ *Universidade do Estado do Rio de Janeiro*

⁶ *National University of Science and Technology*

⁷ *Fermi National Laboratory*

⁸ *Kyungpook National University*

⁹ *CERN*

¹⁰ *Florida International University*

¹¹ *Brookhaven National Laboratory*

¹² *Korea Institute of Science and Technology Information*

¹³ São Paulo State University

Corresponding Authors: paul.ralph.avery@cern.ch, fvlingen@caltech.edu

UltraLight is a collaboration of experimental physicists and network engineers whose purpose is to provide the network advances required to enable and facilitate petabyte-scale analysis of globally distributed data. Existing Grid-based infrastructures provide massive computing and storage resources, but are currently limited by their treatment of the network as an external, passive, and largely unmanaged resource. This paper will give an overview of the recent advances made within the UltraLight collaboration over the last 3 years within the different work areas of the project which include: the UltraLight testbed, transportation layer (FAST TCP and MAX net), transfer applications (FDT), network aware command and control systems (VINCI), network centric storage clouds (LStore), and physics applications (data streaming and distributed analysis). Several of the technologies developed within the UltraLight project are currently being deployed and field tested to support efficient transfer of data and prepare for LHC startup. The core of the tools rely on globally distributed publish and subscribe infrastructures and end-to-end monitoring, to prevent single points of failure, increase robustness and improve scalability.

Grid middleware and tools / 431

Advances in Monitoring of Grid Services in WLCG

Author: James Casey¹

Co-author: Ian Neilson ¹

¹ CERN

Corresponding Author: james.casey@cern.ch

During 2006, the Worldwide LHC Computing Grid Project (WLCG) constituted several working groups in the area of fabric and application monitoring with the mandate of improving the reliability and availability of the grid infrastructure through improved monitoring of the grid fabric.

This talk will discuss the 'Grid Service Monitoring' Working Group. This has the aim to evaluate the existing monitoring system and create a coherent architecture that would let the existing system run, while increasing the quality and quantity of monitoring information gathered.

We will describe in detail the stakeholders in this project, and focus in particular on the needs of the site administrators, which were not well satisfied by existing solutions.

Several standards for service metric gathering and grid monitoring data exchange, and the place of each in the architecture will be shown.

Finally we will describe the use of a Nagios-based prototype deployment for validation of our ideas, and the progress on turning this prototype into a production-ready system.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

WLCG

Grid middleware and tools / 443

AliEn2: the ALICE grid Environment

Authors: Andreas Peters¹ ; Catalin Cistoiu¹ ; Costin Grigoras¹ ; Federico Carminati¹ ; Fons Rademakers¹ ; Latchezar Betev¹ ; Pablo Saiz¹ ; Patricia mendez Lorenzo¹ ; Predrag Buncic¹ ; Stefano Bagnasco²

¹ CERN

² INFN Torino

Corresponding Authors: pablo.saiz@cern.ch, federico.carminati@cern.ch

Starting from the end of this year, the ALICE detector will collect data at a rate that, after two years, will reach 4PB per year. To process such a large quantity of data, ALICE has developed over the last seven years a distributed computing environment, called AliEn, integrated in the WLCG environment. The ALICE environment presents several original solutions, which have shown their viability in a number of large exercises of increasing complexity called data challenges. This talk will describe the architecture of the ALICE distributed computing environment, focusing on the challenges to be faced and on the technical solutions chosen to meet them. The job submission system and the data management infrastructure will be described, as well as the user interface. The current status of AliEn will be illustrated, as well as the performance of the system during the data challenges. The presentation will describe also the development roadmap of the system.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ALICE

Poster 1 - Board: 12 / 319

Alignment data streams for the ATLAS Inner Detector.

Authors: Antonio Amorim¹ ; Belmiro Antonio Venda Pinto¹ ; Jochen Schieck² ; Markus Elsing³ ; Paulo Pereira¹ ; Richard Hawkings³ ; Salvador Garcia⁴

Co-authors: Andre Anjos⁵ ; Hong Ma⁶ ; John Baines⁷ ; R.D. Schaffer Schaffer⁸

¹ Faculdade de Ciencias - Universidade de Lisboa

² Max-Planck-Institut für Physik

³ CERN

⁴ Instituto de Física Corpuscular, València

⁵ University of Wisconsin

⁶ Brookhaven National Laboratory (BNL)

⁷ Rutherford Appleton Laboratory

⁸ LAL Orsay, France

Corresponding Author: belmiro.pinto@cern.ch

The ATLAS experiment uses a complex trigger strategy to be able to achieve the necessary Event Filter rate output, making possible to optimize the storage and processing needs of these data. These needs are described in the ATLAS Computing Model which embraces Grid concepts. The output coming from the Event Filter will consist of four main streams: the physical stream, express stream, calibration stream, and a diagnostic stream. The calibration stream will be transferred to the Tier-0 facilities which will provide the prompt reconstruction of this stream with a

minimum latency of 8 hours, producing calibration constants of sufficient quality to permit a first-pass processing. The Inner Detector community is developing and testing an independent common calibration stream selected at the Event Filter after track reconstruction. It is composed of raw data, in byte-stream format, contained in ROB's with hit information of the selected tracks, and it will be used to derive and update a set of calibration and alignment constants after every fill. This option was selected because makes use of the Byte Stream Converter infrastructure and possibly give us a better bandwidth usage and storage capability's. Processing is done using specialized algorithms running in Athena framework in dedicated Tier-0 resources, and the alignment constants will be stored and distributed using the COOL conditions database infrastructure. The work is addressing in particular the alignment requirements, the needs for track and hit selection and the timing issues.

Event processing / 329

Alignment of the CMS Silicon Tracker using MillePede II

Author: Markus Stoye¹

¹ *Inst. f. Experimentalphysik, Universitaet Hamburg*

Corresponding Author: markus.stoye@cern.ch

The CMS silicon tracker comprises about 17000 silicon modules. Its radius and length of 120 cm and 560 cm, respectively, make it the largest silicon tracker ever built. To fully exploit the precise hit measurements, it is necessary to determine the positions and orientations of the silicon modules to the level of μm and mrad , respectively.

Among other track based alignment algorithms, the CMS collaboration studied MillePede II, developed by V. Blobel. This experiment independent program offers several methods to solve the large system of linear equations which arises from a global χ^2 -minimisation.

Studies show that MillePede II is indeed capable to align the about 45000 degrees of freedom of the CMS silicon tracker that have sensible influence on track reconstruction. This result is achieved utilising complementary data sets like muons from Z- or W-decays and cosmic rays, vertex and mass constraints. A hierarchical parametrisation allows to make full use of survey measurements accomplished during construction.

In a realistic case study, all elements of the tracker have been aligned simultaneously. The precision reached is close to 1 μm for the pixel detector and about 20 μm in the endcaps of the strip detector. Remarkably, using the GMRES method to solve the matrix equation takes less than 2 hours on a standard 64-bit PC and requires only 2 GB of memory.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS Offline group

Online computing / 348

Alignment strategy for the CMS tracker

Author: Martin Weber¹

Co-author: Tracker Group CMS ²

¹ *RWTH Aachen, Germany*

² CMS

Corresponding Author: martin.weber@cern.ch

The full-silicon tracker of the CMS experiment with its 15148 strip and 1440 pixel modules is of an unprecedented size. For optimal track-parameter resolution, the position and orientation of its modules need to be determined with a precision of a few micrometer.

Starting from the inclusion of survey measurements, the use of a hardware alignment system, and track based alignment, this talk details the strategy that is used to align the CMS tracker and reports recent results. These include the usage of novel algorithms that allow to solve the optimization problem with the required accuracy in manageable time, the selection of special data samples to constrain weak modes, and the overall layout of the software alignment framework, database model and data flow for alignment.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS Tracker

Poster 1 - Board: 46 / 294

An Inconvenient Truth: file-level metadata and in-file metadata caching in the (file-agnostic) ATLAS distributed event store

Authors: Arthur Schaffer¹ ; David Malon² ; Peter van Gemmeren² ; Richard Hawkings³

¹ LAL Orsay

² Argonne National Laboratory

³ CERN

Corresponding Author: malon@anl.gov

In the ATLAS event store, files are sometimes “an inconvenient truth.” From the point of view of the ATLAS distributed data management system, files are too small—datasets are the units of interest. From the point of view of the ATLAS event store architecture, files are simply a physical clustering optimization: the units of interest are event collections—sets of events that satisfy common conditions or selection predicates—and such collections may or may not have been accumulated into files that contain those events and no others.

It is nonetheless important to maintain file-level metadata, and to cache metadata in event data files. When such metadata may or may not be present in files, or when values may have been updated after files are written and replicated, a clear and transparent model for metadata retrieval from the file itself or from remote databases is required. In this paper we describe how ATLAS reconciles its file and non-file paradigms, the machinery for associating metadata with files and event collections, and the infrastructure for metadata propagation from input to output for provenance record management and related purposes.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Poster 2 - Board: 27 / 365

An SSH Key management system: easing the pain of managing key/user association

Authors: Dmitry ARKHIPKIN¹ ; Jerome Lauret²

Co-authors: Alexander SHIRYAEV¹ ; Wayne BETTS²

¹ *Particle Physics Laboratory - Dubna*

² *BROOKHAVEN NATIONAL LABORATORY*

Corresponding Author: jlauret@bnl.gov

Secure access to computing facilities has been increasingly on demand of practical tools as the world of cyber-security infrastructure has changed the landscape to access control via gatekeepers or gateways. However, the venue of two factor authentication (SSH keys for example) preferred over simpler Unix based login has introduced the challenging task of managing private keys and its association with individual users. Moreover, while a facility could simplify their model as one key one remote user therefore one local user and deploy a strategy along the lines of LDAP-SSH (Darwin project), such approach would not work for facilities allowing mapping between one “real” remote user and many local accounts adding to that the complexity and dimension of possibly multiple servers.

We will present an SSH key management system we developed, tested and deployed to address the one to many dilemma in the RHIC/STAR experiment. We will explain its use in an online computing context and explain the problems it addresses amongst which, making possible the management and tracing of group account access spread over many sub-system components (data acquisition, slow control, trigger groups) without the need of publicly known passwords (while keeping track at all times who/where).

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

STAR

Poster 2 - Board: 102 / 407

An effective XML based name mapping mechanism within StoRM

Authors: Alberto Forti¹ ; Ezio Corso² ; Luca Magnoni¹ ; Riccardo Zappi¹

¹ *INFN-CNAF*

² *ICTP-EGRID*

Corresponding Author: luca.magnoni@cnaf.infn.it

In a Grid environment the naming capability allows users to refer to specific data resources in a physical storage system using a high level logical identifier. This logical identifier is typically organized in a file system like structure, a hierarchical tree of names. Storage Resource Manager (SRM) services map the logical identifier to the physical location of data evaluating a set of parameters as the desired quality of services and the VOMS attributes specified in the requests.

StoRM is a SRM service developed by INFN and ICTP-EGRID to manage file and space on standard POSIX and high performing parallel and cluster file systems.

An upcoming requirement in the Grid data scenario is the orthogonality of the logical name and the physical location of data, in order to refer, with the same identifier,

to different copies of data archived in various storage areas with different quality of service.

The mapping mechanism proposed in StoRM is based on a XML document that represents the different storage components managed by the service, the storage areas defined by the site administrator, the quality of service they provide and the Virtual Organization that want to use the storage area.

An appropriate directory tree is realized in each storage component reflecting the XML namespace schema.

In this scenario StoRM is able to identify the physical location of a requested data evaluating the logical identifier and the specified attributes following the XML schema, without querying any database service.

This paper presents the namespace schema defined, the mapping mechanism and the technical details of the StoRM implementation.

Software components, tools and databases - Board: 37 / 38

An interface for GEANT4 simulation using ROOT geometry navigation.

Authors: Andrei Gheata¹ ; Mihaela Gheata¹

¹ CERN/ISS

Corresponding Author: andrei.gheata@cern.ch

The ROOT geometry modeller (TGeo) offers powerful tools for detector geometry description. The package provides several functionalities like: navigation, geometry checking, enhanced visualization, geometry editing GUI and many others, using ROOT I/O. A new interface module g4root was recently developed to take advantage of ROOT geometry navigation optimizations in the context of GEANT4 simulation. The interface can be used either by native GEANT4-based simulation applications or in the more general context of the Virtual Monte Carlo (VMC) framework developed by ALICE offline and ROOT teams. The latter allows running GEANT3, GEANT4 and FLUKA simulations without changing the geometry description nor the user code.

The interface was tested and stressed in the context of ALICE simulation framework. A description of the interface, its usage as well as recent results in terms of reliability and performance will be presented. Some benchmarks will be compared for ROOT-TGeo or GEANT4 based navigation.

Summary:

Navigation interface of GEANT4 with a ROOT TGeo geometry. **Submitted on behalf of Collaboration (ex, BaBar, ATLAS):**

ALICE Offline Collaboration

Poster 1 - Board: 82 / 188

An interface to tape for disk pool managers

Author: Jos Van Wezel¹

Co-authors: Doris Ressmann² ; Silke Halstenberg² ; Stephanie Meier²

¹ Forschungszentrum Karlsruhe (FZK/GridKa)

² Forschungszentrum Karlsruhe (FZK)

Corresponding Author: jos.van.wezel@cern.ch

The disk pool managers in use in the HEP community focus on managing disk storage but at the same time rely on a mass storage i.e. tape based system either to offload data that has not been touched for a long time or for archival purposes. Traditionally tape handling systems like HPSS by IBM or Enstore developed at FNAL are used because they offer specialized features to overcome the limitations of the sequential data access of tape. Not all centers have the resources to support special purpose tape handling systems but in many environments like FZK, Tivoli Storage Manager is in use for regular desktop backups. The paper describes the dCache to TSM interface TSS, that has been developed and is in use for over a year at FZK/GridKa. It served during the last WLCG services challenges and the recent experiment data challenges with peak rates of 300 MB/s into 8 tape drives. TSS and TSM make use of SAN connected data movers that write in parallel to tape. The TSS interface can be used for dCache and xrootd and offers a queuing layer between the disk cache and the tape backend in order to enhance store and pre-staging operations. The current status of the project and achieved data rates as well as future enhancements are presented.

Poster 2 - Board: 44 / 136

An original model for the simulation of the stopping power of negative hadrons

Authors: Maria Grazia Pia¹ ; Stephane Chauvie²

¹ INFN GENOVA

² INFN Genova

Corresponding Author: maria.grazia.pia@cern.ch

An original model is presented for the simulation of the energy loss of negatively charged hadrons: it calculates the stopping power by regarding the target atoms as an ensemble of quantum harmonic oscillators.

This approach allows to account for charge dependent effects in the stopping power, which are relevant at low energy: the differences between the stopping powers of positive and negative charged hadrons may amount to approximately a factor two, which is significant when high precision is required for the evaluation of energy deposit distributions. Related use case are the tails of energy distributions in the development of showers, the environment of radiation background monitors, experiments for antimatter searches, and the recent interest demonstrated for the possible therapeutic applications of antiproton beams.

The resulting antiproton stopping powers for different elements are compared against measurements at CERN antiproton experiments and are shown to be in satisfactory agreement with experimental data.

The model described is implemented in the Low Energy Electromagnetic package of the Geant4 Toolkit; it represents a significant improvement for the accurate simulation of low energy negative hadrons with respect to previously available models.

Software components, tools and databases / 259

Analysing CMS software performance using IgProf, OProfile and callgrind

Author: Lassi Tuura¹

Co-authors: Giulio Eulisse¹ ; Vincenzo Innocente²

¹ *Northeastern University*

² *CERN*

Corresponding Author: lassi.tuura@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. Understanding the performance of such a complex system is a very challenging task, not the least because there are extremely few developer tools capable of profiling software systems of this scale, or producing useful reports.

CMS has mainly used IgProf, valgrind, callgrind and OProfile for analysing the performance and memory usage patterns of our software. We describe the challenges, at times rather extreme ones, faced as we've analysed the performance of our software and how we've developed an understanding of the performance features. We outline the key lessons learnt so far and the actions taken to make improvements. We describe why an in-house general profiler tool still ends up besting a number of renowned open-source tools, and the improvements we've made to it in the recent year.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS

Event processing / 225

Analysis Environments for CMS

Author: Christopher Jones¹

Co-authors: Benedikt Hegner² ; Luca Lista³

¹ *Cornell University*

² *DESY*

³ *Istituto Nazionale di Fisica Nucleare*

The CMS offline software suite uses a layered approach to provide several different environments suitable for a wide range of analysis styles.

At the heart of all the environments is the ROOT-based event data model file format. The simplest environment uses "bare" ROOT to read files directly, without the use of any CMS-specific supporting libraries. This is useful for performing simple checks on a file or plotting simple distributions (such as the momentum distribution of tracks).

The second environment supports use of the CMS framework's smart pointers that read data on demand, as well as automatic loading of the libraries holding the object interfaces. This environment fully supports interactive

ROOT sessions in either CINT or PyROOT. The third environment combines ROOT's TSelector with the data access

API of the full CMS framework, facilitating sharing of code between the ROOT environment and the full framework.

The final environment is the full CMS framework that is used for all data production activities as well as full access

to all data available on the Grid. By providing a layered approach to analysis environments, physicists can choose the environment that most closely matches their individual work style.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS

Plenary / 491

Analysis tools for the LHC experiments

Corresponding Author: dietrich.liko@cern.ch

Dietrich Liko:

Dietrich Liko is researcher at the Institute for High Energy Physics of the Austrian Academy of Sciences. He is currently on leave to participate in the development of analysis tools for the grid with the EGEE project and as ATLAS Distributed Analysis Coordinator.

Software components, tools and databases / 217

Assessment of Data Quality in ATLAS

Authors: Andreas Hoecker¹ ; Beate Heinemann² ; Claude Guyot³ ; Else Lytken¹ ; Michael Hauschild¹ ; Michael Wilson¹ ; Richard Hawkings¹ ; Robert McPherson⁴ ; Rolf Seuster⁴

¹ *European Organisation for Nuclear Research (CERN)*

² *Lawrence Berkeley National Lab. (LBNL)*

³ *DAPNIA, Centre d'Etudes de Saclay (CEN Saclay)*

⁴ *University of Victoria*

Corresponding Author: michael.g.wilson@cern.ch

Assessing the quality of data recorded with the Atlas detector is crucial for commissioning and operating the detector to achieve sound physics measurements.

In particular, the fast assessment of complex quantities obtained during event reconstruction and the ability to easily track them over time are especially important given the large data throughput and the distributed nature of the analysis environment.

The data are processed once on a computer farm comprising O(1000) nodes before being distributed on the Grid, and reliable, centralized methods must be used to organize, merge, present, and archive data-quality metrics for performance experts and analysts.

A review of the tools and approaches employed by the detector and physics groups in this environment and a summary of their performances during commissioning are presented.

Poster 2 - Board: 51 / 208

Automatic Model Selection Using Machine Learning Techniques for Event Selection in Particle Physics

Author: Ricardo Vilalta¹

Co-authors: Paul Padley² ; Pedrame Bargassa²

¹ *University of Houston*

² *Rice University*

Corresponding Author: vilalta@cs.uh.edu

Advances in statistical learning have placed at our disposal a rich set of classification algorithms (e.g., neural networks, decision trees, Bayesian classifiers, support vector machines, etc.) with little or no guidelines on how to select the analysis technique most appropriate for the task at hand. In this paper we present a new approach for the automatic selection of predictive models based on the characteristics of the data under analysis. According to the particular data distribution, our methodology may decide to choose a learning algorithm able to delineate complex decision boundaries over the variable space (but exhibiting an inevitable high variance), or rather instead use an algorithm less complex that delineates coarse decision boundaries (but exhibiting a desirable low bias). Our experimental analysis looks for the identification of stop1 signal at energy of 1.96 TeV. The problem is inherently difficult because of the existence of background data with identical signal signatures. We report results using several metrics (e.g., accuracy, efficiency), and compare the performance of our methodology to a model produced by a domain expert that separates manually signal events from background events.

Poster 2 - Board: 4 / 86

Automatic processing of CERN video, audio and photo archives

Author: Michal Kwiatek¹

¹ *CERN*

Corresponding Author: michal.kwiatek@cern.ch

The digitalization of CERN audio-visual archives, a major task currently in progress, will generate over 40 TB of video, audio and photo files. Storing these files is one issue, but a far more important challenge is to provide long-time coherence of the archive and to make these files available on line with minimum manpower investment.

An infrastructure, based on standard CERN services, has been implemented whereby master files, stored in the CERN Distributed File System (DFS), are discovered and scheduled for encoding into lightweight web formats based on predefined profiles. Changes in master files, conversion profiles or in the metadata database (read from CDS, the CERN Document Server) are automatically detected and the media re-encoded whenever necessary. The encoding processes are run on virtual servers provided on-demand by the CERN Server Self Service Center, so that new servers can be easily configured to adapt to higher load. Finally, the generated files are made available from the CERN standard web servers with streaming implemented using Windows Media Services.

This paper describes the architecture in detail and analyses its advantages and limitations.

Computer facilities, production grids and networking / 425**BNL dCache Status and Plan**

Author: Dantong Yu¹

Co-authors: Carlos Gamboa¹; Christopher Hollowell¹; Hiro Ito¹; Jason Smith¹; Jay Packard¹; John Hover¹; Michael Ernst¹; Ofer Rind¹; Razvan Popescu¹; Robert Petkus¹; Wu Yingzi¹; Zhenping Liu¹

¹ BROOKHAVEN NATIONAL LABORATORY

Corresponding Author: dtyu@bnl.gov

BNL ATLAS Computing Facility needs to provide a Grid-based storage system with these requirements: a total of one gigabyte per second of incoming and outgoing data rate between BNL and ATLAS T0, T1 and T2 sites, thousands of reconstruction/analysis jobs accessing locally stored data objects, three petabytes of disk/tape storage in 2007 scaling up to 25 petabytes by 2011, and a cost-effective storage solution.

BNL's dCache implementation utilizes three types of storage media: directly attached disk storage to pool nodes via fiber channel to preserve precious data, strategic local disks on worker nodes to provide cost-effective petascale on-line storage, and an HPSS providing archival and redundancy. Dual home GridFtp door nodes bypassing the BNL firewall allow any internal pool node to send and receive data to and from remote users without exposing these nodes to the Internet. We upgraded the critical components such as PNFS and SRM, distributed loads to separate machines, and fine-tuned various parameters, resulting in 300 MB/s between CERN and Tier 1 during the ATLAS data export exercise.

We have started exercising two critical data transfer tasks to validate the readiness of BNL USATLAS data storage in terms of stability and performance, namely: 1) for WAN data transfer, running basic transfer (e.g. SRMCP w/ and w/o FTS) and data replications based on ATLAS DDM between BNL and USATLAS Tier 2 sites, and 2) exercising the LAN based dCache Posix I/O functionalities (e.g. dcap, TDCacheFile, and Tfile) and measuring the performance of concurrent read access to the same data set by a large number of analysis jobs. We will also test SRM V2 when it becomes available and implement its storage classes to support ATLAS reconstruction jobs requiring pre-staged RAW data from tape to disk.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

Brookhaven National Lab, ATLAS Computing Facility

Distributed data analysis and information management / 48**BaBar MC Production on the Canadian Grid using a Web Services Approach**

Authors: Ashok Agarwal¹; Daniel Vanderster¹; Ian Gable¹; Randall Sobie¹; Ron Desmarais¹; Sergey Popov¹; Tristan Sullivan¹

¹ University of Victoria

Corresponding Author: aka@uvic.ca

The present paper highlights the approach used to design and implement a web services based BaBar Monte Carlo (MC) production grid using Globus Toolkit version 4. The grid integrates the resources of two clusters at the University of Victoria, using the ClassAd mechanism provided by the Condor-G metascheduler. Each cluster uses the Portable Batch System (PBS) as its local resource management system (LRMS). Resource

brokering is provided by the Condor matchmaking process, whereby the job and resource attributes are expressed as ClassAds. The important features of the grid are automatic registering of resource ClassAds to the central registry, ClassAds extraction from the registry to the metascheduler for matchmaking, and the incorporation of input/output file staging. Web-based monitoring is employed to track the status of grid resources and the jobs for an efficient operation of the grid. The performance of this new grid for BaBar jobs, and a comparison with the existing Canadian computational grid (Gridx1) based on Globus Toolkit version 2 is presented.

Collaborative tools / 115

Benchmarks of medical dosimetry on the grid

Authors: Anton Lechner¹ ; Francesco Dirosa² ; Giacomo Cuttone³ ; Giorgio Russo² ; Jakub Moscicki⁴ ; Maria Grazia Pia⁵ ; Pablo Cirrone² ; Patricia Mendez Lorenzo⁴

¹ *Tech. Univ. Wien*

² *INFN LNS*

³ *INFN*

⁴ *CERN*

⁵ *INFN Genova*

Corresponding Authors: maria.grazia.pia@cern.ch, anton.lechner@cern.ch

Computational tools originating from high energy physics developments provide solutions to common problems in other disciplines: this study presents quantitative results concerning the application of HEP simulation and analysis tools, and of the grid technology, to dosimetry for oncological radiotherapy.

The study concerned all the three major radiotherapy techniques: therapy with external beams from a medical linear accelerator (in particular, the modern technique or intensity modulated radiotherapy), brachytherapy with internal or superficial radioactive sources, and hadrontherapy with a proton beam. Geant4-based simulation applications developed for three realistic use case highlight the high precision dose calculation achievable; the simulation is complemented by AIDA-compliant analysis tools for the manipulation of dose distributions relevant to clinical usage. The application design exploits DIANE for transparent execution in sequential and parallel mode, either on a local farm or on the grid and GANGA as a grid user interface.

Benchmarks for execution on the grid are presented; they highlight the capabilities and current limitations for exploiting the grid in real-life applications in all major branches of oncological radiotherapy.

Computational tools for dosimetry based on HEP software systems couple high precision, speed suitable to clinical environments and low cost: therefore, they represent an alternative to commercial products of particular interest to developing countries for oncological radiotherapy or radiation protection applications.

Plenary / 487

Benefits of Grid Technology to Running Experiments

Corresponding Author: fkw@fnal.gov

Poster 1 - Board: 96 / 291

Beyond Grid Security

Author: Bruno Hoefft¹

Co-author: Ursula Epting¹

¹ *Forschungszentrum Karlsruhe*

Corresponding Author: hoefft@iwr.fzk.de

While many fields relevant to Grid security are already covered by existing working groups, their remit rarely goes beyond the scope of the Grid infrastructure itself. However, security issues pertaining to the internal set-up of compute centres have at least as much impact on Grid security. Thus, this talk will present briefly the EU ISSeG project (Integrated Site Security for Grids). To complement groups such as OSCT (Operational Security Coordination Team) and JSPG (Joint Security Policy Group), the purpose of ISSeG is to provide a holistic approach to security for Grid computer centres, from strategic considerations to an implementation plan and its deployment. The generalised methodology of Integrated Site Security (ISS) is based on the knowledge gained during its implementation at several sites as well as through security audits, and this will be briefly discussed. Several examples of ISS implementation tasks at the Forschungszentrum Karlsruhe will be presented, including segregation of the network for administration and maintenance and the implementation of Application Gateways. Furthermore, the web-based ISSeG training material will be introduced. This aims to offer ISS implementation guidance to other Grid installations in order to help avoid common pitfalls.

Summary:

While many fields relevant to Grid security are already covered by existing working groups, their remit rarely goes beyond the scope of the Grid infrastructure itself. However, security issues pertaining to the internal set-up of compute centres have at least as much impact on Grid security. Thus, this talk will present briefly the EU ISSeG project (Integrated Site Security for Grids)

Software components, tools and databases / 275

Booting ROOT with BOOT

Author: Rene Brun¹

¹ *CERN*

Corresponding Author: rene.brun@cern.ch

The BOOT project was introduced at CHEP06 and is gradually implemented in the ROOT project.

A first phase of the project has consisted in an important restructuring of the ROOT core classes such that only a small subset is required when starting a ROOT application (including user libraries). Thanks to this first phase, the virtual address space required by the interactive version has been reduced by a factor 3.

A second phase of the project has the objective to eliminate a substantial fraction of the dictionary code generated by the preprocessor rootcint. This code will be replaced by persistent objects stored in a ROOT file and

the CINT stub functions replaced by direct calls to the compiled code. Prototypes have been developed early 2007 and a full implementation is expected at the time of CHEP07.

A third phase under prototyping will automatize the autoloading of the code on demand from a central source repository and online compilation of this code with local caches. This will facilitate and speed-up the installation of new versions. A by-product of this phase has been the implementation of a ROOT file cache (presented in another talk) improving drastically the performance of the ROOT I/O in high latency networks.

When all these phases will be completed, it should be possible to install and run a ROOT based application from a web browser (a BROOTER).

Distributed data analysis and information management / 161

Building a Scalable Event-Level Metadata System for ATLAS

Authors: David Malon¹ ; Florbela Tique Aires Viegas² ; Helen McGlone³ ; Jack Cranshaw¹ ; Luc Goossens²

¹ *Argonne National Laboratory*

² *CERN*

³ *University of Glasgow/CERN*

Corresponding Author: helen.mcglone@cern.ch

The ATLAS TAG database is a multi-terabyte event-level metadata selection system, intended to allow discovery, selection of and navigation to events of interest to an analysis. The TAG database encompasses file- and relational-database-resident event-level metadata, distributed across all ATLAS Tiers.

An oracle hosted global TAG relational database, containing all ATLAS events, implemented in Oracle, will exist at Tier 0. Implementing a system that is both performant and manageable at this scale is a challenge.

A 1 TB relational Tag database has been deployed at Tier 0 using simulated tag data. The database contains one billion events, each described by two hundred event metadata attributes, and is currently undergoing extensive testing in terms of queries, population and manageability. These 1 TB tests aim to demonstrate and optimise the performance and scalability of an Oracle TAG database on a global scale.

Partitioning and indexing strategies are crucial to well-performing queries and manageability of the database and have implications for database population and distribution, so these are investigated. Physics query patterns are anticipated, but a crucial feature of the system must be to support a broad range of queries across all attributes.

Concurrently, event tags from ATLAS Computing System Commissioning distributed simulations are accumulated in an Oracle-hosted database at CERN, providing an event-level selection service valuable for user experience and gathering information about physics query patterns.

In this paper we describe the status of the Global TAG relational database scalability work and highlight areas of future direction.

Summary:

An oracle hosted global TAG relational database prototype, deployed at cern as part of multi-terabyte event-level metadata selection system, is discussed.

This is an ATLAS Offline Computing abstract, Distributed data analysis and information management track assigned by the ATLAS speakers committee.

Poster 1 - Board: 81 / 183

Building a production-grade Grid infrastructure at DESY

Author: Andreas Gellrich¹

Co-authors: Birgit Lewendel¹; Christoph Wissing¹; Martin Gasthuber¹; Patrick Fuhrmann¹; Sven Sternberger¹; Uwe Ensslin¹; Volker Guelzow¹; Yves Kemp¹

¹ DESY

Corresponding Author: andreas.gellrich@desy.de

As a partner of the international EGEE project in the German/Switzerland federation (DECH) and as a member of the national D-GRID initiative, DESY operates a large-scale production-grade Grid infrastructure with hundreds of CPU cores and hundreds of Terabytes of disk storage.

As Tier-2/3 center for ATLAS and CMS DESY plays a leading role in Grid computing in Germany. DESY strongly support non-LHC VOs and fosters the Grid usage in other eScience fields. The DESY Grid infrastructure is the home of a number of global, regional, and local VOs, among them are the HERA experiments H1 and ZEUS ('hone', 'zeus'), the ILC community ('calice', 'ilc'), and for the International Lattice Data Grid ('ildg'). All of them are heavily using the Grid for event simulation and data analysis.

The DESY Grid infrastructure incorporates all necessary Grid services to host its VOs and provides computing and storage resources for all supported VOs in the same Grid infrastructure. Main emphasis has been put on embedding the Grid infrastructure seamlessly in the DESY computer center. Crucial aspects are the choice of batch system and storage technologies (dCache). Important aspects are reliability of services by providing redundancies, monitoring, and administrative tools as well as scalable installation and updating procedures.

In the contribution the CHEP07 we will describe the Grid set-up at DESY and discuss in detail concepts and implementation to achieve a scalable, pervasive, and easy to maintain system which meets the requirements of a global Grid.

Grid middleware and tools / 47

Building a robust distributed system: some lessons from R-GMA

Authors: A Paventhan¹; Adebisi Kuseju¹; Alastair Duncan¹; Antony Wilson¹; Ming Jiang¹; Parminder Bhatti¹; Steve Fisher¹

¹ RAL

Corresponding Author: s.m.fisher@rl.ac.uk

R-GMA, as deployed by LCG, is a large distributed system. We are currently addressing some design issues to make it highly reliable, and fault tolerant.

In validating the new design, there were two classes of problems to consider: one related to the flow of data and the other to the loss of

control messages. R-GMA streams data from one place to another; there is a need to consider the behaviour when data is being inserted more rapidly into the system than taken out and more generally how to deal with bottlenecks. In the original R-GMA design the system tried hard to deliver all control messages; those messages that were not delivered quickly were queued for retry later. In the case of badly configured firewalls, network problems or very slow machines this led to long queues of messages, some of which were superseded by later messages that were also queued. In the new design no individual control message is critical; the system just needs to know if each message was received successfully.

The system should also avoid single points of failure. However this can require complex code resulting in a system that is actually less reliable.

We describe how we have dealt with bottlenecks in the flow of data, loss of control messages and the elimination of single points of failure to produce a robust R-GMA design. The work presented, though in the context of R-GMA, is applicable to any large distributed system.

Grid middleware and tools / 107

Building the WLCG file transfer service

Authors: Gavin McCance¹ ; Paolo Badino¹

¹ *CERN*

Corresponding Authors: markus.schulz@cern.ch, paolo.badino@cern.ch, gavin.mccance@cern.ch

A key feature of WLCG's multi-tier model is a robust and reliable file transfer service that efficiently moves bulk data sets between the various tiers, corresponding to the different stages of production and user analysis. We describe in detail the file transfer service both the tier-0 data export and the inter-tier data transfers, discussing the transition and lessons learned in moving from a reliable software product, the gLite FTS, to a full production service based on that software. The focus is upon the deployment and operational experience of the service gained during the 2006 and 2007 experiment production activities and dress rehearsals. We discuss the software and operational features that have been deployed to meet the reliability and performance needs of the service, and integration of the service with the WLCG and experiment operations.

Poster 2 - Board: 70 / 103

Bulk file transfer and storage with GridSite and SlashGrid

Authors: Andrew McNab¹ ; Yibiao LI²

¹ *University of Manchester*

² *UNIVERSITY OF MANCHESTER*

GridSite has extended the industry-standard Apache webserver for use within Grid projects, by adding support for Grid security credentials such as GSI and VOMS. With the addition of

the GridHTTP protocol for bulk file transfer via HTTP and the development of a mapping between POSIX filesystem operations and HTTP requests we have extended this scope of GridSite into bulk data transfer and storage. We present measurements of the performance obtained from HTTP over the wide and local area networks and compare this with other available protocols. Finally, we show how the SlashGrid component of GridSite provides transparent access to remote HTTP(S) URLs which applications can use via the operating system's POSIX file access API. SlashGrid/GridSite therefore implement a distributed filesystem based on HTTP(S) and GridSite filesystems, that can be used within a local area network or across the internet. This system solely relies on Grid credentials such as GSI and VOMS and implements a fine-grained access control model without the need to manipulate underlying Unix accounts or permissions.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

GridPP

Software components, tools and databases / 99

CABS3 - CLEO Analysis by Script @ Belle

Author: Nobuhiko Katayama¹

¹ *High Energy Accelerator Research Organization*

Corresponding Author: nobu.katayama@kek.jp

We developed the original CABS language more than 10 years ago. The main objective of the language was to describe a decay of a particle as simply as possible in the context of usual HEP data analysis. A decay mode, for example, can be defined as follows:

```
define Cand Dzerobar kpi 2 { K+ identified pi- identified }
hist 1d inv_mass 0 80 1.5 2.3 "all momentum"
cut inv_mass .ge. 1.83 .and. inv_mass .le. 1.90
...
end-define
```

where Dzerobar kpi together defines a particle made of K+ and pi-. At the execution time of the translated C++ code, the list of Dzerobar candidate combinations are generated from the list of K+ and pi- candidate lists.

The reason we invented the scripting language was the following. Although C++ allows us to define new class and operators for them and let us write native C++ code like "Dzerobar = Kplus + piminus" to make combinations, we thought there was no way to extend it to the cases with more than two particles; Dzerobar = Kplus + piminus + piplus + piminus to make combinations out of four particles as the language only defines the binary operator.

After more than ten years, the author learned that it is indeed possible to define set of classes in C++ and write a single line of code to automatically generate the list of candidate combinations from generic cases with N(>0) particles. The same particle can appear more than once at any order. This noble idea comes from the delayed evaluation of a purely functional language, Haskell. CABS3 is purely templated so that it incorporates any List class derived from .

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

Belle

Computer facilities, production grids and networking / 62**CASTOR2: design and development of a scalable architecture for a hierarchical storage system at CERN**

Authors: Dennis Waldron¹ ; Giulia Taurelli¹ ; Giuseppe Lo Presti² ; Olof Barring¹ ; Rosa Maria Garcia Rioja¹ ; Sebastien Ponce¹

¹ CERN

² CERN/INFN

Corresponding Author: giuseppe.lopresti@cern.ch

In this paper we present the architecture design of the CERN Advanced Storage system (CASTOR) and its new disk cache management layer (CASTOR2). Mass storage systems at CERN have evolved over time to meet growing requirements, both in terms of scalability and fault resiliency. CASTOR2 has been designed as a Grid-capable storage resource sharing facility, with a database-centric architecture, to keep the whole status of the system, and stateless daemons. We present an overview of the software architecture upon which CASTOR2 daemons are built, and the UML based software process that is in place to speed up and automate code development. We also demonstrate how external policies may be plugged into the framework to ease the operation of a CASTOR2 system, which is now being used in production at CERN as well as at a number of Tier1 sites since more than one year.

Poster 2 - Board: 21 / 299

CDF Monte Carlo data transfer and storage with Grid tools

Author: Douglas Benjamin¹

¹ Duke University

The CDF experiment at Fermilab produces Monte Carlo data files using computing resources on both the Open Science Grid (OSG) and LHC Computing Grid (LCG) grids. This data produced must be brought back to Fermilab for archival storage. In the past CDF produced Monte Carlo data on dedicated computer farms through out the world. The data files were copied directly from the worker nodes to a few file servers located at FNAL using rcp and Kerberos authentication. As the experiment has moved from dedicated resources to shared resources on the grid, this technique has proven to be problematic. We plan to show how changing our data delivery model to one of concentrating the data in a few sites located in different regions throughout the world and shipping it back in a coordinated method back to FNAL improved our use of opportunistic CPU cycles. The details of the evaluation process and implemented solution including metrics demonstrating our performance will also be presented.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CDF collaboration

Distributed data analysis and information management / 297

CDF experience with Monte Carlo production using LCG Grid

Authors: Daniele Cesini¹ ; Donatella Lucchesi² ; Gabriele Compostella³ ; Igor Sfiligoi⁴ ; Simone Pagan Griso⁵ ; francesco delli paoli⁶

¹ INFN-CNAF

² INFN Padova

³ University Of Trento INFN Padova

⁴ Fermilab

⁵ University and INFN Padova

⁶ INFN of Padova

Corresponding Authors: simone.pagan@pd.infn.it, donatella.lucchesi@pd.infn.it

The upgrades of the Tevatron collider and of the CDF detector have considerably increased the demand on computing resources in particular for Monte Carlo production for the CDF experiment. This has forced the collaboration to move beyond the usage of dedicated resources and start exploiting Grid resources.

The CDF Analysis Farm (CAF) model has been reimplemented into LcgCAF in order to access Grid resources by using the LCG/EGEE Middleware components. Many sites in Italy and in Europe are accessed via this portal in order to produce Monte Carlo data and in one year of operations we expect about 100,000 Grid jobs submitted by the CDF users.

We review here the setup used to submit jobs to Grid sites and retrieve the output, including the Grid components CDF-specific configuration. The batch and interactive monitor tools developed to allow users to verify the jobs status during their lifetimes in the Grid environment are described.

We analyze the efficiency and typical failure modes of the current Grid infrastructure reporting the performances of different parts of the used system.

Computer facilities, production grids and networking / 427

CDF offline computing'07: computing of a HEP experiment in a mature stage

Author: Pavel Murat¹

¹ Fermilab

Corresponding Author: murat@fnal.gov

CDFII detector at Fermilab is taking physics data since 2002. The architecture of the CDF computing system has substantially evolved during the years of the data taking and currently it reached stable configuration which will allow experiment to process and analyse the data until the end of Run II.

We describe major architectural components of the CDF offline computing - dedicated reconstruction and analysis farms, GRID-based

Monte Carlo Production system, distributed databases, distributed hierarchical storage system, code development and distribution system.

We present technical parameters of the CDF computing system and projected needs of the CDF computing for the next several years.

We summarize the operational experience accumulated over the course of Run II and highlight the challenges the experiment had to overcome to reach the state where the CDF physicists are routinely using GRID in their daily analysis work.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CDF

Software components, tools and databases - Board: 80 / 172

CERN Database Services for the LHC Computing Grid

Author: Maria Girone¹

¹ CERN

Corresponding Author: maria.girone@cern.ch

Physics meta-data stored in relational databases play a crucial role in the Large Hadron Collider (LHC) experiments and also in the operation of the Worldwide LHC Computing Grid (WLCG) services. A large proportion of non-event data such as detector conditions, calibration, geometry and production bookkeeping relies heavily on databases. Also, the core Grid services that catalogue and distribute LHC data cannot operate without a reliable database infrastructure at CERN and elsewhere. The Physics Services and Support group at CERN provides database services for the physics community. With an installed base of several TB-sized database clusters, the service is designed to accommodate growth for data processing generated by the LHC experiments and LCG services. During the last year, the physics database services went through a major preparation phase for LHC start-up and are now fully based on Oracle clusters on Intel/Linux. Over 100 database server nodes are deployed today in some 15 clusters serving almost 2 million database sessions per week. This talk will detail the architecture currently deployed in production and the results achieved in the areas of high availability, consolidation and scalability. Service evolution plans for the LHC start-up will also be discussed.

Collaborative tools / 55

CERN Single Sign On solution

Author: Emmanuel Ormancey¹

¹ CERN

Corresponding Author: emmanuel.ormancey@cern.ch

The need for Single Sign On has always been restricted by the lack of cross platform solutions: a single sign on working only on one platform or technology is nearly useless. The recent improvements in Web Services Federation (WS-Federation) standard enabling federation of identity, attribute, authentication and authorization information can now provide real extended Single Sign On solutions.

CERN has investigated various options and now provides a Web SSO solution using some parts of WS-Federation technology. By using Shibboleth Service Provider module for Apache hosted web sites and Microsoft ADFS as identity provider linked to Active Directory user, users can now authenticate on any web application using a single authentication platform, providing identity, user information (building, phone...) as well as group membership enabling authorization possibilities. A typical scenario: a CERN user can now authenticate on a Linux/Apache website using Windows Integrated credentials, and his Active Directory group membership can be checked before allowing access to a specific web page.

Poster 2 - Board: 18 / 277

CMS CSA06 experience at INFN

Author: Daniele Bonacorsi¹

Co-authors: Alessandra Fanfani²; Chiara Latini²; Chiara Rovelli³; Cristina Bulfon³; Ezio Torassa⁴; Federico Calzolari⁵; Filippo Ambroglini⁶; Giacinto Donvito⁷; Giorgio Maggi⁷; Giovanni Organtini³; Giuseppe Bagliesi⁵; Giuseppe Codispoti²; Guido Cuscela⁷; Livio Fano⁶; Luciano Barone³; Marcello Abbrescia⁷; Marco Paganoni⁸; Martina Malberti⁸; Massimo Biasotto⁴; Nicola De Filippis⁷; Paolo Bartalini⁹; Paolo Capiluppi²; Paolo Meridiani³; Pietro Govoni⁸; Riccardo Paramatti³; Roberto Tenchini⁵; Salvatore My⁷; Shahram Rahatlou³; Simone Gennai⁵; Stefano Belforte¹⁰; Subir Sarkar⁵; Tommaso Boccali⁵; Tommaso Tabarelli De Fatis⁸; William Bacchi²; martino margoni⁴; ugo gasparini⁴

¹ INFN-CNAF, Bologna, Italy

² Università di Bologna, INFN-Bologna, Italy

³ Università di Roma-1 and INFN-Roma, Italy

⁴ Università di Padova, INFN-Legnaro and Laboratori Nazionali di Legnaro, Padova, Italy

⁵ Università di Pisa, Scuola Normale Superiore and INFN-Pisa, Italy

⁶ Università di Perugia, INFN-Perugia, Italy

⁷ Università di Bari, Politecnico di Bari and INFN-Bari, Italy

⁸ Università degli Studi Milano-Bicocca, INFN-Milano, Italy

⁹ University of Florida, Gainesville, USA

¹⁰ Università di Trieste and INFN-Trieste, Italy

The CMS experiment operated a Computing, Software and Analysis Challenge in 2006 (CSA06). This activity is part of the constant work of CMS in computing challenges of increasing complexity to demonstrate the capability to deploy and operate a distributing computing system at the desired scale in 2008. The CSA06 challenge was a 25% exercise, and included several workflow elements: event reconstruction at the CERN Tier-0 center; data distribution to Tier-1's, for archiving and data serving purposes; data skimming, driven by CMS physics groups, and re-reconstruction at Tier-1's; serving of re-processed data to Tier-2's and Grid submission of physics analysis jobs. The INFN region joined this exercise. An overview of the software tools and the computing architecture deployed at INFN for the challenge is given in this report, as well as a summary of the operational experience at INFN-CNAF Tier-1

and INFN Tier-2 centers. A description of the operational experience at INFN during CSA07 challenge - a CMS exercise at a scale of 50% started in July 2007 - are also presented and discussed.

Distributed data analysis and information management / 260

CMS Centers for Control, Monitoring, Offline Operations and Analysis

Author: Lucas Taylor¹

Co-authors: Erik Gottschalk²; Kaori Maeshima²; Patricia McBride²

¹ *Northeastern University, Boston*

² *Fermilab*

Corresponding Author: lucas.taylor@cern.ch

The CMS experiment is about to embark on its first physics run at the LHC. To maximize the effectiveness of physicists and technical experts at CERN and worldwide and to facilitate their communications, CMS has established several dedicated and inter-connected operations and monitoring centers. These include a traditional "Control Room" at the CMS site in France, a "CMS Centre" for up to fifty people on the CERN main site in Switzerland, and remote operations centers, such as the "LHC@FNAL" center at Fermilab.

We describe how this system of centers coherently supports the following activities: (1) CMS data quality monitoring, prompt sub-detector calibrations, and time-critical data analysis of express-line and calibration streams; and (2) operation of the CMS computing systems for processing, storage and distribution of real CMS data and simulated data, both at CERN and at offsite centers.

We describe the physical infrastructure that has been established, the computing and software systems, the operations model, and the communications systems that are necessary to make such a distributed system coherent and effective.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS

Distributed data analysis and information management / 322

CMS Conditions Data Access using FroNTier

Author: Lee Lueking¹

Co-authors: Barry Blumenfeld²; David Dykstra³; Eric Wicklund³

¹ *FERMILAB*

² *Johns Hopkins University*

³ *Fermilab*

Corresponding Author: lueking@fnal.gov

The CMS experiment at the LHC has established an infrastructure using the FroNTier framework to deliver conditions (i.e. calibration, alignment, etc.) data to processing clients worldwide. FroNTier is a simple web service approach providing client HTTP access to a central database service. The system for CMS has been developed to work with POOL which provides object relational mapping between the C++ clients and various database technologies. Because of the read only nature of the data, Squid proxy caching servers are maintained near clients and these caches provide high performance data access. Several features have been developed to make the system meet the needs of CMS including careful attention to cache coherency with the central database, and low latency loading required for the operation of the online High Level Trigger. The ease of deployment, stability of operation, and high performance make the FroNTier approach well suited to the GRID environment being used for CMS offline, as well as for the online environment used by the CMS High Level Trigger (HLT). The use of standard software, such as Squid and various monitoring tools, make the system reliable, highly configurable and easily maintained. We describe the architecture, software, deployment, performance, monitoring and overall operational experience for the system.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS Offline Data Management Group

Event processing / 253

CMS Event Display and Data Quality Monitoring for LHC Startup

Author: Ianna Osborne¹

¹ *Northeastern University*

Corresponding Author: ianna.osborne@cern.ch

The event display and data quality monitoring visualisation systems are especially crucial for commissioning CMS in the imminent CMS physics run at the LHC. They have already proved invaluable for the CMS magnet test and cosmic challenge.

We describe how these systems are used to navigate and filter the immense amounts of complex event data from the CMS detector and prepare clear and flexible views of the salient features to the shift crews and offline users. These allow shift staff and experts to navigate from a top-level general view to very specific monitoring elements in real time to help validate data quality and ascertain causes of problems.

We describe how events may be accessed in the higher level trigger filter farm, at the CERN Tier-0 centre, and in offsite centres to help ensure good data quality at all points in the data processing workflow. Emphasis has been placed on deployment issues in order to ensure that experts and general users may use the visuslisation systems at CERN, in remote operations and monitoring centers offsite, and from their own desktops.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS Offline grou

Computer facilities, production grids and networking / 369

CMS Experiences with Computing Software and Analysis Challenges

Author: Ian Fisk¹

¹ *FNAL*

Corresponding Author: ian.fisk@cern.ch

In preparation for the start of the experiment, CMS has conducted computing, software, and analysis challenges to demonstrate the functionality, scalability, and useability of the computing and software components. These challenges are designed to validate the CMS distributed computing model by demonstrating the functionality of many components simultaneously. In the challenges CMS has had participation from approximately 40 computing centers and has demonstrated event processing, data transfers, and analysis processing across the globally distributed computing environment. In this presentation, CMS will describe the types of tests performed including the scale achieved from each CMS component and externally provided component, the functionality demonstrated in the test, and the functionality left to validate before the experiment begins. The successes and lessons learned will be summarized and the directions for the future will be outlined.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

Ian Fisk, for the CMS Offline and Computing Groups

Computer facilities, production grids and networking / 235

CMS MC Production System Development & Design

Authors: Alessandra Fanfani¹ ; Carlos Kavka² ; Dave Evans³ ; Dave Mason³ ; Frank van Lingen⁴ ; Giulio Eulisse⁵ ; Giuseppe Codispoti¹ ; Jose Hernandez⁶ ; Nicola De Filippis⁷ ; Peter Elmer⁸ ; William Bacchi¹

¹ *INFN Sezione di Bologna and University of Bologna*

² *INFN Sezione di Trieste*

³ *Fermi National Laboratory*

⁴ *CALIFORNIA INSTITUTE OF TECHNOLOGY*

⁵ *North Eastern University*

⁶ *CIEMAT*

⁷ *INFN - Sezione di Bari*

⁸ *Princeton University*

Corresponding Authors: evansde@fnal.gov, fvlingen@caltech.edu, alessandra.fanfani@cern.ch, peter.elmer@cern.ch

The CMS production system has undergone a major architectural upgrade from its predecessor, with the goals of reducing the operations manpower requirement and preparing for the large scale production required by the CMS physics plan.

This paper discusses the CMS Monte Carlo Workload Management architecture. The system consist of 3 major components: ProdRequest, ProdAgent, and ProdMgr and can be deployed in various distributed configurations to prevent and minimize single points of failures. The user and request management interaction will take place on the ProdRequest level. ProdAgents are responsible for job submission and tracking over multiple Grid and Farm computing resources.

The ProdAgents themselves consist of autonomous components and communicate via asynchronous messages, thereby enhancing the robustness of the ProdAgent. Delayed and queued message functionality enables the ProdAgent to adequately deal with 3rd party component interaction (CMS catalogs, transfer systems) even when these components go offline for a while. ProdMgr provides the accounting functionality of the system keeping track of request progress and dividing the work between ProdAgents which request it. Various complementary (self) monitoring systems provide end-2-end monitoring of the system to track down (potential) problems.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS

Computer facilities, production grids and networking / 285

CMS Monte Carlo production in the WLCG Computing Grid

Authors: Ajit Mohapatra¹; Alessandra Fanfani²; Alexander Flossdorf³; Alexis Pompili⁴; Andrei Khomitch⁵; Carlos Kavka⁶; Carsten Hof⁵; Christos Lazaridis¹; Dave Evans⁷; Frank Van Lingen⁸; Giorgio Maggi⁹; Giulio Eulisse¹⁰; Giuseppe Codispoti²; Gregory Hammad¹¹; Ilaria Vilella¹²; Joris Maes¹²; Jose Afonso Lajas Sanches¹³; Jose Caballero¹⁴; Jose Hernandez Calama¹⁴; Marcello Abbrescia⁴; Nicola De Filippis¹⁵; Peter Elmer¹⁶; Peter Kreuzer⁵; Petra Van Mulders¹²; Salvatore My⁹; Sergey Kalinin⁵; Stijn De Weirdt¹²; Stuart Wakefield¹⁷; Subir Sarkar¹⁸; Wen Guan¹⁹; William Bacchi²

¹ *University of Wisconsin*

² *INFN and University of Bologna*

³ *DESY*

⁴ *University of Bari*

⁵ *RWTH*

⁶ *INFN Trieste*

⁷ *FNAL*

⁸ *California Institute of Technology*

⁹ *INFN and University of Bari*

¹⁰ *Northeastern University*

¹¹ *ULB*

¹² *VUB*

¹³ *UERJ*

¹⁴ *CIEMAT*

¹⁵ *INFN Bari*

¹⁶ *Princeton University*

¹⁷ *Imperial College*

¹⁸ *INFN Pisa*

¹⁹ *IHEP*

Corresponding Authors: chema.hernandez.calama@gmail.com, peter.kreuzer@cern.ch

Monte Carlo production in CMS has received a major boost in performance and scale since last CHEP conference. The production system has been re-engineered in order to incorporate the experience gained in running the previous system and to integrate production with the new CMS event data model, data management system and data processing framework. The system is interfaced to the two major computing Grids used by CMS, the LHC Computing Grid (LCG) and the Open Science Grid (OSG).

Operational experience and integration aspects of the new CMS Monte Carlo production system is presented together with an analysis of production

statistics. The new system automatically handles job resubmission, resource monitoring, job queuing, job distribution according to the available resources, data merging, registration of data into the data bookkeeping, data location, data transfer and placement systems. Compared to the previous production system it considerably improves automation, reliability and performance, eventually leading to a system that can be run and monitored by a small number of production operators. A more efficient use of computing resources and a better handling of the inherent Grid unreliability have resulted in an increase of production scale by about an order of magnitude, capable of running in parallel at the order of ten thousand jobs and yielding more than a million events per day.

Collaborative tools / 266

CMS Offline Web Tools

Author: Simon Metson¹

Co-authors: Alessandra Fanfani²; Brian Bockelman³; Carlos Kavka⁴; Dave Evans⁵; Dave Newbold¹; Derek Feichtinger⁶; Frank van Lingen⁷; Giulio Eulisse⁸; Katarzyna Dziedziniwicz⁹; Peter Elmer¹⁰; Ricky Egeland¹¹; Stefano Belforte⁴; Valentin Kuznetsov¹²

¹ *H.H. Wills Physics Laboratory, Bristol University*

² *Universita degli Studi di Bologna*

³ *University of Nebraska-Lincoln*

⁴ *INFN, Sezione di Trieste*

⁵ *Fermi National Accelerator Lab.*

⁶ *Paul Scherrer Institut*

⁷ *Caltech*

⁸ *Northeastern University*

⁹ *Warsaw University of Technology*

¹⁰ *Princeton University*

¹¹ *University of Minnesota*

¹² *Cornell*

Corresponding Authors: giulio.eulisse@cern.ch, simon.metson@cern.ch

We describe a relatively new effort within CMS to converge on a set of web based tools, using state of the art industry techniques, to engage with the CMS offline computing system. CMS collaborators require tools to monitor various components of the computing system and interact with the system itself. The current state of the various CMS web tools is described along side current planned developments. The CMS collaboration comprises of nearly 3000 people from all over the world. As well as its collaborators, its computing resources are spread all over globe and are accessed via the LHC grid to run analysis, large scale production and data transfer tasks. Due to the distributed nature of collaborators effective provision of collaborative tools is essential to maximise physics exploitation of the CMS experiment, especially when the size of the CMS data set is considered. CMS has chosen to provide such tools over the world wide web as a top level service, enabling all members of the collaboration to interact with the various offline computing components. Traditionally web interfaces have been added in HEP experiments as an afterthought. In the CMS offline we have decided to put web interfaces, and the development of a common CMS web framework, on an equal footing with the rest of the offline development.

Tools exist within CMS to transfer and catalogue data (PhEDEx and DBS/DLS), run Monte Carlo production (ProdAgent) and submit analysis (CRAB). Effective human interfaces to these systems are required for users with different agendas and practical

knowledge of the systems to effectively use the CMS computing system. The CMS web tools project aims to provide a consistent interface to all these tools.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS Offline group

Online computing / 221

CMS Online Web Based Monitoring

Author: William Badgett¹

¹ *Fermilab*

Corresponding Author: william.badgett@cern.ch

We present the Online Web Based Monitoring (WBM) system of the CMS experiment, consisting of a web services framework based on Jakarta/Tomcat and the Root data display package. Due to security concerns, many monitoring applications of the CMS experiment cannot be run outside of the experimental site. As such, in order to allow remote users access to CMS experimental status information, we implement a set of Tomcat/Java servlets running in conjunction with Root applications to present current and historical status information to the remote user on their web browser. The WBM services act as a portal to activity at the experimental site. In addition to HTML, java_scripts are used to mark up the results in a convenient folder schema. No special browser options are necessary on the client side. The primary source of data used by WBM is the online Oracle database; the WBM tools provide browsing and transformation functions to convert database entries into HTML tables, graphical plot representations, XML, text and Root based object output. The Root object output includes histogram objects and n-tuple data containers appropriate for download and further analysis by the user. We have devised a system of meta-data entries describing the heterogeneous database which allows the user to plot arbitrary database quantities, including multiple value versus time plots and time correlation plots. The tools can easily be extended to allow detector case-specific displays, with examples from the CMS Tracker and Hadronic Calorimeter shown.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS

Poster 1 - Board: 91 / 252

CMS Software Deployment on the Open Science Grid (OSG)

Author: Bockjoo Kim Kim¹

Co-authors: Frank Wuerthwein² ; Michael Thomas³ ; Paul Avery¹

¹ *University of Florida*

² *University of California at San Diego*

³ *Caltech*

Corresponding Author: bockjoo@phys.ufl.edu

The CMS experiment will begin data collection at the end of 2007 and released its software with new framework since the end of 2005. The CMS experiment employs a tiered distributed computing based on the Grids, the LHC Computing Grid (LCG) and the Open Science Grid (OSG). There are approximately 37 tiered CMS centers around the world. The number of the CMS software releases was three per month on average. This corresponds to roughly 100 CMS software installations.

A set of software deployment tools has been developed for the installation, verification, and deprecation of a release. The tools are mainly targeted for deployment on the OSG.

The main features are the capability of the instant release deployment and the corrective resubmission of the installation jobs. We also use an independent web-based deployment portal with the Grid security infrastructure login mechanism.

We have been deploying approximately 220 CMS softwares releases and one gLite User Interface at each site using the tools on the OSG to provide LCG-OSG interoperability. We found the tools are reliable and can be adaptable to cope with problems with changes in the Grid computing environment and the software releases. We will expand the set of sites at which we maintain CMS software deployments on the OSG

We will present the design of the tools, an analysis of statistics that we gathered during the operation of the tools, and our experience with the CMS software deployment on the OSG Grid computing environment.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS

Poster 2 - Board: 15 / 185

CMS Tier Structure and Operation of the Experiment-specific Tasks in Germany

Author: Andreas Nowack¹

¹ *III. Physikalisches Institut (B), RWTH Aachen*

Corresponding Author: andreas.nowack@cern.ch

In Germany, several university institutes and research centres take part in the CMS experiment. Concerning the data analysis, a couple of computing centres at different Tier levels, ranging from Tier 1 to Tier 3, exists at these places. The German Tier 1 centre GridKa at the research centre at Karlsruhe serves all four LHC experiments as well as for four non-LHC experiments. With respect to the CMS experiment, GridKa is mainly involved in central tasks. The Tier 2 centre in Germany consists of two sites, one at the research centre DESY at Hamburg and one at RWTH Aachen University, forming a federated Tier 2 centre. Both parts cover different aspects of a Tier 2 centre. The German Tier 3 centres are located at the research centre DESY at Hamburg, at RWTH Aachen University, and at the University of Karlsruhe. Furthermore a German user analysis facility is planned to be built.

Since the CMS community in German is rather small, a good cooperation between the different sites is essential. This cooperation includes physical topics as well as technical and operational issues. All available communication channels such as email, phone, monthly video conferences, and regular personal meetings are used. For example, the distribution of data sets is coordinated globally within Germany. Also the CMS-specific services such as the data transfer tool PhEDEx or the Monte Carlo production are operated by people from different sites in order to spread the knowledge widely and increase the redundancy in terms of operators.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

German members of CMS

Poster 1 - Board: 95 / 290

CMS Tier0 - design, implementation and first experiences

Author: Dirk Hufnagel¹

¹ *for the CMS Offline/Computing group*

Corresponding Author: dirk.hufnagel@cern.ch

With the upcoming LHC engineering run in November, the CMS Tier0 computing effort will be the one of the most important activities of the experiment. The CMS Tier0 is responsible for all data handling and processing of real data events in the first period of their life, from when the data is written by the DAQ system to a disk buffer at the CMS experiment site to when it is transferred from CERN to the Tier1 computer centers. The CMS Tier0 accomplishes three principle processing tasks: the realization of the data streaming model of CMS, the automated production of calibration and alignment constants and the first full reconstruction of the raw data. The presentation will describe the data streaming model of CMS and how this is implemented in the CMS trigger/DAQ and the Tier0. For the Tier0 this implementation means a reorganization of the data from a format determined by the demands of the CMS trigger/DAQ to a format that is determined by physics demands. We will also describe the design and implementation of the Prompt Calibration and Prompt Reconstruction workflows. The data flow underlying these workflows will be shown and first results from data challenges and scale tests will be presented.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS

Software components, tools and databases / 302

CMS packaging system or: how I learned stop worrying and love RPM spec files

Author: Giulio Eulisse¹

Co-authors: Lassi Tuura² ; Peter Elmer³ ; Stefano Argiro⁴

¹ *Northeastern University of Boston*

² *NORTHEASTERN UNIVERSITY OF BOSTON (MA) U.S.A.*

³ Princeton University

⁴ Università di Torino

Corresponding Author: giulio.eulisse@cern.ch

CMS software depends on over one hundred external packages, it's therefore obvious that being able to manage the way they are built, deployed and configured and their dependencies (both among themselves and with

respect to core CMS software) is a critical part of the system.

We present a completely new system used to build and distribute CMS software which has enabled us to go from monthly releases of the distribution kit to multiple releases during a single week, for multiple computing architectures.

The system gives full reproducibility of the building and deployment process from the build of the compiler to the actual experiment software, being it offline reconstruction, online filter farm or computing / web interfaces one.

The system is based on industry standard technologies, such as RPM and APT, with minimized custom additions to

workaround limitations or to address specific needs imposed by a large high energy physics experiment like CMS.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

Giulio Eulisse on behalf of CMS Offline group

Poster 1 - Board: 39 / 205

COOL Performance Tests and Optimization

Author: Andrea Valassi¹

¹ CERN

Corresponding Authors: marco.clemencic@cern.ch, andrea.valassi@cern.ch

The COOL software has been chosen by both Atlas and LHCb as the base of their conditions database infrastructure. The main focus of the COOL project in 2007 will be the deployment, test and validation of Oracle-based COOL database services at Tier0 and Tier1. In this context, COOL software development will concentrate on service-related issues, and in particular on the optimization of software performance for data insertion and retrieval from the database. This will involve several activities in parallel, such as improvements in the C++ client CPU consumption, improvements in the SQL query strategy for data insertion and retrieval from the database, and multi-client stress tests of different software releases and server setups. In this presentation, we will review the most important tests and optimizations which have been performed in this context.

Software components, tools and databases / 204

COOL Software Development and Service Deployment Status

Author: Andrea Valassi¹

¹ CERN

Corresponding Authors: marco.clemencic@cern.ch, andrea.valassi@cern.ch

The COOL project provides software components and tools for the handling of the LHC experiment conditions data. COOL software development is the result of a collaboration between the CERN IT Department and Atlas and LHCb, the two experiments that have chosen it as the base of their conditions database infrastructure. COOL supports persistency for several relational technologies (Oracle, MySQL and SQLite), based on the CORAL Relational Abstraction Layer. For both experiments, Oracle will be the backend used for the deployment of COOL database services at Tier0 (both online and offline) and Tier1 sites. While software development is still ongoing, especially in the area of performance optimization for data insertion and retrieval, the deployment and test of Oracle database services for COOL will be the main focus of the project in 2007. In this presentation, we will review the status and plans of both software development and service deployment of COOL database services at the time of the CHEP conference, just before LHC startup.

Distributed data analysis and information management / 314

CRAB (CMS Remote Analysis Builder)

Author: Daniele Spiga¹

¹ *Universita degli Studi di Perugia*

Corresponding Author: daniele.spiga@cern.ch

Starting from 2007 the CMS experiment will produce several Pbytes of data each year, to be distributed over many computing centers located in many different countries. The CMS computing model defines how the data are to be distributed such that CMS physicists can access them in an efficient manner in order to perform their physics analyses. CRAB (CMS Remote Analysis Builder) is a specific tool, designed and developed by the CMS collaboration, that facilitates access to the distributed data in a very transparent way. The tool's main feature is the possibility of distributing and parallelizing the local CMS batch data analysis processes over different Grid environments without any specific knowledge of the underlying computational infrastructures. More specifically CRAB allows the transparent usage of WLCG, gLite and OSG middleware. CRAB interacts with both the local user environment, with CMS Data Management services and with the Grid middleware. CRAB has been in production and in routine use by end-users since Spring 2004. It has been extensively used during studies to prepare the Physics Technical Design Report (PTDR) and in the analysis of reconstructed event samples generated during the Computing Software and Analysis Challenge (CSA06). This involved generating thousands of jobs per day at peak rates. In this work we discuss the current implementation of CRAB, experience with using it in production and plans for improvements in the immediate future.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS Offline Computing

Grid middleware and tools / 228

CRONUS: A Condor Glide-in Based ATLAS Production Executor

Author: Sanjay Padhi¹

Co-author: Rodney Walker²

¹ *University of Wisconsin-Madison*

² *TRIUMF*

Corresponding Authors: sanjay.padhi@cern.ch, rodwalker@triumf.ca

With the evolution of various Grid Technologies along with foreseen first LHC collision this year, a homogeneous and interoperable Production system for ATLAS is a necessity. We present the CRONUS, which a Condor Glide-in based ATLAS Production Executor. The Condor glide-in daemons traverse to the Worker nodes, submitted via Condor-G or gLite RB. Once activated, they preserve the Master-Worker relationships, with the worker pulling the production jobs sequentially until the expiry of their lifetimes. The initial startup glide-ins not only ensures a guaranteed ATLAS software environment but also provides a homogeneous large pools of resources across different grid flavors. The structure of CRONUS and how it handles job management, resource selection, security etc. is described. The requirement of a secondary worker from the same Glide-in daemon for DATA Transfer or any other maintenance jobs is also discussed.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Online computing / 338

Calibration workflow and dataflow in CMS

Author: Luca Malgeri¹

¹ *CERN*

Corresponding Author: luca.malgeri@cern.ch

The Calibration software framework is a crucial ingredient for all LHC experiments. In this report we shall focus on the technical challenges of this effort in the CMS experiment. It spans between careful design of the DataBase infrastructure for a quick and safe storing and retrieving of calibration constants and algorithm optimization to cope with the time and workflow constraints of High Level Triggers and prompt reconstruction of express physics streams. An overview of such aspects will be given, focusing on performance, integration and monitoring issues. Results from work and data-flow tests performed during the commissioning will be presented as well as the strategy for real data taking. As working examples, a particular emphasis on the calibration framework for the electro-magnetic calorimeter will be given.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS

Plenary / 484

Canadian Cyberinfrastructure

Online computing / 276

Cathode Strip Chamber (CSC) Raw Data Unpacking and Packing using bit field data classes

Author: Tumanov Alexander¹

Co-author: Wilkinson Rick²

¹ *T.W. Bonner Nuclear Laboratory*

² *Caltech*

Corresponding Author: tumanov@rice.edu

Unprecedented data rates that are expected at the LHC put high demand on the speed of the detector data acquisition system. The CSC subdetector located in the Muon Endcaps of the CMS detector has a data readout system equivalent in size to that of a whole Tevatron detector (60 VME crates in the CSC DAQ equal to the whole D0 DAQ size).

As a part of the HLT, the CSC data unpacking runs online and it needs to be able to cope with high data rates online. Early versions of the unpacking code used bit shifts and masks to unpack binary data. To reduce the unpacking time we decided to switch to bit field based data unpacking. The switch allowed us to gain an order of magnitude in speed. In this presentation we explain how bit field data unpacking works and why it is dramatically faster compared to conventional bit shift and mask methods.

Summary:

We would like to submit either a presentation or a poster on this topic

Poster 1 - Board: 110 / 419

Circuit Oriented End-to-End Services for HEP Data Transfers

Authors: Artur Barczyk¹ ; Dan Nae¹ ; Harvey Newman¹ ; Iosif Legrand¹ ; Sylvain Ravot¹ ; Yang Xia¹

¹ *Caltech*

Corresponding Author: artur.barczyk@cern.ch

Most of today's data networks are a mixture of packet switched and circuit switched technologies, with Ethernet/IP on the campus and in data centers, and SONET/SDH over the wide area infrastructure.

SONET/SDH allows creating dedicated circuits with bandwidth guarantees along the path, suitable for the use of aggressive transport protocols optimised for fast data transfer and without fairness constraints. On the downside, a provisioned, but under-utilised circuit may result in poor overall network utilisation, as the reserved bandwidth cannot be used by another flow. Addressing this issue, Virtual Concatenation (VCAT) and Link Capacity Adjustment Scheme (LCAS) are recent addition to SONET/SDH, and allow dynamic creation and hitless bandwidth adjustment of virtual circuits.

Caltech and CERN have deployed optical multiplexing equipment supporting VCAT/LCAS on their US LHCNet transatlantic network, and agent-based grid and network monitoring software based on Caltech's MonALISA system, to provide on-demand end-to-end bandwidth guarantees for data transfers between Tier-N centres, following the GLIF concept for control plane interaction between the participating networks.

This is being coordinated with the LHC experiments' management software for dataset distribution, and with the circuit segment-provisioning developments of ESnet, Internet2, Fermilab, BNL, GEANT2 and collaborators to form the end-to-end network paths. MonALISA is used to oversee the progress, troubleshoot and mitigate problems

associated with dynamic provisioning in response to multiple transfer requests. We present our experience with operating VCAT/LCAS enabled network for transatlantic connections, along with the details of a first implementation of circuit oriented end-to-end services for data transfers between data centres.

Collaborative tools / 35

Collaborative Tools and the LHC: An Update

Author: Steven Goldfarb¹

¹ *University of Michigan*

Corresponding Author: steven.goldfarb@cern.ch

I report on major current activities in the domain of Collaborative Tools, focusing on development for the LHC collaborations and HEP, in general, including audio and video conferencing, web archiving, and more. This presentation addresses the follow-up to the LCG RTAG 12 Final Report (presented at CHEP 2006), including the formation of the RCTF (Remote Collaboration Task Force) to steer planning and development, the installation of prototype facilities at CERN, and current funding scenarios. I also summarize presentations and discussion made during the Shaping Collaboration 2006 conference held in Geneva in December 2006, and present proposals offered by the participants to specifically address major issues facing the LHC collaborations in the coming years.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

LHC (ATLAS) **Summary:**

During the CHEP 2006 conference in Mumbai, I presented the findings and recommendations of the LCG RTAG 12 Final Report (CERN-LCG-PEB-2005-07), which was the culmination of a year of research and discussions by representatives of CERN, the LHC, and experts in the field. Since that time, a significant effort has been made at CERN to follow the recommendations, despite severe budget limitations on the part of laboratory. Efforts at CERN have included the equipping of prototype conference rooms and the development, implementation and integration of commonly used applications. Outside of CERN, significant advancements have been made in the development of new tools for video conferencing, web lecture archiving, and elsewhere. A three-day conference bringing together experts in the field of Collaborative Tools and LHC Users examined these developments, looked at future possibilities, and discussed issues most relevant to the functioning of the collaborations. My presentation will attempt to summarize these important and valuable interactions and to spark discussion on the needs of the LHC.

Online computing / 52

Commissioning of the ALICE Data-Acquisition System

Author: Sylvain Chapeland¹

Co-authors: Csaba Soós² ; Ervin Dénes³ ; Ferhat Ozok⁴ ; Filimon Roukoutakis² ; Franco Carena² ; Irina Makhlyueva² ; Klaus Schossmaier² ; Ozgur Cobanoglu⁵ ; Pierre Vande Vyvre² ; Roberto Divià² ; Sergio Vergara⁶ ; Tivadar Kiss³ ; Tome Anticic⁷ ; Ulrich Fuchs² ; Vasco Barroso² ; Wisla Carena²

¹ CERN

² CERN, Geneva, Switzerland

³ KFKI-RMKI, Budapest, Hungary

⁴ Istanbul University, Istanbul, Turkey

⁵ INFN Turin, on leave from Istanbul University, Istanbul, Turkey

⁶ Benemerita Universidad Autonoma de Puebla, Mexico

⁷ Ruder Bošković Institute, Zagreb, Croatia

Corresponding Author: sylvain.chapeland@cern.ch

ALICE (A Large Ion Collider Experiment) is the heavy-ion detector designed to study the physics of strongly interacting matter and the quark-gluon plasma at the CERN Large Hadron Collider (LHC). A large bandwidth and flexible Data Acquisition System (DAQ) has been designed and deployed to collect sufficient statistics in the short running time available per year for heavy ion and to accommodate very different requirements originated from the 18 sub-detectors. The Data Acquisition and Test Environment (DATE) is the software framework of the DAQ, handling the data from the detector electronics up to the mass storage.

This paper reviews the DAQ software and hardware architecture, including the latest features of the final design, such as the handling of the numerous calibration procedures in a common framework. We also discuss the large scale tests conducted on the real hardware to assess the standalone DAQ performances, its interfaces with the other online systems and the extensive commissioning performed in order to be fully prepared for physics data taking scheduled to start in November 2007. The test protocols followed to integrate and validate each sub-detector with DAQ and Trigger hardware synchronized by the Experiment Control System are described. Finally, we give an overview of the experiment logbook, and some operational aspects of the deployment of our computing facilities.

The implementation of a Transient Data Storage able to cope with the 1.25 GB/s recorded by the event-building machines is covered in a separate paper.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ALICE

Event processing / 131

Commissioning of the ATLAS Inner Detector with cosmic rays

Author: Christian Schmitt¹

¹ CERN

Corresponding Authors: helen.hayward@cern.ch, christian.schmitt@cern.ch

The inner detector of the ATLAS experiment is in the process of being commissioned using cosmic ray events. First tests were performed in the SR1 assembly hall at CERN with both barrel and endcaps for all different detector technologies (pixels and microstrips silicon detectors as well as straw tubes with additional transition radiation

detection). Integration with the rest of the ATLAS sub-detectors is now being done in the ATLAS cavern.

The full software chain has been set up in order to reconstruct and analyse this kind of events. Final detector decoders have been developed, different pattern recognition algorithms and track fitters have been validated as well as the various alignment and calibration methods. The infrastructure to deal with conditions data coming from the data acquisition, detector control system and calibration runs has been put in place, allowing also to apply alignment and calibration constants.

The software has also been essential to monitor the detector performance during data taking. Detector efficiencies, noise occupancies and resolutions have been studied in detail and compared with those obtained from simulation.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS Offline Computing

Event processing / 128

Commissioning of the ATLAS offline software with cosmic rays

Author: Maria Jose Costa¹

¹ *Instituto de Fisica Corpuscular (IFIC) UV-CSIC*

Corresponding Authors: haleh.hadavand@cern.ch, maria.jose.costa@cern.ch

The ATLAS experiment of the LHC is now taking its first data by collecting cosmic ray events. The full reconstruction chain including all sub-systems (inner detector, calorimeters and muon spectrometer) is being commissioned with this kind of data for the first time.

Specific adaptations to deal with particles not coming from the interaction point and not synchronized with the readout clock were needed. Data decoders and the infrastructure to deal with conditions data as those coming from the data acquisition configuration, detector control system, calibration and alignment corrections were developed and validated as well.

Detailed analysis are being performed in order to provide ATLAS with its first alignment and calibration constants and to study the combined muon performance. Combined monitoring tools and event displays have also been developed to ensure the good data quality.

A simulation of cosmic events according to the different detector and trigger setups has also been provided to verify it gives a good description of the data.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Poster 1 - Board: 13 / 330

Commissioning with cosmic rays of the Muon Spectrometer of the ATLAS experiment at the Large Hadron Collider

Author: Rosy Nikolaidou¹

¹ DAPNIA

Corresponding Author: rosy.nicolaidou@cern.ch

The Muon Spectrometer of the ATLAS experiment is made of a large toroidal magnet, arrays of high-pressure drift tubes for precise tracking and dedicated fast detectors for the first-level trigger. All the detectors in the barrel toroid have been installed and commissioning has started with cosmic rays. These detectors are arranged in three concentric rings and the total area is about 7000 square meters. During the installation and commissioning of the detectors, data are usually taken with the magnet off, but a dedicated run took place with the magnetic field of the barrel toroid turned on.

We present the procedure to control the response of the single detectors installed in the barrel toroid, Monitored Drift Tubes and Resistive Plate Chambers, and results of the first tests done with cosmic rays triggered by the first-level processor and read-out through the ATLAS data acquisition. A comparison of the detector performance in magnetic field on and off will be presented together with a measurement of the cosmic ray flux in the underground experimental area.

Details on the installation and commissioning schedule will be given in view of the completion of the instrumentation of the muon spectrometer for the first period of data taking with proton-proton collisions.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

Atlas muon group

Poster 2 - Board: 77 / 173

Complete Distributed Computing Environment for a HEP Experiment: Experience with ARC-Connected Infrastructure for ATLAS

Authors: Adrian Taga¹; Alexander Read¹; Bjørn Hallvard Samset¹; David Cameron¹; Farid Ould-Saada¹; Katarina Pajchel¹

Co-authors: Andrej Filipcic²; Mattias Ellert³; Sigve Haug⁴

¹ University of Oslo, Department of Physics

² Jozef Stefan Institute, Dept. for Experimental Particle Physics

³ University of Uppsala, Department of nuclear and particle physics

⁴ University of Bern

Corresponding Author: alexander.lincoln.read@cern.ch

Computing and storage resources connected by the Nordugrid ARC middleware in the Nordic countries, Switzerland and Slovenia are a part of the ATLAS computing grid. This infrastructure is being commissioned with the ongoing ATLAS Monte Carlo simulation production in preparation for the commencement of data taking in late 2007. The unique non-intrusive architecture of ARC, its straightforward interplay with the ATLAS Production System via the Dulcinea executor, and its performance during the commissioning exercise will be described. ARC support for flexible and powerful

end-user analysis within the GANGA distributed analysis framework will be shown. Whereas the storage solution for this grid was earlier based on a large, distributed collection of gridftp-servers, the ATLAS computing design includes a structured SRM-based system with a limited number of storage endpoints. The characteristics, integration and performance of the old and new storage solutions are described. Although the hardware resources in this grid are quite modest, it has provided more than double the agreed contribution to the ATLAS production with an efficiency above 95% during long periods of stable operation. If time allows, the full chain will be demonstrated.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

Particle Physics group, University of Oslo

Poster 1 - Board: 57 / 439

Computationally efficient algorithms for the two-dimensional Kolmogorov-Smirnov test

Authors: Ivan D. Reid¹ ; Peter R. Hobson¹ ; Raul H. C. Lopes¹

¹ *School of Design and Engineering - Brunel University, UK*

Corresponding Authors: ivan.reid@brunel.ac.uk, raul.lopes@brunel.ac.uk

Goodness-of-fit statistics measure the compatibility of random samples against some theoretical probability distribution function. The classical one-dimensional Kolmogorov-Smirnov test is a non-parametric statistic for comparing two empirical distributions, which defines the largest absolute difference between the two cumulative probability distribution functions as a measure of disagreement. Adapting this test to more than one dimension is a challenge because there are $2^k - 1$ independent ways of defining a cumulative probability distribution function when k dimensions are involved. However there are many applications in experimental physics where comparing two-dimensional data sets is important.

In this paper we discuss Peacock's version [3] of the Kolmogorov-Smirnov test for two-dimensional data sets which computes the differences between cumulative probability distribution functions in $4n^2$ quadrants and runs in $O(n^3)$, for a sample of size n . We also discuss Fasano and Franceschini's variation [2] of Peacock's test that runs in $O(n^2)$, Cooke's algorithm [1] for Peacock's test and ROOT's version of the two-dimensional Kolmogorov-Smirnov.

We establish a lower-bound limit on the work for computing Peacock's test in $W(n^2 \lg n)$, which contradicts Cooke's claim that it possible to perform this test with an $O(n \lg n)$ algorithm. We also establish a lower-bound limit of $W(n \lg n)$ for Fasano and Franceschini's test, and present an optimal sequential algorithm for it.

We finally discuss experimental results comparing two parallel algorithms implementing Peacock's test and an optimal algorithm for Fasano and Franceschini's test, and contrast these with the algorithm in ROOT which is based on the calculation of a mean of two one-dimensional Kolmogorov-Smirnov tests.

References

[1]

A. Cooke. The muac algorithm for solving 2d ks test.
<http://www.acooke.org/jara/muac/algorithm.html>.

[2]

G. Fasano and A. Franceschini. A multidimensional of the Kolmogorov-Smirnov test. Monthly Notices Royal Astronomy Society, 225:155-170, 1987.

[3]

J. A. Peacock. Two-dimensional goodness-of-fit testing in astronomy. Monthly Notices Royal Astronomy Society, 202:615-627, 1983.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS

Computer facilities, production grids and networking / 281

Computing Operations at CMS Facilities

Author: Daniele Bonacorsi¹

¹ *INFN-CNAF, Bologna, Italy*

Corresponding Author: bonacorsi@bo.infn.it

The CMS experiment is gaining experience towards the data taking in several computing preparation activities, and a roadmap towards a mature computing operations model stands as a primary target. The responsibility of the Computing Operations projects in the complex CMS computing environment spawns a wide area and aims at integrating the management of the CMS Facilities Infrastructure, so to provide and maintain a working distributed computing fabric with a consistent environment for the Data Operations and the users. Coordination of Operations at facilities (Tier-0, CMS CERN Analysis Facility, WLCG Tiers), resource needs planning and tracking, coordination of operations and liaison to external projects and organisations (WLCG and OSG) are involved in this process, and are described in this paper. A review of the practical experience by CMS in the coordinated operations effort across many regional centers is presented, and the lessons learned in the summer 2007 CMS data challenge (CSA07) are discussed.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS Computing group

Distributed data analysis and information management / 23

Computing and Ground Data Handling for AMS-02 Mission

Author: Vitaly Choutko¹

¹ *Massachusetts Institute of Technology (MIT)*

Corresponding Author: vitaly.choutko@cern.ch

The AMS-02 detector will be installed on ISS ifor at least 3 years. The data will be transmitted from ISS to NASA Marshall Space Flight Center (MSFC, Huntsville, Alabama) and transfered to CERN (Geneva Switzerland) for processing and analysis. We are presenting the AMS-02 Ground Data Handling scenario and requirements to AMS ground centers: the Payload Operation and Control Center (POCC) and the Science Operation Center (SOC). The Payload Operation and Control Center is where AMS operations take place, including commanding, storage and analysis of house keeping data and partial science data analysis for rapid quality control and feed back. The AMS Science Data Center receives and stores all AMS science and house keeping data, as well as ancillary data from NASA. It ensures full science data reconstruction, calibration and alignment; it keeps data available for physics analysis and archives

all data.

We also discuss the AMS-02 distributed data management between 25 Universities and Labs in Europe, USA and Asia.

Poster 1 - Board: 22 / 463

Computing in IceCube

Author: Georges Kohnen¹

¹ *Université de Mons-hainaut*

Corresponding Author: kohnen@umh.ac.be

The IceCube neutrino telescope is a cubic kilometer Cherenkov detector currently under construction in the deep ice at the geographic South Pole. As of 2007, it has reached more than 25 % of its final instrumented volume and is actively taking data. We will briefly describe the design and current status, as well as the physics goals of the detector. The main focus will, however, be on the unique computing structure of the experiment, due to the remoteness of the detector and the limited access to satellites for data transfer. We will also outline the structure of the software framework used for filtering, reconstruction and simulation.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

IceCube

Event processing / 192

Concepts, Design and Implementation of the New ATLAS Track Reconstruction (NEWT)

Author: Andreas Salzburger¹

Co-author: Markus Elsing²

¹ *University of Innsbruck & CERN*

² *CERN*

Corresponding Author: andreas.salzburger@cern.ch

The track reconstruction of modern high energy physics experiments is a very complex task that puts stringent requirements onto the software realisation. The ATLAS track reconstruction software has been in the past dominated by a collection of individual packages, each of which incorporating a different intrinsic event data model, different data flow sequences and calibration data. The ATLAS track reconstruction has undergone a major design revolution to ensure maintainability during the long lifetime of the ATLAS experiment and the flexibility needed for the startup phase. The entire software chain has been re-organised in modular components and a common Event Data Model has been deployed during the last three years. A complete new track reconstruction that concentrates on common tools aimed to be used by both ATLAS tracking devices, the Inner Detector and the Muon System, has been established. The common components approach has been extended to cover the tracking part of the highest level software-based trigger, the ATLAS Event Filter. The New Tracking has been already used during many large scale tests with data from Monte Carlo simulation and from detector commissioning projects such as the combined test beam 2004 and cosmic ray events. The design, concepts and implementation of the newly developed track reconstruction will be presented and overview on the performance for various different applications will be given.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):**ATLAS Offline Computing Summary:**

The ATLAS detector consists of two independent tracking devices: the Inner Detector (ID) close to the interaction region and the Muon System (MS). While the Inner Detector reconstruction has to deal with the high track density that imposes a large number of combinatorial track candidates, the Muon System track reconstruction is mainly limited by the huge amount of inert material, the cavern background and the highly inhomogeneous magnetic field.

In the past - especially during the design phase of the ATLAS experiment - the ATLAS track reconstruction software consisted for both, ID and MS, of several competing reconstruction programs, each of which incorporating its own event data model, different reconstruction geometries, varying concepts in material integration and separate philosophies in algorithmic sequence and steering. A new track reconstruction (NEWT) has been deployed that is based on a common Event Data Model and the definition of interfaces that represent the various tasks (of different hierarchical level) of the reconstruction application. It is designed in a component pattern structure that eases the integration of new developed track reconstruction modules and guarantees maintainability during the long lifetime of the ATLAS experiment. Being just half a year before the startup of the ATLAS experiment, a review of NEWT in both design and performance will be presented.

Plenary / 479

Conference Closeout

Plenary / 478

Conference Summary

Corresponding Author: matthias.kasemann@cern.ch

Poster 1 - Board: 109 / 410

Construct an LHC scale national analysis facility

Author: Martin Gasthuber¹

¹ *Deutsches Elektronen Synchrotron (DESY)*

Based on today's understanding of LHC scale analysis requirements and the clear dominance of fast and high capacity random access storage, this talk will present a generic architecture for a national facility based on existing components from various computing domains. The following key areas will be discussed in detail and solutions will be proposed, building the overall architecture.

1. large scale cluster filesystems known in the HPC community. This community has shown the mature character of this technologies during the last months and evidently this success could be made usable for the demanding HEP community.

2. methods and systems for decentralized/delegated user administration. This also includes the integration with VOMS to allow seamless management with existing system.
3. data access to/from T2 (generally TierX) facilities by using SRM enabled cluster filesystems and high speed data access to local TierX storage resources if available.
4. batch process integration presenting exactly the same interface/behaviour to users with respect to their interactive ones.
5. accounting and monitoring system to enable true sharing of the single resource between all participating VOs.
6. criteria showing the scaling properties in the PB data region and more than 1000 CPUs.

Poster 1 - Board: 41 / 236

Control and monitoring of alignment data for the ATLAS endcap Muon Spectrometer at the LHC

Author: Craig Dowell¹

¹ *Univ. of Washington*

The ATLAS Muon Spectrometer is constructed out of 1200 drift tube chambers with a total area of nearly 7000 square meters. It must determine muon track positions to a very high precision despite its large size necessitating complex real-time alignment measurements. Each chamber, as well as approximately 50 alignment reference bars in the endcap region, are equipped with CCD cameras, laser sources, and LED-illuminated masks which optically link chambers and bars in a three dimensional grid. This permits micron-level determination of chamber-to-chamber positions and chamber distortions. This information is used to correct drift tube positions and shape for muon track reconstruction. The endcap optical system produces about 8000 83 kB images during each 20 minute readout cycle. The optical data acquisition and image analysis are performed by a hardware/software system (LWDAQ) developed at Brandeis University. The system is segmented so that six processes running on several computers perform the optical readout and image analysis in parallel. We describe the architecture and implementation of the control system; monitoring of the optical readout processes; evaluation of the validity of images; display of results, validity parameters, and error conditions; and storage of analysis results and quality in an Oracle database. The distributed control architecture includes a Linux-based control and communication process and a PVSS SCADA system for the user interface, display functions, and database storage. Details of the architecture, communications, and performance will be presented.

Poster 1 - Board: 99 / 309

Creating End-to-End Guaranteed Bandwidth Network Paths Across Multiple Domains with TeraPaths

Author: Dimitrios Katramatos¹

Co-authors: Dantong Yu¹; Jay Packard¹; John DeStefano¹; Michael Ernst¹; Shawn McKee²

¹ Brookhaven National Laboratory

² University of Michigan

Corresponding Author: dkat@bnl.gov

Supporting reliable, predictable, and efficient global movement of data in high-energy physics distributed computing environments requires the capability to provide guaranteed bandwidth to selected data flows and schedule network usage appropriately. The DOE-funded TeraPaths project at Brookhaven National Laboratory (BNL), currently in its second year, is developing methods and tools that enable scientists to reserve network bandwidth for specific time windows and dedicate this bandwidth to important data transfers. The TeraPaths software does this through the creation of an end-to-end (or source computer host to destination computer host) virtual network path with guaranteed bandwidth for the duration of a data transfer. The path is set up through the direct configuration of network devices within end-site LANs and indirect configuration of WAN network devices through the automated invocation of WAN provider services. The software accommodates the end user with an easy to use and secure web interface, as well as an API, for submitting reservations. An ongoing development effort will provide access to TeraPaths services from within popular data transfer tools, and enlarge the current set of supported network devices. With the collaboration of ESnet and Internet 2 development teams, the TeraPaths-capable infrastructure, which currently extends from BNL to the University of Michigan, will be expanded to US ATLAS Tier 2 sites and beyond, with the goal of creating a production-quality data transfer environment for the benefit of the high-energy physics community.

Grid middleware and tools / 401

Cross middlewares Virtual Organization Authorization

Authors: Andrea Ceccanti¹; Antonia Ghiselli¹; Federico Stagni¹; Valerio Venturi¹; Vincenzo Ciaschini¹

¹ INFN

Corresponding Author: valerio.venturi@cnaf.infn.it

The Virtual Organization Membership Service (VOMS) is a system for managing users in a Virtual Organization. It manages and releases user's information such as group membership, roles, and other authorization data. VOMS was born with the aim of supporting dynamic, fine grained, and multi-stakeholder access control to enable coordinate sharing in virtual organizations.

The current software releases Attribute Certificates (ACs) conforming to RFC 3821. In the most adopted use pattern, ACs are embedded in proxy certificates. This proved to be a very convenient way of making user's attributes available for driving authorization of grid services. In these years VOMS has established as one of the main tools for authorization on two of the major grid infrastructure (EGEE, OSG) and as a central component in the respective grid middlewares (gLite, VDT). VOMS is also supported by GT4 Authorization framework.

In the last years, the Security Assertion Markup Language (SAML) has emerged as a central standard in the field of Web Services security. We are extending VOMS to provide SAML support. This is going to make VOMS based authorization available on a

larger number of grid middlewares, and especially on those which don't use proxy certificates. Following this, within the OMII-Europe project, UNICORE will integrate VOMS. Support for SAML is also going to make interoperability with Shibboleth easier. The final aim is to provide VO's with authorization tools that are consistent and homogeneous across different grid middlewares and infrastructures.

Poster 2 - Board: 115 / 177

DIRAC Agents and Services

Authors: Adrian Casajus Ramo¹ ; Ricardo Graciani Diaz¹

¹ *Universidad de Barcelona*

Corresponding Author: graciani@ecm.ub.es

DIRAC Services and Agents are defined in the context of the DIRAC system (the LHCb's Grid Workload and Data Management system), and how they cooperate to build functional sub-systems is presented. How the Services and Agents are built from the low level DIRAC framework tools is described.

Practical experience in the LHCb production system has directed the creation of the current DIRAC framework tools. The framework has been designed to ease the development and deployment of DIRAC components, and to provide a certain level of protection against external misuse. Guidelines for service robustness and real-time response used currently in DIRAC will be discussed.

Measurements on the current performance of the running DIRAC system from the point of view of the functionality provided by the framework are presented.

Grid middleware and tools / 168

DIRAC Data Management: consistency, integrity and coherence of data

Authors: Andrew Cameron Smith¹ ; Marianne Bargiotti¹

¹ *European Organization for Nuclear Research (CERN)*

Corresponding Author: marianne.bargiotti@cern.ch

The DIRAC Data Management System (DMS) relies on both WLCG Data Management services (LCG File Catalogues, Storage Resource Managers and FTS) and LHCb specific components (Bookkeeping Metadata File Catalogue).

The complexity of both the DMS and its interactions with numerous WLCG components as well as the instability of facilities concerned, has turned frequently into unexpected problems in data moving and/or data registration, preventing to have a coherent picture of datasets.

Several developments in LHCb have been done in order to avoid data corruptions, data missing, data incoherence and inconsistencies among Catalogues and physical storages both through safety measures at data management level (failover mechanisms, check sums, roll back mechanisms) and extensive background checks.

In this paper all the tools developed for checking data integrity and consistency will be presented, as well as a Storage Resource Checker, whose aim is to produce an up-to-date accounting of all LHCb storage usage using the LFC mirror database. Goal of this activity is the development of a generic tool suite able to categorize, analyze and systematically cure the disparate problems affecting DM in order to

maintain a consistent picture of the main catalogues (Bookkeeping and LFC) and the Storage Elements.

Poster 2 - Board: 114 / 178

DIRAC Framework for Distributed Computing

Authors: Adrian Casajus Ramo¹ ; Ricardo Graciani Diaz¹

¹ *Universitat de Barcelona*

Corresponding Author: adria@ecm.ub.es

The DIRAC system is made of a number of cooperating Services and Agents that interact between them with a Client-Server architecture. All DIRAC components rely on a low level framework that provides the necessary basic functionality. In the current version of DIRAC these components have been identified as: DISET, the secure communication protocol for remote procedure call and file transfer; Configuration System, providing redundant distributed mechanism for configuration and service discovery; Logging and Monitoring System, a uniform way for all components to report their status and activities, and present that information to the users. The current functionality is the result of the experience collected during the last years of running with DIRAC for LHCb. The design principles and the resulting architecture currently implemented in DIRAC will be presented.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

LHCb

Poster 2 - Board: 118 / 340

DIRAC Job Prioritization and Fair Share in LHCb

Author: Gianluca Castellani¹

¹ *European Organization for Nuclear Research (CERN)*

Corresponding Author: gianluca.castellani@cern.ch

LHCb accesses Grid through DIRAC, its WorkLoad and Data Management system. In DIRAC all the jobs are stored in central task queues and then pulled onto worker nodes via generic Grid jobs called Pilot Agents. These task queues are characterized by different requirements about CPUtime and destination. Because the whole LHCb community is divided in sets of physicists, developers, production and software managers it is important to assign different priorities to different jobs. A special DIRAC agent is responsible for dynamic calculation of the job priorities. The goals of the job priority agent are to ensure that the whole LHCb community fairly uses and shares the Grid resources and to minimize the overall time spent by the high priority jobs in the DIRAC task queues. In the paper the possible technical approaches to define the job priority are evaluated. In particular, the use of the MAUI scheduler is studied. The results of the application of the proposed job prioritization procedure to real jobs are presented. The current limitations and future work are discussed.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

LHCb

Grid middleware and tools / 174**DIRAC Optimized Workload Management****Author:** Stuart Paterson¹¹ CERN**Corresponding Author:** stuart.paterson@cern.ch

The LHCb DIRAC Workload and Data Management System employs advanced optimization techniques in order to dynamically allocate resources. The paradigms realized by DIRAC, such as late binding through the Pilot Agent approach, have proven to be highly successful. For example, this has allowed the principles of workload management to be applied not only at the time of user job submission to the Grid but also to optimize the use of computing resources once jobs have been acquired. Along with the central application of job priorities, DIRAC minimizes the system response time for high priority tasks. This paper will describe the recent developments to support Monte Carlo simulation, data processing and distributed user analysis in a consistent way across disparate compute resources including individual PCs, local batch systems, and the Worldwide LHC Computing Grid. The Grid environment is inherently unpredictable and whilst short-term studies have proven to deliver high job efficiencies, the system performance over an extended period of time will be considered here in order to convey the experience gained so far.

Poster 2 - Board: 113 / 189**DIRAC: A community grid solution**

Authors: Adrian CASAJUS RAMO¹ ; Andrei Tsaregorodtsev² ; Andrew Cameron SMITH³ ; Carmine CIOFFI⁴ ; Gennady KUZNETSOV⁵ ; Gianluca CASTELLANI³ ; Joel CLOSIER³ ; Marcos SECO MIGUELEZ⁶ ; Marianne BARGIOTTI³ ; Nicolas BROOK⁷ ; Philippe CHARPENTIER³ ; Raja NANDAKUMAR⁵ ; Ricardo GRACIANI-DIAZ¹ ; Roberto SANTINELLI³ ; Sergio GOMEZ JIMENEZ⁸ ; Stuart PATERSON³

¹ *University of Barcelona*² *CNRS-IN2P3-CPPM, Marseille*³ *CERN*⁴ *Oxford University*⁵ *Rutherford Appleton Laboratory*⁶ *University of Santiago de Compostela*⁷ *H. H. Wills Physics Laboratory, Bristol*⁸ *University Rovira i Virgili***Corresponding Author:** atsareg@in2p3.fr

The DIRAC system was developed in order to provide a complete solution for using distributed computing resources of the LHCb experiment at CERN for data production and analysis. It allows a concurrent use of over 10K CPUs and 10M file replicas distributed over many tens of sites. The sites can be part of a computing grid such as WLCG or standalone computing clusters all integrated in a single management structure.

DIRAC is a generic system with the LHCb specific functionality incorporated through a number of plug-in modules. It can be easily adapted to the needs of other communities. A special attention is paid to the resilience of the DIRAC components to allow an efficient use of non-reliable resources. The DIRAC production management components provide a framework for building

highly automated data production systems including data distribution and data driven workload scheduling.

In this paper we give an overview of the DIRAC system architecture and design choices. We show how different components are put together to compose an integrated data processing system including all the aspects of the LHCb experiment - from the MC production and raw data reconstruction to the final user analysis.

Distributed data analysis and information management / 194

DIRAC: Data Production Management

Author: Andrew Cameron Smith¹

¹ *CERN*

Corresponding Author: andrew.cameron.smith@cern.ch

The LHCb Computing Model describes the dataflow model for all stages in the processing of real and simulated events and defines the role of LHCb associated Tier1 and Tier2 computing centres. The WLCG ‘dressed rehearsal’ exercise aims to allow LHC experiments to deploy the full chain of their Computing Models, making use of all underlying WLCG services and resources, in preparation for real data taking. During this exercise simulated RAW physics data, matching the properties of eventual real data, will be uploaded from the LHCb Online storage system to Grid enabled storage. This data will then be replicated to LHCb Tier1s and subsequently processed (reconstructed and stripped). The product of this processing is user analysis data that are distributed to all LHCb Tier1 sites. DIRAC, LHCb’s Workload and Data Management System, supports the implementation of the Computing Model in a data driven, real time and coordinated fashion.

In this paper the LHCb Computing Model will be reviewed and the DIRAC components providing the needed functionality to support the Computing Model will be detailed. The experience gained during WLCG’s ‘dressed rehearsal’ exercise will be given along with an evaluation of the preparedness for real data taking.

Poster 2 - Board: 116 / 195

DIRAC: Reliable Data Management for LHCb

Author: Andrew Cameron Smith¹

¹ *CERN*

Corresponding Author: andrew.cameron.smith@cern.ch

DIRAC, LHCb’s Grid Workload and Data Management System, utilises WLCG resources and middleware components to perform distributed computing tasks satisfying LHCb’s Computing Model. The Data Management System (DMS) handles data transfer and data access within LHCb. Its scope ranges from the output of the LHCb Online system to Grid-enabled storage for all data types. It supports metadata for these files in replica and bookkeeping catalogues, allowing dataset selection and localisation. The DMS controls the movement of files in a redundant fashion whilst providing utilities for accessing all metadata. To do these tasks effectively the DMS requires complete self integrity between its components and external physical storage. The DMS

provides highly redundant management of all LHCb data to leverage available storage resources and to manage transient errors in underlying services. It provides data driven and reliable distribution of files as well as reliable job output upload, utilising VO Boxes at LHCb Tier1 sites to prevent data loss.

In this paper the evolution of the DIRAC Data Management System will be presented highlighting successful design choices and limitations discovered.

Poster 2 - Board: 96 / 389

DPM Status and Next Steps

Authors: Akos Frohner¹ ; Jean-Philippe Baud¹ ; Krzysztof Nienartowicz¹ ; Lana Abadie¹ ; Remi Mollon¹ ; Smith David¹ ; Sophie Lemaitre¹

¹ CERN

Corresponding Authors: lana.abadie@cern.ch, akos.frohner@cern.ch

The DPM (Disk Pool Manager) provides a lightweight and scalable managed disk storage system. In this paper, we describe the new features of the DPM.

It is integrated in the grid middleware and is compatible with both VOMS and grid proxies. Besides the primary/secondary groups (or roles), the DPM supports ACLs adding more flexibility in setting file permissions.

Tools to interact with the DPM at different levels have been extended so that site managers can more dynamically configure and manage their DPM in a consistent way. In addition to rfio and gsiftp, users can now use the xrootd and https protocols to access the DPM.

A new version of Storage Resource Manager (SRM) interface, v2.2 has been implemented. One of the novelties is the reserve space concept, useful to guarantee space for a specific user or a group during a given period of time.

DPM has been deployed in roughly 80 Tier-2 sites and in several medical institutes. Unlike physics data, medical data is very sensitive. The DPM will offer the possibility to encrypt data throughout the process in a very secure way by implementing a key-distributed system.

Performance has been improved by the use of bulk queries. Stressing tests have shown a good robustness of the DPM against concurrent accesses.

Grid middleware and tools / 227

DZero Data-Intensive Computing on the Open Science Grid

Author: Gabriele Garzoglio¹

Co-authors: Andreii Baranovski¹ ; Brad Abbot² ; Mike Diesburg¹ ; Parag Mhashilkar¹ ; Tibor Kurca³

¹ FERMIONAL ACCELERATOR LABORATORY

² University of Oklahoma

³ IN2P3

Corresponding Authors: cope@fnal.gov, garzoglio@fnal.gov

High energy physics experiments periodically reprocess data, in order to take advantage of improved understanding of the detector and the data processing code. Between February and May 2007, the DZero experiment will reprocess a substantial fraction of its dataset. This consists of half a billion events, corresponding to more than 100 TB of data, organized in 300,000 files.

The activity utilizes resources from sites around the world, including a dozen sites participating to the Open Science Grid consortium (OSG). About 1,500 jobs are run every day across the OSG, consuming and producing hundreds of Gigabytes of data. OSG computing and storage resources are coordinated by the SAM-Grid system. This system organizes job access to a complex topology of data queues and job scheduling to clusters, using a SAM-Grid to OSG job forwarding infrastructure.

For the first time in the lifetime of the experiment, a data intensive production activity is managed on a general purpose grid, such as OSG. This paper describes the implications of using OSG, where all resources are granted following an opportunistic model, the challenges of operating a data intensive activity over such large computing infrastructure, and the lesson learned throughout the few months of the project.

Poster 1 - Board: 7 / 158

Data Acquisition Backbone Core DABC

Author: Hans G. Essel¹

Co-authors: Jörn Adamczewski¹; Nikolaus Kurz¹; Serguey Linev¹

¹ GSI

Corresponding Author: h.essel@gsi.de

European FP6 program “HadronPhysics”, JRA1 “FutureDAQ” contract number RII3-CT-2004-506078)

For the new experiments at FAIR like CBM new concepts of data acquisition systems have to be developed like the distribution of self-triggered, time stamped data streams over high performance networks for event building. The DAQ backbone DABC is designed for FAIR detector tests, readout components test, data flow investigations, and DAQ controls. All kinds of data channels (front-end systems) are connected by plug-ins into functional components of DABC like data input, combiner, scheduler, event builder, analysis and storage. Several software packages (uDAPL, OpenFabric IB verbs, MPI2) have been used to measure the performance of InfiniBand on a cluster. One can achieve about 80% of the bidirectional nominal IB bandwidth (measured with up to 22 nodes).

The XDAQ package developed for the CMS experiment has been chosen as infrastructure for DABC. The IB transport has been implemented in XDAQ. Measurements showed that the XDAQ transport causes too much overhead. Therefore a new faster (zero copy) transport layer is implemented. To be flexible in the selection of DAQ controls a DIM server has been implemented connecting the XDAQ InfoSpace to arbitrary DIM clients. For EPICS and LabView such clients are implemented.

Summary:

General purpose DAQ backbone for high speed event building over fast networks, i.e. InfiniBand, supports synchronized or self-triggered data channels. Data flow engine uses application plug-ins for custom front-end systems.

Plenary / 481

Data Acquisition at the LHC experiments

Corresponding Author: sylvain.chapeland@cern.ch

The CERN Large Hadron Collider (LHC) is one of the most awesome science tool ever built. To fully exploit the potential of this great instrument, a huge design and development effort has been initiated in order to ensure that measurements can optimally flow out from the detectors in terms of quantity, selectivity, and integrity, be accessible for online monitoring and be recorded for analysis and long-term archive.

This effort is now reaching the end of its initial development phase and evolving towards operation. We will give an overview of the resulting Trigger and Data Acquisition (DAQ) systems designed to harvest and store the precious stream of LHC data. We will in particular review some of the technology choices made to address the specific requirements of each experiment, covering both hardware and software aspects.

Poster 2 - Board: 11 / 88

Data Analysis through the DIANA Meta-Scheduling Approach

Author: Richard McClatchey¹

Co-authors: Ashiq Anjum¹ ; Ian Willers²

¹ UWE

² CERN

Corresponding Author: richard.mcclatchey@cern.ch

We introduce the concept, design and deployment of the DIANA meta-scheduling approach to solving the challenge of the data analysis being faced by the CERN experiments. The DIANA meta-scheduler supports data intensive bulk scheduling, is network aware and follows a policy centric meta-scheduling that will be explained in some detail.

In this paper, we describe a Physics analysis case study using the DIANA meta-scheduler. We demonstrate that a decentralized and dynamic meta-scheduling approach is an effective strategy to cope with the increasing numbers of users, jobs and datasets. We present “quality of service” related statistics for the physics analysis through the application of a policy centric fair-share scheduling model. The DIANA meta-schedulers create a peer-to-peer hierarchy of schedulers to accomplish resource management. They employ scheduling approaches that are able to change with evolving loads and adapt following the dynamic and volatile nature of the resources. The DIANA meta-scheduler also acknowledges the important role of networks in the modern day distributed systems and treats them an equally important resource in the scheduling decisions along with the compute and data resources.

This topic formed the contents of a Ph.D. thesis done partly at CERN and partly at the University of the West of England in the United Kingdom. The Physics analysis case study was based on the computing model that was presented in the CMS Technical Design Report. Our results show that Physics analysis scheduling can be made robust by employing fault tolerant and decentralized meta-scheduling.

Poster 2 - Board: 64 / 434

Data Quality Monitoring and Visualization for the CMS Silicon Strip Tracker

Author: Domenico Giordano¹

¹ *Dipartimento Interateneo di Fisica*

Corresponding Author: domenico.giordano@cern.ch

The CMS Silicon Strip Tracker (SST), consisting of more than 10 millions of channels, is organized in about 16,000 detector modules and it is the largest silicon strip tracker ever built for high energy physics experiments.

In the first half of 2007 the CMS SST project is facing the important milestone of commissioning and testing a quarter of the entire SST with cosmic muons.

The full standard CMS software is deployed for the data acquisition, reconstruction, monitoring and event display.

For the first time the detector performance is monitored using an advanced Data Quality Monitoring (DQM) system capable of running on a variety of online and offline environments, in the control room as well as in remote sites.

More than 100.000 monitorable quantities are managed by the DQM system that organizes them in a hierarchical structure reflecting the detector arrangement in subcomponents and the various levels of data processing.

Monitorable quantities computed at the level of individual detectors are processed to extract automatic quality checks and summary results that can be visualized with specialized graphical user interfaces.

In view of the great complexity of the CMS Tracker detector the standard visualization tools based on histograms have been complemented with 2 and 3 dimensional graphical images of the subdetector that can show the whole detector down to single channel resolution.

The functionalities of the CMS Silicon Strip Tracker DQM system and the experience acquired during the SST commissioning are discussed here.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS

Poster 1 - Board: 10 / 268

Data Quality Monitoring for the CMS Electromagnetic Calorimeter

Authors: Alessio Ghezzi¹ ; Benigno Gobbo² ; Giovanni Franzoni³ ; Giuseppe Della Ricca⁴

¹ *Universita' and INFN, Milano Bicocca & ETH Zurich*

² *INFN, Trieste*

³ *University of Minnesota*

⁴ *Univ. of Trieste and INFN*

Corresponding Author: giuseppe.della-ricca@cern.ch

The electromagnetic calorimeter of the Compact Muon Solenoid experiment will play a central role in the achievement of the full physics performance of the detector at the LHC. The detector performance will be monitored using applications based on the CMS Data Quality Monitoring (DQM) framework and running on the High-Level Trigger Farm as well as on local DAQ systems. The monitorable quantities are organized into hierarchical structures based on the physics content. The information produced is delivered to client applications according to their subscription requests. The client

applications process the received quantities, according to pre-defined analyses, thus making the results immediately available, and store the results in a database, and in the form of static web pages, for subsequent studies. We describe here the functionalities of the CMS ECAL DQM applications and report about their use in a real environment. In particular we detail the usage of the DQM during the data collection campaigns at the 2006 electron calibration test beams, at the cosmic muon calibration stand (2005-2007), at the CMS slice test (2006 Magnet Test and Cosmic Challenge), and during the installation and commissioning of the calorimeter in the CMS experimental area.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

on behalf of the CMS Electromagnetic Calorimeter Group

Poster 1 - Board: 5 / 138

Data Stream handling in the LHCb experiment

Author: Niko Neufeld¹

¹ *CERN*

Corresponding Author: niko.neufeld@cern.ch

Events selected by LHCb's online event filtering farm will be assembled into raw data files of about 2 GBs. Under nominal conditions about 2 such files will be produced per minute. These files must be copied to tape storage and made available online to various calibration and monitoring tasks. The life cycle and state transitions of each files are managed by means of a dedicated database. Transfer of the files to permanent storage is using the Data Management System of DIRAC, slightly adapted for the purposes of the Online system. Once files are copied to CERN's mass-storage, replication to the LHCb Tier1 centres is initiated.

This system has been designed for extreme robustness and reliability with redundancy in almost every component and various failover mechanisms. The software, hardware and protocols designed for this purpose will be discussed in detail. Performance figures of data-challenges done in Spring 2007 will be shown.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

LHCb Online

Distributed data analysis and information management / 96

Data access performance through parallelization and vectored access. Some results.

Author: Fabrizio Furano¹

Co-author: Andrew Hanushevsky²

¹ *INFN sez. di Padova*

² *SLAC - Stanford Linear Accelerator Center*

Corresponding Author: furano@pd.infn.it

HEP data processing and analysis applications typically deal with the problem of accessing and processing data at high speed.

Recent study, development and test work has shown that the latencies due to data access can often be hidden by parallelizing them with the data processing, thus giving the ability to have applications which process remote data with a high level of efficiency. Techniques and algorithms able to reach this result have been implemented in the client side of the Scalla/xrootd system, and in this contribution we also describe the results of some tests done in order to compare their performance and characteristics. These techniques, if used together with multiple streams data access, can also be effective in making possible to efficiently and transparently deal with data repositories accessible via a Wide Area Network.

Distributed data analysis and information management / 344

Data management in BaBar

Author: Douglas Smith¹

Co-author: Tim Adye²

¹ *Stanford Linear Accelerator Center*

² *Rutherford Appleton Laboratory, Chilton, Didcot, Oxon, United Kingdom*

Corresponding Author: douglas@slac.stanford.edu

The BaBar high energy experiment has been running for many years now, and has resulted in a data set of over a petabyte in size, containing over two million files. The management of this set of data has to support the requirements of further data production along with a physics community that has vastly different needs. To support these needs the BaBar bookkeeping system was developed, and within this datasets are defined for data access and use. Datasets are defined in such a way to keep data separate for the hundreds of concurrent analyses, produced from many production cycles, and to keep similar data together for any specific use. In the development of this system, data has been modeled as a flow of information, that constantly changes. This system has been in use now for many years, and has been very successful in meeting these disparate needs. The methods for defining and managing datasets which will undergo constant changes will be discussed. The needs of production also require the distribution of data to computing centers, and the control of production with datasets will be mentioned. With the needs of a constantly changing dataset, the ability to analyze data from a known state, and then add to the analysis changes in the dataset at a future time will also be presented.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

BaBar Computing Group

Poster 1 - Board: 73 / 141

Data storage layout and architecture of the German T1

Author: Doris Ressmann¹

Co-authors: Jos van Wezel¹ ; Silke Halstenberg¹ ; Stephanie Meier¹

¹ *Forschungszentrum Karlsruhe*

Corresponding Author: doris.ressmann@iwr.fzk.de

The grid era brings upon new and steeply rising demands in data storage. The GridKa project at Forschungszentrum Karlsruhe delivers its share of the computation and storage requirements of all LHC and 4 other HEP experiments. Access throughput from the worker nodes to the storage can be as high as 2 GB/s. At the same time a continuous throughput in the order of 300-400 MB/s into and out from GridKa must be guaranteed for several months without interruption. The scalable storage and networking concept is based on modular storage units that offer dCache, xrootd and NFS access to over 1000 clustered hosts.

dCache directs over 300 pools with a total of 600 TB of disk storage.

The tape connection via separate SAN is managed by Tivoli Storage Manager (TSM) using storage agents. The talk describes software and hardware components, their integration and interconnects, then focuses on the design criteria of the architecture. Plans for enhancements and directions based on current experiences, future expansion are also discussed.

Poster 1 - Board: 63 / 462

Database architecture for the calibration of ATLAS Monitored Drift Tube Chambers

Authors: Domizia Orestano¹; Elisabetta Vilucchi²; Florbela Tique Aires Viegas³; Gancho Dimitrov⁴; Joe Rothberg⁵; Manuela Cirilli⁶; Monica Verducci³; Paola Celio⁷; Paolo Bagnaia⁸; Shawn Mc Kee⁶

¹ *Universita' di Roma Tre & INFN Roma Tre*

² *INFN Frascati*

³ *CERN*

⁴ *Lawrence Berkeley National Lab*

⁵ *University of Washington*

⁶ *University of Michigan*

⁷ *INFN Roma Tre*

⁸ *Universita' di Roma "La Sapienza" & INFN Roma*

Corresponding Author: manuela.cirilli@cern.ch

The calibration of the 375000 ATLAS Monitored Drift Tubes will be a highly challenging task: a dedicated set of data will be extracted from the second level trigger of the experiment and streamlined to three remote Tier-2 Calibration Centres.

This presentation reviews the complex chain of databases envisaged to support the MDT Calibration and describes the actual status of the implementation and the tests that are being performed to ensure a smooth operation at the LHC start-up.

Summary:

The size and complexity of LHC experiments raise unprecedented challenges not only in terms of detector design, construction and operation, but also in terms of software models and data persistency. One of the more challenging tasks is the calibration of the 375000 Monitored Drift Tubes, that will be used as precision tracking detectors in the Muon Spectrometer of the ATLAS experiment. A high rate of muon tracks is required to determine the tube position/time (or Rt) calibration constants in order to reach the design average resolution of 80 microns.

In this context, data suitable for MDT calibration will be extracted from the second level trigger and then streamed to three remote Tier-2 Calibration Centres. The Calibration sites will also need the ATLAS conditions data that are relevant for the calculation of MDT calibrations: part of the Conditions DataBase will thus be replicated at the remote sites via ORACLE streams. At each centre, the computation of the actual calibration constants will be performed in several steps, including severe validation and data quality checks. All information produced at every stage of the calibration procedure will be stored in local ORACLE Calibration databases that will be replicated to a central database located at CERN using ORACLE streams: this will allow each Calibration site to access the data produced by the others and to eventually provide back-up should one site become unavailable for any reason. The validated calibration constants will be extracted from the CERN Calibration DB and stored into the ATLAS Conditions database for subsequent use in reconstruction and data analysis. This presentation reviews the complex chain of databases envisaged to support the MDT Calibration and describes the actual status of the implementation and the tests that are being performed to ensure a smooth operation at the LHC start-up at the end of this year.

Poster 1 - Board: 53 / 387

Dataharvester - a library for reading and writing “hierarchic tuples” from/to various file formats

Author: Wolfgang Waltenberger¹

Co-authors: Gerald Richter¹ ; Herbert Valerio Riedel¹

¹ *Hephy Vienna*

Corresponding Author: walten@cern.ch

A tool is presented that is capable of reading from and writing to several different file formats. Currently supported file formats are ROOT, HBook, HDF, XML, Sqlite3 and a few text file formats. A plugin mechanism decouples the file-format specific “backends” from the main library.

All data are internally represented as “heterogenous hierarchic tuples”; no other data structure exists in the harvester.

The tool is written in C++; a python interface exists, as well.

It is a design goal of the tool to make writing tuples as simple as possible – a feature that is intended make it a superb tool for debugging e.g. algorithmic code.

To this end, the following features are implemented in the dataharvester:

- Files are opened implicitly, not explicitly
- the structure of the tuples is defined implicitly: defining and filling a tuple is one step only
- the data types are defined implicitly, not explicitly.

The dataharvester is fully autotooled. Debian packages exist; an rpm spec file is work in progress.

Grid middleware and tools / 121

Deploying HEP Applications Using Xen and Globus Virtual Workspaces

Authors: Andre Charbonneau¹ ; Angela Norton² ; Ashok Agarwal² ; Daniel Vanderster² ; David Grundy² ; Duncan Penfold-Brown² ; Ian Gable² ; Randall Sobie² ; Rolf Seuster² ; Ron Desmarais² ; Ryan Enge² ; Tristan Sullivan²

¹ *National Research Council of Canada*

² *University of Victoria*

Corresponding Author: igable@uvic.ca

Deployment of HEP application in heterogeneous grid environments can be challenging because many of the applications are dependent on specific OS versions and have a large number of complex software dependencies. Virtual machine monitors such as Xen could ease the deployment burden by allowing applications to be packaged complete with their execution environments. Our previous work has shown HEP applications running within Xen to suffer little or no performance penalty as a result of virtualization. However, a practical strategy is required for remotely deploying, booting, and controlling virtual machines on a remote cluster. One tool that promises to overcome the deployment hurdles using standard grid technology is the Globus Virtual Workspaces project. We investigate strategies for the deployment of Xen virtual machines using Globus Virtual Workspace middleware that simplify the deployment of HEP applications. Further, we study the feasibility of deploying user-constructed virtual machines for the purpose of executing custom physics analyses.

Computer facilities, production grids and networking / 251

Deploying perfSONAR-based End-to-End Monitoring for Production US-CMS Networking

Author: Maxim Grigoriev¹

Co-authors: Andrey Bobyshev¹; Don Petravick¹; Matt Crawford¹; Philip DeMar¹; Vyto Grigaliunas¹

¹ *FERMILAB*

Corresponding Author: maxim@fnal.gov

End-to-end (E2E) circuits are used to carry high impact data movement into and out of the US CMS Tier-1 Center at Fermilab. E2E circuits have been implemented to facilitate the movement of raw experiment data from Tier-0, as well as processed data to and from a number of the US Tier-2 sites. Troubleshooting and monitoring those circuits presents a challenge, since the circuits typically cross multiple research and education networks, each with its own management domain and customized monitoring capabilities. The perfSONAR monitoring project was established develop and deploy a common monitoring infrastructure across multiple network management domains. Fermilab has deployed perfSONAR across its E2E circuit infrastructure and enhanced the product with several tools that ease the monitoring and management of those circuits.

This paper will present the current state of perfSONAR monitoring at Fermilab and detail our experiences using perfSONAR to manage our current E2E circuit infrastructure. We will describe how production network circuits are monitored by perfSONAR E2E Monitoring Points (MPs), and the benefits it has brought to production US CMS networking support.

Poster 2 - Board: 94 / 384

Design Principles of a Web Interface for Monitoring Tools

Authors: Antonio Pierro¹ ; Cristina Aiftimiei² ; Enrico Fattibene³ ; Giacinto Donvito⁴ ; Giorgio Pietro Maggi⁵ ; Giuseppe Misurelli³ ; Guido Cuscela⁴ ; Sergio Andreozzi³ ; Sergio Fantinel⁶ ; Stefano Dal Pra⁷ ; Vihang Dudhalkar⁵

¹ INFN, Bari, Italy

² INFN, Padova, Italy, on leave from NIPNE-HH, Romania

³ INFN-CNAF, Bologna, Italy

⁴ INFN, Bari, Italy

⁵ INFN, Bari also affiliated to Politecnico di Bari, Italy

⁶ INFN, Padova also affiliated to INFN-LNL, Padova Italy

⁷ INFN, Padova, Italy

Corresponding Authors: enrico.fattibene@cnafe.infn.it, giuseppe.misurelli@cnafe.infn.it

A monitoring tool for complex Grid systems can gather a huge amount of information that have to be presented to the users in the most comprehensive way. Moreover different types of consumers could be interested in inspecting and analyzing different subsets of data. The main goal in designing a Web interface for the presentation of monitoring information is to organize the huge amount of data in a simple, user-friendly and usable structure. One more problem is to consider different approaches, skills and interests that all the possible categories of users have in looking for the desired information.

Starting from the information architecture guidelines for the Web, it is possible to design Web interfaces towards a better user experience and to deal with an advanced user interaction by exploiting the most recent Web standards.

In this paper, we present a number of principles for the design of Web interface for monitoring tools that provide a wider, richer range of possibilities for what concerns the user interaction. These principles are based on an extensive review of the current literature in Web design, on the experience with the development of the GridICE monitoring tool and on an analysis of the design choices of other monitoring systems. The described principles can drive the evolution of the Web interface of Grid monitoring tools.

Software components, tools and databases / 181

Development Status and Plans for the LCG Common Database Access Layer (CORAL)

Author: Giacomo Govi¹

¹ CERN

Corresponding Authors: dirk.duellmann@cern.ch, giacomo.govi@cern.ch

The CORAL package has been developed as part of the LCG Persistency Framework project, to provide the LHC experiments with a single C++ access layer supporting a variety of relational database systems.

In the last two years, CORAL has been integrated as database foundation in several LHC experiment frameworks and is used in both offline and online domains. Also, the other LCG Persistency Framework components such as POOL and COOL are now using CORAL to implement their higher-level database operations in a consistent way for all supported database back-ends.

This presentation will summarise the CORAL functionality and the experience gained in large-scale physics production activities. We present recent developments, such as support for multi-threaded applications, a python scripting interface and tools for copying data between different databases. Finally, an overview of the remaining development and consolidation activities to prepare for full LHC production will be presented.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CORAL team

Poster 1 - Board: 103 / 370

Development of the Tier-1 Facility at Fermilab

Author: Ian Fisk¹

Co-authors: Burt Holzman²; David Fagan²; Jon Bakken²; Joseph Kaiser²; Lisa Giacchetti²; Merina Albert²; Paul Tater²; Richard Thompson²; Tim Messer²; Yujun Wu²

¹ *Fermi National Accelerator Laboratory (FNAL)*

² *Fermilab*

Corresponding Author: ian.fisk@cern.ch

CMS is preparing seven remote Tier-1 computing facilities to archive and serve experiment data. These centers represent the bulk of CMS's data serving capacity, a significant resource for reprocessing data, all of the simulation archiving capacity, and operational support for Tier-2 centers and analysis facilities. In this paper we present the progress on deploying the largest remote Tier-1 facility for CMS, located at Fermilab. We will present the development, procurement and operations experiences during the final year of preparation. We will discuss the results of scale tests and system design. We will outline the hardware selection and procurement and plans for the future to meet the needs of the experiment and the constraints of the physical facility. We will also discuss the successes and challenges associated with enabling a mass storage system to meet the various experimental needs at a significant increase in scale over what is currently achievable. Finally we will discuss the model to support US Tier-2 centers from the Tier-1 facility.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS

Distributed data analysis and information management / 186

Development, Deployment and Operations of ATLAS Databases

Authors: Alexandre Vaniachine¹; Hans von der Schmitt²

¹ *Argonne National Laboratory*

² *Max-Planck-Institut für Physik*

Corresponding Author: vaniachine@anl.gov

In preparation for ATLAS data taking in ATLAS database activities a coordinated shift from development towards operations has occurred. In addition to development and commissioning activities in databases, ATLAS is active in the development and deployment (in collaboration with the WLCG 3D project) of the tools that allow the worldwide distribution and installation of databases and related datasets, as well as the actual operation of this system on ATLAS multi-grid infrastructure.

We describe development and commissioning of major ATLAS database applications for online and offline: Trigger DB, Luminosity DB, Geometry DB, Conditions DB, Metadata DB,

and Tag DB. We present the ramp-up schedule over the initial LHC years of operations towards the nominal year of ATLAS running, when the database storage volumes are expected to reach 6.1 TB for the Tag DB and 0.8 TB for the Conditions DB.

ATLAS database applications require robust operational infrastructure for data replication between online and offline at Tier-0, and for the distribution of the offline data to Tier-1 and Tier-2 computing centers. We describe ATLAS experience with Oracle Streams and other technologies for coordinated replication of databases in the framework of the WLCG 3D services.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Distributed data analysis and information management / 346

Developments in BaBar simulation - life without a database

Author: Douglas Smith¹

Co-authors: Peter Kim¹ ; Wilko Kroeger¹

¹ *Stanford Linear Accelerator Center*

Corresponding Author: douglas@slac.stanford.edu

There is a need for a large dataset of simulated events for use in analysis of the data from the BaBar high energy physics experiment. The largest cycle of this production in the history of the experiment was just completed in the past year, simulating events against all detector conditions in the history of the experiment, resulting in over eleven billion events in eighteen months. This computing effort was distributed to almost twenty different computing centers in North America and Europe. The history of this production will be discussed in the talk. This was the second cycle of production for BaBar to produce data as a set of root files, where at the start of the experiment data produced into Objectivity databases. But even though Objectivity was removed from use in data storage, it was still in use for detector conditions. For the next cycle of production, which has recently begun in the experiment, the use of an Objectivity database for the detector conditions was removed, and condition data was distributed with the jobs as a set of root files. The results of this latest stage in the development of simulation production in BaBar will be discussed, and its effect on the computing effort.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

The BaBar Computing Group

Distributed data analysis and information management / 287

Distributed Analysis using GANGA on the EGEE/LCG infrastructure

Author: Johannes Elmsheuser¹

Co-authors: Adrian Muraru² ; Alexander Soroko³ ; Benjamin Gaidioz² ; Chun Lik Tan⁴ ; Dietrich Liko² ; Frederic Brochu⁵ ; Hurng-Chung Lee⁶ ; Jakub Moscicki² ; Karl Harrison⁷ ; Ulrik Egede⁸ ; Vladimir Romanovsky⁹

¹ *Ludwig-Maximilians-Universität München*

² CERN

³ University of Oxford

⁴ University of Birmingham

⁵ University of Cambridge

⁶ ASGC, Taipei and CERN

⁷ High Energy Physics Group, Cavendish Laboratory

⁸ Imperial College London

⁹ State Res. Center of Russian Feder. Inst. f. High Energy Phys. (IFVE)

The distributed data analysis using Grid resources is one of the fundamental applications in high energy physics to be addressed and realized before the start of LHC data taking. The needs to manage the resources are very high. In every experiment up to a thousand physicist will be submitting analysis jobs into the Grid. Appropriate user interfaces and helper applications have to be made available to assure that all users can use the Grid without too much expertise in Grid technology. These tools enlarge the number of grid users from a few production administrators to potentially all participating physicists.

The GANGA job management system (<http://cern.ch/ganga>), developed as a common project between the ATLAS and LHCb experiments provides and integrates these kind of tools.

GANGA provides a simple and consistent way of preparing, organizing and executing analysis tasks within the experiment analysis framework, implemented through a plug-in system. It allows trivial switching between running test jobs on a local batch system and running large-scale analyzes on the Grid, hiding Grid technicalities.

We will be reporting on the plug-ins and our experiences of distributed data analysis using GANGA within the ATLAS experiment and the EGEE/LCG infrastructure. The integration with the ATLAS data management system DQ2 into GANGA is a key functionality. In combination with the job splitting mechanism large amounts of jobs can be sent to the locations of data following the ATLAS computing model. GANGA supports tasks of user analysis with reconstructed data and small scale production of Monte Carlo data.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS Offline Computing

Computer facilities, production grids and networking / 202

Distributed Cluster dynamic storage: A comparison of dcache, xrootd and slashgrid storage systems running on batch nodes.

Author: Alessandra Forti¹

Co-authors: Andrew McNab¹; Colin Morey¹; David Bailey¹; Richard Hughes-Jones¹; Roger Barlow¹

¹ University of Manchester

Corresponding Author: alessandra.forti@cern.ch

The HEP department of the University of Manchester has purchased a 1000 nodes cluster. The cluster is dedicated to run EGEE and LCG software and is currently

supporting 12 active VOs. Each node is equipped with 2x250 GB disks for a total amount of 500 GB and there is no tape storage behind nor raid arrays are used. Three different storage solutions are currently being deployed to exploit this space: dcache, xrootd, slashgrid (HTTP based). In this paper we will present a comparison of their ease of use and their performance from different perspectives: System management perspective (ease of installation and maintainance); user perspective (type of functionality and reliability); and from the performance point of view, with random and streamed access of files. The comparisons have been done with different conditions of load on the worker nodes, and with different file sizes. Test executables, user analysis jobs accessing data from real HEP experiments and file transfers clients have been used in these tests.

Summary:

Comparison of 3 different distributed storage systems (dcache,xrootd,slashgrid) running on the batch nodes of a grid cluster.

Distributed data analysis and information management / 176**Distributed Data Analysis in LHCb**

Authors: Andrew Maier¹ ; Stuart Paterson¹

¹ CERN

Corresponding Author: stuart.paterson@cern.ch

The LHCb distributed data analysis system consists of the Ganga job submission front-end and the DIRAC Workload and Data Management System. Ganga is jointly developed with ATLAS and allows LHCb users to submit jobs on several backends including: several batch systems, LCG and DIRAC. The DIRAC API provides a transparent and secure way for users to run jobs to the Grid and is the default mode of submission for the LHCb VO. This is exploited by Ganga to perform distributed user analysis for LHCb. This system provides LHCb with a consistent, efficient and simple user experience in a variety of heterogeneous environments and facilitates the incremental development of user analysis from local test jobs to the Worldwide LHC Computing Grid.

With a steadily increasing number of users, the LHCb distributed analysis system has been tuned and enhanced over the past two years. This paper will describe the recent developments to support distributed data analysis for the LHCb experiment on WLCG.

Grid middleware and tools / 182**Distributed Database Access in the LHC Computing Grid with CORAL**

Author: Giacomo Govi¹

¹ CERN

Corresponding Authors: dirk.duellmann@cern.ch, giacomo.govi@cern.ch

The CORAL package is the LCG Persistency Framework foundation for accessing relational databases. From the start CORAL has been designed to facilitate the deployment of the LHC experiment database applications in a distributed

computing environment. This contribution focuses on the description of CORAL features for distributed database deployment.

In particular we cover

- improvements to database service scalability by client side connection management
- improved application reliability provided by automated reconnection and fail-over strategies
- a secure authentication and authorisation scheme integrated with existing grid services

We will summarize the deployment experience from several experiment productions using the distributed database infrastructure, which is now available in LCG. Finally, we present perspectives for future developments in this area.

Poster 2 - Board: 26 / 358

Distributed Interactive Access to Large Amount of Relational Data

Author: Julius Hrivnac¹

¹ LAL

Corresponding Author: julius.hrivnac@cern.ch

LCG experiments will contain large amount of data in relational databases. Those data will be spread over many sites (Grid or not). Fast and easy access will required not only from the batch processing jobs, but also from the interactive analysis.

While many system have been proposed and developed for access to file-based data in the distributed environment, methods of efficient access to relational database resources have been so far largely underdeveloped.

Presentation will describe an architecture allowing seamless access to distributed relational data from end-user clients.

Implementation is based on the standard protocols and tools, extended with HEP-specific functionality to be easily integrated in LCG frameworks.

Major constituents of the architecture are:

- Java as the implementation language and JDBC as the databases access protocol
- transparent access from other languages using standard interfaces
- standardly available Sequoia tool to offer single access point to distributed database resources
- specific plugins to accommodate HEP-specific access patterns and provide resource usage optimization via networks of transparent proxy caches
- integration in several end-user analysis tools (JAS, jHEPWork,...)

Poster 1 - Board: 20 / 432

Distributed database services in PHENIX - what it takes to support a Petabyte experiment.

Author: Irina Sourikova¹

¹ BROOKHAVEN NATIONAL LABORATORY

Corresponding Author: irina@bnl.gov

After seven years of running and collecting 2 Petabytes of physics data, PHENIX experiment at the Relativistic Heavy Ion Collider (RHIC) has gained a lot of experience with database management systems (DBMS).

Serving all of the experiment's operations - data taking, production and analysis - databases provide 24/7 access to calibrations and book-keeping information for hundreds of users at several computing centers worldwide and face the following challenges:

- Simultaneous data taking, production and analysis result in hundreds of concurrent database connections and heavy server I/O load.
- Online data production at remote sites requires a high degree of Master-Slave server synchronization.
- Database size (presently 100GB with half of data added in the last few months) raises scalability concerns.
- Long life of modern HENP experiments and fast development of database technologies make prediction of the best DBMS provider 5-10 years down the road difficult and require investments in design and support of good APIs.

In this talk PHENIX solutions to the above problems will be presented and the trade-offs discussed.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

PHENIX

Poster 1 - Board: 54 / 399

ETICS Meta-Data Software Editing - From Check Out To Commit Operations

Author: Elisabetta Ronchieri¹

Co-authors: Guillermo Diez-Andino Sancho² ; Marc-Ellian Begin² ; Matteo Selmi¹ ; Saverio Da Ronco³

¹ INFN CNAF

² CERN

³ INFN Padova

Corresponding Author: elisabetta.ronchieri@cnaif.infn.it

People involved in modular projects need to improve the build software process, planning the correct execution order and detecting circular dependencies. The lack of suitable tools may cause delays in the development, deployment and maintenance of the software.

Experience in such projects has shown that the arranged use of version control and build systems is not able to support the development of the software efficiently, due to the large number of errors that cause the breaking of the build process.

In this paper, we describe a possible solution implemented in ETICS, an integrated infrastructure for the automated configuration, build and test of Grid and distributed software. ETICS has defined meta-data software abstractions, from which it is possible to download, build and test software projects, setting for instance dependencies, environment variables and properties. Furthermore, the meta-data information is managed by ETICS reflecting the version control system philosophy, thanks to the existence of a meta-data repository and the handling of a list of operations, such as check out and commit. Because of this, all the information related to a specific software are stored in the repository only when they

are considered to be correct.
 By adopting this solution, we show a reduction of errors at build time.
 Moreover, by introducing this functionality, ETICS will be a version control system like for the management of the meta-data.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ETICS

Software components, tools and databases / 406

ETICS: the international software engineering service for the grid

Author: Alberto Di Meglio¹

¹ *CERN*

Corresponding Author: alberto.di.meglio@cern.ch

The ETICS system is a distributed software configuration, build and test system designed to fulfill the needs to improve the quality, reliability and interoperability of distributed software in general and grid software in particular. The ETICS project is a consortium of five partners (CERN, INFN, Engineering Ingegneria Informatica, 4D Soft and the University of Wisconsin-Madison). The ETICS service consists of a build and test job execution system based on the NMI/Metronome software and an integrated set of web services and software engineering tools to design, maintain and control build and validation scenarios. The ETICS system allows taking into account complex dependencies among applications and middleware components and provides a rich environment to perform static and dynamic analysis of the software and execute deployment, system and interoperability tests. This presentation gives an overview of the system architecture and functionality set and then describes how the EGEE, DILIGENT and OMII-Europe projects are using the software engineering services to build, validate and distribute their software. Finally a number of significant use and test cases will be described to show how ETICS can be used in particular to perform interoperability tests of grid middleware using the grid itself.

Poster 1 - Board: 55 / 417

EVIO - A Lightweight Object-Oriented Event I/O Package

Author: Elliott Wolin¹

Co-authors: Carl Timmer¹ ; David Abbott¹ ; David Lawrence¹ ; Edward Jastrzembski¹ ; Graham Heyes¹ ; Vardan Gurjyan¹

¹ *Jefferson Lab*

Corresponding Author: wolin@jlab.org

EVIO is a lightweight event I/O package consisting of an object-oriented layer on top of a pre-existing, highly efficient, C-based event I/O package. The latter, part of the JLab CODA package, has been in use in JLab high-speed DAQ systems for many years, but other underlying disk I/O packages could be substituted. The event format on disk, a packed tree-like hierarchy of banks, maps directly

to XML, so notions such as stream and DOM parsing directly apply. The EVIO package transparently maps the packed binary representation on disk to/from an object hierarchy or DOM tree in memory. The in-memory tree can then be queried or modified using STL-like algorithms, function objects, etc. Utility programs can transform between binary and real XML (ASCII) format. EVIO will be used by the next generation of JLab online and offline software systems.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

Jefferson Lab Data Acquisition Group

Collaborative tools / 36

EVO (Enabling Virtual Organizations), the Next Generation Grid- enable Collaborative

Author: Philippe Galvez¹

Co-author: Harvey Newman¹

¹ *California Institute of Technology*

Corresponding Author: philippe.galvez@cern.ch

The EVO (Enabling Virtual Organizations) system is based on a new distributed and unique architecture, leveraging the 10+ years of unique experience of developing and operating the large distributed production based VRVS collaboration system. The primary objective being to provide to the High Energy and Nuclear Physics experiments a system/service that meet their unique requirements of usability, quality, scalability, reliability, and cost necessary for nationally and globally distributed research organizations.

The EVO system, which will be officially released during March/April 2007 includes a better-integrated and more convenient user interface, a richer feature set including higher resolution video and instant messaging, greater adaptability to all platforms and operating systems, and higher overall operational efficiency and robustness. All of these aspects will be particularly important as we approach and then enter the startup period of the LHC because the community will require an unprecedented level of daily collaboration. There will be intense demand for long distance scheduled meetings, person-to-person communication, group-to-group discussions, broadcast meetings, workshops and continuous presence at important locations such as control rooms and experimental areas. The need to have the collaboration tools totally integrated in the physicists' working environments will gain great importance. Beyond all these user-features, another key enhancement is the collaboration infrastructure network created by EVO, which covers the entire globe and which is fully redundant and resilient to failure. The EVO infrastructure automatically adapts to the prevailing network configuration and status, so as to ensure that the collaboration service runs without disruption. Because we are able to monitor the end-user's node, we are able to inform the user of any potential or arising problems (e.g. excessive CPU load or packet loss) and, where possible, to fix the problems automatically and transparently on behalf of the user (e.g. by switching to another server node in the network, by reducing the number of video streams received, et cetera). The integration of the MonALISA architecture into this new EVO architecture was an important step in the evolution of the service towards a globally distributed dynamic system that is largely autonomous.

The EVO system is intended to become the primary collaboration system used by the High Energy and Nuclear Physics community going forward.

Distributed data analysis and information management / 284

Efficient Access to Remote Data in High Energy Physics

Authors: Fons Rademakers¹ ; Leandro Franco¹ ; Rene Brun¹

¹ *CERN*

Corresponding Author: leandro.franco.londono@cern.ch

Particle accelerators produce huge amounts of information in every experiment and such quantity cannot be stored easily in a personal computer. For that reason, most of the analysis is done using remote storage servers (this will be particularly true when the Large Hadron Collider starts its operation in 2007). Seeing how the bandwidth has increased in the last few years, the biggest problem of this approach at the moment is latency, which hurts considerably the performance of the analysis process.

Fortunately, particle events are independent of each other, which allows us to transfer the information that must be processed in the future while analyzing the data at hand. The independent nature also allows us to transfer many events instead of the single one needed at a given time. Such ideas are implemented in the data analysis framework ROOT, and its file servers (rootd, xrootd and a http plugin).

Among the techniques used, we have pre-reads, pre-fetching, parallel streams and atomic readings for multiple requests.

All these strategies present an enormous advantage and will facilitate processing remote files at almost the same speed as local ones, as long as the bandwidth does not present any limitations.

Poster 1 - Board: 84 / 226

Efficient, Large Scale Data Transfers in Wide Area Network

Authors: Ciprian Dobre¹ ; Harvey Newman² ; Iosif Legrand² ; Ramiro Voicu²

¹ *Polytechnic University of Bucharest*

² *CALTECH*

Corresponding Authors: iosif.legrand@cern.ch, ramiro.voicu@cern.ch

The efficient use of high-speed networks to transfer large data sets is an essential component for many scientific applications including CERN's LCG experiments. We present an efficient data transfer application, Fast Data Transfer (FDT), and a distributed agent system (LISA) able to monitor, configure, control and globally coordinate complex, large scale data transfers.

FDT is an Application for Efficient Data Transfers which is capable of reading and writing at disk speed over wide area networks (with standard TCP). If it is used for memory to memory transfers between two computers it can saturate a 10Gb/s WAN link. Disk to disk transfers between several servers can saturate a 10Gb/s in both directions. It is written in Java, runs on all major platforms and it is easy to use.

FDT is based on an asynchronous, flexible multithreaded architecture able to balance and optimize the access to disks and to control the flow of data in the network through multiple streams. It streams datasets continuously, using a managed pool of buffers through one several TCP sockets in parallel and is using independent threads to read and write on each physical device.

LISA (The Localhost Information Service Agent) is lightweight dynamic services that provides complete system monitoring and is capable to dynamically configure the system or running applications. It provides the functionality to orchestrate and

optimize distributed large data transfers at the Supercomputing 2006 Bandwidth Challenge between more than 200 systems around the world.

Poster 2 - Board: 48 / 193

Electron reconstruction in CMS

Author: Claude Charlot¹

¹ *Ecole Polytechnique*

Corresponding Author: charlot@poly.in2p3.fr

We describe the strategy developed for electron reconstruction in CMS. Emphasis is put on isolated electrons and on recovering the bremsstrahlung losses due to the presence of the material before the ECAL. Following the strategy used for the high level triggers, a first filtering is obtained building seeds from the clusters reconstructed in the ECAL. A dedicated trajectory building is then used to collect hits up to the ECAL front face. The fit of the electron trajectories involves a Gaussian sum filter, allowing to take into account the bremsstrahlung losses. The different topologies observed in the ECAL and the difference between the momentum at the innermost and outermost track positions are then used to classify electrons according to their patterns regarding radiative losses. These electron classes are finally used to correct the momentum estimate and to improve the electron identification.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS

Online computing / 355

Electronic Shift and handover log and much more, STAR Elog

Authors: Jerome LAURET¹ ; Levente Hajdu¹

¹ *BROOKHAVEN NATIONAL LABORATORY*

Corresponding Authors: lbhajdu@bnl.gov, jlauret@bnl.gov

Keeping a clear and accurate experiment log is important for any scientific experiment. The concept is certainly not new but keeping accurate while useful records for a Nuclear Physics experiment such as RHIC/STAR is not a priori a simple matter – STAR operates 24 hours a day for six months out of the year with more than 24 shift crews operating 16 different subsystems (some located remotely).

To meet the challenge of not only logging the information but passing it in a concise manner from one shift to another, the STAR experiment has designed an electronic shift Log, a flexible application written in Java and interfacing with the Data Acquisition tools, Quality Assurance reporting, Online shift crews or remote personnel and experts as well as including features such as shift change-over (or handover) forms, tailored to the sub-group of interest. We will present an overview of STAR's Electronic Log a system that is clear, reliable, safe, consistent, easy to use, and globally viewable in real time with secure connection.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

STAR

Poster 1 - Board: 56 / 433

Embedding Python- a new approach to specifying program options

Author: Toby Burnett¹¹ *University of Washington***Corresponding Author:** tburnett@phys.washington.edu

Applications often need to have many parameters defined for execution. A few can be done with the command line, but this does not scale very well. I present a simple use of embedded Python that makes it easy to specify configuration data for applications, avoiding wiring in constants, or writing elaborate parsing difficult to justify for small, or one-off applications. But the capability extends far beyond that, in that the full power of Python is available for computation, consistency checking, importing default parameters.

Poster 2 - Board: 101 / 398

Enabling a priority-based FairShare in the EGEE Infrastructure

Authors: Alessandro Italiano¹ ; Andrea Ferraro¹ ; Antonia Ghiselli¹ ; Daniele Cesini¹ ; Davide Salomoni¹ ; Sergio Andreozzi¹ ; Vincenzo Ciaschini¹¹ *INFN CNAF***Corresponding Authors:** vincenzo.ciaschini@cnaf.infn.it, andrea.ferraro@cnaf.infn.it

While starting to use the grid in production, applications have begun to demand the implementation of complex policies regarding the use of resources. Some want to divide their users in different priority brackets and classify the resources in different classes, others again content themselves with considering all users and resources equal. Resource managers have to work into enabling these requirements on their site, in addition to the work necessary to implement policies regarding the use of their resources, to ensure compliance with AUPs.

These requirements prescribe the existence of a security framework not only capable of satisfying them, but also flexible enough not to require continuous and unnecessary low-level tweaking of configurations every time the requirement change, and that should also do so in a scalable mode. Anything else would only be detrimental when things are seen from the site administrator point of view.

Here we will describe the layout used in several Italian sites of the EGEE infrastructure to deal with these requirements, along with a complete rationale of our choices, with the intent of clarifying what issues administrators may run into when dealing with priority requirements, and what common pitfalls should be avoided at any cost.

Beyond the feedback on interfaces for policy management, from VO and site administrators, we will especially report on the aspects coming from the mapping of grid level policies to local computing resource authorization mechanisms at sites like CNAF T1, and how they interfere from management and security point of view.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

EGEE

Poster 1 - Board: 89 / 247

End-to-End Network/Application Performance Troubleshooting Methodology

Author: Wenji Wu¹

Co-authors: Andrey Bobyshev¹; Don Petravick¹; Mark Bowden¹; Matt Crawford¹; Maxim Grigoriev¹; Philip DeMar¹; Vyto Grigaliunas¹

¹ FERMILAB

Corresponding Author: wenji@fnal.gov

The computing models for LHC experiments are globally distributed and grid-based. In such a computing model, the experiments' data must be reliably and efficiently transferred from CERN to Tier-1 regional centers, processed, and distributed to other centers around the world. Obstacles to good network performance arise from many causes and can be a major impediment to the success of this complex, multi-tiered data grid. Factors that affect overall network/application performance exist on the network end systems themselves (application software, operating system, hardware), in the local area networks that support the end systems, and within the wide area networks. Since the computer and network systems are globally distributed, it can be very difficult to locate and identify the factors that are hurting application performance.

In this paper, we present an end-to-end network/application performance troubleshooting methodology developed and in use at Fermilab. The core of our approach is to narrow down the problem scope with a divide and conquer strategy. The overall complex problem is split into two distinct sub-problems: network end system diagnosis and tuning, and network path analysis. After satisfactorily evaluating, and if necessary resolving, each sub-problem, we conduct end-to-end performance analysis and diagnosis. The paper will discuss tools we use as part of the methodology. The long term objective of the effort is to enable end users to conduct much of the troubleshooting themselves, before (or instead of) calling upon network and end system "wizards," who are always in short supply.

Grid middleware and tools / 405

Enriched namespace to support content-aware authorization policies

Authors: Alberto Forti¹; Antonia Ghiselli¹; Federico Stagni²; Luca Magnoni¹; Riccardo Zappi¹

¹ INFN-CNAF

² INFN

Corresponding Author: riccardo.zappi@cnaf.infn.it

In the near future, data on the order of hundred of Petabytes will be spread in multiple storage systems worldwide dispersed in, potentially, billions of replicated data items. Users, typically, are agnostic about the location of their data and they want to get access by either specifying logical names or using some lookup mechanism.

A global namespace is a logical layer that allows the view of data resources independently from the physical location. Usually, the naming scheme is designed to be easily interpreted by humans and it is organized into a purely user-defined directory hierarchy. Within this model, a data resource is uniquely addressed by file name and path. Nevertheless, this hierarchical structures of logical namespace lacks adequate flexibility to manage sophisticated organization of data. In particular the implicit classification of the data item derived from the path is not enough meaningful to classify data objects when different orthogonal dimensions are considered.

In this paper we expose an enriched namespace able to support a new type of data access authorization policy based on tags. The tags are organized in well-defined hierarchies providing a simple representation of the domain ontology. Only authorized users can label data resources with different tags taken from the domain tag hierarchies. In this way an overlay of classical hierarchical structure of logical namespace with faceted hierarchical tags provides a semantics classification of data entities. Authorization policies defined in respect of tags are content-aware.

Poster 2 - Board: 117 / 232

Ensuring GRID resource availability with the SAM framework in LHCb

Author: Joel Closier¹

¹ CERN

Corresponding Author: joel.closier@cern.ch

The LHCb experiment has chosen to use the SAM framework (Service Availability Monitoring Environment) provided by the WLCG developers to make extensive tests of the LHCb environment at all the accessible grid resources. The availability and the proper definition of the local Computing and Storage Elements, user interfaces as well as the WLCG software environment are checked.

The same framework is also used to pre-install the LHCb applications in the shared software area provided by each site. The deployment of the LHCb applications is based on a python tool developed inside the experiment. It is used for software management including incremental installation of interdependent packages and clean package removal. The software installation is followed by a full application run tests.

According to the results of the experiment specific SAM tests, the sites are allowed to be used in the LHCb production system called DIRAC. The possibility of automated dynamic site certification using the SAM test suite is explored.

This paper will describe the various ways of the LHCb use of the SAM framework. Practical experience in the recent production runs, current limitations and future developments will be presented.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

LHCb

Poster 1 - Board: 33 / 120

Evaluating Disk Hardware for Small Deployments of PostgreSQL

Author: Alexander Withers¹

Co-authors: Antonio Chan ¹ ; Robert Petkus ¹

¹ *Brookhaven National Laboratory*

Corresponding Author: alexw@bnl.gov

The PostgreSQL database is a vital component of critical services at the RHIC/USATLAS Computing Facility such as the Quill subsystem of the Condor Project and both PNFS and SRM within dCache. Current deployments are relatively unsophisticated, utilizing default configurations on small-scale commodity hardware. However, a substantial increase in projected growth has exposed deficiencies in this model. Our goal, therefore, is to ensure the scalability and continued availability of our database servers while minimizing costs and administrative overhead. To attain this goal we tested database I/O throughput across a range of inexpensive server and local/external disk configurations in order to determine which was optimal for our environment. This evaluation considered processor type (AMD vs. Intel), disk family (SATA vs. SAS), RAID configuration, and the amount of system memory. Finally, while our evaluation was designed to solve specific problems, we believe that the results of our general tests can be applied to similar deployments of PostgreSQL elsewhere.

Poster 1 - Board: 32 / 111

Evaluation of Goodness-of-Fit tests in physical use cases

Authors: Alberto Ribon¹ ; Andreas Pfeiffer¹ ; Maria Grazia Pia² ; Paolo Viarengo³

¹ *CERN*

² *INFN Genova*

³ *Univ. Napoli*

Corresponding Author: maria.grazia.pia@cern.ch

The Statistical Toolkit provides an extensive collection of algorithms for the comparison of two data samples: in addition to the chisquared test, it includes all the tests based on the empirical distribution function documented in literature for binned and unbinned distributions.

Some of these tests, like the Kolmogorov-Smirnov one, are widely used; others, like the Anderson-Darling or the Cramer-von Mises tests, have been previously used in more sophisticated physics data analyses; nevertheless, several of the tests in the Statistical Toolkit are largely unknown in high energy physics applications, and a few of them, like the weighted formulations of the Kolmogorov-Smirnov test, are available for the first time in an open-source software tool for statistical analysis.

No systematic evaluation of the power of the goodness-of-fit tests has been documented so far in literature. The present work presents a comprehensive study of the power of all the goodness-of-fit tests in rich collection of the Statistical Toolkit in a variety of use cases specific to high energy physics experiments; it highlights the relative merits and deficiencies of the algorithms to deal with peculiar characteristics of the configurations under study. The results of this study provide guidance for the selection of the most appropriate algorithm in experimental analyses.

The Statistical Toolkit user layer is interfaced to AIDA-compliant analysis tools and to ROOT.

Online computing / 339

Event reconstruction algorithms for the ATLAS trigger

Author: Teresa Maria Fonseca Martin¹

¹ CERN

Corresponding Author: teresa.fonseca.martin@cern.ch

The ATLAS experiment under construction at CERN is due to begin operation at the end of 2007. The detector will record the results of proton-proton collisions at a centre-of-mass energy of 14 TeV. The trigger is a three-tier system designed to identify in real-time potentially interesting events that are then saved for detailed offline analysis. The trigger system will select approximately 200 Hz of potentially interesting events out of the 40 MHz bunch-crossing rate (with $\sim 10^9$ interactions per second at the nominal luminosity).

Algorithms used in the trigger system to identify different event features of interest will be described, as well as their expected performance in terms of selection efficiency, background rejection and computation time per event. The talk will concentrate on recent improvements and on performance studies, using a very detailed simulation of the ATLAS detector and electronics chain that emulates the raw data as it will appear at the input to the trigger system. Checks on the robustness of the algorithms to detector misalignment and miscalibration will also be discussed.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS Trigger Algorithms, Performance and Menus group

Poster 2 - Board: 61 / 382

Event reconstruction in the Forward and Backward Silicon detectors of HERA experiment H1

Authors: Sergey Gorbunov¹ ; alexander glazov²

¹ GSI

² DESY

Corresponding Authors: glazov@mail.desy.de, sgorboun@mail.desy.de

Stand-alone event reconstruction was developed for the Forward and the Backward Silicon Trackers of the H1 experiment at HERA. The reconstruction module includes the pattern recognition algorithm, a track fitter and primary vertex finder. The reconstruction algorithm shows high efficiency and speed. The detector alignment was performed to within an accuracy of 10 μm which corresponds to the spatial hit resolution. The reconstruction software has been used for on-line and off-line event analysis of the silicon strip detector's data.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

H1 collaboration

Poster 1 - Board: 93 / 288

Exercising CMS dataflows and workflows in computing challenges at the Spanish Tier-1 and Tier-2 sites

Authors: Antonio Delgado Peris¹; F. Javier Rodriguez Calonge¹; Francisco Matorras²; Gonzalo Merino³; Iban Cabrillo²; Isidro Gonzalez Caballero²; Jose Caballero¹; Jose Hernandez¹; Josep Flix³; Nicanor Colino¹; Rafael Marco²

¹ CIEMAT

² IFCA

³ PIC/CIEMAT

Corresponding Author: chema.hernandez.calama@gmail.com

CMS undertakes periodic computing challenges of increasing scale and complexity to test its computing model and Grid computing systems. The computing challenges are aimed at establishing a working distributed computing system that implements the CMS computing model based on an underlying multi-flavour grid infrastructure. CMS dataflows and data processing workflows are exercised during a period of about a month targeting specific performance and scale goals. Performance values are measured, problems are identified and feedback into the design, integration and operation of the computing system is provided.

The CMS computing architecture is based on a tier-organised structure of computing resources, based on a Tier-0 centre at CERN, 7 Tier-1 centres for organized mass data processing, and about 30 Tier-2 centres where user physics analysis is performed. The Tier-0 is in charge of storing the data coming from the detector onto mass storage, performs a prompt reconstruction of the data and distributes the data among the Tier-1 centres. The Tier-1 sites archive on mass storage its share of data, run data reprocessing, organized group physics analysis for data selection and distribute down the selected data to Tier-2's for user analysis. Tier-1 centres also have the responsibility of storing Monte Carlo data produced at the Tier-2 sites. The above mentioned workflows have been exercised during October 2006 at a scale of 25% of what is needed for operations in 2008, and at a 50% scale during July 2007.

An overview of the data- and workflows conducted at the Spanish Tier-1 and Tier-2 sites during the CMS computing challenges since last CHEP conference is presented. The focus is on presenting achieved results, operational experience and lessons learnt during the challenges.

Poster 1 - Board: 94 / 289

Experience Running a Distributed Tier-2 for the ATLAS Experiment

Authors: LUIS MARCH¹; MOHAMMED KACI²; SANTIAGO GONZALEZ²

Co-authors: ALEJANDRO LAMAS³; ALVARO FERNANDEZ³; ANDREU PACHECO⁴; FARIDA FASSI³; GABRIEL AMOROS³; HEGOI GARITAONANDIA⁵; JAVIER SANCHEZ³; JORDI NADAL⁵; JOSE DEL PESO⁶; JOSE SALT⁷; JUAN PARDO⁶; LUIS MUÑOZ⁶; MARIA LLUISA MIR⁵; PABLO FERNANDEZ⁶; SERGE SHUSKOV⁵; XAVIER ESPINAL⁵

¹ Instituto de Fisica Corpuscular

² INSTITUTO DE FISICA CORPUSCULAR

³ IFIC-VALENCIA

⁴ INSTITUTO DE FISICA DE ALTAS ENERGIAS

⁵ IFAE-BARCELONA

⁶ UNIVERSIDAD AUTONOMA DE MADRID

⁷ *INSITUTO DE FISICA CORPUSCULAR*

Corresponding Authors: luis.march@ific.uv.es, sgonzale@ific.uv.es, kaci@ific.uv.es, jose.salt@ific.uv.es, delpeso@uam.es, pacheco@ifae.es

The Spanish ATLAS Tier-2 is geographically distributed between three HEP institutes. They are IFAE (Barcelona) and IFIC (Valencia) and UAM (Madrid). Currently it has a computing power of about 400 kSI2k CPU, a disk storage capacity of 40 TB and a network bandwidth connecting the three sites and the nearest Tier-1 of 1 Gb/s. These resources will increase with time in parallel to those of all the ATLAS Tier-2. They will be about 875 kSI2k CPU and 387 TB disk capacity at the LHC startup in 2008.

The main roles of the Tier-2 are to provide resources for Production of Simulated Events and Physics Distributed Data Analysis. Since 2002, it has been participating to the different Data Challenge exercises. Currently, it is achieving around 1.5% of the whole ATLAS production in the framework of the Computing System Commissioning exercise. On the other hand, a prototype of Distributed Data Analysis system is being tested, integrated, and deployed. Distributed Data Management is also arising as an important issue in the daily activities of the Tier-2.

The distribution in three sites has shown to be useful due to an increasing service redundancy, a faster solution of problems and the share of computing expertise and know-how. Experience gained running the Distributed Tier-2 and the preparations to perform successfully at the LHC startup will be presented.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Software components, tools and databases / 110

Experience and Lessons learnt from running high availability databases on Network Attached Storage

Author: Juan Manuel Guijarro¹

¹ *CERN*

Corresponding Author: juan.manuel.guijarro@cern.ch

The Database and Engineering Services Group of CERN's Information Technology Department provides the Oracle based Central Data Base services used in many activities at CERN.

In order to provide High Availability and ease management for those services, a NAS (Network Attached Storage) based infrastructure has been set up. It runs several instances of the Oracle RAC (Real Application Cluster) using NFS as share disk space for RAC purposes and Data hosting. It is composed of two private LAN's to provide access to the NAS file servers and Oracle RAC interconnect, both using network bonding. NAS nodes are configured in partnership to prevent having single points of failure and to provide automatic NAS fail-over.

This presentation describes that infrastructure and gives some advice

on how to automate its management and setup using a Fabric Management framework such as Quattor. It also covers aspects related with NAS Performance and Monitoring as well Data Backup and Archive of such facility using already existing infrastructure at CERN.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CERN IT/DES Summary:

Experience and Lessons learnt from running high availability databases on Network Attached Storage.

Grid middleware and tools / 170**Experience from a Pilot based system for ATLAS**

Author: Paul Nilsson¹

¹ *UT-Arlington*

The PanDA software provides a highly performant distributed production and distributed analysis system. It is the first system in the ATLAS experiment to use a pilot based late job delivery technique. In this talk, we will describe the architecture of the pilot system used in Panda. Unique features have been implemented for high reliability automation in a distributed environment. Performance of Panda will be analyzed from one and half year of experience of performing distributed computing on the OSG infrastructure. Experience with pilot delivery mechanism using CondorG, and a GlideIn factory developed under OSG will be described.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Poster 1 - Board: 67 / 59

Experience with monitoring of Prague T2 Site

Author: Tomas Kouba¹

¹ *Institute of Physics - Acad. of Sciences of the Czech Rep. (ASCR)*

Corresponding Author: tomas.kouba@cern.ch

Each tier 2 site is monitored by various services from outside. The Prague T2 is monitored by SAM tests, GSTAT monitoring, RTM from RAL, regional nagios monitoring and experiment specific tools. Besides that we monitor our own site for hardware and software failures and middleware status. All these tools produce an output that must be regularly checked by site administrators. We will present our solution built on nagios that allows our administrators to check just one service (nagios) that encapsulates the results from all monitoring tools (outside and inside) and presents them in a single web page. The solution is based on simple plugin for every service. Nagios uses these plugins to check the results of monitoring tools. We use plugins developed at SRCE, RAL and Prague. We will also present automatic configuration scripts that allow us to generate nagios configuration from local database of servers and services.

Computer facilities, production grids and networking / 303**Experience with the gLite Workload Management System in ATLAS Monte Carlo Production on LCG**

Author: Simone Campana¹

¹ *CERN/IT/PSS*

Corresponding Author: simone.campana@cern.ch

The ATLAS experiment has been running continuous simulated events production since more than two years. A considerable fraction of the jobs is daily submitted and handled via the gLite Workload Management System, which overcomes several limitations of the previous LCG Resource Broker. The gLite WMS has been tested very intensively for the LHC experiments use cases for more than six months, both in terms of performance and reliability. The tests were carried out by the LCG Experiment Integration Support team (in close contact with the experiments) together with the EGEE integration and certification team and the gLite middleware developers. A pragmatic iterative and interactive approach allowed a very quick rollout of fixes and their rapid deployment, together with new functionalities, for the ATLAS production activities. The same approach is being adopted for other middleware components like the gLite and CREAM Computing Elements. In this contribution we will summarize the learning from the gLite WMS testing activity, pointing out the most important achievements and the open issues. In addition, we will present the current situation of the ATLAS simulated event production activity on the EGEE infrastructure based on the gLite WMS, showing the main improvements and benefits from the new middleware. Finally, some preliminary results on the new flavors of Computing Elements usage will be shown, trying to identify possible advantages not only in terms of robustness and performance, but also functionality for the experiment activities.

Event processing - Board: 66 / 457

Experience with validating GEANT4 v7 and v8 against v6 in BaBar

Author: Swagato Banerjee¹

Co-authors: Dennis Wright²; Peter Kim²; William Lockman³

¹ *University of Victoria*

² *SLAC*

³ *University of California, Santa Cruz*

Corresponding Author: swagato.banerjee@louisville.edu

BaBar Abstract #8 - Track 2 (Event processing)

Experience with validating GEANT4 v7 and v8 against v6 in BaBar

S. Banerjee, P. Kim, W. Lockman, and D. Wright for the BaBar Computing Group

The BaBar experiment at SLAC has been using the GEANT 4 package version 6 for simulation of the detector response to passage of particles through its material.

Since 2005 and 2006, respectively, GEANT 4 versions 7 and 8 have been available, providing: improvements in modeling of multiple scattering; corrections to muon ionization and improved MIP signature; widening of the core of electro-magnetic shower shape profiles; newer implementation of elastic scattering for hadronic processes; exact implementation of Bertini cascade models for kaons and lambdas, and updated hadronic cross-sections from calorimeter beam tests.

The effects of these changes in simulation are studied in terms of closer agreement of simulation with respect to data distributions of:

the hit residuals of tracks in the silicon-vertex tracker; the shower shapes of photons and K_L particles in the electro-magnetic calorimeter; the ratio of energy deposited in the electro-magnetic calorimeter and the flux return of the magnet instrumented with a muon detection system composed of resistive plate chambers, and limited streamer-tubes; and the muon identification efficiency in the muon detector system of the BaBar detector.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

BaBar

Poster 2 - Board: 12 / 116

Experience with xrootd/PROOF solution as a data access model for a Computing Center

Author: Trunov Artem¹

Co-author: Jean-Yves Nief²

¹ CC-IN2P3 (Lyon) and EKP (Karlsruhe)

² CC-IN2P3 (Lyon)

Corresponding Author: artem.trunov@cern.ch

We present our experience in setting up an xrootd storage cluster at CC-IN2P3 - a LCG Tier-1 computing Center. The solution consists of xrootd storage cluster made of NAS boxes and includes an interface to dCache/SRM, and Mass Storage System. The feature of this system is integration of PROOF for facilitation of analysis. The setup allows to take advantage of ease of administrative burden, scalability and performance of xrootd, reduction of total cost due to integration of PROOF, while satisfying requirements for SRM interface, strong authentication for transfer and complex data management applied by LCG to site's storage. The same setup can be used at large or small HEP computing centers as analysis facility and data access systems. Variations of this setup for smaller sites are also considered. Feedback from experiments on this setup is also reported.

Computer facilities, production grids and networking / 65

Experiences with gStore, a scalable Mass Storage System with Tape Backend

Author: Horst Goeringer¹

Co-authors: Matthias Feyerabend¹; Sergei Sedykh¹

¹ GSI

Corresponding Author: h.goeringer@gsi.de

GSI in Darmstadt (Germany) is a center for heavy ion research and hosts an Alice Tier2 center. For the future FAIR experiments at GSI, CBM and Panda, the planned data rates will reach those of the current LHC experiments at Cern. Since more than ten years gStore, the GSI Mass Storage System, is successfully in operation.

It is a hierarchical storage system with a unique name space. Its core consists of several tape libraries from different vendors and currently ~20 data mover nodes connected within a SAN network. The gStore clients transfer data via fast socket connections from/to the disk cache of the data movers (~40TB currently). Each tape is accessible from any data mover, fully transparent to the users. The tapes and libraries are managed by commercial software (IBM Tivoli Storage Manager TSM), whereas the disk cache management and the TSM and user interfaces are provided by GSI software. For Alice users all gStore data are worldwide accessible via Alice grid software, and in a test environment the Alice Tier2 xrootd system has been integrated successfully with gStore. For 2007 it is planned to provide ~200TB via xrootd backed with gStore. Our experiences show that it's possible to develop, maintain and operate successfully a large scale mass storage system with mainly 2 FTEs. As gStore is completely hardware independent and fully scalable in data capacity and I/O bandwidth, we are optimistic to fulfill also the dramatically increased mass storage requirements of the FAIR experiments in 2014, which will be several orders of magnitude higher than today.

Grid middleware and tools / 362

Experiences with the GLUE information schema in the LCG/EGEE production Grid

Authors: Laurence Field¹ ; Sergio Andreozzi² ; Stephen Burke³

¹ *CERN, Geneva, Switzerland*

² *INFN-CNAF, Bologna, Italy*

³ *Rutherford Appleton Laboratory, UK*

Corresponding Author: s.burke@rl.ac.uk

A common information schema for the description of Grid resources and services is an essential requirement for interoperating Grid infrastructures, and its implementation interacts with every Grid component. In this context, the GLUE information schema was originally defined in 2002 as a joint project between the European DataGrid and DataTAG projects and the US iVDGL (the predecessors of the current EGEE/LCG and OSG Grids). It has since had three backward-compatible upgrades, with the latest version (1.3) being deployed this year. The schema has major components to describe Computing and Storage Elements, and also generic Service and Site information. It has been used extensively in the LCG/EGEE Grid, for job submission, data management, service discovery and monitoring. In this paper we present the experience gained over the last five years, highlighting both successes and problems. In particular, we consider the importance of having a clear definition of schema attributes; the construction of standard information providers and difficulties encountered in mapping an abstract schema to diverse real systems; the configuration of publication in a way which suits system managers and the varying characteristics of Grid sites; the validation of published information; the ways in which information can be used (and misused) by Grid services and users; and issues related to managing schema upgrades in a large distributed system.

Poster 2 - Board: 106 / 448

Experimental Evaluation of Job Provenance in ATLAS environment

Authors: Ales Krenek¹; Ludek Matyska¹

Co-authors: Frantisek Dvorak¹; Jiri Chudoba¹; Jiri Filipovic¹; Jiri Sitera¹; Laura Perini²; Milos Mulac¹; Miroslav Ruda¹; Simone Campana³; Zdenek Salvet¹; Zdenek Sustr¹

¹ *CESNET*

² *INFN*

³ *CERN*

Corresponding Author: ludek.matyska@muni.cz

Grid middleware stacks, including gLite, matured into the state of being able to process upto millions of jobs per day. Logging and Bookkeeping, the gLite job-tracking service keeps pace with this rate, however it is not designed to provide a long-term archive of executed jobs.

ATLAS—representative of large user community— addresses this issue with its own job catalogue (prodDB). Development of such a customized service took considerable effort which is not easily affordable by smaller communities and is not easily reused.

On the contrary, Job Provenance (JP) is a generic gLite service designed for long-term archive of information on executed jobs. Its design priorities are:

- (i) scalability – store data on billions of jobs;
- (ii) extensibility – virtually any data format can be uploaded and handled by plugins;
- (iii) uniform data view – all data are logically transformed into RDF-like data model, using appropriate namespaces to avoid ambiguities;
- (iv) configurability – highly customizable components maintaining pre-cooked queries provide efficient query interface.

We present first results of experimental JP deployment for the ATLAS production infrastructure. JP installation was fed with a part of ATLAS production jobs (thousands of jobs per day). We provide a functional comparison of JP and ATLAS prodDB, discuss reliability, performance and scalability issues, and focus on the application level functionality as opposed to pure Grid middleware functions.

The main outcome of this work is a demonstration that JP can complement large-scale application-specific job catalogue services, as well as serve similar purpose where these are not available.

Software components, tools and databases / 292

Explicit state representation and the ATLAS event data model: theory and practice

Authors: Arthur Schaffer¹; David Malon²; Kyle Cranmer³; Marcin Nowak³; Peter van Gemmeren²; Scott Snyder³; Sebastien Binet⁴

¹ *LAL Orsay*

² *Argonne National Laboratory*

³ *Brookhaven National Laboratory*

⁴ *Lawrence Berkeley National Laboratory*

Corresponding Authors: marcin.nowak@cern.ch, malon@anl.gov

In anticipation of data taking, ATLAS has undertaken a program of work to develop an explicit state representation of the experiment's complex transient event data model. This effort has provided both an opportunity to consider explicitly the structure, organization, and content of the ATLAS persistent event store before writing tens of petabytes of data (replacing simple streaming, which uses the persistent store as a core dump of transient memory), and a locus for support of event data model evolution, including significant refactoring, beyond the automatic schema evolution capabilities of underlying persistence technologies. ATLAS has encountered the need for such non-trivial schema evolution on several occasions already.

This paper describes the state representation strategy (transient/persistent separation) and its implementation, including both the payoffs that ATLAS has seen (significant and sometimes surprising space and performance improvements, the extra layer notwithstanding, and extremely general schema evolution support) and the costs (additional and relatively pervasive additional infrastructure development and maintenance). The paper further discusses how these costs are mitigated, and how ATLAS is able to implement this strategy without losing the ability to take advantage of the (improving!) automatic schema evolution capabilities of underlying technology layers when appropriate.

Implications of state representations for direct ROOT browsability, and current strategies for associating physics analysis views with such state representations, are also described.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Poster 2 - Board: 30 / 402

Exploring Capable Systems for Fast Data Distribution over End-to-End 10Gb/s Paths

Author: Kyu Park¹

Co-authors: Alan George²; Dimitri Bourilkov³; Paul Avery³; Richard Cavanaugh³; Shawn McKee⁴

¹ *Department of Electrical and Computer Engineering, University of Florida*

² *Department of Electrical and Computer Engineering, University of Florida*

³ *Department of Physics, University of Florida*

⁴ *Department of Physics, University of Michigan*

Corresponding Author: park@hcs.ufl.edu

A primary goal of the NSF-funded UltraLight Project is to expand existing data-intensive grid computing infrastructures to the next level by enabling a managed network that provides dynamically constructed end-to-end paths (optically or virtually, in whole or in part). Network bandwidth used to be the primary limiting factor, but with the recent advent of 10Gb/s network paths end-to-end, the end system has become the bottleneck for fast data distribution. As an additional goal of UltraLight we have been exploring tradeoff issues with relatively inexpensive solutions for capable end systems. The candidate system should be capable of driving 10 Gb/s WAN links in order to provide fast data cache capabilities for LHC(Large Hadron Collider) computing model.

In this paper, we perform various synthetic and application benchmarks on each architectural component in the data path of a disk-to-disk transfer. First, we find that disk subsystems are usually the main limiting factor for fast data distribution. Therefore, the candidate platforms are required to provide wide I/O paths to accommodate high-performance I/O devices. Second, we investigate a broad range of tunable parameters in the operating system and their impact on the throughput of disk-to-disk transfer. We observe improvement exceeding a factor of 2 in disk I/O throughput which directly impacts the throughput of large disk-to-disk transfers. We also find that disk and memory access patterns together with the size of various buffers play critical roles in maximizing the throughput of disk-to-disk transfers.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

UltraLight

Poster 2 - Board: 119 / 296

Extension of the DIRAC workload-management system to allow use of distributed Windows resources

Author: Ying Ying Li¹

Co-authors: Andrei Tsaregorodtsev²; Karl Harrison¹; Michael Andrew Parker¹; Vassily Lyutsarev³

¹ *University of Cambridge*

² *Université d'Aix - Marseille II*

³ *Microsoft Research*

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

The DIRAC workload-management system of the LHCb experiment allows coordinated use of globally distributed computing power and data storage. The system was initially deployed only on Linux platforms, where it has been used very successfully both for collaboration-wide production activities and for single-user physics studies. To increase the resources available to LHCb, DIRAC has been extended so that it also allows use of Microsoft Windows machines. As DIRAC is mostly written in Python, a large part of the code base was already platform independent, but Windows-specific solutions have had to be found in areas such as certificate-based authentication and secure file transfers, where .NetGridFTP has been used. In addition, new code has been written to deal with the way that jobs are run and monitored under Windows, enabling interaction with Microsoft Windows Compute Cluster Server 2003 on sets of machines where this is available. The result is a system that allows users' transparent access to Linux and Windows distributed resources. This paper gives details of the Windows-specific developments for DIRAC, outlines the experience gained deploying the system at a number of sites, and reports on the performance achieved running the LHCb data-processing applications.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

LHCb

Collaborative tools / 426

Extra Dimensions

Author: Norman Graf¹

¹ *SLAC*

Corresponding Author: norman.graf@slac.stanford.edu

High energy physics is replete with multi-dimensional information which is often poorly represented by the two dimensions of presentation slides and print media. Past efforts to disseminate such information to a wider audience have failed for a number of reasons, including a lack of standards which are easy to implement and have broad support. Adobe's Portable Document Format (PDF) has in recent years become the de facto standard for secure, dependable electronic information exchange. It has done so by creating an open format, providing support for multiple platforms and being reliable and extensible. By providing support for the ECMA standard Universal 3D (U3D) file format in its free Adobe Reader software, Adobe has made it easy to distribute and interact with 3D content. By providing support for scripting and animation, temporal data can also be easily distributed to a wide audience.

In this talk, we present examples of HEP applications which take advantage of this functionality. We demonstrate how 3D detector elements can be documented, using either CAD drawings or other sources such as GEANT visualizations as input. We then show how higher dimensional data, such as LEGO plots or time-dependent information, can be included in PDF files.

We finally synthesize these elements by showing how a complete event display, with full interactivity, can be incorporated into a PDF file. This allows the end user not only to customize the view and representation of the data, but to access the underlying data itself.

Summary:

We demonstrate a number of HEP applications which make use of Adobe's support for the ECMA standard U3D file format in its Portable Document Format.

Poster 1 - Board: 3 / 108

FPGA based Compute Nodes for High Level Triggering in PANDA

Author: Wolfgang Kuehn¹

¹ *Univ. Giessen, II. Physikalisches Institut*

Corresponding Author: wolfgang.kuehn@cern.ch

PANDA is a new universal detector for antiproton physics at the HESR facility at FAIR/GSI. The PANDA data acquisition system has to handle interaction rates of the order of 10^{17} /s and data rates of several 100 Gb /s. FPGA based compute nodes with multi-Gb/s bandwidth capability using the ATCA architecture are designed to handle tasks such as event building, feature extraction and high level trigger processing. Data connectivity is provided via optical links as well as multiple Gbit Ethernet ports. The boards will support trigger algorithms such as pattern recognition for RICH detectors, EM shower analysis, fast tracking algorithms and global event characterization. A high level hardware description language (Handel-C) will be used to implement the firmware.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

PANDA

Event processing / 156

FairRoot : the FAIR simulation and analysis framework

Author: Denis Bertini¹**Co-authors:** Florian Uhlig¹ ; Mohammad Al-Turany¹ ; Peter Malzacher¹¹ GSI**Corresponding Authors:** d.bertini@gsi.de, p.malzacher@gsi.de

The experiments design studies at FAIR are done using a ROOT based simulation and analysis framework : FairRoot. The framework is using the Virtual Monte Carlo concept which allows to perform simulation using Geant3, Geant4 or Fluka without changing the user code. The same framework is then used for data analysis. An Oracle database with a build-in versioning management is used to efficiently store the detector geometry, materials and parameters. A generic track follower based on Geane has been implemented which allows precise and fast tracking algorithm development. Moreover a geometry interface which uses different input format (ascii, root , oracle and step format) is also implemented. The status and results of the main FAIR experiments, CBM (compressed baryonic matter) and PANDA (antiproton annihilation at the high energy storage ring) design studies will be presented as well as the comparison between different Monte Carlo transport code.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

GSI IT Summary:

The experiments design studies at FAIR are done using a ROOT based simulation and analysis framework : FairRoot. The framework is using the Virtual Monte Carlo concept which allows to perform simulation using Geant3, Geant4 or Fluka without changing the user code. The same framework is then used for data analysis

Poster 1 - Board: 107 / 394

Farm Logbook, a tool to keep/trace all operations on a big computing farm

Author: Alessandro Italiano¹**Co-authors:** Andrea Chierici¹ ; Davide Salomoni¹ ; Felice Rosso¹¹ INFN-CNAF**Corresponding Author:** alessandro.italiano@cnafe.infn.it

Every day operations on a big computer center farm like that of a Tier1 can be numerous. Opening or closing a host, changing batch system configuration, replacing a disk, reinstalling a host and so on, is just a short list of what can and will really happen. In these conditions remembering all that has been done could be really difficult. Typically a big farm is managed by a team so it can happen to forget which operations have been performed by some other colleague; in the worst case, one could be completely unaware about the changes. There is therefore a real need to keep trace as much as possible of all operations on the cluster in

order to have
the farm status under full control.

With a tool like “Farm Logbook”, developed and deployed at the INFN Tier-1, you can easily track down information about events occurred in the past or ongoing. All information are obtained in an automatic way in order to avoid consistency problems due to someone missing updates. The technologies used to develop this tool, such as CVS, MySQL, PHP, AJAX and RSS, provide smart access to this information in order to grant anytime to farm administrators a realistic snapshot of the cluster. Potentially therefore this tool can improve farm management and troubleshooting.

Poster 2 - Board: 62 / 383

Fast Simulations for the PANDA Experiment at FAIR

Author: Klaus Goetzen¹

¹ *GSI Darmstadt*

Corresponding Author: k.goetzen@gsi.de

As one of the primary experiments to be located at the new Facility for Antiproton and Ion Research in Darmstadt the PANDA experiment aims for high quality hadron spectroscopy from antiproton proton collisions. The versatile and comprehensive projected physics program requires an elaborate detector design. The detector for the PANDA experiment will be a very complex machine consisting of a large number of different subdetectors like various tracking detectors, electromagnetic calorimeters, different devices for particle identification etc.

The simulation of such a system is a very demanding task in terms of computing power as well as man-power, since in this early stage of the experiment the detector design is not completely fixed and different options for the realization of various subsystems have to be taken into account.

For optimization and tuning of the individual detector components a reliable simulation is mandatory, which allows to investigate acceptance and efficiency of various physics benchmark channels for large numbers of events.

Therefore, a fast simulation providing effective parametrization of detector acceptance and resolution was implemented. Presented will be the techniques for various aspects of this simulation model as well as applications.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

PANDA

Event processing / 97

Fast shower simulation in ATLAS Calorimeter

Author: Wolfgang Ehrenfeld¹

¹ *Univ. of Hamburg/DESY*

Corresponding Author: wolfgang.ehrenfeld@cern.ch

The simulation of the ATLAS detector is largely dominated by the showering of electromagnetic particles in the heavy parts of the detector, especially the electromagnetic barrel and endcap calorimeters. Two procedures have been developed to accelerate the processing time of EM particles in these regions: (1) a fast shower parameterization and (2) a frozen shower library. Both work by generating the response of the calorimeter to electrons and positrons with Geant 4, and then re-importing the response into the simulation at run-time.

In the fast shower parameterization technique, a parameterization is tuned to single electrons and used later by simulation. In the frozen shower technique, actual showers from low-energy particles are imported into the simulation. Simulation in the presence of frozen showers is then required to develop the shower down to ~ 1 GeV, at which point the shower is terminated by substituting a frozen shower. Judicious use of both techniques over the entire electromagnetic portion of the ATLAS calorimeter produces an important improvement of CPU time. We discuss the algorithms and their performance in this talk.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Event processing / 187

Feicim: A browser for data and algorithms.

Author: Ronan McNulty¹

Co-authors: Lazar Zsolt²; Tahar Kechadi³

¹ *University College Dublin, School of Physics*

² *University College Dublin (UCD), School of Physics*

³ *University College Dublin (UCD), School of Computer Science*

Corresponding Author: ronan.mculty@cern.ch

As programming and their environments become increasingly complex, more effort must be invested in presenting the user with a simple yet comprehensive interface. Feicim is a tool that unifies the representation of data and algorithms. It provides resource discovery of data-files, data-content and algorithm implementation through an intuitive graphical user interface. It allows local or remote data stored on Grid type platforms to be accessed by the users, the viewing and creation of user-defined or collaboration-defined algorithms, the implementation of algorithms, and the production of output data-files and/or histograms.

An application of Feicim is illustrated using the LHCb data. It provides a graphical view of the Gaudi architecture, LHCb event data model, and interfaces to the file catalogue. Feicim is particularly suited to such frameworks as Gaudi which consider algorithms as objects. Instant viewing of any LHCb data will be of particular value in the commissioning of the detector and for quickly familiarising newcomers to the data and software environment.

Poster 1 - Board: 86 / 234

FermiGrid - Experience and Future Plans

Author: Chadwick Keith¹

Co-authors: Eileen Berman¹ ; Ruth Pordes¹

¹ *Fermilab*

Corresponding Author: chadwick@fnal.gov

Fermilab supports a scientific program that includes experiments and scientists located across the globe. In order to better serve this community, Fermilab has placed its production computer resources in a Campus Grid infrastructure called 'FermiGrid'. The FermiGrid infrastructure allows the large experiments at Fermilab to have priority access to their own resources, enables sharing of these resources in an opportunistic fashion, and movement of work (jobs, data) between the Campus Grid and National Grids such as Open Science Grid and the WLCG.

FermiGrid resources support multiple Virtual Organizations (VOs), including VOs from the Open Science Grid (OSG), EGEE and the Worldwide LHC Computing Grid Collaboration (WLCG). Fermilab also makes leading contributions to the Open Science Grid in the areas of accounting, batch computing, grid security, job management, resource selection, site infrastructure, storage management, and VO services.

Through the FermiGrid interfaces, authenticated and authorized VOs and individuals may access our core grid services, the 10,000+ Fermilab resident CPUs, near-petabyte (including CMS) online disk pools and the multi-petabyte Fermilab Mass Storage System. These core grid services include a site wide Globus gatekeeper, VO management services for several VOs, Fermilab site authorization services, grid user mapping services, as well as job accounting and monitoring, resource selection and data movement services.

Access to these services is via standard and well-supported grid interfaces.

We will report on the user experience of using the FermiGrid campus infrastructure interfaced to a national cyberinfrastructure - the successes and the problems.

Event processing / 133

Final results of the precision validation of Geant4 models in the pre-equilibrium and nuclear de-excitation phase

Authors: Anton Lechner¹ ; Francesco Ri Rosa² ; Giacomo Cuttone² ; Giorgio Russo² ; Maria Grazia Pia³ ; Pablo Cirrone²

¹ *Tech. Univ. Wien*

² *INFN LNS*

³ *INFN Genova*

Corresponding Authors: maria.grazia.pia@cern.ch, cirrone@lns.infn.it

A project is in progress for a systematic, quantitative validation of Geant4 physics models against experimental data.

Due to the complexity of Geant4 physics, the validation of Geant4 hadronic models proceeds according to a bottom-up approach (i.e. from the lower energy range up to higher energies): this approach, which is different from the one adopted in the LCG Simulation Validation Project, allows establishing the accuracy of individual Geant4 models specific to a given energy range on top of already validated models pertinent to a lower energy.

Results are presented concerning the lower energy hadronic interaction phases: the nuclear de-excitation and pre-equilibrium (up to 100 MeV).

All relevant Geant4 electromagnetic and hadronic physics models, and pre-packaged

physics configurations distributed by the Geant4 Collaboration (PhysicsLists) have been included in the validation test. The hadronic models for inelastic scattering involve Nuclear De-excitation in two variants (default and GEM), Precompound (with or without Fermi break-up), Bertini and Binary Cascade, and parameterised models. Elastic scattering includes parameterised models and the newly developed Bertini Elastic model. Various prepackaged PhysicsLists are also subject to the same validation process.

The validation is performed against experimental data measured with 2% accuracy. The quantitative comparison of simulated and experimental data distributions exploits a rigorous goodness-of-fit statistical analysis.

The final results from high statistics production on the grid are presented: they compare both the relative accuracy and the execution performance of all the options considered. These results provide guidance to users about the choice of Geant4 electromagnetic and hadronic physics models.

Poster 1 - Board: 59 / 449

Fluka and Geant4 simulations using common geometry source and digitization algorithms

Author: Manuel Gallas¹

¹ CERN

Corresponding Author: manuel.venancio.gallas.torreira@cern.ch

Based on the ATLAS TileCal 2002 test-beam setup example, we present here the technical, software aspects of a possible solution to the problem of using two different simulation engines, like Geant4 and Fluka, with the common geometry and digitization code. The specific use case we discuss here, which is probably the most common one, is when the Geant4 application is already implemented. Our goal then is to run the same simulation using the Fluka package by re-using the maximum number of the existing components.

For simple setups, a tool (FLUGG) already exists that allows to use the Fluka engine while the navigation is performed with Geant4, starting from a description of the geometry in terms of Geant4 classes. In complex applications, however, the geometry is often built up at run time from the information stored in a database, and in these cases such a tool cannot be used directly; furthermore, it does not deal with sensitive detectors and digitization. We show how it is possible to overcome these two problems by building around FLUGG a set of tools for reading common Geometry Description Markup Language (GDML) files as well as for generating the output in the format allowing common processing algorithms.

Poster 2 - Board: 65 / 453

Fluka and Geant4 simulations using common geometry source and digitization algorithms

Authors: Alberto Ribon¹ ; Manuel Venancio Gallas Torreira¹ ; Witold Pokorski¹

¹ CERN

Corresponding Author: manuel.venancio.gallas.torreira@cern.ch

Based on the ATLAS TileCal 2002 test-beam setup example, we present here the technical, software aspects of a possible solution to the problem of using two different simulation engines, like Geant4 and Fluka, with the common geometry and digitization code. The specific use case we discuss here, which is probably the most common one, is when the Geant4 application is already implemented. Our goal then is to run the same simulation using the Fluka package by re-using the maximum number of the existing components.

For simple setups, a tool (FLUGG) already exists that allows to use the Fluka engine while the navigation is performed with Geant4, starting from a description of the geometry in terms of Geant4 classes. In complex applications, however, the geometry is often built up at run time from the information stored in a database, and in these cases such a tool cannot be used directly; furthermore, it does not deal with sensitive detectors and digitization. We show how it is possible to overcome these two problems by building around FLUGG a set of tools for reading common Geometry Description Markup Language (GDML) files as well as for generating the output in the format allowing common processing algorithms.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

LCG Applications Area

Poster 2 - Board: 53 / 220

Full-scale CMS Tracker application of the Kalman Alignment Algorithm

Author: Edmund Widl¹

Co-author: Rudolf Frühwirth¹

¹ *Institut für Hochenergiephysik (HEPHY Vienna)*

Corresponding Author: edmund.widl@cern.ch

The Kalman alignment algorithm (KAA) has been specifically developed to cope with the demands that arise from the specifications of the CMS Tracker. The algorithmic concept is based on the Kalman filter formalism and is designed to avoid the inversion of large matrices.

Most notably, the KAA strikes a balance between conventional global and local track-based alignment algorithms, by restricting the computation of alignment parameters not only to alignable objects hit by the same track, but also to all other alignable objects that are significantly correlated. Nevertheless, this feature also comes with various trade-offs: Mechanisms are needed that affect which alignable objects are significantly correlated and keep track of these correlations. Due to the large amount of alignable objects involved at each update (at least compared to local alignment algorithms), the time spent for retrieving and writing alignment parameters as well as the required user memory (RAM) becomes a significant factor.

The full-scale test presented here, i.e., the employment of the KAA to the (misaligned) CMS Tracker, demonstrates the feasibility of the algorithm in a realistic scenario. It is shown that both the computation time and the amount of required user memory are within reasonable bounds, given the available computing resources, and that the obtained results are satisfactory.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS

Plenary / 482

Future of Grid Computing

Corresponding Author: miron@cs.wisc.edu

Poster 2 - Board: 98 / 391

GFAL and LCG-Util

Authors: Akos Frohner¹ ; David Smith¹ ; Jean-Philippe Baud¹ ; Krzysztof Nienartowicz¹ ; Lana Abadie¹ ; Remi Mollon¹ ; Sophie Lemaitre¹

¹ CERN

Corresponding Author: akos.frohner@cern.ch

GFAL, or Grid File Access Library, is a C library developed by LCG to give a uniform POSIX interface to local and remote Storage Elements on the Grid. LCG-Util is a set of tools to copy/replicate/delete files and register them in a Grid File Catalog.

In order to match experiment requirements, these two components had to evolve. Thus, the new Storage Resource Manager interface, SRM v2.2, is now supported.

In addition to that, important requirements were to have a python interface to GFAL/LCG-Util, and to fully support Logical File Name (LFN) at GFAL level.

Data privacy is a very important issue for some people. Therefore, we have to integrate Hydra client into GFAL. It allows to manage encrypted files and the cryptographic keys. Others important topics of development are to optimize the number of requests to BDII, and also to provide Perl API to GFAL and LCG-Util.

Grid middleware and tools / 118

GRATIA, a resource accounting system for OSG

Authors: Christopher Green¹ ; Philippe Canal¹

Co-authors: Jeff Mack¹ ; Penelope CONSTANTINA¹

¹ FERMILAB

Corresponding Author: philippe.canal@cern.ch

We will review the architecture and implementation of the accounting service for the Open Science Grid. Gratia's main goal is to provide the OSG stakeholders with a reliable and accurate set of views of the usage of resources across the OSG.

We will review the status of deployment of Gratia across the OSG and its upcoming development. We will also discuss some aspects of current OSG usage as illustrated by information provided by Gratia.

Poster 2 - Board: 88 / 313

GSIMF: A Grid Software Installation Management Framework

Authors: Alexandre Vaniachine¹ ; Balamurali Ananthan² ; Edward May¹ ; Gerald Gieraltowski¹ ; Nanbor Wang²

¹ *Argonne National Laboratory*

² *Tech-X Corporation*

Corresponding Author: vaniachine@anl.gov

To process the vast amount of data from high energy physics experiments, physicists rely on Computational and Data Grids; yet, the distribution, installation, and updating of a myriad of different versions of different programs over the Grid environment is complicated, time-consuming, and error-prone.

We report on the development of a Grid Software Installation Management Framework (GSIMF) for managing versioned files with software applications and file-based databases over the Grid infrastructure. A set of Grid services and tools automates the software and databases installation management process by installing and removing software packages on behalf of users.

We have developed key prototype Grid services for querying available software packages and installing software on distributed Grid computing elements, demonstrated the prototype services using an ATLAS analysis software package, and explored and investigated other software and databases installation management strategies. We report on hardening the implementations and enhancing the functionality of the prototype Grid services and studies of the applicability of GSIMF for the installation and management of database releases for the ATLAS experiment at the LHC.

The new Grid management framework should enable users to remotely install programs and tap into the computing power provided by Grids. The new management framework should find use in various data intensive and collaborative applications, such as nuclear physics experiments, space science observations, and climate modeling.

Distributed data analysis and information management / 146

Ganga - a job management and optimising tool

Author: Andrew Maier¹

Co-authors: Adrian MURARU¹ ; Alexander SOROKO² ; Benjamin GAIDIOZ¹ ; Chun Lik TAN³ ; Dietrich LIKO¹ ; Frederic BROCHU⁴ ; Hurng-Chun LEE⁵ ; Jakub MOSCKICKI¹ ; Johannes ELMSHEUSER⁶ ; Karl HARRISON⁷ ; Ulrik EGEDE⁸ ; Vladimir ROMANOVSKY⁹

¹ *CERN*

² *University of Oxford*

³ *University of Birmingham*

⁴ *University of Cambridge*

⁵ *ASGC, Taipei and CERN*

⁶ *Ludwig-Maximilians-Universität, München*

⁷ *High Energy Physics Group, Cavendish Laboratory*

⁸ *Imperial College London*

⁹ *State Res.Center of Russian Feder. Inst.f.High Energy Phys. (IFVE)*

Corresponding Author: andrew.maier@cern.ch

Ganga, the job-management system (<http://cern.ch/ganga>), developed as an ATLAS- LHCb common project, offers a simple, efficient and consistent user experience in a variety of heterogeneous environments: from local clusters to global Grid systems. Ganga helps end-users to organise their analysis activities on the Grid by providing automatic persistency of the job's metadata. A user has full access to the job history including their configuration and input/output. It is however important that users can see a single environment for developing and testing algorithms locally and for running on large data samples on the Grid. The tool allows for some basic monitoring and a steadily increasing number of users of more than 300 users have been confirmed, both in HEP, as well as in non-HEP applications. The paper will introduce the Ganga philosophy, the Ganga architecture and current and future strategy. It will use the example of how an LHCb user performs his analysis using Ganga and will describe the experiences gathered so far with the tool in LHCb.

Grid middleware and tools / 68

Geographical failover for the EGEE-WLCG Grid collaboration tools

Authors: Alessandro Cavalli¹ ; Alfredo Pagano¹ ; Cyril L'Orphelin² ; Gilles Mathieu² ; Osman Aidel²

Co-author: Rafal Lichwala³

¹ INFN/CNAF, Bologna, Italy

² IN2P3/CNRS Computing Centre, Lyon, France

³ PSNC, Poznan, Poland

Corresponding Authors: alfredo.pagano@cnaf.infn.it, alessandro.cavalli@cnaf.infn.it, gilles.mathieu@in2p3.fr, cyril.lorphelin@cc.in2p3.fr, oaidel@cc.in2p3.fr

Worldwide grid projects such as EGEE and WLCG need services with high availability, not only for grid usage, but also for associated operations. In particular, tools used for daily activities or operational procedures are considered critical. In this context, the goal of the work done to solve the EGEE failover problem is to propose, implement and document well-established mechanisms and procedures to limit service outages for the operations and monitoring tools used by regional and global grid operators to control the status of the EGEE grid.

The operations activity of EGEE relies on different tools developed by teams from different countries. For each tool, only one instance was deployed prior to this work, thus representing single points of failure. In our work, we solved the problem by replicating tools in different sites, using specific DNS features to automatically swap a given service instance in case of failures. After a DNS test phase in a virtual machine (vm) environment focused on nsupdate, NS/zone configuration and fast TTLs, a new domain for grid operations (gridops.org) was registered. In addition, replication of databases, web servers and web services have also been investigated and configured.

In this paper, we describe the technical mechanism used in our approach. We also show the replication procedure implemented for the EGEE/WLCG CIC Operations Portal use case. Furthermore, we present the interest in failover procedures in the context of other grid projects and grid services. Future plans for improvements of the procedures are also described.

Poster 1 - Board: 29 / 100

Glance Project: a database retrieval mechanism for the ATLAS detector

Authors: Carmen Maidantchik¹ ; Felipe Fink Grael¹ ; Kaio Karam Galvao¹ ; Kathy Pomes²

¹ *Universidade Federal do Rio de Janeiro (UFRJ)*

² *CERN*

Corresponding Authors: kathy.pomes@cern.ch, felipe.grael@cern.ch

During the construction and commissioning phases of the ATLAS Collaboration, data related to the installation, testing and performance of the equipment are stored in distinctive databases. Each group acquires information and saves them in repositories placed in different servers, using diverse technologies. Both data modeling and terminology may vary among the storage areas. The development of retrieval systems for each data set would require too much effort and high maintenance cost.

The goal of the Glance Project is to provide navigation mechanisms among the databases, which is independent of both the technology used to build them and their relationship. The browsing over the data sets results in hypertext tables and links on the Web. The user chooses the database and the system shows its structure. After selecting a partition of the repository, the system automatically creates a retrieval interface that allows the specification of search parameters. The interface can be customized for specific needs and saved in order to be easily accessed later. Glance corresponds to a single system that handles distinctive recovery mechanisms among diverse databases. Therefore, further knowledge about how data is organized and labeled is not required to perform queries. Maintenance costs are also minimized. This paper describes the Glance conception, its development and functionalities. The system usage is illustrated with some examples. Current status and future work are also discussed.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Computer facilities, production grids and networking / 76

Global Grid User Support - Building a worldwide distributed user support infrastructure

Author: Torsten Antoni¹

Co-authors: Alistair Mills² ; Diana Bosio²

¹ *Forschungszentrum Karlsruhe*

² *CERN*

The organization and management of the user support in a global e-science computing infrastructure such as EGEE is one of the challenges of the grid. Given the widely distributed nature of the organisation, and the spread of expertise for installing, configuring, managing and troubleshooting the grid middleware services, a standard centralized model could not be deployed in EGEE. This paper presents the model used in EGEE for building a reliable infrastructure for user, virtual organisation and operations support. The model for supporting a production quality infrastructure for scientific applications will be described in detail. The advantages of the chosen model will be presented and the possible difficulties will be discussed. We will describe the ongoing efforts to build a worldwide grid user support infrastructure in the framework of WLCG by achieving interoperability between the EGEE and OSG user support systems. In this paper we will also describe a scheme of how knowledge

management can be used in grid user support and first steps towards a realisation in the framework of the EGEE user support infrastructure.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

EGEE/WLCG

Globus BOF / 498

Globus BOF

Corresponding Author: fraser@mcs.anl.gov

Globus software was developed to enable previously disconnected communities to securely share computational resources and data that span organizational boundaries. As a community driven project, the Globus community is continually creating and enhancing Grid technology to make it easier to administer Grids as well as lowering the barriers to entry for both Grid users and Grid developers. In this presentation we will give a brief snapshot of Globus technology today and provide a look into the future of Globus development and how communities are contributing toward an even brighter future tomorrow.

Dr. Dan Fraser is the Director of the Community Development & Improvement program for Globus Software, and is also head of the Globus GridFTP team. Formerly he was the Senior Architect for Grid Middleware at Sun Microsystems. He has over a decade of experience in designing and implementing grid solutions. He has a Ph.D. in physics from Utah State University. His current research interests include Service Oriented Science, grid metrics, and in making Grid technologies easier to use and generally more accessible.

497

Globus BOF Sessions

Globus software was developed to enable previously disconnected communities to securely share computational resources and data that span organizational boundaries. As a community driven project, the Globus community is continually creating and enhancing Grid technology to make it easier to administer Grids as well as lowering the barriers to entry for both Grid users and Grid developers. In this presentation we will give a brief snapshot of Globus technology today and provide a look into the future of Globus development and how communities are contributing toward an even brighter future tomorrow.

Dr. Dan Fraser is the Director of the Community Development & Improvement program for Globus Software, and is also head of the Globus GridFTP team. Formerly he was the Senior Architect for Grid Middleware at Sun Microsystems. He has over a decade of experience in designing and implementing grid solutions. He has a Ph.D. in physics from Utah State University. His current research interests include Service Oriented Science, grid metrics, and in making Grid technologies easier to use and generally more accessible.

Poster 2 - Board: 111 / 461

Grid Information system for EGEE, scalability and performance assessment and plans

Authors: Felix Ehm¹ ; Laurence Field¹

¹ CERN

Corresponding Author: laurence.field@cern.ch

Grid Information Systems are mission-critical components for production grid infrastructures. They provide detailed information which is needed for the optimal distribution of jobs, data management and overall monitoring of the Grid. As the number of sites within these infrastructure continues to grow, it must be understood if the current systems have the capacity to handle the extra load. EGEE is the worlds largest production grid infrastructure and hence puts the greatest demands on its information system.

Currently 230 sites publish 23 MBytes every 2 minutes into the system and single top level nodes have seen 2 million queries per day. We expect this to grow in the next two years by a factor 5 to 10.

This paper describes the current requirements for the EGEE information system obtained from monitoring the usage patterns. From this data, a number of test conditions are derived for determining the subsequent performance and scalability of a grid information system. The tests are then applied to the BDII, which is the information system implementation used by the EGEE infrastructure. The results are given and show the limits of the underlying OpenLDAP Server implementation used within BDII. As a consequence, new techniques and concepts for enhancing the performance are suggested.

Grid middleware and tools / 134

Grid Interoperability: Joining Grid Information Systems

Author: Martin Flechl¹

Co-author: Laurence Field²

¹ IKP, Uppsala Universitet

² IT Department, CERN

Corresponding Author: martin.flechl@cern.ch

A Grid is defined as being “coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations”. Over recent years a number of grid projects, many of which have a strong regional presence, have emerged to help coordinate institutions and enable grids. Today, we face a situation where a number of grid projects exist, most of which have slightly different middleware. Grid interoperation is trying to bridge these differences and enable virtual organizations to access resources at the institutions independent of the grid project affiliation.

Grid interoperation is usually a bilateral activity between two grid infrastructures. Recently within the Open Grid Forum, the Grid Interoperability Now (GIN) Community Group is trying to build upon these bilateral activities. The GIN group is a focal point where all the infrastructures can come together to share ideas and experiences on grid interoperation. It is hoped that each bilateral activity will bring us one step closer to the overall goal of a uniform grid landscape.

A fundamental aspect of a grid is the information system, which is used to find available grid services. As different grids use different information systems, interoperation between these systems is crucial for grid interoperability. This paper describes the work carried out between a number of grid projects to overcome these differences. It focuses on the different techniques used and highlights the important areas for future standardization.

Summary:

The work carried out between large grid projects (EGEE, NDGF, OSG, Teragrid, PRAGMA, NAREGI) to reach interoperation between their information systems is described. The different information systems are presented, techniques used to overcome the differences between the grid projects are explained and important aspects for future standardization are identified.

Plenary / 486

Grid Interoperability: The Interoperations Cookbook

Corresponding Author: laurence.field@cern.ch

Over recent years a number of grid projects have emerged which have built grid infrastructures that are now the computing backbones for various user communities. A significant number of these user communities are artificially limited to only one grid due the different middleware used in each grid project. Grid interoperation is trying to bridge these differences and enable virtual organizations to access resources independent of the grid project affiliation.

The presentation gives an overview of grid interoperation and describes the current methods used to bridge the differences between grids. Actual use cases encountered during the last three years are discussed and the most important interfaces required for interoperation are highlighted. A summary of the standardisation efforts in these areas is given and arguments are made for moving more aggressively towards standards.

Grid middleware and tools / 306

Grid Monitoring from the VO/User perspective. Dashboard for the LHC experiments.

Authors: Benjamin Gaidioz¹ ; Catalin Cirstoiu¹ ; Gerhild Maier¹ ; Juha Herrala¹ ; Julia Andreeva¹ ; Pablo Saiz¹ ; Ricardo Rocha¹

Co-authors: Fu-Ming Tsai² ; Shih-Chun Chiu² ; Tao-Sheng Chen²

¹ CERN

² ASGC

Corresponding Author: julia.andreeva@cern.ch

The goal of the Grid is to provide a coherent access to distributed computing resources. All LHC experiments are using several Grid infrastructures and a variety of the middleware flavors. Due to the complexity and heterogeneity of a distributed system the monitoring represents a challenging task. Independently of the underlying platform, the experiments need to have a complete and uniform picture of their activities on the Grid ideally seen by the users as a single powerful computing resource. Overall operation of the infrastructure used by experiments is defined both by the quality of the Grid and the quality of the tools and services developed/used by the

experiments. Correspondingly the required monitoring information should combine both Grid-related and experiment/application specific data. On the other hand, users of the LHC experiments have various roles and need different levels of details regarding monitoring data.

The paper will focus on the Grid monitoring from the experiment/user perspectives with a closer look to the Experiment Dashboard system and its current status.

Experiment Dashboard is currently in use by all four LHC experiments. It provides both Grid-related and Experiment-specific monitoring information and works across several middleware platforms (LCG, Glite, OSG).

Distributed data analysis and information management / 354

Grid data storage on widely distributed worker nodes using Scalla and SRM

Authors: Jerome LAURET¹ ; Pavel Jakl²

Co-authors: Alex SIM³ ; Alexandru ROMOSAN³ ; Andrew HANUSHEVSKY⁴ ; Arie SHOSHANI³

¹ BROOKHAVEN NATIONAL LABORATORY

² Nuclear Physics Institute, Academy of Sciences of the Czech Republic

³ Lawrence Berkeley Laboratory

⁴ Stanford Linear Accelerator Center

Corresponding Authors: pjakl@bnl.gov, jlauret@bnl.gov

Facing the reality of storage economics, NP experiments such as RHIC/STAR have been engaged in a shift in the analysis model, and now heavily rely on using cheap disks attached to processing nodes, as such a model is extremely beneficial over expensive centralized storage. Additionally, exploiting storage aggregates with enhanced distributed computing capabilities such as dynamic space allocation (lifetime of spaces), file management on shared storages (lifetime of files, pinning file), storage policies or a uniform access to heterogeneous storage solutions is not an easy task. The Xrootd/Scalla system allows for storage aggregation. We will present an overview of the largest deployment of Scalla (Structured Cluster Architecture for Low Latency Access) in the world spanning over 1000 CPUs co-sharing the 350 TB Storage Elements and the experience on how to make such a model work in the RHIC/STAR standard analysis framework. We will explain the key features and approach on how to make access to mass storage (HPSS) possible in such a large deployment context. Furthermore, we will give an overview of a fully “gridified” solution using the plug-and-play features of Scalla architecture, replacing standard storage access with grid middleware SRM (Storage resource manager) components designed for space management and will compare the solution with the standard Scalla approach in use in STAR for the past 2 years. Integration details, future plans and status of development will be explained in the area of best transfer strategy between multiple-choice data pools and best placement with respect of load balancing and interoperability with other SRM aware tools or implementations.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

STAR

Grid middleware and tools / 404

Grid reliability

Authors: Benjamin Gaidioz¹ ; Gerhild Maier¹ ; Juha Herrala¹ ; Julia Andreeva¹ ; Pablo Saiz¹ ; Ricardo Rocha¹ ; catalin Cirstoiu¹

¹ CERN

Corresponding Author: pablo.saiz@cern.ch

Thanks to the grid, users have access to computing resources distributed all over the world. The grid hides the complexity and the differences of its heterogeneous components. In order for this to work, it is vital that all the elements are setuped properly, and that they can interact with each other. It is also very important that errors are detected as soon as possible, and that the procedure to solve them is well established.

Our goal is to improve the performance of the grid. In order to do this, we studied two of its main elements: the workload and the data management systems. We developed all tools needed to investigate the efficiency of the different centres. Furthermore, our tools can be used to categorize the most common error messages, and measure their time evolution.

One common reason for job failures is site misconfiguration. Being able to detect such a misconfiguration as soon as possible helps in several ways: first of all, it minimizes the time that it takes to bring the site back to a normal state; moreover, debugging it is easier, since the problem happened in the recent past. This can be specially helpful for new centers, since the tools provide the material needed to get a better understanding of the grid's complexity.

In this contribution we will describe all the tools that we have developed to monitor the grid efficiency. These tools are currently used by the four LHC experiments. We will also describe the results and benefits that the tools have provided.

Poster 1 - Board: 113 / 466

GridKa Tier1 Site Management

Authors: Marc Hemberger¹ ; Sven Gabriel¹

¹ Forschungszentrum Karlsruhe

Corresponding Author: sveng@nikhef.nl

GridKa is the German Tier1 centre in the Worldwide LHC Computing Grid (WLCG). It is part of the Institut für Wissenschaftliches Rechnen (IWR) at the Forschungszentrum Karlsruhe (FZK). It started in 2002 as the successor of the "Regional Data and Computing Centre in Germany" (RDCCG). GridKa supports all four LHC experiments, ALICE, ATLAS, CMS and LHCb, four non-LHC high energy physics experiments as well as an astrophysics experiment and geclipse, a project that concentrates on the development of tools for existing Grid infrastructures..

In this presentation we will give an overview of the site fabric management and fabric management tools. The usage of Rocks tool kit for worker node cluster installation is discussed.

Most of the grid middleware services are deployed in Xen-based virtual machines on a variety of physical machines.

We describe the integration of Yaim and Cfengine for installation and configuration of middleware services.

We will also show how tools like Nagios and Cfengine can be combined to initiate recovery actions to ensure functionality of middleware services.

Poster 2 - Board: 7 / 206

HEP System Management Working Group and WEB site

Author: Alessandra Forti¹

Co-author: Michel Jouvin²

¹ *University of Manchester*

² *LAL (IN2P3)*

Corresponding Author: alessandra.forti@cern.ch

System Management Working Group (SMWG) of sys admins from Hepix and grid sites has been setup to address the fabric management problems that HEP sites might have. The group is open and its goal is not to implement new tools but to share what is already in use at sites according to existing best practices. Some sites are already publicly sharing their tools and sensors and some other sites do write very good documentation and share it. The aim is to extend this to a general practice and in a more organised way and avoid the duplication of effort that occurs when system administrators are solving mostly the same problems over and over. The result has been the creation of a WEB site (www.sysadmin.hep.ac.uk) that hosts a subversion repository for management and monitoring tools and a wiki. It works as a file sharing system and single entry point for documentation distributed in other sites. The site, based on gridsite. This paper describes how the group is working and what has been achieved so far.

Summary:

Cooperative effort to collect documentation and tools from HEP sites sys admin in a single accessible place.

Plenary / 494

HEP and Non-HEP Computing at a Laboratory in Transition

Corresponding Author: richard.mount@slac.stanford.edu

Computer facilities, production grids and networking / 347

HEP grids face IPv6: A readiness study

Author: Matt Crawford¹

¹ *FERMILAB*

Corresponding Author: matt.crawford@cern.ch

Due to shortages of IPv4 address space - real or artificial - many HEP computing installations have turned to NAT and application gateways. These workarounds carry a high cost in application complexity and performance. Recently a few HEP facilities have begun to deploy IPv6 and it is expected that many more must follow within several years. While IPv6 removes the problem of address shortages and its painful workarounds, it comes at some initial price in software and network infrastructure evolution.

Routers and host protocol stacks have been ready for IPv6 for quite some years, many major backbone networks carry IPv6 and peer with each other, and site network management applications are available. Now application and security considerations are on the critical path to full exploitation of IPv6. We examine the steps required for grid applications (storage and computation) and security mechanisms and site network infrastructure (DNS, DHCP, access control policies) to move to a mixed v4/v6 environment.

Summary:

Requirements for moving grid facilities to mixed IPv4/v6 operation.

Poster 1 - Board: 43 / 270

HEPTrails: An Analysis Workflow and Provenance Tracking Application

Author: Christopher Jones¹

¹ *Cornell University*

When doing an HEP analysis, physicists typically repeat the same operations over and over while applying minor variations. Doing the operations as well as remembering the changes done during each iteration can be a very tedious process. HEPTrails is an analysis application written in Python and built on top of the University of Utah's VisTrails system which provides workflow and full provenance tracking for scientific analysis. HEPTrails adds substantial extension to VisTrails in order to accommodate HEP analysis. These extensions include: a streaming workflow engine and with very fine grain modules, a lab notebook style presentation window capable of showing preliminary results, and local area distributed computing. Although HEPTrails is still in early development, it already shows great promise in aiding HEP physicists.

Poster 1 - Board: 106 / 388

HLRmon: a Role-based Grid Accounting Reporting Web Tool

Authors: Andrea Guarise¹; Cristina Vistoli²; Davide Salomoni²; Enrico Fattibene²; Federico Pescarmona¹; Felice Rosso²; Giuseppe Misurelli²; Giuseppe Patania¹; Luciano Gaido¹; Paolo Veronesi²; Riccardo Brunetti¹; Stefano Dal Pra³

¹ *INFN-Torino, Italy*

² *INFN-CNAF, Bologna, Italy*

³ *INFN-Padova, Italy*

Corresponding Authors: stefano.dalpra@pd.infn.it, enrico.fattibene@cnaf.infn.it, giuseppe.misurelli@cnaf.infn.it, federico.pescarmona@to.infn.it

In production quality Grid infrastructure accounting data play a key role on the possibility to spot out how the allocated resources have been used. The different types of Grid user have to be taken into account in order to provide different subsets of accounting data based on the specific role covered by a Grid user.

Grid end users, VO (Virtual Organization) managers, site administrators and Grid operators need different ways to get resource usage statistics about jobs executed in a given time period at various different levels, depending on their specific Grid role. In the framework of the Italian production Grid, HLRmon aims at generating suitable reports for various categories of Grid users and has been designed to serve them within an unified layout.

Thanks to its ability to authenticate web users through personal certificate and related authorization rights, it can a-priori restrict the selectable items range offered, so that proper information can only be provided to specifically enabled people. Information are gathered by HLRmon from the accounting database (HLR, Home Location Register), which stores complete accounting data in a per job basis.

Depending on the kind of reports that are to be generated, it directly queries the HLR using an ad-hoc DGAS (Distributed Grid Accounting System) query tool (typically user's level detail info), or a local RDBMS table with daily aggregated information in a per day, site, VO basis, thus saving connection delay time and needless load on the HLR.

Online computing / 286

HLT Online Calibration framework in ALICE

Author: Sebastian Robert Bablok¹

Co-authors: Bernhard Skaali²; Dieter Röhrich³; Gaute Øvrebekk³; Harald Appelshäuser⁴; Håvard Helstrup⁵; Jochen Thäder⁶; Kjetil Ullaland³; Mateusz Ploskon⁴; Matthias Richter³; Per Thomas Hille²; Sorina Popescu⁶; Timm Steinbeck⁶; Torsten Alt⁶; Volker Lindenstruth⁶

¹ *Department of Physics and Technology, University of Bergen*

² *Department of Physics, University of Oslo, Norway*

³ *Department of Physics and Technology, University of Bergen, Norway*

⁴ *Institute for Nuclear Physics, University of Frankfurt, Germany*

⁵ *Faculty of Engineering, Bergen University College, Norway*

⁶ *Kirchhoff Institute of Physics, University of Heidelberg, Germany*

Corresponding Author: sebastian.bablok@ift.uib.no

The ALICE HLT is designed to perform event analysis including calibration of the different ALICE detectors online. The detector analysis codes process data using the latest calibration and condition settings of the experiment. This requires a high reliability on the interfaces to the various other systems operating ALICE.

In order to have a comparable analysis with the results from Offline, HLT requests the same storage for calibration data, Offline Calibration Database (OCDB). A local caching of its content guarantees a fast and permanent availability of the calibration data during a run. In addition, interactions with the other ALICE online systems (Detector Control System (DCS) and Experiment Control System (ECS)) provide current running conditions like temperatures, trigger settings or the current run number, and allow for synchronizing among the current states.

Calibration objects, which are produced online in the HLT cluster, have to be stored in the OCDB after each run, before they can be reused inside the HLT. This guarantees proper versioning of the data and correct assignments of the produced results to the applied settings.

A set of dedicated portal nodes of the HLT cluster cover these tasks and take care of the internal distribution and collection of the data as well as of communication and data transfer with the other online and offline systems. Fail safety and redundancy in the design of these interfaces avoids single points of failure and reduces the risk of time delays or data loss.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ALICE

Plenary / 483

HPC at the Petascale and Beyond

IBM's Blue Gene/L system had demonstrated that it is now feasible to run applications at sustained performances of 100's of teraflops. The next generation Blue Gene/P system is designed to scale up to a peak performance of 3.6 Petaflops. This talk will look at some of the key application successes already achieved at the 100TF scale. It will then address the emerging petascale architectures and look at the challenges which arise as the HPC world now starts to consider designing 100 Petaflop and Exaflop Systems. These challenges are very significant and include power, memory bandwidth, network bandwidth, reliability, systems software, and applications.

Summary:

James Sexton:

(B.A. [Mod.], Theoretical Physics, Trinity College Dublin, 1980; Ph.D., Theoretical Physics, Columbia University, 1986) is Research Staff Member at the Thomas J. Watson Research Center of IBM Corporation, where he is responsible for applications for IBM Blue Gene architectures. Dr. Sexton has had prior appointments at Fermilab (1984-1986), at the Princeton Institute for Advanced Studies (1986-1988), at IBM T. J. Watson Research Center (1988-1991).

Prior to his arrival at IBM in 2004, Sexton was Professor of Physics at Trinity College Dublin. Among his honors, he is a multiple winner of the Gordon Bell Prize in high-performance computing (two awards in 2006, one in 2005) in collaboration with LLNL. Although a theoretical high-energy physicist by background, his areas of interest cover broad ground. His output includes publications in quantum field theory, quantum chromodynamics (QCD), statistical mechanics, computational physics, molecular dynamics, and medicine.

505

HPC at the Petascale and Beyond

Author: James Sexton¹

¹ IBM

IBM's Blue Gene/L system had demonstrated that it is now feasible to run applications at sustained performances of 100's of teraflops. The next generation Blue Gene/P system is designed to scale up to a peak performance of 3.6 Petaflops. This talk will look at some of the key application successes already achieved at the 100TF scale. It will then address the emerging petascale architectures and look at the challenges which arise as the HPC world now starts to consider designing 100 Petaflop and Exaflop Systems. These challenges are very significant and include power, memory

bandwidth, network bandwidth, reliability, systems software, and applications.

Software components, tools and databases / 203

HepData: restructuring the archive for LHC data

Author: Mike Whalley¹

Co-author: Andy Buckley¹

¹ *IPPP, University of Durham*

Corresponding Authors: james.william.monk@cern.ch, m.r.whalley@durham.ac.uk

The Durham HepData database has for many years provided an up-to-date archive of published numerical data from HEP experiments worldwide. In anticipation of the abundance of new data expected from the LHC, the database is undergoing a complete metamorphosis to add new features and improve the scope for use of the database by external applications. The core of the HepData restructuring is the use of a relational database server in place of the legacy hierarchical system, and the use of a Java object model to abstract the database operations into object relationships. Additionally, an XML dialect, HepML, has been developed to describe data records: this provides a rich description of HEP datasets for use by the database migration system and by experiments wishing to submit their own data records. Standard Java persistency systems are used both for the object-relational and object-XML mappings. A new user front end is being developed, using Java Web application technology. This will provide easy user access to HepData's records and flexible output formats, including data comparisons. Furthermore, methods are in development to allow experimental collaborations to input and maintain their own data in a secure way. The re-development of HepData is part of the CEDAR project, which also involves the JetWeb and Rivet event generator tuning systems. HepData's role as a reference source for generator tunings is pivotal to the success of JetWeb and Rivet. In this paper we describe the current status of the development of the new HepData database.

Online computing / 312

High Level Trigger Configuration and Handling of Trigger Tables in the CMS Filter Farm

Authors: David Lange¹; Emilio Meschi²; Jonathan Hollar¹; Philipp Schieferdecker²; Ulf Behrens³

Co-authors: Alexander Oh²; Andrea Carboni²; Andrea Petrucci⁴; Attila Racz²; Christoph Paus⁵; Christoph Schwick²; Dimitrios Tsirigkas²; Dominique Gigi²; Elliot David Lipeles⁵; Eric Cano²; Esteban Gutierrez Mlot⁵; Frank Glege²; Frans Meijers²; Gaetano Maron⁴; Gerry Bauer⁵; Hannes Sakulin²; Ichiro Suzuki⁶; James Branson⁷; Jin Cheol Kim⁸; Joao Varela⁹; Johannes Gutleber²; Juan Antonio Lopez Perez²; Konstanty Sumorok⁵; Luciano Orsini²; Lucien Pollet²; Marco Pieri⁷; Marek Ciganek²; Markus Klute⁵; Matteo Sani⁷; Michele Gulmini⁴; Robert Gomez-Reino²; Roland Moser²; Samim Erhan¹⁰; Sergio Cittolin²; Steven Murray⁶; Vincent Boyer²

¹ *LLNL*

² *CERN*

³ *DESY*

⁴ *INFN Legnaro*

⁵ *MIT*

⁶ *FNAL*

⁷ UCSD

⁸ Kyungpook National University

⁹ LIP

¹⁰ CERN, UCLA

Corresponding Author: emilio.meschi@cern.ch

The CMS experiment at the CERN Large Hadron Collider is currently being commissioned and is scheduled to collect the first pp collision data towards the end of 2007.

CMS features a two-level trigger system. The Level-1 trigger, based on custom hardware, is designed to reduce

the collision rate of 40 MHz to approximately 100 kHz. Data for events accepted by the Level-1 trigger are read

out and assembled by an Event Builder through a complex of switched networks. The High Level Trigger (HLT),

running on a computing farm consisting of standard CPU (Filter Farm), employs a set of sophisticated software

algorithms, based on the same full-fledged reconstruction framework used for offline, to analyze the complete

event information, and further reduce the accepted event rate by approximately three orders of magnitude.

This paper describes the design and implementation of the HLT configuration management system. The creation of an HLT configuration, consisting of many software modules organized in a number of trigger

paths, and its deployment into the distributed online environment consisting of O(1000) CPU, are centered

around a robust database design, abstracting the features of the algorithms and their organization in a trigger

table. The evolution of the underlying code, and the issues related to migration of existing tables across software

releases, are addressed by a thin code parsing layer. The population of tables, using a dedicated GUI, their

retrieval by the Run Control System for deployment in the HLT, and access to historic data all use a unique

interface. Reformatting and deployment are decoupled from the database, thus permitting the target configuration grammar to evolve independently.

Referential integrity and data consistency are expected to be guaranteed by this system across the entire lifetime

of the experiment.

First experiences from the commissioning of the HLT system are also reported.

Poster 2 - Board: 91 / 357

High Performance Data Analysis for Particle Physics using the Gfarm File System

Authors: Ichiro Adachi¹ ; Nobuhiko Katayama² ; Osamu Tatebe³ ; Shohei Nishida¹

¹ HIGH ENERGY ACCELERATOR RESEARCH ORGANIZATION

² High Energy Accelerator Research Organization

³ Tsukuba University

Corresponding Author: nobu.katayama@kek.jp

The Belle experiment operates at the KEKB accelerator, a high luminosity asymmetric energy $e^+ e^-$ collider. The Belle collaboration studies CP violation in decays of B meson to answer one of the fundamental questions of Nature, the matter-anti-matter asymmetry. Currently, Belle accumulates more than one million B \bar{B} meson pairs that correspond to about 1.2 TB of raw data in one day. The amount of raw data is expected to increase by 50 times after an upgrade of the KEKB accelerator.

The challenge is how to realize required high performance data access and scalable data computing. Our solution is the Gfarm file system. It is a commodity-based Grid-wide network shared file system that federates local storage of cluster nodes; moreover it provides scalable I/O performance with distributed data access. We constructed a Gfarm file system with 26 TB capacity and 52 GB/sec I/O bandwidth, integrating local disks of 1112 compute nodes in the KEKB computing facility, and measured scalability of disk I/O performance up to more than 1000 nodes. We also performed a real Belle data analysis program using more than 700 nodes at the speed of 24GB/s, reducing the Belle data analysis time by a factor of about 1,000.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

Belle

Poster 1 - Board: 70 / 81

High Performance Storage Tests At The INFN Pisa Computing Centre

Author: Enrico Mazzoni¹

Co-authors: Alberto Ciampa¹ ; Maurizio Davini² ; Silvia Arezzini¹ ; Subir Sarkar¹

¹ *INFN Pisa*

² *Universita' di Pisa*

Corresponding Author: enrico.mazzoni@pi.infn.it

We report about the tests performed in the INFN Pisa Computing Centre with some of the latest generation storage devices. Fibre Channel and NAS solutions have been tested in a realistic environment, both participating in Worldwide CMS's Service Challenges, and simulating analysis patterns with more than 500 jobs accessing concurrently]data files. Both usage pattern have evidenced the ability to use today's storage links at 10 Gbit/s, allowing a steady transfer rate exceeding what a single Gbit/s interface cannot guarantee when the number of concurrent users increase over a few hundred nodes.

Summary:

We report about the tests performed in the INFN Pisa Computing Centre with some of the latest generation storage devices. Fibre Channel and NAS solutions have been tested in a realistic environment, both participating in Worldwide CMS's Service Challenges, and simulating analysis patterns with more than 500 jobs accessing concurrently]data files. Both usage pattern have evidenced the ability to use today's storage links at 10 Gbit/s, allowing a steady transfer rate exceeding what a single Gbit/s interface cannot guarantee when the number of concurrent users increase over a few hundred nodes.

Poster 2 - Board: 41 / 113

High precision physics simulation: experimental validation of Geant4 Atomic Relaxation

Authors: Alfonso Mantero¹ ; Maria Grazia Pia²

¹ *INFN Genova*

² *INFN GENOVA*

Corresponding Author: maria.grazia.pia@cern.ch

A component of the Geant4 toolkit is responsible for the simulation of atomic relaxation: it is part of a modelling approach of electromagnetic interactions that takes into account the detailed atomic structure of matter, by describing particle interactions at the level of the atomic shells of the target material.

The accuracy of Geant4 Atomic Relaxation has been evaluated against the experimental measurements of the NIST Physical Reference Data, which include a systematic review of the experimental body of knowledge collected and evaluated over several decades of experimental activity.

The validation study concerns X-ray and Auger transition energies.

The comparison of the simulated and experimental data with rigorous statistical methods demonstrates the excellent accuracy of the Geant4 simulation models; precision better than 0.5 % is achieved in most cases.

The results of this validation study are relevant to various experimental fields, both for elemental analysis studies and for precise simulation of energy deposit distributions; they are also important for the design and optimization of novel tracking detectors based on nanotechnologies, which are sensitive to the effects of charged particles travelling short path lengths, like Auger electrons.

Poster 1 - Board: 26 / 66

High-Performance Stream Computing for Particle Beam Transport Simulations

Authors: David Bailey¹ ; Robert Appleby¹

¹ *University of Manchester*

Corresponding Author: david.bailey-2@manchester.ac.uk

Understanding modern particle accelerators requires simulating charged particle transport through the machine elements. These simulations can be very time consuming due to the large number of particles and the need to consider many turns of a circular machine. Stream computing offers an attractive way to dramatically improve the performance of such simulations by calculating the simultaneous transport of many particles using dedicated hardware. Modern Graphics Processing Units (GPUs) are powerful and affordable stream computing devices. The results of simulations of particle transport through a FODO-cell transfer line, including an aperture model, using an NVidia GeForce 7900 GPU are compared to conventional transport codes. Accuracy and potential speed increases are compared and the prospects for future work in the area are discussed.

Plenary / 490

How good is the match between LHC software and current/future processors?

Corresponding Author: sverre.jarp@cern.ch

In the CERN openlab we have looked at how well LHC software matches the execution capabilities of current and, to some extent, future processors. Thanks to current silicon processes, transistor counts in the billions (10^9) have become commonplace and microprocessor manufacturers have been deploying transistors in multiple ways to increase performance. In this talk I will review the various architectural enhancements we have observed in the past and comment on the usefulness for HEP software. I will also make some suggestions for tuning our software, and finally speculate on how well our software will fit some of the possible future processor designs.

Summary:

Sverre Jarpe is CTO in the CERN openlab, a joint collaboration with leading industrial partners in order to assess cutting-edge information technology for the Large Hadron Collider's Computing Grid. He has been working in computing at CERN for over 30 years and has held various managerial and technical positions promoting advanced but cost-effective computing solutions. S. Jarpe holds a degree in Theoretical Physics from the Norwegian University of Science and Technology in Trondheim.

Collaborative tools / 343

HyperNews use in HEP - bigger and better

Author: Douglas Smith¹

Co-authors: Peter Elmer²; Terry Hung¹

¹ *Stanford Linear Accelerator Center*

² *Princeton University*

Corresponding Author: douglas@slac.stanford.edu

International multi-institutional high energy physics experiments require easy means for collaborators to communicate coherently in a global community. To fill this need, the HyperNews system has been widely used in HEP. HyperNews is a discussion management system which is a hybrid between a web-base forum system and a mailing list system. Its goal is to provide a tool for distributed collaborators to easily follow all discussions in a project, and not be limited to only the use of email. The HyperNews system was first presented at CHEP 2006, although it has been in use in HEP for more than a decade. Over the past year and a half it has been adopted by other sites and experiments, even outside of HEP. A number of new features have been developed, including features to handle attachments of files to postings, and improved management of members and access. Following increased problems with spam in communication systems features have been implemented to reduce spam in discussions. Experience with use in larger communities using HyperNews, as well as new and planned features will be presented.

Poster 1 - Board: 108 / 400

INFN Tier-1 status report

Authors: Alessandro Italiano¹ ; Cristina Vistoli¹ ; Davide Salomoni¹ ; Luca dell'Agnello¹ ; Mirco Mazzucato²

¹ *INFN-CNAF*

² *INFN-PADOVA*

INFN CNAF is a multi experiment computing center acting as Tier-1 for LCG but also supporting other HEP and non HEP experiments and Virtual Organizations.

The CNAF Tier-1 is one of the main Resource Centers of the Grid Infrastructure (WLCG/EGEE); the preferred access method to the center is through WLCG/EGEE and INFNGRID middleware and services.

Critical issues to be addressed to meet the requirements of the LHC experiments are stability, robustness and scalability of the services provided: in particular technical infrastructural services like power supply and cooling, CPU, data handling systems and network infrastructure to cope with the amount of data foreseen in the LHC era, database and experiment-specific services, monitoring, authorization and accounting, and 24x7 support.

In this poster we present the current status of the INFN Tier-1 infrastructure and services, an assessment of the experiences gained so far, and the expected evolution leading to the startup of LHC experiments.

ISSeG / 500

ISSEG: Welcome

ISSeG / 503

ISSeG: Feedback/questions

ISSeG / 501

ISSeG: How to sell Security - and get resources for it

Effective security needs resources and support from senior management. This session will look at some ways of gaining that support by establishing a common understanding of risk.

ISSeG / 504

ISSeG: Recommendations

This session will look at some of the emerging recommendations that can be used at sites to improve security.

ISSeG / 502

ISSeG: Risk assessments

This session will look to establish a common understanding of risk and introduce the ISSeG risk assessment questionnaire.

Plenary / 485

Identity Management

Corresponding Author: alberto.pace@cern.ch

54

Identity Management

Author: Alberto Pace¹

¹ CERN

Corresponding Author: alberto.pace@cern.ch

This talk will introduce identity management concepts and discuss various issues associated with its implementation. The presentation will try to highlight technical, legal, and social aspects that must be foreseen when defining the numerous processes that an identity management infrastructure must support.

Poster 1 - Board: 19 / 421

Implementation and Performance of the ATLAS Second Level Jet Trigger

Author: Patricia Conde Muíño¹

¹ LIP-Lisbon

Corresponding Author: patricia.conde.muino@cern.ch

ATLAS is one of the four major LHC experiments, designed to cover a wide range of physics topics. In order to cope with a rate of 40MHz and 25 interactions per bunch crossing, the ATLAS trigger system is divided in three different levels. The first one (LVL1, hardware based) identifies signatures in 2 microseconds that are confirmed by the the following trigger levels (software based). The Second Level Trigger (LVL2)

only looks at a region of the space around the LVL1 signature (called Region of Interest or ROI), confirming/rejecting the event in about 10 ms, while the Event Filter (Third Level Trigger, EF) has potential full event access and larger processing times, of the order of 1 s.

The jet selection starts at the LVL1 with dedicated processors that search for high ET hadronic energy depositions. At the LVL2, the jet signatures are verified with the execution of a dedicated, fast jet reconstruction algorithm. Given the fact that the main jet's background are jets, the energy calibration at the LVL2 is one of the major difficulties of this trigger, allowing to distinguish low/high energy jets. The algorithm for the calibration has been chosen to be fast and robust, with a good performance.

The other major difficulty is the execution time of the algorithm, dominated by the data unpacking time due to the large sizes of the jet ROI. In order to reduce the execution time, three possible granularities have been proposed and are being evaluated: cell based (standard), energy sums calculated at each Fron-End Board (FEB) and the use of the LVL1 Trigger Towers. The FEB and Trigger Tower granularities are also being used/evaluated for the reconstruction of the missing ET triggers at the Event Filter, given the short times available to process the full event. In this presentation, the design and implementation of the jet trigger of ATLAS will be discussed in detail, emphasising the major difficulties of each selection step. The performance of the jet algorithm, including timing, efficiencies and rates will also be shown, with detailed comparisons of the different unpacking modes.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS HLT group

Poster 2 - Board: 39 / 79

Implementation of INCL4 cascade with ABLA evaporation in Geant4

Author: Aatos Heikkinen¹

Co-authors: Alain Boudard²; Pekka Kaitaniemi¹

¹ *Helsinki Institute of Physics, HIP*

² *CEN, Saclay*

Corresponding Author: aatos.heikkinen@cern.ch

We introduce a new implementation of Liege cascade INCL4 with ABLA evaporation in Geant4.

INCL4 treats hadron, Deuterium, Tritium, and Helium beams up to 3 GeV energy, while ABLA provides treatment for light evaporation residues.

The physics models in INCL4 and ABLA and are reviewed with focus on recent additions.

Implementation details, such as first version of object oriented design,

are presented, and C++ performance is compared with original FORTRAN implementation.

Testing framework validating the FORTRAN-C++ translation is based on ROOT software.

Some of the advanced features in testing environment,

such as ROOT scripting and automatic documentation features are explained.

In addition, we introduce a new Geant4 example application using INCL4 and ABLA models,

that demonstrate the physics and compare results against other models

previously made available in Geant4, such as Bertini cascade.

Finally we outline the future development of Liege cascade, INCL5,

in the context of Geant4 hadronic physics framework.

Poster 2 - Board: 46 / 165

Implementation of chamber mis-alignments and deformations in the ATLAS Muon Spectrometer description and estimate of the muon reconstruction performance reconstruction performance

Authors: Ahmimed Ouraou¹ ; Daniela Rebutzi² ; Ivan Logachenko³ ; James Miller³ ; Jean Francois Laporte¹ ; Laurent Chevalier⁴ ; Matthias Schott⁵ ; Monica Verducci⁶ ; Nectarios Benekos⁷ ; Niels Van Eldik⁸ ; Robert Harrington³ ; Stefania Spagnolo⁹ ; Stephane Willocq⁸ ; Steven Goldfarb¹⁰

¹ *Centre d'Etudes de Saclay*

² *INFN Pavia and Pavia University*

³ *Boston University*

⁴ *DAPNIA*

⁵ *Ludwig-Maximilians-Univ. Muenchen*

⁶ *CERN*

⁷ *Max-Planck-Institut fuer Physik*

⁸ *University of Massachusetts*

⁹ *Universita' di Lecce, Italy*

¹⁰ *University of Michigan*

Corresponding Author: daniela.rebutzi@pv.infn.it

The Atlas Muon Spectrometer is designed to reach a very high transverse momentum resolution for muons in a p_T range extending from 6 GeV/c up to 1 TeV/c. The most demanding design goal is an overall uncertainty of 50 microns on the sagitta of a muon with $p_T = 1$ TeV/c. Such precision requires an accurate control of the positions of the muon detectors and of their movements during the experiment operation. Moreover, the light structure of the Muon Spectrometer, consisting mainly of drift tubes assembled in three layers of stations, imply sizable distortions of the nominal layout of individual chambers, due to mechanical stress and thermal gradients. Corrections for mis-alignments and deformations, which will be provided run-time by an optical alignment system, must be integrated in the software chain leading to track reconstruction and momentum measurement. Here we discuss the implementation of run-time dependent corrections for alignment and distortions in the detector description of the Muon Spectrometer along with the strategies for studying such effects in dedicated simulations. Some preliminary results obtained in the context of the ATLAS Condition Data Challenge effort are also presented.

Computer facilities, production grids and networking / 331

Implementing SRM V2.2 Functionality in dCache

Author: Timur Perelmutov¹

¹ *FERMILAB*

Corresponding Author: timur@fnal.gov

The Storage Resource Manager (SRM) and WLCG collaborations recently defined version 2.2 of the SRM protocol, with the goal of satisfying the requirement of the LCH experiments. The dCache team has now finished the implementation of all SRM v2.2 elements required by the WLCG. The new functions include space reservation, more advanced data transfer, and new namespace and permission functions. Implementation of these features required an update of the dCache architecture and

evolution of the services and core components of dCache Storage System.

Implementation of SRM Space Reservation led to new functionality in the Pool Manager and the development of the new Space Manager component of dCache, responsible for accounting, reservation and distribution of the storage space in dCache. SRM's "Bring Online" function required redevelopment of the Pin Manager service, responsible for staging files from the back-end tape storage system and keeping these files on disk for the duration of the Online state.

The new SRM concepts of AccessLatency and RetentionPolicy led to the definition of new dCache file attributes and new dCache pool code that implements these abstractions. SRM permission management functions led to the development of the Access Control List support in the new dCache namespace service, Chimera. I will discuss these new features and services in dCache, provide motivation for particular architectural decisions and describe their benefits to the Grid Storage Community.

Summary:

SRM v2.2 concepts and functions have induced an evolution of the dCache architecture and services.

Poster 2 - Board: 22 / 308

Implementing SRM V2.2 Functionality in dCache

Author: Timur Perelmutov¹

¹ *FERMI NATIONAL ACCELERATOR LABORATORY*

Corresponding Author: timur@fnal.gov

The Storage Resource Manager (SRM) and WLCG collaborations recently defined version 2.2 of the SRM protocol, with the goal of satisfying the requirement of the LCH experiments. The dCache team has now finished the implementation of all SRM v2.2 elements required by the WLCG. The new functions include space reservation, more advanced data transfer, and new namespace and permission functions. Implementation of these features required an update of the dCache architecture and evolution of the services and core components of dCache Storage System.

Implementation of SRM Space Reservation led to new functionality in the Pool Manager and the development of the new Space Manager component of dCache, responsible for accounting, reservation and distribution of the storage space in dCache. SRM's "Bring Online" function required redevelopment of the Pin Manager service, responsible for staging files from the back-end tape storage system and keeping these files on disk for the duration of the Online state.

The new SRM concepts of AccessLatency and RetentionPolicy led to the definition of new dCache file attributes and new dCache pool code that implements these abstractions. SRM permission management functions led to the development of the Access Control List support in the new dCache namespace service, Chimera. I will discuss these new features and services in dCache, provide motivation for particular architectural decisions and describe their benefits to the Grid Storage Community.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

dCache

Poster 1 - Board: 58 / 447

Implementing a Modular Framework in a Conditions Database Explorer for ATLAS

Authors: Antonio Amorim¹ ; Igor Soloviev² ; Joao Batista¹ ; Joao De Almeida Simoes¹ ; Lourenco Lopes¹ ; Paulo Pereira¹ ; Ricardo Neves¹ ; Serguei Kolos³

¹ *Universidade de Lisboa (SIM and FCUL, Lisbon)*

² *PNPI (Petersburg Nuclear Physics Institute)*

³ *University of California, Irvine (UCI, Irvine)*

Corresponding Authors: aamorim@sim.fc.ul.pt, jalmeida@mail.cern.ch

The ATLAS conditions databases will be used to manage information of quite diverse nature and level of complexity. The infrastructure in being built using the LCG COOL infrastructure and provides a powerful information sharing gateway upon many different systems. The nature of the stored information ranges from temporal series of simple values to very complex objects describing the configuration of systems like the TDAQ infrastructure including also associations to large objects managed outside of the database infrastructure. While a unified graphical user interface is crucial for browsing the different data, it must understand and display many different types of information in a flexible way suggesting the use of run time specific plugins. This extension mechanism was heavily used in the KTIDBEXPLORER application that defines and implements not only abstract interfaces to connections to databases and files but also supports extended ROOT and online configuration (OKS) objects in the database. The application is aware of the relations between database objects so it's relations (links) can be explored. The core application, built using QT, displays the hierarchical folder view and powerful table widgets with panels for selecting the database queries. An important example of this architecture is the oNline Objects extended Database browsEr (NODE), that is designed to access and display all data, including histograms and data tables, available in the ATLAS Monitoring Archive. To cope with the special nature of the monitoring objects, a plugin from the MDA framework to the Time managed science Instrument Databases (TIDB2) is used. The database browser is extended, in particular to include operations on histograms like display, overlap, comparisons as well as commenting and local storage.

Software components, tools and databases / 199

Improvements in ROOT I/O's Functionality and Performance

Authors: Anna Kreshuk¹ ; Leandro Franco¹ ; Paul Russo² ; Philippe Canal²

¹ *CERN*

² *FERMILAB*

Corresponding Author: philippe.canal@cern.ch

For the last several months the main focus of development in the ROOT I/O package has been code consolidation and performance improvements.

Access to remote files is affected both by bandwidth and latency. We introduced a pre-fetch mechanism to minimize the number of transactions between client and server and hence reducing the effect of latency. We will review the implementation and how well it works in different conditions (gain of an order of magnitude for remote file access).

We will also review new utilities, including a faster implementation of TTree cloning (gain of an order of magnitude), a generic mechanism for object references, and a new entry list mechanism tuned both for small and large number of selections.

In addition to reducing the coupling with the core module and becoming its own library (libRIO) (as part of the general restructuring of the ROOT libraries), the I/O package has been enhanced in the area of XML and SQL support, thread safety, schema evolution, TTreeFormula, and many other areas.

We will also discuss various ways, ROOT will be able to benefit from multi-core architecture to improve I/O performances.

Event processing / 423

Individual Particle Reconstruction

Authors: Norman Graf¹ ; Steve Magill²

¹ *SLAC*

² *Argonne National Lab*

Corresponding Author: norman.graf@slac.stanford.edu

The International Linear Collider (ILC) promises to provide electron-positron collisions at unprecedented energy and luminosities. The relative democracy with which final states are produced at these high energies places a premium on the efficiency and resolution with which events can be reconstructed.

In particular, the physics program places very demanding requirements on the dijet invariant mass resolution. Collider detectors have successfully improved their jet energy resolutions by augmenting the calorimetric measurements with the momenta of charged particles measured in their trackers a posteriori.

We present studies which apply this paradigm to the design of ILC detectors, proposing to achieve the requisite performance by measuring the charged particle contribution to the jet energy using the track momenta and only using calorimetric information for neutral particles. Designing detectors to implement this algorithm requires a combined approach to the detector as a whole, but since the crux of this technique is the ability

to uniquely identify and assign energy depositions to individual particle showers, the calorimetry is emphasized. In this talk, we present results based on a flexible simulation and analysis framework (slic & org.lcsim). We describe a templated approach to the reconstruction which allows various clustering and track-cluster association algorithms to be quickly and efficiently implemented and compared. We demonstrate the performance of the reconstruction on a number of detector models with different choices of calorimeter absorber, active media and readout segmentation as well as overall detector parameters such as the strength of the magnetic field, magnet bore, aspect ratio, hermeticity, etc.

Summary:

We present studies of ILC detector designs based on individual particle reconstruction.

Poster 1 - Board: 97 / 293

Initial results from the MAGIC-II Datacenter at PIC

Author: Manuel Delfino Reznicek¹

Co-authors: Esther Acción¹ ; Juan Cortina²

¹ *Port d'Informació Científica (PIC)*

² *Institut de Física d'Altes Energies (IFAE)*

Corresponding Author: delfino@pic.es

A new data center has been deployed for the MAGIC Gamma Ray Telescope, located in the Roque de los Muchachos observatory in the Canary Islands, Spain, at the Port d'Informació Científica in Barcelona. The MAGIC Datacenter at PIC receives all the raw data produced by MAGIC, either via the network or tape cartridges, and provides archiving, rapid processing for quality control and calibration, reconstruction and generation of analysis datasets, and distribution of all levels of datasets to all institutes and users of the collaboration. In 2007, services will be expanded to include the submission of batch analysis by all members of the MAGIC collaboration. Additional services provided to MAGIC by PIC are a CVS server and Web-enabled databases for data catalogue, experimental conditions and calibration constants. Implementation of the Datacenter has required PIC to solve a number of challenges. One of them is the efficient mass storage handling for a large number of small files, which is accomplished by implementing Virtual Volumes based on ISO images stored on tape and mounted dynamically using automounter techniques. The other is coping with the limited WAN bandwidth available to the Roque de los Muchachos observatory and having to combine collection of raw data via network and tape cartridges. This paper summarizes the initial implementation of the MAGIC-II Datacenter at PIC, describes the main technologies deployed and gives the growth prospects for the next years.

Poster 1 - Board: 87 / 238

Integrated RT-AT-Nagios system at BNL USATLAS Tier1 computing center

Author: Tomasz Wlodek¹

Co-authors: Carlos Gamboa¹ ; Dantong Yu¹ ; Jason Smith¹ ; Robert Petkus¹ ; Shigeki Misawa¹ ; Tom Throwe¹ ; Yingzi Wu¹ ; Zhenping Liu¹

¹ *Brookhaven National Laboratory*

Corresponding Author: tomw@bnl.gov

Managing large number of heterogeneous grid servers with different service requirements posts great challenges. We describe a cost-effective integrated operation framework which manages hardware inventory, monitors services, raises alarms with different severity levels and tracks the facility response to them.

The system is based on open source components: RT (Request Tracking) tracks user requests, AT (Asset Tracking) manages site inventory, while Nagios performs facility monitoring. We will discuss the integration of those components.

The AT serves as central repository to store information about machines, services, groups of machines and services, their interdependencies and configuration.

Problem reports sent to RT by users are reflected on asset history stored in AT database.

Nagios system uses AT to obtain information about the components to be monitored. Detected problems are classified according to their severity, reported to experts and fed into RT system, where the progress towards their resolution is tracked.

The paper will describe the AT data model, integration between AT and Nagios and interfacing the RT to other problem tracking systems.

The described system provides a scalable solution to commission grid servers, automate the error-prone manual system configuration, and leverage the existing ticket system for problem tracking. It allows BNL to operate Tier1 facility 7X24, and meets service level agreements for each WLCG grid middleware component with different class of service requirements.

Distributed data analysis and information management / 424

Integrating Xgrid technology into HENP distributed computing model

Authors: Adam Kocoloski¹ ; Jerome LAURET² ; Levente HAJDU² ; Mike MILLER¹

¹ MIT

² BROOKHAVEN NATIONAL LABORATORY

Corresponding Authors: kocolosk@mit.edu, jlauret@bnl.gov

Modern Macintosh computers feature Xgrid, a distributed computing architecture built directly into Apple's OS X operating system. While the approach is radically different from those generally expected by the Unix based Grid infrastructures (Open Science Grid, TeraGrid, EGEE), opportunistic computing on Xgrid is nonetheless a tempting and novel way to assemble a computing cluster with a minimum of additional configuration. In fact, it requires only the default operating system and authentication to a central controller from each node. OS X also implements arbitrarily extensible metadata, allowing an instantly updated file catalog to be stored as part of the filesystem itself.

The low barrier to entry allows an Xgrid cluster to grow quickly and organically. This paper and presentation will detail the steps that can be taken to make such a cluster a viable resource for HENP research computing. We will further show how to provide to users a unified job submission framework by integrating Xgrid through the STAR Unified Meta-Scheduler (SUMS), making tasks and jobs submission effortlessly at reach for those users already using the tool for traditional Grid or local cluster job submission. We will discuss additional steps that can be taken to make an Xgrid cluster a full partner in grid computing initiatives, focusing on Open Science Grid integration. MIT's Xgrid system currently supports the work of multiple research groups in the Laboratory for Nuclear Science, and has become an important tool for

generating simulations and conducting data analyses at the Massachusetts Institute of Technology.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

STAR

Software components, tools and databases / 85

Integration of the ATLAS Tag Database with Data Management and Analysis Components

Author: Caitriana Nicholson¹

Co-authors: David Malon²; Helen McGlone¹; Jack Cranshaw²; Michael Kenyon¹

¹ *University of Glasgow*

² *Argonne National Laboratory*

Corresponding Author: caitriana.nicholson@cern.ch

The ATLAS Tag Database is an event-level metadata system, designed to allow efficient identification and selection of interesting events for user analysis. By making first-level cuts using queries on a relational database, the size of an analysis input sample could be greatly reduced and thus the time taken for the analysis reduced. Deployment of such a Tag database is underway, but to be most useful it needs to be integrated with the distributed data management (DDM) and distributed analysis (DA) components. This means addressing the issue that the DDM system at ATLAS groups files into datasets for scalability and usability, whereas the Tag database points to events in files. It also means setting up a system which could prepare a list of input events and use both the DDM and DA systems to run a set of jobs. The ATLAS Tag Navigator Tool (TNT) has been developed to address these issues in an integrated way and provide a tool that the average physicist can use. Here, the current status of this work is presented and areas of future work are highlighted.

Poster 2 - Board: 68 / 61

Integration of the ATLAS VOMS system with the ATLAS Metadata Interface

Author: Thomas Doherty¹

Co-authors: Fabian Lambert²; Fulachier Jerome²; Solveig Albrand²

¹ *University of Glasgow*

² *CNRS*

Corresponding Author: t.doherty@physics.gla.ac.uk

AMI is an application which stores and allows access to dataset metadata for the ATLAS experiment. It provides a set of generic tools for managing database applications. It has a three-tier architecture with a core that supports a connection to any RDBMS using JDBC and

SQL. The middle layer assumes that the databases have an AMI compliant self-describing structure. It provides a generic web interface and a generic command line interface. A Virtual Organisation Membership Service (VOMS) is an authorisation system for Virtual Organisations (VO's). The ATLAS VO has a VOMS system which contains its own authorisation information.

This presentation provides an account of the development of a Java based solution to integrate the ATLAS VOMS system to the ATLAS Metadata Interface (AMI). The prerequisites authentication and authorisation demand on grid architecture are explained. The current workings of a VOMS system and the resulting requirements this has on a client of this system are discussed. We explore possible solutions to these requirements before detailing the mechanism behind the chosen solution and how it was integrated with the AMI framework.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS and GridPP **Summary:**

This presentation provides an account of the development of a Java based solution to integrate the ATLAS VOMS system to the ATLAS Metadata Interface (AMI). The prerequisites authentication and authorisation demand on grid architecture are explained. The current workings of a VOMS system and the resulting requirements this has on a client of this system are discussed. We explore possible solutions to these requirements before detailing the mechanism behind the chosen solution and how it was integrated with the AMI framework.

It is an ATLAS Offline Computing abstract and a GridPP abstract

Online computing / 28

Integration of the Trigger and Data Acquisition Systems in ATLAS

Author: Imma Riu¹

¹ *Universitat Autònoma de Barcelona. Institut de Física d'Altes Energies*

Corresponding Authors: benedetto.gorini@cern.ch, imma.riu@cern.ch

During 2006 and early 2007, integration and commissioning of trigger and data acquisition (TDAQ) equipment in the ATLAS experimental area have progressed. Much of the work has focussed on a final prototype setup consisting of around 80 computers representing a subset of the full TDAQ system. There have been a series of technical runs using this setup. Various tests have been run including ones where around 4k level 1 preselected simulated proton-proton events have been processed in a loop mode through the trigger and dataflow chains. The system included the readout buffers containing the events, event building, level 2 and event filter trigger algorithms. Quantities critical for the final system, such as trigger rates and event processing times, have been studied using different trigger algorithms as well as different dataflow components.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

On behalf of the ATLAS DAQ and trigger communities

Event processing / 422**Intelligent Design****Author:** Norman Graf¹**Co-authors:** Jeremy McCormick ¹ ; Ron Cassell ¹ ; Tony Johnson ¹¹ SLAC**Corresponding Author:** norman.graf@slac.stanford.edu

The International Linear Collider (ILC) promises to provide electron-positron collisions at unprecedented energy and luminosities. Designing the detectors to extract the physics from these events requires efficient tools to simulate the detector response and reconstruct the events.

The detector response package, slic, is based on the Geant4 toolkit and adds a thin layer of C++ code. This allows the end user to fully describe the detector geometry and readout at runtime using a plain text file in an xml format which extends GDML. It also supports reading in simulated events in stdhep format, and writing out events in the ILC-standard LCIO format.

We then describe org.lcsim, a Java toolkit for full event reconstruction and analysis. The components are fully modular and are available for tasks from digitization of tracking detector signals through to cluster finding, pattern recognition, track fitting, jetfinding, and analysis. The code can be run standalone, for batch or Grid computing, or from within JAS3, which then provides access to the WIRED event display and the AIDA-compliant analysis capabilities.

We present the architecture as well as the implementation for several candidate detector designs, demonstrating both the flexibility and the power of the system.

Summary:

We present an integrated detector response simulation and event reconstruction environment being used to study detector concepts for the International Linear Collider.

Computer facilities, production grids and networking / 356**Interfacing with Sun Utility Computing, experience with on demand physics simulations on SunGrid****Authors:** Jerome LAURET¹ ; Maxim Potekhin¹**Co-authors:** Ari SHAMASH ² ; Gabriele CARCASSI ²¹ BROOKHAVEN NATIONAL LABORATORY² Sun MicroSystem**Corresponding Authors:** maxim.potekhin@cern.ch, jlauret@bnl.gov

The simulation program for the STAR experiment at Relativistic Heavy Ion Collider at Brookhaven National Laboratory is growing in scope and responsiveness to the needs of the research conducted by the Physics Working Groups. In addition, there is a significant ongoing R&D activity aimed at future upgrades of the STAR detector, which also requires extensive simulations

support. The principal computing facility used by STAR to conduct the simulations studies is a farm containing 400 nodes, with a total of 1000 CPUs.

OpenScience Grid (OSG) resources have been successfully used in the past and routinely used in STAR. However, the explosive growth of the computing power and the rapid evolution of the distributed computing landscape demand for the STAR Collaboration to dictate that all available options are considered, from Open source to commercial grids using a thin modular layer interfacing with the many “grids”. Sun Grid from Sun Microsystems aims to deliver enterprise computing power and resources over the Internet, enabling developers, researchers, scientists and businesses to optimize performance, speed time to results, and accelerate innovation without investment in IT infrastructure.

We have successfully run a part of our production jobs on the SunGrid facility and will present our experience with its interface, performance and related issues and discuss ongoing efforts and development to interface it with the STAR Unified Meta-schedule (or SUMS).

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

STAR

Event processing / 366

JAIDA, JAS3, WIRED4 and the AIDA tag library - experience and new developments

Authors: Mark Donszelmann¹ ; Max Turri¹ ; Tony Johnson¹ ; Victor Serbo¹

¹ SLAC

Corresponding Authors: serbo@slac.stanford.edu, tonyj@slac.stanford.edu

JAIDA is a Java implementation of the Abstract Interfaces for Data Analysis (AIDA); it is part of the FreeHEP library. JAIDA allows Java programmers to quickly and easily create histograms, scatter plots and tuples, perform fits, view plots and store and retrieve analysis objects from files. JAIDA can be used either in a non-graphical environment (for batch processing) or with a GUI. Files written with JAIDA adhere to the AIDA IO standards and can be read by any AIDA compliant analysis system. JAIDA can also access data from ROOT, HBOOK/PAW or SQL databases, and can be used from C++ via the AIDA “C++ to Java” adapter (AIDAJNI). JAIDA now includes JMinuit, a complete port of Minuit to Java. JAIDA is used internally by JAS3 which provides a full featured GUI in addition to the above functionality.

The AIDA tag library (AIDATLD) is an open source suite of custom tags that provide access to JAIDA from J2EE applications and JSP pages. It provides the ability to dynamically creating high quality physics and astronomy plots, as well as providing access to histograms and Ntuples stored in any AIDA store (which includes ROOT files via rootd or xrootd) from web applications.

This software is currently used by several experiments and collaborations, including BaBar, GLAST, and Geant4. Experience of using AIDATLD, JAIDA, and JAS3 in experiments, as well as description of new developments will be presented in the talk. In particular we will describe a wide ranging suite of web applications developed using these tools for the GLAST experiment.

JetWeb: A library and server for Monte Carlo validation and data comparisons

Author: Jonathan Butterworth¹

¹ *University College London*

Accurate modelling of high energy hadron interactions is essential for the precision analysis of data from the LHC. It is therefore imperative that the predictions of Monte Carlos used to model this physics are tested against existing and future measurements. These measurements cover a wide variety of reactions, experimental observables and kinematic regions. To make this process more reliable and easily verifiable, the CEDAR collaboration has developed a set of tools for tuning and validating models of such interactions, and for archiving the comparisons so that data from previous experiments can be easily compared to new models. JetWeb is a Java-based server which accesses SQL databases for data and model predictions, and acts as a gateway to LCG resources for CPU-intensive simulation jobs. The simulation jobs are run using the HZTool and Rivet applications, also developed within CEDAR, and the data for comparison is taken from HepData. The server already contains an array of useful data. The status, current application and view plans of the system are described.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CEDAR

Grid middleware and tools / 95

Job Submission and Management Through Web Services: the Experience with the CREAM Service

Authors: Alessio Gianelle¹ ; Alvise Dorigo¹ ; Andrea Sciaba² ; Cristina Aiftimiei¹ ; Luigi Zangrando¹ ; Marco Verlati¹ ; Massimo Sgaravatto¹ ; Moreno Marzolla¹ ; Paolo Andreetto¹ ; Sara Bertocco¹ ; Simone Dalla Fina¹

¹ *INFN Padova*

² *CERN*

Corresponding Authors: luigi.zangrando@pd.infn.it, moreno.marzolla@pd.infn.it

Modern GRID middlewares are built around components providing basic functionality, such as data storage, authentication and security, job management, resource monitoring and reservation. In this paper we describe the Computing Resource Execution and Management (CREAM) service. CREAM provides a Web service-based job execution and management capability for Grid systems; in particular, it is being used within the gLite middleware. CREAM exposes a Web service interface allowing conformant clients to submit and manage computational jobs to a Local Resource Management System. We developed a special component, called ICE (Interface to CREAM Environment) to integrate CREAM in gLite. ICE transfer job submissions and cancellations from the Workload Management System, allowing users to manage CREAM jobs from the gLite User Interface. This paper describes some recent studies aimed at measuring the performance and the reliability of CREAM and ICE, also in comparison with other job submission systems. We discuss recent work towards enhancing CREAM with a BES and JSDL compliant interface.

Software components, tools and databases / 180

LCGO - geometry description for ILC detectors

Author: Frank Gaede¹

Co-authors: Norman Graf² ; Tony Johnson²

¹ DESY IT

² SLAC

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

The ILC is in a very active R&D phase where currently four international working groups are developing different detector designs. Increasing the interoperability of the software frameworks that are used in these studies is mandatory for comparing and optimizing the detector concepts. One key ingredient for interoperability is the geometry description. We present a new package (LCGO) which is suited to be incorporated in the existing frameworks. Compatibility with Java and C++ is achieved through the use of gcj - the GNU Java Compiler that allows the integration of object code written in Java in C++ programs. LCGO uses a driver based approach where the geometry description is based on a combination of code and free parameters. LCGO provides a multi level API that allows the users to query the detector geometry and material properties at the level that is needed by the application. This ensures that only once source of the geometry description is needed throughout the full software chain from simulation to reconstruction, analysis and event displays.

Plenary / 480

LHC Expt Computing

Corresponding Author: ian.fisk@cern.ch

Distributed data analysis and information management / 265

LHCb Distributed Conditions Database

Author: Marco Clemencic¹

¹ European Organization for Nuclear Research (CERN)

Corresponding Author: marco.clemencic@cern.ch

The LHCb Conditions Database project provides the necessary tools to handle non-event time-varying data. The main users of conditions are reconstruction and analysis processes, which are running on the Grid. To allow efficient access to the data, we need to use a synchronized replica of the content of the database located at the same site as the event data file, i.e. the LHCb Tier1. The replica to be accessed is selected from information stored on LFC (LCG File Catalog) and managed with the interface provided by the LCG developed library CORAL. The way we limit the submission of jobs to those sites where the required conditions are available will also be presented.

LHCb applications are using the Conditions Database framework on a production basis since March 2007. We have been able to collect statistics on the performances and effectiveness of both the LCG library COOL (the library providing conditions handling functionalities) and the distribution framework itself. Stress tests on the CNAF

hosted replica of the Conditions Database have been performed and the result will be summarized here.

Online computing / 89

LHCb Online Interface to a Conditions Database

Author: Maria Del Carmen Barandela Pazos¹

¹ *University of Vigo*

Corresponding Author: maria.del.carmen.barandela.pazos@cern.ch

In a High Energy Physics experiment it is fundamental to handle information related to the status of the detector and its environment at the time of the acquired event. This type of time-varying non-event data are often grouped under the term “conditions”. The LHCb’s Experiment Control System groups all the infrastructure for the configuration, control and monitoring of all the components of the online system. It is in this environment where an interface to define and store conditions is needed. These conditions are stored in the Conditions Database. This database will contain a subset of the monitoring data, read from hardware, that are needed for physics processing and also some configuration data, like for example, the trigger settings. The Interface to the Conditions Database has been developed as a component of the LHCb control framework and it is based on a SCADA (Supervisory Control and Data Acquisition) system called PVSSII. It consists in a PVSS panel which allows users to define which data should be stored as a condition, how these data should be packaged and when these data should be updated: when they change, when they change by more than a certain value, regular intervals, etc. Once these data are updated they are sent to a server which is the responsible to write and read the conditions from the database. This system provides a very simple and flexible way to define conditions and it can also be used by any sub detector because the way the information is transferred and stored is completely transparent for the users.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

LHCb

Online computing / 137

LHCb Online event processing and filtering

Author: Niko Neufeld¹

¹ *CERN*

Corresponding Author: niko.neufeld@cern.ch

The first level trigger of LHCb accepts 1 MHz of events per second. After preprocessing in custom FPGA-based boards these events are distributed to a large farm of PC-servers using a high-speed Gigabit Ethernet network. Synchronisation and event management is achieved by the Timing and Trigger system of LHCb. Due to the complex nature of the selection of B-events, which are the main interest of LHCb, a full event-readout is required. Event processing on the servers is parallelised on an event basis. The reduction factor is typically 1 / 500. The remaining events from all farm-nodes are forwarded to a formatting layer, where the raw data files are formed and temporarily stored. A

small part of the events is also forwarded to a dedicated farm for calibration and monitoring. The files are subsequently shipped to the CERN Tier0 facility for permanent storage and from there to the various Tier1 sites for reconstruction. In parallel files are used by various monitoring and calibration processes running within the LHCb Online system. The entire dataflow is controlled and configured by means of a SCADA system and several databases. After an overview of the LHCb data acquisition and its design principles this paper will emphasize the LHCb event filter system, which is now implemented using the final hardware and will be ready for data-taking for the LHC startup. Control, configuration and security aspects will also be discussed.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

LHCb Online

Software components, tools and databases / 213

LHCb experience with LFC database replication

Authors: Angelo Carbone¹ ; Barbara Martelli² ; Federico Bonifazzi² ; Gianluca Peco³

¹ *Italian National Institute of Nuclear Physics (INFN-Bologna), Italian INFN National Center for Telematics and Informatics (CNAF)*

² *Italian INFN National Center for Telematics and Informatics (CNAF)*

³ *Italian National Institute of Nuclear Physics (INFN-Bologna)*

Corresponding Authors: barbara.martelli@cnaif.infn.it, angelo.carbone@bo.infn.it

Database replication is a key topic in the LHC Computing GRID environment to allow processing of data in a distributed environment. In particular LHCb computing model relies on the LHC File Catalog (LFC). LFC is the database catalog which stores informations about files spread across the GRID, their logical names and physical locations of all their replicas. The LHCb computing model requires the LFC to be replicated at Tier1s via Oracle Streams technology.

This paper will describe the LFC replicas deployment at Italian INFN National Center for Telematics and Informatics (CNAF) and at other LHCb Tier1 sites and will present subsequent stress test results. The tests were designed to evaluate any delay on the propagation of the streams and the scalability of the system. The tests show the robustness of the replica implementation with performance going beyond the experiment's requirements.

Computer facilities, production grids and networking / 249

Lambda Station: Alternate Network Path Forwarding for Production SciDAC Applications

Authors: Andrey Bobyshev¹ ; Maxim Grigoriev¹

Co-authors: Conrad Steenberg² ; Dan Nae² ; Don Petravick¹ ; Frank Van Lingen² ; Harvey Newman² ; Julian Bunn² ; Matt Crawford¹ ; Maxim Grigoriev¹ ; Michael Thomas² ; Philip DeMar³ ; Vyto Grigaliunas¹ ; Xun Su² ; Yang Xia²

¹ *FERMILAB*

² *Caltech*

³ *FE*

Corresponding Authors: maxim@fnal.gov, bobyshev@fnal.gov

The LHC experiments will start very soon, creating immense data volumes capable of demanding allocation of an entire network circuit for task-driven applications. Circuit-based alternate network paths are one solution to meeting the LHC high bandwidth network requirements. The Lambda Station project is aimed at addressing growing requirements for dynamic allocation of alternate network paths. Lambda Station orchestrates the re-routing of designated traffic through site LAN infrastructure onto so-called “high-impact” wide-area networks. The prototype Lambda Station developed with Service Oriented Architecture (SOA) will be presented. Lambda Station has been successfully integrated into the production version of the Storage Resource Manager (dCache/SRM), and deployed at US CMS Tier1 center at Fermilab, as well as at US-CMS Tier-2 site at Caltech.

This paper will discuss experiences using the prototype system with production SciDAC applications for data movement between Fermilab and Caltech. The architecture and design principles of the production version Lambda Station software, currently being reimplemented as Java based Web services, will also be presented in this paper.

Poster 1 - Board: 14 / 333

Large Scale Access Tests and Online Interfaces to ATLAS Conditions Databases

Authors: Antonio Amorim¹ ; Doris BURCKHART² ; Igor SOLOVIEV³ ; Johann Georg VON DER SCHMITT⁴ ; João José de Almeida Simões⁵ ; Lourenço Vaz Gil Lopes⁵ ; Mihai CAPRINI⁶ ; Paulo José Sousa Pereira⁵ ; Serguei KOLOS⁷

¹ *Universidade de Lisboa CFNUL - FCUL*

² *CERN*

³ *PNPI (Petersburg Nuclear Physics Institute)*

⁴ *Max-Planck-Institut fuer Physik*

⁵ *UNIVERSIDADE DE LISBOA SIM - FCUL*

⁶ *“Horia Hulubei” National Institute for Physics and Nuclear Engineering (IFIN/HH, Bucharest)*

⁷ *University of California, Irvine (UCI, Irvine)*

Corresponding Authors: aamorim@sim.fc.ul.pt, antonio.amorim@fisica.fc.ul.pt

The ATLAS Trigger and Data Acquisition systems (TDAQ) to the Conditions databases has strong requirements on reliability and performance. Several applications were developed to support the integration of Condition database access with the online services in TDAQ like the interface to the Information Services and to the TDAQ configuration..

The DBStressor was developed to test and stress the access to the Conditions database using the LCG/COOL interface while operating in an integrated way as a TDAQ application. The performance of simultaneous Conditions database read accesses was studied in the context of the ATLAS High Level Trigger large computing farms. A large set of tests were performed involving up to 1000 computing nodes that accessed simultaneously the LCG central database server infrastructure at CERN. Most of the general figures of the results can be explained assuming a simple (model involving) described by a number of threads in the database servers that were providing data at near constant rate.

The information storage requirements were the motivation for the ONline ASynchronous Interface form the Information Service (IS) with LCG/COOL databases. It avoids the backpressure from Online Database servers by managing a local cache. In parallel the OKS2COOL application was developed to store Configuration Databases into an Offline Database with history record.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS - TDAQ

Computer facilities, production grids and networking / 263

Large-scale ATLAS Production on EGEE

Authors: Rodney Walker¹; Simone Campana²; Xavier Espinal³¹ TRIUMF² CERN³ PIC/IFAE**Corresponding Author:** espinal@pic.es

In preparation for first data at the LHC, a series of Data Challenges, of increasing scale and complexity, have been performed. Large quantities of simulated data have been produced on three different Grids, integrated into the ATLAS production system. During 2006, the emphasis moved towards providing stable continuous production, as is required in the immediate run-up to first data, and thereafter. Here, we discuss the experience of the production done on EGEE resources, using submission based on the gLite WMS, CondorG and a system using Condor glide-in's.

The overall walltime efficiency of around 90% is largely independent of the submission method, and the dominant source of wasted cpu comes from data handling issues. The efficiency of grid job submission is significantly worse than this, and the glide-in method benefits greatly from factorising this out.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Distributed data analysis and information management / 307

Latest Developments in the PROOF System

Author: Fons Rademakers¹**Co-authors:** Bertrand Bellenot¹; Gerardo Ganis¹; Jan Iwaszkiewicz¹; Maarten Ballintijn²; Rene Brun¹¹ CERN² MIT**Corresponding Author:** fons.rademakers@cern.ch

The goal of PROOF (Parallel ROOT Facility) is to enable interactive analysis of large data sets in parallel on a distributed cluster or multi-core machine. PROOF represents a high-performance alternative to a traditional batch-oriented computing system.

The ALICE collaboration is planning to use PROOF at the CERN Analysis Facility (CAF) and has been stress testing the system since mid 2006 on a 40 machine pilot cluster. The ALICE CAF is expected to grow to around 500 machines.

The testing by ALICE has allowed us to identify missing functionality and to improve the system in many ways. Areas of significant development include: a dataset manager to optimally distribute data on the cluster; facilities to upload and manage the experiment software; a new "packetizer" which significantly reduces the end-of-query tails; a worker-level

priority-based scheduling system to control the fraction of resources assigned to a group of users; improved error handling and user feedback mechanism; and much more.

The CMS collaboration is also actively investigating PROOF as Tier-2 analysis facility.

Current activities focus on the development of a central scheduling system that uses the OLBD/XROOTD control network as information routing system. This scheduler aims to improve resource sharing in a multi-user environment, taking per-query decisions based on the status of the farm, the query requirements and the history and priorities of the user.

In this paper we will describe in detail the recent developments, the status of the current activities, and outline the future plans to bring PROOF in production for LHC analysis.

Computer facilities, production grids and networking / 31

Lessons Learnt From WLCG Service Deployment

Author: Jamie Shiers¹

¹ CERN

Corresponding Author: jamie.shiers@cern.ch

This talk summarises the main lessons learnt from deploying WLCG production services, with a focus on Reliability, Scalability, Accountability, which lead to both manageability and usability.

Each topic is analysed in turn. Techniques for zero-user-visible downtime for the main service interventions are described, together with pathological cases that need special treatment. The requirements in terms of scalability are analysed, calling for as much robustness and automation in the service as possible. The different aspects of accountability - which covers measuring / tracking / logging / monitoring what is going on - and has gone on - is examined, with the goal of attaining a manageable service.

Finally, a simple analogy is drawn with the Web in terms of usability - what do we need to achieve to cross the chasm from small-scale adoption to ubiquity?

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

WLCG Summary:

A summary of the key lessons learnt from deploying WLCG production services, with a focus on Reliability, Scalability, Accountability, which lead to both manageability and usability.

Event processing / 335

Level-1 RPC Trigger in the CMS experiment - software for emulation and commissioning

Authors: Artur Kalinowski¹ ; Karol Bunkowski¹ ; Marcin Konecki¹ ; Michal Bluj² ; Michal Pietrusinski¹ ; Tomasz Maciej Frueboes¹

¹ *Institute of Experimental Physics - University of Warsaw*

² *Soltan Institute for Nuclear Studies - Warsaw*

Corresponding Author: fruboes@cern.ch

The CMS detector will start its operation in the end of 2007. Until that time great care must be taken in order to assure that hardware operation is fully understood. We present an example of how emulation software helps achieving this goal in the CMS Level-1 RPC Trigger system.

The design of the RPC trigger allows to insert sets of so-called test pulses at any stage of the hardware pipeline. Reading out data from different stages is also possible. Such design allows for easy debugging of the trigger hardware by comparing hardware and software emulation values, since the software and hardware algorithms are identical.

In the talk we would like to present the architecture of our test applications, specific test cases and the RPC trigger emulation software itself.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS

Poster 1 - Board: 9 / 262

Lorentz Angle Calibration for the CMS Pixel Detector

Authors: Lotte Wilke¹ ; Thomas Speer¹ ; Vincenzo Chiochia¹

¹ *Universitat Zurich*

Corresponding Author: vincenzo.chiochia@cern.ch

The CMS Pixel Detector is hosted inside the large solenoid generating a magnetic field of 4 T.

The electron-hole pairs produced by particles traversing the pixel sensors will thus experience the Lorentz force due to the combined presence of magnetic and electric field. This results in a systematic shift of the charge distribution. In order to achieve a high position resolution a correction for this shift, which can be up to $120\mu\text{m}$, has to be applied. At start-up the Lorentz shift for a given bias voltage is well known from beam test studies. Due to irradiation the electric field in the sensors will change and thereby the Lorentz drift as well. Furthermore, since the irradiation will not be uniform across the detector, each sensor will be differently affected. Therefore, the effective Lorentz displacement will be regularly measured using data.

We present a strategy to extract this drift by comparing the cluster shapes of pixel hits in fully reconstructed tracks. The procedure measures the Lorentz displacement as function of the sensor depth and is developed using the CMS simulation and reconstruction software.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS

Poster 1 - Board: 11 / 271

Low-level reconstruction software and regional unpacking for the CMS strip tracker

Authors: Matthew Wingham¹ ; Robert Bainbridge¹

Co-authors: Boris Mangano² ; Christopher D. JONES³ ; Domenico Giordano⁴ ; Joanne Cole⁵ ; Jonathan Fulcher¹

¹ *Imperial College London*

² *University of California, San Diego*

³ *Cornell University*

⁴ *INFN & Università di Bari*

⁵ *Rutherford Appleton Laboratory*

Corresponding Authors: robert.bainbridge@cern.ch, matthew.wingham@cern.ch

The CMS silicon strip tracker is unprecedented in terms of its size and complexity, providing a sensitive area of $>200 \text{ m}^2$ and comprising 10M readout channels. Its data acquisition system is based around a custom analogue front-end ASIC, an analogue optical link system and an off-detector VME board that performs digitization, zero-suppression and data formatting. These data are forwarded to the CMS online computing farm, which performs reconstruction and provides the high-level trigger (HLT) using tools defined within the offline software framework. The strip tracker geometry and high-multiplicity events combine to create very large data volumes, which must be “unpacked” from a custom format into objects handled by the reconstruction chain. This must be done within stringent time quotas imposed by the available online computing resources. We review the issues and requirements for HLT, solutions for optimizing the low-level reconstruction chain, such as regional unpacking, and results based on simulation and the expected final detector configuration.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS

Software components, tools and databases / 435

Maintenance, validation and tuning of Monte Carlo event generators for the LHC experiments in the Generator Services project

Authors: Mikhail Kirsanov¹ ; Witold Pokorski²

Co-authors: Alexander Toropin¹ ; Oleg Zenin³

¹ *Institute for Nuclear Research (INR)*

² *CERN*

³ *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 paper 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.

Poster 1 - Board: 45 / 282

Management System of Event Processing and Data Files Based on XML Software Tools at Belle

Author: Ichiro Adachi¹

Co-authors: Masahiko Yokoyama¹; Nobuhiko Katayama¹; Takashi Okamoto²; Yukimasa Matsuda²; Yutaka Ueshima²

¹ KEK

² Quatre-i science cooperation

Corresponding Author: ichiro.adachi@kek.jp

The Belle experiment has been operational since 1999 and we have processed more than 700/fb of data so far. To cope with ever increasing data, complete automation of the event processing is one of the most critical issues. In addition, unified management in the processing job and the processed data files to be analyzed is very important especially to deal with ~400K data files amounting to Peta-scale data. To do this, we have implemented new application called "R&D Chain Management (RCM)". In the RCM program, we define a "work-flow" as a set of check-points related to the event processing procedures such as job submission, job log verification, output file backup and so on. The work-flow is described in the XML and it is so flexible that we can cover all of actions required in the processing stream. The RCM application automatically surveys each check-point in the work-flow and it reports failure if any. Furthermore, simultaneously the RCM tools can store any necessary information at each check-point into the own database written in the XML. The RCM system provides the web-based interface between the XML database and users for an excellent visualization. Thus, one can quickly pick up relevant information from this system. The processed data files are also taken care by this RCM, and automatic action to take a backup, for instance, can be issued from the RCM tools. We will present our experience using the RCM software tool as well as performance in the Belle data management.

Grid middleware and tools / 64**Managing ATLAS data on a petabyte-scale with DQ2**

Authors: Benjamin Gaidioz¹ ; Birger Koblitz¹ ; Mario Lassnig² ; Massimo Lamanna¹ ; Miguel Branco¹ ; Pedro Salgado³ ; Ricardo Rocha¹ ; Vincent Garonne¹

¹ *CERN*

² *CERN & University of Innsbruck, Austria*

³ *University of Texas at Arlington, USA*

Corresponding Author: mario.lassnig@cern.ch

The ATLAS detector at CERN's Large Hadron Collider presents data handling requirements on an unprecedented scale. From 2008 on the ATLAS distributed data management system (DQ2) must manage tens of petabytes of event data per year, distributed globally via the LCG, OSG and NDGF computing grids, now known as the WLCG.

Since its inception in 2005 DQ2 has continuously managed all datasets for the ATLAS collaboration, which now comprises over 3000 scientists participating from more than 150 universities and laboratories in more than 30 countries.

Fulfilling its primary requirement of providing a highly distributed, fault-tolerant as well as scalable architecture

DQ2 has now been successfully upgraded from managing data on a terabyte-scale to data on a petabyte-scale.

We present improvements and enhancements to DQ2 based on the increasing demands for ATLAS data management. We describe performance issues, architectural changes and implementation decisions, the current state of deployment in test and production as well as anticipated future improvements. Test results presented here show that DQ2 is capable of handling data up to and beyond the requirements of full-scale data-taking.

Collaborative tools / 272**Managing an Institutional Repository with CDS Invenio**

Author: Nicholas Robinson¹

Co-authors: Jean-Yves Le Meur¹ ; Tibor Simko¹

¹ *CERN*

Corresponding Author: nicholas.robinson@cern.ch

CERN has long been committed to the free dissemination of scientific research results and theories. Towards this end, CERN's own institutional repository, the CERN Document Server (CDS) offers access to CERN works and to all related scholarly literature in the HEP domain. Hosting over 500 document collections containing more than 900,000 records, CDS provides access to anything from preprints and articles, to multimedia information such as photographs, movies, posters and brochures. The software that powers this service, CDS

Invenio, is distributed freely under the GNU GPL and is currently used in approximately 15 institutions worldwide.

In this paper, we discuss the use of CDS Invenio to manage a repository of scientific literature. We outline some of the issues faced during the lifecycle of a document from acquisition, processing and indexing to dissemination. In particular, we focus on the features and technology developed to meet the complexities of managing scientific information in the LHC era of large international collaborations each of which has its own distinct needs and requests.

Poster 2 - Board: 103 / 409

Mapping GRID to Site Credentials using GUMS

Authors: Jay Packard¹ ; John Hover¹

¹ *BNL*

Corresponding Author: jpackard@bnl.gov

Identity mapping is necessary when a site's resources do not use GRID credentials natively, but instead use a different mechanism to identify users, such as UNIX accounts or Kerberos principals. In these cases, the GRID credential for each incoming job must be associated with an appropriate site credential. Many sites consist of a heterogeneous environment with multiple gatekeepers, which can make control and security difficult. Therefore, a single site-wide usage policy is desirable. GUMS (Grid User Management System) is such a Grid Identity Mapping Service providing a single site-wide usage policy. It was designed to integrate with the site's local information services (such as HR databases or LDAP) and with GRID information services (such as VOMS). When a request comes in to a gatekeeper, the gatekeeper contacts the GUMS server to verify a user has appropriate membership (via VOMS, LDAP, or manual user groups), and if so, retrieves the mapping to the local identity (via shared group accounts, a pool of accounts, or a manually defined account). GUMS supports extended X509 certificates that include group membership and role, which influence the mappings. It provides a web interface for managing and testing the server as well as a command line tool. It is contained in the OSG RBAC infrastructure. GUMS has been improved to include a more comprehensive web interface, and we plan to implement recyclable accounts to ensure scaling. GUMS is well suited to sites needing secure and centrally managed GRID to site credential mapping capabilities.

Poster 2 - Board: 99 / 392

Medical Data Management Status and Plans

Authors: Akos Frohner¹ ; David Smith¹ ; Jean-Philippe Baud¹ ; Krzysztof Nienartowicz¹ ; Lana Abadie¹ ; Remi Mollon¹ ; Sophie Lemaitre¹

¹ *CERN*

Corresponding Author: akos.frohner@cern.ch

The goal of the Medical Data Management (MDM) task is to provide secure (encrypted and under access control) access to medical

images, which are stored at hospitals in DICOM servers or are replicated to standard grid Storage Elements (SE) elsewhere.

In gLite 3.0 there are three major components to satisfy the requirements: The dCache/DICOM SE is a special SE, which encrypts every requested image with a file specific key. It does not provide a storage area on its own, but interfaces a hospital's DICOM server to the grid. The gLite I/O server with a Fireman catalog service provides the access control by wrapping an SE, which holds medical images. And finally Hydra client library does the en/decryption of the files, using the file specific keys stored in the Hydra keystore.

In gLite R3.1 we are planning to simplify the software stack by relying on richer functionality of the underlying components: as storage elements (for example DPM) provide ACLs on individual files, we can remove the wrapping gLite I/O layer from a storage element and access it directly from the client side.

Refactoring of the dCache/DICOM SE is also necessary to unify the server side en/decryption and access control functionality in a single component.

Finally the Hydra keystore is being split into distributed services for reliability and to reduce the impact of a compromised key server.

Poster 1 - Board: 88 / 245

Metropolitan Area Network Support at Fermilab

Author: Philip DeMar¹

Co-authors: Andrey Bobyshev¹; Chuck Andrews²; Don Petravick¹; Donna Lamore¹; Matt Crawford¹; Orlando Colon¹; Steve Fry¹; Vyto Grigaliunas¹

¹ FERMILAB

² F

Corresponding Author: demar@fnal.gov

Advances in wide area network service offerings, coupled with comparable developments in local area network technology have enabled many HEP sites to keep their offsite network bandwidth ahead of demand. For most sites, the more difficult and costly aspect of increasing wide area network capacity is the local loop, which connects the facility LAN to the wide area service provider(s). Fermilab has chosen to provide its own local loop access through leasing of dark fiber to a nearby network exchange point (StarLight), and procuring dense wave division multiplexing (DWDM) equipment to provide data channels across the fiber. Installing and managing such optical network infrastructure has broadened the Laboratory's network support responsibilities to include operating network equipment that is located off-site, and is technically much different than classic LAN network equipment. Effectively, the Laboratory has assumed the role of a local service provider.

This presentation will cover Fermilab's experiences with deploying and supporting a Metropolitan Area Network (MAN) infrastructure, based on metro-range DWDM equipment, in order to meet its offsite networking needs. The benefits and drawbacks of providing and supporting such a service will be discussed. Issues of scalability, complexity, monitoring, and troubleshooting will also be discussed.

Poster 2 - Board: 104 / 436

Modular grid middleware configuration system

Author: Robert Harakaly¹

Co-authors: Gergely Debreczeni¹ ; Louis Poncet¹ ; Maria Alandes Pradillo¹ ; Oliver Keeble¹

¹ CERN

Corresponding Author: robert.harakaly@cern.ch

Configuration is an essential part of the deployment process of any software product. In the case of Grid middleware the variety and complexity of grid services coupled with multiple deployment scenarios make the provision of a coherent configuration both more important and more difficult. The configuration system must provide a simple interface which strikes a balance between the requirements of small university laboratories and those of large computing centers. It should also be simple to maintain alongside rapid developments in the software it is intended to configure.

This paper describes the evolution of a modular configuration system (YAIM) which is used to configure different Grid middleware products. It summarizes the lessons learned during several years of LCG/gLite middleware production use and presents a new approach chosen to face the forthcoming challenges. Along with the main design considerations and implementations decisions, an example of a fully integrated third party grid service (dCache) is also discussed.

Poster 2 - Board: 72 / 117

MonALISA: An Agent Based, Dynamic Service System to Monitor, Control and Optimize Distributed Systems

Authors: Adrian Muraru¹ ; Catalin Cirstoiu¹ ; Ciprian Dobre² ; Costin Grigoras¹ ; Harvey Newman³ ; Iosif Legrand³ ; Lucian Musat² ; Ramiro Voicu³

¹ CERN

² Polytechnic University of Bucharest

³ CALTECH

Corresponding Author: iosif.legrand@cern.ch

MonaLISA (Monitoring Agents in A Large Integrated Services Architecture) provides a distributed service for monitoring, control and global optimization of complex systems including the grids and networks used by the LHC experiments. MonALISA is based on an ensemble of autonomous multi-threaded, agent-based subsystems which able to collaborate and cooperate to perform a wide range of monitoring and decision tasks in large scale distributed applications, and to be discovered and used by other services or clients that require such information. It is a fully distributed system with no single point of failure.

The system is deployed now at more than 340 sites, serving several large Grid communities (ALICE, CMS, OSG, Ultralight, LCG-Russia...), and it is monitoring around one million parameters in near real-time (complete information for computing nodes, jobs, end to end connectivity, accounting, different grid services, network traffic and topology). MonALISA and its APIs are currently used by different tools in High Energy Physics (CMS job submission systems, Alien, Xrootd, Ganga, Diane...) to collect specific monitoring data which is used as an automatic feedback to different user

communities to understand how these complex systems are used and to detect problems. The system is able to react to specific conditions, triggered by alarm conditions, and thus to automatically select an appropriate action. MonALSIA is also used for optimizing global workflows in distributed systems.

Poster 1 - Board: 92 / 261

Monitoring a WLCG Tier-1 computing facility aiming at a reliable 24/7 service

Author: Andreas Heiss¹

Co-authors: Axel Jaeger²; Bernhard Verstege²; Bruno Hoefft²; Holger Marten²

¹ *Forschungszentrum Karlsruhe*

² *FORSCHUNGSZENTRUM KARLSRUHE*

Corresponding Author: andreas.heiss@iwr.fzk.de

Within the Worldwide LHC Computing Grid (WLCG), a Tier-1 centre like the German GridKa computing facility has to provide significant CPU and storage resources as well as several Grid services with a high level of quality. GridKa currently supports all four LHC Experiments, Alice, Atlas, CMS and LHCb as well as four non-LHC high energy physics experiments, and is about to significantly extend its services for other communities within the German Grid initiative D-Grid. In order to ensure the simultaneous usability of the resources by all VOs as well as the persistent import of data from CERN and the distribution of data to associated Tier-2 sites, a sophisticated monitoring model is essential.

We present the GridKa monitoring concept which is based on the Ganglia and Nagios systems combined with additional tools to monitor Grid services and infrastructure. Due to the complex dependencies between a high number of monitored hosts and services, a clear and simple to use 'dashboard' showing a summarized view of the monitoring information is an essential tool. This 'dashboard' allows for a quick overview of the status and performance of services during the day and will be the first source of information for a deeper problem analysis if an automatic alarm notification is sent during nights and weekends.

Poster 2 - Board: 90 / 341

Monitoring of Grid Behaviour by LHCb

Author: Gianluca Castellani¹

Co-author: roberto santinelli²

¹ *CERN*

² *cern*

Corresponding Author: gianluca.castellani@cern.ch

Facilities offered by WLCG are extensively used by LHCb in all aspects of their computing activity. A real time knowledge of the status of all Grid components involved is needed to optimize their exploitation. This is achieved by employing different monitoring services each one supplying a specific overview of the system. SAME tests are used in LHCb for monitoring the status of CE services hence its availability for LHCb; DIRAC proprietary monitoring tools allow to track experiment application problems; a Dirac transfer monitoring system reports

about failures in data transfers; FTS transfer tests are also closely monitored and the ARDA Dashboard provides complementary information on site reliability. This paper addresses the lack of correlation and analysis of all the available information and proposes a solution for such a central service.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

LHCb

Distributed data analysis and information management / 317

Monitoring the ATLAS Production System

Author: John Kennedy¹

¹ *LMU Munich*

Corresponding Author: john.kennedy@cern.ch

The ATLAS production system is responsible for the distribution of O(100,000) jobs per day to over 100 sites worldwide. The tracking and correlation of errors and resource usage within such a large distributed system is of extreme importance.

The monitoring system presented here is designed to abstract the monitoring information away from the central database of jobs. This approach ensures that the monitoring does not destructively interfere with the production itself and provides faster responses to monitoring queries.

The design and functionality of the system is discussed and the possible future development of monitoring tools for the ATLAS Production System are explored.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Distributed data analysis and information management / 255

Monitoring the Atlas Distributed Data Management System

Authors: Benjamin Gaidioz¹ ; Birger Koblitz¹ ; David Cameron¹ ; Dietrich Liko¹ ; Mario Lassnig¹ ; Massimo Lamanna¹ ; Miguel Branco¹ ; Pedro Salgado¹ ; Ricardo Rocha¹ ; Vincent Garonne¹

¹ *CERN*

Corresponding Author: ricardo.rocha@cern.ch

The ATLAS Distributed Data Management (DDM) system is evolving to provide a production-quality service for data distribution and data management support for production and users' analysis.

Monitoring the different components in the system has emerged as one of the key issues to achieve this goal. Its distributed nature over different grid infrastructures (EGEE, OSG and NDGF) with infrastructure-specific data management components makes the task particularly challenging. Providing simple views over the status of the

DDM components and data to users and site administrators is essential to effectively operate the system under realistic conditions.

In this paper we present the design of the DDM monitor system, the information flow, data aggregation. We discuss the available usage, the interactive functionality for end-users and the alarm system.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Computer facilities, production grids and networking / 379

Monitoring the EGEE/WLCG Grid Services

Authors: Alexandre Duarte¹ ; Antonio Retico² ; Domenico Vicinanza³ ; Piotr Nyczyk²

¹ CERN/Federal University of Campina Grande

² CERN

³ CERN/University of Salerno

Corresponding Authors: antonio.retico@cern.ch, alexandre.duarte@cern.ch

Grids have the potential to revolutionise computing by providing ubiquitous, on demand access to computational services and resources. They promise to allow for on demand access and composition of computational services provided by multiple independent sources. Grids can also provide unprecedented levels of parallelism for high-performance applications. On the other hand, grid characteristics, such as high heterogeneity, complexity and distribution create many new technical challenges.

Among these technical challenges, failure management is a key area that demands much progress. A recent survey revealed that fault diagnosis is still a major problem for grid users. When a failure appears at the user screen, it becomes very difficult for her to identify whether the problem is in the used application, somewhere in the grid middleware, or even lower in the fabric that comprises the grid.

In this paper we present a tool able to check if a given grid service works as expected for a given set of users (Virtual Organisation) on the different resources available on a grid. Our solution deals with grid services as single components that should produce an expected output to a pre-defined input, what is quite similar to unit testing. The tool, called Service Availability Monitoring or SAM, is being currently used by several different Virtual Organizations to monitor more than 300 grid sites belonging to the largest grids available today. We also discuss how this tool is being used by some of those VOs and how it is helping in the operation of the EGEE/WLCG grid.

Grid middleware and tools / 412

Monitoring the user/application activities on the grid

Authors: Antonio Pierro¹ ; Cristina AIFTIMIEL² ; Enrico Fattibene³ ; Giacinto Donvito¹ ; Giorgio Maggi⁴ ; Giuseppe Misurelli³ ; Guido Cuscela¹ ; Sergio Andreozzi³ ; Sergio Fantinel⁵ ; Stefano Dal Pra² ; Vihang Dudalkhar¹

¹ INFN-BARI

² INFN-PADOVA

³ INFN-CNAF

⁴ INFN-BARI – Politecnico di Bari

⁵ INFN-Padova/Legnaro

Corresponding Authors: antonio.pierro@ba.infn.it, giacinto.donvito@ba.infn.it

The monitoring of the grid user activity and application performance is extremely useful to plan resource usage strategies particularly in cases of complex applications. Large VO's, like the LHC ones, do their monitoring by means of dashboards. Other VO's or communities, like for example the BioinforGRID one, are characterized by a greater diversification of the application types: so the effort to provide a dashboard like monitor is particularly heavy.

We present in this paper the improvements introduced in GridICE, a general grid monitoring tool, to provide reports on the resources usage with the details of the VOMS groups, roles and users. By accessing the GridICE Web pages, the grid user can get all information that is relevant to keep track of his activity on the grid.

In the same way, the activity of a VOMS group can be distinguished from the activity of the entire VO.

In this paper we briefly talk about the features and advantages of this approach and, after discussing the requirements, we describe the software solutions, middleware and prerequisite to manage and retrieve the user's credentials.

Poster 2 - Board: 89 / 328

Monitoring with MonAMI: a case study.

Authors: Graeme Stewart¹ ; Greig Cowen¹ ; Paul Millar¹

¹ GridPP

Corresponding Author: p.millar@physics.gla.ac.uk

Computing resources in HEP are increasingly delivered utilising grid technologies, which presents new challenges in terms of monitoring. Monitoring involves the flow of information between different communities: the various resource-providers and the different user communities. The challenge is providing information so everyone can find what they need: from the local site administrators, regional operational centres through to end-users.

To meet this challenge, MonAMI was developed. MonAMI aims to be a universal sensor framework with a plugin architecture. This plugin structure allows flexibility in what is monitored and how the gathered information is reported. MonAMI supports gathering statistics from services like MySQL and Apache. The gathered data can be sent to many monitoring systems, including Ganglia, Nagios, MonaLisa and R-GMA.

This flexibility allows MonAMI to be plumbed into whatever monitoring system is being used. This avoids the current duplication of sensors, allowing gathered statistics to be presented within a greater diversity of monitoring systems.

Using the MonAMI framework, sensors have been developed for the DPM and dCache storage systems, both common at HEP grid centres. The development of these tools specifically to tackle the challenges of high availability storage is described. We illustrate how MonAMI's architecture allows a single sensor to both deliver long term trend information and to trigger alarms in abnormal conditions.

The use of MonAMI within the ScotGrid distributed Tier-2 is examined as a case study, illustrating both the ease with which MonAMI

integrates into existing systems and helps provide a framework for extending monitoring where necessary.

Online computing / 397

Multi-Agent Framework for Experiment Control Systems (AF ECS)

Author: Vardan Gyurjyan¹

Co-authors: Carl Timmer²; David Abbott²; Ed Jastrzembski²; Elliott Wolin²; Graham Heyes²

¹ *Jefferson Lab*

² *JEFFERSON LAB*

Corresponding Author: gurjyan@jlab.org

AF ECS is a pure Java based software framework for designing and implementing distributed control systems. AF ECS creates a control system environment as a collection of software agents behaving as finite state machines. These agents can represent real entities, such as hardware devices, software tasks, or control subsystems. A special control oriented ontology language (COOL), based on RDFS is provided for control system description as well as for agent communications. AF ECS agents can be distributed over a variety of platforms. All communication between the agents and their associated physical components are handled transparently by an underlying publish-subscribe communication system, cMsg, also developed at Jefferson Lab. This framework has been used to design the JLAB data acquisition run control system.

The main features of the framework, the COOL language in particular, as well as recent and near future upgrades will be discussed.

Poster 1 - Board: 23 / 27

Multi-threaded Event Reconstruction with JANA

Author: David Lawrence¹

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. 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 (and soon many-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 multi-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

Plenary / 488

Networks for High Energy Physics and Data Intensive Science, and the Digital Divide

Corresponding Author: harvey.newman@cern.ch

Networks of sufficient and rapidly increasing end-to-end capability, as well as a high degree of reliability are vital for the LHC and other major HEP programs. Our bandwidth usage on the major national backbones and intercontinental links used by our field has progressed by a factor of several hundred over the past decade, and the outlook is for a similar increase over the next decade. This growth is paralleled, and in some ways driven by the rapid development of national, continental and transoceanic networks serving research and education, which have recently made the transition from 10 Gbps to multi-10 Gbps optical infrastructures. Several of the major networks are currently working together towards a new “dynamic circuit” paradigm to meet the needs of HEP and other fields of “data intensive science”, while continuing to meet a broad range of other network needs.

I will review the recent developments, current status, and future directions for the world’s research networks and major international links used by high energy physics along with other scientific communities, and will touch on recent network technology and related advances on which our community depends and in which we have an increasingly important role. I will provide a brief update on the problem of the Digital Divide in our community, which is a primary focus of ICFA’s Standing Committee on Inter-regional Connectivity (SCIC), and highlight progress and approaches to solutions in some world regions.

Poster 2 - Board: 95 / 385

Next Steps in the Evolution of GridICE: a Monitoring Tool for Grid Systems

Authors: Antonio Pierro¹ ; Cristina Aiftimieci² ; Enrico Fattibene³ ; Giacinto Donvito¹ ; Giorgio Maggi¹ ; Giuseppe Misurelli³ ; Guido Cuscela¹ ; Sergio Andreozzi³ ; Sergio Fantinel⁴ ; Stefano Dal Pra⁵ ; Vihang Dudhalkar¹

¹ INFN-Bari

² INFN-Padova (on leave from NIPNE-HH)

³ INFN-CNAF

⁴ INFN-Padova/Legnaro

⁵ INFN-Padova

Corresponding Author: sergio.andreozzi@egi.eu

GridICE is an open source distributed monitoring tool for Grid systems that is integrated in the gLite middleware and provides continuous monitoring of the EGEE infrastructure. The main goals of GridICE are: to provide both summary and detailed view of the status and availability of Grid resource, to highlight a number of pre-defined fault situations and to present usage information. In this paper, we briefly summarize the core ideas and design choices behind this tool. In particular, we describe the 3-layer data distribution architecture composed by intra-site and inter-site collection plus the end-user presentation; furthermore we list the number of supported measurements. After this introduction, we discuss the main focus of the paper, that is the description of the next steps in the GridICE development as driven by user requirements. The purpose is to disseminate our evolution strategy in order to solicit for a community feedback.

Software components, tools and databases / 445

Next generation of OpenGL support in ROOT

Author: Matevz Tadel¹

¹ CERN

Corresponding Author: matevz.tadel@cern.ch

OpenGL has been promoted to become the main 3D rendering engine of ROOT. This required a major re-modularization of OpenGL support on all levels, from basic window-system specific interface to medium-level object-representation and top-level scene management. This new architecture allows seamless integration of external scene-graph libraries into the ROOT OpenGL viewer as well as inclusion of ROOT 3D scenes into external GUI and OpenGL-based 3D-rendering frameworks. Scene representation was removed from inside of the viewer, allowing scene-data to be shared among several viewers and providing for natural implementation of multi-view canvas layouts. The object-graph traversal infrastructure allows free mixing of 3D and 2D-pad graphics and makes implementation of ROOT canvas in pure OpenGL possible. Scene-elements representing ROOT objects trigger automatic instantiation of user-provided rendering-objects based on the dictionary information and class-naming convention. Additionally, a finer, per-object control over scene-updates is available to the user, allowing overhead-free maintenance of dynamic 3D scenes and creation of complex real-time animations. User-input handling was modularized as well, making it easy to support application-specific scene navigation, selection handling and tool management.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ALICE and ROOT project

Poster 1 - Board: 49 / 350

Nightly builds and software distribution in the LCG / AA / SPI project

Authors: Andreas Pfeiffer¹ ; Piotr Kolet¹ ; Stefan Roiser¹

¹ CERN

Corresponding Author: stefan.roiser@cern.ch

The Software Process and Infrastructure project (SPI) of the LCG Applications Area (AA) is responsible for a set of services for software build, software packaging, software distribution, communication and quality assurance. Recently a new tool has been developed in SPI for the automatic configuration and build of the LCG AA software stack which is used for nightly builds. In this talk the design, features and performance of this nightly build system will be presented. Examples for configurations currently in use and their maintenance will be discussed as well as constraints and difficulties in the development of the system (e.g. multi platform, architectures, compilers, configuration tools, performance issues). The nightly build system will also be used to release the whole LHC software stack in one go. There will also be an outlook to future developments like distributed and parallel builds. The latter being important to speed

up the delivery times for bug fixes of the LCG AA software stack to the user / experiments. This will become more important as we are nearing the startup of the LHC.

The second part of the talk will describe policies of software packaging and software distribution in the LHC environment. As we are coming closer to the startup of the LHC it will be more and more important to be able to distribute the LHC software in an easy and transparent way to the users. The tools in the SPI project which aim to achieve this functionality and possibly can also be used by experiments for their software packaging and distribution will be presented.

Poster 2 - Board: 86 / 256

On using a generic framework for integrating advanced batch systems to production-level grid infrastructures

Author: Sylvain Reynaud¹

Co-author: Fabio Hernandez ¹

¹ *IN2P3/CNRS*

Corresponding Author: sreynaud@in2p3.fr

Advanced capabilities available in nowadays batch systems are fundamental for operators of high-performance computing centers in order to provide a high-quality service to their local users. Existing middleware allow sites to expose grid-enabled interfaces of the basic functionalities offered by the site's computing service. However, they do not provide enough mechanisms for addressing the operational and scalability-related issues which are necessary for offering the same quality of service to the site's grid users.

This paper will present an overview of those issues and a proposed set of features required for the batch systems integration to a grid infrastructure. Then, we will describe a generic software framework that facilitates the implementation of those features. The framework is designed to interact with different grid execution services on one hand and with different batch systems on the other hand. Specialization is done via Java plug-ins, XSL plug-ins and XML configuration files.

To illustrate the benefits of the framework features, an example of its usage for interfacing our site's home-grown batch system (BQS) to the EGEE grid execution service (gLite CE) will also be presented. In particular, we will describe how the operational needs are addressed by the reusable components of the framework.

Poster 1 - Board: 6 / 140

Online Data Monitoring in the LHCb experiment

Author: Markus Frank¹

¹ *CERN*

Corresponding Author: markus.frank@cern.ch

The High Level Trigger and Data Acquisition system selects about 2 kHz of events out of the 40 MHz of beam crossings. The selected events are sent to permanent storage for subsequent analysis.

In order to ensure the quality of the collected data, identify possible malfunctions of the detector and perform calibration and alignment checks, a small fraction of the accepted events is sent to a monitoring farm, which consists of a few tens of general purpose PCs.

This contribution introduces the architecture of the data stream splitting mechanism from the storage system to the monitoring farm, where the raw data are analyzed by dedicated tasks. It describes the collaborating software components that are all based on the Gaudi event processing framework.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

LHCb Online

Online computing - Board: 15 / 337

Online and offline software for the CMS strip tracker data acquisition system

Authors: Laurent Mirabito¹; Robert Bainbridge²

Co-authors: Alessandro Giassi³; Christophe Delaere⁴; Domenico Giordano⁵; Frederic Drouhin⁶; Guillaume Beaulieu¹; Joanne Cole⁷; Jonathan Fulcher²; Kristian Hahn⁸; Laurent Gross⁹; Matthew Wingham²; Milan Nikolic¹⁰; Nicholas Cripps²; Sebastien Bel⁴; Slawomir Tkaczyk¹¹; Stefano Mersi¹²

¹ *Institut de Physique Nucleaire de Lyon*

² *Imperial College London*

³ *INFN, Sezione di Pisa*

⁴ *CERN*

⁵ *INFN & Università di Bari*

⁶ *Universite de Haute Alsace*

⁷ *Rutherford Appleton Laboratory*

⁸ *Massachusetts Institute of Technology*

⁹ *Institut Pluridisciplinaire Hubert Curien, CNRS*

¹⁰ *University of California*

¹¹ *Fermi National Accelerator Laboratory*

¹² *INFN & Università di Firenze*

Corresponding Author: robert.bainbridge@cern.ch

The CMS silicon strip tracker, providing a sensitive area of $>200 \text{ m}^2$ and comprising 10M readout channels, is undergoing final assembly at the tracker integration facility at CERN. The strip tracker community is currently working to develop and integrate the online and offline software frameworks, known as XDAQ and CMSSW respectively, for the purposes of data acquisition and detector commissioning. Recent developments have seen the integration of many new services and tools within the online data acquisition system, such as event building, online distributed analysis within CMSSW, an online monitoring framework, and data storage management. We review the various software components that comprise the strip tracker data acquisition system, the software architectures used for “local” and “global” data-taking modes, and our experiences during commissioning and operation of large-scale systems.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS tracker collaboration

Plenary / 467

Opening Session

Poster 2 - Board: 81 / 207

Optimising LAN access to grid enabled storage elements

Authors: Andrew Elwell¹ ; Graeme Stewart¹ ; Greig Cowan² ; Paul Millar¹

¹ *University of Glasgow*

² *University of Edinburgh*

Corresponding Authors: graeme.andrew.stewart@cern.ch, greig.cowan@cern.ch

When operational, the Large Hadron Collider experiments at CERN will collect tens of petabytes of physics data per year. The worldwide LHC computing grid (WLCG) will distribute this data to over two hundred Tier-1 and Tier-2 computing centres, enabling particle physicists around the globe to access the data for analysis. Different middleware solutions exist for effective management of storage systems at collaborating institutes. Two of these have been widely deployed at Tier-2 sites: the Disk Pool Manager (DPM) from EGEE and dCache, a joint project between DESY and FNAL. Two distinct access patterns are envisaged for these systems. The first involves bulk transfer of data between different Grid storage elements using protocols such as GridFTP. The second relates to how physics analysis jobs will read the data while running on Grid computing resources. Such jobs require a POSIX-like interface to the storage so that individual physics events can be extracted. Both DPM and dCache have their own protocols for POSIX access (rfio and gsidcap respectively) and it is essential that these scale with the available computing resources in order to meet the demands of physics analysis in the LHC era. In this paper we study the performance of these protocols as a function of the number of clients that are simultaneously reading data from the storage. We investigate server kernel tuning so as to optimise the performance of LAN access. We also consider the performance of these protocols from the point of view of real ATLAS and CMS analysis jobs.

Poster 1 - Board: 98 / 300

Optimization of dCache MSS tape efficiency through Virtual Volumes

Author: Manuel Delfino Reznicek¹

Co-authors: Daniel Alvarez² ; Esther Acción¹ ; Miquel-Angel Senar²

¹ *Port d'Informació Científica (PIC)*

² *Universitat Autònoma de Barcelona*

Corresponding Author: delfino@pic.es

Small files pose performance issues for Mass Storage Systems, particularly those using magnetic tape. The ViVo project reported at CHEP06 solved some of these problems by using Virtual Volumes based on ISO images containing the small files, and

only storing and retrieving these images from the MSS. Retrieval was handled using Unix automounters, requiring deployment of ISO servers with a separately managed cache. We report developments which extend the use of ISO-based Virtual Volumes to the dCache storage management system, using Castor1 or ENSTORE as the tape back-end. By using the MSS interface already implemented and documented in dCache, we have been able to catalog files into dCache, pack them into ISO volumes, store the ISO images to tape and then transparently allow retrieval of the individual files by dCache. A simple catalog of files and corresponding ISO images allows us to fetch from tape the precise ISO image containing the requested file when it is requested by the dCache MSS interface. We also have developed a cache read-ahead technique that allows the injection of all files in an ISO image from MSS into dCache, triggered by an original single file request at the user level. This can provide a high tape efficiency if the files in each ISO image have been chosen to have a high probability of being requested within a short period of time. This also allows delegation back to dCache of most of the ISO-related (or Virtual Volume related) cache management, simplifying implementation and maintenance.

Software components, tools and databases / 105

Optimizations in Python-based HEP Analysis

Authors: Sebastien Binet¹ ; Wim Lavrijsen¹

¹ LBNL

Corresponding Authors: sebastien.binet@cern.ch, wim.lavrijsen@cern.ch

Python does not, as a rule, allow many optimizations, because there are too many things that can change dynamically. However, a lot of HEP analysis work consists of logically immutable blocks of code that are executed many times: looping over events, fitting data samples, making plots. In fact, most parallelization relies on this. There is therefore room for optimizations.

There are many open source tools available to optimize python code. However, all of them stop short of dealing with calls into extension libraries, which is a major part of any Python-based HEP analysis.

The typical extension library in HEP is written in object-oriented C++ code, and used through interface pointers. In the analysis code, these pointers are then used to call specific functionality (e.g. to retrieve data), as well as simply passed around (e.g. to call a fit on selected data). In both cases, the Python part exists mostly for the convenience to the user in wiring the needed functionality together; it does not add functional code that the C++ extension library needs to be aware of.

The natural division in blocks, and the usage of Python as a conduit from C++ to C++, makes HEP analysis code particularly suited to the kind of partial evaluation and specialization techniques used in Psyco (psyco.sourceforge.net). In this paper, I will show how this is used to achieve automatic optimizations for HEP libraries bound with PyROOT.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Poster 2 - Board: 82 / 210

Optimized tertiary storage access in the dCache SE

Author: Martin Radicke¹

Co-author: Patrick Fuhrmann¹

¹ *DESY Hamburg*

Corresponding Author: martin.radicke@desy.de

The dCache software has become a major storage element in the WLCG, providing high-speed file transfers by caching datasets on potentially thousands of disk servers in front of tertiary storage. Currently dCache's model of separately connecting all disk servers to the tape backend leads to locally controlled flush and restore behavior has shown some inefficiencies in respect of tape drive utilization. This paper reports on enhancements to dCache and its interface to Hierarchical Storage Management Systems through a new module to centrally control the tape backend. The focus is to optimize the mount behavior, parallelization of requests and utilization of the available tape drive data rate. dCache is widely deployed in production instances, so improvements will be phased to maintain backwards compatibility. The first phase has been achieved by introducing the Flushmanager, a module to collect and coordinate all tape write operations. The second phase of implementation will be about collecting metadata and restore requests respectively. This information will be provided to user-added plug-ins via a clean interface, for managing restore requests using customized strategies. The third phase will focus on resource knowledge and drive steering with a generic and feature-rich interface to unify access to common Tertiary Storage Systems. The final phase will be the aggregation of the existing Flush- and Restore-Manager to implement a single module for controlling tape backends through dCache, offering a extensible programming framework to the community.

Poster 1 - Board: 31 / 109

Oracle RAC (Real Application Cluster) application scalability, experience with PVSS and methodology

Author: Eric Grancher¹

Co-author: Anton Topurov¹

¹ *CERN*

Corresponding Author: eric.grancher@cern.ch

Database applications increasingly demand higher performance. This is especially true in the context of the LHC accelerator, LHC experiments, and LHC Computing Grid projects at CERN. Oracle RAC (Real Application Cluster) is a cluster solution which allows a database to be served by several nodes, and is a technology that is being exploited successfully at CERN and at LCG Tier1 sites. Database applications often have initially low scalability, with a growing number of cluster nodes. This paper describes a methodology and innovative ideas developed in order to obtain almost linear scalability for some of the typical database workloads.

This paper describes, amongst others, the work which has been performed on the PVSS "Oracle archiver" (the controls software used for LHC and its experiments), where the performance of the event archiving module has been increased from 1000 to 150000 event changes per second (x150). This has been achieved with several architectural changes (core program separation from the data manipulation, data loading techniques, and database schema). The result is also a near-linear scalability between the number of nodes in the cluster and the performance.

Based on the experience gathered on many database projects, guidelines and tips are provided in order to help with the creation of scalable database applications, or to achieve scalability improvements for already designed database applications.

Summary:

This paper describes a methodology and ideas for achieving Oracle cluster (RAC) application for typical HEP workloads. An example is given with the PVSS (controls system used for LHC and LHC experiments) Oracle archiver scalability increase (x150 gain with a 6 nodes RAC).

Event processing / 414

Overview of the Inner Silicon detector alignment procedure and techniques in the RHIC/STAR experiment

Authors: Gene VAN BUREN¹ ; Jerome LAURET¹ ; Spyridon MARGETIS² ; Yuri Fisyak¹

Co-authors: Ivan KOTOV³ ; Jonathan BOUCHET⁴ ; Rafael Derradi de Souza⁵ ; Victor Perevoztchikov¹

¹ *BROOKHAVEN NATIONAL LABORATORY*

² *Kent State University, USA*

³ *Ohio State University, USA*

⁴ *Subatech, France*

⁵ *Instituto de Fisica da Universidade de Sao Paulo*

Corresponding Authors: fisyak@bnl.gov, jlauret@bnl.gov

The STAR experiment was primarily designed to detect signals of a possible phase transition in nuclear matter. Its layout, typical for a collider experiment, contains a large Time Projection Chamber (TPC) in a Solenoid Magnet, a set of four layers of combined silicon strip and silicon drift detectors for secondary vertex reconstruction plus other detectors. In this presentation, we will report on recent global and individual detector element alignment as well as drift velocity calibration work performed on this STAR inner silicon tracking system. We will show how attention to details positively impacts the Physics capabilities of STAR and explain the iterative procedure conducted to reach such result in low, medium and high track density / detector occupancy.

Grid middleware and tools / 167

PanDA: Distributed production and distributed analysis system for ATLAS

Author: Tadashi Maeno¹

¹ *Brookhaven National Laboratory*

Corresponding Author: tmaeno@bnl.gov

A new distributed software system was developed in the fall of 2005 for the ATLAS experiment at the LHC. This system, called PanDA, provides an integrated service architecture with late binding of jobs, maximal automation through layered services, tight binding with ATLAS distributed data management (DDM) system, advanced error discovery and recovery procedures, and other features. In this talk, we will describe the PanDa software system. Special emphasis will be placed on the evolution of PanDA

based on one and half year of real experience in carrying out CSC data production for ATLAS. The architecture of Panda is well suited for the computing needs of the ATLAS experiment, which is expected to be one of the first HEP experiments to operate at the petabyte scale.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Poster 1 - Board: 60 / 451

Parameterization of the LHCb magnetic field map

Authors: Adlene Hicheur¹ ; Geraldine Conti¹

¹ EPFL

Corresponding Authors: geraldine.conti@epfl.ch, adlene.hicheur@cern.ch

The LHCb warm magnet has been designed to provide an integrated field of 4 Tm for tracks coming from the primary vertex. To insure good momentum resolution of a few per mil, an accurate description of the magnetic field map is needed. This is achieved by combining the information from a TOSCA-based simulation and data from measurements. The paper presents the fit method applied to both the simulation and data to achieve the requirements. It also explains how the corresponding software tool is integrated in the LHCb Gaudi software and shows the relation with the environment in which it is used.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

LHCb

Software components, tools and databases / 316

Perfmon2 - A leap forward in performance monitoring

Authors: Ryszard Jurga¹ ; Sverre Jarp¹

¹ CERN

Corresponding Author: sverre.jarp@cern.ch

A new interface to the performance monitoring hardware of almost all supported hardware processors (AMD, IBM, INTEL, SUN, etc.) is in the process of being added to the Linux 2.6 kernel.

CERN openlab has participated in some of the development together with one of the key developers from HP labs.

In this talk we review the capabilities of this interface on relevant platforms, such as the recent Intel Core 2 based processors. Amongst other things, the interface enables a job profiling capability that is non-intrusive and can cover everything from a single process to an entire system. Dynamic libraries are also handled transparently, even if they are unloaded and reloaded – a key requirement for our applications. We review the design of this interface and we discuss how a test service has been established at CERN for analyzing the performance of the entire frameworks of the LHC experiments with an aim of allowing all developers to get control of the performance of their software. Finally we show some of the early results from this exciting new monitoring capability.

Summary:

Review of a new performance monitoring interface in the Linux kernel and description of how this can help tune LHC software frameworks.

Poster 1 - Board: 38 / 166

Performance Measurement and Monitoring for HENP Applications

Author: Sebastien Binet¹

Co-authors: Alex Undrus² ; Paolo Calafiura¹ ; Wim Lavrijsen¹

¹ LBNL

² BNL

Corresponding Author: sebastien.binet@cern.ch

LHC experiments are entering in a phase where optimization in view of data taking as well as robustness' improvements are of major importance. Any reduction in event data size can bring very significant savings in the amount of hardware (disk and tape in particular) needed to process data. Another area of concern and potential major gains is reducing the memory size and I/O bandwidth requirements of processing nodes, especially with increasing usage of multi-core CPUs. LHC experiments are already collecting abundant performance information about event size, memory and CPU usage, I/O compression and bandwidth requirements. What is missing is a coherent set of tools to present this information in a tailored fashion to release coordinators, package managers and physics algorithm developers.

This paper describes such a toolkit that we are developing in the context of ATLAS computing to harvest performance monitoring information from an extensible set of sources. The challenge is to map performance data with an immediate impact on hardware costs into entities which are relevant for the various users. For example the toolkit allows an ATLAS data model developer to evaluate the impact on resource usage throughout the entire software pipeline of their design decisions for an event data object.

We present the data in a way that highlights potential areas of concerns, allowing experts to drill down to the level of detail they need (the size of a data member of a class, the CPU usage of a component method). A configurable monitoring system allows to set off alarms when a quantity or an histogram goes out of a specified range. This allows a release coordinator to monitor e.g. the global size of a data stream throughout a development cycle and have the developers correct a problem well before a release goes into production.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Poster 2 - Board: 112 / 499

Performance study of multicore systems with the ATLAS software Athena

Authors: Graeme A Stewart¹ ; Marco La Rosa² ; Markus Bischofberger³

¹ *University of Glasgow*

² *The University of Melbourne*

³ *University of Melbourne*

Corresponding Author: mlarosa@physics.unimelb.edu.au

With the proliferation of multi-core x86 processors, it is reasonable to ask whether the supporting infrastructure of the system (memory bandwidth, IO bandwidth etc) can handle as many jobs as there are cores. Furthermore, are traditional benchmarks like SpecINT and SpecFloat adequate for assessing multi-core systems in real computing situations. In this paper we present the results of simulation and full reconstruction jobs with the ATLAS software Athena on both Intel and AMD multi-core systems. The aim of this paper is to examine whether there is a performance penalty associated with multi-core systems. And if there is, the threshold (number of jobs / cores) at which that penalty is first observed.

Computer facilities, production grids and networking / 372

PetaCache: Data Access Unleashed

Author: Richard Mount¹

Co-authors: Andrew Hanushevsky¹ ; Chuck Boeheim¹ ; David Leith¹ ; Jacek Becla¹ ; Randal Melen¹ ; Teela Pulliam¹ ; William Weeks¹

¹ *SLAC*

Corresponding Author: richard.mount@slac.stanford.edu

The PetaCache project started at SLAC in 2004 with support from DOE Computer Science and the SLAC HEP program. PetaCache focuses on using cost-effective solid state storage for the hottest data under analysis. We chart the evolution of metrics such as accesses per second per dollar for different storage technologies and deduce the near inevitability of a massive use of solid-state storage in the near future. We report on the latency and access-rate performance of a DRAM-based prototype constructed in 2005 using commodity hardware and a Flash-based prototype constructed in 2007 using purpose-built hardware. We describe the use of xrootd to cluster individual servers in a highly scalable and fault tolerant approach, and present tests of scalability. A study of access to ATLAS AOD data is reported as a first step in understanding the software issues that will be encountered in achieving unfettered access to objects within HEP events. Finally we examine the cost-benefit outlook for the use of solid-state storage in HEP experiments.

Event processing / 179

Physics Analysis Tools for Beauty Physics in ATLAS

Authors: Maria Smizanska¹ ; Pavel Reznicek²

¹ *Lancaster University*

² *IPNP, Charles University in Prague*

Corresponding Authors: pavel.reznicek@cern.ch, maria.smizanska@cern.ch

The LHC experiments will search for physics phenomena beyond the Standard Model (BSM). Highly sensitive tests of beauty hadrons will represent an alternative approach to this research. The analyzes of complex decay chains of beauty hadrons will require involving several nodes, and detector tracks made by these reactions must be extracted efficiently from other events to make sufficiently precise measurements. This places severe demands on the software used to analyze the B-physics data. The ATLAS B-physics group has written series of tools and algorithms for performing these tasks, to be run within the ATLAS offline software framework ATHENA. The presentation will describe this analysis suite, paying particular attention to mechanisms for handling combinatorics, interfaces to secondary vertex fitting packages, B-flavour tagging tools and finally Monte Carlo truth association to pursue simulation data in process of the software validations which is important part of the development of the Physics Analysis tools.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Software components, tools and databases / 376

Physics Analysis Tools for CMS experiment at LHC

Author: Luca Lista¹

Co-authors: Benedikt Hegner² ; Christopher D Jones³ ; Francesco Fabozzi¹

¹ *INFN Napoli*

² *DESY*

³ *Cornell University*

Corresponding Author: luca.lista@cern.ch

By end of 2007 the CMS experiment will start running and Petabytes of data will be produced every year. To make analysis of this huge amount of data possible the CMS Physics Tools package builds the highest layer of the CMS experiment software. A core part of this package is the Candidate Model providing a coherent interface to different types of data. Standard tasks like combinatorial analyses, generic cuts, MC truth matching and fitting are supported. Advanced template techniques enable the user to add missing features easily. We explain the underlying model, certain details of implementation and present some use cases showing how the tools are currently used in generator and full simulation studies as preparation for analysis of real data.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

for the CMS Offline group

Poster 1 - Board: 74 / 152

Portable Monitoring for Castor2

Authors: Alasdair EARL¹ ; Miguel Coelho Dos Santos¹

Co-authors: Dennis Waldron¹ ; Jan van ELDIK¹ ; Olof Barring¹

¹ CERN

Corresponding Author: miguel.coelho.santos@cern.ch

We present our design, development and deployment of a portable monitoring system for the CERN Archival and Storage System (Castor) based on its existing internal database infrastructure and deployment architecture.

This new monitoring architecture is seen as an important requirement for future development and support. Castor is now deployed at several sites which use different monitoring systems to the LHC Era Monitoring (Lemon) system used at CERN. This includes sites with significant computing resources in the United Kingdom and Italy. Providing a portable monitoring system is seen as desirable as this will reduce development overhead and provide a common framework for understanding the state of the systems and resolving operational tasks.

We present an overview of the reasoning behind this project and its aims; a discussion on the various aspects of the system which have previously been monitored and how moving to this new system improves on this; and discuss development trade-offs and our future plans.

Summary:

CASTOR, Quattor, monitoring, operations management

Plenary / 492

Power and Air Conditioning Challenges in Computer Centres

Corresponding Author: cope@fnal.gov

Plenary / 496

Power, Density, Reliability & Performance: customer-driven evolution of cluster computing and storage, in high energy physics and other scientific fields.

Cluster systems now comprise 50% to 90% of the High Performance Computing (HPC) market. However, with computing and storage needs outpacing Moore's law, the traditional approach of scaling is giving rise to facility, administrative and performance issues. Details of industry trends and unmet customer requirements for cluster computing will be presented. Implications on systems and facility design will also be explored.

Summary:

Eng Lim Goh, Ph.D.
Senior Vice President and Chief Technology Officer , SGI

Dr. Eng Lim Goh has been with SGI for 17 years, becoming one of the chief scientists in 1998 and chief technology officer in 2001. His tenure includes work in computer graphics algorithms and high-performance computing (HPC) architectures.

In HPC, he oversees Project Ultraviolet, the goal of which is to design and build the company's next generation science and high-performance enterprise-driven computer architecture. He is also the coauthor of SGI's recommendation to the high-end

computing revitalization task force (HECRTF) for federal funding of key corresponding technologies. This proposal was reviewed by HECRTF in 2003 and judged to be one of the top submitted papers.

He is a proponent of next-generation computer systems designed specifically for customer applications performance. To this, he advocates computational density and energy efficiency when integrating commodity components; and where appropriate, a balanced multi-paradigm approach, across a globally addressable memory, to architectural design.

In computer graphics, Dr. Goh's current research interest is in the relationships between human visual perception and visual computing. He has been awarded three U.S. patents in this field. He is also leading a small research effort to investigate application-transparent, massively parallel rendering.

In 2005, the IDG publication, InfoWorld, named Dr. Goh one of World's 25 most influential CTOs. That same year, he was also included in the HPCwire list of "15 People to Watch." In 2007, Dr. Goh was named in "Champions 2.0" for the bio-IT industry by BioIT World magazine.

Before joining SGI, Dr. Goh worked for Intergraph Systems, Schlumberger Wireline Netherlands, and Shell Research U.K. A Shell Cambridge University Scholar, he completed his Ph.D. research and dissertation on parallel architectures and computer graphics. He also holds a first-class honors degree in mechanical engineering from Birmingham University, U.K.

Poster 1 - Board: 100 / 323

Preparation for WLCG production from a Tier-1 viewpoint

Authors: Andrew Sansum¹ ; Martin Bly¹

¹ *STFC/RAL*

Corresponding Author: m.j.bly@rl.ac.uk

The GRIDPP Tier-1 Centre at RAL is one of 10 Tier-1 centres worldwide preparing for the start of LHC data taking in late 2007. The RAL Tier-1 is expected to provide a reliable grid-based computing service running thousands of simultaneous batch jobs with access to a multi-petabyte CASTOR-managed disk storage pool and tape silo, and will support the ATLAS, CMS and LHCb experiments as well as many other experiments already taking or analysing data.

The RAL Tier-1 is already well advanced towards readiness for LHC data-taking. We describe some of the reliability and performance issues encountered with various generations of storage hardware in use at RAL and how the problems were addressed.

We describe the networking challenges for shipping late volumes of data into and out of the Tier-1 storage systems, and system to system within the Tier-1, and the changes made to accommodate the expected data volumes.

We describe the scalability and reliability issues encountered with the grid-services and the various strategies used to minimise the impact of problems, including multiplying the number of service hosts, splitting services across a number of hosts, and upgrading services to more resilient hardware.

Primary Vertex Reconstruction in the ATLAS Experiment at the LHC

Authors: Andreas Wildauer¹ ; Giacinto Piacquadio² ; Kirill Prokofiev³

¹ CERN

² Albert-Ludwigs-Universität Freiburg

³ University of Sheffield

Corresponding Authors: kprok@mail.cern.ch, andreas.wildauer@cern.ch

In the harsh environment of the Large Hadron Collider at CERN (design luminosity of 10^{34} cm⁻²s⁻¹) efficient reconstruction of the signal primary vertex is crucial for many physics analyses. Described in this paper are primary vertex reconstruction strategies implemented in the ATLAS software framework Athena. The implementation of the algorithms follows a very modular design based on object oriented C++ and use of abstract interfaces. This guarantees the easy use and exchange of different vertex fitters and finders which are considered for a given analysis. This modular approach relies on a dedicated Event Data Model for vertex reconstruction. The data model has been developed alongside with the reconstruction algorithms. Its design is presented in detail. The performance of the implemented primary vertex reconstruction algorithms has been studied on a variety of Monte Carlo samples and results are presented.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS Offline Computing

Poster 2 - Board: 2 / 74

Printing at CERN

Author: Rafal Otto¹

Co-authors: Juraj Sucik¹ ; Michal Kwiatek¹

¹ CERN

Corresponding Authors: michal.kwiatek@cern.ch, rafal.otto@cern.ch

For many years at CERN we had a very sophisticated print server infrastructure which supported several different protocols (AppleTalk, IPX and TCP/IP) and many different printing standards. Today's situation differs a lot: we have much more homogenous network infrastructure, where TCP/IP is used everywhere and we have less printer models, which almost all work with current standards (i.e. they all provide PostScript drivers). This change gave us a possibility to review the current printing architecture aiming in simplification of the infrastructure we have and in full automation of the service.

The new infrastructure offer both LPD service exposing print queues to Linux and Mac OS X computers and native printing for Windows based clients. The printer driver distribution is automatic and native on Windows and is automated by custom mechanisms on UNIX, where the appropriate Foomatic drivers are configured. Also the process of a printer registration and queue creation is completely automated following the printer registration in the network database.

At the end of 2006 we have migrated all (1100+) CERN printers and all user's connections at CERN to the new service. The talk will describe a new architecture and summarize process of migration.

Poster 2 - Board: 17 / 233

Production Experience Using Resilient dCache

Author: Alexander Kulyavtsev¹

¹ *FNAL*

Corresponding Author: aik@fnal.gov

dCache is a distributed storage system which today stores and serves petabytes of data in several large HEP experiments. Resilient dCache is a top level service within dCache, created to address reliability and file availability issues when storing data for extended periods of time on disk-only storage systems. The Resilience Manager automatically keeps the number of copies within specified bounds by adjusting the number of replicas of each logical file on different units of disk hardware when files disk pool nodes are found to have crashed, been removed from, or added to the system.

We presented design of the dCache Resilience Manager in the CHEP2006 report “Resilient dCache: Replicating Files for Integrity and Availability”. The present paper provides an update on further development of Resilient Manager and experience in the production deployment and operations in US-CMS T1 and T2 centers. The US-CMS T1 center substantially increased the size of their Resilient dCache and added second group of resilient pools for merging short files with production job output before storing files on tape. Two resilient pool groups operate independently of each other and other pool groups (tape-backed or volatile). A few more US-CMS T2 centers started to use Resilient Manager to increase the integrity and size of their systems. Based on experience with the Resilient Manager in US-CMS centers we added new features to drain files from the pools for hardware retirement and to avoid replication of files to the same pool host, while improving the Resilient Manager’s performance and manageability.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

dCache

Computer facilities, production grids and networking / 171

Production Experience with Distributed Deployment of Databases for the LHC Computing Grid

Author: Dirk Duellmann¹

¹ *CERN*

Corresponding Author: dirk.duellmann@cern.ch

Relational database services are a key component of the computing models for the Large Hadron Collider (LHC). A large proportion of non-event data including detector conditions, calibration, geometry and production bookkeeping metadata require reliable storage and query services in the LHC Computing Grid (LCG). Also core grid services to catalogue and distribute data cannot operate without a database infrastructure at CERN and the LCG

sites. The Distributed Deployment of Databases (3D) project is a joint activity between CERN's IT department, the LHC experiments and LCG sites to implement database services that are coherent, scalable and highly available.

This contribution describes the LCG 3D service architecture based on database clusters and data replication and caching techniques, which is now implemented at CERN and ten LCG Tier-1 sites. The experience gained with this infrastructure throughout several experiment conditions data challenges and the LCG dress rehearsal is summarised and an overview of the remaining steps to prepare for full LHC production will be given.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

LCG 3D Project

Poster 2 - Board: 110 / 456

Production of BaBar skimmed analysis datasets using the Grid

Author: Gregory Dubois-Felsmann¹

¹ SLAC

Corresponding Author: gpdf@slac.stanford.edu

The BaBar experiment currently uses approximately 4000 KSI2k on dedicated Tier 1 and Tier 2 compute farms to produce Monte Carlo events and to create analysis datasets from detector and Monte Carlo events. This need will double in the next two years requiring additional resources.

We describe enhancements to the BaBar experiment's distributed system for the creation of skimmed analysis datasets from detector and Monte Carlo-generated data using European LCG and US OSG Grid resources and present the results with regard to the latest production run over BaBar's accumulated data. The benefits of local and Grid-based systems, the ease with which the system is managed and the challenges of integrating the Grid with legacy software will be presented as well as the audit and monitoring software needed to control the system on the Grid. We compare job success rates and manageability issues between Grid and non-Grid production and present an investigation into the relative efficiency of the components of the Grid with particular reference to exporting and accessing input data in a distributed environment.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

BaBar

Grid middleware and tools / 37

Providing a Single View of Heterogeneous Clusters using Platform LSF

Author: Chris Smith¹

¹ *Platform Computing*

Corresponding Authors: rstober@platform.com, csmith@platform.com

Universus refers to an extension to Platform LSF that provides a secure, transparent, one-way interface from an LSF cluster to any foreign cluster. A foreign cluster is a local or remote cluster managed by a non-LSF workload management system. Universus schedules work to foreign clusters as it would to any other execution host.

Beyond its ability to interface with foreign workload management systems, the two most important features of Universus are its security and its transparency. Universus leverages the LSF Kerberos 5 integration to provide user and daemon authentication, and a Kerberized Secure Shell implementation to perform encrypted file transfers and to securely execute commands on remote systems.

Transparency can best be described as making jobs that are actually executing within a foreign cluster ³look and feel² like native LSF jobs from the end users perspective. Universus provides transparency on both the command line and the LSF Web UI level. Universus also provides transparency to the LSF system by obtaining accurate exit status from foreign jobs, even when the foreign cluster provides no such functionality.

Poster 2 - Board: 6 / 135

Publication in scientific journals: the role of HEP computing

Author: Maria Grazia Pia¹

¹ *INFN GENOVA*

Corresponding Author: maria.grazia.pia@cern.ch

Journal publication plays a fundamental role in scientific research, and has practical effects on researchers' academic career and towards funding agencies. An analysis is presented, also based on the author's experience as a member of the Editorial Board of a major journal in Nuclear Technology, of publications about high energy physics computing in refereed journals.

The statistical distribution of papers associated to various fields of HEP computing (simulation, data reconstruction and analysis, online computing, grid computing etc.) published in representative journals is critically analyzed.

The relative contribution of HEP computing is evaluated with respect to published papers in other domains: articles on computing in other closely related physics disciplines, like nuclear physics, space science and medical physics, and on hardware developments for high energy physics experiments.

The statistical results hint to the fact that, in spite of the significant effort invested in high energy physics computing and its fundamental role in the experiments, this research area is underrepresented in scientific literature. HEP computing seems also to be largely absent from the current debate on Open Access publishing in scientific research.

The implications of the picture emerging from this analysis as a perception of computing in high energy physics as a technical service rather than a scientific research domain are discussed, and recommendations for a more effective presence of HEP computing in scientific literature are proposed.

Quattor and QWG Templates : efficient management of (complex) grid sites

Author: Michel Jouvin¹

¹ LAL / IN2P3

Corresponding Author: jouvin@lal.in2p3.fr

Quattor is a tool aimed at efficient management of fabrics with hundred or thousand of Linux machines, still being easy enough to manage smaller clusters. It has been originally developed inside the European Data Grid (EDG) project. It is now in use at more than 30 grid sites running gLite middleware, ranging from small LCG T3 to very large one like CERN.

Main goals and specific features of Quattor are :

- Abstract, service oriented description of machine configuration, based on “templates” that can be heavily factorized.
- Machine configuration described in term of final state, instead of actions required to reach the final state.
- Manage both installation and configuration changes from the same configuration description.
- Configuration description versioning, allowing very easy rollback of any change.

Quattor is particularly well suited for management of complex sites made of several clusters or subsites spread over several geographical locations, like federated T2s.

Quattor ability to factorize common part of description configuration and advanced features of PAN language used to do this description allowed to build and maintained a common set of templates that any site can just import and customize without editing them. This resulted in so-called QWG templates, a complete set of standard templates to configure OS and gLite middleware. This results in a very efficient sharing of installation and configuration tasks around the world.

Even if Quattor can be and is used to manage non grid resources (including desktops), this talk will concentrate on Quattor benefits to manage grid resources running gLite, especially using QWG templates.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

LCG Summary:

This talk will present Quattor unique features leveraging administration of a grid site, in particular in the context of distributed T2s or small sites. It will focus on QWG templates, a collaborative effort to produce a standard configuration description that site can customize without modifying it.

Poster 2 - Board: 23 / 311

RGLite, an interface between ROOT and gLite

Authors: Anar Manafov¹ ; Kilian Schwarz¹ ; Peter Malzacher¹

¹ GSI

Corresponding Author: k.schwarz@gsi.de

After all LHC experiments managed to run globally distributed Monte Carlo productions on the Grid, now the development of tools for equally spread

data analysis stands in the foreground. To grant Physicists access to this world suited interfaces must be provided. As a starting point serves the analysis framework ROOT/PROOF, which enjoys a wide distribution within the HEP community. Using abstract ROOT classes (TGrid, ...) interfaces can be implemented via which Grid access directly from ROOT can be accomplished. A concrete implementation exists already for the ALICE Grid environment AliEn via which also the distribution of PROOF daemons in the Grid can be taken care of. Within the D-Grid project now also an interface to the common Grid middleware of all LHC experiments, gLite, has been created. Herewith it is possible to query Grid File Catalogues directly from inside ROOT for the location of the data to be analysed, Grid jobs can be submitted into a gLite based Grid, the status of the jobs can be asked for, and the results can be obtained. It is shown that it is possible by using RGLite as well to send PROOF daemons as Grid jobs, to start a PROOF session by connecting to the submitted PROOF daemons, and to perform a data analysis using PROOF and gLite.

The possibility to create a PROOF analysis cluster shared by several computing centres using existing middleware versions of gLite and/or Globus is being investigated.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

D-Grid HEPCG Summary:

a ROOT-gLite interface, RGLite, has been implemented via which it is possible to do most interactions with a gLite based Grid directly from within ROOT scripts.

Software components, tools and databases / 198

ROOT Graphics: status and future.

Author: Olivier Couet¹

¹ CERN

Corresponding Author: olivier.couet@cern.ch

The ROOT graphical libraries provide support for many different functions including basic graphics, high-level visualization techniques, output on files, 3D viewing etc. They use well-known world standards to render graphics on screen, to produce high-quality output files, and to generate images for Web publishing. Many techniques allow visualization of all the basic ROOT data types, projected in different dimensions and coordinate systems and the production of high quality output for publication purposes. This paper will present the current status of the ROOT graphics including recent developments in 2D representations and the latest developments in the 3D area based on OpenGL. As OpenGL is becoming the standard cross-platform basic graphics package, the current work being done to base all ROOT screen rendered graphics (2D and 3D) on OpenGL will also be presented. Finally we will present some visualization techniques not (yet) available in the ROOT framework which might be of some interest in the future.

Distributed data analysis and information management / 69

ROOTlets and Pythia: Grid enabling HEP applications using the Clarens Toolkit

Authors: Conrad Steenberg¹ ; Julian Bunn¹

Co-authors: Franciscus van Lingen¹ ; Harvey Newman¹ ; Michael Thomas¹

¹ *Caltech*

Corresponding Authors: conrad@hep.caltech.edu, julian.bunn@caltech.edu

We describe how we have used the Clarens Grid Portal Toolkit to develop powerful application and browser-level interfaces to ROOT and Pythia. The Clarens Toolkit is a codebase that was initially developed under the auspices of the Grid Analysis Environment project at Caltech, with the goal of enabling LHC physicists engaged in analysis to bring the full power of the Grid to their desktops, while at the same time not altering the look, feel and interface of their chosen analysis tool. By wrapping existing applications, and providing a well documented wrapper API, client applications are able to exchange commands, data and results using standard protocols such as XML-RPC, HTTP and HTTPS. In particular, we have implemented a wrapper to the Pythia particle collision simulation code, and developed an encapsulated form of the ROOT environment, called a ROOTlet, that allows convenient access from within ROOT to a collection of Clarens-based ROOTlet servers, distributed around the Grid. An attractive feature of the implementation is that the ROOTlets run standard, unadorned instances of ROOT and the clients only need to load a small runtime plugin to gain access to the loosely-coupled system. This paper describes these developments, work in progress and plans for future enhancements.

Poster 2 - Board: 3 / 80

RSS based CERN Alerter

Authors: Emmanuel Ormancey¹ ; Rafal Otto¹

¹ *CERN*

Corresponding Authors: emmanuel.ormancey@cern.ch, rafal.otto@cern.ch

Nearly every large organization use a tool to broadcast messages and information across the internal campus (messages like alerts announcing interruption in services or just information about upcoming events). The tool typically allows administrators (operators) to send “targeted” messages which is sent only to specific group of users or computers (for instance only those ones located in a specified building or connected to a particular computing service).

CERN has a long history of such tools: CERNVMS’s SPM_“MESSAGE” command, Zephyr and the most recent NICE Alerter based on NNTP protocol. Today, the current NICE Alerter used on all Windows-based computers has to be phased out as a consequence of the NNTP stoppage at CERN. Zephyr still used on Linux workstations will probably be replaced as well.

The new solution to broadcast information messages on the CERN campus will continue to provide the service based on cross-platform technologies, minimizing custom developments and relying on commercial software as much as possible. The new proposal is based on RSS for the transport protocol and will use Microsoft SharePoint as the backend for database and posting interface. The windows-based client will rely on Internet Explorer 7.0 with a custom script to trigger the notifications for new events. Linux and MacOS clients could also rely on any RSS readers to subscribe to targeted notifications.

The talk will cover the architecture and implementation aspects of the new system.

Poster 2 - Board: 24 / 327

Rapid-response Adaptive Computing Environment

Authors: Christos Lazaridis¹ ; Dan Bradley¹ ; Sridhara Dasu² ; Vishal Mehta¹

Co-authors: Ajit Mohapatra¹ ; William Maier¹

¹ UNIVERSITY OF WISCONSIN

² University of Wisconsin

Corresponding Author: dasu@hep.wisc.edu

We describe the ideas and present performance results from a rapid-response adaptive computing environment (RACE) that we setup at the UW-Madison CMS Tier-2 computing center. RACE uses Condor technologies to allow rapid-response to certain class of jobs, while suspending the longer running jobs temporarily. RACE allows us to use our entire farm for long running production jobs, but also harness a portion of it for unpredictable shorter period user analysis jobs. RACE features are ideal at Tier-2 computing centers where farm usage will become less than optimal if a portions of the farm are dedicated to long and short queues.

Summary:

RACE uses Condor technologies to allow rapid-response to a chosen set of jobs, while suspending the longer running jobs temporarily. We have explored two mechanisms, one that is based on computing-on-demand implementation that does not have queueing and another that uses a parallel scheduler. Both mechanisms use the operating system services to suspend and release the existing job process. The suspended jobs free-up both CPU and memory, so the new jobs have access to the complete resources of the system. There is a period of time during which there is some contention for resources. The Condor computing-on-demand implementation minimizes this contention, but it does not provide any accounting nor prioritization of new jobs. We have used computing-on-demand with PROOF. After some improvements to Condor and PROOF classes, we were satisfied with job suspension and resumption times. We will present latency and resumption time results. However, we were not happy with the restricted services on both Condor job scheduling, monitoring and accounting side, and by the PROOF limitation of the analysis jobs to those written in ROOT framework only. Therefore, we have explored an alternate mechanism using multiple schedulers for the same set of virtual machines. Condor was configured such that when higher priority scheduler has jobs to run, it suspends the normal priority jobs. This way both schedulers provided complete Condor services. When the higher priority jobs are done, the normal priority jobs resumed. We have tuned the scheduler performance so that the mechanism can be used in practice. We will also present timing results for this setup.

For high-energy physics usage, large numbers of long running production jobs can be submitted to the normal

priority scheduler, and the ephemeral and chaotically appearing analysis jobs to the high priority scheduler. This way the usage of the computing farms is maximized, and the analysis jobs get processed rapidly. We have written simple scripts that automatically divide the job into small chunks so that large datasets can be processed in a distributed way in a short amount of time. We will provide statistics of usage on our farm where CMS simulation production and CMS high-level trigger exercise related analysis jobs were processed. We will also provide other ideas for configuration or multi-scheduler Condor operational environments.

Event processing / 442

Raw-data display and visual reconstruction validation in ALICE

Author: Matevz Tadel¹

¹ CERN

Corresponding Author: matevz.tadel@cern.ch

ALICE Event Visualization Environment (AliEVE) is based on ROOT and its GUI, 2D & 3D graphics classes. A small application kernel provides for registration and management of visualization objects. CINT scripts are used as an extensible mechanism for data extraction, selection and processing as well as for steering of frequent event-related tasks. AliEVE is used for event visualization in offline and high-level trigger frameworks. The first emphasis of the talk is on visual representations of raw-data for different detector-types. Common infrastructure for thresholding and color-coding of signal/time information, placement of detector-modules in various 2D/3D layouts and for user-interaction with displayed data is presented. Methods for visualization of raw-data on different levels of detail are discussed as they are expected to play an important role during early detector operation with poorly understood detector calibration, occupancy and noise-levels. The second emphasis of the talk is put on tools developed for visual validation of reconstruction code on event-by-event basis. Since September 2006 ALICE applies a regular visual-scanning procedure to simulated proton-proton data to detect any shortcomings in cluster finding, tracking and primary & secondary vertex reconstruction. A high-level of interactivity is required to allow in-depth exploration of event-structure and navigation back to simulation records is supported for debugging purposes. Standard 2D projections and transformations are available for clusters, tracks and simplified detector geometry

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ALICE

Event processing / 239

Readiness of CMS Simulation towards LHC Startup

Author: Sunanda Banerjee¹

¹ *Fermilab/TIFR*

Corresponding Author: sunanda.banerjee@cern.ch

The CMS simulation based on the Geant4 toolkit and the CMS object-oriented framework has been in production for more than three years and has delivered a total of more than 200 M physics events for the CMS Data Challenges and Physics Technical Design Report studies. The simulation software has been successfully ported to the new CMS Event-Data-Model based software framework and is used in understanding the data taken by CMS in test beams as well as in the Magnet Test and Cosmic Challenge. In this paper, we present the experience from years in physics production, the migration process to the new architecture, efforts towards robustness and performance of the simulation chain and different operational scenarios in terms of hit simulation, event mixing and digitization.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS Offline group

Distributed data analysis and information management / 310

Real-time Data Access Monitoring in Distributed, Multi-Petabyte Systems

Author: Tofigh Azemoon¹

Co-authors: Andrew Hanushevsky¹ ; Jacek Becla¹ ; Turri Massimiliano¹

¹ *Stanford Linear Accelerator Center*

Corresponding Author: azemoon@slac.stanford.edu

Petascale systems are in existence today and will become widespread in the next few years. Such systems are inevitably very complex, highly distributed and heterogeneous. Monitoring a petascale system in real time and understanding its status at any given moment without impacting its performance is a highly intricate task. Common approaches and off the shelf tools are either unusable, do not scale, or severely impact the performance of the servers that are monitored. This talk will describe an unobtrusive monitoring software developed at Stanford Linear Accelerator Center (SLAC) and currently deployed by the BaBar Experiment that uses the xrootd file access system to access its highly distributed petascale production data set. The system facilitates central monitoring of all BaBar Tier A centers at SLAC. The talk will describe the employed solutions, the lessons learned, and the issues still to be addressed, and discuss the advantages of such a system in predicting the storage needs and understanding data access patterns. It will further explain how the system can be deployed in other High Energy Physics centers where the data servers may be shared by many experiments and run under a different file access system.

Poster 1 - Board: 16 / 342

Real-time analysis of the operational state of the CMS strip tracker readout system

Authors: Laurent GROSS¹ ; Milan NIKOLIC² ; Robert Bainbridge³ ; Stefano Mersi⁴

Co-authors: Alessandro GIASSI⁵; Christophe DELAERE⁶; Domenico GIORDANO⁷; Frederic DROUHIN⁸; Guillaume BEAULIEU⁹; Joanne COLE¹⁰; Jonathan FULCHER³; Kristian HAHN¹¹; Laurent MIRABITO⁹; Matthew WINGHAM³; Nicholas CRIPPS³; Sebastien Bel⁶; Slawomir TKACZYK¹²

¹ *Institut Pluridisciplinaire Hubert Curien, CNRS*

² *University of California*

³ *Imperial College London*

⁴ *INFN & Università di Firenze*

⁵ *INFN, Sezione di Pisa*

⁶ *CERN*

⁷ *INFN & Università di Bari*

⁸ *Universite de Haute Alsace*

⁹ *Institut de Physique Nucleaire de Lyon*

¹⁰ *Rutherford Appleton Laboratory*

¹¹ *Massachusetts Institute of Technology*

¹² *Fermi National Accelerator Laboratory*

Corresponding Authors: stefano.mersi@cern.ch, robert.bainbridge@cern.ch

The CMS silicon strip tracker comprises a sensitive area of >200 m² and 10M readout channels. Its data acquisition system is based around a custom analogue front-end ASIC, an analogue optical link system and an off-detector VME board that performs digitization, zero-suppression and data formatting. The data acquisition system uses the CMS online software framework, known as XDAQ, to configure, control and monitor the hardware components and steer the data acquisition. Recent developments have seen the integration of the CMS offline software framework, known as CMSSW, within the online data acquisition system. This provides many new features and services within the online environment, such as distributed analysis within CMSSW, access to geometry and conditions data, and a monitoring framework. We review how the monitoring frameworks available within both XDAQ and CMSSW will be used to assess the operational state of the hardware components of the strip tracker readout system during data-taking and provide real-time feedback to shifters in the CMS control room. We will report on the software components, the chosen architecture, the various monitoring streams available, and our experiences of commissioning and operating large-scale systems at the tracker integration facility.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS tracker collaboration

Distributed data analysis and information management / 237

Real-time dataflow and workflow with the CMS Tracker data

Authors: Carsten Noeding¹; Domenico Giordano²; Fabrizio Palla³; Giuseppe Bagliesi³; Nicola De Filippis⁴; Subir Sarkar³; Tommaso Boccali³; Vitaliano Ciulli⁵

Co-authors: Laurent Mirabito⁶; Robert Bainbridge⁷

¹ *Fermi National Accelerator Laboratory, Batavia (IL), USA*

² *Università di Bari e INFN Sezione di Bari*

³ *INFN - Sezione di Pisa*

⁴ *INFN - Sezione di Bari*

⁵ *Università di Firenze e INFN Sezione di Firenze*

⁶ *CERN PH/CMT*

⁷ *High Energy Physics Group Blackett-Lab - Imperial College, London*

Corresponding Author: subir.sarkar@cern.ch

The Tracker detector has been taking real data with cosmics at the Tracker Integration Facility (TIF) at CERN.

First DAQ checks and on-line monitoring tasks are executed at the Tracker Analysis Centre (TAC) which is a dedicated Control Room at TIF with limited computing resources. A set of software agents were developed to perform the real-time data conversion in a standard Event Data Model format, the copy of RAW data to CASTOR storage system at CERN and the registration of them in the official CMS bookkeeping systems.

According to the CMS computing and analysis model, most of the subsequent data processing has to be done in remote Tier-1 and Tier-2 sites, so data are automatically injected for the transfer from the TAC to the sites interested to analyze them, currently Fermilab, Bari and Pisa.

Official reconstruction in the distributed environment is triggered in real-time from Bari by using the ProdAgent tool, currently used with simulated data. Data are reprocessed with the most recent (pre-)releases of the official CMS software to provide immediate feedback to the software developers and the users. Automatic end-user analysis of published data is performed via CRAB tool to derive the distributions of the most important physics variables.

A monitoring system to check all the steps of the processing chain is also under development. An overview of the status of the tools developed is given, together with the evaluation of the real-time performance of the chain of tasks.

Summary:

The Tracker detector has been operated with cosmic events at the Tracker Integration Facility (TIF) at CERN. Data are checked via on-line data quality monitoring tools running at the Tracker Analysis Centre (TAC) which is a dedicated Control Room with limited computing resources.

Procedures are also developed and executed in real-time to make data officially available to the CMS community so raw data are firstly converted in a standard format, then archived on CASTOR storage system at CERN and registered in the official CMS data bookkeeping (DBS) and data location (DLS) systems.

The local storage available at TAC computers is sufficient to cache incoming data for about 10 days, and is clearly the best solution for fast-response analyses and DAQ checks. On the other side, a large community is expected to analyze data taken at the TAC, and this cannot happen at because of the limited resources.

Data are expected to flow from the TAC to the CMS Tier-1 and Tier-2 remote sites and to be accessed using standard CMS tools. Once data are registered in the DBS and DLS they are ready to be transferred in remote sites using the CMS official data movement tool PhEDEx. This operation requires that data are injected in the PhEDEx transfer management database to be routed to destination sites; a set of scripts are developed to perform this operation periodically in order to send data in the sites interested to analyze them, currently Fermilab, Bari and Pisa.

Official reconstruction in the distributed environment is automatic and triggered in real-time from a Bari machine by using a set of scripts optimized to run the ProdAgent tool, just used to reconstruct Monte Carlo simulated data. Data are reprocessed with the most recent releases and prereleases of the official CMS software to provide immediate feedback to the software developers and the users. A parallel reprocessing executed at FNAL with the last package patches is used to test the performances of the track reconstruction algorithms in real-time with read data.

Reconstruction, re-reconstruction, calibration and alignment tasks running at remote sites which requires to access data in the offline database located at CERN are run by using FroNTier software to access those data remotely.

Automatic end-user analysis of published data is performed via CRAB tool to derive

the distributions of the most important physics variables. A monitoring system to check all the steps of the processing chain is also under development. An overview of the status of the tools developed is given, together with the evaluation of the real-time performance of the chain of tasks.

Poster 2 - Board: 97 / 390

Recent Developments in LFC

Authors: Akos Frohner¹ ; David Smith¹ ; Jean-Philippe Baud¹ ; Krzysztof Nienartowicz¹ ; Lana Abadie¹ ; Remi Mollon¹ ; Sophie Lemaitre¹

¹ CERN

Corresponding Authors: sophie.lemaitre@cern.ch, akos.frohner@cern.ch

The LFC (LCG File Catalogue) allows retrieving and registering the location of physical replicas in the grid infrastructure given a LFN (Logical File Name) or a GUID (Grid Unique Identifier). Authentication is based on GSI (Grid Security Infrastructure) and authorization uses also VOMS.

The catalogue has been installed in more than 100 sites. It is essential to provide consistent and user-friendly tools to manage the catalogue. The LFC is based on a hierarchical namespace with a POSIX interface. The LFC API is similar to UNIX commands and includes functions to start/end a session or a transaction. The support of secondary groups and ACLs (Access Control Lists) allow a flexible management of file permissions.

An automated recovery strategy based on a retry mechanism offers a better reliability. Accessed very often by many tools, the catalogue needs to guarantee a fast response time. Performance issues have been studied and tuned by implementing bulk queries. Several other tests are being conducted such as the impact of the size of the communication buffer between the client and a local/remote LFC server on the response time.

Software components, tools and databases / 345

Recent Developments of the ROOT Mathematical and Statistical Software

Author: Lorenzo Moneta¹

¹ CERN

Corresponding Author: lorenzo.moneta@cern.ch

Advanced mathematical and statistical computational methods are required by the LHC experiments to analyze their data. These methods are provided by the Math work package of the ROOT project. We present an overview

of the recent developments of this work package by describing in detail the restructuring of the core mathematical library in a coherent set of new C++ classes and interfaces. We will describe how this new core

Math library has been integrated in the ROOT framework and it is used by the ROOT analysis objects. We will

present as well the achieved improvements, in terms of performances and quality, of numerical methods present in ROOT, such as random number generations, or matrix computations. Furthermore, we will review the new developments in the fitting and minimization packages, where new classes have been introduced to extend the previously existing functionality and to provide consistent interfaces to the users. We will present as well the recent and planned developments of integrating in the ROOT environment new advanced statistical tools required for the analysis of the LHC data.

Event processing / 244

Reconstruction and identification of Tau decays at CMS

Author: Giuseppe Bagliesi¹

¹ *INFN Sezione di Pisa*

Corresponding Author: giuseppe.bagliesi@cern.ch

Tau leptons play surely a key role in the physics studies at the LHC. Interests in using tau leptons include (but are not limited to) their ability to offer a relatively low background environment, a competitive way of probing new physics as well as the possibility to explore new physics regions not accessible otherwise. The Tau identification and reconstruction algorithms developed for the CMS experiment are described, from the first level of the trigger to the off-line reconstruction and selection.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

Simone Gennai, for the CMS Reconstruction group

Poster 2 - Board: 52 / 215

Reconstruction of Converted Photons in CMS

Author: Nancy Marinelli¹

¹ *University of Notre Dame*

Corresponding Author: nancy.marinelli@cern.ch

A seed/track finding algorithm has been developed for reconstruction of e^+e^- from converted photons. It combines the information of the electromagnetic calorimeter with the accurate information provided by the tracker. An Ecal seeded track finding is used to locate the approximate vertex of the conversion. Tracks found with this method are then used as input to further inside-out tracking aiming at reconstructing fully and very efficiently the conversion track pairs. Converted photon objects are finally built from the seeding ECAL energy deposit, a pair of opposite tracks and a conversion vertex, determined by fitting the two tracks to a common vertex. Results on reconstruction efficiency of

converted photons, as well as on fake rate when the photon did not convert, will be shown for single isolated photons and for photons from H→ $\gamma\gamma$ events with 2.10^{33} LHC luminosity.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS

Software components, tools and databases / 122

Relational databases for conditions data and event selection in ATLAS

Authors: Florbela Viegas¹ ; Gancho Dimitrov² ; Richard Hawkings¹

¹ CERN

² LBL / CERN

Corresponding Authors: florbela.tique.aires.viegas@cern.ch, richard.hawkings@cern.ch

The ATLAS experiment at LHC will make extensive use of relational databases in both online and offline contexts, running to O(TBytes) per year. Two of the most challenging applications in terms of data volume and access patterns are conditions data, making use of the LHC conditions database, COOL, and the TAG database, that stores summary event quantities allowing a rapid selection of interesting events. Both of these databases are being replicated to regional computing centres using Oracle Streams technology, in collaboration with the LCG 3D project. Database optimisation, performance tests and first user experience with these applications will be described, together with plans for first LHC data-taking and future prospects.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Poster 1 - Board: 111 / 459

Release Management - the EGEE project's approach

Author: Oliver Keeble¹

Co-authors: Joachim Flammer¹ ; Laurence Field¹ ; Markus Schulz¹

¹ CERN

Corresponding Authors: oliver.keeble@cern.ch, laurence.field@cern.ch

We describe an approach to maintaining a large integrated software distribution, the gLite middleware. We describe why we have moved away from the concept of regular releases of the entire distribution, favoring instead a multispeed approach where components can evolve at their own pace. An overview of our implementation of such a release process is given, explaining the full life cycle of updates as tracked from requirement capture through to deployment. In a broader context, the release strategies of comparable projects are assessed in an attempt

to isolate some useful principles. We conclude with some thoughts on the future direction of releases for gLite.

Poster 1 - Board: 21 / 437

Remote Management of nodes in the ATLAS Online Processing Farms

Authors: Marc Dobson¹ ; Usman Ahmad MALIK²

¹ CERN

² NCP, Quaid-E-Azam University

Corresponding Author: marc.dobson@cern.ch

The ATLAS experiment will use of order three thousand nodes for the online processing farms. The administration of such a large cluster is a challenge especially due to high impact of any down time. The ability to quickly and remotely turn on/off machines, especially following a power cut, and the ability to monitor the hardware health whether the machine be on or off are some of the major issues which the ATLAS SysAdmin Team faced. To solve these problems ATLAS has decided wherever possible to use Intelligent Platform Management Interfaces (IPMI) for its nodes.

This paper will present the mechanisms which were developed to allow the distribution of management and monitoring commands to the cluster machines in parallel. These commands were run simultaneously on the prototype farm and on the small scale final farm already purchased. The commands and their distribution take into account the specificities of the different IPMI versions and implementations, and the network topology of the ATLAS Online system.

Results from timing measurements for the distribution of commands to many nodes will be shown. These measurements will cover the times for booting and for shutting down of the nodes and will be extrapolated to the final cluster size.

Software components, tools and databases / 411

Replication and load Balancing strategy of STAR's RDBM

Authors: Jerome LAURET¹ ; Michael DePhillips¹ ; Mikhail KOPYTINE²

¹ BROOKHAVEN NATIONAL LABORATORY

² Kent State University, USA

Corresponding Authors: dephillips@bnl.gov, jlauret@bnl.gov

Database demands resulting from offline analysis and production of data at The STAR experiment at Brookhaven National Laboratory's Relativistic Heavy-Ion Collider has steadily increased over the last 6 years of data taking activities. With each year STAR more than doubles events taken with an anticipation of reaching a billion event capabilities as early as next year. The challenges faced from producing and analyzing this magnitude of events have raised issues with regard to distribution of calibrations and geometry data, via databases, to STAR's growing global collaboration. Rapid distribution, availability, ensured synchronization and load balancing have become paramount considerations. Both conventional technology and novel approaches are used in parallel to realize these goals. This paper discusses how STAR uses distribution methods via MySQL master slave replication to distribute its databases; the synchronization issues that arise from

this type of distribution and solutions, mostly homegrown, put forth to overcome these issues. Also discussed is a novel approach toward load balancing between slave nodes that assists in maintaining a high availability rate for a veracious community. This load balancing addresses both, pools of nodes internal to given location, as well as balancing the load for remote users between different available locations. Challenges, trade-offs, rationale for decisions and paths forward will be discussed in all cases, presenting a solid production environment with a vision for scalable growth.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

STAR

201

Resource sharing and deployment using virtualization at CERN openlab

Authors: Jose Miguel Dana Perez¹ ; Xavier Grehant¹

¹ CERN

Corresponding Authors: jose.dana@cern.ch, xavier.grehant@cern.ch

Today virtualization is used in computing centers to supply execution environments to a variety of users and applications. Appropriate flavours and configurations can be booted depending on the requirement, and in the same time the resources of a single server can be shared while preserving isolation between the environments.

In order to optimize distributed resource sharing, configuration and deployment we're building on two projects from HP Labs: SmartFrog and Tycoon.

We have developed SmartDomains based on SmartFrog, a framework for distributed configuration and deployment. Our additional components integrate Xen management capabilities into the framework. The system automatically deploys distributed pools of virtual machines according to descriptions in an Architecture Description Language, where additional components for dynamic re-configuration can be plugged in.

We're also working on Tycoon, which is another virtualization platform that optimizes resource usage.

Tycoon is a market-based system for trade of resources. The basic idea is to build dynamic clusters using virtualization and share their resources using an economic point of view. Tycoon deploys virtual machines on demand using a bid system where users can be both client and auctioneer.

Our mission at CERN openlab is to integrate Tycoon with EGEE, obtaining a dynamic Grid ready to share resources with others. Tycoon will run as a new service, deploying and destroying virtual machines (Worker Nodes and Computing Elements) on demand. Tycoon will give to EGEE a dynamic, flexible, secure and transparent platform/service to trade resources more efficiently.

Poster 2 - Board: 49 / 196

Rivet: a system for event generator validation and development

Author: Andy Buckley¹

¹ Durham University

The Rivet system is a framework for validation of Monte Carlo event generators against archived experimental data, and together with JetWeb and HepData forms a core element of the CEDAR event generator tuning programme. It is also an essential tool in the development of next generation event generators by members of the MCnet network. Written primarily in C++, Rivet provides a uniform interface for event generators and an analysis system to replicate experimental analyses.

The generator interface produces events via the HepMC event record and currently supports both Fortran and C++ versions of the Herwig and Pythia generators, as well as the C++ Sherpa generator and the Alpgen and Jimmy subprocess generators. More generators will be added to the package according to demand.

The Rivet analysis system is based on a concept of “event projections”, which project a simulated event into a lower-dimensional quantity such as scalar or tensor event shape variables. Projections can be nested and their results are automatically cached to eliminate duplicate computations. A set of standard projections and analyses which use them are included with the Rivet package, and this collection will grow with subsequent releases. Analysis data is accumulated using the AIDA interfaces, and exported primarily in the AIDA XML histogram format. To complement this, HepData-generated AIDA records for each bundled analysis are included in the Rivet package and can be used to define the binnings of generated data observables: this improves the robustness of analysis implementations and allows easy data-theory comparisons without requiring network access to HepData.

Plenary / 493

Role of Advanced Computation in the Design of the International Linear Collider.

Corresponding Author: quarkpt@slac.stanford.edu

The Global Design Effort for the International Linear Collider (ILC) has made use of modern computing capabilities in a number of areas: modeling the desired (accelerating) and undesired (wakefields, RF deflections) fields in the RF cavities, simulations of accelerator operations and tuning, prediction of accelerator uptime based on component performance and overall site design, and computer assisted design of accelerator components, beamlines, and regions. The use of advanced computing in these areas will be described, emphasizing both the new opportunities and the limitations present in the state of the art. Possible future developments will also be discussed.

Summary:

Peter Tenenbaum is an accelerator physicist at the Stanford Linear Accelerator Center (SLAC). He has worked on linear colliders for 16 years, beginning with the Stanford Linear Collider (SLC) and the Final Focus Test Beam (FFTB) experiment at SLAC. He has played a leading role in the development and use of simulation packages for the study of linear collider beam dynamics and beamline design.

Grid middleware and tools / 349

Role of Digital Forensics in Service Oriented Architectures

Author: Syed Naqvi¹

Co-author: Michel Riguidel²

¹ CoreGRID Network of Excellence

² TELECOM PARIS

Corresponding Author: snaqvi@ieee.org

Security requirements of service oriented architectures (SOA) are reasonably higher than the classical information technology (IT) architectures. Loose coupling – the inherent benefit of SOA – stipulates security as a service so as to circumvent tight binding of the services. The services integration interfaces are developed with minimal assumptions between the sending and receiving parties. This services aggregation approach is highly beneficial for achieving higher performance level; however, bookkeeping and logging of various events of such dynamic architecture are very complex issues. Security architecture requires these trails of events to establish clearly what happened and why in the post-accident scenario. The techniques employed to determine the reasons of security architecture's failure to prevent an incident are collectively known as Digital Forensics. It is necessary to develop digital forensics techniques for SOA so that necessary actions can be taken in the wake of a security breach. In this article we explore the role of digital forensics in SOA especially in the mission-critical SOA applications. We envision digital forensics as a sub-service of the security service in SOA. We propose the use of a monitoring service to generate these logs. We then present a mechanism of efficiently managing the logs of various actions based on the lifecycle of these logs. We finally conclude with the open issues and areas for further improvements.

Poster 2 - Board: 109 / 455

Running BaBar simulation on “public” Grid resources in Italy

Author: Gregory Dubois-Felsmann¹

¹ SLAC

Corresponding Author: gpdf@slac.stanford.edu

The BaBar experiment needs fast and efficient procedure for distributing jobs to produce a large amount of simulated events for analysis purpose. We discuss the benefits/drawbacks gained mapping the traditional production schema on the grid paradigm, and describe the structure implemented on the standard “public” resources of INFN-Grid project. Data access/distribution on sites involved using AMS and Xrootd servers is explained, comparing results obtained in Italy between the standard and the Grid approach.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

BaBar

Computer facilities, production grids and networking / 420

Running CE and SE in a Xen-virtualized environment

Author: Sergey Chechelnskiy¹

¹ Simon Fraser University

Corresponding Author: chech@sfu.ca

SFU is responsible for running two different clusters - one is designed for WestGrid internal jobs with its specific software and the other should run Atlas jobs only. In addition to different software configuration the Atlas cluster should have a diffener networking

configuration. We would also like to have a flexibility of running jobs on different hardware. That is why it has been decided to run two clusters in virtualized environments.

Extensive tests of running CE and SE in a Xen-virtualized environment have been performed. The following configuration has been selected for CE. The performance of Xen virtual machines has been found to be excellent.

Each WN is running two virtual machines - one is running internal WestGrid jobs and the other is running Atlas jobs. Moab is scheduling jobs for both clusters. We work closely with Cluster Resources on different moab configurations. We plan to implement memory management when Moab

changes the amount of assigned memory between two virtualized WNs running on the same hardware.

We plan to create an automated system of Xen images replication from one piece of hardware to another. An optimal Xen and Moab configuration for a heterogeneous clusters should be developed.

Different tests of running dcache SE in Xen environment have been performed.

Poster 1 - Board: 36 / 159

Runtime memory and data monitoring using Reflex for memory usage evaluation and data quality control.

Author: Charles Leggett¹

¹ LAWRENCE BERKELEY NATIONAL LABORATORY

Corresponding Author: charles.g.leggett@gmail.com

Runtime memory usage in experiments has grown enormously in recent years, especially in large experiments like Atlas. However, it is difficult to break down total memory usage as indicated by OS-level tools, to identify the precise users and abusers. Without a detailed knowledge of memory footprints, monitoring memory growth as an experiment evolves in order to control ballooning inflation, is ineffective.

We present a process that make use of Reflex to traverse the contents of all objects in the event store. As each object is encountered, its constituents are analysed using Reflex based introspection, to determine their true memory footprint. Pointers to other constituent objects are recursively followed, and their full sizes are added to the total. Runtime configuration allows specific object types to be ignored. Simultaneously, the values of all data members can be printed, using either their fundamental types, or an optional user supplied method. Address maps are kept to ensure that multiple counting does not occur. By using these procedures, the analysis and reconstruction software can be monitored in a historical fashion, with statistics produced to show which objects and object containers grow or shrink in size, and likewise can be monitored for data quality consistency and release validation.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS Summary:

Using C++ based object introspection to monitor full memory footprint of objects in an event store, for event size control, and dumping of data member contents for data quality and release validation.

Distributed data analysis and information management / 258**Scaling CMS data transfer system for LHC start-up****Author:** Lassi Tuura¹**Co-authors:** Brian Bockelman²; Daniele Bonacorsi³; Derek Feichtinger⁴; Jens Rehn⁵; Ricky Egeland⁶; Simon Metson⁷¹ *Northeastern University*² *University of Nebraska–Lincoln*³ *INFN-CNAF, Bologna, Italy*⁴ *PSI*⁵ *CERN*⁶ *University of Minnesota – Twin Cities*⁷ *University of Bristol***Corresponding Author:** lassi.tuura@cern.ch

The CMS experiment will need to sustain uninterrupted high reliability, high throughput and very diverse data transfer activities as the LHC operations start. PhEDEx, the CMS data transfer system, will be responsible for the full range of the transfer needs of the experiment. Covering the entire spectrum is a demanding task: from the critical high-throughput transfers between CERN and the Tier-1 centres, to high-scale production transfers among the Tier-1 and Tier-2 centres, to managing the 24/7 transfers among all the 170 institutions in CMS and to providing straightforward access to handful of files to individual physicists.

In order to produce the system with confirmed capability to meet the objectives, the PhEDEx data transfer system has undergone rigorous development and numerous demanding scale tests. We have sustained production transfers exceeding 1 PB/month for several months and have demonstrated core system capacity several orders of magnitude above expected LHC levels.

We describe the level of scalability reached, and how we got there, with focus on the main insights into developing a robust, lock-free and scalable distributed database application, the validation stress test methods we have used, and the development and testing tools we found practically useful.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS

Software components, tools and databases / 223**Searching for CMS data: Rapid web development using python and AJAX****Author:** Valentin Kuznetsov¹**Co-author:** Andrew Dolgert¹¹ *Cornell University*

Corresponding Author: valentin.kuznetsov@cern.ch

We discuss the rapid development of a large scale data discovery service for the CMS experiment using modern AJAX techniques and the Python language. To implement a flexible interface capable of accommodating several different versions of the DBS database, we used a “stack” approach. Asynchronous JavaScript and XML (AJAX) together with an SQL abstraction layer, template engine, code generation tool and dynamic queries provide powerful tools for constructing interactive interfaces to large amounts of data. We show how the use of these tools, with rapid development in a modern scripting language, improved the scalability and usability of the search interface for different user communities.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS

Computer facilities, production grids and networking / 41

Security Incidents management in a Grid environment

Author: Romain Wartel¹

¹ CERN

Corresponding Authors: markus.schulz@cern.ch, romain.wartel@cern.ch

Today’s production Grids connect large numbers of distributed hosts using high throughput networks and hence are valuable targets for attackers. In the same way users transparently access any Grid service independently of its location, an attacker may attempt to propagate an attack to different sites that are part of a Grid. In order to contain and resolve the incident, and since such an attack may rapidly span many different administrative domains, efficient sharing of appropriate information between sites is important. Appointing an incident coordinator to obtain, correlate, filter and redistribute relevant information needed by sites in their local investigations has proven to be very effective to rapidly process the massive amount of data involved. Improving the trust between the site security teams, which may have different security standards, is an important factor in obtaining more relevant and accurate information. However, wider distribution increases the risk of voluntarily or involuntarily leaks of sensitive information outside of the community. Such leaks are not only dangerous because they may expose sensitive information, but also because they may discourage other sites from sharing their findings in the future. As a result, an essential part of the incident response relies on processes implementing appropriate and timely management and control of the information flow. This document describes the model adopted by the EGEE infrastructure, as well as issues encountered.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

EGEE Operational Security Coordination Team

Poster 1 - Board: 71 / 91

Service Level Status - a new real-time status display for IT services

Author: Sebastian Lopienski¹

¹ CERN

Corresponding Author: sebastian.lopienski@cern.ch

Nowadays, IT departments provide, and people use, computing services of an increasingly heterogeneous nature. There is thus a growing need for a status display that groups these different services and reports status and availability in a uniform way. The Service Level Status (SLS) system addresses these needs by providing a web-based display that dynamically shows availability, basic information and statistics about various IT services, as well as the dependencies between them.

This paper first introduces the requirements SLS had to meet, and the main concepts behind it, like service availability and status, Key Performance Indicators (KPIs), sub/meta-services, and service dependencies. It then describes the SLS system architecture, and some interesting implementation details, such as the usage of XML Schemas. Since clear visualization of service availability and status is one of the main goals of SLS, emphasis is put on describing the intuitive web-based user interface.

Poster 2 - Board: 8 / 408

Shaping Collaboration 2006: Action Items for the LHC

Authors: Homer A Neal¹ ; Jeremy Herr¹ ; Steven Goldfarb¹

¹ *University of Michigan*

Corresponding Author: steven.goldfarb@cern.ch

“Shaping Collaboration 2006” was a workshop held in Geneva, on December 11-13, 2006, to examine the status and future of collaborative tool technology and its usage for large global scientific collaborations, such as those of the CERN LHC (Large Hadron Collider). The workshop brought together some of the leading experts in the field of collaborative tools (WACE 2006) with physicists and developers of the LHC collaborations and HEP. We highlight important presentations and key discussions held during the workshop, then focus on a large and aggressive set of goals and specific action items targeted at institutes from all levels of the LHC organization. This list of action items, assembled during a panel discussion at the close of the LHC sessions, includes recommendations for the LHC Users, their Universities, Project Managers, Spokespersons, National Funding Agencies and Host Laboratories. We will present this list, along with suggestions for priorities in addressing the immediate and long-term leads of HEP.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

LHC (ATLAS) Summary:

The need for a secure, robust, and high-quality collaborative environment for the LHC has been known for some time. With collaboration membership typically counting in the thousands, and spread over hundreds of institutes located around the globe, active participation in the experiments will push the limits of our current tools, even if there were adequate funding for development, installation and maintenance. The lack of dedicated resources over the past several years leading up to the LHC start-up, however, has left the LHC collaborations in a difficult position to catch up. Work is needed in all aspects of the field, from the proper equipping of conference rooms to the creation of robust video and audio systems, to the development of integrated environments that will

satisfy

the needs of the experiments for security and functionality. The Shaping Collaboration 2006 conference addressed many of these issues and provided an excellent forum for charting a course to bridge the gap in the near and long-term future. Our presentation attempts to summarize this important outcome and to spark further action within the community.

Poster 2 - Board: 69 / 73

Sharing LCG files across different platforms

Author: Yaodong Cheng¹

Co-author: Aigui Liu²

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

² *IHEP, China*

Corresponding Author: chyd@ihep.ac.cn

Currently more and more heterogeneous resources are integrated into LCG. Sharing LCG files across different platforms, including different OS and grid middlewares, is a basic issue. We implemented web service interface for LFC and simulated LCG file access client by using globus Java CoG Kit.

Event processing / 403

Simulation and event reconstruction inside the PandaRoot frameworks.

Author: Stefano Spataro¹

¹ *II Physikalisches Institut, Universität Giessen (Germany)*

Corresponding Author: stefano.spataro@exp2.physik.uni-giessen.de

The PANDA detector will be located at the future GSI accelerator FAIR. Its primary objective is the investigation of strong interaction with anti-proton beams, in the range up to 15 GeV/c as momentum of the incoming anti-proton. The PANDA offline simulation framework is called "PandaRoot", as it is based upon the ROOT 5.12 package. It is characterized by a high versatility; it allows to perform simulation and analysis, to run different event generators (EvtGen, Pluto, UrQmd), different transport models (Geant3, Geant4, Fluka) with the same code, thus to compare the results simply by changing few macro lines without recompiling at all.

Moreover auto-configuration scripts allow installing the full framework easily in different Linux distributions and with different compilers (the framework was installed and tested in more than 10 Linux platforms) without further manipulation. The final data are in a tree format, easily accessible and readable through simple clicks on the root browsers.

The presentation will report on the actual status of the computing development inside the PandaRoot framework, in terms of detector implementation and event reconstruction.

Summary:

The PandaRoot framework is part of the FairRoot project, a common framework for the future FAIR experiments, such as CBM, Panda and HADES upgrade. The talk will explain the general structure of the framework, the several event generators implemented, the possibility to use Geant3 and Geant4 at the same time, thus to compare results (which are substantially different in the case of EM showers), the implemented detectors via the geometry visualization tool, and the reconstruction status (cluster finding in the em calorimeter, pattern recognition and global tracking by using a Kalman filter and the GEANE track follower). **Submitted on behalf of Collaboration (ex, BaBar, ATLAS):**

Panda

Event processing / 26

Simulation readiness for the first data at LHC

Author: Adele Rimoldi¹

¹ *Pavia University & INFN*

Corresponding Author: adele.rimoldi@cern.ch

The ATLAS detector is entering the final phases of construction and commissioning in order to be ready to take data during the first LHC commissioning run, foreseen by the end of 2007. A good understanding of the experiment performance from the beginning is essential to efficiently debug the detector and assess its physics potential in view of the physics runs which are going to take place from 2008 on. The ATLAS Detector Simulation programs have been developed since the ATLAS inception and have been developed for easing the detector optimization and construction: further developments to the simulation suite have recently been introduced to cope with essential factors like misalignment, inefficiencies, imperfections but still maintaining a high level of efficiency and operability to serve the ongoing production exercises. Emphasis in this talk is put on recent developments and new features, on validation and production strategies as well as on performance figures, robustness and maintainability

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Event processing / 374

Software for CMS Reconstruction

Authors: Shahram Rahatlou¹ ; Tommaso Boccali²

¹ *Univ di Roma La Sapienza*

² *INFN Sezione di Pisa*

Corresponding Authors: tommaso.boccali@cern.ch, shahram.rahatlou@cern.ch

At the end of 2007 the first colliding beams from LHC are expected. The CMS Computing model enforces the use of the same software (with different performance settings) for offline and online(HLT) operations; this is particularly true for the reconstruction software: the different settings must allow a processing time per event (typically, numbers for 2×10^{33} luminosity are given) of 50 ms at HLT, while 25 sec are allowed for the offline reconstruction. During 2006 CSA06 data challenge the

focus has been put on the offline reprocessing. The reconstruction software has substantially improved from the end-of-2005 version, which was able to process only local reconstruction, to a full fledged reconstruction program, with high level objects ready for analysis tasks (electrons, jets, muons, b/tau tagged jets). The same software is also ready to process a non-ideal detector, and takes into account hardware inefficiencies and misalignments. This second mode of operation, which is important for the readiness of data taking, has been tested in CMS slice tests and commissioning tasks, and has shown that the algorithms used for MonteCarlo processing are well suited to real-world tasks. During 2007, a second data challenge will explore online/offline reconstruction, and will be used as the base for the ready-for-beam demonstration.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS Summary:

The CMS Reconstruction software is described, pointing out current performance and future developments towards data taking.

Poster 2 - Board: 58 / 301

Software for the Commissioning of the CMS Micro Strip Silicon Tracker

Author: Dorian Kcira¹

¹ *University of Louvain*

Corresponding Author: dorian.kcira@cern.ch

With a total area of more than 200 square meters and about 16000 silicon detectors the Tracker of the CMS experiment will be the largest silicon detector ever built. The CMS silicon Tracker will detect charged tracks and will play a determinant role in lepton reconstruction and heavy flavour quark tagging.

A general overview of the Tracker data handling software, which allows the detector to be configured, calibrated, monitored and its data to be analysed by means of distributed computing resources and databases is given. Results of the Tracker performance as obtained in the various setups where the Tracker is being commissioned are also presented. A first functional version of the Tracker data handling software was tested in the CMS Magnet Test Cosmic Challenge (MTCC). At the MTCC a small fraction of all the CMS subdetectors was operated in the 4T solenoid of the experiment. Cosmic rays were triggered by the muon detectors and all subdetectors were readout with the global data acquisition system. The MTCC Tracker setup represented about 1% of the final system.

A much larger fraction of the Tracker is currently being readout at the Tracker Integration Facility (TIF) where large parts of the full Silicon Strip Detector are already integrated. The data handling software in use at the TIF is close to final.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS [Tracker] Collaboration Summary:

A general overview of the Tracker data handling software, which allows the detector to be configured, calibrated, monitored and its data to be analysed by means of distributed computing resources and databases is given. Results of the Tracker performance as obtained in the various setups where the Tracker is being commissioned are also presented.

Poster 2 - Board: 42 / 127

Software representation of the ATLAS magnetic field

Authors: Paul Miyagawa¹ ; Salvador Marti I Garcia²

Co-author: Jochen Schieck³

¹ *University of Manchester*

² *Instituto de Fisica Corpuscular (IFIC) UV-CSIC*

³ *Max-Planck-Institut fuer Physik Muenchen*

Corresponding Author: salvador.marti@ific.uv.es

The ATLAS solenoid produces a magnetic field which enables the Inner Detector to measure track momentum by track curvature. This solenoidal magnetic field was measured using a rotating-arm mapping machine and, after removing mapping machine effects, has been understood to the 0.05% level. As tracking algorithms require the field strength at many different points, the representation of this magnetic field in Athena (the ATLAS offline software framework) can have a significant impact on the processing time for these algorithms. We review the field models and mathematical techniques used to decouple machine effects from real field features and evaluate the performance of different field representations within Athena.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Poster 1 - Board: 25 / 57

StatPatternRecognition: A C++ Package for Multivariate Classification

Author: Ilya Narsky¹

¹ *California Institute of Technology*

Corresponding Author: narsky@hep.caltech.edu

SPR implements various tools for supervised learning such as boosting (3 flavors), bagging, random forest, neural networks, decision trees, bump hunter (PRIM), multi-class learner, logistic regression, linear and quadratic discriminant analysis, and others. Presented at CHEP 2006, SPR has been extended with several important features since then. The package has been stripped of CLHEP dependency, equipped with autotools and posted at Sourceforge for distribution under general public license: <http://sourceforge.net/projects/statpatrec/> . It is now a standalone package with an optional dependency on Root for data input/output. Several new methods have been included in the package. SPR is now capable of boosting and bagging an arbitrary sequence of included classifiers allowing the user to explore a broad range of classifier combinations. This talk is meant to summarize recent updates to the package and review recent applications of the package to physics analysis. More info on the project is available from <http://www.hep.caltech.edu/~narsky/spr.html> .

Poster 2 - Board: 107 / 450

Steering of GRID production in ATLAS experiment

Author: Pavel Nevski¹

Co-authors: Alexei Klimentov¹ ; Junich Tanaka²

¹ *Brookhaven National Laboratory (BNL)*

² *University of Tokyo*

Corresponding Author: pavel.nevski@cern.ch

In order to be ready for the physics analysis ATLAS experiment is running a world wide Monte Carlo production for many different physics samples with different detector conditions.

Job definition is the starting point of ATLAS production system. This is a common interface for the ATLAS community to submit jobs for processing by the Distributed production system used for all ATLAS-wide operations. It consists of a Task Request interface and Job submission service.

Task Request interface is realised by an Apache server and a Task request Database. It provides user autotification, on-line validation of the request consistency and the request registration. Job submission service makes sure that the input data and requested software are available on GRID, supplies completed job definitions to the Production system database and monitors the execution progress

We will present the experience of more then a year system running with many users submitting ATLAS jobs all over the world

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Grid middleware and tools / 78

Storage Resource Manager version 2.2: design, implementation, and testing experience

Author: Flavia Donno¹

Co-authors: Alex Sim² ; Arie Shoshani² ; Birger Koblitz¹ ; David Smith¹ ; Don Petravick³ ; Ezio Corso⁴ ; Gavin Mccance¹ ; Jean-Philippe Baud¹ ; Junmin Gu² ; Luca Magnoni⁵ ; Maarten Litmaath¹ ; Paolo Badino¹ ; Paolo Tedesco¹ ; Remi Mollon¹ ; Riccardo Zappi⁵ ; Shaun De Witt⁶ ; Sophie Lemaitre¹ ; Timur Perelmutov³ ; Vijaya Natarajan²

¹ *CERN*

² *LBNL*

³ *FNAL*

⁴ *ICTP*

⁵ *CNAF*

⁶ *RAL*

Corresponding Author: flavia.donno@cern.ch

Storage Services are crucial components of the Worldwide LHC Computing Grid (WLCG) infrastructure spanning more than 200 sites and serving computing and storage resources to the High Energy Physics LHC communities. Up to tens of Petabytes of data are collected every year by the 4 LHC experiments at CERN. To process these large data volumes it is important to establish a protocol and a very efficient interface to the various storage solutions adopted by the WLCG sites. In this work we report on

the experience acquired during the definition of the Storage Resource Manager v2.2 protocol. In particular, we focus on the study performed to enhance the interface and make it suitable for use by the WLCG communities. At the moment 5 different storage solutions implement the SRM 2.2 interface: BeStMan (LBNL), CASTOR (CERN, RAL, and INFN), dCache (DESY and FNAL), DPM (CERN), and StoRM (INFN and ICTP). After a detailed inside review of the protocol, various test suites have been written identifying the most effective set of tests: the S2 test suite from CERN and the SRM-Tester test suite from LBNL. Such test suites have helped verifying the consistency and coherence of the proposed protocol and validating existing implementations. We conclude our work describing the results achieved.

Summary:

SRM 2.2 definition and testing experience

Poster 2 - Board: 31 / 429

Storage Technology Evaluations at BNL

Author: Robert Petkus¹**Co-authors:** Jason Smith¹ ; Ofer Rind¹¹ *Brookhaven National Laboratory***Corresponding Author:** rpetkus@bnl.gov

The RHIC/USATLAS Computing Facility at BNL has evaluated high-performance, low-cost storage solutions in order to complement a substantial distributed file system deployment of dCache (>400 TB) and xrootd (>130 TB). Currently, these file systems are spread across disk-heavy computational nodes providing over 1.3 PB of aggregate local storage. While this model has proven sufficient to date, the projected demand for disk storage over the next five years is expected to cap 30 petabytes. At this scale, more cluster nodes may be needed to meet storage requirements than necessary for computation alone, introducing potential architectural inefficiencies. Candidate solutions include extending the existing paradigm (more and larger disks per compute node), augmentation with a cache of file servers, or consolidation of distributed storage onto dedicated file servers. The product of our evaluation process will strive to address these concerns and provide a clear path for the future.

To this end, a spectrum of storage solutions was subjected to both synthetic and production benchmarking in an effort to discern the maximum I/O capacity of each system. The goal was to identify the highest achieving and most versatile storage systems, compare and contrast technology (e.g., RAID implementation, SATA vs. SAS disks, Solaris/ZFS vs. Linux/ext3), and consider cost, ease of management and scalability. Since data center real estate and power concerns are paramount, additional consideration was given to ultra high-density storage systems. Test methodology, results, and analysis are provided.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

RHIC/USATLAS

Computer facilities, production grids and networking / 395

Storage management solutions and performance tests at INFN Tier-1

Authors: Alessandro Italiano¹ ; Andrea Chierici¹ ; Angelo Carbone¹ ; Armando Fella¹ ; Cristina Vistoli¹ ; Davide Salomoni¹ ; Dejan Vitlacil¹ ; Donato Degirolamo¹ ; Fabrizio Furano² ; Felice Rosso¹ ; Giacinto Donvito³ ; Giuseppe Lo Re¹ ; Luca dell’Agnello¹ ; Marco Bencivenni¹ ; Massimo Donatelli¹ ; Pier Paolo Ricci¹ ; Riccardo Veraldi¹ ; Stefano Zani¹ ; Vladimir Sapunenko¹

¹ *INFN-CNAF*

² *INFN-PADOVA*

³ *INFN-BARI*

Performance, reliability and scalability in data access are key issues when considered in the context of HEP data processing and analysis applications. The importance of these topics is even larger when considering the quantity of data and the request load that a LHC data centers has to support.

In this paper we give the results and the technical details of a large scale validation, performance and comparison tests performed at CNAF.

The storage management solutions CASTOR, gpfs, xrootd and dcache have been tested on the CNAF production environment.

Our storage solution is based on Fibre Channel systems organized in a Storage Area Network where disk servers are interconnected to the farm via gigabit LAN: for these tests 24 disk servers (for a total of 220 TB of disk space) and about 260 worker nodes have been used.

The test aim was to evaluate both the sequential and random (reading and writing) access to the data in order to verify efficiency, availability and robustness of the different storage solutions.

Computer facilities, production grids and networking / 153

Streamlining and Scaling Castor2 Operations

Authors: Alasdair EARL¹ ; Ignacio REGUERO¹ ; Jan van ELDIK¹ ; Miguel Coelho Dos Santos¹

¹ *CERN*

Corresponding Authors: jan.van.eldik@cern.ch, miguel.coelho.santos@cern.ch

This paper presents work, both completed and planned, for streamlining the deployment, operation and re-tasking of Castor2 instances. We present a summary of what has recently been done to reduce the human intervention necessary for bringing systems into operation; including the automation of Grid host certificate requests and deployment in conjunction with the CERN Trusted CA and automated configuration using Quattor.

We provide an overview of the software developed for monitoring operations so that various types of problem are quickly identified and remedied. Many of these tasks have been automated in a portable manner so that they can be used by other sites running Castor.

To aid in taking disk servers out of production, for hardware interventions or to retask the machine to another instance, we present the development of a program which can take machines out of production while ensuring that data is reliably replicated.

Summary:

Castor, castorReconcile, disknodeShutdown, operations management

Plenary / 471

Summary of Collaborative Tools and Initiatives

Plenary / 473

Summary of Distributed data analysis and information management track

Corresponding Author: roger.jones@cern.ch

Plenary / 472

Summary of Online Computing Track

Plenary / 29

Summary of WLCG Collaboration Workshop 1-2 September 2007

Author: Jamie Shiers¹

¹ CERN

Corresponding Author: jamie.shiers@cern.ch

This talk summarises the main discussions and issues raised at the WLCG Collaboration workshop held immediately prior to CHEP.

The workshop itself will focus on service needs for initial data taking: commissioning, calibration and alignment, early physics.

Target audience: all active sites plus experiments

We start with a detailed update on the schedule and operation of the accelerator for 2007/2008, followed by similar sessions from each experiment.

The main thrust of this workshop will be to understand the status of the WLCG Services with respect to the 2007 / 2008 requirements, in particular the 'residual services' discussed at the January 2007 workshop.

- * SRM 2.2
- * Full FTS services
- * Distributed Database Services
- * gLite 3.x / SL(C)4 / 64bit
- * VOMS roles

We also review the status / plans of the Dress Rehearsals and initial running.

BOF sessions are foreseen in the following areas:

- * Monitoring
- * Data Management
- * Databases
- * Operations, including experiment / site / Grid operations

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

WLCG Summary:

A summary of the plenary and parallel (BOF) sessions held at the WLCG Collaboration workshop held immediately prior to CHEP:

<http://indico.cern.ch/conferenceDisplay.py?confId=3578>

Plenary / 476

Summary of computing facilities, production grids and networking track

Plenary / 475

Summary of event processing track

Corresponding Author: mcbride@fnal.gov

Plenary / 477

Summary of grid middleware and tools track

Plenary / 474

Summary of software components, tools and databases track

470

Summary of the WLCG Meeting

Corresponding Author: jamie.shiers@cern.ch

Poster 1 - Board: 44 / 274

Survey of the main HEP software components (1970 → 2010)

Author: Rene Brun¹

¹ CERN

Corresponding Author: rene.brun@cern.ch

A poster (two A0 pages) shows the main software systems used in HEP in the period 1970 -> 2010 from their conception to their death. Graphics bands are used to indicate the relative importance of each system or tool in the following categories:

- Machines and Operating systems
- Storage systems and access libraries
- Networking and communication software
- Compiled languages
- Code management systems
- Data structures management and I/O systems. Data bases
- Graphical User Interface and Graphics
- Histograms, Math and Statistics
- Scripting systems and Interpreters
- Interactive Analysis systems
- Detector Geometry and Simulation

Event processing / 49

TMVA - Toolkit for Multivariate Data Analysis

Authors: Andreas Hoecker¹ ; Fedrik Tegenfeldt² ; Helge Voss³ ; Jörg Stelzer⁴

¹ CERN, Switzerland

² Iowa University, USA

³ MPI for Nuclear Physics, Heidelberg, Germany

⁴ CERN

Corresponding Author: joerg.stelzer@cern.ch

In high-energy physics, with the search for ever smaller signals in ever larger data sets, it has become essential to extract a maximum of the available information from the data. Multivariate classification methods based on machine learning techniques have become a fundamental ingredient to most analyses. Also the multivariate classifiers themselves have significantly evolved in recent years. Statisticians have found new ways to tune and to combine classifiers to further gain in performance. Integrated into the analysis framework ROOT, TMVA is a toolkit which holds a large variety of multivariate classification algorithms. They range from rectangular cut optimization using a genetic algorithm and from likelihood estimators over the linear Fisher discriminant and non-linear neural networks, to sophisticated methods like support vector machines, boosted decision trees and rule ensemble fitting that was recently developed. TMVA manages the simultaneous training, testing, and performance evaluation of all these classifiers with a user-friendly interface, and expedites the application of the trained classifiers to data.

Poster 2 - Board: 60 / 324

Testing TMVA software in b-tagging for the search of MSSM Higgs bosons at the LHC

Authors: Aatos Heikkinen¹ ; Tapio Lampen¹ ; Tomas Linden¹

Co-authors: Francisco Garcia ¹; Lauri Wendland ¹; Matti Kortelainen ¹; Pekka Kaitaniemi ¹; Sami Lehti ¹; Veikko Karimaki ¹

¹ *Helsinki Institute of Physics HIP*

Corresponding Authors: tapio.lampen@cern.ch, aatos.heikkinen@cern.ch, tomas.linden@cern.ch

We demonstrate the use of a ROOT Toolkit for Multivariate Data Analysis (TMVA) in tagging b-jets associated with heavy neutral MSSM Higgs bosons at the LHC.

The associated b-jets can be used to extract Higgs events from the Drell-Yan background, for which the associated jets are mainly light quark and gluon jets.

TMVA provides an evaluation for different multivariate classification techniques wrapped in a ROOT-integrated machine learning environment. Background discriminating power is demonstrated for various methods available in TMVA, such as rectangular cut optimisation, projective and multi-dimensional likelihood estimators, linear discriminant analysis with H-Matrix and Fisher discriminants, artificial neural networks and boosted/bagged decision trees.

The effect of choice of variables and variable transformation is described. TMVA working in transparent factory mode guarantees an unbiased performance comparison, since all classifiers are evaluated with the same training and test data. Finally, results are compared against previous studies using neural networks and standard methodology where associated b-jets can be identified using lifetime based tagging algorithms, which rely on displaced secondary vertices and track impact parameters.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS

Poster 2 - Board: 73 / 145

Testing and integrating the WLCG/EGEE middleware in the LHC computing

Authors: Alessandro Di Girolamo¹; Andrea Sciabà¹; Elisa Lanciotti¹; Enzo Miccio¹; Nicolò Magini¹; Patricia Méndez Lorenzo¹; Roberto Santinelli¹; Simone Campana¹

¹ *CERN*

Corresponding Author: andrea.sciaba@cern.ch

The main goal of the Experiment Integration and Support (EIS) team in WLCG is to help the LHC experiments with using proficiently the gLite middleware as part of their computing framework. This contribution gives an overview of the activities of the EIS team, and focuses on a few of them particularly important for the experiments. One activity is the evaluation of the gLite workload management system (WMS) to assess its adequacy for the needs of the LHC computing in terms of functionality, reliability and scalability. We describe in detail how the experiment requirements can be mapped to validation criteria, and the WMS performances are accurately measured under realistic load conditions over prolonged periods of time. Another activity is the integration of the Service Availability Monitoring system (SAM) with the experiment monitoring framework. The SAM system is widely used in the EGEE operations to identify malfunctions in Grid services, but it can be adapted to perform the same function on experiment-specific services. We describe how this has been done for some LHC experiments, which are now using SAM as part of their operations.

Poster 1 - Board: 62 / 460

Testing gLite for releases

Author: Andreas Unterkircher¹

¹ CERN

Corresponding Author: andreas.unterkircher@cern.ch

We describe the methodology for testing gLite releases. Starting from the needs given by the EGEE software management process we illustrate our design choices for testing gLite.

For certifying patches different test scenarios have to be considered: regular regression tests, stress tests and manual verification of bug fixes. Conflicts arise if these tests are all carried out at the same time on the same infrastructure. Thus virtualisation is used and its benefits are shown by several examples. Furthermore we sketch the architecture of our distributed testbed including lessons learnt from such a distributed test environment. Finally we describe the test framework we're using. We also give an overview of the tests that have been developed to test the different gLite services. Some statistics on the patches certified with this process are also presented. Apart from these more conventional testing activities we describe how we address testing maturing complex services for scalability and long term stability which require excessive testbed sizes.

Poster 2 - Board: 57 / 283

Testing suite for validation of Geant4 hadronic generators

Author: Vladimir Ivantchenko¹

¹ CERN, ESA

Corresponding Author: vladimir.ivantchenko@cern.ch

The testing suite for validation of Geant4 hadronic generators with the data of thin target experiments is presented. The results of comparisons with the neutron and pion production data of are shown for different Geant4 hadronic generators for the beam momentum interval 0.5 – 12.9 GeV/c.

Summary:

The developed software of hadronic testing suite allows performing validation of the Geant4 hadronic generators in semiautomatic regime. The afs file system is used to archive results of simulation for different versions of Geant4. The software is used for model development and for final analysis of the simulation precision. It allows building an optimal combination of Geant4 cascade generators and high energy generators for LHC experiments simulation.

Software components, tools and databases / 446

The ALICE Offline Environment

Author: Federico Carminati¹

¹ CERN

Corresponding Author: federico.carminati@cern.ch

Since 1998 the ALICE Offline Project has developed an integrated offline framework (AliRoot) and a distributed computing environment (AliEn) to process the data of the ALICE experiment. These systems are integrated with the LCG computing infrastructure, and in particular with the ROOT system and with the WLCG Grid middleware, but they also present a number of original solutions, which have been developed by the ALICE Offline Project to face the challenges specific to the ALICE, experiment. This talk will review the development and current status of the ALICE Offline. The presentation will describe how this environment has been tested during a series of exercises of increasing complexity carried on over the years. The status of readiness of the systems will be described, as well as the major challenges facing it at the eve of data taking. The lessons learned during this nine-year development will be described and analysed, and the development roadmap will be presented and discussed.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ALICE

Online computing / 39

The ALICE-LHC Online Data Quality Monitoring Framework

Author: Filimon Roukoutakis¹

Co-author: Sylvain Chapeland²

¹ CERN & University of Athens

² CERN

Corresponding Author: filimon.roukoutakis@cern.ch

ALICE is one of the experiments under installation at CERN Large Hadron Collider, dedicated to the study of Heavy-Ion Collisions. The final ALICE Data Acquisition system has been installed and is being used for the testing and commissioning of detectors. Data Quality Monitoring (DQM) is an important aspect of the online procedures for a HEP experiment. In this presentation we overview the architecture, implementation and usage experience of ALICE's AMORE (Automatic MOnitoring Environment), a distributed application aimed to collect, analyze, visualize and store monitoring data in a large, experiment wide scale. AMORE is executed interfaced to the DAQ software framework (DATE) and follows the publish-subscribe paradigm where a large number of batch processes execute detector-specific analysis on raw data samples and publish monitoring results on specialized servers. Clients connected to these servers have the ability to correlate, further analyze and visualize the monitoring data. Provision is taken to archive the most important results so that historic plots can be produced.

Poster 1 - Board: 85 / 230

The ATLAS Canada Network

Author: Ian Gable¹

¹ *University of Victoria*

Corresponding Author: igable@uvic.ca

The ATLAS Canada computing model consists of a Tier-1 computing centre located at the TRIUMF Laboratory in Vancouver, Canada, and two distributed Tier-2 computing centres: one in Eastern Canada and one in Western Canada. Each distributed Tier-2 computing centre is made up of a group of universities. To meet the network requirements of each institution, HEPnet Canada and CANARIE (Canada's National Research Network), in collaboration with other research networks, have constructed a series of lightpaths to connect the institutions involved. TRIUMF is connected to CERN via a 10G circuit, a portion of which will be dropped at TRIUMF's Tier-1 peer SARA/NIKEF in Amsterdam. Current Canadian ATLAS Tier-2 institutions are connected to TRIUMF via 1G lightpaths, and routing between Tier-2s occurs through TRIUMF. We discuss the architecture of the ATLAS Canada network, challenges of building the network, and future plans for the network.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

On behalf of HEPnet Canada Collaboration

Distributed data analysis and information management / 200

The ATLAS Computing Model

Author: Roger Jones¹

¹ *LANCAS*

Corresponding Author: roger.jones@cern.ch

The ATLAS Computing Model was constructed after early tests and was captured in the ATLAS Computing TDR in June 2005. Since then, the grid tools and services have evolved and their performance is starting to be understood through large-scale exercises. As real data taking becomes immanent, the computing model continues to evolve, with robustness and reliability being the watchwords for the early deployment. Particular areas of active development are the data placement and data access, and the interaction between the TAGs, the datasets and the Distributed Data Management issues. The earlier high-level policies and models are now being refined into lower level instantiations.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS Summary:

The ATLAS Computing Model is presented, including recent updates from Computing System Commissioning experience.

Online computing / 332

The ATLAS DAQ System Online Configurations Database Service Challenge

Author: Igor Soloviev¹

Co-authors: Andrei Kazarov² ; Giovanna Lehmann Miotto³ ; Joao Almeida Simoes⁴ ; John Erik Sloper⁵ ; Marc Dobson³ ; Rodrigo Coura Torres⁶

¹ CERN/PNPI

² PNPI

³ CERN

⁴ Universidade de Lisboa

⁵ University of Warwick

⁶ UFRJ

Corresponding Author: igor.soloviev@cern.ch

This paper describes challenging requirements on the configuration service. It presents the status of the implementation and testing one year before the start of the ATLAS experiment at CERN providing details of:

- capabilities of underlying OKS* object manager to store and to archive configuration descriptions, it's user and programming interfaces;
- the organization of configuration descriptions for different types of data taking runs and combinations or participating sub-detectors;
- the scalable architecture to support simultaneous access to the service by thousands of processes during the online configuration stage of ATLAS;
- the results of large scale tests performed on the configuration service and experience of it's usage during test beam and technical runs.

The paper also presents pro and cons of the chosen object-oriented implementation comparing with solutions based on pure relational database technologies, and explains why after several years of usage we continue with it.

- "The OKS in-memory persistent object manager", R. Jones, L. Mapelli, Y. Ryabov and I. Soloviev, RT 1997; IEEE Transactions on Nuclear Science, Volume 45, Issue 4, Part 1, Aug. 1998 Page(s):1958-1964

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Online computing / 143

The ATLAS High Level Trigger Steering

Author: Simon George¹

Co-authors: Andreas Hoecker² ; Carlo Schiavi³ ; Christiane Risler⁴ ; Gordon Fischer⁵ ; Jiri Masik⁶ ; Joerg Stelzer² ; Johannes Haller⁷ ; Martin zur Nedden⁵ ; Nicolas Berger² ; Till Eifert⁸ ; Tomasz Bold⁹ ; Valeria Perez Reale² ; Xin Wu⁸

¹ Royal Holloway

² CERN

³ INFN, Sezione di Genova

⁴ Humboldt-Universitaet Berlin

⁵ Humboldt-Universitaet zu Berlin

⁶ University of Manchester

⁷ Universitaet Hamburg

⁸ Universite de Geneve

⁹ University of California, Irvine

Corresponding Author: s.george@rhul.ac.uk

The High Level Trigger (HLT) of the ATLAS experiment at the Large Hadron Collider receives events which pass the LVL1 trigger at ~75 kHz and has to reduce the rate to ~200 Hz while retaining the most interesting physics. It is a software trigger and performs the reduction in two stages: the LVL2 trigger should take ~10 ms and the Event Filter (EF) ~1 s.

At the heart of the HLT is the Steering software. To minimise processing time and data transfers it implements the novel event selection strategies of seeded, step-wise reconstruction and early rejection. The HLT is seeded by regions of interest identified at LVL1. These and the static configuration determine which algorithms are run to reconstruct event data and test the validity of trigger signatures. The decision to reject the event or continue is based on the valid signatures, taking into account pre-scale and pass-through. After the EF, event classification tags are assigned for streaming purposes.

Several powerful new features for commissioning and operation have been added: comprehensive monitoring is now built in to the framework; for validation and debugging, reconstructed data can be written out; the steering is integrated with the new configuration (presented separately), and topological and global triggers have been added.

This paper will present details of the final design and its implementation, the principles behind it, and the requirements and constraints it is subject to. The experience gained from technical runs with realistic trigger menus will be described.

Summary:

The final version of the software for first LHC collisions will be presented. There have been major changes, significant experience and a new design drawn up and implemented since the CHEP '06 paper.

Poster 2 - Board: 32 / 430

The ATLAS METADATA INTERFACE

Author: Solveig Albrand¹

Co-authors: Fabian LAMBERT¹ ; JEROME FULACHIER¹ ; THOMAS DOHERTY²

¹ LPSC/IN2P3/UJF Grenoble France

² University of Glasgow UK

Corresponding Author: solveig.albrand@lpsc.in2p3.fr

AMI was chosen as the ATLAS dataset selection interface in July 2006. It should become the main interface for searching for ATLAS data using physics metadata criteria. AMI has been implemented as a generic database management framework which allows parallel searching over many catalogues, which may have differing schema. The main features of the web interface will be described; in particular the powerful graphic query builder. The use of XML/XLST technology ensures that all commands can be used either on the web or from a command line interface via a web service.

The presentation will also describe the overall architecture of ATLAS metadata and the different actors and granularity involved, and the place of AMI within this architecture. We will discuss the problems involved in the correlation of metadata of differing granularity, and propose a solution for information mediation.

Poster 1 - Board: 69 / 71

The ATLAS Software Installation System for LCG/EGEE

Author: Alessandro De Salvo¹

Co-authors: Alex Barchiesi²; Andrzej Olszewski³; Carl Gwilliam⁴; Gernot Kroboth⁵; Grigori Rybkine⁶; John Kennedy⁵; Kondo Gnanvo⁷

¹ *Istituto Nazionale di Fisica Nucleare Sezione di Roma 1*

² *Università di Roma I "La Sapienza"*

³ *Henryk Niewodniczanski Institute of Nuclear Physics, Polish Academy of Sciences*

⁴ *University of Liverpool*

⁵ *Ludwig-Maximilians-Universität München*

⁶ *Royal Holloway College*

⁷ *Queen Mary and Westfield College*

Corresponding Author: alessandro.de.salvo@cern.ch

The huge amount of resources available in the Grids, and the necessity to have the most updated experiment software deployed in all the sites within a few hours, have spotted the need for automatic installation systems for the LHC experiments.

In this paper we describe the ATLAS system for the experiment software installation in LCG/EGEE, based on the Lightweight Job Submission Framework for Installation (LJSFi). This system is able to automatically discover, check, install, test and tag the full set of resources made available in LCG/EGEE to the ATLAS Virtual Organization in a few hours, depending on the site availability. The installations or removals may be centrally triggered as well as requested by the end-users for each site. A fallback solution to the manual operations is also available, in case of problems.

The installation data, status and job history are centrally kept in the installation DB and browseable via a web interface. The installation tasks are performed by one or more automatic agents. The ATLAS installation team is automatically notified in case of problems, in order to proceed with the manual operations. Each user may browse or request an installation activity in a site, directly by accessing the web pages, being identified by his personal certificate.

This system has been successfully used by ATLAS since 2003 to deploy about 60 different software releases and has performed more than 75000 installation jobs so far. The LJSFi framework is currently being extended to the other ATLAS Grids (NordGrid and OSG).

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Computer facilities, production grids and networking / 123

The ATLAS T0 Software Suite

Authors: Armin NAIRZ¹; Luc Goossens¹

¹ *CERN*

Corresponding Author: luc.goossens@cern.ch

ATLAS is a multi-purpose experiment at the LHC at CERN, which will start taking data in November 2007.

To handle and process the unprecedented data rates expected at the LHC (at nominal operation, ATLAS will record about

10 PB of raw data per year) poses a huge challenge on the computing infrastructure.

The ATLAS Computing Model foresees a multi-tier hierarchical model to perform this task, with CERN hosting the Tier-0 centre and associated Tier-1, Tier-2, ... centres distributed around the world.

The role of the Tier-0 centre 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 centres.

In this paper we report on the requirements, design and implementation of the ATLAS T0 software suite that has successfully met this challenge, most notably: TOM, the ATLAS T0 Manager and Eowyn, the job supervision component shared with the ATLAS WLCG-based production system.

We also report on the ATLAS Tier-0 scaling tests carried out in 2006/2007, whose goals were to evaluate the ATLAS Tier-0 work- and dataflow model, to test the infrastructure at CERN, and to perform Tier-0 operations up to their nominal rates.

Summary:

In this paper we report on the requirements, design and implementation of the ATLAS T0 software suite, most notably: TOM, the ATLAS T0 Manager and Eowyn, the job supervision component shared with the ATLAS WLCG-based production system. **Submitted on behalf of Collaboration (ex, BaBar, ATLAS):**

ATLAS

Online computing / 45

The ATLAS Trigger: Commissioning with cosmic-rays

Author: Boyd Jamie¹

¹ CERN

Corresponding Author: jamie.boyd@cern.ch

The ATLAS detector at CERN's LHC will be exposed to proton-proton collisions from beams crossing at 40 MHz. At the design luminosity there are roughly 23 collisions per bunch crossing. ATLAS has designed a three-level trigger system to select potentially interesting events. The first-level trigger, implemented in custom-built electronics, reduces the incoming rate to less than 100 kHz with a total latency of less than $2.5\mu\text{s}$. The next two trigger levels run in software on commercial PC farms. They reduce the output rate to 100-200 Hz.

In preparation for collision data-taking which is scheduled to commence in November 2007, several cosmic-ray commissioning runs have been performed. Among the first sub-detectors available for commissioning runs are parts of the barrel muon detector including the RPC detectors that are used in the first-level trigger. Data have been taken with a full slice of the muon trigger and readout chain, from the detectors in one sector of the RPC system, to the

second-level trigger algorithms and the data-acquisition system. The system is being prepared to include the inner-tracking detector in the readout and second-level trigger.

We will present the status and results of these cosmic-ray based commissioning activities. This work will prove to be invaluable not only during the commissioning phase but also for cosmic-ray data-taking during the normal running for detector performance studies.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Poster 2 - Board: 1 / 30

The Aragats Data Acquisition System for Highly Distributed Particle Detecting Networks

Author: Suren Chilingaryan¹

Co-authors: Ashot Chilingarian²; Varuzhan Danielyan²

¹ *The Institute of Data Processing and Electronics, Forschungszentrum Karlsruhe*

² *Cosmic Ray Division of Yerevan Physics Institute*

Corresponding Author: suren.chilingarian@ipe.fzk.de

For the reliable and timely forecasts of dangerous conditions of Space Weather world-wide networks of particle detectors are located at different latitudes, longitudes and altitudes. To provide better integration of these networks the DAS (Data Acquisition System) is facing a challenge to establish reliable data exchange between multiple network nodes which are often located in hardly accessible locations and operated by the different research groups. In this article we want to present DAS for SEVAN (Space Environmental Viewing and Analysis Network) elaborated on the top of free open-source technologies.

Our solution is organized as a distributed network of the uniform components connected by standard interfaces. The main component is URCS (Unified Readout and Control Server) which controls frontend electronics, obtains data and makes preliminary analysis. The URCS operates fully autonomous. Essential characteristics of software components and electronics are remotely controllable via dynamic web interface, the data is stored locally for certain amount of time and distributed to other nodes over web service interface on a request. To simplify data exchange with collaborating groups we use extensible XML based format for the data dissemination. The DAS at ASEC (Aragats Space Environmental Center) in Armenia is in operation from November 2006. The reliability of the service was proved by continuous monitoring of incident cosmic ray flux with 7 particle monitors located at 2000 and 3200 meters above sea level on the distance of 40 and 60 km. from the data analysis servers in main lab.

Online computing / 371

The BaBar Online Detector Control System Upgrade

Author: Matthias Wittgen¹

¹ *SLAC*

Corresponding Author: wittgen@slac.stanford.edu

The BaBar slow control system uses EPICS (Experimental Physics and Industrial Control System) running on 17 VME based single board computers (SBCs). EPICS supports the real-time operating systems vxWorks and RTEMS. During the 2004/05 shutdown BaBar started to install a new detector component, the Limited Streamer Tubes (LST), adding over 20000 high voltage channels and about 350 monitoring tasks to the control system. During 2005 data taking 5 out of 17 SBCs were replaced by PowerPC SBCs running RTEMS. Due to a lack of debugging and monitoring tools and memory and task management limitations on RTEMS, the decision was made to run Linux instead. Only a few EPICS drivers needed to be ported. Running EPICS on Linux provides a very stable environment and better debugging and monitoring tools than vxWorks or RTEMS. Many existing bugs in BaBar detector control applications could be discovered and fixed.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

BaBar

Poster 2 - Board: 50 / 197

The CEDAR event generator validation project

Author: Andy Buckley¹

¹ *Durham University*

Monte Carlo event generators are an essential tool for modern particle physics; they simulate aspects of collider events ranging from the parton-level “hard process” to cascades of QCD radiation in both initial and final states, non-perturbative hadronization processes, underlying event physics and specific particle decays. LHC events in particular are so complex that event generator simulations are essential to background modelling, trigger tuning and the analysis of candidate signal events itself.

However, event generators are not fully predictive: various phenomenological parameters must be tuned to experimental data to bootstrap a general purpose generator before physically meaningful predictions can be obtained. The CEDAR collaboration has developed a system for systematic validation of event generators, using archived data from the HepData archive for comparison with generator predictions for a range of observables. The experimental analyses corresponding to each HepData record are performed on simulated data using a generator interface and analysis toolkit named Rivet. CEDAR provides a Web interface to the Rivet system, JetWeb, which stores known generator parameter settings and their comparisons to HepData’s records in a relational database. As part of CEDAR, HepData has been significantly restructured to allow JetWeb and other applications to automatically retrieve and process the data archives.

CEDAR also provides a free collaborative development facility, HepForge, for HEP projects which aim to provide useful, well-engineered tools to the community. All the CEDAR projects are available through the portal Web site at www.cedar.ac.uk and the HepForge site at www.hepforge.org.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CEDAR

Distributed data analysis and information management / 444

The CERN Analysis Facility - A PROOF Cluster for Day-One Physics Analysis

Author: Jan Fiete Grosse Oetringhaus¹

¹ CERN

Corresponding Author: jan.fiete.grosse-oetringhaus@cern.ch

ALICE (A Large Ion Collider Experiment) at the LHC plans to use a PROOF cluster at CERN (CAF - Cern Analysis Facility) for fast analysis. The system is especially aimed at the prototyping phase of analyses that need a high number of development iterations and thus desire a short response time. Typical examples are the tuning of cuts during the development of an analysis as well as calibration and alignment. Furthermore, the use of an interactive system with very fast response will allow ALICE to extract physics observables out of first data quickly. A test setup consisting of 40 machines exists for evaluation since May 2006. The PROOF system enables the distributed usage and xrootd the access to locally distributed files. An automatic staging system of files migrated to CASTOR and files available in the AliEn Grid has been developed. The talk will present the current setup as well as performance tests that have been performed. The integration of PROOF into ALICE's software framework (AliRoot) will be shown.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ALICE

Software components, tools and databases / 325

The CMS Dataset Bookkeeping Service

Author: Lee Lueking¹

Co-authors: Andrew Dolgert²; Anzar Afaq³; Chris Jones²; Dan Riley²; Sergey Kosyakov³; Valentin Kuznetsov²; Vijay Sekhri³; Yuyi Guo³

¹ FERMILAB

² Cornell University

³ Fermilab

Corresponding Author: lueking@fnal.gov

The CMS Dataset Bookkeeping Service (DBS) has been developed to catalog all CMS event data from Monte Carlo and Detector sources. It includes the ability to identify MC or trigger source, track data provenance, construct datasets for analysis, and discover interesting data. CMS requires processing and analysis activities at various service levels and the system provides support for localized processing or private analysis, as well as global access for CMS users at large. Catalog entries can be moved among the various service levels with a simple set of migration tools, thus forming a loose federation of databases. DBS is available to CMS users via a Python API, Command Line, and a Discovery web page interfaces. The system is built as a multi-tier web application with Java servlets running under Tomcat, with connecting via JDBC to Oracle or MySQL database backends. Clients connect to the service through HTTP or HTTPS with authentication provided by GRID certificates and authorization through VOMS. DBS is an integral part of the overall CMS Data Management and Workflow Management systems. The system has been in operation since March 2007, an overview of

the schema, functionality, deployment details, operational statistics and experience will be presented.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS Offline Data Management Group

Poster 2 - Board: 19 / 280

The CMS LoadTest 2007: an infrastructure to exercise CMS transfer routes among WLCG Tiers

Author: Daniele Bonacorsi¹

¹ *INFN-CNAF, Bologna, Italy*

Early in 2007 the CMS experiment deployed a traffic load generator infrastructure, aimed at providing CMS Computing Centers (Tiers of the WLCG) with a means for debugging, load-testing and commissioning data transfer routes among them. The LoadTest is built upon, and relies on, the PhEDEx dataset transfer tool as a reliable data replication system in use by CMS. On top of PhEDEx, the CMS LoadTest 2007 exploited a light file generator mechanism; a simple system to customize test loads applied to inter-site links and to decouple the testing scopes for various routes. The implementation enables each CMS Tier to plug-in in the test and to exercise its connectivity to other CMS Tiers as desired, so to focus on debugging specific routes until complete commissioning of all links requested by the CMS Computing Model. The monitoring and visualization features of the PhEDEx tool provide an integrated way to overview the LoadTest activities at the same time as production transfers. The CMS LoadTest infrastructure was setup since the beginning of 2007 in preparation to the CMS Computing, Software and Analysis Challenge (CSA07) and then extensively used the winter and spring months to constantly exercise the transfer links to the needed scale, and to help CMS Tiers to address WLCG milestones. Experiences with this infrastructure and the achieved results are reviewed in this paper.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS Computing group

Poster 1 - Board: 17 / 367

The CMS Storage Manager and Data Flow in the High Level DAQ

Author: Elizabeth Sexton-Kennedy¹

¹ *FNAL*

Corresponding Author: sexton@fnal.gov

With the turn-on of the LHC, the CMS DAQ system is expecting to log petabytes of experiment data in the coming years. The CMS Storage Manager system is a part of the

high bandwidth event data handling pipeline of the CMS high level DAQ. It has two primary functions. Each Storage Manager instance collects data from the sub-farm, or DAQ slice of the Event Filter farm it has been assigned to, and logs it to disk. It also serves as an event and histogram server for calibration and monitoring processes. The Event Filter and Storage Manager cooperating systems use a special serialized form of the offline event data model in order to achieve the bandwidth specifications required by CMS. The online format is converted to the standard offline format in the tier zero before storage to the tape archive. This paper will detail the technical implementation and performance achievements of these cooperating systems during the recent data challenges.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

for the CMS Storage Manager group

Poster 1 - Board: 18 / 377

The CMS Tracker Control System

Author: Lorenzo Masetti¹

¹ CERN

Corresponding Author: lorenzo.masetti@cern.ch

The Tracker Control System (TCS) is a distributed control software to operate 2000 power supplies for the silicon modules of the CMS Tracker and monitor its environmental sensors. TCS must thus be able to handle 10^4 power supply parameters, 10^3 environmental probes from the Programmable Logic Controllers of the Tracker Safety System (TSS), 10^5 parameters read via DAQ from the DCUs in all front end hybrids and from CCUs in all control groups. TCS is built on top of an industrial SCADA program (PVSS) extended with a framework developed at CERN (JCOP) and used by all LHC experiments. The logical partitioning of the detector is reflected in the hierarchical structure of the TCS, where commands move down to the individual hardware devices, while states are reported up to the root which is interfaced to the broader CMS control system. The system computes and continuously monitors the mean and maximum values of critical parameters and updates the percentage of currently operating hardware. Automatic procedures switch off selected parts of the detector using detailed granularity and avoiding widespread TSS intervention.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS Tracker DCS

Online computing / 360

The DZERO Run 2 L3/DAQ System Performance

Author: Gordon Watts¹

¹ University of Washington

Corresponding Author: gwatts@uw.edu

The DZERO experiment records proton-antiproton collisions at the Fermilab Tevatron collider. The DZERO Level 3 data acquisition (DAQ) system is required to transfer event fragments of approximately 1-20 kilobytes from 63 VME crate sources to any of approximately 240 processing nodes at a rate

of 1 kHz. It is built upon a Cisco 6509 Ethernet switch, standard PCs, and commodity VME single board computers (SBCs). We will discuss the running experience of the system since 2002, the incremental upgrades, scaling capabilities, and how physics goals have altered the way we run the system.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

DZERO

Poster 1 - Board: 66 / 58

The Design and Implementation of BES Grid Computing System

Authors: Gang Chen¹ ; Yaodong CHENG¹

Co-authors: Jingyan SHI¹ ; Lu WANG¹ ; Xiaomei ZHANG¹

¹ *IHEP, China*

Corresponding Author: chyd@ihep.ac.cn

Beijing Electron Spectrometer (BESIII) experiment will produce 5 PB of data in next five years. Grid is used to solve this challenge. This paper introduces BES grid computing model and specific technologies, including automatic data replication, fine-grained job scheduling and so on.

Poster 1 - Board: 61 / 458

The Distributed Parallel Multi-Platform ATLAS Release Nightly Build System

Author: Obreshkov Emil¹

Co-authors: Alex Undrus² ; Christian Arnault³ ; David Quarrie⁴ ; Vincent Garonne⁵

¹ *INRNE/CERN*

² *BNL*

³ *LAL Orsay*

⁴ *LBNL*

⁵ *CERN*

Corresponding Authors: emil.obreshkov@cern.ch, drquarrie@lbl.gov

The ATLAS offline software comprises over 1000 software packages organized into 10 projects that are built on a variety of compiler and operating system combinations every night. File-level parallelism, package-level parallelism and multi-core build servers are used to perform simultaneous builds of 6 platforms that are merged into a single installation on AFS. This in turn is used to build distribution kits from which remote sites can create cloned installations. Since ATLAS typically has multiple release branches open simultaneously, corresponding to ongoing GRID productions, detector commissioning activities and ongoing software development, several instances of these distributed build clusters operate in parallel. We discuss the various tools that provide performance

gains and the error detection and retry mechanisms that have been developed in order to counteract network and other instabilities that would otherwise degrade the robustness of the system.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Computer facilities, production grids and networking / 248

The EELA Grid Infrastructure and HEP Applications in Latin America

Author: Lukas Nellen¹

¹ *I. de Ciencias Nucleares, UNAM*

Corresponding Author: lukas@nucleares.unam.mx

The EELA project aims at building a grid infrastructure in Latin America and at attracting users to this infrastructure. The EELA infrastructure is based on the gLite middleware, developed by the EGEE project. A test-bed, including several European and Latin American countries, was set up in the first months of the project. Several applications from different areas, especially Bio-medicine and High Energy Physics were deployed immediately and others, from climate and e-learning, were added during the second half of the first year of the project. In High Energy Physics, EELA currently provides resources to ALICE and LHCb. Work on resources for ATLAS in EELA is on its way and collaborations have been established with the Latin American CMS groups.

We will present the experience of the first 18 months of EELA and the current status. Finally, we will present the plans for future grid developments supporting the collaboration between Europe and Latin America, with particular emphasis on setting up a sustainable infrastructure.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

The EELA project

Distributed data analysis and information management / 361

The GLAST Data Handling Pipeline

Author: Dan Flath¹

Co-authors: Igor Pavlin¹; Karen Heidenreich¹; Tony Johnson¹

¹ *SLAC*

Corresponding Authors: dflath@slac.stanford.edu, tonyj@slac.stanford.edu

The Data Handling Pipeline ("Pipeline") has been developed for the Gamma-Ray Large Area Space Telescope (GLAST) launching at the end of 2007. Its goal is to generically process graphs of dependent tasks, maintaining a full record of its state, history

and data products. In cataloging the relationship between data, analysis results, software versions, as well as statistics (memory usage, cpu usage) of the processing it is able to track the complete provenance of all the data products.

The pipeline will be used to automatically process the data down-linked from the satellite and to deliver science products to the GLAST collaboration and the Science Support Center. It is currently used to perform Monte Carlo simulations, and analysis of commissioning data from the instrument. It will be stress tested this summer with “end-to-end” tests of data processing from the satellite and a full 1 year simulation run.

The Pipeline software is written almost entirely in Java and comprises several modules. A set of Java Stored Procedures compiled into the Oracle database allow computations on data to occur without network overhead. The Pipeline Server module accepts user requests, performs remote job scheduling and submission, and processes small “scriptlets” that allow lightweight calculations without the overhead of a batch job. The Pipeline Server submits jobs to the SLAC batch farm (3000+ linux cores), but will soon also submit jobs to a batch farm in France, and via the Grid to a farm in Italy. The “Pipeline Front End” displays live processing statistics via the web. It also provides AIDA charts summarizing CPU and memory usage, average submission wait time and also provides a graphical work-flow representation of the processing logic. Pipeline administrators can interact with the pipeline via web based or line-mode clients.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

GLAST

Event processing / 126

The Geant4 Virtual Monte Carlo

Author: Ivana Hrivnacova¹

¹ *IPN*

Corresponding Author: ivana.hrivnacova@cern.ch

The Virtual Monte Carlo (VMC) provides the abstract interface into the Monte Carlo transport codes: Geant3, Geant4 and Fluka. The user VMC based application, independent from the specific Monte Carlo codes, can be then run with all three simulation programs. The VMC has been developed by the ALICE Offline Project and since then it draw attention in more experimental frameworks.

Since its first release in 2002, the implementation of the VMC for Geant4 (Geant4 VMC) is in continuous maintenance and development, mostly driven by the requirements from new, non ALICE, users. In this presentation we will give an overview and the present status of this interface. We will report on new features, such as support for user defined Geant4 classes (the physics list, detector construction class) or support for Root TGeo geometry definition and G4Root navigation. We will also discuss the aspects specific to Geant4 and make a reflection about the strong and weak points of the VMC approach.

Summary:

We will give an overview and the present status of the Geant4 VMC interface.

Poster 2 - Board: 5 / 104

The GridPP Collaboration Web Site, 2000-2007

Author: Andrew McNab¹

¹ *University of Manchester*

Corresponding Author: andrew.mcnab@man.ac.uk

We describe the operation of www.gridpp.ac.uk, the website provided for GridPP and its precursor, UK HEP Grid, since 2000, and explain the operational procedures of the service and the various collaborative tools and components that were adapted or developed for use on the site. We pay particular attention to the security issues surrounding such a prominent site, and how the GridSite systems was developed in response to the needs of the site's users to allow delegation of management rights to members of particular subgroups within the collaboration. We explain how third party content generation systems have been incorporated into the site's security model, both purely interactive systems such as MediaWiki, and semi-automated tools such as RSS feeds and the Subversion version management system. Finally, we offer a brief survey of future developments relevant to portals and collaborative websites, such as OpenID and Shibboleth, and how they can be incorporated into the wider security framework of High Energy Physics grids.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

GridPP

Grid middleware and tools / 102

The GridSite security architecture

Author: Andrew McNab¹

¹ *University of Manchester*

Corresponding Author: andrew.mcnab@man.ac.uk

Components of the GridSite system are used within WLCG and gLite to process security credentials and access policies. We describe recent extensions to this system to include the Shibboleth authentication framework of Internet2, and how the GridSite architecture can now import a wide variety of credential types, including onetime passcodes, X.509, GSI, VOMS, Shibboleth and OpenID and then apply a single access policy to determine the subset of rights to be granted to particular request or session, controlled by policies written in the GACL or XACML languages. Finally, we provide examples of using GridSite and Apache to host web services for High Energy Physics grids written in C/C++/Scripts, as well as Java, and show how this one architecture has been used for purely interactive websites such as www.gridpp.ac.uk, for sites that are a mixture of human-generated and automated monitoring such as the LCG GOC Database, and for web services for grids such as the gLite WMPProxy service.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

GridPP

Collaborative tools / 87

The Health-e-Child Project: A Grid enabled Platform for European Paediatrics

Author: Richard McClatchey¹

Co-authors: Ashiq Anjum² ; Ian Willers³

¹ *University of the West of England*

² *UWE*

³ *CMS*

Corresponding Author: richard.mcclatchey@cern.ch

The Health-e-Child (HeC) project is an EC Framework Programme 6 Integrated Project that aims at developing an integrated healthcare platform for paediatrics. Through this platform biomedical informaticians will integrate heterogeneous data and perform epidemiological studies across Europe.

The main objective of the project is to gain a comprehensive view of a child's health by 'vertically' integrating biomedical, information and knowledge that spans the entire spectrum from genetic to epidemiological. The resulting Grid enabled biomedical information platform will be supported by robust search, optimization and matching techniques for information collected in hospitals across Europe.

In particular, paediatricians will be provided with decision support, knowledge discovery and disease modelling applications that will access data in hospitals in the UK, Italy and France, integrated via the Grid. For economy of scale, reusability, extensibility, and maintainability, Health-e-Child is being developed on top of an EGEE/gLite based infrastructure that provides all the common data and computation management services required by the applications. The emphasis of the Health-e-Child effort is on universality of information, person-centricity of information, universality of application, multiplicity and variety of biomedical analytics and person-centricity of interaction. Its corner stone is the integration of information across biomedical abstraction whereby layers of biomedical information (i.e., genetic, cell, tissue, organ, individual, and population layer) are integrated to provide a unified view of a person's biomedical and clinical condition.

This paper discusses the major issues and challenges in bio-medical data integration and how these will be resolved in the Health-e-Child system. It establishes the need for the HeC infrastructure and emphasises the importance of user requirements analysis when integrating highly heterogeneous medical information. HeC is presented as an example of how computer science originating from the high energy physics community can be adapted for use by biomedical informaticians to deliver tangible real-world benefits.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

Health-e-Child Grid project

Online computing / 222

The High-Level Trigger at the CMS experiment

Author: Christos Leonidopoulos¹

¹ *CERN*

Corresponding Authors: leonard.apanasevich@cern.ch, c.leonidopoulos@cern.ch

The High Level Trigger (HLT) that runs in the 1000 dual-CPU box Filter Farm of the CMS experiment is a set of sophisticated software tools

for selecting a very small fraction of interesting events in real time. The coherent tuning of these algorithms to accommodate multiple physics channels is a key issue for CMS, one that literally defines the reach of the experiment's physics program. In this presentation we will discuss studies of the performance of the HLT algorithms for preliminary versions of integrated Trigger Menus.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS

Collaborative tools / 438

The IceCube Data Acquisition Software: Lessons Learned during Distributed, Collaborative, Multi-Disciplined Software Development.

Authors: Chuck McParland¹ ; Keith Beattie¹

Co-authors: Chris Day¹ ; Kael Hanson² ; Simon Patton¹

¹ *LBNL*

² *UW Madison*

Corresponding Authors: ksbeattie@lbl.gov, cpmcparland@lbl.gov

In this experiential paper we report on lessons learned during the development of the data acquisition software for the IceCube project - specifically, how to effectively address the unique challenges presented by a distributed, collaborative, multi-institutional, multi-disciplined project such as this. While development progress in software projects is often described solely in terms of technical issues, our experience indicates that non- and quasi-technical interactions play a substantial role in the effectiveness of large software development efforts. These include: selection and management of multiple software development methodologies, the effective use of various collaborative communication tools, project management structure and roles, and the impact and apparent importance of these elements when viewed through the differing perspectives of hardware, software, scientific and project office roles. Even in areas clearly technical in nature, success is still influenced by non-technical issues that can escape close attention. In particular we describe our experiences on language selection, software requirements specification, and selection and use of development, framework and communication tools. Using both anecdotal and detailed software architecture descriptions, we make observations on what tools and techniques have and have not been effective in this geographically disperse (including the South Pole) collaboration and offer suggestions on how similarly structured future projects may build upon our experiences.

Event processing / 60

The LDC Software Framework for the ILC detector

Author: Frank Gaede¹

¹ *DESY IT*

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

The International Linear Collider is the next large accelerator project in High Energy Physics. The Large Detector Concept (LDC) study is one of four

international working groups that are developing a detector concept for the ILC. The LDC uses a modular C++ application framework (Marlin) that is based on the international data format LCIO. It allows the distributed development of reconstruction and analysis software. The framework is mainly used for optimizing the physics performance of the planned detector based on the Particle Flow paradigm. It recently has been adapted to also be applied in various detector prototype test beam studies within the EUDET project. In this talk we give an overview of the core framework where the focus will be on recent developments and improvements since it has been first presented at CHEP2006.

Plenary / 468

The LHC Machine and Experiments: Status and Prospects

Corresponding Author: tejinder.virdee@cern.ch

The current status of the LHC machine and the experiments, especially the general-purpose experiments, will be given. Also discussed will be the preparations for the physics run in 2008. The prospects for physics, with an emphasis on what can be expected with an integrated luminosity of 1 fb⁻¹, will be outlined.

Poster 2 - Board: 120 / 150

The LHCb Computing Data Challenge DC06

Author: Raja Nandakumar¹

Co-authors: Adria Casajus²; Andrei Tsaregorotsev³; Andrew Cameron Smith⁴; Andrew Pickford⁵; Angelo Carbone⁶; Besma M'charek⁷; Carmine Cioffi⁸; Cedric Potterat⁹; Daniela Bortolotti⁶; Davide Perego¹⁰; Gennady Kuznetsov¹; Gianluca Castellani¹¹; Joel Closier¹¹; Johan Blouw¹²; Marco Adinolfi¹³; Marcos Seco Miguelez¹⁴; Marianne Bargiotti¹¹; Nick Brook¹⁵; Philippe Charpentier¹¹; Ricardo Graciani Diaz²; Roberto Santinelli¹¹; Roland Bernet¹⁶; Sergio Gomez²; Stuart Paterson¹¹

¹ Rutherford Appleton Laboratory

² Universitat de Barcelona

³ Université d'Aix - Marseille II

⁴ University of Edinburgh

⁵ University of Glasgow

⁶ Università & INFN, Bologna

⁷ Vrije Universiteit, Amsterdam

⁸ University of Oxford

⁹ LPHE-IPEP, Lausanne

¹⁰ Università degli Studi di Milano Bicocca, Milano

¹¹ CERN

¹² Physikalisches Institut, Heidelberg

¹³ H. H. Wills Physics Laboratory, Bristol

¹⁴ Universidad de Santiago de Compostela

¹⁵ University of Bristol

¹⁶ Universität Zürich

Corresponding Author: raja.nandakumar@cern.ch

The worldwide computing grid is essential to the LHC experiments in analysing the data collected by the detectors. Within LHCb, the computing model aims to simulate data at Tier-2 grid sites as well as non-grid resources. The reconstruction, stripping and analysis of the produced LHCb data will primarily place at the Tier-1 centres. The computing data challenge DC06 started in May 2006 with the primary aims being to exercise the LHCb computing model and to produce events which will be used for analyses in the forthcoming LHCb physics book. This paper gives an overview of the LHCb computing model and addresses the challenges and experiences during DC06. The management of the production of Monte Carlo data on the LCG was done using the DIRAC workload management system which in turn uses the WLCG infrastructure and middleware. We shall report on the amount of data simulated during DC06, including the performance of the sites used. The paper will also summarise the experience gained during DC06, in particular the distribution of data to the Tier-1 sites and the access to this data.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

LHCb

Online computing / 139

The LHCb High Level Trigger Software Framework

Author: Markus Frank¹

¹ CERN

Corresponding Author: markus.frank@cern.ch

The High Level Trigger and Data Acquisition system of the LHCb experiment at the CERN Large Hadron Collider must handle proton-proton collisions from beams crossing at 40 MHz. After a hardware-based first level trigger events have to be processed at the rate of 1 MHz and filtered by software-based trigger applications that run in a trigger farm consisting of up to 2000 PCs. The final rate of accepted events is around 2 kHz.

This contribution describes the architecture used to host the selection algorithms of the high level trigger on each trigger node, that is based on shared memory event buffers. It illustrates the interplay between event building processes, event filter processes and processes sending accepted events to the storage system. It describes these software components that are based on the Gaudi event processing framework.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

LHCb Online

Poster 2 - Board: 45 / 157

The LiC Detector Toy Program

Authors: Manfred Valentan¹ ; Meinhard Regler¹ ; Rudolf Frühwirth¹ ; Rudolf Höfler¹ ; Winfried Mitaroff¹

¹ *Inst. of High Energy Physics, Vienna*

Corresponding Author: rudolf.fruehwirth@oeaw.ac.at

We present the “LiC Detector Toy” (“LiC” for Linear Collider) program, a simple but powerful software tool for detector design, modification and geometry studies. It allows the user to determine the resolution of reconstructed track parameters for the purpose of comparing and optimizing various detector set-ups. It consists of a simplified simulation of the detector measurements, taking into account multiple scattering and measurement errors, followed by full single track reconstruction using the Kalman filter. The detector model is built from geometry files describing the layout of the detector layers, their material, their accuracy and their efficiency. In addition, it contains information about passive scattering layers and their material budget. The reconstructed tracks can be written to a text file and passed on to a vertex reconstruction program. The tool is written in MATLAB and may be installed on a laptop. For the ease of use, the program is integrated into a Graphical User Interface (GUI). We describe the main components of the tool and show results from performance studies with the LDC and SiD detector concepts at the ILC, and Super-BELLE at KEK.

Computer facilities, production grids and networking / 42

The Open Science Grid - Its Status and Implementation Architecture

Author: Ruth Pordes¹

¹ FERMILAB

Corresponding Author: ruth@fnal.gov

The Open Science Grid (OSG) is receiving five years of funding across six program offices of the Department of Energy Office of Science and the National Science Foundation. OSG is responsible for operating a secure production-quality distributed infrastructure, a reference software stack including the Virtual Data Toolkit (VDT), extending the capabilities of the high throughput virtual facility, and supporting an expansion of the user base. OSG also educates existing and potential users.

OSG Consortium members provide the computing and storage resources accessible from the distributed infrastructure and the user applications for its use. Over sixty DOE Lab and University facilities can now be accessed. Access to large storage resources is increasing. The infrastructure relies on ESNET and Internet2 production and advanced networks.

The OSG implementation architecture presents the Virtual Organization (VO) - aka science/research community – as a capable middle tier between the diverse distributed resources and the end users. Implementation of this architecture focusses on: making each resource self-managed, secure, sharable, and accessible locally and remotely; providing secure common services, support and reference software to the communities to enable their effective use of the OSG Facility; and providing end-to-end and facility-wide tools, operational security, and user support.

The OSG implementation architecture is cognizant of federated and intersecting infrastructures – spanning

individually managed facilities, university department clusters, local area shared campus infrastructures, the large national grids, and community scoped distributed environments.

We report on the status of OSG, its implementation architecture today, and plans for the future.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

Open Science Grid

Online computing / 219

The PHENIX Experiment in the RHIC Run 7

Author: Martin Purschke¹

¹ *BROOKHAVEN NATIONAL LABORATORY*

Corresponding Author: purschke@bnl.gov

The PHENIX experiment at the Relativistic Heavy Ion Collider (RHIC) has commissioned several new detector systems which are part of the general readout for the first time in the RHIC Run 7, which is currently under way.

In each of the RHIC Run periods since 2003, PHENIX has collected about 0.5 PB of data. For Run 7 we expect record luminosities for the Au-Au beams, which will lead to even larger data sizes.

We have in the past used GRID tools to transfer substantial data volumes in the order of 400TB off-site in order to make use of remote computing capacity in Japan, France, and the US, and are set up to do the same in Run 7.

Even at the highest expected luminosities we will still be able to log all events triggered by the level-1 trigger in this Run. The data acquisition system can sustain a rate of up to 600MB/s. In order to expedite the analysis of interesting events, we will use the Level-2 trigger system as a filter to select interesting events for priority reconstruction and analysis.

We will give an overview of the online system, our strategies to cope with the data storage demands, near-line transfers to off-site locations, and analysis strategies.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

for the PHENIX Collaboration

Event processing / 326

The Pierre Auger Observatory Offline Software

Authors: Darko Veberic¹; Javier Gonzalez²; Julio Bahilo³; Luis Prado, Jr.⁴; Lukas Nellen⁵; Markus Roth⁶; Michael Unger⁶; Ralf Ulrich⁶; Sergio Barroso⁷; Stefano Argiro⁸; Thomas McCauley⁹; Thomas Paul⁹; Troy Porter¹⁰

¹ *University of Nova Gorica*

² *Louisiana State University*

³ *University of Granada*

⁴ *Universidade Estadual de Campinas*

⁵ *Universidad Nacional Autonoma de Mexico*

⁶ *Karlsruhe Institute of Technology*

⁷ *Centro Brasileiro de Pesquisas Fisicas*

⁸ *INFN and University of Torino*

⁹ *Northeastern University*

¹⁰ *University of California, Santa Cruz*

Corresponding Authors: t.paul@neu.edu, lukas@nucleares.unam.mx

The Pierre Auger Observatory aims to discover the nature and origins of the highest energy cosmic rays. The large number of physicists involved in the project and the diversity of simulation and reconstruction tasks pose a challenge for the offline analysis software, not unlike the challenges confronting software for very large high energy physics experiments. Previously we have reported on the design and implementation of a general purpose but relatively lightweight framework which allows collaborators to contribute algorithms and sequencing instructions to build up the variety of applications they require. In this report, we update the status of this work and describe some of the successes and difficulties encountered over the last few years of use. We explain the machinery used to manage user contributions, to organize the abundance of configuration files, to facilitate multi-format file handling, and to provide access to event and time-dependent detector information residing in various data sources. We also describe the testing procedures used to help maintain stability of the code in the face of a large number of contributions. Foundation classes will also be discussed, including a novel geometry package which allows manipulation of abstract geometrical objects independent of coordinate system choice.

Poster 2 - Board: 83 / 211

The Pre-Production Service in WLCG/EGEE

Authors: Antonio Retico¹ ; Nicholas Thackray¹

¹ *CERN*

Corresponding Author: antonio.retico@cern.ch

The WLCG/EGEE Pre-Production Service (PPS) is a grid infrastructure whose goal is to give early access to new services to WLCG/EGEE users in order to evaluate new features and changes in the middleware before new versions are actually deployed in PPS.

The PPS grid counts about 30 sites providing resources and manpower.

The service contributes to the overall quality of the grid middleware by:

- making new middleware functionalities early available to the VOs in order to tune their applications
- allowing Evaluation of software and deployment procedures in real operational conditions
- providing feedback for early bug fix to be done on the release before it is moved to production.

Finally, as the PPS is well integrated in the framework of the operational procedures in place for production, it results to be also a good context where to practice new procedures and monitoring tools.

The Pre-Production Service has been now running for nearly two years in the WLCG/EGEE context. We give in this paper an overview on the current mandate and status of the service, their past evolutions and the future plans.

Poster 2 - Board: 63 / 428

The RAVE/VERTIGO Vertex Reconstruction Toolkit and Framework

Authors: Bernhard Pflugfelder¹ ; Fabian Moser² ; Herbert V. Riedel² ; Winfried A. Mitaroff² ; Wolfgang Waltenberger²

¹ *Institute of High Energy Physics, Vienna*

² *Institute of High Energy Physics (HEPHY) of the Austrian Academy of Sciences, Vienna*

Corresponding Author: mitaroff@hephy.oeaw.ac.at

A detector-independent toolkit (RAVE) is being developed for the reconstruction of the common interaction vertices from a set of reconstructed tracks. It deals both with “finding” (pattern recognition of track bundles) and with “fitting” (estimation of vertex position and track momenta). The algorithms used so far include robust adaptive filters which are derived from the CMS experiment, and the ZvTop topological vertex finder from the ILC-LCFI project. Further contributions from outside are welcome.

The toolkit is supplemented by a standalone framework (VERTIGO) for testing, analyzing and debugging. Tools include visualisation, histogramming, artificial event generation (“vertex gun”), an LCIO interface, and flexible I/O (“data harvester”, see separate presentation). Emulation of various detector environments is supported by a flexible “skin concept”.

Main design goals have been ease of use, high integrability into existing software environments, extensibility, general openness and platform independence. The toolkit and framework are coded in C++, with interfaces for other languages (Java, Python). RAVE has been successfully embedded into the ILC software packages MarlinReco and (via a wrapper) org.lcsim. VERTIGO is used as a framework for the development of new RAVE algorithms, and for standalone vertex reconstruction of data created by detector simulation packages like the “LiC Toy” (see separate presentation).

Source code and documentation of RAVE and VERTIGO are maintained in a subversion repository accessible via the web. A beta release is available.

Event processing / 63

The Reconstruction and Calibration of the BESIII Drift Chamber

Author: Linghui Wu¹

Co-author: Yao Zhang¹

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

Corresponding Authors: zhangyao@mail.ihep.ac.cn, wulh@mail.ihep.ac.cn

The BESIII detector will be commissioned at the upgraded Beijing Electron Positron Collider (BEPCII) at the end of 2007. The drift chamber(MDC), which is one of the most important sub-detectors of the BESIII detector, is expected to provide good momentum resolution (0.5%@1GeV/c) and tracking efficiency in a

range of 0.1~2.0 GeV/c. This makes stringent demands on the performance of the offline software. The event reconstruction and offline calibration algorithms have been developed in BESIII Offline Software System (BOSS), using C++ language and object-oriented techniques.

The reconstruction consists of tracking and track fitting using Kalman filter method. The tracking uses a pattern matching method to find track segments and then combined into track candidates followed by a least square fit. The track fitting based on Kalman filter is used to handle effects of material and non-uniform magnetic field. The implementation of the tracking and the fitting algorithms and the performance with the Monte Carlo data will be presented.

The study of the offline calibration method using the cosmic ray data, including the calibration of the X-T relation and the software alignment, will be also presented.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

BESIII

Online computing / 279

The Run Control and Monitoring System of the CMS Experiment

Author: Alexander Oh¹

¹ CERN

Corresponding Author: alexander.oh@cern.ch

The CMS experiment at the LHC at CERN will start taking data towards the end of 2007. To configure, control and monitor the experiment during data-taking the Run Control and Monitoring System (RCMS) was developed.

This paper describes the architecture and the technology used to implement the RCMS, as well as the deployment and commissioning strategy of this online software component for the CMS experiment. The RCMS framework is based on a set of web-applications implemented with Java Servlet technology, AJAX and JSP for user interfaces, and supports MySQL and Oracle as DB back-end.

A hierarchical control structure organizes the Run Control in sub-systems. For the DAQ system a set of tools has been developed to manage the flexible generation of configurations with the goal to allow fast reconfiguration of the system, which will comprise about 4000 computing nodes in the full stage of expansion.

A crucial test was passed with RCMS being successfully used in the so called "Magnet Test & Cosmic Challenge" of CMS - a small set of sub-detectors being operated to detect cosmic muons - during 2006.

Towards the first run, RCMS will be tested in another "Cosmic Challenge" exercise with the sub-detectors, DAQ and trigger components in their final position in the underground cavern at the LHC ring.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

Alexander Oh, for the CMS DAQ group

Poster 2 - Board: 59 / 304

The Simulation of the CMS Electromagnetic Calorimeter

Author: Fabio Cossutti¹

¹ INFN

Corresponding Author: fabio.cossutti@ts.infn.it

The CMS Collaboration has developed a detailed simulation of the electromagnetic calorimeter (ECAL), which has been fully integrated in the collaboration software framework CMSSW. The simulation is based on the Geant4 detector simulation toolkit for the modelling of the passage of particles through matter and magnetic field. The geometrical description of the detector is being re-implemented using the DetectorDescription language, combining an XML based description of with the algorithmic definition of the position of the elements. The ECAL simulation software is fully operational and has been validated using real data from the ECAL test beam experiment that took place in summer 2006.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

F. Cossutti for the CMS ECAL group

Poster 1 - Board: 112 / 464

The Status of Grid for Belle Experiment

Author: Go Iwai¹

Co-authors: Hideyuki Nakazawa²; Setsuya Kawabata¹; Shohei Nishida³; Takashi Sasaki¹; Yoshimi Iida¹; Yoshiyuki Watase¹

¹ KEK/CRC

² NCU

³ KEK/IPNS

Corresponding Author: go.iwai@kek.jp

The Belle Experiment is an ongoing experiment with an asymmetric electron-positron collider at KEK and already has a few PB scales of data in total including hundreds TB DST (Data Summary Tape) and MC data. It's too much difficult to export existing data to LCG (LHC Computing Grid) physically because of huge amount of data. We setup a SRB (Storage Resource Broker) server to access them by using SRB-DSI (SRB Data Storage Interface) which is an extension to the GridFTP server to interact with SRB. This approach isn't benefit only for Belle LCG users but also for legacy users. The Belle VO (Virtual Organization) has been federated among 4 countries, 6 institutes, 9 sites, and the analysis software has been installed to every site of this VO. The results of testing functionality, measuring performance and future prospects based on this federation will be shown in this paper.

Poster 1 - Board: 79 / 169

The Swiss ATLAS Grid

Authors: Frederik Orellana¹; Sigve Haug²; Szymon Gadomski²

Co-authors: Allan Clark³; Antonio Ereditato²; Borge Kile Gjelsten²; Cyril Topfel²; Hans Peter Beck²; Ignacio Aracena²; Till Eifert³

¹ *Niels Bohr Institute, Copenhagen*

² *LHEP University of Bern*

³ *DPNC Universite de Geneve*

Corresponding Author: sigve.haug@lhep.unibe.ch

Since 2005 the Swiss ATLAS Grid is in production. It comprises four clusters at one Tier 2 and two Tier 3 sites. About 800 heterogenous cores and 60 TB disk space are connected by a dark fibre network operated at 10 Giga bit per second. Three different operating systems are deployed. The Tier 2 cluster runs both LCG and NorduGrid middleware (ARC) while the Tier 3 clusters run only the latter. As local resource management systems both OpenPBS/Torque and Sun Grid Engine are used. In 2006 about 200 000 CPU hours of central ATLAS production were run on a smaller version of this infrastructure. Local users ran about 60 000 CPU hours for ATLAS physics studies. As end user tools ARC clients, ATLAS DQ2 software, ATLAS Ganga and a new graphical user interface are employed. With emphasis on the heterogenous solutions we report on the current and future infrastructure and usage of this part of the ATLAS grid.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Poster 1 - Board: 68 / 67

The Tier0 road to LHC Data Taking

Author: Tony Cass¹

¹ *CERN*

Corresponding Author: tony.cass@cern.ch

CERN, as other sites, has been preparing computing services for the arrival of LHC data for some time—more than 11 years if everything started at the First LHC Computing Workshop, held in Padova in June 1996. With LHC data taking now just around the corner, this presentation takes a look back at preparations at CERN and considers some of the key choices made along the way. Which were prescient? Which will we live to regret? And, most importantly, are we ready?

Poster 1 - Board: 105 / 386

The US LHCNet Network for HEP

Authors: Artur Barczyk¹; Dan Nae²; Harvey Newman¹; Iosif Legrand¹; Julian Bunn¹; Sylvain Ravot¹; Yang Xia¹

¹ *Caltech*

² *California Institute of Technology (CALTECH)*

Corresponding Author: dan.nae@cern.ch

In this paper we present the design, implementation and evolution of the mission-oriented USLHCNet for HEP research. The design philosophy behind our network is to help meet the dataintensive computing challenges of the next generation of particle physics experiments with a comprehensive, network-focused approach. Instead of treating the network as a static, unchanging and unmanaged set of inter-computer links, we are developing and using it as a dynamic, configurable and closely monitored resource that is managed from end-to-end. We will present our work in the various areas of the project, recent changes in the infrastructure, including the addition of LCAS/VCAT/GFP capable SONET equipment, future plans, transport protocol research and grid application development. Our working methodology is a continuous cycle of evaluating equipment and technologies (servers, networking equipment, new standards) and network application development in order to build a production network for research. Our goal is to construct a nextgeneration network that is able to meet the data processing, distribution, access and analysis needs of the particle physics community.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

USLHCNet

Software components, tools and databases / 124

The Virtual Geometry Model

Author: Ivana Hrivnacova¹

Co-author: Brett Viren²

¹ *IPN, Orsay, France*

² *Physics Department, BNL*

Corresponding Author: ivana.hrivnacova@cern.ch

The Virtual Geometry Model (VGM) was introduced at CHEP in 2004, where its concept, based on the abstract interfaces to geometry objects, has been presented. Since then, it has undergone a design evolution to pure abstract interfaces, it has been consolidated and completed with more advanced features. Currently it is used in Geant4 VMC for the support of TGeo geometry definition with Geant4 native geometry navigation and recently it has been used in the validation of the G4Root tool.

The implementation of the VGM for a concrete geometry model represents a small layer between the VGM and the particular native geometry. In addition to the implementations for Geant4 and Root TGeo geometry models, there is now added the third one for AGDD, which together with the existing XML exporter makes the VGM the most advanced tool for exchanging geometry formats providing 9 ways of conversions between Geant4, TGeo, AGDD and GDML models.

In this presentation we will give the overview and the present status of the tool, we will review the supported features and point to possible limits in converting geometry models.

Summary:

We will give the overview and the present status of the Virtual Geometry Model.

Software components, tools and databases / 50

The configuration system of the ATLAS Trigger

Authors: Andreas Hoecker¹ ; David Berge¹ ; Johannes Haller² ; Jörg Stelzer¹ ; Miroslav Nozicka³ ; Paul Bell⁴ ; Simon Head⁴ ; Thorsten Wengler⁴

¹ *CERN, Switzerland*

² *Hamburg University, Germany*

³ *DESY Hamburg, Germany*

⁴ *Manchester University, UK*

Corresponding Author: joerg.stelzer@cern.ch

The ATLAS detector at CERN's LHC will be exposed to proton-proton collisions at a rate of 40 MHz. To reduce the data rate, only potentially interesting events are selected by a three-level trigger system. The first level is implemented in custom-made electronics, reducing the data output rate to less than 100 kHz. The second and third levels are software triggers with a final output rate of 100 to 200 Hz. A system has been designed and implemented that hosts and records the configuration of all three trigger levels at a centrally maintained location. This system provides consistent configuration information to the online trigger for the purpose of data taking as well as to the offline trigger simulation. The use of relational database technology provides a means of flexible information browsing, easy information distribution across the ATLAS reconstruction sites, and reliable recording of the trigger configuration history over the lifetime of the experiment. The functionality of this design has been demonstrated in dedicated configuration tests of the ATLAS level-1 Central Trigger and of a 600-node software trigger computing farm. We present an overview of the main system components, including a sophisticated, JAVA-based front end to populate and maintain the configuration information, and report on the current status.

Grid middleware and tools / 190

The gLite Workload Management System

Authors: Alessandro Maraschini¹ ; Alessio Gianelle² ; Alvise Dorise² ; Andrea Cavallini¹ ; Andrea Guarise³ ; Andrea Krop¹ ; Andrea Parrini¹ ; David Rebatto⁴ ; Elisabetta Molinari⁴ ; Elisabetta Ronchieri⁵ ; Fabrizio Pacini¹ ; Francesco Giacomini⁵ ; Francesco Prelz⁴ ; Giuseppe Avellino¹ ; Giuseppe Patania³ ; Luca Petronzio¹ ; Luigi Zangrando² ; Marco Cecchi⁵ ; Marco Mezzadri⁴ ; Marco Pappalardo⁶ ; Massimo Sgaravatto² ; Maurizio Porciani¹ ; Moreno Marzolla² ; Paolo Andreetto² ; Roberto Lops⁵ ; Salvatore Monforte⁶ ; Sergio Andreatto⁵ ; Stefano Beco¹ ; Ugo Grandinetti¹ ; Valerio Venturi⁵ ; Vincenzo Ciaschini⁵ ; Vincenzo Martelli⁶

¹ *DATAMAT*

² *INFN PD*

³ *INFN TO*

⁴ *INFN MI*

⁵ *INFN cnaf*

⁶ *INFN CT*

Corresponding Authors: marco.cecchi@cnaf.infn.it, alessandro.maraschini@datamat.it

The gLite Workload Management System (WMS) is a collection of components providing a service responsible for the distribution and management of tasks across resources available on a Grid. The main purpose is to accept a request of execution of a job from a client, find appropriate resources to satisfy it and follow it until completion. Different aspects of job management are accomplished by different WMS

components such as the WMPProxy (a Web Service managing users authentication/authorization and operation requests) and the Workload Manager (which performs the matchmaking on the job's requirements and determines where it has to be actually executed).

Different kinds of job can be described providing needed information through a flexible high-level language called JDL. The most interesting and innovating job types are the Directed Acyclic Graphs (a set of jobs where the input/output/execution of one of more jobs may depend on one or more other jobs), the Parametrics (which allow the submission of a large number of jobs by simply specifying a parametrized description), and the Collections (which represent a possibly huge number of jobs specified within a single description)

Several new functionalities (such as the use of Service Discovery for obtaining new service endpoints to be contacted, the automatic sandbox files archiving/compression and sharing, the bulk-matchmaking support), intense testing and a constant bug fixing activity dramatically increased job submission rate and service stability.

Future developments of the gLite WMS will be focused on reducing external software dependency, improving its portability, robustness and usability.

Event processing / 144

The global chi² track fitter in ATLAS

Author: Thijs Cornelissen¹

¹ CERN

While most high energy experiments use track fitting software that is based on the Kalman technique, the ATLAS offline reconstruction has several global track fitters available. One of these is the global chi² fitter, which is based on the scattering angle formulation of the track fit. One of the advantages of this method over the Kalman fit is that it can provide the scattering angles and related quantities (e.g. the residual derivatives) to the alignment algorithms. The algorithm has been implemented in the new common tracking framework in ATLAS, the philosophy of which is to improve the modularity and flexibility of the tracking software. This flexibility has proven crucial for the understanding of the data from the testbeam and cosmic runs. An overview of recent results will be presented, in particular the results from the combined tracking with the inner detector and the muon spectrometer using the cosmics data.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Software components, tools and databases / 351

The life cycle of HEP offline software

Authors: Chris Jones¹ ; Liz Sexton-Kennedy² ; Peter Elmer³

¹ Cornell University

² FNAL

³ Princeton University

Corresponding Author: peter.elmer@cern.ch

Modern HEP experiments at colliders typically require offline software systems consisting of many millions of lines of code. The software is developed by hundreds of geographically distributed developers and is often used actively for 10-15 years or longer. The tools and technologies to support this HEP software development model have long been an interesting topic at CHEP conferences. In this presentation we look instead at the software project management aspects, and in particular at the time evolution of the offline software projects of experiments over their lifetimes, from the pre-datataking period to the analysis period. We focus on three mature experiments (BaBar, CDF, CLEO) and one experiment about to start taking data (CMS). We examine quantitatively how the software code base and developer participation evolve through the various phases of the experiment. We also explore the impact of functionality increases, requirement changes and the phases of the experiment in order to draw conclusions for experiments at the beginning of their life cycle.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

Multiple (BaBar/CDF/CLEO/CMS)

Event processing / 278

The performance of the Geant4 Standard EM package for LHC and other applications

Author: Vladimir Ivantchenko¹

¹ CERN, ESA

Corresponding Author: vladimir.ivantchenko@cern.ch

Current status of the Standard EM package of the Geant4 toolkit is described. The precision of simulation results is discussed with the focus on LHC experiments. The comparisons of the simulation with the experimental data are shown.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

Geant4 Summary:

All components needed for simulation of electromagnetic processes for LHC experiments are provided by the Geant4 Standard EM package. The software for the testing suite of the package is described. This software allows performing both validations versus experimental data and regression tests for stability of simulation predictions between different Geant4 releases. The results of validation of EM physics with experimental data show a good quality of the package. The comparison of precision versus CPU performance allows to choose an optimal options for the concrete simulation.

Grid middleware and tools / 318

Tools for the management of stored data and transfer of data: DPM and FTS

Author: David Smith¹

¹ CERN

Corresponding Authors: markus.schulz@cern.ch, david.smith@cern.ch

As a part of the EGEE project the data management group at CERN has developed and support a number of tools for various aspects of data management:

A file catalog (LFC), a key store for encryption keys (Hydra), a grid file access library (GFAL) which transparently uses various byte access protocols to access data in various storage systems, a set of utilities (lcg_utils) for higher level operations on data are all supported. However, in this presentation we will focus on giving an overview of two components in particular:

A disc pool manager (DPM) which provides a service to coordinate the storage of files across discs. The DPM features POSIX ACLs on files and pools, file lifetime with garbage collection, optional replication of data within the DPM and authorization based on VOMS grid certificates. The DPM offers an SRM interface, versions 1.1 and 2.2, along with its own control interface. Access to data is supported via gsiftp, rfiio and an optional xrootd module is also available.

A file transfer service (FTS) allows the replication of data from one data store to another. The FTS features individually configurable, unidirectional management channels. The channels allow allocation of parameters such as number of concurrent transfers, number of parallel streams or TCP buffer size.

SRM (version 1.1 or 2.2) is used to send requests to the storage systems. Third party gridftp or SRMCopy initiated transfers are supported.

Summary:

Summary of EGEE data management components FTS and DPM.

Poster 2 - Board: 92 / 380

Towards GLUE 2: Evolution of the Computing Element Information Model

Authors: Balazs Konya¹ ; Laurence Field² ; Sergio Andreatozzi³ ; Stephen Burke⁴

¹ *Lund University*

² *CERN*

³ *INFN-CNAF*

⁴ *RAL*

Corresponding Author: sergio.andreatozzi@egi.eu

A key advantage of Grid systems is the capability of sharing heterogeneous resources and services across traditional administrative and organizational domains. This capability enables the creation of virtual pools of resources that can be assigned to groups of users. One of the problems that the utilization of such pools presents is the awareness of the resources, i.e., the fact that users or user agents need to have knowledge of the existence and state of the resources. This awareness requires the presence of a description of the services and resources typically defined via a community-agreed information model. One of the most popular information models is the GLUE Schema that is providing a common language to describe Grid resources for a number of Grid infrastructures. Other approaches exist undertaking different modeling strategies. The presence of different flavors of information models for Grid resources is a problem for enabling inter-Grid interoperability. In order to solve this problem, the GLUE Working Group in the context of the Open Grid Forum was started. The purpose of the group is a major redesign of the GLUE Schema that should consider: the successful modeling choices and flaws which have emerged from practical experience; use cases and

modeling choices of other information modeling initiatives. In this paper, we present the status of the new model for describing computing resources as the first outcome of the working group. The purpose is to disseminate the result and solicit feedback from the community.

Poster 1 - Board: 48 / 315

Towards a full implementation of a robust solution of a Domain Specific Visual Query Language DSVQL for HEP analysis

Authors: Vasco Amaral¹ ; Vasco Sousa¹

¹ UNL (*Universidade Nova de Lisboa*), Portugal

Corresponding Authors: patricia.conde.muino@cern.ch, vasco.amaral@di.fct.unl.pt

With the project PHEASANT a DSVQL was proposed for the purpose of providing a tools that could increase user's productivity while producing query code for data analysis. The previous project aimed at the proof concept and methodology feasibility by introducing the concept of DSLs. We are now concentrated on implementation issues in order to deploy a final tool.

The concept of domain specific languages has always been implicit in Software Engineering although the development of such languages was never done in a systematic way. The main goal of having DSLs is to rise the level of abstraction, as the main idea is to provide the final user (stakeholder) tools to reason and model the solution by using concepts of the problem domain instead of having to reason with concepts of the problem domain (meaning the implementation details like programming concepts and hardware restrictions). Once we have the model specified, we can use Model Driven Development and Software Product Lines techniques to deploy artifacts in a automatic way (meaning: software products, code, documentation etc).The SE community has been focusing its attention to methodologies and deploy tools for helping DSL developers in their effort to help productivity and efficiency at several application domains such as HEP. These tools start to mature and it worths having a look in order to avoid "redoing the wheel".

In this communication we will present the several technologies for DSLs meta-modeling studied in order to implement the DSVQL proposed by the PHEASANT project.

Poster 2 - Board: 105 / 440

Towards the Integration of StoRM on Amazon Simple Storage Service (S3)

Authors: Alberto Forti¹ ; Luca Magnoni¹ ; Riccardo Zappi¹ ; Sergio Andreozzi¹

¹ INFN-CNAF

Corresponding Author: riccardo.zappi@cnaf.infn.it

In Grid systems, a core resource being shared among geographically-dispersed communities of users is the storage. For this resource, a standard interface specification (Storage Resource Management or SRM) was defined and is being evolved in the context of the Open Grid Forum. By implementing this interface, all storage

resources part of a Grid could be managed in an homogenous fashion. In this work, we consider the extension of StoRM (STOrage Resource Manager, an implementation of SRM v2.2) in order to integrate a new type of storage resource: the Amazon Simple Storage Service (S3). Amazon S3 is a simple Web services interface offering access to the same highly scalable, reliable, fast, inexpensive data storage infrastructure that Amazon uses to run its own global network of Web sites. By performing this integration, we offer to the Grid community the capability to manage and access an incredible amount of storage resources freeing them from considering the costs associated with server maintenance, or whether they have enough storage available. The characteristics of StoRM are suitable for a smooth integration with Amazon S3. In particular, StoRM is designed to be easily adapted to the underlying storage resource via a plug-in mechanism, therefore a new plugin for integration with the Amazon S3 Web Service will be written. As regards the access policies, StoRM translates the Grid authorization rules into the Amazon S3 ones and applies them to the Amazon Web Services identity.

Event processing / 241

Track Reconstruction with the CMS Tracking Detector

Authors: Boris Mangano¹ ; Oliver Gutsche²

¹ *University of California, San Diego*

² *FERMILAB*

Corresponding Authors: boris.mangano@cern.ch, oliver.gutsche@cern.ch

With nominal collision energies of 14 TeV at luminosities of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, the LHC will explore energies an order of magnitude higher than colliders before. This poses big challenges for the tracking system and the tracking software to reconstruct tracks in the primary collision and the ~20 underlying events.

CMS has built a full silicon tracking system consisting of an inner pixel detector and an outer strip detector with over 200 m^2 of active area and over 70 million readout channels. The tracking software for the CMS tracking system has to master the dense environment in LHC collisions and also take into account multiple scattering. On average, a track has to transverse 13 layers of the silicon tracker.

An overview of the tracking system and the tracking software will be given. Both general and specialized tracking algorithms covering for example electron reconstruction will be discussed.

To prepare for the start of data taking expected at the end of 2007, CMS is conducting extensive cosmics tests with the tracking detector outside of the collision hall. An overview of the preliminary results of cosmic muon reconstruction with the CMS tracker using some of the previously described algorithms will be given.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS Tracker group

Event processing / 129

Track based alignment of the ATLAS inner detector

Author: Salvador Marti I Garcia¹

Co-authors: Jochen Schieck² ; Sergio Gonzalez-Sevilla³ ; Tobias GOETTERT²

¹ *Instituto de Fisica Corpuscular (IFIC) UV-CSIC*

² *Max-Planck-Institut fuer Physik Muenchen*

³ *Instituto de Fisica Corpuscular (IFIC) UV-CSIC*

Corresponding Authors: sergio.gonzalez.sevilla@cern.ch, salvador.marti@ific.uv.es

It is foreseen that the Large Hadron Collider will start its operations and collide proton beams during November 2007. ATLAS is one of the four LHC experiments currently under preparation. The alignment of the ATLAS tracking system is one of the challenges that the experiment must solve in order to achieve its physics goals. The tracking system comprises two silicon technologies: pixel and microstrip plus a transition radiation detector. The alignment of the system requires a the determination of more than 36000 degrees of freedom. The precision required for the most sensitive coordinate of the devices is of the order of few microns. This precision should be attained with a track based alignment and from the application of complex alignment algorithms. They require an extensive CPU and memory usage as large matrix inversion and many iterations algorithms are used. The alignment algorithms have been already exercised on several challenges as a Combined Test Beam, Cosmic Ray runs (at the surface and in the pit) and large scale computing simulation of physics samples. This note reports on the methods, their computing requirements and its preliminary results.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Poster 1 - Board: 64 / 465

Track based software package for measurement of the energy deposited in the calorimeters of the ATLAS detector

Authors: Konstantinos Bachas¹ ; Samira Hassani²

¹ *Aristotle University of Thessaloniki*

² *CEA-Dapnia,SACLAY*

Corresponding Author: konstantinos.bachas@cern.ch

The measurement of the muon energy deposition in the calorimeters is an integral part of muon identification, track isolation and correction for catastrophic muon energy losses, which are the prerequisites to the ultimate goal of refitting the muon track using calorimeter information as well. To this end, an accurate energy loss measurement method in the calorimeters is developed which uses only Event Data Model tools and is used by the muon isolation tool in the official ATLAS software, in order to provide isolation related variables at the Event Summary Data level. The strategy of the energy deposition measurement of the track in the calorimeters is described. Inner Detector, or Muon Spectrometer tracks are extrapolated to each calorimeter compartment using existing tools, which take into account multiple scattering and bending due to the magnetic field. The energy deposited in each compartment is measured by summing-up cells, corrected for noise, inside a cone around the track of desired size. The results of the measured energy loss in the calorimeters with this method are validated with Monte Carlo single muon samples.

Poster 2 - Board: 55 / 264

Track reconstruction of real cosmic muon events with CMS tracker detector

Author: Piergiulio Lenzi¹

¹ *Dipartimento di Fisica*

Corresponding Author: piergiulio.lenzi@cern.ch

The first application of one of the official CMS tracking algorithm, known as Combinatorial Track Finder, on cosmic muon real data is described.

The CMS tracking system consists of a silicon pixel vertex detector and a surrounding silicon microstrip detector.

The silicon strip tracker consists of 10 barrel layers and 12 endcap disks on each side. The system is currently going through final assembly stage.

As the construction goes on, big parts of the detector are being powered, controlled and readout as a whole.

A cosmic ray trigger has been setup, thus providing the first real events recorded in the CMS tracker.

Reconstruction was performed with the combinatorial track finder algorithm, based on the Kalman filter technique for trajectory building and track fitting.

A dedicated algorithm for cosmic ray track's seeding has been developed, together with useful tools for the tracking performance analysis; the exercise has been very useful for the understanding and the improvement of the tracking algorithm.

We present the results concerning the tracking performances and the possible applications of cosmic ray tracking for the full CMS tracker commissioning and alignment.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS Tracker group **Summary:**

The first application of one of the official CMS tracking algorithm, known as Combinatorial Track Finder, on cosmic muon real data is described.

Online computing / 378

Trigger Selection Software for Beauty physics in ATLAS

Authors: John Baines¹ ; Maria Smizanska²

¹ *RAL UK*

² *Lancaster University*

Corresponding Authors: d.emelianov@rl.ac.uk, maria.smizanska@cern.ch, john.baines@stfc.ac.uk

The unprecedented rate of beauty production at the LHC will yield high statistics for measurements such as CP violation and Bs oscillation and will provide the opportunity to search for and study very rare decays, such as $B \rightarrow \mu\mu$. The trigger is a vital component for this work and must select events containing the channels of interest from a huge background in order to reduce the 40 MHz bunch crossing rate down to 100-200 Hz for recording, of which only a part will be assigned to B-physics. Requiring a single or di-muon trigger provides the first stage of the B-trigger selection. Track reconstruction is then performed in the Inner Detector, either using

the full detector, at initial luminosity, or within Regions of Interest identified by the first level trigger at higher luminosities. Based on invariant mass, combinations of tracks are selected as likely decay products of the channel of interest and secondary vertex fits are performed. Events are selected based on properties such as fit quality and invariant mass. We present fast vertex reconstruction algorithms suitable for use in the second level trigger and event filter (level three). We discuss the selection software and the flexible trigger strategies that will enable ATLAS to pursue a B-physics programme from the first running at a luminosity of about $10^{31} \text{ cm}^{-2}\text{s}^{-1}$ through to the design luminosity running at $10^{34} \text{ cm}^{-2}\text{s}^{-1}$.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ATLAS

Computer facilities, production grids and networking / 75

UK Grid Computing for High-Energy Physics

Author: Jeremy Coles¹

¹ RAL

Corresponding Author: j.coles@rl.ac.uk

Over the last few years, UK research centres have provided significant computing resources for many high-energy physics collaborations under the guidance of the GridPP project. This paper reviews recent progress in the Grid deployment and operations area including findings from recent experiment and infrastructure service challenges. These results are discussed in the context of how GridPP is dealing with the important areas of networking, data storage and user education and support. Throughout, the paper offers feedback on observed successes and problems with the gLite middleware and the experiment specific software which is required to make the Grid usable. The paper moves on to examine current thinking on how ready GridPP sites are for LHC startup with an emphasis on some of the remaining challenges such as meeting strict WLCG availability targets and the scaling of resources at many sites to levels well beyond previous levels. The paper ends with a discussion of the decisions being taken to ensure a stable service and project through LHC startup until GridPP funding ends in 2011.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

GridPP

Poster 1 - Board: 101 / 353

US CMS Tier-2 Computing

Author: Kenneth Bloom¹

¹ University of Nebraska-Lincoln

Corresponding Author: kenbloom@unl.edu

The CMS computing model relies heavily on the use of “Tier-2” computing centers. At LHC startup, the typical Tier-2 center will have 1 MSpecInt2K of CPU resources, 200 TB of disk for data storage, and a WAN connection of at least 1 Gbit/s. These centers will be the primary sites for the production of large-scale simulation samples and for the hosting of experiment data for user analysis –

an interesting mix of experiment-controlled and user-controlled tasks. As a result, there are a wide range of services that must be deployed and commissioned at these centers, which are responsible for tasks such as dataset transfer, management of datasets, hosting of jobs submitted through Grid interfaces, and several varieties of monitoring. We discuss the development of the seven CMS Tier-2 computing centers in the United States, with a focus on recent operational performance and preparations for the start of data-taking at the end of 2007.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS

Grid middleware and tools / 254

Unified Storage Systems for Distributed Tier-2 Centres

Authors: Andrew Elwell¹ ; Graeme A Stewart¹ ; Greig A Cowan²

¹ *University of Glasgow*

² *University of Edinburgh*

Corresponding Authors: greig.cowan@cern.ch, graeme.andrew.stewart@cern.ch

The start of data taking this year at the Large Hadron Collider will herald a new era in data volumes and distributed processing in particle physics. Data volumes of 100s of Terabytes will be shipped to Tier-2 centres for analysis by the LHC experiments using the Worldwide LHC Computing Grid (WLCG).

In many countries Tier-2 centres are distributed between a number of institutes, e.g., the geographically spread Tier-2s of GridPP in the UK. This presents a number of challenges for experiments in terms of the use of such centres, as CPU and storage resources may be sub-divided and exposed in smaller units than the experiment would ideally want to work with. In addition, unhelpful mismatches between storage and CPU at the individual centres may be seen, which make efficient exploitation of a Tier-2's resources difficult.

One method of addressing this is to unify the storage across a distributed Tier-2, presenting the centres' aggregated storage as a single system. This greatly simplifies the data management for the VO, which then can access a greater amount of data across the Tier-2. However, such an approach will lead to scenarios where analysis jobs on one site's batch system must access data hosted on another site.

We investigate this situation using the Glasgow and Edinburgh clusters, which are part of the ScotGrid distributed Tier-2. In particular we look at how to mitigate the problems associated with "distant" data access and discuss the security implications of having LAN access protocols traverse the WAN between centres.

Collaborative tools / 415

University of Michigan Lecture Archiving

Author: Jeremy Herr¹

¹ *University of Michigan*

Corresponding Author: jeremy.herr@cern.ch

Large scientific collaborations as well as universities have a growing need for multimedia archiving of meetings and courses. Collaborations need to disseminate training and news to their wide-ranging members, and universities seek to provide their students with more useful studying tools. The University of Michigan ATLAS Collaboratory Project has been involved in the recording and archiving of multimedia lectures since 1999. Our software and hardware architecture has been used to record events for CERN, ATLAS, many units inside the University of Michigan, Fermilab, the American Physical Society and the Int'l Conference on Systems Biology at Harvard. Until 2006 our group functioned primarily as a tiny research/development team with special commitments to the archiving of certain ATLAS events. In 2006 we formed the MScribe project, using a larger scale, highly automated recording system to record and archive eight University courses in a wide array of subjects. Several robotic carts are wheeled around campus by unskilled student helpers to automatically capture and post to the Web audio, video, slides and chalkboard images. The advances the MScribe project has made in automation of these processes, including a robotic camera operator and automated video processing, are now being used to record ATLAS Collaboration events, making them available more quickly than before and enabling the recording of more events.

Poster 1 - Board: 52 / 373

Usage and Extension of Maven to build, release and distribute Java and Native Code Projects

Author: Mark Donszelmann¹

Co-authors: Max Turri¹ ; Tony Johnson¹ ; Victor Serbo¹

¹ *SLAC*

Corresponding Author: mark.donszelmann@slac.stanford.edu

Maven is a software project management and comprehension tool. Based on the concept of a project object model (POM), Maven can manage a project's build, reporting and documentation from a single XML file which declaratively specifies the project's properties. In short, Maven replaces Make or Ant, adds the handling of dependencies and generates documentation and a project website. Maven is an Open Source tool written in Java and mainly used for Java. It can easily be extended by writing common or project specific plugin modules.

One of Maven's main features is the handling of your project's (transitive) binary dependencies. A simple declaration that your project needs a library and that library gets automatically downloaded from a nearby mirror of the Maven repository and installed for local usage. While downloading platform-independent Java libraries is relatively straightforward, Maven did not originally provide any support for the harder problem of dealing with platform specific native code dependencies. We extended Maven with the Native ARchive Plugin (NAR) to allow it to compile and link Native (C, C++ and Fortran) code and handle Native libraries as dependencies.

In this talk we will describe the main features and benefits of Maven, how these benefits can be extended to native code using the NAR plugin, and describe how we use Maven as the project management tool for both Java and C++ components of the FreeHEP library.

Poster 1 - Board: 30 / 101

Usage of LSF for the batch farms at CERN

Author: Ulrich Schwickerath¹

¹ CERN

Corresponding Author: ulrich.schwickerath@cern.ch

LSF 7, the latest version of Platform's batch workload management system, addresses many issues which limited the ability of LSF 6.1 to support large scale batch farms, such as the lxbatch service at CERN. In this paper we will present the status of the evaluation and deployment of LSF 7 at CERN, including issues concerning the integration of LSF 7 with the gLite grid middleware suite and, in particular, the steps taken to endure an efficient reporting of the local batch system status and usage to the Grid Information System

Computer facilities, production grids and networking / 246

Use of Alternate Path WAN Circuits for CMS

Author: Philip DeMar¹

Co-authors: Andrey Bobyshev¹ ; Don Petravick¹ ; Matt Crawford¹ ; Maxim Grigoriev¹ ; Vyto Grigaliunas¹

¹ FERMILAB

Corresponding Author: demar@fnal.gov

Fermilab hosts the American Tier-1 Center for the LHC/CMS experiment. In preparation for the startup of CMS, and building upon extensive experience supporting TeVatron experiments and other science collaborations, the Laboratory has established high bandwidth, end-to-end (E2E) circuits with a number of US-CMS Tier2 sites, as well as other research facilities in the collaboration. These circuits provide preferred network paths for movement of high volumes of CMS data and represent a departure from the traditional approach of utilizing the general research and education (R&E) network infrastructure for movement of science data. All circuits are statically configured and are based on a variety of underlying network technologies. These circuits are presumed to provide, and generally do provide, more predictable performance, and they avoid the traffic contention concerns of general-use R&E network links. But the circuits also add significant complexity and effort for the Laboratory's wide area network support.

This presentation will discuss Fermilab's experiences with deploying, managing, and utilizing E2E circuits as preferred network paths in parallel with the general IP R&E network infrastructure. Alternate path routing techniques, monitoring issues, troubleshooting, and failover concerns will be covered. Issues of scalability and evolution toward dynamic circuits will also be discussed.

Poster 1 - Board: 83 / 214

Use of Cfengine for deploying LCG/gLite middle-ware

Author: Colin Morey¹

¹ *Manchester University*

Corresponding Author: moreyc@hep.man.ac.uk

Cfengine is a middle to high level policy language and autonomous agent for building expert systems to administrate and configure large computer clusters. It is ideal for large-scale cluster management and is highly portable across varying computer platforms, allowing the management of multiple architectures and node types within the same farm.

As well as being a highly capable configuration manager, Cfengine can also augment a network based service manager by providing an alerting mechanism which can monitor the system for abnormal behaviour, eg. watching and maintaining the permissions on files and directories, to monitoring available file-system space.

A number of sites within the GridPP collaboration and across the wider EGEE project are using Cfengine for the deployment and maintenance of their clusters and LCG/gLite middle-ware installations. This paper will collect together the experiences of these sites with both the initial setup and the ongoing maintenance of their Cfengine installations. This paper will also discuss some of the common implementation scenarios that may be encountered. This paper will also provide some solutions to common issues that the various sites have faced.

Poster 1 - Board: 90 / 250

Use of Flow Data for Traffic Analysis and Network Performance Characterization

Author: Andrey Bobyshev¹

Co-authors: Maxim Grigoriev¹ ; Philip DeMar¹ ; Vyto Grigaliunas¹

¹ *FERMILAB*

Corresponding Author: bobyshev@fnal.gov

At Fermilab, there is a long history of utilizing network flow data collected from site routers for various analyses, including network performance characterization, anomalous traffic detection, investigation of computer security incidents, network traffic statistics and others. Fermilab's flow analysis model is currently built as a distributed system that collects flow data from the site network border routers, as well as from internal core routers & aggregation switches. The flow data is complete, not sampled, with a daily volume of approximately 10GBytes. Despite the high volume of collected information, large scale analysis is conducted in near real-time to satisfy demands of the user community for timely availability of the analyzed data.

In this paper, we present Fermilab's Netflow Collection and Analysis system, as well as tools developed to analyze the flow data. Tools presented will include traffic characterization and network performance estimation for the US-CMS Tier1 Center, verification of path symmetry for network traffic re-routed over alternate path circuits, and profiling of traffic patterns for individual systems to characterize their typical behavior and enable identification of anomalous behavior.

Poster 2 - Board: 38 / 40

Use of GEANE for tracking in Virtual Monte Carlo

Author: Andrea Fontana¹

Co-authors: Alberto Panzarasa² ; Alberto Rotondi² ; Denis Bertini³ ; Lia Lavezzi² ; Mohammad Al-Turany³ ; Pablo Genova¹

¹ *INFN-Pavia*

² *Dept. of Nuclear and Theoretical Physics, University of Pavia*

³ *GSI*

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

The concept of Virtual Monte Carlo allows to use different Monte Carlo programs to simulate particle physics detectors without changing the geometry definition and the detector response simulation. In this context, to study the reconstruction capabilities of a detector, the availability of a tool to extrapolate the track parameters and their associated errors due to magnetic field, straggling in energy loss and Coulomb multiple scattering plays a central role: GEANE is an old program written in Fortran 15 years ago that performs this task through dense materials and that is still successfully used by many modern experiments in its native form. Among its features there are the capability to read directly the geometry and the magnetic field map from the simulation and to use different track representations. In this work we have “rediscovered” GEANE in the context of the Virtual Monte Carlo: the talk will show how GEANE has been integrated in the FairROOT framework, firmly based on the VMC, by keeping the old features in the new ROOT geometry modeler. Moreover new features have been added to GEANE to allow its use also for low density materials, i.e. for gaseous detectors, and preliminary results will be shown and discussed. The tool is now used by the PANDA and CBM collaborations at GSI as the first step for the global reconstruction algorithms, based on a Kalman filter which is currently under development.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

Panda Summary:

The use of GEANE as track follower for the Virtual Monte Carlo with the ROOT geometry modeler is shown. New features that extend the capabilities of GEANE to low density materials are also discussed. The tool is now used by the PANDA and CBM collaborations at GSI.

Poster 1 - Board: 102 / 363

User Centric Monitoring (UCM) information service for the next generation of Grid-enabled scientists

Authors: David Alexander¹ ; Jerome LAURET²

Co-authors: CHUANG LI¹ ; Valeri FINE²

¹ *Tech-X Corporation*

² *BROOKHAVEN NATIONAL LABORATORY*

Corresponding Author: jlauret@bnl.gov

Nuclear and high-energy physicists routinely execute data processing and data analysis jobs on a Grid and need to be able to monitor their jobs execution at an arbitrary site at any time. Existing Grid monitoring tools provide abundant

information about the whole system, but are geared towards production jobs and well suited for Grid administrators, while the information tailored towards an individual user is not readily available in a user-friendly and user-centric way. Such User Centric information includes monitoring information such as the status of the submitted job, queue position, time of the start/finish, percentage of being done, error messages, standard output, and reasons for failure.

We proposed to develop a framework centered on Grid service technology that allows scientists to track and monitor their jobs easily from a user-centric view. The proposed framework aims to be flexible so that it can be applied by any Grid Virtual Organization (VO) with various ways of collecting the user-centric job monitoring information built into the framework. Furthermore, the framework provides a rich and reusable set of methods of presenting the information to the user from within a Web browser and other clients.

In this presentation, we will give an architectural overview of the UCM service, show an example implementation in the RHIC/STAR experiment context and discuss limitations and future collaborative work.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

STAR

Poster 2 - Board: 20 / 295

Using Parrot to access the CDF software on LCG Grid sites

Authors: Aureliano Rama¹ ; Donatella Lucchesi² ; Douglas Thain³ ; Gabriele Compostella⁴ ; Igor Sfiligoi⁵ ; Simone Pagan Griso²

¹ *INFN of Padova*

² *University and INFN Padova*

³ *University of Notre Dame*

⁴ *University Of Trento INFN Padova*

⁵ *Fermilab*

Corresponding Author: donatella.lucchesi@pd.infn.it

When the CDF experiment was developing its software infrastructure, most computing was done on dedicated clusters. As a result, libraries, configuration files, and large executables were deployed over a shared file system.

As CDF started to move into the Grid world, the assumption of having a shared file system showed its limits. In a widely distributed computing model, such as the Grid, the CDF software will be not available natively on each worker node.

In order to overcome this problems CDF investigated several solutions and finally Parrot was adopted by LcgCAF, the CDF portal to the LCG Grid resources.

Parrot is a virtual filesystem for performing POSIX-like I/O on remote data services. It supports several protocols, including HTTP, FTP, RFIO, and other protocols common in Grid computing.

CDF has chosen to use the HTTP protocol, since it can be easily cached by Squid caches, already deployed at, or close to big Grid sites.

The current configuration used by CDF in production on LCG sites will be presented together with the different performance benchmarks, both with and without a local Squid. The experience and problems found by using Parrot in a system with several hundred of concurrent users will be

discussed.

Finally, the problem of cache coherence due to daily code updates will be analyzed and the possible solutions discussed

Poster 1 - Board: 51 / 364

Using ROOT with Modern Programming Languages

Author: Gordon Watts¹

¹ *University of Washington*

Corresponding Author: gwatts@uw.edu

ROOT is firmly based on C++ and makes use of many of its features – templates and multiple inheritance, in particular. Many modern languages like Java and C# and python are missing these features or have radically different implementations. These programming languages, however, have many advantages to offer scientists including improved programming paradigms, development environments, and full blown GUI development frameworks. Python is well served by the PyROOT project which gives full access to ROOT's capabilities. The bindings between ROOT and Python are built on the fly by the PyROOT infrastructure from the CINT dictionaries. This poster reports on progress towards implementing a similar infrastructure for the C# and the .NET family of languages with an eventual goal towards helping with physics analysis.

Poster 1 - Board: 4 / 119

Using the Grid for Large Scale and Nightly Testing of the ATLAS Trigger & Data Acquisition System

Authors: Alessandra Forti¹ ; Hegoi Garitaonandia Elejabarrieta² ; Jiri Masik¹ ; Sarah Wheeler³ ; Thorsten Wengler¹

¹ *University of Manchester*

² *Instituto de Fisica de Altas Energias (IFAE)*

³ *California University*

Corresponding Author: hegoi.garitaonandia@cern.ch

The ATLAS Trigger & Data Acquisition System has been designed to use more than 2000 CPUs. During the current development stage it is crucial to test the system on a number of CPUs of similar scale. A dedicated farm of this size is difficult to find, and can only be made available for short periods. On the other hand many large farms have become available recently as part of computing grids, leading to the idea of using them to test the TDAQ. However the task of adapting the TDAQ to run on the Grid is not trivial, as the TDAQ system requires full access to the computing resources it runs on and real-time interaction. Moreover the Grid virtualises the resources to present a common interface to the user. We will describe the implementation and first tests of a scheme that resolves these issues using a pilot job mechanism. The Tier2 cluster in Manchester was successfully used to run a full TDAQ system on 400 nodes using this implementation. Apart from the tests described above, this scheme also has great potential for other applications, like running Grid remote farms to perform detector calibration and monitoring in real-time, and automatic nightly testing of the TDAQ.

Poster 1 - Board: 72 / 114

VINCI : Virtual Intelligent Networks for Computing Infrastructures

Authors: Ciprian Dobre¹ ; Harvey Newman² ; Iosif Legrand² ; Ramiro Voicu²

¹ *Polytechnic University of Bucharest*

² *CALTECH*

Corresponding Authors: harvey.newman@cern.ch, iosif.legrand@cern.ch

The main objective of the VINCI project is to enable data intensive applications to efficiently use and coordinate shared, hybrid network resources, to improve the performance and throughput of global-scale grid systems, such as those used in high energy physics. VINCI uses a set of agent-based services implemented in the MonALISA framework to enable the efficient use of network resources, coordinated with computing and storage resources. VINCI is an integrated network service system that provides client authentication and authorization, discovery of services and the topology of connections, workflow scheduling, global optimization and monitoring.

The distributed agent system provides dynamically, on demand, end to end optical (or layer two VLAN) connections in less than one second independent of the location and the number of optical switches involved. It monitors and supervises all the created connections and is able to automatically generate an alternative path in case of connectivity errors. The alternative path is set up rapidly enough to avoid a TCP timeout, and thus to allow the transfer to continue uninterrupted.

Dedicated agents are used to monitor the client systems and to detect hardware and software configuration. They can perform end to end performance measurements and if necessary to configure the systems.

We are developing agents able to interact with GMPLS and CIENA's G.ASON network control plane protocols and to integrate this functionality into the network services provided by the VINCI framework.

Poster 2 - Board: 84 / 231

Virtual Organization Trustworthiness in the Grid world

Author: Irwin Gaines¹

Co-authors: Don Petravick¹ ; Igor Sfiligoi¹

¹ *FNAL*

Corresponding Author: gaines@fnal.gov

Computing in High Energy Physics and other sciences is quickly moving toward the Grid paradigm, with resources being distributed over hundreds of independent pools scattered over the five continents. The transition from a tightly controlled, centralized computing paradigm to a shared, widely distributed model, while bringing many benefits, has also introduced new problems, a major one being the handling of trust between participating parties.

The trust problem has been recognized since the beginning of the Grid movement, and a lot of thought has been put into developing the infrastructure for handling trust between resource providers and users. In particular, recognizing the size of the problem, the trust handling has been split into two pieces;

- a) between final users and Virtual Organizations (VOs), and
- b) between VOs and resource providers.

However, the above mentioned split has only been tackling the scalability issue, and very little thought has gone into understanding the trust relationship problems that a VO itself introduces. In particular, most VOs run dozens of services, many of them handling user binaries and user credentials.

Such services are obviously critical both for the final users as well as for the security health of the whole Grid; a compromised service could easily generate a major security incident. In spite of this, there is very little, if any, formal process in place to maintain the necessary level of trust.

This presentation will give an introduction to the problem of VO trust as well as an overview of the possible solutions.

Poster 1 - Board: 65 / 32

Virtualization applications at the Brookhaven Computing Facility

Authors: Alexander Withers¹ ; Christopher Hollowell¹ ; Tony Chan¹

¹ *BROOKHAVEN NATIONAL LAB*

Corresponding Author: tony@bnl.gov

The Brookhaven Computing Facility provides for the computing needs of the RHIC experiments, supports the U.S. Tier 1 center for the ATLAS experiment at the LHC and provides computing support for the LSST experiment. The multi-purpose mission of the facility requires a complex computing infrastructure to meet different requirements and can result in duplication of services with a large number of single-purpose servers for narrowly-defined applications. The facility is investigating potential applications of the open-source Xen virtualization package to allow the consolidation of services and servers. This presentation also discusses using Xen to virtualize the bulk of our Linux-based computing cluster. This is being integrated with Condor, the dCache-managed distributed storage system and the new multi-core CPU's to improve availability and increase effective usage of our facility resources by virtualizing a wide array of software support packages to meet the needs of various applications. Virtualization support (both hardware and software) can be an important element for efficient operations in a heterogeneous computing environment with increasing reliance on distributed computing models.

Poster 2 - Board: 10 / 46

WEB Based Online Event Displays for KASCADE-Grande

Authors: Andreas Weindl¹ ; Harald Schieler¹

¹ *FZ Karlsruhe / IK*

Corresponding Authors: schieler@ik.fzk.de, weindl@ik.fzk.de

The KASCADE-Grande experiment is a multi-detector installation at the site of the Forschungszentrum Karlsruhe, Germany, to measure and study extensive air showers induced in the atmosphere by primary cosmic rays in the energy range from 10^{14} to 10^{18} eV. For three of the detector components, WEB based online event displays have been implemented. They provide in a fast and simplified way actual information about energy deposits and arrival times of measured events and about the overall detector

status. Besides the aspect of being able to show air shower events to interested people wherever there is an internet access available, these event displays are an easy and highly useful tool for controlling and maintaining tasks from remote places. The event displays are designed as client-server applications, with the server running as independent part of the local data acquisition. Simplified event data are distributed via socket connections directly to the java applets acting as clients. These clients can run in any common browser on any computer somewhere on the planet.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

KASCADE-Grande

Plenary / 469

WLCG - Where we are now and the Real Data Challenges ahead

Corresponding Author: les.robertson@cern.ch

The talk will review the progress so far in setting up the distributed computing services for LHC data handling and analysis and look at some of the challenges we face when the real data begins to flow.

Grid middleware and tools / 240

WLCG scale testing during CMS data challenges

Author: Oliver Gutsche¹

¹ FERMILAB

Corresponding Author: oliver.gutsche@cern.ch

The CMS computing model to process and analyze LHC collision data follows a data-location driven approach and is using the WLCG infrastructure to provide access to GRID resources. As a preparation for data taking beginning end of 2007, CMS tests its computing model during dedicated data challenges.

Within the CMS computing model, user analysis plays an important role in the CMS computing strategy and poses a special challenge for the infrastructure with its random distributed access patterns. For this purpose, CMS developed the CMS Remote Analysis Builder (CRAB). CRAB handles all interactions with the WLCG infrastructure transparently for the user.

During the 2006 challenge, CMS set its goal to test the infrastructure at a scale of 50,000 user jobs per day using CRAB. Both direct submissions by individual users and automated submissions by robots were used to achieve this goal. A report will be given about the outcome of the user analysis part of the challenge and observations made during these tests using both the EGEE and OSG parts of the WLCG will be presented. The test infrastructure will be described and improvements made during the challenge to reach the target scale will be discussed.

In particular, the most prominent difference in the submission structure between both GRID middlewares will be discussed with regard to its impact on the challenge. EGEE uses a resource broker submission approach while OSG uses direct Condor-G submissions.

For 2007, CMS plans to increase the scale of the tests by a factor of 2. A report on work done in 2007 in the context of preparation for the summer 2007 data challenge will be given and first results will be presented.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

CMS Computing group

Poster 1 - Board: 50 / 359

Web System to support analysis for experimental equipment commissioning

Authors: Alexandre FARIA DE MELO¹ ; Andrea Dotti² ; Carmen MAIDANTCHIK¹ ; Felipe FINK GRAEL¹ ; Fernando GUIMARAES FERREIRA¹ ; Kaio KARAM GALVAO¹ ; Lawrence Price³

¹ *Universidade Federal do Rio de Janeiro (UFRJ)*

² *INFN*

³ *Argonne National Laboratory*

Corresponding Authors: andrea.dotti@pi.infn.it, kaio.karam.galvao@cern.ch, lawrence.price@cern.ch

During the ATLAS detector commissioning phase, installed readout electronics must pass performance standards tests. The resulting data must be analyzed to ensure correct operation. For the Tile Calorimeter, developers plug their code into a specific framework for physics data-processing. Collaboration members, taking shifts on commissioning work, interpret the results, in thousands of readout channels, to identify potential problems that may need correction during commissioning.

The Tile Commissioning Web System (TCWS) facilitates the repetitive data analysis and quality control by encapsulating all necessary steps to retrieve information, execute programs, access the outcomes, register statements, and verify the equipment status. TCWS integrates different applications, each presenting a particular view of the commissioning process. The TileComm Analysis application stores plots and analysis results, provides equipment-oriented visualization, collects information regarding equipment performance, and summarizes its status. The Timeline application provides equipment status history in a chronological way. The Web Interface for Shifters application supports monitoring tasks by managing test parameters, graphical views of the calorimeter performance, and information status of all equipment that was used in each test. Finally, equipment quality control data can be filled, stored, modified, and retrieved as hypertext forms through the ATLASMonitor application. These applications are also connected with other commissioning programs that allow an automatic gathering of the commissioning data.

This paper describes in detail the programs that compose the TCWS and how they are integrated within the Tile Calorimeter commissioning. Current status and future work are also discussed

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

Tile

Poster 2 - Board: 28 / 368

Xrootd in BaBar

Author: Wilko Kroeger¹

¹ SLAC

Corresponding Author: wilko@slac.stanford.edu

The BaBar Experiment stores its reconstructed event data in root files which amount to more than one petabyte and more than two million files. All the data are stored in the mass storage system (HPSS) at SLAC and part of the data is exported to Tier-A sites.

Fast and reliable access to the data is provided by Xrootd at all sites. It integrates with a mass storage system and files that are not on disk are automatically staged to disk if requested by a user. The built-in fault tolerance allows clients to survive failures, for example a crashed data server, and eases the maintenance of the cluster. With Xrootd it is easy to repair a broken machine without any impact on users.

At SLAC we have two Xrootd clusters. One read-only cluster with 59 data servers which is used by analysis jobs. The second is a production cluster with 16 data servers which is used for reading and writing files. More than 4000 cpus access these two clusters.

We will discuss the setup, usage and experience we have with maintaining Xrootd data clusters for a running experiment. We also will describe current development of connecting Xrootd clusters between Tier-A sites.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

BaBar

Distributed data analysis and information management / 336

ZEUS Grid Usage: Monte Carlo Production and Data Analysis

Authors: Hartmut Stadie¹ ; Krzysztof Wrona²

¹ Universitaet Hamburg

² DESY

Corresponding Author: stadie@mail.desy.de

The detector and collider upgrades for the HERA-II running at DESY have considerably increased the demand on computing resources for the ZEUS experiment. To meet the demand, ZEUS commissioned an automated Monte Carlo (MC) production capable of using Grid resources in November 2004. Since then, more than one billion events have been simulated and reconstructed on the Grid which corresponds to two thirds of the overall MC production in this period.

Based on the experience gained in this successful and efficient production on the Grid, a system has been developed for also running analysis jobs on the Grid. It allows standard ZEUS executables to be submitted to the Grid without any changes and with similar commands as used for the batch analysis. With this approach, the users can easily switch to running their analyses on the Grid and benefit from the additional resources without any Grid middleware knowledge.

We present the design and implementation of the Monte Carlo production

and the analysis systems which are both based on the ZEUS Grid-toolkit. Furthermore, we report on the status of these systems and our experience with large scale production on the Grid.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

ZEUS

Poster 1 - Board: 47 / 305

‘Anvil’ a case study of the Unified Software Development Process

Author: Simon Patton¹

Co-author: Christopher Day¹

¹ *LAWRENCE BERKELEY NATIONAL LABORATORY*

Corresponding Author: sjpatton@lbl.gov

The Unified Software Development Process (USDP) defines a process for developing software from the initial inception to the final delivery. The process creates a number of difference models of the final deliverable; the use case, analysis, design, deployment, implementation and test models. These models are developed using an iterative approach that breaks down into four main phases; inception, elaboration, construction and transition.

The ‘anvil’ project is required to produce the Experiment Control sub-system for the IceCube Neutrino Telescope based at the South Pole. This project used the USDP as the basis of its development. It turned out to be small enough and self-contained enough to act as a valuable case study of the application of the USDP in HEP software.

This paper shows how the various USDP models evolved during the development of the final deliverable. It also demonstrates the various phases of development and how these were used to mitigate risk. Finally, this paper discusses the final set of artifacts created by the USDP and presents a number of templates to help with their creation in future projects.

Summary:

Use the ‘anvil project, which is the Experiment Control sub-system for IceCube, as a worked example of the Unified Software Development Process.

Poster 1 - Board: 8 / 191

cMsg - A general purpose, publish-subscribe, interprocess communication implementation and framework

Author: Carl Timmer¹

Co-authors: David Abbott¹ ; Ed Jastrzembski¹ ; Elliott Wolin¹ ; Graham Heyes¹ ; Vardan Gyurjyan¹

¹ *TJNAF*

Corresponding Author: timmer@jlab.org

cMsg is software used to send and receive messages in the Jefferson Lab online and runcontrol systems. It was created to replace the several IPC

software packages in use with a single API. cMsg is asynchronous in nature, running a callback for each message received. However, it also includes synchronous routines for convenience. On the framework level, cMsg is a thin API layer in Java, C, or C++ that can be used to wrap most message-based interprocess communication protocols. The top layer of cMsg uses this same API and multiplexes user calls to one of many such wrapped protocols (or domains) based on a URL-like string which we call a Uniform Domain Locator or UDL. One such domain is a complete implementation of a publish-subscribe messaging system using network communications and written in Java (user APIs in C and C++ too). This domain is built in a way which allows it to be used as a proxy server to other domains (protocols). Performance is excellent allowing the system not only to be used for messaging but also as a data distribution system.

Poster 2 - Board: 71 / 106

dCache NFSv4.1 - true nfs server with SRM interface and HSM support

Author: Tigran Mkrtchyan Mkrtchyan¹

¹ *Deutsches Elektronen-Synchrotron DESY*

Corresponding Author: kofemann@gmail.com

Starting June 2007, all WLCG data management services have to be ready and prepared to move terabytes of data from CERN to the Tier 1 centers world wide, and from the Tier 1s to their corresponding Tier 2s. Reliable file transfer services, like FTS, on top of the SRM v2.2 protocol are playing a major role in this game. Nevertheless, moving large junks of data is only part of the challenge. As soon as the LHC experiments go online, thousands of physicists across the world will start data analysis, to provide first results as soon as possible. At that point in time, local file access becomes crucial. Currently, large numbers of local file access protocols are supported by various Storage Systems – dcap, gsidcap, rfio-dpm, rfio-castor, http and xrootd. A standard protocol, usable by any unmodified application, assuming POSIX data access, is highly desirable. The NFSv4.1 protocol, defined by IETF and implemented by various Operating System and Storage Box vendors, e.g. EMC, IBM, Linux, NetApp, Panasas and SUN, provides all necessary functionality: security mechanism negotiation (GSS-API, GSI, X509, UNIX), data access protocol negotiation (NFSv4 mandatory), clear distinction between metadata (namespace) and data access, support of multiple dataservers, ACLs, client and server crash recovery and much more. The client modules are being developed for AIX, Linux, and the Solaris kernels.

NFSv4.1 is an open standard, industry backed protocol which easily integrates into the dCache architecture. Together with the new namespace provider, Chimera, dCache provides a native NFSv4.1 implementation. At the most recent NFS “Bakeathon” at CITI-UMICH September 2006, dCache has proven to be compatible to all existing clients. At the time of this presentation, at least two production level installations are available. The one for the DESY, H1 experiment and for the Tier 2 center at University of Michigan. The official production version will be available as soon as the NFSv4.1 protocol has been released.

Grid middleware and tools / 242

dCache, the Challenge.

Author: Patrick Fuhrmann¹

¹ DESY

Corresponding Author: patrick.fuhrmann@desy.de

With the start of the Large Hadron Collider at CERN, end of 2007, the associated experiments will feed the major share of their data into the dCache Storage Element technology at most of the Tier I centers and many of the Tier IIs including the larger sites.

For a project, not having its center of gravity at CERN, and receiving contributions from various loosely coupled sites in Europe and the US, this is certainly an organizational and financial challenge.

The presentation is meant to be the comprehensive overview of the numerous sub-projects collectively known as dCache.org. Beside a brief introduction into the dCache technology itself we will provide an insight into the project structure and the various organizations dCache contributes to and gets contributions by.

The technical part will focus on topics like the SRM 2.2 development, the integration of the improved file name space provider (Chimera), including the ACL plug-in, the pluggable authorization module (gPlazma), the various file location and scheduling mechanisms, the variety of dCache access protocols and many more. Last but not least we are eager to give an outlook on the upcoming short and mid term topics.

These are the generic posix data access to dCache via the nfs4.1 protocol, the refurbished tertiary storage connector module, the dCache back-to-back transfer mechanisms and the enhanced meta data storage technology ,provided by chimera.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

dCache.ORG

Poster 2 - Board: 67 / 56

gPlazma in dCache

Author: Ted Hesselroth¹

¹ *Fermi National Accelerator Laboratory*

Corresponding Author: tdh@fnal.gov

gPlazma is the authorization mechanism for the distributed storage system dCache. Clients are authorized based on a grid proxy and may be allowed various privileges based on a role contained in the proxy. Multiple authorization mechanisms may be deployed through gPlazma, such as legacy dcache-kpwd, grid-mapfile, grid-vorolemap, or GUMS. Site-authorization through SAZ is also supported. Services within dCache requesting authorization contact gPlazma through the dCache cell mechanism and receive a mapping of user credentials and a set of obligations which define the user's privilege.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS):

dCache Collaboration

Grid middleware and tools / 43

glExec - gluing grid computing jobs to the Unix world

Author: David Groep¹

Co-authors: Gerben Venekamp¹; Oscar Koeroo¹

¹ *NIKHEF*

Corresponding Author: davidg@nikhef.nl

The majority of compute resources in today's scientific grids are based on Unix and Unix-like operating systems. In this world, user and user-group management is based around the well-known and trusted concepts of 'user IDs' and 'group IDs' that are local to the resource; in contrast, the grid concepts of user and group management are centered around globally assigned user identities and VO membership and structures that are entirely independently of the resource where the actual work is done.

To this end gatekeepers have been deployed traditionally at the fabric boundary to translate grid identities to Unix user IDs – usually in the form of 'map files' that translate (many) grid identity names to (many or a few) Unix user IDs. New job submission frameworks, such as the (java-based) execution web services and the introduction of late binding of the user jobs in a grid-wide overlay network of 'pilot' jobs, push the fabric boundary ever further into the resource. This necessitates the introduction glExec, a secure and light-weight (and thereby auditable) credential mapping system, that can be run both on fabric boundary, as part of an execution web service, and on the worker node in a late-binding scenario.

In this contribution we describe the rationale for glExec, how it interacts with the site authorization and credential mapping frameworks such as LCAS, LCMAPS and GUMS, and how it can be used to improve site control and traceability in a pilot-job system.

Grid middleware and tools / 94

glideinWMS - A generic pilot-based Workload Management System

Author: Igor Sfiligoi¹

¹ *FNAL*

Grids are making it possible for Virtual Organizations (VOs) to run hundreds of thousands of jobs per day. However, the resources are distributed among hundreds of independent Grid sites. A higher level Workload Management System (WMS) is thus necessary.

glideinWMS is a pilot-based WMS, inheriting several useful features:

- 1) Late binding: Pilots are sent to all suitable Grid sites. Only once pilots start are real jobs selected for that resources. No forecasting is needed.
- 2) Reliability: A broken Grid site will either kill pilot jobs or pilots will detect the problem at startup. Real jobs only start on well-behaved resources.
- 3) Grid-wide fair share: The relative priorities between jobs of the same VO are set inside the WMS. Grid sites only manage priorities between different VOs.

glideinWMS is based on the Condor glidein concept, i.e. a regular Condor pool, with the Condor daemons (startd) being started by pilot jobs. The real jobs are vanilla, standard or MPI universe jobs.

glideinWMS is composed of Glidein Factories and VO Frontends, communicating using Condor ClassAds:

Factories publish the available Grid sites, Frontends match the Grid attributes to job attributes and publish a request for a stream of glideins to suitable Grid sites

* Factories pick up the requests and submit the glideins

A detailed description of the system will be presented, along with the currently deployed systems inside USCMS production and user analysis frameworks. Integration with frameworks of other VOs will also be presented, as well as the measured scalability limits.