

# CHEP04

Monday 27 September 2004 - Friday 01 October 2004

Interlaken, Switzerland



## Book of Abstracts



<BR>

<BR>

<BR>

<BR>

This document and all information concerning the conference programme can be accessed online via the CHEP04 web site: <http://www.chep04.org>. The information on the web will be updated in the event on any last minute changes in the programme. These will also be communicated directly at the conference.

<BR>

<BR>

<BR>

<BR><br>

<BR>

This document contains all abstracts of the talks and posters to be presented at CHEP 2004. They are ordered by <b>Abstract Number</b>. Information on when, and in which session, the presentation is due to be made is also given here for completeness. The Conference Programme should be consulted to view the complete schedule of talks and cross reference to this document made using the Abstract Number.



# Contents

Detector-independent vertex reconstruction toolkit (VERTIGO) . . . . .	1
Development of algorithms for cluster finding and track reconstruction in the forward muon spectrometer of ALICE experiment . . . . .	1
Central Reconstruction System on the RHIC Linux Farm in Brookhaven Laboratory . . . . .	2
PyBus – A Python Software Bus . . . . .	2
Experience producing simulated events for the DZero experiment on the SAM-Grid . . . . .	2
A distributed, Grid-based analysis system for the MAGIC telescope . . . . .	3
Multi-Terabyte EIDE Disk Arrays running Linux RAID5 . . . . .	4
AMS-02 Computing and Ground Data Handling . . . . .	4
Genetic Programming and its application to HEP . . . . .	5
Test of data transfer over an international network with a large RTT . . . . .	5
Building the LCG: from middleware integration to production quality software . . . . .	6
CMD-3 Project Offline Software Development . . . . .	7
Automated Tests in NICOS Nightly Control System . . . . .	7
Federating Grids: LCG meets Canadian HEPGrid . . . . .	8
OptorSim: a Simulation Tool for Scheduling and Replica Optimisation in Data Grids . . . . .	9
IGUANA Interactive Graphics Project: Recent Developments . . . . .	9
Performance of an operating High Energy Physics Data grid, D0SAR-grid . . . . .	10
Adaptive Multi-vertex fitting . . . . .	10
An Embedded Linux System Based on PowerPC . . . . .	11
The ALICE Experiment Control System . . . . .	11
IgProf profiling tool . . . . .	12
Composite Framework for CMS Applications . . . . .	13

Monitoring CMS Tracker construction and data quality using a grid/web service based on a visualization tool . . . . .	13
CMS Tracker Visualisation Tools . . . . .	14
The simulation for the ATLAS experiment: present status and outlook . . . . .	14
Resource Predictors in HEP Applications . . . . .	15
Tracking of long lived hyperons in silicon detector at CDF. . . . .	15
Performance analysis of Cluster File System on Linux . . . . .	16
The application of PowerPC/VxWorks to the read-out subsystem of the BESIII DAQ . . . .	16
The LHCb Configuration Database . . . . .	17
A kinematic and a decay chain reconstruction library . . . . .	17
Deterministic Annealing for Vertex Finding at CMS . . . . .	18
Using the reconstruction software, ORCA, in the CMS datachallenge . . . . .	18
Toward a Grid Technology Independent Programming Interface for HEP Applications . .	19
The CEDAR Project . . . . .	19
Raw Ethernet based hybrid control system for the automatic control of suspended masses in gravitational waves interferometric detectors . . . . .	20
Recent Developments in the ROOT I/O . . . . .	20
Job Submission & Monitoring on the PHENIX Grid* . . . . .	21
Aspect-Oriented Extensions to HEP Frameworks . . . . .	21
A Globally Distributed Real Time Infrastructure for World Wide Collaborations . . . . .	22
MonALISA: An Agent Based, Dynamic Service System to Monitor, Control and Optimize Grid based Applications. . . . .	23
MONARC2: A Processes Oriented, Discrete Event Simulation Framework for Modelling and Design of Large Scale Distributed Systems. . . . .	24
New distributed offline processing scheme at Belle . . . . .	24
Managing software licences for a large research laboratory . . . . .	25
Domain Specific Visual Query Language for HEP analysis or How far can we go with user friendliness? . . . . .	25
The introduction to BES computing environment . . . . .	26
The Network Security Protection System at IHEP-Net . . . . .	26
RDBC: ROOT DataBase Connectivity . . . . .	27
Carrot ROOT Apache Module . . . . .	27

A Gaussian-sum filter for vertex reconstruction . . . . .	28
Usage statistics and usage patterns on the NorduGrid: Analyzing the logging information collected on one of the largest production Grids of the world. . . . .	28
Applications of the FLUKA Monte Carlo code in High Energy and Accelerator Physics . . . . .	29
Authentication/Security services in the ROOT framework . . . . .	30
Event Data Model in ATLAS . . . . .	31
Vertex finding and B-tagging algorithms for the ATLAS Inner Detector . . . . .	31
Storage Resource Manager . . . . .	32
The Athena Control Framework in Production, New Developments and Lessons Learned . . . . .	32
The Track Extrapolation package in the new ATLAS Tracking Realm . . . . .	33
Test of the ATLAS Inner Detector reconstruction software using combined test beam data . . . . .	33
Simulation and reconstruction of heavy ion collisions in the ATLAS detector. . . . .	33
BFD: A software management & deployment tool for mixed-language distributed projects . . . . .	34
Testing the CDF Distributed Computing Framework . . . . .	34
Experience integrating a General Information System API in LCG Job Management and Monitoring Services . . . . .	35
ATLAS Distributed Analysis . . . . .	36
Experiment Software Installation experience in LCG-2 . . . . .	36
Implementation and Performance of the High-Level Trigger electron and photon selection for the ATLAS experiment at the LHC . . . . .	37
AutoBlocker: A system for detecting and blocking of network scanning based on analysis of netflow data. . . . .	38
The DAQ system for the Fluorescence Detectors of the Pierre Auger Observatory . . . . .	38
Transparently managing time varying conditions and detector data on ATLAS. . . . .	39
Networking for High Energy and Nuclear Physics as Global E-Science . . . . .	39
A Scalable Grid User Management System for Large Virtual Organization . . . . .	40
Grid Deployment Experiences: The path to a production quality LDAP based grid information system . . . . .	41
XNNetFile, a fault tolerant extension of ROOT TNetFile . . . . .	41
Production of simulated events for the BaBar experiment by using LCG . . . . .	42
The Configurations Database Challenge in the ATLAS DAQ System . . . . .	42
The TAG Collector - A Tool for Atlas Code Release Management . . . . .	43

ATLAS Metadata Interfaces (AMI) and ATLAS Metadata Catalogs . . . . .	44
Panoramix . . . . .	44
Tier-1 and Tier-2 Real-time Analysis experience in CMS Data Challenge 04 . . . . .	45
CMS Software Installation . . . . .	45
64-Bit Opteron systems in High Energy and Astroparticle Physics . . . . .	46
Don Quijote - Data Management for the ATLAS Automatic Production System . . . . .	46
FAMOS, a FAsT MOnte-Carlo Simulation for CMS . . . . .	47
Network Architecture: lessons from the past, vision for the future . . . . .	47
HEP Applications Experience with the European DataGrid Middleware and Testbed . . . . .	48
GROSS: an end user tool for carrying out batch analysis of CMS data on the LCG-2 Grid. . . . .	49
Simplified deployment of an EDG/LCG cluster via LCFG-UML . . . . .	50
InGRID - Installing GRID . . . . .	50
An Object-Oriented Simulation Program for CMS . . . . .	50
Mantis: the Geant4-based simulation specialization of the CMS COBRA framework . . . . .	52
Jefferson Lab Data Acquisition Run Control System . . . . .	52
Cluster architectures used to provide CERN central CVS services . . . . .	53
Optimizing Selection Performance on Scientific Data by utilizing Bitmap Indices . . . . .	53
The ATLAS Computing Model . . . . .	54
Software agents in data and workflow management . . . . .	54
Using Tripwire to check cluster system integrity . . . . .	55
Using Nagios for intrusion detection . . . . .	56
Experience with CORBA communication middleware in the ATLAS DAQ . . . . .	56
The deployment mechanisms for the ATLAS software. . . . .	57
The Atlantis event visualisation program for the ATLAS experiment . . . . .	58
Guidelines for Developing a Good GUI . . . . .	58
ROOT Graphical User Interface . . . . .	59
Grid3: An Application Grid Laboratory for Science . . . . .	59
How to build an event store - the new Kanga Event Store for BaBar . . . . .	60
Fast tracking for the ATLAS LVL2 Trigger . . . . .	60
Recent evolutions of CMT. Multi-project and activity management. . . . .	61



The SEAL Component Model . . . . .	61
Breaking the 1 GByte/sec Barrier? High speed WAN data transfers for science . . . . .	62
A database prototype for managing computer systems configurations . . . . .	62
Grid Enabled Analysis : Architecture, prototype and status . . . . .	63
Porting CLEO software to Linux . . . . .	64
The Clarens Grid-enabled Web Services Framework: Services and Implementation . . . . .	64
Experience with Deployment and Operation of the ATLAS Production System and the Grid3+ Infrastructure at Brookhaven National Lab . . . . .	65
Operating the LCG and EGEE Production Grids for HEP . . . . .	65
Distributed Filesystem Evaluation and Deployment at the US-CMS Tier-1 Center . . . . .	66
Secure Grid Data Management Technologies in ATLAS . . . . .	66
Managed Data Storage and Data Access Services for Data Grids . . . . .	67
The Virtual Organization Membership Service eXtension project (VOX) . . . . .	68
The Open Science Grid (OSG) . . . . .	69
G-PBox: a Policy Framework for Grid Environments . . . . .	69
The CMS User Analysis Farm at Fermilab . . . . .	70
Lattice QCD Data and Metadata Archives at Fermilab and the International Lattice Data Grid . . . . .	70
Specifying Selection Criteria using C++ Expression Templates . . . . .	71
EventStore: Managing Event Versioning and Data Partitioning using Legacy Data Formats . . . . .	72
Paths: Specifying Multiple Job Outputs via Filter Expressions . . . . .	72
Implementation of a reliable and expandable on-line storage for compute clusters . . . . .	73
Parallel implementation of Parton String Model event generator . . . . .	73
FroNtier: High Performance Database Access Using Standard Web Components in a Scal- able Multi-tier Architecture . . . . .	74
Database Usage and Performance for the Fermilab Run II Experiments . . . . .	74
ScotGrid: A prototype Tier 2 centre . . . . .	75
Mass Storage Management and the Grid . . . . .	75
Experience with Real Time Event Reconstruction Farm for Belle Experiment . . . . .	76
The evolution of the distributed Event Reconstruction Control System in BaBar . . . . .	76

New compact hierarchical mass storage system at Belle realizing a peta-scale system with inexpensive ice-raid disks and an S-ait tape library . . . . .	77
ATLAS Detector Description Database Architecture . . . . .	77
OpenPAW . . . . .	78
SRB system at Belle/KEK . . . . .	78
Performance of the ATLAS DAQ DataFlow system . . . . .	79
On Distributed Database Deployment for the LHC Experiments . . . . .	80
The ATLAS DAQ system . . . . .	81
Rio . . . . .	81
Integrating Mutiple PC Farms into an uniform computing System with Maui . . . . .	82
The SEAL C++ Reflection System . . . . .	83
A Hardware Based Cluster Control And Management System . . . . .	83
Evaluation of Grid Security Solutions using Common Criteria . . . . .	84
A database perspective on CMS detector data . . . . .	84
Beyond Persistence: Developments and Directions in ATLAS Data Management . . . . .	85
Muon Reconstruction Software in CMS . . . . .	85
HEP@HOME - A distributed computing system based on BOINC . . . . .	86
Fast reconstruction of tracks in the inner tracker of the CBM experiment . . . . .	87
CASTOR: Operational issues and new Developments . . . . .	87
Production Management Software for the CMS Data Challenge . . . . .	88
dCache, LCG Storage Element and enhanced use cases . . . . .	88
The role of legacy services within ATLAS DC2 . . . . .	89
CMS Detector Description: New Developments . . . . .	89
Future processors: What is on the horizon for HEP farms? . . . . .	90
H1OO - an analysis framework for H1 . . . . .	91
CERN's openlab for Datagrid applications . . . . .	91
Status and Plans of the LCG PI Project . . . . .	92
Software management infrastructure in the LCG Application Area . . . . .	92
A general and flexible framework for virtual organization application tests in a grid system . . . . .	93

Validation of the GEANT4 Bertini Cascade model and data analysis using the Parallel ROOT Facility on a Linux cluster . . . . .	93
Production Experience of the Storage Resource Broker in the BaBar Experiment . . . . .	94
Volume-based representation o the magnetic field . . . . .	94
Middleware for the next generation Grid infrastructure . . . . .	95
LCIO persistency and data model for LC simulation and reconstruction . . . . .	96
The GridSite authorization system . . . . .	96
Installing and Operating a Grid Infrastructure at DESY . . . . .	97
Portable Gathering System for Monitoring and Online Calibration at Atlas . . . . .	98
New experiences with the ALICE High Level Trigger Data Transport Framework . . . . .	98
Experiences with Data Indexing services supported by the NorduGrid middleware . . . . .	99
A Control Software for the ALICE High Level Trigger . . . . .	99
Ion transport simulation using Geant4 hadronic physics . . . . .	100
The Binary Cascade . . . . .	101
Porting LCG Applications . . . . .	101
Control and state logging for the PHENIX DAQ System . . . . .	102
Adding Kaons to the Bertini Cascade Model . . . . .	102
CHIPS based hadronization of quark-gluon strings . . . . .	103
Addressing the persistency patterns of the time evolving HEP data in the ATLAS/LCG MySQL Conditions Databases . . . . .	103
The WEB interface for the ATLAS/LCG MySQL Conditions Databases and performance constraints in the visualisation of extensive scientific/technical data . . . . .	104
Conditions Databases: the interfaces between the different ATLAS systems . . . . .	105
New Applications of PAX in Physics Analyses at Hadron Colliders . . . . .	105
Testbed Management for the ATLAS TDAQ . . . . .	106
GoToGrid - A Web-Oriented Tool in Support to Sites for LCG Installations . . . . .	107
Developments of Mathematical Software Libraries for the LHC experiments . . . . .	107
”RecPack”, a general reconstruction tool-kit . . . . .	108
A Pattern-based Continuous Integration Framework for Distributed EGEE Grid Middleware Development . . . . .	108
Job Interactivity using a Steering Service in an Interactive Grid Analysis Environment . . . . .	109

Job Monitoring in Interactive Grid Analysis Environment . . . . .	110
Predicting Resource Requirements of a Job Submission . . . . .	110
Design and Implementation of a Notification Model for Grid Monitoring Events . . . . .	111
Offline Software for the ATLAS Combined Test Beam . . . . .	112
Distributed Testing Infrastructure and Processes for the EGEE Grid Middleware . . . . .	112
Managing third-party software for the LCG . . . . .	113
Evolution of LCG-2 Data Management . . . . .	113
The GeoModel Toolkit for Detector Description . . . . .	114
Update On the Status of the FLUKA Monte Carlo Transport Code . . . . .	114
Migrating PHENIX databases from object to relational model . . . . .	115
INDICO - the software behind CHEP 2004 . . . . .	116
The Description of the Atlas Detector . . . . .	116
On-demand Layer VPN Support for Grid Applications . . . . .	117
Super scaling PROOF to very large clusters . . . . .	118
Parallel compilation of CMS software . . . . .	118
LHC data files meet mass storage and networks: going after the lost performance . . . . .	119
Automatic Procedures as Generated Analysis Tool . . . . .	119
IceTray: a Software Framework for IceCube . . . . .	120
JIM Deployment for the CDF Experiment . . . . .	121
Data management services of NorduGrid . . . . .	121
Realization of a stable network flow with high performance communication in high bandwidth- delay product network . . . . .	122
A Multidimensional Approach to the Analysis of Grid Monitoring Data . . . . .	123
New specific solids definitions in the Geant4 geometry modeller . . . . .	123
ROOT : detector visualization . . . . .	124
The ARDA Prototypes . . . . .	124
HEPBook - A Personal Collaborative HEP notebook . . . . .	125
Control in the ATLAS TDAQ System . . . . .	126
The ROOT Linear Algebra . . . . .	126

The Design, Installation and Management of a Tera-Scale High Throughput Cluster for Particle Physics Research . . . . .	127
Experiences with the gLite Grid Middleware . . . . .	127
Use of Condor and GLOW for CMS Simulation Production . . . . .	128
ARDA Project Status Report . . . . .	128
A GRID approach for Gravitational Waves Signal Analysis with a Multi-Standard Farm Prototype . . . . .	129
Dynamic Matched Filters for Gravitational Waves Detection . . . . .	130
OpenScientist. Status of the project. . . . .	130
CLHEP Infrastructure Improvements . . . . .	130
Remote Shifting at the CLEO Experiment . . . . .	131
Geant4: status and recent developments . . . . .	131
Synergia: A Modern Tool for Accelerator Physics Simulation . . . . .	132
CDB - Distributed Conditions Database of BaBar Experiment . . . . .	133
Linux for the CLEO-c Online system . . . . .	133
The STAR Unifid Meta-Scheduler project, a front end around evolving technologies for user analysis and data production. . . . .	134
Grid Collector: Using an Event Catalog to Speed up User Analysis in Distributed Environment . . . . .	134
Overview and new developments in Geant4 electromagnetic physics . . . . .	135
SPHINX: A Scheduling Middleware for Data Intensive Applications on a Grid . . . . .	135
Extending EGS with SVG for Track Visualization . . . . .	136
From Geant 3 to Virtual Monte Carlo: Approach and Experience . . . . .	137
Improving Standard C++ for the Physics Community . . . . .	137
Disk storage technology for the LHC T0/T1 centre at CERN . . . . .	138
An intelligent resource selection system based on neural network for optimal application performance in a grid environment . . . . .	138
The Next Generation Root File Server . . . . .	139
The PHENIX Event Builder . . . . .	139
A Regional Analysis Center at the University of Florida . . . . .	140
Integration of ATLAS Software in the Combined Beam Test . . . . .	141

Production data export and archiving system for new data format of the BaBar experiment. .....	141
A statistical toolkit for data analysis .....	141
Production mode Data-Replication framework in STAR using the HRM Grid .....	142
Tools for GRID deployment of CDF offline and SAM data handling systems for Summer 2004 computing. ....	143
Comparative study of the power of goodness-of-fit algorithms .....	144
Alibaba: A heterogeneous grid-based job submission system used by the BaBar experiment .....	144
BaBar Bookkeeping - a distributed meta-data catalog of the BaBar event store. ....	145
BaBar simulation production - A millennium of work in under a year .....	145
A Dynamically Reconfigurable Data Stream Processing System .....	146
Patriot: Physics Archives and Tools required to Investigate Our Theories .....	147
Precision validation of Geant4 electromagnetic physics .....	147
Chimera - a new, fast, extensible and Grid enabled namespace service .....	148
Storage Resource Managers at Brookhaven .....	149
Lattice QCD Clusters at Fermilab .....	149
The new BaBar Analysis Model .....	150
Wide Area Network Monitoring system for HEP experiments at Fermilab .....	150
Distributed computing and oncological radiotherapy: technology transfer from HEP and experience with prototype systems .....	151
The BABAR Analysis Task Manager .....	152
Precision electromagnetic physics in Geant4: the atomic relaxation models .....	152
Experience with the Unified Process .....	153
The Offline Framework of the Pierre Auger Observatory .....	153
The Geometry Package for the Pierre Auger Observatory .....	154
The ZOOM Minimization Package .....	155
The GANGA user interface interface for physics analysis on distributed resources .....	155
SRM AND GFAL TESTING FOR LCG2 .....	156
LambdaStation: A forwarding and admission control service to interface production net- work facilities with advanced research network paths .....	157
Muon Event Filter Software for the ATLAS Experiment at LHC .....	157

Quality Assurance and Testing in LCG . . . . .	158
Data Rereprocessing on Worldwide Distributed Sytems . . . . .	158
On the Management of Certification Authority in Large Scale GRID Infrastructure . . . . .	159
LEXOR, the LCG-2 Executor for the ATLAS DC2 Production System . . . . .	160
DIRAC Workload Management System . . . . .	160
Globally Distributed User Analysis Computing at CDF . . . . .	161
Supporting the Development Process of the DataGrid Workload Management System Software with GNU autotools, CVS and RPM . . . . .	162
DIRAC Lightweight information and monitoring services using XML-RPC and Instant Messaging . . . . .	162
WAN Emulation Development and Testing at Fermilab . . . . .	163
Methodologies and techniques for analysis of network flow data . . . . .	164
The D0 Virtual Center and planning for large scale computing . . . . .	164
Run II computing . . . . .	165
Operation of the CERN Managed Storage environment; current status and future directions . . . . .	165
Geant4 as Simulation Toolkit addressed to interplanetary manned missions studies: required developments and improvements . . . . .	166
DIRAC - The Distributed MC Production and Analysis for LHCb . . . . .	166
Pixel Reconstruction in the CMS High-Level Trigger . . . . .	167
Python-based physics analysis environment for LHCb . . . . .	168
GILDA: a Grid for dissemination activities . . . . .	168
POOL Development Status and Plans . . . . .	169
POOL Integration into three Experiment Software Frameworks . . . . .	169
Experience with POOL from the LCG Data Challenges of the three LHC experiments . . . . .	170
The LCG Savannah software development portal . . . . .	171
Cross Experiment Workflow Management: The Runjob Project . . . . .	171
Self-Filling Histograms: An object-oriented analysis framework . . . . .	172
The Virtual Geometry Model . . . . .	172
A Lightweight Monitoring and Accounting System for LHCb DC'04 Production . . . . .	173
Deploying and operating LHC Computing Grid 2 (LCG2) During Data Challenges . . . . .	174

The Condor based CDF CAF . . . . .	174
The FEDRA - Framework for Emulsion Data Reconstruction and Analysis in OPERA experiment. . . . .	175
File-Metadata Management System for the LHCb Experiment . . . . .	175
Development and use of MonALISA high level monitoring services for the star unified Meta-Scheduler . . . . .	176
Using HEP Systems to Provide Storage for Biologists . . . . .	177
WIRED 4 - A generic Event Display plugin for JAS 3 . . . . .	177
Job-monitoring over the Grid with GridIce infrastructure. . . . .	178
AIDA, JAIDA and AIDAJNI: Data Analysis using interfaces . . . . .	179
JASSimApp plugin for JAS3: Interactive Geant4 GUI . . . . .	179
The FreeHEP Java Library Root IO package . . . . .	180
Interactive Data Analysis on the Grid using Globus 3 and JAS3 . . . . .	180
Writing Extension Modules (Plug-ins) for JAS3 . . . . .	181
Results of the LHCb experiment Data Challenge 2004 . . . . .	182
Visualisation of the ROOT geometries (TVolume and TGeoVolume) with Coin-based model. . . . .	182
Full Event Reconstruction in Java . . . . .	183
Monte Carlo Event Generation in a Multilanguage, Multiplatform Environment . . . . .	183
LCG Generator . . . . .	184
Status of the alignment calibrations in the ATLAS-Muon experiment . . . . .	184
The High Level Filter of the H1 Experiment at HERA . . . . .	185
XML I/O in ROOT . . . . .	186
Data Management in EGEE . . . . .	186
Software Management in the HARP experiment . . . . .	187
Information and Monitoring Services within a Grid Environment . . . . .	187
Go4 analysis design . . . . .	188
Distributed Tracking, Storage, and Re-use of Job State Information on the Grid . . . . .	188
Practical approaches to Grid workload and resource management in the EGEE project . . . . .	189
The ZEUS Global Tracking Trigger Barrel Algorithm . . . . .	190
The Architecture of the ZEUS Second Level Global Tracking Trigger . . . . .	191



The ZEUS Global Tracking Trigger . . . . .	191
The AliRoot framework, status and perspectives . . . . .	191
GraXML . . . . .	192
Aspects . . . . .	192
Persistence for Analysis Objects . . . . .	193
Global Distributed Parallel Analysis using PROOF and AliEn . . . . .	194
Usage of ALICE Grid middleware for medical applications . . . . .	194
Fault Tolerance and Fault Adaption for High Performance Large Scale Embedded Systems	195
The Virtual MonteCarlo : status and applications . . . . .	195
The ALICE High Level Trigger . . . . .	196
Track reconstruction in high density environment . . . . .	197
The LCG-AliEn interface, a realization of a MetaGrid system . . . . .	197
Simulations and Prototyping of the LHCb L1 and HLT Triggers . . . . .	198
Participation of Russian sites in the Data Challenge of ALICE experiment in 2004 . . . .	199
Building Global HEP Systems on Kerberos . . . . .	199
The architecture of the AliEn system . . . . .	200
D0 data processing within EDG/LCG . . . . .	200
AliEn Web Portal . . . . .	201
The High Level Trigger software for the CMS experiment . . . . .	201
Bayesian Approach for Combined Particle Identification in ALICE Experiment at LHC . .	202
ALICE Multi-site Data Transfer Tests on a Wide Area Network . . . . .	203
LCG Conditions Database Project Overview . . . . .	203
Harp data and software migration from Objectivity to Oracle . . . . .	204
DZERO Data Acquisition Monitoring and History Gathering . . . . .	205
Bringing High-Performance Networking to HEP Users . . . . .	205
SAMGrid Monitoring Service and its Integration with MonALisa . . . . .	206
Application of the SAMGrid Test-Harness for Performance Evaluation and Tuning of a Distributed Cluster Implementation of Data Handling Services . . . . .	206
K5 @ INFN.IT: an infrastructure for the INFN cross REALM & AFS cell authentication. . .	207

Gfarm v2: A Grid file system that supports high-performance distributed and parallel data computing . . . . .	208
SAMGrid Integration of SRMs . . . . .	208
StoRM: grid middleware for disk resource management . . . . .	209
The SAMGrid Database Server Component: Its Upgraded Infrastructure and Future Development Path . . . . .	210
Mis-use Cases for the Grid . . . . .	211
The status of Fermilab Enstore Data Storage System . . . . .	211
Network Information and Management Infrastructure Project . . . . .	212
Online Monitoring and online calibration/reconstruction for the PHENIX experiment . . . . .	212
Boosting the data logging rates in run 4 of the PHENIX experiment . . . . .	213
Deployment of SAM for the CDF Experiment . . . . .	213
A New STAR Event Reconstruction Chain . . . . .	214
Monitoring a Petabyte Scale Storage System . . . . .	214
Designing a Useful Email System . . . . .	215
Experiences Building a Distributed Monitoring System . . . . .	215
Patching PCs . . . . .	216
CHOS, a method for concurrently supporting multiple operating system . . . . .	216
The DZERO Run II Level 3 Trigger and Data Acquisition System . . . . .	216
Generic logging layer for the distributed computing . . . . .	217
SAMGrid Experiences with the Condor Technology in Run II Computing . . . . .	217
A Level-2 trigger algorithm for the identification of muons in the Atlas Muon Spectrometer . . . . .	218
A high-level language for specifying detector coincidences . . . . .	219
Monitoring the CDF distributed computing farms . . . . .	219
Enterasys - Networks that Know . . . . .	220
A Data Grid for the Analysis of Data from the Belle Experiment . . . . .	220
InfiniBand for High Energy Physics . . . . .	221
CERN Modular Physics Screensaver or Using spare CPU cycles of CERN's Desktop PCs . . . . .	221
Current Status of Fabric Management at CERN . . . . .	222
Grid2003 Monitoring, Metrics, and Grid Cataloging System . . . . .	223

The BIRN Project: Distributed Information Infrastructure and Multi-scale Imaging of the Nervous System (BIRN = Biomedical Informatics Research Network) . . . . .	223
Computing for Belle . . . . .	224
Physics validation of the simulation packages in a LHC-wide effort . . . . .	225
Reflection-Based Python-C++ Bindings . . . . .	225
The Project CampusGrid . . . . .	226
Developing & Managing a large Linux farm - the Brookhaven Experience . . . . .	226
Distributed Computing Grid Experiences in CMS DC04 . . . . .	227
Role of Tier-0, Tier-1 and Tier-2 Regional Centres in CMS DC04 . . . . .	228
Performance of the NorduGrid ARC and the Dulcinea Executor in ATLAS Data Challenge 2 . . . . .	229
Housing Metadata for the Common Physicist Using a Relational Database . . . . .	230
ATLAS Production System in ATLAS Data Challenge 2 . . . . .	231
BaBar computing - From collisions to physics results . . . . .	232
ATLAS Data Challenge Production on Grid3 . . . . .	232
Global Grid User Support for LCG . . . . .	233
Concepts and technologies used in contemporary DAQ systems . . . . .	233
The ALICE Data Challenge 2004 and the ALICE distributed analysis prototype . . . . .	234
The Evolving Wide Area Network Infrastructure in the LHC era . . . . .	234
Physics Validation of the LHC Software . . . . .	235
Summary - Online Computing . . . . .	235
Summary - Event Processing . . . . .	235
Summary - Core Software . . . . .	235
Summary - Distributed Computing Services . . . . .	236
Summary - Distributed Computing Systems and Experiences . . . . .	236
Summary - Computer Fabrics . . . . .	236
Summary - Wide Area Networking . . . . .	236
Conference Conclusions . . . . .	236
Conference Closeout . . . . .	236
Welcome Address . . . . .	236

50 years of Computing at CERN . . . . .	237
The LCG Project - Preparing for Startup . . . . .	237
The impact of e-science . . . . .	237
The role of scientific middleware in the future of HEP computing . . . . .	238
The Evolution of Computing: Slowing down? Not Yet! . . . . .	238
Grand Challenges facing Storage Systems . . . . .	238
Evolution and Revolution in the Design of Computers Based on Nanoelectronics . . . . .	238
The IBM Research Global Technology Outlook . . . . .	239
Griding The Nordic Supercomputing Infrastructure . . . . .	239
Grid Security . . . . .	240
Computing Models and Data Challenges of the LHC experiments . . . . .	240
Defining a Semantic Web Initiative for High Energy Physics . . . . .	240
INTAS Discussion . . . . .	241
EU Grid Research - Projects and Vision . . . . .	241
Huge Memory systems for data-intensive science . . . . .	242

Poster Session 3 - Board: 63 / 31

## Detector-independent vertex reconstruction toolkit (VERTIGO)

**Authors:** G. Richter<sup>1</sup>; W. Mitaroff<sup>1</sup>; W. Waltenberger<sup>1</sup>

<sup>1</sup> *Austrian Academy of Sciences // Institute of High Energy Physics*

**Corresponding Author:** mitaroff@hephy.oeaw.ac.at

A proposal is made for the design and implementation of a detector-independent vertex reconstruction toolkit and interface to generic objects (VERTIGO). The first stage aims at re-using existing state-of-the-art algorithms for geometric vertex finding and fitting by both linear (Kalman filter) and robust estimation methods. Prototype candidates for the latter are a wide range of adaptive filter algorithms being developed for LHC/CMS, as well as proven ones (like ZVTOP of SLC/SLD). In a second stage, also kinematic constraints will be included for the benefit of complex multi-vertex topologies.

The design is based on modern object-oriented techniques. A core (RAVE) is surrounded by a shell of abstract interfaces (using adaptors for access from/to the particular environment) and a set of analysis and debugging tools. The implementation follows an open source approach and is easily adaptable to future standards.

Work has started with the development of a specialized visualisation tool, following the model-view-controller (MVC) paradigm; it is based on COIN3D and may also include interactivity by PYTHON scripting. A persistency storage solution, intended to provide a general data structure, was originally based on top of ROOT and is currently being extended for AIDA and XML compliance; interfaces to existing or future event reconstruction packages are easily implementable. Flexible linking to a math library is an important requirement; at present we use CLHEP, which could be replaced by a generic product.

Poster Session 3 - Board: 13 / 32

## Development of algorithms for cluster finding and track reconstruction in the forward muon spectrometer of ALICE experiment

**Authors:** A. Vodopianov<sup>1</sup>; A. Zinchenko<sup>2</sup>; G. Chabratova<sup>1</sup>; V. Pismennaia<sup>1</sup>

<sup>1</sup> *Joint Institute for Nuclear Research, Dubna, Russia*

<sup>2</sup> *JOINT INSTITUTE FOR NUCLEAR RESEARCH (JINR), DUBNA, RUSSIA*

**Corresponding Author:** alexandre.zintchenko@cern.ch

A simultaneous track finding / fitting procedure based on Kalman filtering approach has been developed for the forward muon spectrometer of ALICE experiment.

In order to improve the performance of the method in high-background conditions of the heavy ion collisions the “canonical” Kalman filter has been modified and supplemented by a “smoother” part. It is shown that the resulting “extended” Kalman filter gives better tracking results and offers higher flexibility.

To further improve the tracking performance in a high occupancy environment a new algorithm for cluster / hit finding in cathode pad chambers of the muon spectrometer has been developed. It is based on the expectation maximization procedure for a shape deconvolution of overlapped clusters. It is demonstrated that the proposed method allows to reduce the loss of the coordinate reconstruction accuracy for high hit multiplicities and

achieve better tracking results.  
Both the hit finding and track reconstruction algorithms have been implemented within the AliRoot software framework.

**Poster Session 2 - Board: 65 / 33**

## Central Reconstruction System on the RHIC Linux Farm in Brookhaven Laboratory

**Authors:** A. Withers<sup>1</sup>; C. Hollowell<sup>2</sup>; G. Tsai<sup>1</sup>; R. Hogue<sup>2</sup>; R. Popescu<sup>1</sup>; T. Chan<sup>3</sup>; T. Wlodek<sup>2</sup>

<sup>1</sup> *Brookhaven National Laboratory*

<sup>2</sup> *Brookhaven National Lab*

<sup>3</sup> *BROOKHAVEN NATIONAL LAB*

**Corresponding Author:** tony@bnl.gov

A description of a Condor-based, Grid-aware batch software system configured to function asynchronously with a mass storage system is presented. The software is currently used in a large Linux Farm (2700+ processors) at the RHIC and ATLAS Tier 1 Computing Facility at Brookhaven Lab. Design, scalability, reliability, features and support issues with a complex Condor-based batch system are addressed within the context of a Grid-like, distributed computing environment.

**Poster Session 3 - Board: 29 / 34**

## PyBus – A Python Software Bus

**Author:** W. Lavrijsen<sup>1</sup>

<sup>1</sup> *LBNL*

**Corresponding Author:** wim.lavrijsen@cern.ch

A software bus, just like its hardware equivalent, allows for the discovery, installation, configuration, loading, unloading, and run-time replacement of software components, as well as channeling of inter-component communication. Python, a popular open-source programming language, encourages a modular design on software written in it, but it offers little or no component functionality. However, the language and its interpreter provide sufficient hooks to implement a thin, integral layer of component support. This functionality can be presented to the developer in the form of a module, making it very easy to use. This paper describes a Python module, PyBus, with which the concept of a 'software bus' can be realised in Python. It demonstrates, within the context of the Atlas software framework Athena, how PyBus can be used for the installation and (run-time) configuration of software, not necessarily Python modules, from a Python application in a way that is transparent to the end-user.

## Experience producing simulated events for the DZero experiment on the SAM-Grid

**Authors:** A. Nishandar<sup>1</sup>; G. Garzoglio<sup>2</sup>; I. Terekhov<sup>2</sup>; J. Snow<sup>3</sup>; S. Jain<sup>1</sup>

<sup>1</sup> *University of Texas at Arlington*

<sup>2</sup> *FERMI NATIONAL ACCELERATOR LABORATORY*

<sup>3</sup> *Langston University*

**Corresponding Author:** kennedy@fnal.gov

Most of the simulated events for the DZero experiment at Fermilab have been historically produced by the “remote” collaborating institutions. One of the principal challenges reported concerns the maintenance of the local software infrastructure, which is generally different from site to site. As the understanding of the community on distributed computing over distributively owned and shared resources progresses, it becomes increasingly interesting the adoption of grid technologies to address the production of monte-carlo events for high energy physics experiments. The SAM-Grid is a software system developed at Fermilab, which integrates standard grid technologies for job and information management with SAM, the data handling system of the DZero and CDF experiments. During the past few months, this grid system has been tailored for the monte-carlo production of DZero. Since the initial phase of deployment, this experience has exposed an interesting series of requirements to the SAM-Grid services, the standard middleware, the resources and their management and to the analysis framework of the experiment. As of today, the inefficiency due to the grid infrastructure has been reduced to as little as 1%. In this paper, we present our statistics and the “lesson learned” in running large high energy physics applications on a grid infrastructure.

**Distributed Computing Systems and Experiences / 43**

## A distributed, Grid-based analysis system for the MAGIC telescope

**Authors:** A. Forti<sup>1</sup>; A. deAngelis<sup>1</sup>; C. Bigongiari<sup>2</sup>; G. Cabras<sup>1</sup>; H. Kornmayer<sup>3</sup>; M. Delfino<sup>4</sup>; M. Frailis<sup>1</sup>; M. Hardt<sup>3</sup>; M. Kunze<sup>3</sup>; M. Mazzucato<sup>5</sup>; M. Piraccini<sup>1</sup>

<sup>1</sup> *University of Udine*

<sup>2</sup> *INFN, Padua*

<sup>3</sup> *FORSCHUNGSZENTRUM KARLSRUHE (FZK)*

<sup>4</sup> *PIC, Barcelona*

<sup>5</sup> *INFN-Padova*

The observation of high-energetic gamma-rays with ground based air cerenkov telescopes is one of the most exciting areas in modern astro particle physics. End of the year 2003 the MAGIC telescope started operation. The low energy threshold for gamma-rays together with different background sources leads to a considerable amount of data. The analysis will be done in different institutes spread over Europe. The production of Monte Carlo events including the simulation of Cerenkov light in the atmosphere is very computing intensive and another challenge for a collaboration like MAGIC. Therefore the MAGIC telescope collaborations will take the opportunity to use Grid technology to set up a distributed computational and data intensive analysis system with nowadays available technology. The basic architecture of such a distributed, Europe wide Grid system will be presented. First implementation results will be shown. This Grid might be the starting point for a wider distributed astro particle Grid in Europe.

Poster Session 2 - Board: 52 / 47

## Multi-Terabyte EIDE Disk Arrays running Linux RAID5

**Authors:** D. Petravick<sup>1</sup>; D. Sanders<sup>2</sup>; D. Summers<sup>2</sup>; L. Cremaldi<sup>2</sup>; M.D. Joy<sup>2</sup>; R. Godang<sup>2</sup>; V. Eschenburg<sup>2</sup>

<sup>1</sup> *Fermilab*

<sup>2</sup> *UNIVERSITY OF MISSISSIPPI*

**Corresponding Author:** sanders@phy.olemiss.edu

High-energy physics experiments are currently recording large amounts of data and in a few years will be recording prodigious quantities of data. New methods must be developed to handle this data and make analysis at universities possible. Grid Computing is one method; however, the data must be cached at the various Grid nodes. We examine some storage techniques that exploit recent developments in commodity hardware. Disk arrays using RAID level 5 (RAID5) include both parity and striping. The striping improves access speed. The parity protects data in the event of a single disk failure, but not in the case of multiple disk failures. We report on tests of dual-processor Linux Software RAID5 arrays and Hardware RAID5 arrays using the 12-disk 3ware controller, in conjunction with 300 GB disks, for use in offline high-energy physics data analysis. The price of IDE disks is now less than \$1/GB. These RAID5 disk arrays can be scaled to sizes affordable to small institutions and used when fast random access at low cost is important.

Distributed Computing Systems and Experiences / 48

## AMS-02 Computing and Ground Data Handling

**Authors:** A. Klimentov<sup>1</sup>; M. Pohl<sup>2</sup>; V. Choutko<sup>2</sup>

<sup>1</sup> *A*

<sup>2</sup> *E*

**Corresponding Author:** alexei.klimentov@cern.ch

AMS-02 Computing and Ground Data Handling.

V.Choutko (MIT, Cambridge), A.Klimentov (MIT, Cambridge) and M.Pohl (Geneva University)

AMS (Alpha Magnetic Spectrometer) is an experiment to search in space for dark matter and antimatter on the International Space Station (ISS). The AMS detector had a precursor flight in 1998 (STS-91, June 2-12, 1998). More than 100M events were collected and analyzed.

The final detector (AMS-02) will be installed on ISS in the fall of 2007 for at least 3 years. The data will be transmitted from ISS to NASA Marshall Space Flight Center (MSFC, Huntsville, Alabama) and transferred 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 MC production currently running in 15 Universities and Labs in Europe, USA and Asia, with automatic jobs submission and control from one central place (CERN). The software uses CORBA technology to control and monitor MC production and an ORACLE relational database, to keep catalogues, event description as well as production and monitoring information.

## Event Processing / 49

### Genetic Programming and its application to HEP

**Author:** E. Vaandering<sup>1</sup>

<sup>1</sup> VANDERBILT UNIVERSITY

**Corresponding Author:** ewv@fnal.gov

Genetic programming is a machine learning technique, popularized by Koza in 1992, in which computer programs which solve user-posed problems are automatically discovered. Populations of programs are evaluated for their fitness of solving a particular problem. New populations of ever increasing fitness are generated by mimicking the biological processes underlying evolution. These processes are principally genetic recombination, mutation, and survival of the fittest.

Genetic programming has potential advantages over other machine learning techniques such as neural networks and genetic algorithms in that the form of the solution is not specified in advance and the program can grow as large as necessary to adequately solve the posed problem.

This talk will give an overview and demonstration of the genetic programming technique and show a successful application in high energy physics: the automatic construction of an event filter for FOCUS which is more powerful than the experiment's usual methods of event selection. We have applied this method to the study of doubly Cabibbo suppressed decays of charmed hadrons ( $D^+$ ,  $D_s^+$ , and  $\Lambda_c^+$ ).

## Wide Area Networking / 51

### Test of data transfer over an international network with a large RTT

**Authors:** H. Matsumoto<sup>1</sup>; H. Sakamoto<sup>1</sup>; I. Ueda<sup>1</sup>; J. Tanaka<sup>1</sup>; M. Ishino<sup>1</sup>; S.Y. SUZUKI<sup>2</sup>; T. Mashimo<sup>1</sup>

<sup>1</sup> ICEPP, UNIVERSITY OF TOKYO

<sup>2</sup> KEK

**Corresponding Author:** junichi.tanaka@cern.ch

We have measured the performance of data transfer between CERN and our laboratory, ICEPP, at the University of Tokyo in Japan. The ICEPP will be one of the so-called regional centers for handling the data from the ATLAS experiment which will start data taking in 2007. More than petabytes of data are expected to be generated from the experiment each year. It is therefore essential to achieve a high throughput of data transfer over the long-distance network connection between CERN and ICEPP. A connection with several gigabits per second is now available between the two sites. The round trip time, however, reaches about 300 msec. Moreover the connection is not dedicated to us. Due to the large latency and other traffic on the same network, it is not easy to fully exploit the available bandwidth. We have measured the performance of the network connection using tools such as iperf, bftpl, and gridftp with various TCP parameters, Linux kernel versions and so on. We have examined factors limiting the speed and tried to improve the throughput of the data transfer. In this talk we report on the results of our measurements and investigations.

**Poster Session 2 - Board: 47 / 52**

## Building the LCG: from middleware integration to production quality software

**Authors:** C. Osuna<sup>1</sup>; D. Qing<sup>2</sup>; D. Smith<sup>3</sup>; F. Chollet<sup>4</sup>; G. Grosdidier<sup>5</sup>; J-P. Baud<sup>3</sup>; L. Poncet<sup>6</sup>; M. Litmaath<sup>3</sup>; M. Serra<sup>7</sup>; P. Bettini<sup>7</sup>; Z. Sekera<sup>3</sup>

<sup>1</sup> University of Granada (Spain)

<sup>2</sup> Academia Sinica (Taiwan)

<sup>3</sup> CERN

<sup>4</sup> LAPP-IN2P3

<sup>5</sup> LAL-IN2P3

<sup>6</sup> LAL-IN2p3

<sup>7</sup> INFN-CERN

**Corresponding Author:** louis.poncet@cern.ch

In the last few years grid software (middleware) has become available from various sources. However, there are no standards yet which allow for an easy integration of different services. Moreover, middleware was produced by different projects with the main goal of developing new functionalities rather than production quality software. In the context of the LHC Computing Grid project (LCG) an integration, testing and certification activity is ongoing which aims at producing a stable coherent set of services. Here we report on the processes employed to produce the LCG middleware release and related activities, including the infrastructures used, the activities needed to integrate the various components and the certification process. Our certification process consists of a continuous iterative cycle that also involves feedback from the LCG production system and input from the software providers.

The architecture of the LCG middleware is described, including additional components developed by LCG to improve scalability and performance.

Other associated activities include packaging for deployment, porting to different platforms, debugging and patching of the software. Functionality and stress tests are performed via a large test-bed infrastructure that allows for benchmarking of different configurations. We describe also the results of our tests and our experience collected during the building of the LCG infrastructure.

**Poster Session 3 - Board: 66 / 53**

## **CMD-3 Project Offline Software Development**

**Author:** A. Zaytsev<sup>1</sup>

**Co-authors:** E. Algaer<sup>2</sup>; N. Stuly<sup>2</sup>; S. Pirogov<sup>1</sup>

<sup>1</sup> *BUDKER INSTITUTE OF NUCLEAR PHYSICS*

<sup>2</sup> *NOVOSIBIRSK STATE UNIVERSITY*

CMD-3 is the general purpose cryogenic magnetic detector for VEPP-2000 electron-positron collider, which is being commissioned at Budker Institute of Nuclear Physics (BINP, Novosibirsk, Russia). The main aspects of physical program of the experiment are study of known and search for new vector mesons, study of the  $p\bar{p}$  and  $n\bar{n}$  production cross sections in the vicinity of the threshold and search for exotic hadrons in the region of center-of-mass energy below 2 GeV. The essential upgrade of CMD-2 detector (designed for VEPP-2M collider at BINP) farm and distributed data storage management software is required to satisfy new detector needs and scheduled to perform in near future.

In this talk I will present the general design overview and status of implementation of CMD-3 offline software for reconstruction, visualization, data farm management and user interfaces. Software design standards for this project are object oriented programming techniques, C++ as a main language, Geant4 as an only simulation tool, Geant4 and GDML based detector geometry description, WIRED and HepRep based visualization, CLHEP library based primary generators and Linux as a main platform. The dedicated software development framework (Cmd3Fwk) was implemented in order to be the basic software integration solution and persistency manager. We also look forward to achieve high level of integration with ROOT framework and Geant4 toolkit.

**Poster Session 3 - Board: 59 / 54**

## **Automated Tests in NICOS Nightly Control System**

**Author:** A. Undrus<sup>1</sup>

<sup>1</sup> *BROOKHAVEN NATIONAL LABORATORY, USA*

**Corresponding Author:** undrus@bnl.gov

Software testing is a difficult, time-consuming process that requires technical sophistication and proper planning. This is especially true for the large-scale software projects of High Energy Physics where constant modifications and enhancements are typical. The automated nightly testing is the important component of NICOS, Nightly Control System, that manages the multi-platform nightly builds based on the

recent versions of software packages. It facilitates collective work in collaborative environment and provides four benefits to developers: repeatability (tests can be executed more than once), accumulation (results are stored and reflected on NICOS web pages), feedback (automatic e-mail notifications about test failures), user friendly setup (configuration parameters can be encrypted in the body of test scripts). The modular structure of NICOS allows plugging in other validation and organization tools, such as QMTest and CppUNIT. NICOS classifies tests according to their granularity level and purpose. The low level structural tests reveal compilation problems, inconsistencies in package configuration, such as circular dependencies, and simple isolated bugs. The results for these three groups of tests are published for each package of the software project. The integrated (or behavioral) tests find bugs at levels of users scenarios and NICOS generates the special web page with their results. The NICOS tool is currently used to coordinate the efforts of more than 100 developers for the ATLAS project at CERN and included in the tool library of the LHC computing proje

**Poster Session 2 - Board: 67 / 55**

## **Federating Grids: LCG meets Canadian HEPGrid**

**Authors:** A. Agarwal<sup>1</sup>; A. Dimopoulos<sup>1</sup>; B. Caron<sup>2</sup>; C. Lindsay<sup>1</sup>; D. Vanderster<sup>1</sup>; G. Mateescu<sup>3</sup>; L. Klektau<sup>1</sup>; M. Vetterli<sup>4</sup>; R. Impey<sup>3</sup>; R. Sobie<sup>1</sup>; R. Walker<sup>5</sup>

<sup>1</sup> *University of Victoria*

<sup>2</sup> *University of Alberta/TRIUMF*

<sup>3</sup> *Ottawa*

<sup>4</sup> *Simon Fraser University/TRIUMF*

<sup>5</sup> *Simon Fraser University*

A large number of Grids have been developed, motivated by geo-political or application requirements. Despite being mostly based on the same underlying middleware, the Globus Toolkit, they are generally not inter-operable for a variety of reasons. We present a method of federating those disparate grids which are based on the Globus Toolkit, together with a concrete example of interfacing the LHC grid(LCG) with HEPGrid. HEPGrid consists of shared resources, at several Canadian research institutes, which are exposed via Globus gatekeepers, and makes use of Condor-G for resource advertisement, matchmaking and job submission. An LCG Computing Element(CE) based at the TRIUMF Laboratory hosts a HEPGrid User Interface(UI) which is contained within a custom jobmanager. This jobmanager appears in the LCG information system as a normal CE publishing an aggregation of the HEPGrid resources. The interface interprets the incoming job in terms of HEPGrid UI usage, submits it onto HEPGrid, and implements the jobmanager 'poll' and 'remove' methods, thus enabling monitoring and control across the grids. In this way non-LCG resources are integrated into LCG, without the need for LCG middleware on those resources. The same method can be used to create interfaces between other grids, with the details of the child-Grid being fully abstracted into the interface layer. The LCG-HEPGrid interface is operational, and has been used to federate 1300 CPU's at 4 sites into LCG for the Atlas Data Challenge (DC2).

Poster Session 2 - Board: 43 / 56

## OptorSim: a Simulation Tool for Scheduling and Replica Optimisation in Data Grids

**Authors:** C. Nicholson<sup>1</sup>; D. Cameron<sup>1</sup>; F. Zini<sup>2</sup>; K. Stockinger<sup>3</sup>; P. Millar<sup>1</sup>; R. Carvajal-Schiaffino<sup>2</sup>

<sup>1</sup> UNIVERSITY OF GLASGOW

<sup>2</sup> ITC-irst

<sup>3</sup> Lawrence Berkeley National Laboratory

**Corresponding Author:** c.nicholson@physics.gla.ac.uk

In large-scale Grids, the replication of files to different sites is an important data management mechanism which can reduce access latencies and give improved usage of resources such as network bandwidth, storage and computing power.

In the search for an optimal data replication strategy, the Grid simulator OptorSim was developed as part of the European DataGrid project. Simulations of various HEP Grid scenarios have been undertaken using different job scheduling and file replication algorithms, with the experimental emphasis being on physics analysis use-cases. Previously, the CMS Data Challenge 2002 testbed and UK GridPP testbed were among those simulated; recently, our focus has been on the LCG testbed. A novel economy-based strategy has been investigated as well as more traditional methods, with the economic models showing distinct advantages in terms of improved resource usage.

Here, an overview of OptorSim's design and implementation is presented with a selection of recent results, showing its usefulness as a Grid simulator both in its current features and in the ease of extensibility to new scheduling and replication algorithms.

Event Processing / 57

## IGUANA Interactive Graphics Project: Recent Developments

**Authors:** G. Alverson<sup>1</sup>; G. Eulisse<sup>1</sup>; I. Osborne<sup>2</sup>; L. Taylor<sup>1</sup>; L. Tuura<sup>1</sup>; S. MUZAFFAR<sup>3</sup>

<sup>1</sup> NORTHEASTERN UNIVERSITY

<sup>2</sup> Northeastern University, Boston, USA

<sup>3</sup> NorthEastern University, Boston, USA

**Corresponding Author:** ianna.osborne@cern.ch

This paper describes recent developments in the IGUANA (Interactive Graphics for User ANALysis) project. IGUANA is a generic framework and toolkit, used by CMS and D0, to build a variety of interactive applications such as detector and event visualisation and interactive GEANT3 and GEANT4 browsers.

IGUANA is a freely available toolkit based on open-source components including Qt, OpenInventor (Coin3D) and OpenGL and LCG services.

New features we describe since the last CHEP conference include:

multi-document architecture;

user interface to Python scripting;

2D visualisation with auto-generation of slices/projections from 3D data;

per-object actions such as clipping, slicing, lighting or animation;

correlated actions (e.g. picking) for multiple views;

production of high-quality and compact vector postscript output from any

OpenGL display, with surface shading and invisible surface culling (together with the gl2ps project).

We compare the IGUANA rendering, memory performance, and porting issues for various platforms including: Linux on x86, Windows, and Mac OSX with its native Quartz-Extreme rendering system.

**Poster Session 2 - Board: 48 / 58**

## Performance of an operating High Energy Physics Data grid, D0SAR-grid

**Authors:** B. Abbott<sup>1</sup>; B. Quinn<sup>2</sup>; D. Meyer<sup>3</sup>; E. Gregores<sup>4</sup>; H. Kim<sup>3</sup>; H. Severini<sup>1</sup>; J. Snow<sup>5</sup>; J. Yu<sup>6</sup>; M. Sosebee<sup>3</sup>; N. Mondal<sup>7</sup>; P. Baringer<sup>8</sup>; P. Skubic<sup>1</sup>; S. Novaes<sup>4</sup>; T. Bolton<sup>9</sup>; Z. Greenwood<sup>10</sup>

<sup>1</sup> *University of Oklahoma*

<sup>2</sup> *The University of Mississippi*

<sup>3</sup> *University of Texas at Arlington*

<sup>4</sup> *Universidade Estadual Paulista, Brazil*

<sup>5</sup> *Langston University*

<sup>6</sup> *UNIVERSITY OF TEXAS AT ARLINGTON*

<sup>7</sup> *Tata Institute of Fundamental Research, India*

<sup>8</sup> *University of Kansas*

<sup>9</sup> *Kansas State University*

<sup>10</sup> *Louisiana Tech. University*

**Corresponding Author:** jaehoonyu@uta.edu

The D0 experiment at Fermilab's Tevatron will record several petabytes of data over the next five years in pursuing the goals of understanding nature and searching for the origin of mass. Computing resources required to analyze these data far exceed the capabilities of any one institution. Moreover, the widely scattered geographical distribution of collaborators poses further serious difficulties for optimal use of human and computing resources. These difficulties will be exacerbated in future high energy physics experiments, like those at the LHC. The computing grid has long been recognized as a solution to these problems. This technology is being made a more immediate reality to end users by developing a fully realized grid in the D0 Southern Analysis Region (D0SAR). D0SAR consists of eleven universities in the Southern US, Brazil, Mexico and India. The centerpiece of D0SAR is a data and resource hub, a Regional Analysis Center (RAC). Each D0SAR member institution constructs an Institutional Analysis Center (IAC), which acts as a gateway to the grid for users within that institution. These IACs combine dedicated rack-mounted servers and personal desktop computers into a local physics analysis cluster. D0SAR has been working on establishing an operational regional grid, D0SAR-Grid, using all available resources within it and a home-grown local task manager, McFarm. In this talk, we will describe the architecture of the D0SAR-Grid implementation, the use and functionality of the grid, and the experiences of operating the grid for simulation, reprocessing and analysis of data from a currently running HEP experiment.

**Poster Session 3 - Board: 62 / 59**

## Adaptive Multi-vertex fitting

**Authors:** R. Fruehwirth<sup>1</sup>; W. Waltenberger<sup>2</sup>

<sup>1</sup> *HEPHY Vienna*

<sup>2</sup> *HEPHY VIENNA*

**Corresponding Author:** walten@cern.ch

State of the art in the field of fitting particle tracks to one vertex is the Kalman technique. This least-squares (LS) estimator is known to be ideal in the case of perfect assignment of tracks to vertices and perfectly known Gaussian errors. Experimental data and detailed simulations always depart from this perfect model. The imperfections can be expected to be larger in high luminosity experiments like at the LHC. In such a context vertex fitting algorithms will have to be able to deal with mis-associated tracks and mis-estimated or non-Gaussian track errors. We present a vertex fitting technique that is insensitive to outlying observations and mis-estimated track errors, while it retains close-to-optimal results in the case of perfect data; it adapts to the data. This is realized by introducing weights that are associated to the tracks and reflect the compatibility of the tracks with the vertex. Outliers are no longer simply discarded - as is done in most of the classical robustification techniques - but rather downweighted. The algorithm will be presented in detail, and comparisons with classical methods will be shown in relevant physics cases.

**Online Computing / 61**

## **An Embedded Linux System Based on PowerPC**

**Authors:** C. YUANPING<sup>1</sup>; M. Ye<sup>2</sup>; T. KEJUN<sup>1</sup>; Z. JINGWEI<sup>3</sup>

<sup>1</sup> *IHEP, Beijing, China*

<sup>2</sup> *INSTITUTE OF HIGH ENERGY PHYSICS, ACADEMIA SINICA*

<sup>3</sup> *IHEP, Beijing, CHINA*

**Corresponding Author:** yem@ihep.ac.cn

This article introduces a Embedded Linux System based on vme series PowerPC as well as the base method on how to establish the system. The goal of the system is to build a test system of VMEbus device. It also can be used to setup the data acquisition and control system. Two types of compiler are provided by the developer system according to the features of the system and the PowerPC. At the top of the article some typical embedded Operation system will be introduced and the features of different system will be provided. And then the method on how to build a embedded Linux system as well as the key technique will be discussed in detail. Finally a successful data acquisition example will be given based on the test system.

**Online Computing / 62**

## **The ALICE Experiment Control System**

**Authors:** A. VASCOTTO<sup>1</sup>; C. SOOS<sup>1</sup>; F. Carena<sup>2</sup>; J.-C. MARIN<sup>1</sup>; K. SCHOSSMAIER<sup>1</sup>; P. VANDE VYVRE<sup>1</sup>; R. DIVIA<sup>1</sup>; S. CHAPELAND<sup>1</sup>; W. CARENA<sup>1</sup>

<sup>1</sup> CERN, Geneva, Switzerland

<sup>2</sup> CERN

**Corresponding Author:** franco.carena@cern.ch

The Experiment Control System (ECS) is the top level of control of the ALICE experiment.

Running an experiment implies performing a set of activities on the online systems that control the operation of the detectors. In ALICE, online systems are the Trigger, the Detector Control Systems (DCS), the Data-Acquisition System (DAQ) and the High-Level Trigger (HLT).

The ECS provides a framework in which the operator can have a unified view of all the online systems and perform operations on the experiment seen as a set of detectors.

ALICE has adopted a hierarchical -yet loose- architecture, in which the ECS is a layer sitting above the online systems, still preserving their autonomy to operate independently. The interface between the ECS and the online systems applies a powerful paradigm based on inter-communicating objects. The behavioural aspects of the ECS are described using a finite-state machine model.

The ALICE experiment must be able to run either as a whole (during the physics production) or as a set of independent detectors (for installation and commissioning). The ECS provides all the features necessary to split the experiment into partitions, containing one or more detectors, which can be operated independently and concurrently.

This paper will present the architecture of the ALICE ECS, its current status and the practical experience acquired at the test beams.

**Core Software / 63**

## IgProf profiling tool

**Authors:** G. Eulisse<sup>1</sup>; L. Tuura<sup>1</sup>

<sup>1</sup> NORTHEASTERN UNIVERSITY OF BOSTON (MA) U.S.A.

**Corresponding Author:** giulio.eulisse@cern.ch

A fundamental part of software development is to detect and analyse weak spots of the programs to guide optimisation efforts. We present a brief overview and usage experience on some of the most valuable open-source tools such as valgrind and oprofile. We describe their main strengths and weaknesses as experienced by the CMS experiment.

As we have found that these tools do not satisfy all our needs, CMS has also developed a tool of its own called "igprof". It complements the other tools, allowing us to profile memory usage, CPU usage, memory leaks and file descriptor usage of large complex applications such as the CMS reconstruction and analysis software. It is requires no instrumentation and works with multi-threaded programs and with all shared libraries, including dynamically loaded ones.

We describe this new tool, it's features and output, and experience including improvements gained in CMS.



**Core Software / 64****Composite Framework for CMS Applications****Authors:** G. Eulisse<sup>1</sup>; I. Osborne<sup>2</sup>; L. Taylor<sup>1</sup>; L. Tuura<sup>1</sup>; S. Muzaffar<sup>1</sup>; V. Innocente<sup>3</sup><sup>1</sup> *Northeastern University*<sup>2</sup> *Northeastern University, Boston, USA*<sup>3</sup> *CERN***Corresponding Author:** ianna.osborne@cern.ch

We present a composite framework which exploits the advantages of the CMS data model and uses a novel approach for building CMS simulation, reconstruction, visualisation and future analysis applications. The framework exploits LCG SEAL and CMS COBRA plug-ins and extends the COBRA framework to pass communications between the GUI and event threads, using SEAL callbacks to navigate through the metadata and event data interactively in a distributed environment.

We give examples of current applications based on this framework, including CMS test-beams, geometry description debugging, GEANT4 simulation, event reconstruction, and the verification of reconstruction and higher level trigger algorithms.

**Poster Session 2 - Board: 44 / 65****Monitoring CMS Tracker construction and data quality using a grid/web service based on a visualization tool****Authors:** A. Regano<sup>1</sup>; G. Zito<sup>2</sup>; M.S. Mennea<sup>3</sup><sup>1</sup> *University Bari*<sup>2</sup> *INFN BARI*<sup>3</sup> *UNIVERSITY & INFN BARI***Corresponding Author:** giuseppe.zito@ba.infn.it

The complexity of the CMS Tracker (more than 50 million channels to monitor) now in construction in ten laboratories worldwide with hundreds of interested people, will require new tools for monitoring both the hardware and the software. In our approach we use both visualization tools and Grid services to make this monitoring possible. The use of visualization enables us to represent in a single computer screen all those million channels at once. The Grid will make it possible to get enough data and computing power in order to check every channel and also to reach the experts everywhere in the world allowing the early discovery of problems.

We report here on a first prototype developed using the Grid environment already available now in CMS i.e. LCG2. This prototype consists on a Java client which implements the GUI for Tracker Visualization and a few data servers connected to the tracker construction database, to Grid catalogs of event datasets or directly to test beam setups data acquisition. All the communication between client and servers is done using data encoded in xml and standard Internet protocols.

We will report on the experience acquired developing this prototype and on possible future developments in the framework of an interactive Grid and a virtual counting room allowing complete detector control from everywhere in the world.

**Poster Session 3 - Board: 34 / 66**

## **CMS Tracker Visualisation Tools**

**Authors:** A. Regano<sup>1</sup>; G. Zito<sup>2</sup>; I. Osborne<sup>3</sup>; M.S. Mennea<sup>4</sup>

<sup>1</sup> *University Bari*

<sup>2</sup> *INFN Bari*

<sup>3</sup> *Northeastern University, Boston, USA*

<sup>4</sup> *UNIVERSITY & INFN BARI*

**Corresponding Author:** maria.santa.mennea@cern.ch

This document will review the design considerations, implementations and performance of the CMS Tracker Visualization tools. In view of the great complexity of this subdetector (more than 50 millions channels organized in 17000 modules each one of these being a complete detector), the standard CMS visualisation tools (IGUANA and IGUANACMS) that provide basic 3D capabilities and integration within CMS framework respectfully have been complemented with additional 2D graphics objects and a detailed object model of the tracker. Based on the experience acquired by using this software to debug and understand both hardware and software during the construction phase, we will propose possible future improvements to cope with online monitoring and event analysis during data taking.

**Event Processing / 67**

## **The simulation for the ATLAS experiment: present status and outlook**

**Authors:** A. DELL'ACQUA<sup>1</sup>; A. NAIRZ<sup>1</sup>; A. Rimoldi<sup>2</sup>; D. COSTANZO<sup>3</sup>; J. BOUDREAU<sup>4</sup>; M. GALLAS<sup>5</sup>; V. TSULAI<sup>4</sup>

<sup>1</sup> *CERN, Geneva, Switzerland*

<sup>2</sup> *PAVIA UNIVERSITY & INFN*

<sup>3</sup> *LBL, USA*

<sup>4</sup> *University of Pittsburgh, USA*

<sup>5</sup> *CERN*

**Corresponding Author:** adele.rimoldi@cern.ch

The simulation for the ATLAS experiment is presently operational in a full OO environment and it is presented here in terms of successful solutions to problems dealing with application in a wide community using a common framework. The ATLAS experiment is the perfect scenario where to test all applications able to satisfy the different needs of a big community. Following a well stated strategy of transition from the GEANT3 to the GEANT4-based simulation, a good validation programme during the last months confirmed the characteristics of reliability, performance and robustness of this new tool in comparison with the results of the previous simulation. Generation, simulation and digitization steps on different full sets of physics events were tested in terms of performance and robustness in comparisons with the same samples undergoing the old GEANT3-based simulation. The simulation program is simultaneously tested on all different testbeam setups characterizing the R&D programme of all subsystems belonging to the ATLAS detector with comparison to real data in order to validate the physics content and the reliability in the detector description of each component.

Distributed Computing Services / 69

## Resource Predictors in HEP Applications

**Authors:** J. Huth<sup>1</sup>; J. Schopf<sup>2</sup>; S. Grinstein<sup>1</sup>

<sup>1</sup> HARVARD UNIVERSITY

<sup>2</sup> Argonne National Lab

**Corresponding Author:** sgrinstein@ifae.es

The ATLAS experiment uses a tiered data Grid architecture that enables possibly overlapping subsets, or replicas, of the original set to be located across the ATLAS collaboration. The full set of experiment data is located at a single Tier 0 site, and then subsets of the data are located at national Tier 1 sites, smaller subsets at smaller regional Tier 2 sites, and so on. In order to understand the data needs, both in terms of access, replication policy, and storage capacity, we need good estimations of resource needs for data manipulation. Specifically, we envision a time when a user will want to determine which is more expedient, downloading a replica from a site or recreating it from scratch.

This paper presents our technique to predict the behavior of ATLAS applications, and then to combine this information with Internet link bandwidth estimation to improve resource usage in the ATLAS Grid environment. We studied the parameters that affect the execution time performance of event generation, detector simulation, and event reconstruction. Our results show that we can achieve predictions within 10-40% of the execution time (depending on the application), better than many other pragmatic prediction techniques. We implemented a software package to provide data transfer bandwidth estimation and execution time prediction that can be used with the Chimera software to aid in managing application execution and to improve resource usage for ATLAS.

Poster Session 3 - Board: 17 / 70

## Tracking of long lived hyperons in silicon detector at CDF.

**Authors:** D. Litvintsev<sup>1</sup>; E. Gerchtein<sup>2</sup>

<sup>1</sup> FNAL

<sup>2</sup> CMU

**Corresponding Author:** lena@fnal.gov

Long lived charged hyperon,  $\Xi$  and  $\Omega$ , are capable of travelling significant distances producing hits in the silicon detector, before decaying into  $\Lambda^0\pi$  and  $\Lambda^0K$  pairs, respectively. This gives unique opportunity of reconstructing hyperon tracks. We have developed a dedicated “outside-in” tracking algorithm that is seeded by 4-momentum and decay vertex of the long lived hyperon reconstructed by its decay products. The tracking of hyperons in the silicon detector results in a dramatic reduction of the combinatorial background and an improvement of the momentum resolution compared with the standard reconstruction using final decay products.

Using a super clean sample of  $\Xi$  hyperons CDF observed charmed-strange baryon isodoublet  $\Xi_c^0$  and  $\Xi_c^+$  for the first time in  $p\bar{p}$  collision.  
 $\Xi$  hyperons were used for the search for exotic  $S = -2$  baryons decaying into  $\Xi\pi$ .

**Computer Fabrics / 72**

## Performance analysis of Cluster File System on Linux

**Authors:** D. XU<sup>1</sup>; G. CHEN<sup>1</sup>; L. WANG<sup>1</sup>; W. WU<sup>1</sup>; Y. CHENG<sup>1</sup>

<sup>1</sup> COMPUTING CENTER, INSTITUTE OF HIGH ENERGY PHYSICS, CHINESE ACADEMY OF SCIENCES

With the development of Linux and improvement of PC's performance, PC cluster used as high performance computing system is becoming much popular. The performance of I/O subsystem and cluster file system is critical to a high performance computing system. In this work the basic characteristics of cluster file systems and their performance are reviewed. The performance of four distributed cluster file systems, AFS, NFS, PVFS and CASTOR, were measured. The measurements were carried out on CERN version RedHat 7.3.3 Linux using standard I/O performance benchmarks. Measurements show that for single-server single client configuration, NFS, CASTOR and PVFS have better performance and write rate slightly increases while the record length becomes larger. CASTOR has the best throughput when the number of write processes increases. PVFS and CASTOR are tested on multi-server and multi-client system. The two file systems nicely distribute data I/O to all servers. CASTOR RFIO protocol shows the best utilization of network bandwidth and optimized to large data size files. CASTOR also has the better scalability as a cluster file system. Based on the test some methods are proposed to improve the performance of cluster file system.

**Poster Session 3 - Board: 67 / 73**

## The application of PowerPC/VxWorks to the read-out subsystem of the BESIII DAQ

**Authors:** N. tao<sup>1</sup>; YP. chu<sup>1</sup>; Z. JINGWEI<sup>2</sup>

<sup>1</sup> IHEP

<sup>2</sup> IHEP, Beijing, China

**Corresponding Author:** ninatao@mail.ihep.ac.cn

This article describes the simulation of the read-out subsystem which will be subject to the BESIII data acquisition system. According to the purpose of the BESIII, the event rate will be about 4000Hz, and the data rate up to 50Mbytes/sec after Level 1 trigger. The read-out subsystem consists of some read-out crates and read-out computer whose principle function is to collect event data from the front-end electronics after Level 1 trigger, to transfer data fragments from each VME read-out crate to online computer farm through two levers of computer pre-processing and high speed network transmission. The read-out implementation is based on commercial single board computer MVME5100 running VxWorks operating system.

The article outlines the structure of the simulative platform which included hardware components and software components. It puts

emphasis on the framework of the read-out subsystem, data process flow and test method.

Especially, it enumerates key technologies in the process of design and analyses of the test result. In addition, results which summaries the performance of the single board computer from the data transferring aspects will be presented.

Key word: BESIII read-out subsystem MVME5100 VxWorks VMEbus DMA read-out computer

## Online Computing / 74

### The LHCb Configuration Database

**Authors:** B. Jost<sup>1</sup>; C. Gaspar<sup>1</sup>; E. Van Herwijnen<sup>1</sup>; L. Abadie<sup>1</sup>; N. Neufeld<sup>1</sup>; R. Jacobsson<sup>1</sup>

<sup>1</sup> CERN

The aim of the LHCb configuration database is to store all the controllable devices of the detector. The experiment's control system (that uses PVSS) will configure, start up and monitor the detector from the information in the configuration database. The database will contain devices with their properties, connectivity and hierarchy. The ability to rapidly store and retrieve huge amounts of data, and the navigability between devices are important requirements. We have collected use cases to ensure the completeness of the design.

Using the entity relationship modeling technique we describe the use cases as classes with attributes and links. We designed the schema of the tables using the relational diagrams. This methodology has been applied to describe and store the connectivity of the devices in the TFC (switches) and DAQ system. Other parts of the detector will follow later.

The database has been implemented using Oracle to benefit from central CERN database support. The project also foresees the creation of tools to populate, maintain, and configure the configuration database. To communicate between the control system and the database we have developed a system which sends queries to the database and displays the results in PVSS. This database will be used in conjunction with the configuration database developed by the CERN JCOP project for PVSS.

## Event Processing / 75

### A kinematic and a decay chain reconstruction library

**Authors:** K. Prokofiev<sup>1</sup>; T. Speer<sup>1</sup>

<sup>1</sup> Physik-Institut Universitat Zurich

**Corresponding Author:** kirill.prokofiev@cern.ch

A kinematic fit package was developed based on Least Means Squared minimization with Lagrange multipliers and Kalman filter techniques and implemented in the framework of the CMS reconstruction program. The package allows full decay chain reconstruction from final state

to primary vertex according to the given decay model. The class framework allowing decay tree description on every reconstruction step will be described in details. Package extension to any type of physics object reconstructed in CMS, integration to general CMS reconstruction framework and related questions will be discussed. Examples of decay chain models, constraints and their application on Bs reconstruction will be presented.

**Poster Session 3 - Board:** 12 / 76

## Deterministic Annealing for Vertex Finding at CMS

**Authors:** E. Chabanat<sup>1</sup>; N. Estre<sup>1</sup>

<sup>1</sup> *IN2P3*

**Corresponding Author:** chabanat@ipnl.in2p3.fr

CMS and others LHC experiments offer a new challenge for the vertex reconstruction: the elaboration of efficient algorithms at high-luminosity beam collisions. We present here a new algorithm in the vertex finding field : Deterministic Annealing (DA). This algorithm comes from information theory by analogy to statistical physics and has already been used in clustering and classification problems. In our purpose, the main job is to code information of a set of tracks into prototypes which will be our vertices at the end of the process. The advantages of such a technique is to globally search all vertices at one time and a priori knowledge of the expected number of vertices is not required: the algorithm creates new vertices by a phase transition mechanism which will be describe in this contribution. Thus, the first part of this talk is devoted to a short description of the DA algorithm and to the necessary introduction of the concept of apex points which stand for tracks in this method ; then a discussion of vertex reconstruction efficiencies follows consisting finding DA's internal parameters and making a comparison between DA and the most popular vertex finding algorithm. This comparison is done considering 4000 bbar events generated in the detector central region without pile-up in a first approach ; primary and secondary vertices reconstruction results are shown. Then performances of DA in regional vertex search with regional tracks reconstruction is also presented and lead to a short study of 500 bbar event with pile-up at low luminosity.

**Event Processing / 77**

## Using the reconstruction software, ORCA, in the CMS datachallenge

**Author:** S. Wynhoff<sup>1</sup>

<sup>1</sup> *PRINCETON UNIVERSITY*

**Corresponding Author:** stephan.wynhoff@cern.ch

We report on the software for Object-oriented Reconstruction for CMS Analysis, ORCA. It is based on the Coherent Object-oriented Base for Reconstruction, Analysis and simulation (COBRA) and used for digitization and reconstruction of simulated Monte-Carlo events as well as testbeam data.

For the 2004 data challenge the functionality of the software has been extended to store collections of reconstructed objects (DST) as

well as the previously storable quantities (Digis) in multiple, parallel streams.

We describe the structure of the DST, the way to ensure and store the configuration of reconstruction algorithms that fill the collections of reconstructed objects as well as the relations between them. Also the handling of multiple streams to store parts of selected events is discussed. The experience from the implementation used early 2004 and the modifications for future optimization of reconstruction and analysis are presented.

**Poster Session 2 - Board: 16 / 78**

## **Toward a Grid Technology Independent Programming Interface for HEP Applications**

**Authors:** A. Delgado Peris<sup>1</sup>; A. Sciaba<sup>1</sup>; F. Donno<sup>1</sup>; P. Mendez Lorenzo<sup>1</sup>; R. Santinelli<sup>1</sup>; S. Campana<sup>1</sup>

<sup>1</sup> CERN

**Corresponding Author:** flavia.donno@cern.ch

In the High Energy Physics (HEP) community, Grid technologies have been accepted as solutions to the distributed computing problem. Several Grid projects have provided software in the last years. Among of all them, the LCG - especially aimed at HEP applications - provides a set of services and respective client interfaces, both in the form of command line tools as well as programming language APIs in C, C++, Java, etc.

Unfortunately, the programming interface presented to the end user (the physicist) is often not uniform or provides different levels of abstractions. In addition, Grid technologies face a constant change and an improvement process and it is of major importance to shield changes of underlying technology to the end users. As services evolve and new ones are introduced, the way users interact with them also changes.

These new interfaces are often designed to work at a different level and with a different focus than the original ones. This makes it hard for the end user to build Grid applications.

We have analyzed the existing LCG programming environment and

identified several ways to provide high-level technology independent interfaces. In this article, we describe the use cases we were presented by the LCG experiments and the specific problems we encountered in documenting existing APIs and providing usage examples. As a main contribution, we also propose a prototype high-level interface for the information, authentication and authorization systems that is now under test on the LCG EIS testbed by the LHC experiments.

**Poster Session 3 - Board: 64 / 79**

## **The CEDAR Project**

**Authors:** B. Waugh<sup>1</sup>; J. Butterworth<sup>1</sup>; J. Stirling<sup>2</sup>; M. Whalley<sup>3</sup>; S. BUTTERWORTH<sup>4</sup>

<sup>1</sup> UCL, London

<sup>2</sup> IPPP, Durham

<sup>3</sup> IPPP, UNIVERSITY OF DURHAM

<sup>4</sup> UCL, London, UK

**Corresponding Author:** m.r.whalley@durham.ac.uk

We will describe the plans and objectives of the recently funded PPARC(UK) e-science project, the Combined E-Science Data Analysis Resource for High Energy Physics (CEDAR), which will combine the strengths of the well established and widely used HEPDATA library of HEP data and the innovative JETWEB Data/Monte Carlo comparison facility built on the HZTOOL package and which exploits developing grid technology. The current status and future plans of both of these individual sub-projects within the CEDAR framework are described showing how they will cohesively provide a) an extensive archive of Reaction Data, b) validation and tuning of Monte Carlo programmes against the Reaction Data sets, and c) a validated code repository for a wide range of HEP code such as parton distribution functions and other calculation codes used by particle physicists. Once established it is envisaged CEDAR will become an important GRID tool used by LHC experimentalists in their analyses and may well serve as a model in other branches of science which have need to compare data and complex simulations

**Poster Session 1 - Board: 6 / 84**

## **Raw Ethernet based hybrid control system for the automatic control of suspended masses in gravitational waves interferometric detectors**

**Authors:** A. Eleuteri<sup>1</sup>; F. Acernese<sup>2</sup>; F. Barone<sup>3</sup>; G. RUSSO<sup>4</sup>; I. Ricciardi<sup>4</sup>; K. Qipiani<sup>5</sup>; L. Milano<sup>4</sup>; R. De Rosa<sup>4</sup>; S. Pardi<sup>6</sup>

<sup>1</sup> DIPARTIMENTO DI SCIENZE FISICHE - UNIVERSITÀ DI NAPOLI FEDERICO II

<sup>2</sup> Dipartimento di Scienze Fisiche - Università di Salerno

<sup>3</sup> Dipartimento di Scienze Farmaceutiche - Università di Salerno

<sup>4</sup> Dipartimento di Scienze Fisiche - Università di Napoli Federico II

<sup>5</sup> INFN - sez. Napoli

<sup>6</sup> Dipartimento di Matematica ed Applicazioni "R.Caccioppoli"

**Corresponding Author:** antonio.eleuteri@na.infn.it

In this paper we examine the performance of the raw Ethernet protocol in deterministic, low-cost, real-time communication. Very few applications have been reported until now, and they focus on the use of the TCP and UDP protocols, which however add a sensible overhead to the communication and reduce the useful bandwidth. We show how low-level Ethernet access can be used for peer-to-peer, short distance communication, and how it allows the writing of applications requiring large bandwidth. We show some examples running on the Lynx real-time OS and on Linux, both in mixed and homogeneous environments. As an example of application of this technique, we describe the architecture of a hybrid Ethernet based real-time control system prototype we implemented in Napoli, discussing its characteristics and performances. Finally we discuss its application to the real-time control of a suspended mass of the mode cleaner of the 3m prototype optical interferometer for gravitational wave detection operational in Napoli.



**Core Software / 85****Recent Developments in the ROOT I/O**

**Authors:** F. Rademakers<sup>1</sup>; P. Canal<sup>2</sup>; R. Brun<sup>1</sup>

<sup>1</sup> *CERN*

<sup>2</sup> *FERMILAB*

**Corresponding Author:** philippe.canal@cern.ch

Since version 3.05/02, the ROOT I/O System has gone through significant enhancements.

In particular, the STL container I/O has been upgraded to support splitting, reading without existing libraries and using directly from TTreeFormula (TTree queries).

This upgrade to the I/O system is such that it can be easily extended (even by the users) to support the splitting and querying of almost any collections. The ROOT TTree queries engine has also been enhanced in many ways including an increase performance, better support for array printing and histogramming, addition of the ability to call any external C or C++ functions, etc.

We improved the I/O support for classes not inheriting from TObject, including support for automatic schema evolution without using an explicit class version. ROOT now support generating files larger than 2Gb. We also added plugins for several of the mass storage servers (Castor, DCache, Chirp, etc.).

We will describe in details these new features and their implementation.

**Distributed Computing Systems and Experiences / 86****Job Submission & Monitoring on the PHENIX Grid\***

**Authors:** A. Shevel<sup>1</sup>; A. Withers<sup>1</sup>; B. Jacak<sup>1</sup>; D. Morrison<sup>2</sup>; I. Sourikova<sup>3</sup>; M. Reuter<sup>1</sup>; R. Lacey<sup>1</sup>; T. Thomas<sup>4</sup>

<sup>1</sup> *STATE UNIVERSITY OF NEW YORK AT STONY BROOK*

<sup>2</sup> *Brookhaven National Laboratory*

<sup>3</sup> *Brookhaven National Laboratory*

<sup>4</sup> *State University of New Mexico*

**Corresponding Author:** shevel@mail.chem.sunysb.edu

The PHENIX collaboration records large volumes of data for each experimental run (now about 1/4 PB/year). Efficient and timely analysis of this data can benefit from a framework for distributed analysis via a growing number of remote computing facilities in the collaboration. The grid architecture has been, or is being deployed at most of these facilities.

The experience being obtained in the transition to the Grid infrastructure with minimum of manpower is presented with particular emphasis on job monitoring and job submission in multi cluster environment. The integration of the existing subsystems (from Globus project, from several HEP collaborations), large application libraries, and other software tools to render the resulting architecture stable, robust, and useful for the end user is also discussed.

**Core Software / 87****Aspect-Oriented Extensions to HEP Frameworks****Authors:** C. Tull<sup>1</sup>; P. Calafiura<sup>1</sup><sup>1</sup> *LBNL/ATLAS***Corresponding Author:** cetull@lbl.gov

In this paper we will discuss how Aspect-Oriented Programming (AOP) can be used to implement and extend the functionality of HEP architectures in areas such as performance monitoring, constraint checking, debugging and memory management. AOP is the latest evolution in the line of technology for functional decomposition which includes Structured Programming (SP) and Object-Oriented Programming (OOP). In AOP, an Aspect can contribute to the implementation of a number of procedures and objects and is used to capture a concern such as logging, memory allocation or thread synchronization that crosscuts multiple modules and/or types. We have chosen Gaudi as a representative HEP architecture because it is a component architecture and has been successfully adopted by several HEP experiments. Since most HEP frameworks are currently implemented in C++, for our study we have used AspectC++, an extension to C++ that allows the use of AOP techniques without adversely affecting software performance. We integrated AspectC++ in the development environment of the Atlas experiment, and we will discuss some of the configuration management issues that may arise in a mixed C++/AspectC++ environment. In this study we have focused on “Development Aspects”, i.e. aspects that are intended to facilitate program development but can be transparently removed from the production code, such as execution tracing, constraint checking and object lifetime monitoring. We will briefly discuss possible “Production Aspects” related to cache management and object creation. For each of the concerns we have examined we will discuss how traditional SP or OOP techniques compare to the AOP solution we developed. We will conclude discussing the short and medium term feasibility of introducing AOP, and AspectC++ in particular, in the complex software systems of the LHC experiments.

**Wide Area Networking / 88****A Globally Distributed Real Time Infrastructure for World Wide Collaborations****Authors:** D. Adamczyk<sup>1</sup>; D. Lattka<sup>1</sup>; G. Denis<sup>1</sup>; H. Newman<sup>1</sup>; I. Legrand<sup>2</sup>; J. Fernandes<sup>3</sup>; J. SUCIK<sup>1</sup>; K. Wei<sup>1</sup>; P. FARKAS<sup>1</sup>; P. Galvez<sup>2</sup><sup>1</sup> *Caltech*<sup>2</sup> *CALTECH*<sup>3</sup> *Caltech/CERN***Corresponding Author:** galvez@caltech.edu

VRVS (Virtual Room Videoconferencing System) is a unique, globally scalable next-generation system for real-time collaboration by small workgroups, medium and large teams engaged in research, education and outreach. VRVS operates over an ensemble of national and international networks. Since it went into production service in early 1997, VRVS has become a standard part of the toolset used daily by a large sector of HENP, and it is used increasingly for other DoE/NSF-supported programs. Today, the VRVS Web-based system is regularly accessed by more than 30,000 registered hosts running the VRVS software in more than 103 countries. There are currently 78 VRVS “reflectors” that create the

interconnections and manage the traffic flow, in the Americas, Europe and Asia. New reflectors recently have been installed in Brazil, China, Pakistan, Australia and Slovakia.

VRVS is global in scope: it covers the full range of existing and emerging protocols and the full range of client devices for collaboration, from mobile systems through desktops to installations in large auditoria. VRVS will be integrated with the Grid-enabled Analysis Environment (GAE) now under development at Caltech in partnership with the GriPhyN, iVDGL and PPDG projects in the US, and Grid projects in Europe.

A major architectural change is currently in development. The new version v4.0, is expected to be deployed in early 2005. We will describe the current operational state of the VRVS service and provide a description of the new architecture including all the new and advanced functionalities that will be added.

### Distributed Computing Services / 89

## MonALISA: An Agent Based, Dynamic Service System to Monitor, Control and Optimize Grid based Applications.

**Authors:** C Grigoras<sup>1</sup>; C. Cirstoiu<sup>1</sup>; C. Dobre<sup>1</sup>; H. Newman<sup>2</sup>; I. Legrand<sup>3</sup>; M. Toarta<sup>1</sup>; R. Voicu<sup>4</sup>

<sup>1</sup> Polytechnic University Bucharest

<sup>2</sup> Caltech

<sup>3</sup> CALTECH

<sup>4</sup> CERN

**Corresponding Author:** iosif.legrand@cern.ch

The MonALISA (MONitoring Agents in A Large Integrated Services Architecture) system is a scalable Dynamic Distributed Services Architecture which is based on the mobile code paradigm.

An essential part of managing a global system, like the Grids, is a monitoring system that is able to monitor and track the many site facilities, networks, and all the task in progress, in real time. MonALISA is designed to easily integrate existing monitoring tools and procedures and to provide this information in a dynamic, self describing way to any other services or clients.

The monitoring information gathered is essential for developing higher level services that provide decision support, and eventually some degree of automated decisions, to help maintain and optimize workflow through the Grid.

MonALISA is an ensemble of autonomous multi-threaded, agent-based subsystems which are registered as dynamic services and are able to collaborate and cooperate in performing a wide range of monitoring, data processing and control tasks in large scale distributed applications. We also present the development of specialized higher level services, implemented as distributed mobile agents in the MonALISA framework to control and globally optimize tasks as grid scheduling, real-time data streaming or effective file replication.

The system is currently used to monitor several large scale systems and provides detailed information for computing nodes, LAN and WAN network components, job execution and applications specific parameters. This distributed system proved to be reliable, able to correctly handle connectivity problems and is running around the clock on more than 120 sites.

Poster Session 2 - Board: 35 / 90

## MONARC2: A Processes Oriented, Discrete Event Simulation Framework for Modelling and Design of Large Scale Distributed Systems.

**Authors:** H. Newman<sup>1</sup>; I. Legrand<sup>2</sup>

<sup>1</sup> *Caltech*

<sup>2</sup> *CALTECH*

**Corresponding Author:** iosif.legrand@cern.ch

The design and optimization of the Computing Models for the future LHC experiments, based on the Grid technologies, requires a realistic and effective modeling and simulation of the data access patterns, the data flow across the local and wide area networks, and the scheduling and workflow created by many concurrent, data intensive jobs on large scale distributed systems.

This paper presents the latest generation of the MONARC (MOdels of Networked Analysis at Regional Centers) simulation framework, as a design and modelling tool for large scale distributed systems applied to HEP experiments. A process-oriented approach for discrete event simulation is used for describing concurrent running programs, as well as the stochastic arrival patterns that characterize how such systems are used. The simulation engine is based on Threaded Objects, (or Active Objects) which offer great flexibility in simulating the complex behavior of distributed data processing programs. The engine provides an appropriate scheduling mechanism for the Active Objects with efficient support for interrupts.

The framework provides a complete set of basic components (processing nodes, data servers, network components) together with dynamically loadable decision units (scheduling or data replication modules) for easily building complex Computing Model simulations.

Examples of simulating complex data processing systems, specific for the LHC experiments (production tasks associated with data replication and interactive analysis on distributed farms) are presented, and the way the framework is used to compare different decision making algorithms or to optimize the overall Grid architecture.

Poster Session 2 - Board: 4 / 91

## New distributed offline processing scheme at Belle

**Authors:** F. Ronga<sup>1</sup>; I. Adachi<sup>1</sup>; N. Katayama<sup>1</sup>

<sup>1</sup> *KEK*

**Corresponding Author:** ichiro.adachi@kek.jp

The Belle experiment has accumulated an integrated luminosity of more than 240fb<sup>-1</sup> so far, and a daily logged luminosity has exceeded 800pb<sup>-1</sup>. This requires more efficient and reliable way of event processing. To meet this requirement, new offline processing scheme has been constructed, based upon technique employed for the Belle online reconstruction farm. Event processing is performed at PC farms, which consists of 60 quad(0.7GHz) and 225 dual(1.3GHz or 3.2GHz) CPU PC nodes. Raw event data are read from a Solaris tape server connected to a DTF2 tape drive, and they are distributed over all PC nodes. Reconstructed events are recorded onto 8 file servers,

which are newly installed last year. To maximize processing capabilities, various optimizations such as PC clustering, job control, output data management and so on have been done. As a result, processing power with this scheme has been more than doubled, which corresponds to that more than 3 fb<sup>-1</sup> of beam data per day can be processed. In this talk, stable operation of our new system, together with a description of the Belle offline computing model, will be demonstrated by showing computing performance obtained from experience in processing beam data.

Poster Session 1 - Board: 12 / 93

## Managing software licences for a large research laboratory

**Authors:** A. Thys<sup>1</sup>; H. Hansbakk<sup>1</sup>; M. Katzarova<sup>1</sup>; N. Hoemyr<sup>1</sup>; P. Baehler<sup>1</sup>; P. Saiz<sup>1</sup>; T. Osborne<sup>1</sup>

<sup>1</sup> CERN IT

**Corresponding Author:** nils.hoimyr@cern.ch

The Product Support (PS) group of the IT department at CERN distributes and supports more than one hundred different software packages, ranging from tools for computer aided design, field calculations, mathematical and structural analysis to software development. Most of these tools, which are used on a variety of Unix and Windows platforms by different user populations, are commercial packages requiring a licence. The group is also charged with license negotiations with the software vendors.

Keeping track of large number and variety of licences is no easy task, so in order to provide a more automated and more efficient service, the PS group has developed a database system to both track detailed licence configurations and to monitor their use. The system is called PSLicmon (PS Licence Monitor) and is based on an earlier development from the former CE group.

PSLicmon consists of four main components: report generation, data loader, Oracle product database and a PHP-based Web-interface. The license log parser/loader is implemented in Perl and loads reports from the different license managers into the Oracle database. The database contains information about products, licenses and suppliers and is linked to CERN's human resource database. The web-interface allows for on the fly generation of statistics plots as well as data entry and updates. The system also includes an alarm system for licence expiry.

Thanks to PSLicmon, the support team is able to better match licence acquisitions with the diverse needs of its user community, and to be in control of migration and phaseout scenarios between different products and/or product versions. The tool has proved to be a useful aid when making decisions regarding product support policy and licence acquisitions, in particular ensuring the provision of the correct number of often expensive software licences to match CERN's needs.

Poster Session 3 - Board: 35 / 94

## Domain Specific Visual Query Language for HEP analysis or How far can we go with user friendliness?

**Author:** V M. Moreira do Amaral<sup>1</sup>

**Co-authors:** G. Moerkotte<sup>1</sup>; S. Helmer<sup>1</sup>

<sup>1</sup> *UNIVERSITY OF MANNHEIM*

**Corresponding Author:** amaral@pi3.informatik.uni-mannheim.de

There is a permanent quest for user friendliness in HEP Analysis. This growing need is directly proportional to the analysis frameworks' interface complexity. In fact, the user is provided with an analysis framework that makes use of a General Purpose Language to program the query algorithms. Usually the user finds this overwhelming, since he or she is presented with the complexity of the intricacies of the systems. This way the final user of HEP experiments becomes a forced programmer or an application developer.

In our opinion this inflicts directly or indirectly in the query system performances. For this reason we have decided to invest in a line of research to find a solution that balances the complexity and variability of the analysis queries with the need for simpler query systems interfaces. The ultimate goal is to save time on query algorithms production and to have a way to increase efficiency.

In this communication we are going to present how we explored the hypothesis of generating a visual query language specific for the HEP high level analysis domain. The prototyped framework developed so far, PHEASANT, is supporting our arguments in the feasibility of this approach. Therefore, like in any young Human Centric development project, this raises the need of a broad discussion in order to validate it. We believe to be opening a new fruitful research topic among the community and we expect to motivate both computer science and physics experts into the same discussion.

**Online Computing / 95**

## **The introduction to BES computing environment**

**Authors:** D. XU<sup>1</sup>; G. CHEN<sup>1</sup>

<sup>1</sup> *COMPUTING CENTER, INSTITUTE OF HIGH ENERGY PHYSICS, CHINESE ACADEMY OF SCIENCES*

BES is an experiment on Beijing Electron-Positron Collider (BEPC). BES computing environment consists of PC/Linux cluster and mainly relies on the free software. OpenPBS and Ganglia are used as job schedule and monitor system. With helps from CERN IT Division, CASTOR was implemented as storage management system. BEPC is being upgraded and luminosity will increase one hundred times comparing to current machine. The data produced by new BES-III detector will be about 700 Terabytes per year. To meet the computing demand, we proposed a solution based on PC/Linux/Cluster and SAN technology. CASTOR will be used to manage the storage resources of SAN. We started to develop a graphical interface for CASTOR. Some tests on data transmission performance of SAN environment were carried out. The result shows that I/O performance of SAN is better than that of traditional storage connection method including IDE, SCSI etc and it can satisfy BESIII experiment's demand for data processing.

**Wide Area Networking / 96**

## The Network Security Protection System at IHEP-Net

**Author:** L. Ma<sup>1</sup>

**Co-authors:** B. Liu <sup>2</sup>; C. Wu <sup>2</sup>; C. Yu <sup>2</sup>; D. An <sup>2</sup>; G. Chen <sup>2</sup>; R. Liu <sup>2</sup>; R. Xu <sup>2</sup>

<sup>1</sup> *INSTITUTE OF HIGH ENERGY PHYSICS*

<sup>2</sup> *Institute of High Energy Physics*

**Corresponding Author:** ma@ihep.ac.cn

Network security at IHEP is becoming one of the most important issues of computing environment. To protect its computing and network resources against attacks and viruses from outside of the institute, security measures to combat these are implemented. To enforce security policy the network infrastructure was re-configured to one intranet and two DMZ areas. New rules to control the access between intranet and DMZ areas are applied. All hosts at IHEP are divided into three types according to their security levels. Hosts of the first type are isolated in the institute and can just access the hosts inside of IHEP. The second type hosts access Internet through NAT. The third type hosts will directly connect to outside. An intrusion detection system works with firewall so that all packets from outside IHEP are checked and filtered. Access from outside will go through firewall or VPN. In order to prevent virus spread at IHEP and reduce the number of spam mail we installed a virus filter and spam filter system. All of these measures make the network at IHEP more secure. Attacks, virus and spam mails decrease dramatically.

**Poster Session 3 - Board: 39 / 98**

## RDBC: ROOT DataBase Connectivity

**Author:** V. Onuchin<sup>1</sup>

**Co-authors:** Fons Rademakers <sup>2</sup>; P. Canal <sup>3</sup>; Rene Brun <sup>2</sup>

<sup>1</sup> *CERN, IHEP*

<sup>2</sup> *CERN*

<sup>3</sup> *FNAL*

**Corresponding Author:** valeri.onouchine@cern.ch

The RDBC (ROOT DataBase Connectivity) library is a C++ implementation of the The Java Database Connectivity Application Programming Interface. It provides a DBMS-independent interface to relational databases from ROOT as well as a generic SQL database access framework. RDBC also extends the ROOT TSQL abstract interface. Currently it is used in two large experiments:  
- in Minos as interface to MySQL and Oracle databases  
- in Phenix as interface to PostgreSQL database.  
In this paper we will describe the main features and applicability of this library.

**Poster Session 3 - Board: 38 / 99**

## Carrot ROOT Apache Module

**Authors:** F. Rademakers<sup>1</sup>; R. Brun<sup>1</sup>; V. Onuchin<sup>2</sup>

<sup>1</sup> CERN

<sup>2</sup> CERN, IHEP

**Corresponding Author:** [valeri.onouchine@cern.ch](mailto:valeri.onouchine@cern.ch)

Carrot is a scripting module for the Apache webserver. Based on the ROOT framework, it has a number of powerful features, including the ability to embed C++ code into HTML pages, run interpreted and compiled C++ macros, send and execute C++ code on remote web servers, browse and analyse the remote data located in ROOT files with the web browser, access and manipulate databases, and generate graphics on-the-fly, among many others.

In this talk we will describe and demonstrate the main features of Carrot. We will also discuss the future development of this module in context of GRID and integration with PROOF and xrootd/rootd.

More information about Carrot is available from the Carrot website at: <http://carrot.cern.ch>

**Event Processing / 100**

## A Gaussian-sum filter for vertex reconstruction

**Authors:** R. Fruehwirth<sup>1</sup>; T. Speer<sup>2</sup>

<sup>1</sup> *Institut für Hochenergiephysik der OAW, Vienna, Austria*

<sup>2</sup> *UNIVERSITY OF ZURICH, SWITZERLAND*

**Corresponding Author:** [thomas.speer@cern.ch](mailto:thomas.speer@cern.ch)

A vertex fit algorithm was developed based on the Gaussian-sum filter (GSF) and implemented in the framework of the CMS reconstruction program. While linear least-squares estimators are optimal in case all observation errors are Gaussian distributed, the GSF offers a better treatment of the non-Gaussian distribution of track parameter errors when these are modeled by Gaussian mixtures.

In addition, when using electron tracks reconstructed with an electron-reconstruction Gaussian-sum filter, the full mixture can be used rather than the approximation by a single Gaussian.

Properties, results and performance of this filter with simulated data will be shown, and compared to the Kalman filter and to robust filters.

**Poster Session 2 - Board: 34 / 102**

## Usage statistics and usage patterns on the NorduGrid: Analyzing the logging information collected on one of the largest production Grids of the world.

**Authors:** A. KONSTANTINOV<sup>1</sup>; A. WAANANEN<sup>2</sup>; B. Konya<sup>3</sup>; F. OULD-SAADA<sup>4</sup>; J.R. HANSEN<sup>2</sup>; J.L. Nielsen<sup>2</sup>; K. PAJCHEL<sup>4</sup>; M. Ellert<sup>5</sup>; O. SMIRNOVA<sup>6</sup>; P. Eerola<sup>6</sup>; T. EKELOF<sup>7</sup>; U. ERKARSLAN<sup>4</sup>

<sup>1</sup> *Vilnius University, Lithuania*



<sup>2</sup> *Niels Bohr Institute, Copenhagen, Denmark*

<sup>3</sup> *LUND UNIVERSITY, NORDUGRID PROJECT*

<sup>4</sup> *University of Oslo, Norway*

<sup>5</sup> *CERN*

<sup>6</sup> *Lund University, Sweden*

<sup>7</sup> *Uppsala University, Sweden*

**Corresponding Author:** balazs.konya@hep.lu.se

The Nordic Grid facility (Nordugrid) came into production operation during the summer of 2002 when the Scandinavian Atlas HEP group started to use the Grid for the Atlas Data Challenges and was thus the first Grid ever contributing to an Atlas production. Since then, the Grid facility has been in continuous 24/7 operation offering an increasing number of resources to a growing set of active users coming from various scientific areas including chemistry, biology, informatics. As of today the Grid has grown into one of the largest production Grids of the world continuously running Grid jobs on the more than 30 Grid-connected sites which offer over 2000 CPUs.

This article will start with a short overview of the design and implementation of the Advanced Resource Connector (ARC), the Nordugrid middleware, which delivers reliable Grid services to the Nordugrid production facility. This will be followed by a presentation of the logging facility of Nordugrid, describing the logging service and the collected information. The main part of the talk will focus on the analysis of the collected logging information: usage statistics, usage patterns (what is a typical grid job on the Nordugrid looks like?). Use cases from different application domains will also be discussed.

References:

-Nordugrid live: [www.nordugrid.org](http://www.nordugrid.org) -> Grid Monitor

-Atlas Data-Challenge 1 on Nordugrid: <http://arxiv.org/abs/physics/0306013>

**Event Processing / 103**

## Applications of the FLUKA Monte Carlo code in High Energy and Accelerator Physics

**Authors:** A. Empl<sup>1</sup>; A. Fasso<sup>2</sup>; A. Ferrari<sup>3</sup>; A. Ottolenghi<sup>4</sup>; D. Scannicchio<sup>5</sup>; E. Gadioli<sup>6</sup>; F. Ballarini<sup>7</sup>; F. Cerutti<sup>8</sup>; G. Battistoni<sup>9</sup>; J. Ranft<sup>10</sup>; K. TSOULOU<sup>11</sup>; L. Pinsky<sup>12</sup>; M V. Garzelli<sup>13</sup>; M. PELLICIONI<sup>14</sup>; P. Sala<sup>15</sup>; S. MURARO<sup>9</sup>; S. Roesler<sup>16</sup>; T. RANCATT<sup>9</sup>; V. PARINI<sup>17</sup>; V. Vlachoudis<sup>3</sup>

<sup>1</sup> *University of Houston, Texas, USA*

<sup>2</sup> *SLAC, Stanford, USA*

<sup>3</sup> *CERN, Geneva, Switzerland*

<sup>4</sup> *Universita' and INFN Pavia*

<sup>5</sup> *Univeristy and INFN, Pavia, Italy*

<sup>6</sup> *University and INFN, Milano, Italy*

<sup>7</sup> *University and INFN Pavia, Italy*

<sup>8</sup> *Univeristy and INFN Milano, Italy*

<sup>9</sup> *INFN Milano, Italy*

<sup>10</sup> *University of Siegen, Germany*

<sup>11</sup> *CERN Geneva, Switzerland*

<sup>12</sup> *University of Houston, Texas, USA*

<sup>13</sup> *Univeristy and INFN, Milano, Italy*

<sup>14</sup> *INFN, Frascati, Italy*

<sup>15</sup> *ETH, Zurich, Switerland*

<sup>16</sup> *CERN, Geneva, Switzerland*

<sup>17</sup> *INFN and University di Pavia, Italy*

**Corresponding Author:** giuseppe.battistoni@mi.infn.it

The FLUKA Monte Carlo transport code is being used for different applications in High Energy, Cosmic Ray and Accelerator Physics. Here we review some of the ongoing projects which are based on this simulation tool.

In particular, as far as accelerator physics is concerned, we wish to summarize the work in progress for the LHC and the CNGS project. From the point of view of experimental activity, a part the activity going

in the framework of LHC detectors, we wish to discuss as a major example the application of FLUKA to the ICARUS Liquid Argon TPC.

Upgrades in cosmic ray calculations, to demonstrate the capability of FLUKA to reproduce existing experimental data, are also presented

**Grid Security / 104**

## **Authentication/Security services in the ROOT framework**

**Author:** G. GANIS<sup>1</sup>

<sup>1</sup> *CERN*

**Corresponding Author:** gerardo.ganis@cern.ch

The new authentication and security services available in the ROOT framework for client/server applications will be described.

The authentication scheme has been designed with the purpose to make the system complete and flexible, to fit the needs of the coming clusters and facilities.

Three authentication methods have been made available: Globus/GSI, for GRID-awareness; SSH, to allow using a secure and very popular protocol; a fast identification method for intrinsically secure situations.

A mechanism to allow server access control has been implemented, allowing to model the authorization schemes according to the needs.

A lightweight mechanism for client/server method negotiation has been introduced, to adapt to heterogeneous situations.

The forward of the authentication credentials in the PROOF system has been fully automatized.

The modularity of the code has been improved to ease maintenance and reuse in new ROOT modules. In particular, a plug-in library for the new Xrootd file server daemon has been designed and implemented.

Authentication support has been extended to the main socket server class, allowing to run a ROOT interactive session as a full-featured daemon.

Security services have also been added to ROOT. The exchange of sensitive information, e.g. passwords, has been secured.

New socket classes supporting SSL-secured connections have been provided for encryption of all the information exchanged with the remote host.

**Event Processing / 105****Event Data Model in ATLAS****Authors:** E. MOYSE<sup>1</sup>; F. Akesson<sup>2</sup><sup>1</sup> *CERN, Geneva, Switzerland*<sup>2</sup> *CERN***Corresponding Author:** fredrik.akesson@cern.ch

The event data model (EDM) of the ATLAS experiment is presented. For large collaborations like the ATLAS experiment common interfaces and data objects are a necessity to insure easy maintenance and coherence of the experiments software platform over a long period of time. The ATLAS EDM improves commonality across the detector subsystems and subgroups such as trigger, test beam reconstruction, combined event reconstruction, and physics analysis. The object oriented approach in the description of the detector data allows the possibility to have one common raw data flow. Furthermore the EDM allows the use of common software between online data processing and offline reconstruction. One important component of the ATLAS EDM is a common track class which is used for combined track reconstruction across the innermost tracking subdetectors and is also used for tracking in the muon detectors. The structure of the track object and the variety of track parameters are presented. For the combined event reconstruction a common particle class is introduced which serves as the interface between event reconstruction and physics analysis.

**Event Processing / 106****Vertex finding and B-tagging algorithms for the ATLAS Inner Detector****Authors:** A. Wildauer<sup>1</sup>; F. Akesson<sup>2</sup><sup>1</sup> *UNIVERSITY OF INNSBRUCK*<sup>2</sup> *CERN***Corresponding Author:** andreas.wildauer@cern.ch

For physics analysis in ATLAS, reliable vertex finding and fitting algorithms are important. In the harsh environment of the LHC (~ 23 inelastic collisions every 25 ns) this task turns out to be particularly challenging. One of the guiding principles in developing the vertexing packages is a strong focus on modularity and defined interfaces using the advantages of object oriented C++. The benefit is the easy expandability of the vertexing with additional fitting strategies integrated in the Athena framework.

Various implementations of algorithms and strategies dedicated to primary and secondary vertex reconstruction using the full reconstruction of simulated ATLAS events are presented.

Primary and secondary vertex finding is essential for the identification of b-jets in a reconstructed event. Results from a modular and expandable b-tagging algorithm are shown using the presented strategies for vertexing.

**Computer Fabrics / 107****Storage Resource Manager**

**Authors:** D. Petravick<sup>1</sup>; J. Bakken<sup>1</sup>; T. Perelmutov<sup>1</sup>

<sup>1</sup> *FERMI NATIONAL ACCELERATOR LABORATORY*

**Corresponding Author:** timur@fnal.gov

Storage Resource Managers (SRMs) are middleware components whose function is to provide dynamic space allocation and file management on shared storage components on the Grid. SRMs support protocol negotiation and reliable replication mechanism. The SRM standard allows independent institutions to implement their own SRMs, thus allowing for a uniform access to heterogeneous storage elements. SRMs leave the policy decision to be made independently by each implementation at each site. Resource Reservations made through SRMs have limited lifetimes and allow for automatic collection of unused resources thus preventing clogging of storage systems with “forgotten” files.

The storage systems can be classified on basis of their longevity and persistence of their data. Data can also be temporary or permanent. To support these notions, SRM defines Volatile, Durable and Permanent types of files and spaces. Volatile files can be removed by the system to make space for new files upon the expiration of their lifetimes. Permanent files are expected to exist in the storage system for the lifetime of the storage system. Finally Durable files have both the lifetime associated with them and a mechanism of notification of owners and administrators of lifetime expiration, but cannot be deleted automatically by the system and require explicit removal.

Fermilab’s data handling system uses the SRM management interface, the dCache Distributed Disk Cache and the Enstore Tape Storage System as key components to satisfy current and future user requests.

Storage Resource Manager specification is a result of international collaborative effort by representatives of JLAB, LBNL, FNAL, EDG-WP2 and EDG-WP5.

**Core Software / 108****The Athena Control Framework in Production, New Developments and Lessons Learned**

**Authors:** C. Leggett<sup>1</sup>; D. Quarrie<sup>1</sup>; M. Marino<sup>1</sup>; P. Calafiura<sup>1</sup>; W. Lavrijsen<sup>1</sup>

<sup>1</sup> *LBNL*

**Corresponding Author:** pcalafiura@lbl.gov

Athena is the Atlas Control Framework, based on the common Gaudi architecture, originally developed by LHCb. In 2004 two major production efforts, the Data Challenge 2 and the Combined Test-beam reconstruction and analysis were structured as Athena applications. To support the production work we have added new features to both Athena and Gaudi: an “Interval of Validity” service to manage time-varying conditions and detector data; a History service, to manage the provenance information of each event data object; and a toolkit to simulate and analyze the overlay of multiple collisions during the detector sensitive time (pile-up). To support the analysis of simulated and test-beam data in athena we have introduced a python-based scripting interface, based on the CERN LCG tools PyLCGDict, PyRoot and PyBus. The scripting interface allows to fully configure any athena component, interactively

browse and modify this configuration, as well as examine the content of any data object in the event or detector store.

**Poster Session 3 - Board: 49 / 109**

## **The Track Extrapolation package in the new ATLAS Tracking Realm**

**Author:** A. Salzburger<sup>1</sup>

<sup>1</sup> *UNIVERSITY OF INNSBRUCK*

**Corresponding Author:** andreas.salzburger@cern.ch

The ATLAS reconstruction software requires extrapolation to arbitrary oriented surfaces of different types inside a non-uniform magnetic field. In addition multiple scattering and energy loss effects along the propagated trajectories have to be taken into account. A good performance in respect of computing time consumption is crucial due to hit and track multiplicity in high luminosity events at the LHC and the small time window of the ATLAS high level trigger.

Therefore stable and fast algorithms for the propagation of the track parameters and their associated covariance matrices in specific representations to different surfaces in the detector are required.

The recently developed track extrapolation package inside the new ATLAS offline tracking software is presented. Timing performance studies, integration tests with client algorithms and results on ATLAS 2004 Combined Test Beam data are given.

**Event Processing / 110**

## **Test of the ATLAS Inner Detector reconstruction software using combined test beam data**

**Authors:** M.J. Costa<sup>1</sup>; W. Liebig<sup>1</sup>

<sup>1</sup> *CERN*

**Corresponding Authors:** wolfgang.liebig@cern.ch, maria.jose.costa@cern.ch

The athena software framework for event reconstruction in ATLAS will be employed to analyse the data from the 2004 combined test beam. In this combined test beam, a slice of the ATLAS detector is operated and read out under conditions similar to future LHC running, thus providing a test-bed for the complete reconstruction chain. First results for the ATLAS InnerDetector will be presented.

In particular, the reading of the bytestream data inside athena, the monitoring tasks, the alignment techniques and all the different online and offline reconstruction algorithms will be fully tested with real data. Their performance will be studied and results compared to simulated data, which has been generated specifically for the test beam layout.

**Poster Session 3 - Board: 60 / 111**

## Simulation and reconstruction of heavy ion collisions in the ATLAS detector.

**Author:** P. Nevski<sup>1</sup>

<sup>1</sup> *BROOKHAVEN NATIONAL LABORATORY*

**Corresponding Author:** nevski@bnl.gov

The ATLAS detector is a sophisticated multi-purpose detector with over 10 million electronics channels designed to study high-pT physics at LHC. Due to their high multiplicity, reaching almost hundred thousand particles per event, heavy ion collisions pose a formidable computational challenge. A set of tools have been created to realistically simulate and fully reconstruct the most difficult case of central Pb-Pb collisions (impact parameter < 1 fm) in the ATLAS detector.

A number of issues concerning extensive memory management, CPU versus memory optimization, tradeoff between data volume and physics analysis capacity have been formulated and solved. As a result we are able to predict and optimise the physics performance of the experiment and its sub-systems. We will describe the optimal dataflow organization and solutions which allowed flexible system tuning during the massive simulated data production and the analysis of tens of thousand of multi-megabyte events.

**Poster Session 3 - Board:** 54 / 112

## BFD: A software management & deployment tool for mixed-language distributed projects

**Author:** M. Stoufer<sup>1</sup>

**Co-author:** S. Patton<sup>1</sup>

<sup>1</sup> *LAWRENCE BERKELEY NATIONAL LAB*

**Corresponding Author:** mcstoufer@lbl.gov

As any software project grows in both its collaborative and mixed codebase nature, current tools like CVS and Maven start to sag under the pressure of complex sub-project dependencies and versioning. A developer-wide failure in mastery of these tools will inevitably lead to an unrecoverable instability of a project. Even keeping a single software project stable in a large collaborative environment has proved a difficult venture in which numerous home-spun and commercial tools have yet to fully succeed.

BFD looks to solve the problems inherent with large scale software projects that span multiple software mixed-language projects. This is accomplished two-fold. BFD extends the versioning methodology of CVS or Maven by enforcing a rich data type format for its version tags. BFD also improves on the naive build ideologies of the developers IDE by being able to resolve complex dependencies between non-related projects as well as knowing when incompatible dependencies cannot be resolved. The concept of the Meta project has also been introduced to allow projects to be grouped together in a logical manner. Thus allowing varying versions of said projects to be kept track of by an overarching framework.

Poster Session 2 - Board: 11 / 113

## Testing the CDF Distributed Computing Framework

**Authors:** A. Baranovski<sup>1</sup>; A. Kreymer<sup>2</sup>; A. Lyon<sup>2</sup>; A. Sill<sup>3</sup>; F. Ratnikov<sup>4</sup>; G. Garzoglio<sup>2</sup>; I. Terekhov<sup>1</sup>; J. Trumbo<sup>1</sup>; L. Loebel Carpenter<sup>1</sup>; L. Lueking<sup>2</sup>; M. Burgon-Lyon<sup>5</sup>; M. Leslie<sup>6</sup>; R. Herber<sup>2</sup>; R. Illingworth<sup>2</sup>; R. Kennedy<sup>2</sup>; R. St.Denis<sup>5</sup>; S. Belforte<sup>7</sup>; S. Stonjek<sup>8</sup>; S. Veseli<sup>1</sup>; S. White<sup>1</sup>; U. Kerzel<sup>9</sup>; V. Bartsch<sup>10</sup>; W. Merritt<sup>2</sup>

<sup>1</sup> Fermilab

<sup>2</sup> FERMI NATIONAL ACCELERATOR LABORATORY

<sup>3</sup> Texas Tech. University

<sup>4</sup> Rutgers University

<sup>5</sup> Glasgow University

<sup>6</sup> Oxford University

<sup>7</sup> INFN, Trieste

<sup>8</sup> FNAL

<sup>9</sup> Karlsruhe University

<sup>10</sup> OXFORD UNIVERSITY

To distribute computing for CDF (Collider Detector at Fermilab) a system managing local compute and storage resources is needed. For this purpose CDF will use the DCAF (Decentralized CDF Analysis Farms) system which is already at Fermilab. DCAF has to work with the data handling system SAM (Sequential Access to data via Metadata). However, both DCAF and SAM are mature systems which have not yet been used in combination, and on top of this DCAF has only been installed at Fermilab and not on local sites. Therefore tests of the systems are necessary to test the interplay of the data handling with the farms, the behaviour of the off-site DCAFs and the user friendliness of the whole system. The tests are focussed on the main tasks of the DCAFs, like Monte Carlo generation and stores, as well as the readout of data files and connected data handling. To achieve user friendliness the SAM station environment has to be common to all stations and adaptations to the environment have to be made.

Poster Session 2 - Board: 41 / 114

## Experience integrating a General Information System API in LCG Job Management and Monitoring Services

**Authors:** A. Delgado Peris<sup>1</sup>; A. Sciaba<sup>1</sup>; F. Donno<sup>1</sup>; P. Mendez Lorenzo<sup>1</sup>; R. Santinelli<sup>1</sup>; S. Campana<sup>1</sup>

<sup>1</sup> CERN IT/GD

**Corresponding Author:** patricia.mendez@cern.ch

In a Grid environment, the access to information on system resources is a necessity in order to perform common tasks such as matching job requirements with available resources, accessing files or presenting monitoring information. Thus both middleware service, like workload and data management, and applications, like monitoring tools, require an interface to the Grid information service which provides that data.

Even though a unique schema for the published information is defined, actual implementations use different data models, and define different access protocols. Applications interacting with the information service must therefore deal with several APIs, and be aware of the underlying technology in order to use the appropriate syntax for their queries or to publish new information.

We have produced a new high level C++ API that accommodates several existing implementations of the information service such as Globus MDS(LDAP based), MDS3(XML

based) an R-GMA(SQL based). It allows applications to access information in a transparent manner loading the needed implementation specific library on demand. Features allowing for the adding and removal of dynamic information have been included as well. A general query language to make the API compatible with future protocols has been used.

In this paper we described the design of this API and the results obtained integrating this API in the Workload Management system and in the GridIce monitoring system of LCG.

## Distributed Computing Systems and Experiences / 115

### ATLAS Distributed Analysis

**Authors:** A. Soroko<sup>1</sup>; C. Haerberli<sup>2</sup>; C. Kannan<sup>3</sup>; C.L. Tan<sup>4</sup>; D. Adams<sup>5</sup>; D. Liko<sup>6</sup>; F. Fassi<sup>7</sup>; F. Orellana<sup>6</sup>; G. Rybkin<sup>None</sup>; J. Fulachier<sup>8</sup>; J. Lozano<sup>7</sup>; K. Harrison<sup>9</sup>; M. Branco<sup>6</sup>; N. Chetan<sup>3</sup>; S. Albrand<sup>8</sup>; V. Sambamurthy<sup>3</sup>; W. Deng<sup>5</sup>

<sup>1</sup> *Oxford University*

<sup>2</sup> *Bern*

<sup>3</sup> *Stonybrook*

<sup>4</sup> *Birmingham*

<sup>5</sup> *BNL*

<sup>6</sup> *CERN*

<sup>7</sup> *IFIC Valencia*

<sup>8</sup> *LPSC, Grenoble*

<sup>9</sup> *Cambridge University*

**Corresponding Author:** dladams@bnl.gov

The ATLAS distributed analysis (ADA) system is described. The ATLAS experiment has more than 2000 physicists from 150 institutions in 34 countries. Users, data and processing are distributed over these sites. ADA makes use of a collection of high-level web services whose interfaces are expressed in terms of AJDL (abstract job definition language) which includes descriptions of datasets, transformations and jobs. The high-level services are implemented using generic parts of these objects while clients and endpoint applications additionally make use of experiment-specific extensions. The key high-level service is the analysis service which receives a generic job request and creates and runs a corresponding job, typically as a collection of sub-jobs each handling a subset of the input dataset. The submitting client is able to monitor the progress of the job including partial results. The system is capable of running a wide range of applications but the emphasis is on event processing, in particular simulation, reconstruction and analysis of ATLAS data. Other high-level services include catalogs and dataset splitters and mergers. The ATLAS production system has been used to construct an analysis service that makes production activities available to ATLAS users. An analysis service with interactive response is provided by DIAL. Another analysis service based on the EGEE middleware is being constructed in the context of the ARDA project. All are accessible from ROOT and python command lines and from the user-friendly graphical interface provided by GANGA.



## Experiment Software Installation experience in LCG-2

**Authors:** A. Delgado Peris<sup>1</sup>; A. Sciaba<sup>1</sup>; F. Donno<sup>1</sup>; P. Mendez Lorenzo<sup>2</sup>; R. santinelli<sup>1</sup>; S. Campana<sup>1</sup>

<sup>1</sup> CERN/IT/GD

<sup>2</sup> CERN/IT/GD

**Corresponding Author:** roberto.santinelli@cern.ch

The management of Application and Experiment Software represents a very common issue in emerging grid-aware computing infrastructures.

While the middleware is often installed by system administrators at a site via customized tools that serve also for the centralized management of the entire computing facility, the problem of installing, configuring and validating Gigabytes of Virtual Organization (VO) specific software or frequently changing user applications remains an open issue.

Following the requirements imposed by the experiments, in the LHC Computing Grid (LCG) Experiment Software Managers (ESM) are designated people with privileges of installing, removing and validating software for a specific VO on a per site basis.

They can manage univocally identifying tags in the LCG Information System to announce the availability of a specific software version.

Users of a VO can then select, via the published tag, sites to run their jobs.

The solution adopted by LCG has mainly served its purpose but it presents many problems.

The requirement imposed by the present solution for the existence of a shared file-system in a computing farm poses performance, reliability and scalability issues for large installations.

With this work we present a more flexible service based on P2P

technology that has been designed to tackle the limitation of the current system.

This service allows the ESM to propagate the installation occurring in a given WN to the rest of the farm elements.

We illustrate the deployment, the design, preliminary results obtained and the feedback from the LHC experiments and sites that have adopted it.

**Event Processing / 117**

## Implementation and Performance of the High-Level Trigger electron and photon selection for the ATLAS experiment at the LHC

**Author:** M. DIAZ GOMEZ<sup>None</sup>

The ATLAS experiment at the Large Hadron Collider (LHC) will face the challenge of efficiently selecting interesting candidate events in pp collisions at 14 TeV center-of-mass energy, whilst rejecting the enormous number of background events, stemming from an interaction rate of about  $10^9$  Hz. The Level-1 trigger will reduce the incoming rate to around  $O(100)$  kHz. Subsequently, the High-Level Triggers (HLT), which are comprised of the second level trigger and the event filter, will need to reduce this rate further by a factor of  $O(10^3)$ . The HLT selection is software based and will be implemented on commercial CPUs using a common framework, which is based on the standard ATLAS object-oriented software architecture. In this talk an overview of the current implementation of the selection for electrons and photons in the trigger is given. The performance of this implementation has been evaluated using Monte Carlo simulations in terms of the efficiency for the signal channels, the rate expected for the selection, the data preparation times, and the algorithm execution times. Besides the efficiency and rate estimates, some physics examples will be discussed, showing that the triggers are well adapted for the physics programme envisaged at LHC. The electron/gamma trigger software has been also integrated in the ATLAS 2004 combined test-beam, to validate the chosen selection architecture in a real on-line environment.

Poster Session 1 - Board: 1 / 118

## **AutoBlocker: A system for detecting and blocking of network scanning based on analysis of netflow data.**

**Authors:** A. Bobyshev<sup>1</sup>; D. Lamore<sup>1</sup>; P. Demar<sup>1</sup>

<sup>1</sup> FERMILAB

**Corresponding Author:** bobyshev@fnal.gov

In a large campus network, such as Fermilab's ten thousand nodes, scanning initiated from either outside of or within the campus network raises security concerns, may have very serious impact on network performance, and even disrupt normal operation of many services. In this paper we introduce a system for detecting and automatic blocking of excessive traffic of different nature, scanning, DoS attacks, virus infected computers. The system, called AutoBlocker, is a distributed computing system based on quasi-real time analysis of network flow data collected from the border router and core routers. AutoBlocker also has an interface to accept alerts from the IDS systems (e.g. BRO, SNORT) that are based on other technologies. The system has multiple configurable alert levels for the detection of anomalous behavior and configurable trigger criteria for automated blocking of the scans at the core or border routers. It has been in use at Fermilab for about 2 years, and become a very valuable tool to curtail scan activity within the Fermilab campus network.

Online Computing / 119

## **The DAQ system for the Fluorescence Detectors of the Pierre Auger Observatory**

**Authors:** A. KOPMANN<sup>1</sup>; H.-J. Mathes<sup>2</sup>; O. MARTINEAU<sup>3</sup>; S. Argiro<sup>4</sup>

<sup>1</sup> Forschungszentrum Karlsruhe, Germany

<sup>2</sup> FORSCHUNGSZENTRUM KARLSRUHE, INSTITUT FÜR KERNPHYSIK

<sup>3</sup> Forschungszentrum Karlsruhe, Germany

<sup>4</sup> INFN Sezione Torino

**Corresponding Author:** hermann-josef.mathes@cern.ch

S.Argiro<sup>(1)</sup>, A. Kopmann<sup>(2)</sup>, O.Martineau<sup>(2)</sup>, H.-J. Mathes<sup>(2)</sup>  
for the Pierre Auger Collaboration

(1) INFN, Sezione Torino

(2) Forschungszentrum Karlsruhe

The Pierre Auger Observatory currently under construction in Argentina will investigate extensive air showers at energies above  $10^{18}$  eV. It consists of a ground array of 1600 Cherenkov water detectors and 24 fluorescence telescopes to discover the nature and origin of cosmic rays at these ultra-high energies.

The ground array is overlooked by 4 different fluorescence buildings which are equipped with 6 telescopes each. An independent local data acquisition (DAQ) is running in each building to readout 480 channels per telescope. In addition, a central DAQ merges data coming from the water detectors and all fluorescence buildings.

The system architecture follows the object oriented paradigm and has been implemented using several of the most widespread open source tools for interprocess communication, data storage and user interfaces.

Each local DAQ is connected with further sub-systems for calibration, for monitoring of atmospheric parameters and slow control. The latter is responsible for general safety functions and the experiment control.

After a prototype phase to validate the system concept the Observatory is taking data in the final setup since September 2003. The data taking will continue during the construction phase and the integration of all sub-systems.

We present the design and the present status of the system currently running in two different buildings with a total of 8 telescopes installed.

**Poster Session 3 - Board: 30 / 120**

## **Transparently managing time varying conditions and detector data on ATLAS.**

**Author:** C. Leggett<sup>1</sup>

<sup>1</sup> *LAWRENCE BERKELEY NATIONAL LABORATORY*

**Corresponding Author:** charles.g.leggett@gmail.com

It is essential to provide users transparent access to time varying data, such as detector misalignments, calibration parameters and the like. This data should be automatically updated, without user intervention, whenever it changes. Furthermore, the user should be able to be notified whenever a particular datum is updated, so as to perform actions such as re-caching of compound results, or performing computationally intensive task only when necessary. The user should only have to select a particular calibration scheme or time interval, without having to worry about explicitly updating data on an event by event basis. In order to minimize database activity, it is important that the system only manage the parameters that are actively used in a particular job, making updates only on demand. For certain situations however, such as testbeam environments, pre-caching of data is essential, so the system must also be able to pre-load all relevant data at the start of a run, and avoid further updates to the data.

In this talk we present the scheme for managing time varying data and their associated intervals of validity, as used in the Athena framework on ATLAS, which features automatic updating of conditions data occurring invisibly to the user; automatic and explicit registration of objects of interest; callback function hierarchies; and abstract conditions database interfaces.

**Wide Area Networking / 121**

## **Networking for High Energy and Nuclear Physics as Global E-Science**

**Author:** H. Newman<sup>1</sup>

**Co-author:** (Does Not Apply) on Behalf of the ICFA Standing Committee on Inter-Regional Connectivity<sup>2</sup>

<sup>1</sup> *Caltech*

<sup>2</sup> *ICFA*

**Corresponding Author:** harvey.newman@cern.ch

Wide area networks of sufficient, and rapidly increasing end-to-end capability are vital for every phase of high energy physicists' work. Our bandwidth usage, and the typical capacity of the major national backbones and intercontinental links used by our field have progressed by a factor of more than 1000 over the past decade, and the outlook is for a similar increase over the next decade, as we enter the era of LHC physics served by Grids on a global scale. Responding to these trends, and the emerging need to provide rapid access and distribution of Petabyte-scale datasets, physicists working with network engineers and computer scientists are learning to use networks effectively in the 1-10 Gigabit/range, placing them among the leading developers of global networks.

In this talk I review the network requirements and usage trends, and present a bandwidth roadmap for HEP and other fields of "data intensive" science. I give an overview of the status and outlook for the world's research networks, technology advances, and the problem of the Digital Divide, based on the recent work of ICFA's Standing Committee on Inter-regional Connectivity (SCIC). Finally, I discuss the role of high speed networks in the next generation of Grid systems that are now being constructed to support data analysis for the LHC experiments.

[This is a candidate Plenary Presentation.]

**Grid Security / 122**

## **A Scalable Grid User Management System for Large Virtual Organization**

**Authors:** D. Yu<sup>1</sup>; G. Carcassi<sup>1</sup>; G. Smith<sup>1</sup>; J. Smith<sup>1</sup>; J. Spiletic<sup>1</sup>; T. Carter<sup>1</sup>; T. Wlodek<sup>1</sup>; X. Zhao<sup>1</sup>; Z. Liu<sup>1</sup>

<sup>1</sup> *BROOKHAVEN NATIONAL LABORATORY*

**Corresponding Authors:** carcassi@bnl.gov, dtyu@bnl.gov

We present a work-in-progress system, called GUMS, which automates the processes of Grid user registration and management and supports policy-aware authorization as well. GUMS builds on existing VO management tools (LDAP VO, VOMS and VOMRS) with a local grid user management system and a site database which stores user credentials, accounting history and policies in XML format. We use VOMRS, being developed by Fermilab, to collect user information and register legitimate users into the VOMS server.

Our local grid user management system jointly retrieves user information and VO policies from multiple VO databases based on site security policies. Authorization can be done by mapping the user's credential to local accounts. Four different mapping schemes have been implemented: user's existing account, recyclable pool account, non-recyclable pool account and group shared account. The mapping selection is determined by the type of target resource and its usage policies. We already deployed our automatic grid mapfile generators on the BNL Grid Gatekeeper, GridFtp server and HPSS mass storage system. Work is in progress to enable "single-sign-on"

based upon X509 certificate credential for job execution and access to both disk and tape storage resources.

Poster Session 2 - Board: 22 / 126

## Grid Deployment Experiences: The path to a production quality LDAP based grid information system

**Authors:** L. Field<sup>1</sup>; M. Schulz<sup>1</sup>

<sup>1</sup> CERN

**Corresponding Author:** laurence.field@cern.ch

This paper reports on the deployment experience of the defacto grid information system, Globus MDS, in a large scale production grid. The results of this experience led to the development of an information caching system based on a standard openLDAP database. The paper then describes how this caching system was developed further into a production quality information system including a generic framework for information providers. This includes the deployment and operation experience and the results from performance tests on the information system to assess the scalability limits of it.

Poster Session 2 - Board: 24 / 128

## XTNetFile, a fault tolerant extension of ROOT TNetFile

**Authors:** A. Dorigo<sup>1</sup>; A. Hanushevsky<sup>2</sup>; F. Furano<sup>3</sup>; P. Elmer<sup>4</sup>

<sup>1</sup> INFN PADOVA

<sup>2</sup> SLAC

<sup>3</sup> INFN Padova

<sup>4</sup> Princeton University

**Corresponding Author:** alvise.dorigo@pd.infn.it

This paper describes XTNetFile, the client side of a project conceived to address the high demand data access needs of modern physics experiments such as BaBar using the ROOT framework. In this context, a highly scalable and fault tolerant client/server architecture for data access has been designed and deployed which allows thousands of batch jobs and interactive sessions to effectively access the data repositories basing on the XROOTD data server, a complex extension of the rootd daemon. The majority of the communication problems are handled by the design of the client/server mechanism and the communication protocol.

This allows us to build distributed data access systems which are highly robust, load balanced and scalable to an extent which allows 'no jobs to fail'.

Furthermore XTNetFile ensures backward compatibility with the 'old' rootd server by using same API as the existing ROOT TFile/TNetFile classes.

The code is designed with a high degree of modularity that allows to build other interfaces, such as administrative tools, based on the

same communication layer. In addition the client plugin can also be used to read other types of (non-ROOT I/O) data files, providing the same benefits.

Poster Session 2 - Board: 10 / 129

## Production of simulated events for the BaBar experiment by using LCG

**Authors:** C. Bozzi<sup>1</sup>; D. Andreotti<sup>2</sup>; E. Antonioli<sup>2</sup>; E. Luppi<sup>2</sup>; M. Melani<sup>3</sup>; P. Veronesi<sup>2</sup>

<sup>1</sup> INFN SEZIONE DI FERRARA

<sup>2</sup> INFN Sezione di Ferrara

<sup>3</sup> SLAC

**Corresponding Author:** bozzi@fe.infn.it

The BaBar experiment has been taking data since 1999. In 2001 the computing group started to evaluate the possibility to evolve toward a distributed computing model in a Grid environment. In 2003, a new computing model, described in other talks, was implemented, and ROOT I/O is now being used as the Event Store. We implemented a system, based on the LHC Computing Grid (LCG) tools, to submit full-scale MonteCarlo simulation jobs in this new BaBar computing model framework. More specifically, the resources of the LCG implementation in Italy, grid.it, are used as computing elements (CE) and Worker Nodes (WN). A Resource Broker (RB) specific for the BaBar computing needs was installed. Other BaBar requirements, such as the installation and usage of an object-oriented (Objectivity) Database to read detector conditions and calibration constants, were accommodated by using non-gridified hardware in a subset of grid.it sites. The BaBar simulation software was packed and installation on Grid elements was centrally managed with LCG tools. Sites were geographically mapped to Objectivity databases, and conditions were read by the WN either locally or remotely. An LCG User Interface (UI) has been used to submit simulation tests by using standard JDL commands. The ROOT I/O output files were retrieved from the WN and stored in the closest Storage Element (SE). Standard BaBar simulation production tools were then installed on the UI and configured such that the resulting simulated events can be merged and shipped to SLAC, like in the standard BaBar simulation production setup. Final validation of the system is being completed. This gridified approach results in the production of simulated events on geographically distributed resources with a large throughput and minimal, centralized system maintenance.

Poster Session 1 - Board: 29 / 130

## The Configurations Database Challenge in the ATLAS DAQ System

**Authors:** A. AMORIM<sup>1</sup>; A. Kazarov<sup>2</sup>; D. BURCKHART-CHROMEK<sup>3</sup>; D. KLOSE<sup>4</sup>; D. Liko<sup>5</sup>; E. BADESCU<sup>6</sup>; I. ALEXANDROV<sup>7</sup>; I. Soloviev<sup>8</sup>; J. FLAMMER<sup>3</sup>; L. MAPELLI<sup>3</sup>; L. PEDRO<sup>4</sup>; M. CAPRINI<sup>6</sup>; M. DOBSON<sup>3</sup>; M. Mineev<sup>9</sup>; N. FIUZA DE BARROS<sup>4</sup>; R. JONES<sup>3</sup>; S. KOLOS<sup>5</sup>; S. KOROBV<sup>7</sup>; V. Kotov<sup>9</sup>; Y. RYABOV<sup>10</sup>

<sup>1</sup> Faculdade de Ciencias, Lisbon, Portugal

<sup>2</sup> Petersburg Nuclear Physics Institute (PNPI)

<sup>3</sup> CERN, Geneva, Switzerland

<sup>4</sup> Universidade de Lisboa, Portugal

<sup>5</sup> CERN

<sup>6</sup> National Institute of Physics and Nuclear Engineering, Bucharest, Romania

<sup>7</sup> *JINR, Dubna, Russia*

<sup>8</sup> *CERN/PNPI*

<sup>9</sup> *Joint Institute for Nuclear Research (JINR)*

<sup>10</sup> *PNPI, St-Petersburg, Russia*

**Corresponding Author:** igor.soloviev@cern.ch

The ATLAS data acquisition system uses the database to describe configurations for different types of data taking runs and different sub-detectors. Such configurations are composed of complex data objects with many inter-relations. During the DAQ system initialisation phase the configurations database is simultaneously accessed by a large number of processes. It is also required that such processes be notified about database changes that happen during or between data-taking runs.

The paper describes the architecture of the configurations database. It presents the set of graphical tools which are available for the database schema design and the data editing. The automatic generation of data access libraries for C++ and Java languages is also described. They provide the programming interfaces to access the database either via a common file system or via remote database servers, and the notification mechanism on data changes.

The paper presents results of recent performance and scalability tests, which allow a conclusion to be drawn about the applicability of the current configurations database implementation in the future DAQ system.

**Poster Session 3 - Board: 2 / 133**

## **The TAG Collector - A Tool for Atlas Code Release Management**

**Authors:** F. Lambert<sup>1</sup>; J. Collot<sup>1</sup>; J. Fulachier<sup>1</sup>; S. Albrand<sup>1</sup>

<sup>1</sup> *LPSC*

**Corresponding Author:** solveig.albrand@lpsc.in2p3.fr

The Tag Collector is a web interfaced database application for release management. The tool is tightly coupled to CVS, and also to CMT, the configuration management tool. Developers can interactively select the CVS tags to be included in a build, and the complete build commands are produced automatically. Other features are provided such as verification of package CMT requirements files, and direct links to the package documentation, making it a useful tool for all ATLAS users. The software for the Atlas experiment contains about 1 MSLOC. It is organized in over 50 container packages containing about 500 source code packages. One or several developers maintain each package. ATLAS developers are widely distributed geographically.

The Tag Collector was designed and implemented during the summer of 2001, in response to a near crisis situation. It has been in use since September 2001. Until this time the ATLAS librarian constructed a build of the software release after a cascade of e-mails from developers; communicating the correct CVS code repository version tag of their respective packages. This was subject to all sorts of human errors, and inefficient in our multi-time zone environment. In addition, it was difficult to manage the contents of a release. It was all too easy for a prolific developer to introduce a well-intentioned change in his package just before a build, often with unsuspected border effects. Developers were also asking for regular, and frequent developer builds.

The tool has proved extremely successful, and features that are outside the scope of the original design have been requested. Requirements for a new version were

collected during 2003, culminating in a formal review in December 2003. The new version is currently being designed. It will be more flexible and easier to maintain.

#### Core Software / 134

## ATLAS Metadata Interfaces (AMI) and ATLAS Metadata Catalogs

**Authors:** J. Fulachier<sup>1</sup>; S. Albrand<sup>1</sup>

<sup>1</sup> LPSC

**Corresponding Author:** solveig.albrand@lpsc.in2p3.fr

The ATLAS Metadata Interface (AMI) project provides a set of generic tools for managing database applications. AMI 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. The top layer contains application specific features. The principal uses of AMI are the ATLAS Data Challenge dataset bookkeeping catalogs, and Tag Collector, a tool for release management.

The first AMI Web service client was introduced in early 2004. It offers many advantages over earlier clients because:

- Web services permit multi-language and multi-operating system support
- The user interface is very effectively de-coupled from the implementation.

Most upgrades can be implemented on the server side; no redistribution of client software is needed. In 2004 this client will be used for the ATLAS Data Challenge 2, for the ATLAS combined test beam offline bookkeeping, and also in the first prototypes of ARDA compliant analysis interfaces.

#### Event Processing / 135

## Panoramix

**Author:** G B. Barrand<sup>1</sup>

<sup>1</sup> CNRS / IN2P3 / LAL

**Corresponding Author:** barrand@lal.in2p3.fr

Panoramix is an event display for LHCb. LaJoconde is an interactive environment over DaVinci, the analysis software layer for LHCb. We shall present global technological choices behind these two softwares : GUI, graphic, scripting, plotting. We shall present the connection to the framework (Gaudi), how we can integrate other tools like hippodraw. We shall present the overall capabilities to these systems and their today status. We shall outline also how good part of choosen technologies may be reused to build the same kind of interactive environments for ATLAS (LAL Agora prototype).



**Distributed Computing Systems and Experiences / 136****Tier-1 and Tier-2 Real-time Analysis experience in CMS Data Challenge 04**

**Authors:** A. Fanfani<sup>1</sup>; A. Pierro<sup>2</sup>; C. Grandi<sup>1</sup>; D. Bonacorsi<sup>3</sup>; F. Ambroglini<sup>4</sup>; F. Fanzago<sup>5</sup>; G. Donvito<sup>6</sup>; J. Hernandez<sup>7</sup>; L. Silvestris<sup>2</sup>; M. Corvo<sup>5</sup>; N. De Filippis<sup>8</sup>; V. Innocente<sup>9</sup>

<sup>1</sup> *INFN Bologna*

<sup>2</sup> *INFN Bari*

<sup>3</sup> *CNAF-INFN Bologna*

<sup>4</sup> *Universita' di Perugia and INFN*

<sup>5</sup> *INFN Padova*

<sup>6</sup> *Universita' di Bari*

<sup>7</sup> *CIEMAT Madrid*

<sup>8</sup> *UNIVERSITA' DEGLI STUDI DI BARI AND INFN*

<sup>9</sup> *CERN, PH/SFT*

During the CMS Data Challenge 2004 a realtime analysis was attempted at INFN and PIC Tier-1 and Tier-2s in order to test the ability of the instrumented methods to quickly process the data.

Several agents and automatic procedures were implemented to perform the analysis at the Tier-1/2 synchronously with the data transfer from Tier-0 at CERN. The system was implemented in the Grid LCG-2 environment and allowed on-the-fly job preparation and subsequent submission to the Resource Broker as new data come along.

Running job accessed data from the Storage Elements through POOL via remote file protocol, whenever possible, or copying them locally with gridftp.

Job monitoring and bookkeeping was performed using BOSS.

Details of the procedures adopted to run the analysis jobs and the expected results are described.

An evaluation of the ability of the system to maintain an analysis rate at Tier-1 and Tier-2 comparable with the data transfer rate is also presented.

The results on the analysis timeline, the statistics of submitted jobs, the overall efficiency of the GRID services and the overhead introduced by the agents/procedures are reported. Performances and possible bottlenecks of the whole procedure are discussed.

**Poster Session 3 - Board: 47 / 137**

**CMS Software Installation**

**Authors:** A. Nowack<sup>1</sup>; A. Sciaba<sup>2</sup>; J. Weng<sup>3</sup>; K. Rabbertz<sup>4</sup>; M. Corvo<sup>5</sup>; S. MUZAFFAR<sup>6</sup>; S. Wynhoff<sup>7</sup>

<sup>1</sup> *RWTH Aachen*

<sup>2</sup> *CERN*

<sup>3</sup> *CERN & University of Karlsruhe*

<sup>4</sup> *UNIVERSITY OF KARLSRUHE*

<sup>5</sup> *INFN Padova*

<sup>6</sup> *NorthEastern University, Boston, USA*

<sup>7</sup> *Princeton University*

**Corresponding Author:** klaus.rabbertz@cern.ch

For data analysis in an international collaboration it is important to have an efficient procedure to distribute, install and update the centrally maintained software. This is even more true when not only locally but also grid accessible resources are to be exploited. A practical solution will be presented that has been successfully employed for CMS software installations on systems ranging from physicists' notebooks up to LCG2 enabled clusters. It is based on perl for an automated production of rpm's and xcmsi, a tool written in perl and perl/Tk, to facilitate installing, updating and verifying our rpm packaged software.

**Computer Fabrics / 138**

## 64-Bit Opteron systems in High Energy and Astroparticle Physics

**Authors:** P. Wegner<sup>1</sup>; S. Wiesand<sup>1</sup>

<sup>1</sup> *DESY*

**Corresponding Authors:** stephan.wiesand@desy.de, peter.wegner@desy.de

64-Bit commodity clusters and farms based on AMD technology meanwhile have been proven to achieve a high computing power in many scientific applications. This report first gives a short introduction into the specialties of the amd64 architecture and the characteristics of two-way Opteron systems. Then results from measuring the performance and the behavior of such systems in various Particle Physics applications as compared to the classical 32-Bit systems are presented. The investigations cover analysis tools like ROOT, Astrophysics simulations based on CORSIKA and event reconstruction programs. Another field of investigations are parallel high performance clusters for Lattice QCD calculations, and n-loop calculations based on perturbative methods in quantum field theory using the formula manipulation program FORM. In addition to the performance results the compatibility of 32- and 64-Bit architectures and Linux operating system issues, as well as the impact on fabric management are discussed. It is shown that for most of the considered applications the recently available 64-bit commodity computers from AMD are a viable alternative to comparable 32-Bit systems.

**Distributed Computing Services / 142**

## Don Quijote - Data Management for the ATLAS Automatic Production System

**Author:** M. Branco<sup>1</sup>

<sup>1</sup> *CERN*

**Corresponding Author:** miguel.branco@cern.ch

As part of the ATLAS Data Challenges 2 (DC2), an automatic production system was introduced and with it a new data management component.

The data management tools used for previous Data Challenges were built as separate components from the existing Grid middleware. These tools relied on a database of its own which acted as a replica catalog.

With the extensive use of Grid technology expected for the most part of the DC2 production, no longer can a data management tool be independent of the Grid middleware. Each Grid relies on its own replica catalog and not on an ATLAS specific tool.

ATLAS DC will attempt to use uniformly the resources provided by three Grids: NorduGrid, US Grid3 and LCG-2. Legacy system will be supported as well.

The proposed solution was to build a data management proxy system which consists of a common high-level interface, whose implementation depends on each Grid's replica and metadata catalog as well as the storage backend (mainly "classic" GridFTP servers and SRM).

Don Quijote provides management of replicas in a services oriented architecture, across the several "flavours" of Grid middleware used by ATLAS DC.

With a higher-level interface common across several Grids (and legacy systems) a user (such as the new automatic production system) can seamlessly manage replicas independently of their hosting environment. Given the services-based architecture, a lightweight command line tool is capable of interacting uniformly within each Grid and between Grids (e.g. moving files from LCG-2 to US Grid 3 while maintaining attributes such as the Global Unique Identifier).

## Event Processing / 143

### FAMOS, a FASt MOnte-Carlo Simulation for CMS

**Author:** F. Beaudette<sup>1</sup>

<sup>1</sup> CERN

**Corresponding Author:** [florian.beaudette@cern.ch](mailto:florian.beaudette@cern.ch)

An object-oriented FASt MOnte-Carlo Simulation (FAMOS) has recently been developed for CMS to allow rapid analyses of all final states envisioned at the LHC while keeping a high degree of accuracy for the detector material description and the related particle interactions. For example, the simulation of the material effects in the tracker layers includes charged particle energy loss by ionization and multiple scattering, electron Bremsstrahlung and photon conversion. The particle showers are developed in the calorimeters with an emulation of GFLASH, finely interfaced with the calorimeter geometry (e.g., crystal positions, cracks, rear leakage, etc).

As the same software framework is used for FAMOS and ORCA (the full Object-oriented Reconstruction software for CMS Analysis), the various Physics Objects (electrons, photons, muons, taus, jets, missing ET, charged particle tracks, ...) can be accessed with a similar code with both fast and full simulation, thus allowing any analysis algorithm to be transported from FAMOS to ORCA (and later, to data analysis and DST reading) or vice-versa without any additional work.

Altogether, a gain in CPU time of about a hundred can be achieved with respect to the full simulation, with little loss in precision.

## Plenary / 144

## Network Architecture: lessons from the past, vision for the future

**Author:** F. Fluckiger<sup>1</sup>

<sup>1</sup> CERN

**Corresponding Author:** francois.fluckiger@cern.ch

The Architectural Principles of the Internet have dominated the past decade. Orthogonal to the telecommunications industry principles, they dramatically changed the networking landscape because they relied on iconoclastic ideas. First, the Internet end-to-end principle, which stipulates that the network should intervene minimally on the end-to-end traffic, pushing the complexity to the end-systems. Second, the ban of centralized functions: all the Internet techniques (routing, DNS, management) are based on distributed, decentralized mechanisms. Third, the absolute domination of connectionless (stateless) protocols (as with IP, HTTP).

However, when facing new requirements: multimedia traffic, security, Grid applications, these principles appear sometimes as architectural barriers. Multimedia requires QoS guarantees, but stateless systems are not good at QoS. Security requires active, intelligent networks, but dumb routers or plain end-to-end mail systems are insufficient. Grid applications require middleware overlay networks, often with centralized functions.

Attempts to overcome these deficiencies may lead to excessively complicated hybrid solutions, distorting the initial principles (the QoS Pandora box). Middleware solutions are sometimes difficult to deploy (e.g for large scale PKI deployment). “Lambda on-demand” technologies are conceptually nothing else than old switched circuits, that we never managed to satisfactorily integrate with IP networks.

Where is all this going? To help forming a vision of the future, the paper will refer to several observations that the author has formulated over the past 30 years: the “breathing law” (a succession of decentralization and recentralization phases), the perpetual and oscillating mismatch of the bandwidth offer-demand, the conceptual antagonisms between resource level and complexity, between scaling and QoS.

**Distributed Computing Systems and Experiences / 147**

## HEP Applications Experience with the European DataGrid Middleware and Testbed

**Authors:** A. De Salvo<sup>1</sup>; A. Fanfani<sup>2</sup>; A. Tsaregorodtsev<sup>3</sup>; C. Grandi<sup>4</sup>; D. Boutigny<sup>5</sup>; D. Colling<sup>6</sup>; D. Groep<sup>7</sup>; E. Luppi<sup>8</sup>; F. Brochu<sup>9</sup>; F. Carminati<sup>10</sup>; F. Harris<sup>11</sup>; G. Negri<sup>12</sup>; I. Augustin<sup>10</sup>; I. Stokes-Rees<sup>13</sup>; J.J. Blaising<sup>5</sup>; J. Closier<sup>10</sup>; K. Bos<sup>7</sup>; L. Perini<sup>12</sup>; M. Reale<sup>1</sup>; O. Maroney<sup>14</sup>; O. Smirnova<sup>15</sup>; P. Capiluppi<sup>2</sup>; P. Cerello<sup>16</sup>; R. Barbera<sup>17</sup>; S. Bagnasco<sup>16</sup>; S. Burke<sup>18</sup>; S. Resconi<sup>12</sup>; V. GARONNE<sup>19</sup>; W. van Leeuwen<sup>7</sup>

<sup>1</sup> INFN Rome 1

<sup>2</sup> INFN Bologna

<sup>3</sup> Marseille

<sup>4</sup> INFN-Bologna

<sup>5</sup> LAPP Annecy

<sup>6</sup> Imperial

<sup>7</sup> NIKHEF

<sup>8</sup> INFN Sezione di Ferrara

<sup>9</sup> *Cambridge*

<sup>10</sup> *CERN*

<sup>11</sup> *CERN/Oxford*

<sup>12</sup> *INFN Milan*

<sup>13</sup> *Oxford*

<sup>14</sup> *Bristol*

<sup>15</sup> *Lund*

<sup>16</sup> *INFN Turin*

<sup>17</sup> *INFN Catania*

<sup>18</sup> *Rutherford Appleton Laboratory*

<sup>19</sup> *CNRS-IN2P3-CPPM, MARSEILLE*

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

The European DataGrid (EDG) project ran from 2001 to 2004, with the aim of producing middleware which could form the basis of a production Grid, and of running a testbed to demonstrate the middleware. HEP experiments (initially the four LHC experiments and subsequently BaBar and D0) were involved from the start in specifying requirements, and subsequently in evaluating the performance of the middleware, both with generic tests and through increasingly complex data challenges. A lot of experience has therefore been gained which may be valuable to future Grid projects, in particular LCG and EGEE which are using a substantial amount of the middleware developed in EDG. We report our experiences with job submission, data management and mass storage, information and monitoring systems, Virtual Organisation management and Grid operations, and compare them with some typical Use Cases defined in the context of LCG. We also describe some of the main lessons learnt from the project, in particular in relation to configuration, fault-tolerance, interoperability and scalability, as well as the software development process itself, and point out some areas where further work is needed. We also make some comments on how these issues are being addressed in LCG and EGEE.

**Poster Session 2 - Board:** 59 / 148

## **GROSS: an end user tool for carrying out batch analysis of CMS data on the LCG-2 Grid.**

**Authors:** B. MacEvoy<sup>1</sup>; D. Colling<sup>1</sup>; H. Tallini<sup>1</sup>; W. Stuart<sup>1</sup>

<sup>1</sup> *IMPERIAL COLLEGE LONDON*

**Corresponding Authors:** h.tallini@imperial.ac.uk, stuart.wakefield@imperial.ac.uk

GROSS (GRidified Orca Submission System) has been developed to provide CMS end users with a single interface for running batch analysis tasks over the LCG-2 Grid. The main purpose of the tool is to carry out job splitting, preparation, submission, monitoring and archiving in a transparent way which is simple to use for the end user. Central to its design has been the requirement for allowing multi-user analyses, and to accomplish this all persistent information is stored on a backend MySQL database. This database is additionally shared with BOSS, to which GROSS interfaces in order to provide job submission and real time monitoring capability.

In this paper we present an overview of GROSS's architecture and functionality and report on first user tests of the system using CMS Data Challenge 2004 data (DC04).

Poster Session 1 - Board: 9 / 151

## Simplified deployment of an EDG/LCG cluster via LCFG-UML

**Authors:** A. Garcia<sup>1</sup>; G. Tsouloupas<sup>2</sup>; M. Hardt<sup>3</sup>

<sup>1</sup> KARLSRUHE RESEARCH CENTER (FZK)

<sup>2</sup> UNIVERSITY OF CYPRUS (UCY)

<sup>3</sup> FORSCHUNGSZENTRUM KARLSRUHE (FZK)

**Corresponding Authors:** garcia@iwr.fzk.de, hardt@kit.edu

The clusters using DataGrid middleware are usually installed and managed by means of an “LCFG” server. Originally developed by the Univ. of Edinburgh and extended by DataGrid, this is a complex piece of software. It allows for automated installation and configuration of a complete grid site. However, installation of the “LCFG”-Server takes most of the time, thus hinder widespread use.

Our approach was to set up and preconfigure the LCFG-server inside a “User Mode Linux” (UML) instance in order to make deployment faster. The result is the “UML-LCFG-Sserver”. It is provided as a prebuilt root-filesystem image which can be up and running within only with few configuration steps. Detailed instructions and experience are also provided on the basis of tests within the CrossGrid project. Altogether UML-LCFG makes it easier for a new site to join an EDG/LCG based Grid by bypassing most of the LCFG server installation.

Poster Session 1 - Board: 32 / 152

## InGRID - Installing GRID

**Authors:** F.M. Taurino<sup>1</sup>; G. Tortone<sup>2</sup>; R. Esposito<sup>2</sup>

<sup>1</sup> INFN - INFN

<sup>2</sup> INFN

**Corresponding Author:** taurino@na.infn.it

The “gridification” of a computing farm is usually a complex and time consuming task. Operating system installation, grid specific software, configuration files customization can turn into a large problem for site managers. This poster introduces InGRID, a solution used to install and maintain grid software on small/medium size computing farms. Grid elements installation with InGRID consists in three steps. In the first step nodes are installed using RedHat Kickstart, an installation method that automate most of a Linux distribution installation, including disk partitioning, boot loader configuration, network configuration, base package selection. Grid specific software is then integrated using apt4rpm, a package management wrapper over the rpm commands. Apt automatically manages packages dependencies, and is able to download, install and upgrade RPMs from a central software repository. Finally, grid configuration files are customized through LCFGng, a system to setup and maintain Unix machines, that can configure many system files, execute scripts, create users, etc.

Event Processing / 153

## An Object-Oriented Simulation Program for CMS

**Authors:** A. De Roeck<sup>1</sup>; A. FANFANI<sup>2</sup>; A. NIKITENDO<sup>3</sup>; A. ZHOKIN<sup>4</sup>; D. ELVIRA<sup>5</sup>; F. FERRO<sup>6</sup>; G. EULISSE<sup>7</sup>; H.P. WELLSCH<sup>8</sup>; I. Osborne<sup>9</sup>; J. WENG<sup>10</sup>; K. LASSILA-PERINI<sup>11</sup>; L. TUURA<sup>12</sup>; M. CASE<sup>13</sup>; M. Liendl<sup>1</sup>; M. Stavriakou<sup>14</sup>; M. ZANETTI<sup>15</sup>; N. AMAPANE<sup>16</sup>; P. ARCE<sup>17</sup>; P. ZYCH<sup>18</sup>; S. ABDOULLINE<sup>19</sup>; S. BANERJEE<sup>20</sup>; S. DUTTA<sup>21</sup>; S. MUZAFFAR<sup>7</sup>; S. WYNHOFF<sup>22</sup>; T. BOCCALI<sup>23</sup>; T. TODOROV<sup>24</sup>; T. WILDISH<sup>22</sup>; V. ANDREEV<sup>25</sup>

<sup>1</sup> CERN

<sup>2</sup> University of Bologna, Italy

<sup>3</sup> Imperial College, London, UK

<sup>4</sup> ITEP, Moscow, Russia

<sup>5</sup> FNAL, Chicago, USA

<sup>6</sup> INFN, Genova, Italy

<sup>7</sup> NorthEastern University, Boston, USA

<sup>8</sup> CERN, Geneva, Switzerland

<sup>9</sup> Northeastern University, Boston, USA

<sup>10</sup> Karlsruhe University, Germany

<sup>11</sup> HIP, Helsinki, Finland

<sup>12</sup> NorthEastern University, Boston, USA

<sup>13</sup> UC Davis, USA

<sup>14</sup> FNAL

<sup>15</sup> INFN, Italy

<sup>16</sup> University of Torino, Italy

<sup>17</sup> CIEMAT, Madrid, Spain

<sup>18</sup> Warsaw University, Poland

<sup>19</sup> FNAL, USA

<sup>20</sup> Tata Institute, Bombay, India

<sup>21</sup> University of Pisa, Italy

<sup>22</sup> Princeton University, USA

<sup>23</sup> SNS, Pisa, Italy

<sup>24</sup> IRES, Strasbourg, France

<sup>25</sup> UCLA, USA

**Corresponding Author:** maya.stavriakou@cern.ch

The CMS detector simulation package, OSCAR, is based on the Geant4 simulation toolkit and the CMS object-oriented framework for simulation and reconstruction.

Geant4 provides a rich set of physics processes describing in detail electro-magnetic and hadronic interactions. It also provides the tools for the implementation of the full CMS detector geometry and the interfaces required for recovering information from the particle tracking in the detectors.

This functionality is interfaced to the CMS framework, which, via its “action on demand” mechanisms, allows the user to selectively load desired modules and to configure and tune the final application.

The complete CMS detector is rather complex with more than 12 million readout channels and more than 1 million geometrical volumes.

OSCAR has been validated by comparing its results with test beam data and with results from simulation with a GEANT3-based program.

It has been successfully deployed in the 2004 data challenge for CMS, where ~20 million events for various LHC physics channels were simulated and analysed.

Authors:

S. Abdulline, V. Andreev, P. Arce, S. Arce, S. Banerjee, T. Boccali, M. Case, A. De Roeck, S. Dutta, G. Eulisse, D. Elvira, A. Fanfani, F. Ferro, M. Liendl, S. Muzaffar, A. Nikitenko, K. Lassila-Perini, I. Osborne,

M. Stavrianakou, T. Todorov, L. Tuura, H.P. Wellisch, T. Wildish, S. Wynhoff,  
M. Zanetti, A. Zhokin, P. Zych

**Poster Session 3** - Board: 53 / 155

## **Mantis: the Geant4-based simulation specialization of the CMS COBRA framework**

**Authors:** A. DE ROECK<sup>1</sup>; M. Liendl<sup>2</sup>; M. Stavrianakou<sup>3</sup>; P. ARCE<sup>4</sup>; S. BANERJEE<sup>5</sup>; T. BOCCALI<sup>6</sup>; T. TODOROV<sup>7</sup>; V. Innocente<sup>8</sup>

<sup>1</sup> *CERN, Geneva, Switzerland*

<sup>2</sup> *CERN*

<sup>3</sup> *FNAL*

<sup>4</sup> *CIEMAT, Madrid, Spain*

<sup>5</sup> *Tata Institute, Bombay, India*

<sup>6</sup> *SNS, Pisa, Italy*

<sup>7</sup> *IRES, Strasbourg, France*

<sup>8</sup> *CERN, PH/SFT*

**Corresponding Author:** maya.stavrianakou@cern.ch

The CMS Geant4-based Simulation Framework, Mantis, is a specialization of the COBRA framework, which implements the CMS OO architecture. Mantis, which is the basis for the CMS-specific simulation program OSCAR, provides the infrastructure for the selection, configuration and tuning of all essential simulation elements: geometry construction, sensitive detector and magnetic field management, event generation and Monte Carlo truth, physics, particle propagation and tracking, run and event management, and user monitoring actions.

The experimental setup is built by Mantis using the COBRA Detector Description Database, DDD, which allows transparent instantiation of any layout (full or partial CMS simulation, test beam setups etc).

Persistency, histogramming and other important services are available using the standard COBRA infrastructure and are transparent to user applications.

Authors:

M. Stavrianakou, P. Arce, S. Banerjee, T. Boccali, A. De Roeck, V. Innocente,  
M. Liendl, T. Todorov

**Online Computing** / 157

## **Jefferson Lab Data Acquisition Run Control System**

**Authors:** C. Timmer<sup>1</sup>; D. Abbott<sup>1</sup>; D. LAWRENCE<sup>2</sup>; E. Wolin<sup>1</sup>; Ed Jastrzembski<sup>1</sup>; G. Heyes<sup>1</sup>; V. Gyurjyan<sup>1</sup>

<sup>1</sup> *Jefferson Lab*

<sup>2</sup> *TJNAF, USA*

**Corresponding Author:** gurjyan@jlab.org

A general overview of the Jefferson Lab data acquisition run control system is presented. This run control system is designed to operate the configuration, control, and monitoring of all Jefferson Lab experiments. It controls data-taking activities by coordinating the operation of DAQ sub-systems, online software components and



third-party software such as external slow control systems.

The main, unique feature which sets this system apart from conventional systems is its incorporation of intelligent agent concepts. Intelligent agents are autonomous programs which interact with each other through certain protocols on a peer-to-peer level. In this case, the protocols and standards used come from the domain-independent Foundation for Intelligent Physical Agents (FIPA), and the implementation used is the Java Agent Development Framework (JADE).

A lightweight, RDF (Resource Definition Framework) based language was developed to standardize the description of the run control system for configuration purposes.

Fault tolerance and recovery issues are addressed.

Key features of the system include: subsystem state management, configuration management, agent communication, multiple simultaneous run management and synchronization, and user interfaces. A user interface allowing web-wide monitoring was developed which incorporates a JAS/AIDA data server extensible through Java servlets.

**Poster Session 1 - Board: 11 / 159**

## Cluster architectures used to provide CERN central CVS services

**Authors:** I. Reguero<sup>1</sup>; M. Guijarro<sup>1</sup>; S. Lopienski<sup>1</sup>

<sup>1</sup> CERN

**Corresponding Author:** sebastian.lopienski@cern.ch

There are two cluster architecture approaches used at CERN to provide central CVS services. The first one (<http://cern.ch/cvs>) depends on AFS for central storage of repositories and offers automatic load-balancing and fail-over mechanisms.

The second one (<http://cern.ch/lcgcvs>) is an N + 1 cluster based on local file systems, using data replication and not relying on AFS. It does not provide either dynamic load-balancing or automatic fail-over. Instead a series of tools were developed for repository relocation in case of fail-over and for manual load-balancing.

Both architectures are used in production at CERN and project managers can choose one or the other, depending on their needs. If, eventually, one architecture proves to be significantly better, the other one may be phased out. This paper presents in detail both approaches and describes their relative advantages and drawbacks, as well as some data about them (number of repositories, average repository size, etc).

**Core Software / 160**

## Optimizing Selection Performance on Scientific Data by utilizing Bitmap Indices

**Author:** H. Schmuecker<sup>1</sup>

<sup>1</sup> CERN

**Corresponding Author:** helmut.gerhard.schmucker@cern.ch

Bitmap indices have gained wide acceptance in data warehouse applications handling large amounts of read only data. High dimensional ad hoc queries can be efficiently performed by utilizing

bitmap indices, especially if the queries cover only a subset of the attributes stored in the database. Such access patterns are common use in HEP analysis. Bitmap indices have been implemented by several commercial database management systems. However, the provided query algorithms focus on typical business applications, which are based on discrete attributes with low cardinality. HEP data, which are mostly characterized by non discrete attributes, cannot be queried efficiently by these implementations.

Support for selections on continuously distributed data can be added to the bitmap index technique by extending it with an adaptive binning mechanism. Following this approach a prototype has been implemented, which provides the infrastructure to perform index based selections on HEP analysis data stored in ROOT trees/tuples. For the indices a range encoded design with multiple components has been chosen. This design concept allows to realize a very fine binning granularity, which is crucial to selection performance, with an index of reasonable size. Systematic performance tests have shown that the query processing time and the disk-I/O can be significantly reduced compared to a conventional scan of the data. This especially applies to optimization scenarios in HEP analysis, where selections are slightly varied and performed repetitively on one and same data sample.

**Poster Session 2 - Board: 5 / 161**

## The ATLAS Computing Model

**Author:** R. JONES<sup>1</sup>

<sup>1</sup> LANCAS

**Corresponding Author:** roger.jones@cern.ch

The ATLAS Computing Model is under continuous active development. Previous exercises focussed on the Tier-0/Tier-1 interactions, with an emphasis on the resource implications and only a high-level view of the data and workflow. The work presented here considerably revises the resource implications, and attempts to describe in some detail the data and control flow from the High Level Trigger farms all the way through to the physics user. The model draws from the experience of previous and running experiments, but will be tested in the ATLAS Data Challenge 2 (DC2, described in other abstracts) and in the ATLAS Combined Testbeam exercises.

An important part of the work is to devise the measurements and tests to be run during DC2. DC2 will be nearing completion in September 2004, and the first assessments of the performance of the computing model in scaled slice tests will be presented.

**Distributed Computing Services / 162**

## Software agents in data and workflow management

**Authors:** A. Afaq<sup>1</sup>; A. Fanfani<sup>2</sup>; C. Charlot<sup>3</sup>; C. Grandi<sup>2</sup>; D. Bonacorsi<sup>4</sup>; D. Newbold<sup>5</sup>; I. Fisk<sup>1</sup>; J. Hernandez<sup>6</sup>; J. Rehn<sup>7</sup>; K. Rabbertz<sup>7</sup>; L. Tuura<sup>8</sup>; M. Ernst<sup>1</sup>; O. Maroney<sup>5</sup>; P. GARCIA-ABIA<sup>9</sup>; S. Metson<sup>5</sup>; T. Barrass<sup>10</sup>; T. Wildish<sup>11</sup>; W. Jank<sup>12</sup>; Y. Wu<sup>1</sup>

<sup>1</sup> FNAL

<sup>2</sup> INFN-Bologna

<sup>3</sup> IN2P3, LLR

<sup>4</sup> CNAF-INFN

<sup>5</sup> University of Bristol

<sup>6</sup> CIEMAT Madrid

<sup>7</sup> University of Karlsruhe

<sup>8</sup> Northeastern University

<sup>9</sup> CIEMAT

<sup>10</sup> CMS, UNIVERSITY OF BRISTOL

<sup>11</sup> Princeton

<sup>12</sup> CERN

CMS currently uses a number of tools to transfer data which, taken together, form the basis of a heterogeneous datagrid. The range of tools used, and the directed, rather than optimised nature of CMS recent large scale data challenge required the creation of a simple infrastructure that allowed a range of tools to operate in a complementary way.

The system created comprises a hierarchy of simple processes (named agents) that propagate files through a number of transfer states. File locations and some application metadata were stored in POOL file catalogues, with LCG LRC or MySQL backends. Agents were assigned limited responsibilities, and were restricted to communicating state in a well-defined, indirect fashion through a central transfer management database. In this way, the task of distributing data was easily divided between different groups for implementation.

The prototype system was developed rapidly, and achieved the required sustained transfer rate of ~10 MBps, with  $O(10^6)$  files distributed to 6 sites from CERN. Experience with the system during the data challenge raised issues with underlying technology (MSS write/read, stability of the LRC, maintenance of file catalogues, synchronisation of filesystems) which have been successfully identified and handled. The development of this prototype infrastructure allows us to plan the evolution of backbone CMS data distribution from a simple hierarchy to a more autonomous, scalable model drawing on emerging agent and grid technology.

**Poster Session 2 - Board: 45 / 163**

## Using Tripwire to check cluster system integrity

**Authors:** E. Perez-Calle<sup>1</sup>; F.J. Rodriguez Calonge<sup>1</sup>; M. Cardenas Montes<sup>1</sup>

<sup>1</sup> CIEMAT

**Corresponding Author:** elio.perez@ciemat.es

Expansion of large computing fabrics/clusters throughout the world would create a need for stricter security. Otherwise any system could suffer damages such as data loss, data falsification or misuse.

Perimeter security and intrusion detection system (IDS) are the two main aspects that must be taken into account in order to achieve system security.

The main target of an intrusion detection system is early detection in the previously mentioned cases, as a way to minimize any damage in data contained in the system.

Tripwire is one of the most powerful IDSs and is widely used as a security tool by the community of network administrators. Tripwire is oriented to monitor the status of files and directories, being able to detect the lightest change suffered by them.

At Ciemat, Tripwire has been used to monitor our local clusters, involved in GRID projects such as implementation of LCG prototypes, to guarantee the integrability of data generated, and stored there. It is used as well to monitor any modification of operating system files and any other scientific core software.

**Grid Security / 164**

## Using Nagios for intrusion detection

**Authors:** E. PEREZ-CALLE<sup>1</sup>; F.J. Rodriguez Calonge<sup>1</sup>; M. Cardenas Montes<sup>1</sup>

<sup>1</sup> CIEMAT

**Corresponding Author:** miguel.cardenas@ciemat.es

Implementing strategies for secured access to widely accessible clusters is a basic requirement of these services, in particular if GRID integration is sought for. This issue has two complementary lines to be considered: security perimeter and intrusion detection systems. In this paper we address aspects of the second one.

Compared to classical intrusion detection mechanisms, close monitoring of computer services can substantially help to detect intrusion signs. Having alarms indicating the presence of an intrusion into the system, allows system administrators to take fast actions to minimize damages and stop diffusion towards other critical systems.

One possible monitoring tool is Nagios ([www.nagios.org](http://www.nagios.org)), a powerful GNU tool with capacity to observe and collect information about a variety of services, and trigger alerts.

In this paper we present the work done at CIEMAT, where we have applied these directives to our local cluster. We have implemented a system to monitor the hardware and system sensitive information. We describe the process and show through different simulated security threads how does our implementation respond to it.

**Poster Session 1 - Board: 15 / 166**

## Experience with CORBA communication middleware in the ATLAS DAQ

**Authors:** A. AMORIM<sup>1</sup>; A. KAZAROV<sup>2</sup>; D. BURCKHART-CHROMEK<sup>3</sup>; D. KLOSE<sup>1</sup>; D. Liko<sup>4</sup>; E. BADESCU<sup>5</sup>; I. ALEXANDROV<sup>6</sup>; I. SOLOVIEV<sup>7</sup>; J. FLAMMER<sup>3</sup>; L. MAPELLI<sup>3</sup>; L. PEDRO<sup>1</sup>; M. CAPRINI<sup>5</sup>; M. DOBSON<sup>3</sup>; M. Mineev<sup>8</sup>; N. FIUZA DE BARROS<sup>1</sup>; R. JONES<sup>3</sup>; S. KOROBV<sup>6</sup>; S. Kolos<sup>4</sup>; V. Kotov<sup>8</sup>; Y. RYABOV<sup>7</sup>

<sup>1</sup> Universidade de Lisboa, Portugal

<sup>2</sup> Petersburg Nuclear Physics Institute (PNPI)

<sup>3</sup> CERN, Geneva, Switzerland

<sup>4</sup> CERN

<sup>5</sup> *National Institute of Physics and Nuclear Engineering, Bucharest, Romania*

<sup>6</sup> *JINR, Dubna, Russia*

<sup>7</sup> *Petersburg Nuclear Physics Institute, Gatchina, Russia*

<sup>8</sup> *Joint Institute for Nuclear Research (JINR)*

**Corresponding Author:** serguei.kolos@cern.ch

As modern High Energy Physics (HEP) experiments require more distributed computing power to fulfill their demands, the need for an efficient distributed online services for control, configuration and monitoring in such experiments becomes increasingly important. This paper describes the experience of using standard Common Object Request Broker Architecture (CORBA) middleware for providing a high performance and scalable software, which will be used for the online control, configuration and monitoring in the ATLAS Data Acquisition (DAQ) system. It also presents the experience, which was gained from using several CORBA implementations and replacing one CORBA broker with another.

Finally the paper introduces results of the large scale tests, which have been done on the cluster of more than 300 nodes, demonstrating the performance and scalability of the ATLAS DAQ online services.

These results show that the CORBA standard is truly appropriate for the highly efficient online distributed computing in the area of modern HEP experiments.

**Poster Session 2 - Board: 7 / 167**

## **The deployment mechanisms for the ATLAS software.**

**Authors:** A. De Salvo<sup>1</sup>; C. ARNAULT<sup>2</sup>; G. Rybkine<sup>3</sup>; S. George<sup>4</sup>

<sup>1</sup> *INFN*

<sup>2</sup> *CNRS*

<sup>3</sup> *INR*

<sup>4</sup> *RHUL*

**Corresponding Author:** arnault@lal.in2p3.fr

One of the most important problems in software management of a very large and complex project such as Atlas is how to deploy the software on the running sites. By running sites we include computer sites ranging from computing centers in the usual sense down to individual laptops but also the computer elements of a computing grid organization. The deployment activity consists in constructing a well defined representation of the states of the working software (known as releases), and transporting them to the target sites, in such a way that the installation process can be entirely automated and can take care of discovering the context and adapting itself to it. A set of tools based on both CMT - the basic configuration management tool of ATLAS - and Pacman has been developed. The resulting mechanism now supports the systematic production of distribution kits for various binary conditions of every release, the partial or complete automatic installation of kits on any site and the running of test suites to validate the installed kits. This mechanism is meant to be fully compliant with the Grid requirements and has been tested in the context of LCG. Several issues related with the constraints on the target system, or with the incremental updates of the installation still need to be studied and will be discussed.

**Event Processing / 168****The Atlantis event visualisation program for the ATLAS experiment**

**Authors:** C. Timmermans<sup>1</sup>; D. Petrusca<sup>2</sup>; E. JANSEN<sup>1</sup>; F. Crijns<sup>1</sup>; G. Taylor<sup>3</sup>; H. Drevermann<sup>4</sup>; J. Couchman<sup>5</sup>; J. Drohan<sup>5</sup>; N. Konstantinidis<sup>6</sup>; P. Klok<sup>1</sup>; Z. MAXA<sup>5</sup>

<sup>1</sup> *University of Nijmegen*

<sup>2</sup> *Seigen/CERN*

<sup>3</sup> *UC Santa Cruz*

<sup>4</sup> *CERN*

<sup>5</sup> *University College London*

<sup>6</sup> *UNIVERSITY COLLEGE LONDON*

**Corresponding Author:** nikolaos.konstantinidis@cern.ch

We describe the philosophy and design of Atlantis, an event visualisation program for the ATLAS experiment at CERN. Written in Java, it employs the Swing API to provide an easily configurable Graphical User Interface.

Atlantis implements a collection of intuitive, data-orientated 2D projections, which enable the user to quickly understand and visually investigate complete ATLAS events. Event data is read in from XML files produced by a dedicated algorithm running in the ATLAS software framework ATHENA, and translated into internal data objects. Within the same main canvas area, multiple views of the data can be displayed with varying size and position. Interactions such as zoom, selection and query can occur between these views using Drag and Drop.

Associations between data objects as well as the values of their member variables provide criteria upon which the Atlantis user may filter a full Atlas event. By choosing whether or not to show certain data and, if so, in what colour, a more personalised and useful display may be obtained. The user can dynamically create and manage their own associations and perform context dependent operations upon them.

**Core Software / 169****Guidelines for Developing a Good GUI**

**Authors:** F. Rademakers<sup>1</sup>; I. Antcheva<sup>1</sup>; R. Brun<sup>1</sup>

<sup>1</sup> *CERN*

**Corresponding Author:** ilka.antcheva@cern.ch

Designing a usable, visually-attractive GUI is somewhat more difficult than it appears at a first glance. The users, the GUI designers and the programmers are three important parts involved in this process and everyone has a comprehensive view on the aspects of the application goals, as well as the steps that have to be taken to meet successfully the application requirements. The fundamental GUI design principles and the main programming aspects are discussed in this paper.

Key topics include:

- User requirements: identifying users and support different user profiles - from beginners to advanced users
- Close relationship between the GUI widgets, user actions, tasks and user goals

- Task-analysis methods
- Prototypes development and testing prototypes
- General design considerations
- Effective GUI design keys, guidelines and style guides

**Poster Session 3 - Board: 3 / 170**

## ROOT Graphical User Interface

**Authors:** B. Bellenot<sup>1</sup>; F. Rademakers<sup>2</sup>; I. Antcheva<sup>2</sup>; O. Couet<sup>2</sup>; R. Brun<sup>2</sup>; V. Onuchin<sup>2</sup>

<sup>1</sup> *Alcan*

<sup>2</sup> *CERN*

**Corresponding Author:** ilka.antcheva@cern.ch

The GUI is a very important component of the ROOT framework. Its main purpose is to improve the usability and end-user perception. In this paper, we present two main projects in this direction: the ROOT graphics editor and the ROOT GUI builder.

The ROOT graphics editor is a recent addition to the framework. It provides a state of the art and an intuitive way to create or edit objects in the canvas.

The ROOT GUI builder greatly facilitates the design, the development and the maintenance of any interactive application based on the ROOT framework. GUI objects can be selected, dragged/dropped in the widgets. An automatic code generator can be activated to save the code corresponding to any complex layout. This code can be executed via the CINT interpreter or directly compiled with the user application.

Past surveys indicate that the development of a GUI is a significant undertaking and that the GUI's source code is a substantial portion of the program's overall source base. The new GUI builder in ROOT will enable the rapid construction of simple and complex GUIs.

**Plenary / 171**

## Grid3: An Application Grid Laboratory for Science

**Author:** R. Gardner<sup>1</sup>

<sup>1</sup> *University of Chicago*

**Corresponding Author:** robert.w.gardner@cern.ch

The U.S. Trillium Grid projects in collaboration with High Energy Experiment groups from the Large Hadron Collider (LHC), ATLAS and CMS, Fermi-Lab's BTeV, members of the LIGO, SDSS collaborations and groups from other scientific disciplines and computational centers have deployed a multi-VO, application-driven grid laboratory ("Grid3"). The grid laboratory has sustained for several months the production-level services required by the participating experiments. The deployed infrastructure has been operating since November 2003 with 27 sites, a peak of 2800 processors, work loads from 10 different applications exceeding 1300 simultaneous jobs, and data transfers among sites of greater than 2 TB/day.

The Grid3 infrastructure was deployed from grid level services provided by groups and applications within the collaboration. The services were organized into four distinct “grid level services” including: Grid3 Packaging, Monitoring and Information systems, User Authentication and the iGOC Grid Operations Center. In this paper we describe the Grid3 operational model, deployment strategies, and site installation and configuration procedures. We describe the grid middleware components used, how the components were packaged and deployed on sites each under its own local administrative domain, and how the pieces fit together to form the Grid3 grid infrastructure.

**Event Processing / 172**

## How to build an event store - the new Kanga Event Store for BaBar

**Author:** M. Steinke<sup>1</sup>

<sup>1</sup> *Ruhr Universitaet Bochum*

In the past year, BaBar has shifted from using Objectivity to using ROOT I/O as the basis for our primary event store. This shift required a total reworking of Kanga, our ROOT-based data storage format. We took advantage of this opportunity to ease the use of the data by supporting multiple access modes that make use of many of the analysis tools available in ROOT.

Specifically, our new event store supports: 1) the pre-existing separated transient + persistent model, 2) a transient based load-on-demand model currently being developed, 3) direct access to persistent data classes in compiled code, 4) fully interactive access to persistent data classes from either the ROOT prompt and via interpreted macros.

We will describe key features of Kanga including: 1) the separation and management of transient and persistent representations of data, 2) the implementation of read on demand references in ROOT, 3) the modular and extensible persistent event design, 4) the implementation of schema evolution and 5) BaBar specific extensions to core ROOT classes that we used to preserve the end-user “feel” of ROOT.

**Event Processing / 173**

## Fast tracking for the ATLAS LVL2 Trigger

**Authors:** D. Emeliyanov<sup>1</sup>; F. PARODI<sup>2</sup>; H. Drevermann<sup>3</sup>; J. Baines<sup>1</sup>; M. Sutton<sup>4</sup>; N. Konstantinidis<sup>5</sup>

<sup>1</sup> *RAL*

<sup>2</sup> *INFN Genova, Italy*

<sup>3</sup> *CERN*

<sup>4</sup> *UCL*

<sup>5</sup> *UNIVERSITY COLLEGE LONDON*

**Corresponding Author:** nikolaos.konstantinidis@cern.ch

We present a set of algorithms for fast pattern recognition and track reconstruction using 3D space points aimed for the High Level



Triggers (HLT) of multi-collision hadron collider environments. At the LHC there are several interactions per bunch crossing separated along the beam direction,  $z$ . The strategy we follow is to (a) identify the  $z$ -position of the interesting interaction prior to any track reconstruction; (b) select groups of space points pointing back to this  $z$ -position, using a histogramming technique which avoids performing any combinatorics; and (c) proceed to the combinatorial tracking only within the individual groups of space points. The validity of this strategy will be demonstrated with results in terms of timing and physics performance for the LVL2 trigger of ATLAS at the LHC, although the strategy is generic and can be applied to any multi-collision hadron collider experiment.

In addition, the algorithms are conceptually simple, flexible and robust and hence appropriate for use in demanding, online environments. We will also make qualitative comparisons with an alternative, complimentary strategy, based on the use of look-up tables for handling combinatorics, that has been developed for the ATLAS LVL2 trigger. These algorithms have been used for the results that appear in the ATLAS HLT, DAQ and Controls Technical Design Report, which was recently approved by the LHC Committee.

**Poster Session 3 - Board: 5 / 174**

## **Recent evolutions of CMT. Multi-project and activity management.**

**Author:** C. ARNAULT<sup>1</sup>

<sup>1</sup> CNRS

**Corresponding Author:** [arnault@lal.in2p3.fr](mailto:arnault@lal.in2p3.fr)

Since its introduction in 1999, CMT is now used as a production tool in many large software projects for physics research (ATLAS, LHCb, Virgo, Auger, Planck). Although its basic concepts remain unchanged since the beginning, proving their viability, it is still improving and increasing its coverage of the configuration management mechanisms. Two important evolutions have recently been introduced, one for explicitly supporting multi-project environments, and the other to specify and manage configuration activities.

The existing concept of package area is now extended to cover the support of sub-projects structuring, with the possibility of assigning configuration management properties (typically strategies) to each sub project, allowing for instance to have installation area mechanisms only applicable for some of them.

It is also possible to specify parameterized activities that will be run on demand either through make or through an explicit activation command, which ensures that the runtime environment is properly setup.

**Core Software / 175**

## **The SEAL Component Model**

**Authors:** L. Tuura<sup>1</sup>; P. Mato<sup>2</sup>; R. Chytracek<sup>2</sup>

<sup>1</sup> *Northeastern University*

<sup>2</sup> *CERN*

**Corresponding Authors:** radovan.chytracak@cern.ch, pere.mato@cern.ch

This paper describes the component model that has been developed in the context of the LCG/SEAL project. This component model is an attempt to handle the increasing complexity in the current data processing applications of LHC experiments. In addition, it should facilitate the software re-use by the integration of software components from LCG and non-LCG into the experiment's applications. The component model provides the basic mechanisms and base classes that facilitate the decomposition of the whole C++ object-oriented application into a number of run-time pluggable software modules with well defined generic behavior, inter-component interaction protocols, run-time configuration and user customization.

This new development is based on the ideas and practical experiences of the various software frameworks in use by the different LHC experiments for several years. The design and implementation choices will be described and the practical experiences and difficulties in adopting this model to existing experiment software systems will be outlined.

**Wide Area Networking / 177**

## **Breaking the 1 GByte/sec Barrier? High speed WAN data transfers for science**

**Authors:** D. NAE<sup>1</sup>; H. Newman<sup>1</sup>; J. Bunn<sup>1</sup>; S. Ravot<sup>1</sup>; X. SU<sup>1</sup>; Y. Xia<sup>1</sup>

<sup>1</sup> *Caltech*

**Corresponding Authors:** sylvain.ravot@cern.ch, julian.bunn@caltech.edu

In this paper we describe the current state of the art in equipment, software and methods for transferring large scientific datasets at high speed around the globe. We first present a short introductory history of the use of networking in HEP, some details on the evolution, current status and plans for the Caltech/CERN/DataTAG transAtlantic link, and a description of the topology and capabilities of the research networks between CERN and HEP institutes in the USA. We follow this with some detailed material on the hardware and software environments we have used in collaboration with international partners (including CERN and DataTAG) to break several Internet2 land speed records over the last couple of years. Finally we describe our recent developments in collaboration with Microsoft, Newisys, AMD, Cisco and other industrial partners, in which we are attempting to transfer HEP data files from disk servers at CERN via a 10Gbit network path to disk servers at Caltech's Center for Advanced Computing Research (a total distance of over 11,000 kilometres), at a rate exceeding 1 GByte per second. We describe some solutions being used to overcome networking and hardware performance issues. Whilst such transfers represent the bleeding edge of what is possible today, they are expected to be commonplace at the start of LHC operations in 2007.

**Poster Session 1 - Board: 33 / 180**

## **A database prototype for managing computer systems configurations**

**Authors:** N. Sinanis<sup>1</sup>; Z. Toteva<sup>2</sup>

<sup>1</sup> *CERN/CMS*

<sup>2</sup> Sofia University/CERN/CMS

**Corresponding Authors:** zhechka.toteva@cern.ch, nick.sinanis@cern.ch

We describe a database solution in a web application to centrally manage the configuration information of computer systems. It extends the modular cluster management tool Quattor with a user friendly web interface.

System configurations managed by Quattor are described with the aid of PAN, a declarative language with a command line and a compiler interface. Using a relational schema, we are able to build a database for efficient data storage and configuration data processing. The relational schema ensures the consistency of the described model while the standard database interface ensures the fast retrieval of configuration information and statistic data.

The web interface simplifies the typical administration and routine operations tasks, e.g. definition of new types, configuration comparisons and updates etc. We present a prototype built on the above ideas and used to manage a cluster of developer workstations and specialised services in CMS.

## Distributed Computing Systems and Experiences / 182

### Grid Enabled Analysis : Architecture, prototype and status

**Authors:** A. Anjum<sup>1</sup>; C. Steenberg<sup>2</sup>; D. Bourilkov<sup>3</sup>; F. van Lingen<sup>4</sup>; H. Newman<sup>5</sup>; I. LeGrand<sup>2</sup>; J. Bunn<sup>2</sup>; J. Uk In<sup>3</sup>; L. Chitnis<sup>3</sup>; M. Kulkarni<sup>6</sup>; M. Thomas<sup>2</sup>; P. Avery<sup>3</sup>; R. Cavanaugh<sup>3</sup>; T. Azim<sup>1</sup>

<sup>1</sup> NUST, Pakistan

<sup>2</sup> California Institute of Technology

<sup>3</sup> University of Florida

<sup>4</sup> CALIFORNIA INSTITUTE OF TECHNOLOGY

<sup>5</sup> Caltech

<sup>6</sup> UNIVERSITY OF FLORIDA

**Corresponding Author:** fvlingen@caltech.edu

In this paper we report on the implementation of an early prototype of distributed high-level services supporting grid-enabled data analysis within the LHC physics community as part of the ARDA project within the context of the GAE (Grid Analysis Environment) and begin to investigate the associated complex behaviour of such an end-to-end system. In particular, the prototype integrates a typical physics user interface client (ROOT), a uniform web-services interface to grid services (Clarens), a virtual data service (Chimera), a request scheduling service (Sphinx), a monitoring service (MonALISA), a workflow execution service (Virtual Data Toolkit Client), a remote data file service (Clarens), a grid resource service (Virtual Data Toolkit Server), a replica location service/meta data catalog (RLS/POOL), an analysis session management system (CAVES) and a fine grain monitor system for job submission (BOSS).

For testing and evaluation purposes, the prototype is deployed across a modest sized U.S. regional CMS Grid Test-bed (consisting of sites in California, Florida, Fermilab) and is in the early stages of exhibiting interactive remote data access demonstrating interactive workflow generation and collaborative data analysis using virtual data and data provenance, as well as showing non-trivial examples of policy based scheduling of requests in a resource

constrained grid environment. In addition, the prototype is used to characterize the system performance as a whole, including the determination of request-response latencies in a distributed service model and the classification of high-level failure modes in a complex system.

**Poster Session 3 - Board: 28 / 183**

## Porting CLEO software to Linux

**Authors:** C. Jones<sup>1</sup>; V. Kuznetsov<sup>1</sup>

<sup>1</sup> CORNELL UNIVERSITY

**Corresponding Author:** valentin.kuznetsov@cern.ch

Linux operating system has become the platform of choice in the HEP community. However, the migration process from another operating system to Linux can be a tremendous effort for developers and system administrators. The ultimate goal of such a transition is to maximize agreement between the final results of identical calculations on the different platforms. Apart from the fine tuning of the existing software the following issues need to be resolved: choice of Linux distribution, development tools (compiler, debugger, profilers etc.), compatibility with 3d party software, and deployment strategy. It would be ideal to develop, run and test software using office desktops, local farm systems, or personal laptop regardless of the Linux distribution chosen. To accomplish this task you need to have flexible package management system which is capable to install/upgrade/verify/uninstall necessary software components without particular knowledge of remote system configuration and user privileges. We discuss how Linux became the third official computing platform of the CLEO collaboration, outlining the details of the transition from OSF and Solaris operating systems to Linux, software model and deployment strategy employed.

**Distributed Computing Services / 184**

## The Clarens Grid-enabled Web Services Framework: Services and Implementation

**Authors:** A. Anjum<sup>1</sup>; C. Steenberg<sup>2</sup>; F. van Lingen<sup>3</sup>; H. Newman<sup>4</sup>; I. Legrand<sup>5</sup>; J. Bunn<sup>3</sup>; M. Thomas<sup>2</sup>; T. Azim<sup>1</sup>

<sup>1</sup> National University of Science and Technology, Rawalpindi

<sup>2</sup> California Institute of Technology

<sup>3</sup> CALIFORNIA INSTITUTE OF TECHNOLOGY

<sup>4</sup> Caltech

<sup>5</sup> CALTECH

**Corresponding Authors:** conrad@hep.caltech.edu, fvlingen@caltech.edu

Clarens enables distributed, secure and high-performance access to the worldwide data storage, compute, and information Grids being constructed in anticipation of the needs of the Large Hadron Collider at CERN. We report on the rapid progress in the development of a second server implementation in the Java language, the evolution of a peer-to-peer network of Clarens servers, and general improvements in client and server implementations.

Services that are implemented at this time include read/write file access, service lookup and discovery, configuration management, job execution, Virtual Organization Management, an LHCb Information Service, as well as web service interfaces to POOL replica location and metadata catalogs, MonaLISA monitoring information, CMS MCRunjob workflow management, BOSS job monitoring and bookkeeping, Sphinx job scheduler and Chimera virtual data systems.

Commodity web service protocols allows a wide variety of computing platforms and applications to be used to securely access Clarend services, including a standard web browser, Java applets and stand-alone applications, the ROOT data analysis package, as well as libraries that provide programmatic access from the Python, C/C++ and Java languages.

**Poster Session 2 - Board: 69 / 185**

## **Experience with Deployment and Operation of the ATLAS Production System and the Grid3+ Infrastructure at Brookhaven National Lab**

**Authors:** A. Chan<sup>1</sup>; D. Yu<sup>1</sup>; G. Carcassi<sup>1</sup>; J. Smith<sup>1</sup>; R. Baker<sup>1</sup>; T. Wlodek<sup>1</sup>; W. Deng<sup>1</sup>; X. Zhao<sup>1</sup>; Z. Liu<sup>1</sup>

<sup>1</sup> *Brookhaven National Laboratory*

**Corresponding Author:** xzhao@bnl.gov

This paper describes the deployment and configuration of the production system for ATLAS Data Challenge 2 starting in May 2004, at Brookhaven National Laboratory, which is the Tier1 center in the United States for the International ATLAS experiment. We will discuss the installation of Windmill (supervisor) and Capone (executor) software packages on the submission host and the relevant security issues. The Grid3+ infrastructure and information service are used for the deployment of grid enabled ATLAS transformations on the Grid3+ computing elements. The Tier 1 hardware configuration includes 95 dual processor Linux compute nodes, 24 TB of NFS disk and an HPSS mass storage system. VOMS server maintains both VO services for US ATLAS and BNL local site policies. This paper describes the work of optimizing the performance and efficiency of this configuration.

**Plenary / 186**

## **Operating the LCG and EGEE Production Grids for HEP**

**Author:** I. Bird<sup>1</sup>

<sup>1</sup> *CERN*

**Corresponding Author:** ian.bird@cern.ch

In September 2003 the first LCG-1 service was put into production at most of the large Tier 1 sites and was quickly expanded up to 30 Tier 1 and Tier 2 sites by the end of the year. Several software upgrades were made and the LCG-2 service was put into production in time for the experiment data challenges that began in February 2004 and continued for several months. In particular LCG-2 introduced transparent access to mass storage and managed disk-only storage elements, and a first release

of the Grid File Access library. Much valuable experience was gained during the data challenges in all aspects from the functionality and use of the middleware, to the deployment, maintenance, and operation of the services at many sites. Based on this experience a program of work to address the functional and operational issues is being implemented. The goal is to focus on essential areas such as data management and to build by the end of 2004 a basic grid system capable of handling the basic needs of LHC computing, providing direction for future middleware and service development.

The LCG-2 infrastructure also forms the production service of EGEE. This involves supporting new application communities, bringing in new sites not associated with HEP and evolving a full scale 24x7 user and operational support structure. We will describe the EGEE infrastructure, how it supports and interacts with LCG, and how we expect the infrastructure to evolve over the next year of the EGEE project.

**Poster Session 1 - Board: 16 / 187**

## **Distributed Filesystem Evaluation and Deployment at the US-CMS Tier-1 Center**

**Authors:** D. Fagan<sup>1</sup>; J. kaiser<sup>1</sup>; L. Lisa Giacchetti<sup>1</sup>; M. Ernst<sup>2</sup>; R. Pasetes<sup>1</sup>

<sup>1</sup> *FERMILAB*

<sup>2</sup> *DESY*

**Corresponding Author:** michael.ernst@desy.de

The scalable serving of shared filesystems across large clusters of computing resources continues to be a difficult problem in high energy physics computing. The US CMS group at Fermilab has performed a detailed evaluation of hardware and software solutions to allow filesystem access to data from computing systems.

The goal of the evaluation was to arrive at a solution that was able to meet the growing needs of the US-CMS Tier-1 facility. The system needed to be scalable and be able to grow with the increasing size of the facility, load balanced and with high performance for data access, reliable and redundant with protection against failures, and manageable and supportable given a reasonable level of effort.

Over the course of a one year evaluation the group developed a suite of tools to analysis performance and reliability under load conditions, and then applied these tools to evaluations systems at Fermilab. In this presentation we will describe the suite of tools developed, the results of the evaluation process, the system and architecture that were eventually chosen, and the experience so far supporting a user community.

**Grid Security / 189**

## **Secure Grid Data Management Technologies in ATLAS**

**Authors:** A. Vaniachine<sup>1</sup>; D. Malon<sup>1</sup>; M. Branco<sup>2</sup>

<sup>1</sup> ANL<sup>2</sup> CERN**Corresponding Author:** miguel.branco@cern.ch

In a resource-sharing environment on the grid both grid users and grid production managers call for security and data protection from unauthorized access. To secure data management several novel grid technologies were introduced in ATLAS data management. Our presentation will review new grid technologies introduced in HEP production environment for database access through the Grid Security Infrastructure (GSI): secure GSI channel mechanisms for database services delivery for reconstruction on grid clusters behind closed firewalls; grid certificate authorization technologies for production database access control and scalable locking technologies for the chaotic 'on-demand' production mode. We address the separation of file transfer process from the file catalog interaction process (file location registration, file metadata querying, etc.), database transactions capturing data integrity and the high availability fault-tolerant database solutions for the core data management tasks. We discuss the complementarities of the security model for the online and the offline computing environments; best practices (and realities) of the database users' roles: administrators, developers, data writers, data replicators and data readers, need for elimination of the clear-text passwords; stateless and stateful protocols for the binary data transfers over secure grid data transport channels in heterogeneous grids. We present the security policies and technologies integrated in the ATLAS Production Data Management System - Don Quijote (GSI-enabled services oriented architecture, GSI proxy certificate delegation) and approaches for seamless integration of Don Quijote with POOL event collections and tag databases - while making the system non-intrusive to end-users.

**Distributed Computing Services / 190**

## Managed Data Storage and Data Access Services for Data Grids

**Authors:** D. Petravick<sup>1</sup>; I. Fisk<sup>1</sup>; J. Bakken<sup>1</sup>; M. Ernst<sup>2</sup>; P. Fuhrmann<sup>2</sup>; T. Mkrtchyan<sup>2</sup>; T. Perelmutov<sup>1</sup><sup>1</sup> FERMILAB<sup>2</sup> DESY**Corresponding Authors:** michael.ernst@desy.de, ian.fisk@cern.ch

The LHC needs to achieve reliable high performance access to vastly distributed storage resources across the network. USCMS has worked with Fermilab-CD and DESY-IT on a storage service that was deployed at several sites. It provides Grid access to heterogeneous mass storage systems and synchronization between them. It increases resiliency by insulating clients from storage and network failures, and facilitates file sharing and network traffic shaping.

This new storage service is implemented as a Grid Storage Element (SE). It consists of dCache as the core storage system and an implementation of the Storage Resource Manager (SRM), that together allow both local and Grid based access to the mass storage facilities. It provides advanced functionalities for managing, accessing and distributing collaboration data.

USCMS is using this system both as Disk Resource Manager at Tier-1 and Tier-2 sites, and as Hierarchical Resource Manager with Enstore as tape back-end at the Fermilab Tier-1. It is used for providing shared managed disk pools at sites and for streaming data between the CERN Tier-0, the Fermilab Tier-1 and U.S. Tier-2 centers

Applications can reserve space for a time period, ensuring space availability when the application runs. Worker nodes without WAN connection can trigger data replication to the SE and then access data via the LAN. Moving the SE functionality off the worker nodes reduces load and improves reliability of the compute farm elements significantly.

We describe architecture, components, and experience gained in CMS production and the DC04 Data Challenge.

**Grid Security / 191**

## The Virtual Organization Membership Service eXtension project (VOX)

**Authors:** A. Heavey<sup>1</sup>; D. Skow<sup>2</sup>; E. Berman<sup>1</sup>; G. Carcassi<sup>3</sup>; G. Graham<sup>2</sup>; I. Fisk<sup>2</sup>; J. Kaiser<sup>2</sup>; J. Weigand<sup>1</sup>; L. Bauerdick<sup>2</sup>; L. Grundhoefer<sup>4</sup>; R. Baker<sup>3</sup>; R. Gardner<sup>5</sup>; R. Pordes<sup>2</sup>; T. Levshina<sup>1</sup>; V. Sekhri<sup>1</sup>; Y. Wu<sup>1</sup>

<sup>1</sup> *FERMILAB*

<sup>2</sup> *Fermilab*

<sup>3</sup> *BNL*

<sup>4</sup> *University of Indiana*

<sup>5</sup> *University of Chicago*

**Corresponding Author:** [ian.fisk@cern.ch](mailto:ian.fisk@cern.ch)

Current grid development projects are being designed such that they require end users to be authenticated under the auspices of a “recognized” organization, called a Virtual Organization (VO). A VO must establish resource-usage agreements with grid resource providers. The VO is responsible for authorizing its members for grid computing privileges. The individual sites and resources typically enforce additional layers of authorization.

The VOX project developed at Fermilab is an extension of VOMS, developed jointly for DataTAG by INFN and for DataGrid by CERN. The Virtual Organization Membership Registration Service (VOMRS) is a major component of the VOX project. VOMRS is a service that provides the means for registering members of a VO, and coordination of this process among the various VO and grid administrators. It consists of a database to maintain user registration and institutional information, a server to handle members’ notification and synchronization with various interfaces, web services and a web user interface for the input of data into the database and manipulation of that data.

The VOX project also includes a component for the Site AuthoriZation (SAZ), which allows security authorities at a site to control access to site resources and a component for the Local Resource Administration (LRAS), which associates the VO member with the local account and local resources on a grid cluster.



The current state of deployment and future steps to improve the prototype and implement some new features will be presented.

## Distributed Computing Systems and Experiences / 192

### The Open Science Grid (OSG)

**Author:** R. Pordes<sup>1</sup>

<sup>1</sup> *FERMILAB*

**Corresponding Author:** ruth@fnal.gov

The U.S.LHC Tier-1 and Tier-2 laboratories and universities are developing production Grids to support LHC applications running across a worldwide Grid computing system. Together with partners in computer science, physics grid projects and running experiments, we will build a common national production grid infrastructure which is open in its architecture, implementation and use.

The OSG model builds upon the successful approach of last year's joint Grid2003 project. The Grid3 shared infrastructure has for over eight months given significant computational resources and throughput to more than six applications, including ATLAS and CMS data challenges, SDSS, LIGO and Biology analyses and computer science demonstrators.

To move towards LHC-scale data management, access and analysis capabilities, we will need to increase the scale, services, and sustainability of the current infrastructure by an order of magnitude. This requires a significant upgrade in its functionalities and technologies.

The OSG roadmap is a strategy and work plan to build the U.S.LHC computing enterprise as a fully usable, sustainable and robust grid, which is part of the LHC global computing infrastructure and open to partners.

The approach is to federate with other application communities in the U.S. to build a shared infrastructure open to other sciences and capable of being modified and improved to respond to needs of other applications, including CDF, D0, BaBar and RHIC experiments.

We describe the application driven engineered services of the OSG, short term plans and status, and the roadmap for a consortium, its partnerships and national focus.

## Grid Security / 194

### G-PBox: a Policy Framework for Grid Environments

**Authors:** A. Caltroni<sup>1</sup>; A. Ferraro<sup>2</sup>; A. Ghiselli<sup>3</sup>; G. Rubini<sup>2</sup>; R. Zappi<sup>2</sup>; V. Ciaschini<sup>2</sup>

<sup>1</sup> *INFN PD*

<sup>2</sup> *INFN CNAF*

<sup>3</sup> *INFN-CNAF*

**Corresponding Author:** vincenzo.ciaschini@cnaf.infn.it

A key feature of Grid systems is the sharing of its resources among multiple Virtual Organizations (VOs). The sharing process needs a policy framework to manage the resource access and usage. Generally Policy frameworks exist for farms or local systems only, but now, for Grid environments, a general, and distributed policy system is necessary.

Generally VOs and local systems have contracts that regulate the resource usage, hence complex relationships among these entities implying different kind of policies may exist: VOs oriented, local systems oriented, and a mix of these ones. We propose an approach to the representation, and management of such policies: the Grid Policy Box (G-PBox) framework. The approach is based on a set of databases belonging hierarchically-organised levels distributed onto the Grid and VOs structures. Each level contains only policies regarding itself. These levels have to communicate among themselves to accommodate for mixed policies, originating the need for a secure communication service framework, - for privacy reasons, - with the ability to sort and dispatch various kind of policies to the involved parties.

In this paper we present our first implementation of the G-PBox, and its architecture details, and we discuss the plans for G-PBox-related application and research.

**Poster Session 1 - Board: 7 / 195**

## The CMS User Analysis Farm at Fermilab

**Authors:** H. Wenzel<sup>1</sup>; I. Mandrichenko<sup>1</sup>; J. Kaiser<sup>1</sup>; L. Giacchetti<sup>1</sup>; N. Ratnikova<sup>1</sup>

<sup>1</sup> FERMILAB

US-CMS is building up expertise at regional centers in preparation for analysis of LHC data. The User Analysis Farm (UAF) is part of the Tier 1 facility at Fermilab. The UAF is being developed to support the efforts of the Fermilab LHC Physics Center (LPC) and to enable efficient analysis of CMS data in the US.

The support, infrastructure, and services to enable a local analysis community at a computing center which is remote from the physical detector and the majority of the collaboration present unique challenges.

The current UAF is a farm running the LINUX operating system providing interactive and batch computing for users. Load balancing, resource and process management are realized with FBSNG, the batch system developed at Fermilab. Over the course of the next three years the UAF must grow in size and functionality, while continuing to support simulated analysis activities and test beam applications.

In this presentation we will describe the development of the current cluster, the technology choices made, the services required to support regional analysis activities, and plans for the future.

**Distributed Computing Services / 196**

## Lattice QCD Data and Metadata Archives at Fermilab and the International Lattice Data Grid

**Authors:** E. Neilsen<sup>1</sup>; J. Simone<sup>1</sup>

<sup>1</sup> *FERMI NATIONAL ACCELERATOR LABORATORY*

**Corresponding Author:** neilsen@fnal.gov

The lattice gauge theory community produces large volumes of data. Because the data produced by completed computations form the basis for future work, the maintenance of archives of existing data and metadata describing the provenance, generation parameters, and derived characteristics of that data is essential not only as a reference, but also as a basis for future work. Development of these archives according to uniform standards both in the data and metadata formats provided and in the software interfaces to the component services could greatly simplify collaborations between institutions and enable the dissemination of meaningful results.

This paper describes the progress made in the development of a set of such archives at the Fermilab lattice QCD facility. We are coordinating the development of the interfaces to these facilities and the formats of the data and metadata they provide with the efforts of the international lattice data grid (ILDG) metadata and middleware working groups, whose goals are to develop standard formats for lattice QCD data and metadata and a uniform interface to archive facilities that store them. Services under development include those commonly associated with data grids: a service registry, a metadata database, a replica catalog, and an interface to a mass storage system. All services provide GSI authenticated web service interfaces following modern standards, including WSDL and SOAP, and accept and provide data and metadata following recent XML based formats proposed by the ILDG metadata working group.

**Poster Session 3 - Board: 23 / 197**

## Specifying Selection Criteria using C++ Expression Templates

**Author:** C. Jones<sup>1</sup>

<sup>1</sup> *CORNELL UNIVERSITY*

Generic programming as exemplified by the C++ standard library makes use of functions or function objects (objects that accept function syntax) to specialize generic algorithms for particular uses. Such separation improves code reuse without sacrificing efficiency. We employed this same technique in our combinatoric engine: DChain. In DChain, physicists combine lists of child particles to form a list of parent hypotheses. E.g.,  $d_0 = \text{pi.plus()} * \text{K.minus()}()$ . The selection criteria for the hypothesis is defined in a function or function object that is passed to the list's constructor.

However, C++ requires that functions and class declarations be defined outside the scope of a function.

Therefore physicists are forced to separate the code that defines the combinatorics from the code that sets the selection criteria. We will discuss a technique using C++ expression templates to allow users to define function objects using a mathematical expression directly in

their main function, e.g.,  
`func = (sqrt( beamEnergy*beamEnergy - vPMag*vPMag) >= 5.1*k_GeV).`

Use of such techniques can greatly decrease the coding ‘excess’ needed to perform an analysis.

**Core Software / 199**

## **EventStore: Managing Event Versioning and Data Partitioning using Legacy Data Formats**

**Author:** C. Jones<sup>1</sup>

**Co-authors:** D. Riley <sup>1</sup>; G. Sharp <sup>1</sup>; V. Kuznetsov <sup>1</sup>

<sup>1</sup> CORNELL UNIVERSITY

HEP analysis is an iterative process. It is critical that in each iteration the physicist’s analysis job accesses the same information as previous iterations (unless explicitly told to do otherwise). This becomes problematic after the data has been reconstructed several times. In addition, when starting a new analysis, physicists normally want to use the most recent version of reconstruction. Such version control is useful for data managed by a single physicist using a laptop or small groups of physicists at a remote institution in addition to the collaboration wide managed data.

In this presentation we will discuss our implementation of the EventStore which uses a data location, indexing and versioning service to manage legacy data formats (e.g. an experiment’s existing proprietary file format or Root files). A plug-in architecture is used to support adding additional file formats. The core of the system is used to implement three different sizes of services: personal, group and collaboration.

**Poster Session 3 - Board: 24 / 200**

## **Paths: Specifying Multiple Job Outputs via Filter Expressions**

**Author:** C. Jones<sup>1</sup>

<sup>1</sup> CORNELL UNIVERSITY

A common task for a reconstruction/analysis system is to be able to output different sets of events to different permanent data stores (e.g. files). This allows multiple related logical jobs to be grouped into one process and run using the same input data (read from a permanent data store and/or created from an algorithm). In our system, physicists can specify multiple output ‘paths’, where each path contains a group of filters followed by output ‘operations’. The filters are combined using a physicist specified boolean expression; only if the expression evaluates to true will the output operation be performed for that event. Paths do not explicitly contain the order that data objects should be

created as our system uses a 'data on demand' mechanism which causes data to be created the first time the data is requested. Separating the data dependencies from the event selection criteria vastly simplifies the task of creating a path, thereby making the facility more accessible to physicists.

### Computer Fabrics / 201

## Implementation of a reliable and expandable on-line storage for compute clusters

**Author:** J. VanWezel<sup>1</sup>

<sup>1</sup> FORSCHUNGSZENTRUM KARLSRUHE

**Corresponding Author:** jvw@iwr.fzk.de

The HEP experiments that use the regional center GridKa will handle large amounts of data. Traditional access methods via local disks or large network storage servers show limitations in size, throughput or data management flexibility.

High speed interconnects like Fibre Channel, iSCSI or Infiniband as well as parallel file systems are becoming increasingly important in large cluster installations to offer the scalable size and throughput needed for PetaByte storage. At the same time the reliable and proven NFS protocol allows local area storage access via traditional Ethernet very cost effectively.

The cluster at GridKa uses the General Parallel File System (GPFS) on a 20 node file server farm that connects to over 1000 FC disks via a Storage Area Network. The 130 TB on-line storage is distributed to the 390 node cluster via NFS. A load balancing system ensures an even load distribution and additionally allows for on-line file server exchange.

Discussed are the components of the storage area network, specific Linux tools, and the construction and optimisation of the cluster file system along with the RAID groups. A high availability is obtained and measurements prove high throughput under different conditions. The use of the file system administration and management possibilities is presented as is the implementation and effectiveness of the load balancing system.

### Poster Session 1 - Board: 18 / 203

## Parallel implementation of Parton String Model event generator

**Authors:** A. Asryan<sup>1</sup>; A. Zarochensev<sup>2</sup>; G. Feofilov<sup>3</sup>; P. Naumenko<sup>1</sup>; S. Nemnyugin<sup>4</sup>; V. Solodkov<sup>5</sup>; V. Vechernin<sup>3</sup>; V. Zolotarev<sup>3</sup>

<sup>1</sup> *Postgraduate*

<sup>2</sup> *Researcher*

<sup>3</sup> *Professor*

<sup>4</sup> *ASSOCIATE PROFESSOR*

<sup>5</sup> *Student*

**Corresponding Author:** nemnugin@mph.phys.spbu.ru

We report the results of parallelization and tests of the Parton String Model event generator at the parallel cluster of St.Petersburg State University Telecommunication center.

Two schemes of parallelization were studied. In the first approach master process coordinates work of slave processes, gathers and analyzes data. Results of MC calculations are saved in local files. Local files are sent to the host computer on which the program of data processing is started. The second approach uses the parallel write in the common file shared between all processes. In this case the load of a communication subsystem of the cluster grows. Both approaches are realized with MPICH library. Some problems including the pseudorandom number generation in parallel computations were solved.

The modified parallel version of the PSM code includes a number of the additional possibilities: a selection of the impact parameter windows, the account of acceptance of the experimental setup and trigger selection data, and the calculation of various long range correlations between such observables as mean transverse momentum and charged particles multiplicity.

**Distributed Computing Services / 204**

## **FroNtier: High Performance Database Access Using Standard Web Components in a Scalable Multi-tier Architecture**

**Authors:** B. Blumenfeld<sup>1</sup>; D. Litvintsev<sup>2</sup>; J. Kowalkowski<sup>3</sup>; L. Lueking<sup>4</sup>; M. Mathis<sup>1</sup>; M. Paterno<sup>3</sup>; P. Maksimovic<sup>1</sup>; S. Kosyakov<sup>3</sup>; S.P. White<sup>3</sup>

<sup>1</sup> *John Hopkins University*

<sup>2</sup> *FNAL*

<sup>3</sup> *Fermilab*

<sup>4</sup> *FERMILAB*

**Corresponding Author:** lueking@fnal.gov

A high performance system has been assembled using standard web components to deliver database information to a large number (thousands?) of broadly distributed clients. The CDF Experiment at Fermilab is building processing centers around the world imposing a high demand load on their database repository. For delivering read-only data, such as calibrations, trigger information and run conditions data, we have abstracted the interface that clients use to retrieve database objects. A middle tier is deployed that translates client requests into database specific queries and returns the data to the client as HTTP datagrams. The database connection management, request translation, and data encoding are accomplished in servlets running under Tomcat. Squid Proxy caching layers are deployed near the Tomcat servers as well as close to the clients to significantly reduce the load on the database and provide a scalable deployment model. This system is highly scalable, readily deployable, and has a very low administrative overhead for data delivery to a large, distributed audience. Details of how the system is built and used will be presented including its architecture, design, interfaces, administration, and performance measurements.

**Poster Session 2 - Board: 37 / 205**

## Database Usage and Performance for the Fermilab Run II Experiments

**Authors:** A. Kumar<sup>1</sup>; D. Bonham<sup>1</sup>; D. Box<sup>1</sup>; D. Litvintsev<sup>2</sup>; E. Gallas<sup>1</sup>; E. Wicklund<sup>1</sup>; J. Kowalkowski<sup>1</sup>; J. Trumbo<sup>1</sup>; L. Lueking<sup>3</sup>; M. Vittone-Wiersma<sup>1</sup>; N. Stanfield<sup>1</sup>; P. Maksimovic<sup>4</sup>; R. Jetton<sup>1</sup>; S. Kovich<sup>1</sup>; S.P. White<sup>1</sup>; T. Yasuda<sup>1</sup>; Y. Guo<sup>1</sup>

<sup>1</sup> *Fermilab*

<sup>2</sup> *FNAL*

<sup>3</sup> *FERMILAB*

<sup>4</sup> *John Hopkins University*

**Corresponding Author:** lueking@fnal.gov

The Run II experiments at Fermilab, CDF and D0, have extensive database needs covering many areas of their online and offline operations. Delivery of the data to users and processing farms based around the world has represented major challenges to both experiments. The range of applications employing databases includes data management, calibration (conditions), trigger information, run configuration, run quality, luminosity, and others. Oracle is the primary database product being used for these applications at Fermilab and some of its advanced features have been employed, such as table partitioning and replication. There is also experience with open source database products such as MySQL for secondary databases. A general overview of the operation, access patterns, and transaction rates is examined and the potential for growth in the next year presented. The two experiments, while having similar requirements for availability and performance, employ different architectures for database access. Details of the experience for these approaches will be compared and contrasted, as well as the evolution of the delivery systems throughout the run. Tools employed for monitoring the operation and diagnosing problems will also be described.

**Computer Fabrics / 207**

## ScotGrid: A prototype Tier 2 centre

**Authors:** A. Earl<sup>1</sup>; P. Clark<sup>2</sup>; S. Thorn<sup>None</sup>

<sup>1</sup> *UNIVERSITY OF EDINBURGH*

<sup>2</sup> *University of Edinburgh*

**Corresponding Author:** aearl@ed.ac.uk

ScotGrid is a prototype regional computing centre formed as a collaboration between the universities of Durham, Edinburgh and Glasgow as part of the UK's national particle physics grid, GridPP. We outline the resources available at the three core sites and our optimisation efforts for our user communities. We discuss the work which has been conducted in extending the centre to embrace new projects both from particle physics and new user communities and explain our methodology for doing this.

**Poster Session 2 - Board: 62 / 208**

## Mass Storage Management and the Grid

**Authors:** A. Earl<sup>1</sup>; P. Clark<sup>1</sup>

<sup>1</sup> *University of Edinburgh*

**Corresponding Author:** aearl@ed.ac.uk

The University of Edinburgh has an significant interest in mass storage systems as it is one of the core groups tasked with the roll out of storage software for the UK's particle physics grid, GridPP. We present the results of a development project to provide software interfaces between the SDSC Storage Resource Broker, the EU DataGrid and the Storage Resource Manager. This project was undertaken in association with the eDikt group at the National eScience Centre, the Universities of Bristol and Glasgow, Rutherford Appleton Laboratory and the San Diego Supercomputing Center.

**Online Computing / 209**

## **Experience with Real Time Event Reconstruction Farm for Belle Experiment**

**Authors:** I. Adachi<sup>1</sup>; M. Nakao<sup>1</sup>; N. Katayama<sup>1</sup>; R. Itoh<sup>1</sup>; S. Suzuki<sup>1</sup>; T. Higuchi<sup>1</sup>

<sup>1</sup> *KEK*

**Corresponding Author:** ryosuke.itoh@kek.jp

A sizeable increase in the machine luminosity of KEKB accelerator is expected in coming years. This may result in a shortage in the data storage resource for the Belle experiment in the near future and it is desired to reduce the data flow as much as possible before writing the data to the storage device.

For this purpose, a realtime event reconstruction farm has been installed in the Belle DAQ system. The farm consists of 60 linux-operated PC servers with dual CPUs. Every event from the event builder is distributed to one of the servers through a socket connection. A full event reconstruction is done on each server so that a sophisticated event selection can be performed to reduce the data flow. The same event reconstruction program as that used in the offline DST production runs on each farm server. Selected events are collected through socket connections and written to a fast disk array.

The farm has been being operated in the beam runs from the beginning of this year and processing the data at an average L1 trigger rate of 450Hz. The experience of the operation is reported at the conference. In particular, the performance of the full event reconstruction and selection is discussed in detail. A scheme to monitor the quality of processed data in real time is also described.

**Distributed Computing Systems and Experiences / 210**

## **The evolution of the distributed Event Reconstruction Control System in BaBar**

**Authors:** A. Ceseracciu<sup>1</sup>; T. Pulliam<sup>2</sup>

<sup>1</sup> *SLAC / INFN PADOVA*

<sup>2</sup> *Ohio State University*

The Event Reconstruction Control System of the BaBar experiment was redesigned in 2002, to satisfy the following major requirements: flexibility and scalability.



Because of its very nature, this system is continuously maintained to implement the changing policies, typical of a complex, distributed production environment. In 2003, a major revolution in the BaBar computing model, the Computing Model 2, brought a particularly vast set of new requirements in various respects, many of which had to be discovered during the early production effort, and promptly dealt with. Particularly, the reconstruction pipeline was expanded with the addition of a third stage. The first fast calibration stage was kept running at SLAC, USA, while the two stages doing most of the computation were moved to the ~400 CPU reconstruction facility of INFN, Italy.

In this paper, we summarize the extent and nature of the evolution of the Control System, and we demonstrate how the modular, well engineered architecture of the system allowed to efficiently adapt and expand it, while making great reuse of existing code, leaving virtually intact the core layer, and exploiting the “engineering for flexibility” philosophy.

**Poster Session 1 - Board: 14 / 211**

## **New compact hierarchical mass storage system at Belle realizing a peta-scale system with inexpensive ice-raid disks and an S-ait tape library**

**Authors:** J. Hikita<sup>1</sup>; K. Goto<sup>1</sup>; K. Tamura<sup>1</sup>; M. Makino<sup>2</sup>; M. Yokoyama<sup>3</sup>; N. Katayama<sup>4</sup>; T. Hibino<sup>3</sup>

<sup>1</sup> SONY Co.

<sup>2</sup> Systemworks Co.

<sup>3</sup> HIGH ENERGY ACCELERATOR RESEARCH ORGANIZATION

<sup>4</sup> KEK

The Belle experiment has accumulated an integrated luminosity of more than 240fb<sup>-1</sup> so far, and a daily logged luminosity now exceeds 800pb<sup>-1</sup>. These numbers correspond to more than 1PB of raw and processed data stored on tape and an accumulation of the raw data at the rate of 1TB/day. To meet these storage demands, a new cost effective, compact hierarchical mass storage system has been constructed. The system consists of commodity RAID systems using IDE disks and Linux PC servers as the front-end and a tape library system using the new high density SONY S-AIT tape as the back-end. The SONY Peta Serv software manages migration and restoration of the files between tapes and disks. The capacity of the tape library is, at the moment, 500 TB in three 19 inch racks and the RAID system, 64 TB in two 19 inch racks. An extension of the system to 1.2 PB tape library in eight racks with 150 TB RAID in four racks is planned. In this talk, experiences with the new system will be discussed and the performance of the system when used for data processing and physics analysis of the Belle experiment will be demonstrated.

**Poster Session 3 - Board: 61 / 212**

## **ATLAS Detector Description Database Architecture**

**Authors:** A. Vaniachine<sup>1</sup>; D. Malon<sup>1</sup>; K. Bernardet<sup>2</sup>; P. Nevski<sup>3</sup>; S. Baranov<sup>4</sup>

<sup>1</sup> ANL

<sup>2</sup> CPPM

<sup>3</sup> *BROOKHAVEN NATIONAL LABORATORY*

<sup>4</sup> *Freiburg University*

**Corresponding Author:** vaniachine@anl.gov

In addition to the well-known challenges of computing and data handling at LHC scales, LHC experiments have also approached the scalability limit of manual management and control of the steering parameters ("primary numbers") provided to their software systems. The laborious task of detector description benefits from the implementation of a scalable relational database approach. We have created and extensively exercised in the ATLAS production environment a primary numbers database utilizing NOVA relational database technologies. In our report we describe the architecture of the relational database deployed for the storage, management, and uniform treatment of primary numbers in ATLAS detector description. We describe the benefits of the ATLAS software framework (Athena) on-demand data access architecture, and an automatic system for code generation of more than 300 classes (about 10% of ATLAS offline code) for primary numbers access from the Athena framework. Integration with the LHC Interval-of-Validity database infrastructure, measures for tighter primary numbers database input control, experience with ATLAS Combined Testbeam geometry and conditions payload storage using NOVA technologies integrated with the LHC ConditionsDB implementation, methods for application-side resource pooling, new user tools for knowledge discovery, navigation and browsing, and plans for new primary numbers database developments, are also described.

**Poster Session 3 - Board: 7 / 213**

## OpenPAW

**Author:** G B. Barrand<sup>1</sup>

<sup>1</sup> *CNRS / IN2P3 / LAL*

**Corresponding Author:** barrand@lal.in2p3.fr

OpenPAW is for people that definitively do not want to quit the PAW command prompt, but seek anyway an implementation based over more modern technologies. We shall present the OpenScientist/Lab/opaw program that offers a PAW command prompt by using the OpenScientist tools (then C++, Inventor for doing graphic, Rio for doing the IO, OnX for the GUI, etc...). The OpenScientist/Lab package being also AIDA compliant, we shall show that it is possible to marry AIDA with a PAW command prompt.

**Computer Fabrics / 216**

## SRB system at Belle/KEK

**Authors:** A Manabe<sup>1</sup>; H Kuraishi<sup>2</sup>; I. Adachi<sup>3</sup>; K Ishikawa<sup>4</sup>; M. Mei<sup>5</sup>; N. Katayama<sup>1</sup>; S Honma<sup>2</sup>; S Kawabata<sup>1</sup>; S Yashiro<sup>1</sup>; S. Suzuki<sup>1</sup>; T Nakajima<sup>2</sup>; T. Sasaki<sup>3</sup>; Y Watase<sup>1</sup>; Y. Iida<sup>3</sup>

<sup>1</sup> *KEK*

<sup>2</sup> *Fujitsu*

<sup>3</sup> *HIGH ENERGY ACCELERATOR RESEARCH ORGANIZATION*

<sup>4</sup> *ISE, Chiba*

<sup>5</sup> *IHEP, Beijing*

The Belle experiment has accumulated an integrated luminosity of more than 240fb<sup>-1</sup> so far, and a daily logged luminosity now exceeds 800pb<sup>-1</sup>. These numbers correspond to more than 1PB of raw and processed data stored on tape and an accumulation of the raw data at the rate of 1TB/day. The processed, compactified data, together with Monte Carlo simulation data for the final physics analyses amounts to more than 100TB. The Belle collaboration consists of more than 55 institutes in 14 countries and at most of the collaborating institutions, active physics data analysis programs are being undertaken. To meet these storage and data distribution demands, we have tried to adopt a resource broker, SRB. We have installed the SRB system at KEK, Australia, and other collaborating institutions and have started to share data. In this talk, experiences with the SRB system will be discussed and the performance of the system when used for data processing and physics analysis of the Belle experiment will be demonstrated.

Online Computing / 217

## Performance of the ATLAS DAQ DataFlow system

**Authors:** A. CORSORADU<sup>1</sup>; A. DOS ANJOS<sup>2</sup>; A. KACZMARSKA<sup>3</sup>; A. KUGEL<sup>4</sup>; A. LANKFORD<sup>5</sup>; A. MISIEJUK<sup>6</sup>; B. DIGIROLAMO<sup>1</sup>; B. GORINI<sup>1</sup>; B. GREEN<sup>6</sup>; B. MARTIN<sup>1</sup>; C. HAEBERLI<sup>7</sup>; C. HINKELBEIN<sup>4</sup>; C. MEIROSU<sup>1</sup>; D. BOTTERILL<sup>8</sup>; D. FRANCIS<sup>1</sup>; D. PRIGENT<sup>1</sup>; E. PALENCIA CORTEZON<sup>1</sup>; E. PASQUALUCCI<sup>9</sup>; F. WICKENS<sup>8</sup>; G. CRONE<sup>10</sup>; G. KIEFT<sup>11</sup>; G. LEHMANN<sup>1</sup>; G. MORNACCHI<sup>1</sup>; G. unel<sup>12</sup>; H. ZOBERNIG<sup>13</sup>; H.P. BECK<sup>7</sup>; J. DAWSON<sup>14</sup>; J. PETERSEN<sup>1</sup>; J. SCHLERETH<sup>14</sup>; J. STRONG<sup>6</sup>; J. VERMEULEN<sup>11</sup>; J.A. BOGAERTS<sup>1</sup>; K. KORCYL<sup>3</sup>; K. NAKAYOSHI<sup>15</sup>; L. MAPELLI<sup>1</sup>; L. TREMBLET<sup>1</sup>; M. ABOLINS<sup>16</sup>; M. BERETTA<sup>17</sup>; M. CIOBOTARU<sup>1</sup>; M. GRUWE<sup>1</sup>; M. JOOS<sup>1</sup>; M. LE VINE<sup>18</sup>; M. LOSADA MAIA<sup>2</sup>; M. MULLER<sup>4</sup>; M. SHIMOJIMA<sup>19</sup>; M. YU<sup>4</sup>; M. ZUREK<sup>3</sup>; M.L. FERRER<sup>20</sup>; P. DE MATOS LOPES PINTO<sup>1</sup>; P. GOLONKA<sup>1</sup>; P. WERNER<sup>1</sup>; R. BLAIR<sup>14</sup>; R. CRANFIELD<sup>10</sup>; R. DOBINSON<sup>1</sup>; R. FERRARI<sup>21</sup>; R. HAUSER<sup>16</sup>; R. Hughes-Jones<sup>22</sup>; R. MCLAREN<sup>1</sup>; R. MOMMSEN<sup>5</sup>; R. SPIWOKS<sup>1</sup>; S. GADOMSKI<sup>7</sup>; S. GAMEIRO<sup>1</sup>; S. HAAS<sup>1</sup>; S. STANCU<sup>1</sup>; T. MAENO<sup>1</sup>; V. PEREZ REALE<sup>7</sup>; W. VANDELLI<sup>21</sup>; Y. ERMOLINE<sup>16</sup>; Y. HASEGAWA<sup>23</sup>; Y. NAGASAKA<sup>24</sup>; Y. YASU<sup>15</sup>

<sup>1</sup> CERN, Geneva, Switzerland

<sup>2</sup> Universidade Federal do Rio de Janeiro, COPPE/EE, Brazil

<sup>3</sup> Polish Academy of Sciences, Cracow, Poland

<sup>4</sup> Universitaet Mannheim, Germany

<sup>5</sup> University of California, Irvine, USA

<sup>6</sup> University of London, Egham, UK

<sup>7</sup> University of Bern, Switzerland

<sup>8</sup> RAL, Didcot, UK

<sup>9</sup> INFN Roma, Italy

<sup>10</sup> University College London, UK

<sup>11</sup> NIKHEF, The Netherlands

<sup>12</sup> UNIVERSITY OF CALIFORNIA AT IRVINE AND CERN

<sup>13</sup> University of Wisconsin, USA

<sup>14</sup> Argonne National Laboratory, USA

<sup>15</sup> KEK, Japan

<sup>16</sup> Michigan State University, USA

<sup>17</sup> INFN Frascati, Italy

<sup>18</sup> BNL, USA

<sup>19</sup> Nagasaki Institute of Applied Science, Japan

<sup>20</sup> INFN Frascati, Italy

<sup>21</sup> INFN Pavia, Italy

<sup>22</sup> THE UNIVERSITY OF MANCHESTER

<sup>23</sup> Shinshu University, Japan

<sup>24</sup> Hiroshima Institute of Technology, Japan

**Corresponding Author:** gokhan.unel@cern.ch

The ATLAS Trigger and DAQ system is designed to use the Region of Interest (RoI) mechanism to reduce the initial Level 1 trigger rate of 100 kHz down to about 3.3 kHz Event Building rate.

The DataFlow component of the ATLAS TDAQ system is responsible for the reading of the detector specific electronics via 1600 point to point readout links, the collection and provision of RoI to the Level 2 trigger, the building of events accepted by the Level 2 trigger and their subsequent input to the Event Filter system where they are subject to further selection criteria.

To validate the design and implementation of the DAQ DataFlow system, a prototype setup representing 20% of the final system, has been put together at CERN. This baseline prototype contains 68 PCs running Linux, and exchanging data via a 64-port and a 31-port Gigabit Ethernet switches for Event Building and RoI Collection. The system performance is measured by playing back simulated data through the system and running prototype algorithms in the Level 2 trigger. In parallel a full discrete event model of the system has been developed and tuned to the testbed results as an aid to studying the system performance at and beyond the size of the prototype setup.

Measurements will be presented on the performance of the prototype setup, showing that the components of the current integrated system implementation can already sustain their nominal ATLAS requirements using existing hardware and Gigabit network technology: 20 kHz RoI Collection rate per readout link, 3 kHz Event Building rate and 70 Mbyte/s throughput per event building node. The use of these results to calibrate the model will also be presented along with the model predictions for the performance of the final DAQ DataFlow system.

**Distributed Computing Services / 218**

## On Distributed Database Deployment for the LHC Experiments

**Authors:** A. Vaniachine<sup>1</sup>; D. Duellmann<sup>2</sup>; D. Malon<sup>1</sup>; I. Bird<sup>2</sup>; J. Shiers<sup>2</sup>; M. Girone<sup>2</sup>; M. Lamanna<sup>2</sup>

<sup>1</sup> ANL

<sup>2</sup> CERN

**Corresponding Authors:** dirk.duellmann@cern.ch, jamie.shiers@cern.ch

While there are differences among the LHC experiments in their views of the role of databases and their deployment, there is relatively widespread agreement on a number of principles:

1. Physics codes will need access to database-resident data. The need for database access is not confined to middleware and services: physics-related data will reside in databases.
2. Database-resident data will be distributed, and replicated. A single, centralized database, at CERN or elsewhere, does not suffice.
3. Distributed deployment infrastructure should be open to the use of different technologies as appropriate at the various Tier N sites.

A variety of approaches to distributed deployment have been explored in the context of individual experiments; indeed, a degree of distributed deployment has been integral to the computing model tests of some experiments (cf. ATLAS) in their 2004 data challenges. Approaches to replication have also been investigated in the context of specific databases, often with vendor-specific replication tools (e.g., Oracle Replication via Streams for the LCG File Catalog and the Oracle instantiation of the LCG conditions database; MySQL tools for replication in the MySQL instantiation of the LCG conditions database). XML exchange mechanisms have also been discussed. Distributed database deployment, though, is more than a middleware and applications software issue—a successful strategy must involve those who will be responsible for systems deployment and administration at LHC grid sites.

We describe the status of ongoing work in this area, and discuss the prospects for components of a common approach to distributed deployment in the time frame of the 2005 LHC data challenges.

**Poster Session 1 - Board: 34 / 219**

## The ATLAS DAQ system

**Authors:** G. unel<sup>1</sup>; TDAQ ATLAS<sup>2</sup>

<sup>1</sup> UNIVERSITY OF CALIFORNIA AT IRVINE AND CERN

<sup>2</sup> ATLAS

**Corresponding Author:** gokhan.unel@cern.ch

The 40 MHz collision rate at the LHC produces ~25 interactions per bunch crossing within the ATLAS detector, resulting in terabytes of data per second to be handled by the detector electronics and the trigger and DAQ system. A Level 1 trigger system based on custom designed and built electronics will reduce the event rate to 100 kHz.

The DAQ system is responsible for the readout of the detector specific electronics via 1600 point to point links hosted by Readout Subsystems, the collection and provision of “Region of Interest data” to the Level 2 trigger, the building of events accepted by the Level 2 trigger and their subsequent input to the Event Filter system where they are subject to further selection criteria. Also the DAQ provides the functionality for the configuration, control, information exchange and monitoring of the whole ATLAS detector.

The baseline ATLAS DAQ architecture and its implementation will be introduced. In this implementation, the configuration, control, information exchange and monitoring functionalities are provided with CORBA; the control aspects are handled by an expert system based on CLIPS and the data connection between 150 Readout Subsystems, up to 500 Level 2 Processing Units and to 80 Event building nodes is done Gigabit Ethernet network technology.

The experience from using the DAQ system in a combined test beam environment where all ATLAS subdetectors are participating will be presented. The current performances of some DAQ components as measured in the laboratory environment will be summarized. Some results from the large scale functionality tests, on a system of a 300 nodes, aimed at understanding the scalability of the current implementation will also be shown.

**Poster Session 3 - Board: 8 / 220**

## Rio

**Author:** G B. Barrand<sup>1</sup>

<sup>1</sup> CNRS / IN2P3 / LAL

**Corresponding Author:** barrand@lal.in2p3.fr

Rio (for ROOT IO) is a rewriting of the file IO system of ROOT. We shall present our strong motivations of doing this tedious work. We shall present the main choices done in the Rio implementation (then by opposition to what we don't like in ROOT). For example, we shall say why we believe that an IO package is not a drawing package (no TClass::Draw) ; why someone should use pure abstract interfaces in such package (for example to open cleanly to various dictionaries) ; how we can have a more reliable system than ROOT (for example, by simply protect the various buffer overflows).

We shall cover the today role of Rio within OpenScientist to store histograms and tuples. We shall present the effort done around Gaudi, at the beginning of 2003, to read LHCb events with Rio (then in the "before POOL" system).

We shall present our views about the LCG proposed solution for storage, that is to say POOL over ROOT, and why the author believe that this coarse graining assembly is simply poor software engineering. We shall explain why CERN, due to its fermionic sociology, is going to miss an essential target : an appealing open source object oriented data base for HEP. We shall explain then how to do it without this lab, then passing from Rio to RioGrande...

**Poster Session 1 - Board:** 30 / 221

## **Integrating Mutiple PC Farms into an uniform computing System with Maui**

**Author:** G. Sun<sup>1</sup>

<sup>1</sup> *INSTITUE OF HIGH ENERGY PHYSICS*

**Corresponding Author:** sungx@mail.ihep.ac.cn

These are several on-going experiments at IHEP, such as BES, YBJ, and CMS collaboration with CERN. each experiment has its own computing system, these computing systems run separately. This leads to a very low CPU utilization due to different usage period of each experiment. The Grid technology is a very good candidate for integrating these separate computing systems into a "single image", but it is too early to be put into a production system as it is not stable and user-friendly as well. A realistic choice is to implement such an integration and sharing with Maui, an advacned scheduler. Each PC farm is thought as a partition, which is assigned high priority to its owner users with preemtor feature. this paper will describe the detail of implementation with Maui scheduler, as well as the entire system architecture and configuration and fuctions.

**Core Software / 222****The SEAL C++ Reflection System****Authors:** P. MATO<sup>1</sup>; S. Roiser<sup>1</sup><sup>1</sup> CERN**Corresponding Author:** stefan.roiser@cern.ch

The C++ programming language has very limited capabilities for reflection information about its objects. In this paper a new reflection system will be presented, which allows complete introspection of C++ objects and has been developed in the context of the CERN/LCG/SEAL project in collaboration with the ROOT project.

The reflection system consists of two different parts. The first part is a code generator that produces automatically reflection information from existing C++ classes. This generation of the reflection information is done in a non intrusive way, which means that the original C++ classes definition do not need to be changed or instrumented. The second part of the reflection system is able to load/build this information in memory and provides an API to the user.

The user can query reflection information from any C++ class and also interact generically with the objects, like invocation of functions, setting and getting data members or constructing and deleting objects. When designing the different packages it was taken care of having minimal dependencies on external software and a possibility to port the software to different platforms/compilers.

A quick overview of the current implementation in use by the LCG SEAL and POOL projects will be given. A more detailed description of the new model, which aims to reflect the complete C++ language and to be a common reflection system used also by the ROOT framework, will be given.

**Online Computing / 223****A Hardware Based Cluster Control And Management System****Authors:** H. TILSNER<sup>1</sup>; L. Hess<sup>2</sup>; R. Panse<sup>2</sup>; T. Alt<sup>2</sup>; V. Lindenstruth<sup>2</sup><sup>1</sup> *Kirchhoff Institute of Physics, Heidelberg, Germany*<sup>2</sup> *KIRCHHOFF INSTITUTE FOR PHYSICS - UNIVERSITY OF HEIDELBERG***Corresponding Author:** panse@kip.uni-heidelberg.de

Super-computers will be replaced more and more by PC cluster systems. Also future LHC experiments will use large PC clusters. These clusters will consist of off-the-shelf PCs, which in general are not built to run in a PC farm. Configuring, monitoring and controlling such clusters requires a serious amount of time consuming and administrative effort.

We propose a cheap and easy hardware solution for this issue. The main item of our cluster control system is the Cluster Interface Agent card (CIA).

The CIA card is a low-cost PCI expansion card equipped with a network interface. With the aid of the CIA card the computer can be fully controlled remotely, independent of the state of the node itself. The card combines a number of feature needed for this remote

control, including power management and reset. The card operates entirely independent of the PC and can remain powered while the PC may even be powered down. It offers a wide range of automatization features, including automatic installation of the operating system, changing BIOS settings or booting a rescue disk and also to monitor and debug the node. With the aid of PCI scans and hardware tests errors and pending failures can be easily detected in an early stage.

Working prototypes exist. The presentation will outline the status of the project and first implementation results of the preproduction devices, currently being built.

**Grid Security / 224**

## **Evaluation of Grid Security Solutions using Common Criteria**

**Authors:** M. RIGUIDEL<sup>1</sup>; S. NAQVI<sup>1</sup>

<sup>1</sup> TELECOM PARIS

**Corresponding Author:** naqvi@enst.fr

In the evolution of computational grids, security threats were overlooked in the desire to implement a high performance distributed computational system. But now the growing size and profile of the grid require comprehensive security solutions as they are critical to the success of the endeavour. A comprehensive security system, capable of responding to any attack on grid resources, is indispensable to guarantee its anticipated adoption by both the users and the resource providers. Some security teams have started working on establishing in-depth security solutions. The evaluation of their grid security solutions requires excellent criteria to assure sufficient security to meet the needs of its users and resource providers. Grid community's lack of experience in the exercise of the Common Criteria (CC), which was adopted in 1999 as an international standard for security product evaluation, makes it imperative that efforts be exerted to investigate the prospective influence of the CC in advancing the state of grid security. This article highlights the contribution of the CC to establishing confidence in grid security, which is still in need of considerable attention from its designers. The process of security evaluation is outlined and the roles each part of the evaluation may play in obtaining confidence are examined.

**Poster Session 3 - Board: 1 / 225**

## **A database perspective on CMS detector data**

**Authors:** A.T.M. Aerts<sup>1</sup>; F. Glege<sup>2</sup>; M. Liendl<sup>2</sup>

<sup>1</sup> TU Eindhoven

<sup>2</sup> CERN

**Corresponding Authors:** martin.liendl@cern.ch, a.t.m.aerts@tue.nl

Building a state of the art high energy physics detector like CMS requires strict interoperability and coherency in the design and construction of all sub-systems comprising the detector. This issue is especially critical for the many database components that are planned for storage of the various categories of data related to



the construction, operation, and maintenance of the detector like event data, slow control data, conditions data, calibration data, event meta data, etc ... . The data structures needed to operate the detector as a whole need to be present in the database before the data is entered. Changing these structures for a database system that already contains a substantial amount of data is a very time and labour consuming exercise that needs to be avoided. Cases where the detector needs to be treated as a whole are detector operation (control, error tracking, conditions) and the interfacing of there construction and simulation software.

In this paper we propose to use the detector geometry as the structure connecting the various elements. The design and implementation of a relational database that captures the CMS detector geometry and the detector components is discussed. The detector geometry can serve as a core component in several other databases in order to make them interoperable. It also provides a common viewpoint between the physical detector and its image in the reconstruction software. Some of the necessary extensions to the detector description are discussed.

**Poster Session 2 - Board: 39 / 226**

## **Beyond Persistence: Developments and Directions in ATLAS Data Management**

**Authors:** A. Schaffer<sup>1</sup>; D. Malon<sup>2</sup>

<sup>1</sup> *LAL Orsay*

<sup>2</sup> *ANL*

As ATLAS begins validation of its computing model in 2004, requirements imposed upon ATLAS data management software move well beyond simple persistence, and beyond the “read a file, write a file” operational model that has sufficed for most simulation production. New functionality is required to support the ATLAS Tier 0 model, and to support deployment in a globally distributed environment in which the preponderance of computing resources—not only CPU cycles but data services as well—reside outside the host laboratory.

This paper takes an architectural perspective in describing new developments in ATLAS data management software, including the ATLAS event-level metadata system and related infrastructure, and the mediation services that allow one to distinguish writing from registration and selection from retrieval, in a manner that is consistent both for event data and for time-varying conditions. The ever-broader role of databases and catalogs, and issues related to the distributed deployment thereof, are also addressed.

**Event Processing / 227**

## **Muon Reconstruction Software in CMS**

**Author:** N. Neumeister<sup>1</sup>

<sup>1</sup> *CERN / HEPHY VIENNA*

**Corresponding Author:** [norbert.neumeister@cern.ch](mailto:norbert.neumeister@cern.ch)

The CMS detector has a sophisticated four-station muon system made up of tracking chambers (Drift Tubes, Cathode Strip Chambers) and dedicated trigger chambers. A muon reconstruction software based on Kalman filter techniques has been developed which reconstructs muons in the standalone muon system, using information from all three types of muon detectors, and links the resulting muon tracks with tracks reconstructed in the silicon tracker. The software is designed to work for both, offline reconstruction and for online event selection within the CMS High-Level Trigger (HLT). Since the quality of the selection algorithms used in the HLT system is of utmost importance the software has been designed using modern object-oriented software techniques and is implemented within the CMS reconstruction software framework. The system should be able to select events with final-state muons, indicating interesting physics. The design implementation and performance of the CMS muon reconstruction software is presented. We will show that offline code with little modifications can be used in the HLT system, by making use of the concepts of regional and conditional reconstruction. The implementation and performance of possible HLT selection algorithms are illustrated.

## Distributed Computing Systems and Experiences / 228

### HEP@HOME - A distributed computing system based on BOINC

**Authors:** A. Amorim<sup>1</sup>; J. Villate<sup>2</sup>; P. Andrade<sup>3</sup>

<sup>1</sup> *FCUL Portugal / CERN*

<sup>2</sup> *FEUP Portugal*

<sup>3</sup> *University of Porto*

Project SETI@HOME has proven to be one of the biggest successes of distributed computing during the last years. With a quite simple approach SETI manages to process huge amounts of data using a vast amount of distributed computer power.

To extend the generic usage of these kinds of distributed computing tools, BOINC (Berkeley Open Infrastructure for Network Computing) is being developed. In this communication we propose a BOINC version tailored to the specific requirements of the High Energy Physics (HEP) community - the HEP@HOME

The HEP@HOME will be able to process large amounts of data using virtually unlimited computing power, as BOINC does, and it should be able to work according to HEP specifications.

One of the main applications of distributed computing is distributed data analysis. In HEP the amounts of data to be analyzed have a large order of magnitude. Therefore, one of the design principles of this tool is to avoid data transfer - computation is done where data is stored. This will allow scientists to run their analysis applications even if they do not have a local copy of the data to be analyzed, taking advantage of either very large farms of dedicated computers or using their colleagues desktop PCs. This tool also satisfies other important requirements in HEP, namely, security, fault-tolerance and monitoring.

**Event Processing / 229****Fast reconstruction of tracks in the inner tracker of the CBM experiment****Author:** I. Kisel<sup>1</sup><sup>1</sup> UNIVERSITY OF HEIDELBERG, KIRCHHOFF INSTITUTE OF PHYSICS

Typical central Au-Au collision in the CBM experiment (GSI, Germany) will produce up to 700 tracks in the inner tracker. Large track multiplicity together with presence of nonhomogeneous magnetic field make reconstruction of events complicated.

A cellular automaton method is used to reconstruct tracks in the inner tracker. The cellular automaton algorithm creates short track segments in neighbored detector planes and links them into tracks. Being essentially local and parallel the cellular automaton avoids exhaustive combinatorial search, even when implemented on conventional computers. Since the cellular automaton operates with highly structured information, the amount of data to be processed in the course of the track search is significantly reduced. The method employs a very simple track model which leads to utmost computational simplicity and fast algorithm.

Efficiency of track reconstruction for particles detected in at least three stations is presented. Tracks of high momentum particles are reconstructed very well with efficiency about 98%, while multiple scattering in detector material leads to lower reconstruction efficiency of slow particles.

**Computer Fabrics / 230****CASTOR: Operational issues and new Developments****Authors:** B. Couturier<sup>1</sup>; J-D. Durand<sup>1</sup>; O. Barring<sup>1</sup>; S. Ponce<sup>1</sup><sup>1</sup> CERN**Corresponding Author:** jean-damien.durand@cern.ch

The Cern Advanced STORAGE (CASTOR) system is a scalable high throughput hierarchical storage system developed at CERN. CASTOR was first deployed for full production use in 2001 and has expanded to now manage around two PetaBytes and almost 20 million files. CASTOR is a modular system, providing a distributed disk cache, a stager, and a back end tape archive, accessible via a global logical name-space.

This paper focuses on the operational issues of the system currently in production, and first experiences with the new CASTOR stager which has undergone a significant redesign in order to cope with the data handling challenges posed by the LHC, which will be commissioned in 2007.

The design target for the new stager was to scale to another order of magnitude above the current CASTOR, namely to be able to sustain peak rates of the order of 1000 file open requests per second for a PetaByte disk pool. The new developments have been inspired by the problems which arose managing massive installations of commodity storage hardware. The farming of disk servers poses new challenges to the disk cache management: request scheduling; resource sharing and partitioning; automated configuration and monitoring; and fault tolerance of unreliable hardware

Management of the distributed component based CASTOR system across a large farm, provides an ideal example of the driving forces for the development of automated management suites. Quattor and Lemon frameworks naturally address CASTOR's operational requirements, and we will conclude by describing their deployment on the masstorage systems at CERN.

**Distributed Computing Systems and Experiences / 231**

## Production Management Software for the CMS Data Challenge

**Authors:** A. Anzar<sup>1</sup>; A. Fanfani<sup>2</sup>; C. Grandi<sup>3</sup>; G. Graham<sup>1</sup>; G. Guglielmo<sup>1</sup>; J. Andreeva<sup>4</sup>; N. Darmenov<sup>5</sup>; N. Ratnikova<sup>1</sup>; P. GARCIA-ABIA<sup>6</sup>; T. Wildish<sup>7</sup>; V. Lefebure<sup>5</sup>

<sup>1</sup> *Fermilab*

<sup>2</sup> *INFN, Bologna*

<sup>3</sup> *INFN-Bologna*

<sup>4</sup> *UC Riverside*

<sup>5</sup> *CERN*

<sup>6</sup> *CIEMAT*

<sup>7</sup> *Princeton University*

**Corresponding Author:** julia.andreeva@cern.ch

One of the goals of CMS Data Challenge in March-April 2004 (DC04) was to run reconstruction for sustained period at 25 Hz input rate with distribution of the produced data to CMS T1 centers for further analysis.

The reconstruction was run at the T0 using CMS production software, of which the main components are RefDB (CMS Monte Carlo 'Reference Database' with Web interface) and McRunjob (a framework for creation and submission of large numbers of Monte Carlo jobs).

This paper presents an overview of CMS production cycle, describing production tools, covering data processing, bookkeeping and publishing issues, in the context of their use during the T0 reconstruction part of DC04.

**Computer Fabrics / 233**

## dCache, LCG Storage Element and enhanced use cases

**Author:** P. Fuhrmann<sup>1</sup>

<sup>1</sup> *DESY*

**Corresponding Author:** patrick.fuhrmann@desy.de

The dCache software system has been designed to manage a huge amount of individual disk storage nodes and let them appear under a single file system root. Beside a variety of other features, it supports the GridFtp dialect, implements the Storage Resource Manager interface (SRM V1) and can be linked against the CERN GFAL software layer. These abilities makes dCache a perfect Storage Element in the context of LCG and possibly future grid initiatives as well.

During the last year, dCache has been deployed at dozens of Tier-I and Tier-II centers for the CMS and CDF experiments in the US and Europe, including Fermilab, Brookhaven, San Diego, Karlsruhe and CERN. The largest implementation, the CDF system at FERMI, provides 150 TeraBytes of disk space and delivers up to 50 TeraBytes/day to its clients.

Sites using the LCG dCache distribution are more or less operating the cache as black box and little knowledge is available about customization and enhanced features.

This presentation is therefore intended to make non dCache users curious and enable dCache users to better integrate dCache into their site specific environment. Beside many other topics, paper will touch on the possibility of dCache to closely cooperate with tertiary storage systems, like Enstore, Tsm and HPSS. It will describe the way dCache can be configured to attach different pool nodes to different user groups but let them all use the same set of fall back pools. We will explain how dCache takes care of dataset replication, either by configuration or by automatic detection of data access hot spots. Finally we will report on ongoing development plans.

**Poster Session 2 - Board: 32 / 234**

## **The role of legacy services within ATLAS DC2**

**Authors:** DC2 production system team<sup>1</sup>; J. Kennedy<sup>2</sup>

<sup>1</sup> ATLAS

<sup>2</sup> LMU Munich

**Corresponding Author:** john.kennedy@cern.ch

This paper presents an overview of the legacy interface provided for the ATLAS DC2 production system. The term legacy refers to any non-grid system which may be deployed for use within DC2. The reasoning behind providing such a service for DC2 is twofold in nature. Firstly, the legacy interface provides a backup solution should unforeseen problems occur while developing the grid based interfaces. Secondly, this system allows DC2 to use resources which have yet to deploy grid software, thus increasing the available computing power for the Data Challenge.

The aim of the legacy system is to provide a simple framework which is easily adaptable to any given computing system. Here the term computing system refers to the batch system provided at a given site and also to the structure of the computing and storage systems at that site. The legacy interface provides the same functionality as the grid based interfaces and is deployed transparently within the DC2 production system. Following the push-pull model implemented for DC2 the system pulls jobs from a production database and pushes them onto a given computing/batch system.

In a world which is becoming increasingly grid orientated this project allows us to evaluate the role of non-grid solutions in dedicated production environments. Experiences, both good and bad, gained during DC2 are presented and the future of such systems is discussed.

Core Software / 235

## CMS Detector Description: New Developments

**Author:** M. Case<sup>1</sup>

**Co-authors:** A.J. Muhammad<sup>2</sup>; A.T.M. Aerts<sup>3</sup>; M. Liendl<sup>4</sup>

<sup>1</sup> UNIVERSITY OF CALIFORNIA, DAVIS

<sup>2</sup> University of the West of England

<sup>3</sup> Eindhoven University of Technology

<sup>4</sup> CERN

**Corresponding Author:** case@ucdhep.ucdavis.edu

The CMS Detector Description Database (DDD) consists of a C++ API and an XML based detector description language. DDD is used by the CMS simulation (OSCAR), reconstruction (ORCA), and visualization (IGUANA) as well by test beam software that relies on those systems. The DDD is a sub-system within the COBRA framework of the CMS Core Software. Management of the XML is currently done using a separate Geometry project in CVS.

We give an overview of the DDD integration and report on recent developments concerning detector description in CMS software:

- The ability of client software to describe sub-detectors by providing an algorithm plug-in in C++ based on SEAL plug-in facilities. A typical algorithm plug-in makes use of the DDD API to describe detector properties. Through the API seamless access to data defined via the XML description language is ensured.
- An Oracle schema was recently developed and the database populated by a DDD application. The geometrical structure of the detector is seen as a skeleton to which conditions or configuration data can be attached.
- A C++ streaming mechanism to output the geometry as binary files was developed. This representation can be read into memory much more rapidly than the XML files can be parsed.

The DDD API shields clients from each of the possible input sources. Even the simultaneous use of several different input sources is possible through various configuration options in the framework COBRA.

Poster Session 1 - Board: 13 / 237

## Future processors: What is on the horizon for HEP farms?

**Author:** S. Jarp<sup>1</sup>

<sup>1</sup> CERN

**Corresponding Author:** sverre.jarp@cern.ch

In 1995 I predicted that the dual-processor PC would start invading HEP computing and a couple of years later the x86-based PC was omnipresent in our computing facilities. Today, we cannot imagine HEP computing without thousands of PCs at the heart. This talk will look at some of the reasons why we may one day be forced to leave this sweet-spot. This would be not because we (the HEP community) want to, but rather because other market forces may pull in different directions. Amongst such forces, I will review the new generation of powerful game consoles where IBM's Power processor

is currently making strong inroads. Then I will look at the huge mobile market where low-powered processing rules rather than power-hungry DP Xeon/Xeon-like processors, and thirdly I will explore in my talk the promise of enterprise servers with a large number of processors on each die (so-called Core Multi-Processors). For all the scenarios, we must, of course, keep in mind that HEP can only move when the price-performance ratio is right.

## Event Processing / 238

### H100 - an analysis framework for H1

**Author:** J. Katzy<sup>1</sup>

<sup>1</sup> DESY, HAMBURG

**Corresponding Author:** judith.katzy@desy.de

During the years 2000 and 2001 the HERA machine and the H1 experiment performed substantial luminosity upgrades. To cope with the increased demands on data handling an effort was made to redesign and modernize the analysis software. Main goals were to lower turn-around time for physics analysis by providing a single framework for data storage, event selection, physics analysis and event display. The new object oriented analysis environment is using C++ and is based on the RooT framework. Data layers with a high level of abstraction are defined, i.e. physics particles, event summary information and user specific information.

A generic interface makes the use of reconstruction output stored in BOS format transparent to the user. Links between all data layers and partial event reading allow correlating quantities of different abstraction levels with high performance. Detailed physics analysis is performed by passing transient data between different analysis modules. Binding of existing fortran based libraries on demand allows the use of existing utility functions and interface to the existing data base. On this basis tools with enhanced functionality are provided. This framework has become standard for data analyses of the previously and currently collected data.

## Computer Fabrics / 239

### CERN's openlab for Datagrid applications

**Author:** S. Jarp<sup>1</sup>

<sup>1</sup> CERN

**Corresponding Author:** sverre.jarp@cern.ch

For the last 18 months CERN has collaborated closely with several industrial partners to evaluate, through the opencluster project, technology that may (and hopefully will) play a strong role in the future computing solutions, primarily for LHC but possibly also for other HEP computing environments. Unlike conventional field testing where solutions from industry are evaluated rather independently, the openlab principle is based on active collaboration between all partners, with the common goal of constructing a coherent system.

The talk will discuss our experience to date with the following hardware

- 64-bit computing (in our case represented by the Itanium processor). This will also include the porting of applications and Grid software to 64 bits.
- Rack mounted servers
- The use of 10 Gbps Ethernet for both LAN and WAN connectivity
- An iSCSI-based Storage System that promises to scale to Petabyte dimensions
- The use of 10 Gbps Infiniband as a cluster interconnect

On the software side we will review our experience with the latest grid-enabled release of Oracle, the so-called release "10g".

The talk will review the results obtained so far, either in stand alone tests or as part of the larger LCG testbed, and it will describe the plans for the future in this three-year collaboration with industry.

**Poster Session 3 - Board: 42 / 240**

## Status and Plans of the LCG PI Project

**Authors:** A. Pfeiffer<sup>1</sup>; H.C. Lee<sup>2</sup>; L. Moneta<sup>1</sup>; V. Innocente<sup>1</sup>; W. Ueng<sup>2</sup>

<sup>1</sup> CERN, PH/SFT

<sup>2</sup> Academia Sinica, Taipei

**Corresponding Author:** andreas.pfeiffer@cern.ch

In the context of the LHC Computing Grid (LCG) project, the Applications Area develops and maintains that part of the physics applications software and associated infrastructure that is shared among the LHC experiments.

The Physicist Interface (PI) project of the LCG Application Area encompasses the interfaces and tools by which physicists will directly use the software. In collaboration with users from the experiments, work has concentrated on the Analysis Services subsystem, where implementations of the AIDA interfaces for (binned and unbinned) histogramming, fitting and minimization as well as manipulation of tuples have been developed and adapted. In addition, bindings of these interfaces to the Python interpreted language have been done using the dictionary subsystem of the SEAL project.

The actual status and the future planning of the project will be presented.

**Core Software / 241**

## Software management infrastructure in the LCG Application Area

**Authors:** A. Pfeiffer<sup>1</sup>; E. Poinsignon<sup>2</sup>; S. Ashby<sup>2</sup>; Y. Patois<sup>2</sup>

<sup>1</sup> CERN, PH/SFT

<sup>2</sup> CERN

**Corresponding Author:** andreas.pfeiffer@cern.ch

In the context of the SPI project in the LCG Application Area, a centralized s/w management infrastructure has been deployed. It comprises of a suite of scripts handling the building and validating of the releases of the various projects as well as providing a customized packaging of the released s/w. Emphasis



was put on the flexibility of the packaging and distribution solution as it should cover a broad range of use-cases and needs, ranging from full packages for developers in the projects and experiments to a minimal set of libraries and binaries for specific applications running, e.g., on grid nodes. In addition, regular reviews of the QA analysis of the releases of the projects are performed and fed back to the project leaders to improve the overall quality of the software produced. The present status and future perspectives of this activity will be presented and we will show examples of quality improvement in the projects.

**Poster Session 2 - Board: 19 / 243**

## **A general and flexible framework for virtual organization application tests in a grid system**

**Authors:** A. Pierro<sup>1</sup>; G. Donvito<sup>1</sup>; G. Maggi<sup>1</sup>; T. Coviello<sup>1</sup>

<sup>1</sup> INFN Via E. Orabona 4 I - 70126 Bari Italy

**Corresponding Author:** tommaso.coviello@ba.infn.it

A grid system is a set of heterogeneous computational and storage resources, distributed on a large geographic scale, which belong to different administrative domains and serve several different scientific communities named Virtual Organizations (VOs). A virtual organization is a group of people or institutions which collaborate to achieve common objectives. Therefore such system has to guarantee the coexistence of different VO's applications providing them the suitable run-time environment. Hence tools are needed both at local and central level for testing and detecting eventually bad software configuration on a grid site.

In this paper we present a web based tool which permits to a Grid Operational Centre (GOC) or a Site Manager to test a grid site from the VO viewpoint.

The aim is to create a central repository for collecting both existing and emerging VO tests. Each VO test may include one or more specific application tests, and each test could include one or more subtests, arranged in a hierarchic structure.

A general and flexible framework is presented capable to include VO tests straightforwardly by means of a description file. Submission of a bunch of tests to a particular grid site is made available through a web portal. On the same portal, past and current results and logs can be browsed.

**Poster Session 3 - Board: 31 / 244**

## **Validation of the GEANT4 Bertini Cascade model and data analysis using the Parallel ROOT Facility on a Linux cluster**

**Authors:** A. Heikkinen<sup>1</sup>; T. Linden<sup>2</sup>

<sup>1</sup> Helsinki Institute of Physics HIP

<sup>2</sup> HELSINKI INSTITUTE OF PHYSICS

**Corresponding Author:** tomas.linden@helsinki.fi

Validation of hadronic physics processes of the Geant4 simulation toolkit is a very important task to ensure adequate physics results for the experiments being built at the Large Hadron Collider. We report on simulation results obtained using the Geant4 Bertini cascade double-differential production cross-sections for various target materials and incident hadron kinetic energies between 0.1-10 GeV [1, 2].

The cross-section benchmark study in this work has been performed using a Linux cluster set up with the Red Hat Linux based NPACI Rocks Cluster Distribution. For analysis of the validation data we have used the Parallel ROOT Facility (PROOF). PROOF has been designed for setting up a parallel data analysis environment in an inhomogeneous computing environment. Here we use a homogeneous Rocks cluster and automatic class generation for PROOF event data-analysis [3].

[1] J. Beringer, “(p, xn) Production Cross Sections: A benchmark Study for the Validation of Hadronic Physics Simulation at LHC”, CERN-LCGAPP-2003-18.

[2] A. Heikkinen, N. Stepanov, and J.P. Wellisch, “Bertini intra-nuclear cascade implementation in Geant4”, arXiv: nucl-th/0306008

[3] F. Rademakers, M. Goto, P. Canal, R. Brun, “ROOT Status and Future Developments”, arXiv: cs.SE/0306078

**Poster Session 2 - Board: 29 / 245**

## Production Experience of the Storage Resource Broker in the BaBar Experiment

**Authors:** A. Hasan<sup>1</sup>; J-Y. Nief<sup>2</sup>; W. Kroeger<sup>1</sup>

<sup>1</sup> SLAC

<sup>2</sup> ccin2p3

**Corresponding Author:** hasan@slac.stanford.edu

We describe the production experience gained from implementing and using exclusively the San Diego Super Computer Center developed Storage Resource Broker (SRB) to distribute the BaBar experiment's production event data stored in ROOT files from the experiment center at SLAC, California, USA to a Tier A computing center at ccinp3, Lyon France. In addition we outline how the system can be readily expanded to include more sites.

**Poster Session 3 - Board: 58 / 246**

## Volume-based representation of the magnetic field

**Authors:** N. AMAPANE<sup>1</sup>; T. Todorov<sup>2</sup>; V. Andreev<sup>3</sup>; V. Drollinger<sup>1</sup>; V. Karimaki<sup>4</sup>; V. Klyukhin<sup>5</sup>

<sup>1</sup> INFN

<sup>2</sup> CERN/IReS

<sup>3</sup> UCLA

<sup>4</sup> HELSINKI-HIP

<sup>5</sup> MSU/CERN

**Corresponding Author:** theodore.todorov@cern.ch

The simulation, reconstruction and analysis software access to the magnetic field has large impact both on CPU performance and on accuracy.

An approach based on a volume geometry is described. The volumes are constructed in such a way that their boundaries correspond to field discontinuities, which are due to changes in magnetic permeability of the materials. The field in each volume is contiguous.

The field in each volume is interpolated from a regular grid of values resulting from a TOSCA calculation. In case a parameterization is available for some volumes it is used instead of the grid interpolation.

Global access to the magnetic field values requires efficient search for the volume that contains a global point. An algorithm that exploits explicitly the layout and the symmetries of the detector is presented.

The main clients of the magnetic field, which are the simulation (geant4) and propagation of track parameters and errors in the reconstruction, can be made aware of the magnetic field volumes by connecting the per-volume magnetic field providers in their respective geometries to the corresponding volume in the magnetic field geometry. In this way the global volume search is by-passed and the access to the field is sped up significantly.

**Distributed Computing Services / 247**

## Middleware for the next generation Grid infrastructure

**Authors:** A. Aimar<sup>1</sup>; A. Di Meglio<sup>1</sup>; A. Edlund<sup>2</sup>; D. Groep<sup>3</sup>; E. Laure<sup>1</sup>; F. Hemmer<sup>1</sup>; F. Pacini<sup>4</sup>; F. Prelz<sup>5</sup>; L. Guy<sup>1</sup>; M. Barroso<sup>1</sup>; M. Livny<sup>6</sup>; M. Sgaravatto<sup>5</sup>; O. Mulmo<sup>7</sup>; P. Buncic<sup>1</sup>; P. Kunszt<sup>1</sup>; S. Beco<sup>4</sup>; S. Fisher<sup>8</sup>

<sup>1</sup> CERN

<sup>2</sup> KTH

<sup>3</sup> NIKHEF

<sup>4</sup> DATAMAT

<sup>5</sup> INFN

<sup>6</sup> UNIVERSITY OF WISCONSIN

<sup>7</sup> KTH, Sweden

<sup>8</sup> RAL

**Corresponding Author:** erwin.laure@cern.ch

The aim of the EGEE (Enabling Grids for E-Science in Europe) is to create a reliable and dependable European Grid infrastructure for e-Science. The objective of the Middleware Re-engineering and Integration Research Activity is to provide robust middleware components, deployable on several platforms and operating systems, corresponding to the core Grid services for resource access, data management, information collection, authentication & authorization, resource matchmaking and brokering, and monitoring and accounting.

For achieving this objective, we developed an architecture and design of the next generation Grid middleware leveraging experiences and existing components mainly from AliEn, EDG, and VDT. The

architecture follows the service breakdown developed by the LCG ARDA RTAG. Our goal is to do as little original development as possible but rather re-engineer and harden existing Grid services. The evolution of these middleware components towards a Service Oriented Architecture (SOA) adopting existing standards (and following emerging ones) as much as possible is another major goal of our activity.

A rapid prototyping approach has been adopted, providing a sequence of more sophisticated prototypes to the EGEE candidate applications coming from the LHC HEP experiments and the Biomedical field. The close feedback loop with applications via these prototypes is indispensable for achieving our ultimate goals of providing a reliable and dependable Grid infrastructure.

In this paper we will report on the architecture and design of the main Grid components and report on our experiences with early prototype systems.

**Core Software / 248**

## **LCIO persistency and data model for LC simulation and reconstruction**

**Authors:** F. Gaede<sup>1</sup>; H. Vogt<sup>2</sup>; N. Graf<sup>3</sup>; R. Cassell<sup>3</sup>; T. Behnke<sup>2</sup>; T. Johnson<sup>3</sup>

<sup>1</sup> *DESY IT*

<sup>2</sup> *DESY*

<sup>3</sup> *SLAC*

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

LCIO is a persistency framework and data model for the next linear collider. Its original implementation, as presented at CHEP 2003, was focused on simulation studies. Since then the data model has been extended to also incorporate prototype test beam data, reconstruction and analysis. The design of the interface has also been simplified. LCIO defines a common abstract user interface (API) in Java, C++ and Fortran in order to fulfill the needs of the global linear collider community. It is designed to be lightweight and flexible without introducing additional dependencies on other software packages.

User code is completely separated from the concrete persistency implementation. SIO, a simple binary format that supports data compression and pointer retrieval is the current choice. LCIO is implemented in such a way that it can also be used as the transient data model in any linear collider application, e.g. a modular reconstruction program can use the LCIO event class (LCEvent) as the container for the modules' input and output data. As LCIO offers a common API for three languages it is also possible to construct a multi-language reconstruction framework that would facilitate the integration of already existing algorithms.

A number of groups has already incorporated LCIO in their software frameworks and others plan to do so.

We present the design and implementation of LCIO, focusing on new developments and uses.

**Grid Security / 249****The GridSite authorization system**

**Authors:** A. McNab<sup>1</sup>; S. Kaushal<sup>1</sup>

<sup>1</sup> UNIVERSITY OF MANCHESTER

**Corresponding Author:** andrew.mcnab@man.ac.uk

We describe the GridSite authorization system, developed by GridPP and the EU DataGrid project for access control in High Energy Physics grid environments with distributed virtual organizations. This system provides a general toolkit of common functions, including the evaluation of access policies (in GACL or XACML), the manipulation of digital credentials (X.509, GSI Proxies or VOMS attribute certificates) and utility functions for protocols such as HTTP.

GridSite also provides a set of extensions to the Apache web server to permit it to function in a Grid security environment, including access control, fileserver / webservice management and a lightweight Virtual Organization service.

Using Apache as an example, we explain how Grid security can be added to an existing service using our toolkit. We then outline some of the other uses to which components have been put in the deployed Grids of GridPP, the EU DataGrid and the LHC Computing Grid.

**Poster Session 2 - Board: 26 / 250****Installing and Operating a Grid Infrastructure at DESY**

**Authors:** A. Campbell<sup>1</sup>; A. Gellrich<sup>1</sup>; B. Lewendel<sup>1</sup>; C. Wissing<sup>2</sup>; F. Brasolin<sup>3</sup>; J Ferrando<sup>4</sup>; K. Wrona<sup>1</sup>; M. Ernst<sup>1</sup>; M. Vorobiev<sup>5</sup>; M. de Riese<sup>1</sup>; P. Fuhrmann<sup>1</sup>; R. Mankel<sup>1</sup>; S. Padhi<sup>1</sup>; U. Ensslin<sup>1</sup>; V. Gulzow<sup>1</sup>

<sup>1</sup> DESY

<sup>2</sup> Dortmund University

<sup>3</sup> INFN, Bologna

<sup>4</sup> Glasgow University

<sup>5</sup> ITEP Moscow

**Corresponding Author:** andreas.gellrich@desy.de

DESY is one of the world-wide leading centers for research with particle accelerators and a center for research with synchrotron light.

The hadron-electron collider HERA houses four experiments which are taking data and will be operated until 2006 at least.

The computer center manages a data volumes of order 1 PB and is the home for around 1000 CPUs.

In 2003 DESY started to set up a Grid infrastructure on site. Monte Carlo production is the primer HEP application candidate for the Grid at DESY. The experiments have started major tests.

A first Grid Testbed was based on EDG 1.4. Some effort was taken to install the binary distribution of the middleware on SuSE based Linux systems at DESY.

With the first fixed LCG-2 release in spring 2004, the Grid Testbed2 was installed, which serves as the basis for all further DESY activities.

The contribution to CHEP2004 will start by briefly summarizing the status of the Grid activities at DESY in the context of EGEE and D-GRID, in which DESY takes a leading role.

In the following, we will discuss the integration of Grid components in the infrastructure of the DESY computer center.

This includes technical aspects of the operating system, such as SuSE versus RedHat Linux, the interaction with the mass storage system, and the management of Virtual Organizations.

We will finish with discussing installation and operation experiences of Grid middleware at DESY, also having in mind HEP and future synchrotron light experiments in the X-FEL era.

**Poster Session 1 - Board: 4 / 251**

## **Portable Gathering System for Monitoring and Online Calibration at Atlas**

**Authors:** ATLAS TDAQ HLT group<sup>None</sup>; P. Conde MUINO<sup>1</sup>

<sup>1</sup> CERN

**Corresponding Author:** patricia.conde.muino@cern.ch

During the runtime of any experiment, a central monitoring system that detects problems as soon as they appear has an essential role. In a large experiment, like Atlas, the online data acquisition system is distributed across the nodes of large farms, each of them running several processes that analyse a fraction of the events. In this architecture, it is necessary to have a central process that collects all the monitoring data from the different nodes, produces full statistics histograms and analyses them. In this paper we present the design of such a system, called the “gatherer”. It allows to collect any monitoring object, such as histograms, from the farm nodes, from any process in the DAQ, trigger and reconstruction chain. It also adds up the statistics, if required, and processes user defined algorithms in order to analyse the monitoring data. The results are sent to a centralized display, that shows the information online, and to the archiving system, triggering alarms in case of problems.

The innovation of our approach is that conceptually it abstracts the several communication protocols underneath, being able to talk with different processes using different protocols at the same time and, therefore, providing maximum flexibility. The software is easily adaptable to any trigger-DAQ system.

The first prototype of the gathering system has been implemented for Atlas and will be running during this year’s combined test beam. An evaluation of this first prototype will also be presented.

**Online Computing / 252**

## **New experiences with the ALICE High Level Trigger Data Transport Framework**

**Authors:** H. Tilsner<sup>1</sup>; T.M. Steinbeck<sup>1</sup>; V. Lindenstruth<sup>1</sup>

<sup>1</sup> KIRCHHOFF INSTITUTE OF PHYSICS, RUPRECHT-KARLS-UNIVERSITY HEIDELBERG, for the Alice Collaboration

**Corresponding Author:** timm.steinbeck@kip.uni-heidelberg.de

The Alice High Level Trigger (HLT) is foreseen to consist of a cluster of 400 to 500 dual SMP PCs at the start-up of the experiment. It's input data rate can be up to 25GB/s. This has to be reduced to at most 1.2 GB/s before the data is sent to DAQ through event selection, filtering, and data compression. For these processing purposes, the data is passed through the cluster in several stages and groups for successive merging until, at the last stage, fully processed complete events are available. For the transport of the data through the stages of the cluster, a software framework is being developed consisting of multiple components. These components can be connected via a common interface to form complex configurations that define the data flow in the cluster. For the framework, new benchmark results are available as well as experience from tests and data challenges run in Heidelberg. The framework is scheduled to be used during upcoming testbeam experiments.

**Distributed Computing Services / 253**

## Experiences with Data Indexing services supported by the NorduGrid middleware

**Authors:** A. Konstantinov<sup>1</sup>; A. Waananen<sup>2</sup>; B. Konya<sup>3</sup>; F. Ould-Saada<sup>4</sup>; J. Renner Hansen<sup>2</sup>; J.L. Nielsen<sup>2</sup>; M. Ellert<sup>5</sup>; O. Smirnova<sup>3</sup>; P. Eerola<sup>3</sup>; T. Ekelof<sup>6</sup>

<sup>1</sup> Vilnius University, Lithuania

<sup>2</sup> Niels Bohr Institute, Copenhagen, Denmark

<sup>3</sup> Lund University, Sweden

<sup>4</sup> University of Oslo, Norway

<sup>5</sup> CERN

<sup>6</sup> Uppsala University, Sweden

**Corresponding Authors:** oxana.smirnova@hep.lu.se, langgard@nbi.dk

The NorduGrid middleware, ARC, has integrated support for querying and registering to Data Indexing services such as the Globus Replica Catalog and Globus Replica Location Server. This support allows one to use these Data Indexing services for for example brokering during job-submission, automatic registration of files and many other things. This integrated support is complemented by a set of command-line tools for registering to and querying these Data Indexing services.

In this talk we will describe experiences with these Data Indexing services both from a daily work point of view and in production environments such as the Atlas Data-Challenges 1 and 2. We will describe the advantages of such Data Indexing services as well as their shortcomings. Finally we will present a proposal for an extended Data Indexing service which should deal with the shortcomings described. The development of such a Data Indexing service is being planned at the moment.

**Online Computing / 254**

## A Control Software for the ALICE High Level Trigger

**Authors:** H. Tilsner<sup>1</sup>; T.M. Steinbeck<sup>1</sup>; V. Lindenstruth<sup>1</sup>

<sup>1</sup> *KIRCHHOFF INSTITUTE OF PHYSICS, RUPRECHT-KARLS-UNIVERSITY HEIDELBERG, for the Alice Collaboration*

**Corresponding Author:** timm.steinbeck@kip.uni-heidelberg.de

The Alice High Level Trigger (HLT) cluster is foreseen to consist of 400 to 500 dual SMP PCs at the start-up of the experiment. The software running on these PCs will consist of components communicating via a defined interface, allowing flexible software configurations. During Alice's operation the HLT has to be continuously active to avoid detector dead time. To ensure that the several hundred software components, distributed throughout the cluster, operate and interact properly, a control software was written that is presented here. It was designed to run distributed over the cluster and to support control program hierarchies. Distributed operation avoids central performance bottlenecks and single-points-of-failures. The last point is of particular importance, as each of the commodity type PCs in the HLT cluster cannot be relied upon to operate continuously. Control hierarchies in turn are relevant for scalability over the required number of nodes. The software makes use of existing and widely used technologies: Configurations of programs to be controlled are saved in XML, while Python is used as a scripting language and to specify actions to execute. Interface libraries are used to access the controlled components, presenting a uniform interface to the control program. Using these mechanisms the control software remains generic and can be used for other purposes as well. It is being used for HLT data challenges in Heidelberg and is planned for use during upcoming beam tests.

**Event Processing / 255**

## Ion transport simulation using Geant4 hadronic physics

**Authors:** B. TRIEU<sup>1</sup>; G. Folger<sup>2</sup>; H.P. Wellisch<sup>2</sup>; I. CORNELIU<sup>3</sup>; P. TRUSCOTT<sup>4</sup>; T. Koi<sup>5</sup>

<sup>1</sup> *CERN, Geneva, Switzerland*

<sup>2</sup> *CERN*

<sup>3</sup> *INFN, Torino, Italy*

<sup>4</sup> *QinetiQ, Farnborough, UK*

<sup>5</sup> *SLAC*

**Corresponding Authors:** tkoi@slac.stanford.edu, hpw@free.fr

The transportation of ions in matter is subject of much interest in not only high-energy ion-ion collider experiments such as RHIC and LHC but also many other field of science, engineering and medical applications. Geant4 is a tool kit for simulation of passage of particles through matter and its OO designs makes it easy to extend its capability for ion transports. To simulate ions interaction, we had to develop two major functionalities to Geant4. One is cross section calculators and the other is final stage generators for ion-ion interactions. For cross sections calculator, several empirical cross section formulas for the total reaction cross section of ion-ion interactions were investigated. And for final stage generator, binary cascade and quark-gluon string model of Geant4 were improved so that ions reaction with matter can also be calculated. Having successfully developed both functionalities, Geant4 can be applied to ion transportation problems. In the presentation we will explain cross section and final stage generator in detail and show comparisons with experimental data.



Poster Session 3 - Board: 16 / 256

## The Binary Cascade

**Authors:** G. Folger<sup>1</sup>; H.P. Wellisch<sup>1</sup>

<sup>1</sup> CERN

**Corresponding Authors:** gunter.folger@cern.ch, hpw@free.fr

Geant4 is a toolkit for the simulation of the passage of particles through matter. Amongst its applications are hadronic calorimeters of LHC detectors and simulation of radiation environments. For these types of simulation, a good description of secondaries generated by inelastic interactions of primary nucleons and pions is particularly important.

The Geant4 Binary Cascade is a hybrid between a classical intra-nuclear cascade and a QMD model for the simulation of inelastic scattering of pions, protons and neutrons, and light ions of intermediate energies off nuclei. The nucleus is modeled by individual nucleons bound in the nuclear potential. Binary collisions of projectiles or projectile constituents and secondaries with single nucleons, resonance production, and decay are simulated according to measured, parametrised or calculated cross sections. Pauli's exclusion principle, i.e. blocking of interactions due to Fermi statistics, reduces the free cross section to an effective intra-nuclear cross section. Secondary particles are allowed to further interact with remaining nucleons.

We will describe the modeling, and give an overview of the components of the model, their object oriented design, and implementation.

Poster Session 3 - Board: 48 / 257

## Porting LCG Applications

**Authors:** I. Reguero<sup>1</sup>; J A. Lopez-Perez<sup>1</sup>

<sup>1</sup> CERN, IT DEPARTMENT

**Corresponding Author:** ignacio.reguero@cern.ch

Our goal is two fold. On one hand we wanted to address the interest of CMS users to have LCG Physics analysis environment on Solaris. On the other hand we wanted to assess the difficulty of porting code written in Linux without particular attention to portability to other Unix implementations. Our initial assumption was that the difficulty would be manageable even for a very small team. This is because the implicit respect by Linux of most Unix interfaces and standards such as the IEEE (PASC) 1003.1 1003.2 specifications.

We started with the LCG External software (<http://spi.web.cern.ch/spi/extsoft/platform.html>) in order to use it to build the LCG applications such as POOL and SEAL (<http://lcgapp.cern.ch/project/>).

We will discuss the main problems found with the system interfaces as well

as the advantages and disadvantages of using the GNU compilers and development environment versus the vendor provided ones.

**Poster Session 1** - Board: 19 / 258

## Control and state logging for the PHENIX DAQ System

**Author:** E. desmond<sup>1</sup>

<sup>1</sup> *BROOKHAVEN NATIONAL LAB*

**Corresponding Author:** desmond2@bnl.gov

The PHENIX DAQ system is managed by a control system responsible for the configuration and monitoring of the PHENIX detector hardware and readout software. At its core, the control system, called Runcontrol, is a process that manages the various components by way of a distributed architecture using CORBA. The control system, called Runcontrol, is a set of process that manages virtually all detector components through a distributed architecture base on CORBA. A key aspect of the distributed control system, the messaging system, is the ability to access critical detector state information, and deliver it to operators and applications of the control system. The goal of the system is to concentrate all output messages of the distributed processes, which would normally end up in log files or on a terminal, in a central place. The messages may originate from or be received by applications running on any of the multiple platforms which are in use including Linux, Windows, Solaris, and VxWorks. Listener applications allow the DAQ operators to get a comprehensive overview of all messages they are interested in, and also allows scripts or other programs to take automated action in response to certain messages. Messages are formatted to contain information about the source of the message, the message type, and its severity. Applications written to provide filtering of messages by the DAQ operators by type, severity and source will be presented. We will discuss the mechanism underlying this system, present examples of the use, and discuss performance and reliability issues.

**Event Processing** / 259

## Adding Kaons to the Bertini Cascade Model

**Authors:** A. Heikkinen<sup>1</sup>; D. Wright<sup>2</sup>

<sup>1</sup> *Helsinki Institute of Physics HIP*

<sup>2</sup> *SLAC*

A version of the Bertini cascade model for hadronic interactions is part of the Geant4 toolkit, and may be used to simulate pion-, proton-, and neutron-induced reactions in nuclei. It is typically valid for incident energies of 10 GeV and below, making it especially useful for the simulation of hadronic calorimeters. In order to generate the intra-nuclear cascade, the code depends on tabulations of exclusive channel cross section data, parameterized angular distributions and phase-space generation of multi-particle final states. To provide a more detailed treatment of hadronic calorimetry, and kaon interactions in general, this model is being extended to

include incident kaons up to an energy of 15 GeV. Exclusive channel cross sections, up to and including six-body final states, will be included for  $K^+$ ,  $K^-$ ,  $K^0$ ,  $K^0\bar{}$ ,  $\lambda$ ,  $\sigma^+$ ,  $\sigma^0$ ,  $\sigma^-$ ,  $\xi^0$  and  $\xi^-$ .  $K^+$ -nucleon and  $K^-$ -nucleon cross sections are taken from various cross section catalogs, while most of the cross sections for incident  $K^0$ ,  $K^0\bar{}$  and hyperons are estimated from isospin and strangeness considerations. Because there is little data for incident hyperon cross sections, use of the extended model will be restricted to incident  $K^+$ ,  $K^-$ ,  $K^0S$  and  $K^0L$ . Hyperon cross sections are included only to handle the secondary interactions of hyperons created in the intra-nuclear cascade.

## Event Processing / 261

### CHIPS based hadronization of quark-gluon strings

**Author:** M. Kosov<sup>1</sup>

<sup>1</sup> CERN

**Corresponding Author:** kosov@cern.ch

Quark-gluon strings are usually fragmented on the light cone in hadrons (PITHIA, JETSET) or in small hadronic clusters which decay in hadrons (HERWIG). In both cases the transverse momentum distribution is parameterized as an unknown function. In CHIPS the colliding hadrons stretch Pomeron ladders to each other and, when the Pomeron ladders meet in the rapidity space, they create Quasmons (hadronic clusters bigger than Amati-Veneziano clusters of HERWIG). The Quasmon size and the corresponding transverse momentum distributions are tuned by the Drell-Yan  $\mu^+\mu^-$  pairs. The final Quasmon fragmentation in CHIPS is tuned by the  $e^+e^-$  and proton-antiproton annihilation, which is already published.

## Core Software / 262

### Addressing the persistency patterns of the time evolving HEP data in the ATLAS/LCG MySQL Conditions Databases

**Authors:** A. Amorim<sup>1</sup>; A. Valassi<sup>2</sup>; D. Duellmann<sup>2</sup>; D. Klose<sup>1</sup>; L. Pedro<sup>1</sup>; N. Barros<sup>1</sup>; T. Franco<sup>1</sup>

<sup>1</sup> FACULTY OF SCIENCES OF THE UNIVERSITY OF LISBON

<sup>2</sup> CERN

**Corresponding Authors:** antonio.amorim@fisica.fc.ul.pt, luis.pedro@cern.ch

The size and complexity of the present HEP experiments represents an enormous effort in the persistency of data. These efforts imply a tremendous investment in the databases field not only for the event data but also for data that is needed to qualify this one - the Conditions Data.

In the present document we'll describe the strategy for addressing the Conditions data problem in the ATLAS experiment, focusing in the ConditionsDB MySQL for the ATLAS/LCG project. The need for a persistent engine for structured conditions data has motivated the studies for an relational backend that maps transient structured objects in the relational database persistent engine. This paper illustrate the proposal for the storage of Conditions data

in the LCG framework using it both to store only the Interval Of Validity (IOV) and a reference that represents the 'path' to an external persistent storage mechanism, and to archive the IOV and the data in relational tables mapping the customizable CondDBTable objects. This allow to take advantages of all the relational features and also to directly map between transient objects and tables in the database server.

The issue of distributed data storage and partitioning, is also analyzed in this paper, taking into account the different levels of indirection that are provided by the ConditionsDB MySQL implementation. These features represent a very important built in functionality in terms of scalability, data balance in a system that aims to be completely distributed and with a very high performance for hundreds of users.

**Poster Session 3 - Board: 26 / 263**

## **The WEB interface for the ATLAS/LCG MySQL Conditions Databases and performance constraints in the visualisation of extensive scientific/technical data**

**Authors:** A. Amorim<sup>1</sup>; A. Vaniachine<sup>2</sup>; D. Klose<sup>3</sup>; J. Castilho<sup>1</sup>; L. Pedro<sup>1</sup>; N. Barros<sup>1</sup>; T. Franco<sup>1</sup>

<sup>1</sup> *FACULTY OF SCIENCES OF THE UNIVERSITY OF LISBON*

<sup>2</sup> *Argonne National Laboratory*

<sup>3</sup> *Universidade de Lisboa, Portugal*

**Corresponding Author:** luis.pedro@cern.ch

A common LCG architecture for the Conditions Database for the time evolving data enables the possibility to separate the interval-of-validity (IOV) information from the conditions data payload. The two approaches can be beneficial in different cases and separation presents challenges for efficient knowledge discovery, navigation and data visualization. In our paper we describe the conditions data browser - CondDBrowser - a tool deployed in ATLAS for scientific analysis and visualization of this data.

A wide availability and access to the overall distributed conditions data repository was achieved through a seamless integration of the IOV and the payload data to the user a unifying web interface that hides the persistency storage details. Another user-friendly feature of the tool is a simplified querying language similar to QBE (Query by Example).

Our case study is based on the web interface developed for the ATLAS/LCG ConditionsDB. The interaction with other payload storage technologies, external to the ConditionsDB, will also be presented. In particular, the integration of the NOVA database technologies.

We will discuss how the information is gathered from the ConditionsDB and the corresponding extensions needed to enable data browsing in the external repositories, how it is organized, and what kind of operations (search and visualization) are allowed. We'll also present how this interface uses the C++ API extending it to a similar PHP interface, that can be used to browse data collected using any of the ConditionsDB implementations.

Performance constraints are also presented and will be discussed in detail.

Poster Session 3 - Board: 25 / 264

## Conditions Databases: the interfaces between the different ATLAS systems

**Authors:** A. Amorim<sup>1</sup>; A. Kazarov<sup>2</sup>; A. Perus<sup>3</sup>; A. Schaffer<sup>3</sup>; D. Burckhart-Chromek<sup>4</sup>; D. KLOSE<sup>5</sup>; D. Liko<sup>4</sup>; D. Malon<sup>6</sup>; E. Badescu<sup>7</sup>; I. Soloviev<sup>2</sup>; J. Cook<sup>4</sup>; J. Flammer<sup>4</sup>; L. Mapelli<sup>4</sup>; L. Pedro<sup>1</sup>; M. Caprini<sup>7</sup>; M. Dobson<sup>4</sup>; M. Mineev<sup>8</sup>; N. Barros<sup>1</sup>; R. Hawkings<sup>4</sup>; R. Jones<sup>9</sup>; S. Kolos<sup>2</sup>; T. Franco<sup>1</sup>; V. Kotov<sup>8</sup>; Y. Alexandrov<sup>10</sup>; Y. Ryabov<sup>11</sup>

<sup>1</sup> FACULTY OF SCIENCES OF THE UNIVERSITY OF LISBON

<sup>2</sup> Petersburg Nuclear Physics Institute (PNPI)

<sup>3</sup> Laboratoire de l'Accelérateur Lineaire (LAL) (IN2P3) (LAL) Université de Paris-Sud (ParisXI)

<sup>4</sup> CERN

<sup>5</sup> Universidade de Lisboa, Portugal

<sup>6</sup> ANL

<sup>7</sup> Institute of Atomic Physics National Institute for Physics and Nuclear Engineering (IFIN-HH)

<sup>8</sup> Joint Institute for Nuclear Research (JINR)

<sup>9</sup> CERN, Geneva, Switzerland

<sup>10</sup> P.N. Lebedev Institute of Physics (FIAN) Russian Academy of Sciences

<sup>11</sup> High Energy Physics Division (HEPD) Theory Division Petersburg Nuclear Physics Institute (PNPI)

**Corresponding Author:** luis.pedro@cern.ch

Conditions Databases are beginning to be widely used in the ATLAS experiment. Conditions data are time-varying data describing the state of the detector used to reconstruct the event data. This includes all sorts of slowly evolving data like detector alignment, calibration, monitoring and data from Detector Control System (DCS).

In this paper we'll present the interfaces between the ConditionsDB and the DCS, Trigger and Data Acquisition (TDAQ) and offline control framework (Athena).

In the DCS case, a PVSS API Manager was developed based on the C++ interface for the ConditionsDB. The Manager links to a selection of datapoints and stores any value change in the ConditionsDB. The structure associated to each datapoint is mapped to a table that reflects this structure and is stored in the database.

The ConditionsDB Interface to the TDAQ (CDI) is a service provided by the Online Software that acts as an intermediary between TDAQ producers and consumers of conditions data. CDI provides the pathway to the ConditionsDB information regarding the present or past condition of the detector and trigger system as well as all the operational and monitoring data. It will provide the link between the Information Service (IS) and the ConditionsDB

Conditions database integration into the ATLAS Athena framework is also described, including connections to Athena's transient interval-of-validity management, conversion services to support conditions data I/O into Athena transient stores, and mechanisms by which the conditions databases may be used for timestamp-mediated access to data stored in other technologies such as NOVA and POOL.

Poster Session 3 - Board: 51 / 265

## New Applications of PAX in Physics Analyses at Hadron Colliders

**Authors:** A. Schmidt<sup>1</sup>; C. Jung<sup>1</sup>; C. Weiser<sup>1</sup>; D. Hirschbuehl<sup>1</sup>; G. Quast<sup>1</sup>; J. Rehn<sup>1</sup>; K. Rabbertz<sup>1</sup>; M. Erdmann<sup>1</sup>; M. Kirsch<sup>1</sup>; P. Schemitz<sup>1</sup>; S. Kappler<sup>1</sup>; S. Schalla<sup>1</sup>; T. Walter<sup>1</sup>; U. FELZMANN<sup>2</sup>

<sup>1</sup> *Institut fuer Experimentelle Kernphysik, Karlsruhe University, Germany*

<sup>2</sup> *University of Karlsruhe, Germany*

**Corresponding Author:** alexander.schmidt@cern.ch

At CHEP03 we introduced “Physics Analysis eXpert” (PAX), a C++ toolkit for advanced physics analyses in High Energy Physics (HEP) experiments. PAX introduces a new level of abstraction beyond detector reconstruction and provides a general, persistent container model for HEP events. Physics objects like fourvectors, vertices and collisions can easily be stored, accessed and manipulated. Bookkeeping of relations between these objects (like decay trees, vertex and collision separation, including deep copies etc.) is fully provided by a “relation manager”. Event container and associated objects represent a uniform interface for algorithms and facilitate the parallel development and evaluation of different physics interpretations of individual events. So called “analysis factories”, which actively identify and distinguish different physics processes, can easily be constructed with the PAX toolkit.

PAX has been officially released to the experiments CDF (Tevatron) and CMS (LHC) during the last year. It is being explored by a growing user community and applied in various complex physics analyses. We report about the successful application in studies of  $t\bar{t}$  production at the Tevatron and Higgs searches in the channel  $t\bar{t}H$  at the LHC.

**Online Computing / 266**

## Testbed Management for the ATLAS TDAQ

**Author:** M. ZUREK<sup>1</sup>

<sup>1</sup> *CERN, IFJ KRAKOW*

**Corresponding Author:** marian.zurek@cern.ch

The talk presents the experience gathered during the testbed administration (~100 PC and 15+ switches) for the ATLAS Experiment at CERN.

It covers the techniques used to resolve the HW/SW conflicts, network related problems, automatic installation and configuration of the cluster nodes as well as system/service monitoring in the heterogeneous dynamically changing cluster environment.

Techniques range from manual actions to the fully automated procedures based on tools like Kickstart, SystemImager, Nagios, MRTG and Spectrum. Booting diskless nodes using EtherBoot, PXEboot is also investigated as a possible technique of managing Atlas Production Farms.

Kernel customization techniques (building, deploying, distribution policy) allow users to freely choose proffered kernel flavors without sysadmin intervention. At the same time administrator retains full control over entire testbed.

The overall experience has shown that the proper use of the open-source tools addresses very well the needs of the ATLAS Trigger DAQ community. This approach may also be interesting for addressing certain aspects of GRID Farm Management.

Poster Session 2 - Board: 50 / 267

## GoToGrid - A Web-Oriented Tool in Support to Sites for LCG Installations

**Authors:** A. Retico<sup>1</sup>; A. Usai<sup>1</sup>; G. Diez-Andino Sancho<sup>1</sup>; I. tkachev<sup>1</sup>; M. Schulz<sup>1</sup>; O. Keeble<sup>1</sup>; P. Nyczyk<sup>1</sup>

<sup>1</sup> CERN

**Corresponding Author:** antonio.retico@cern.ch

The installation and configuration of LCG middleware, as it is currently being done, is complex and delicate.

An “accurate” configuration of all the services of LCG middleware requires a deep knowledge of the inside dynamics and hundreds of parameters to be dealt with. On the other hand, the number of parameters and flags that are strictly needed in order to run a working “default” configuration of the middleware is relatively small, due to the fact that the values to be set mainly deal with environment configuration and with a limited set of possible operation scenarios.

This “default” configuration appears to be the most suitable for sites joining LCG for the first time.

The GoToGrid system is aimed to support Site Administrators to easily perform such a configuration.

G2G combines the gathering of configuration information, provided by sites, with the dynamic adaptive creation of customized documentation and installation tools.

By using a web interface and being requested only for the relevant configuration information, site Administrators will be able to design the desired configuration of their own LCG site.

Site configuration data is collected and stored in a well defined format liable to be used as the interface to different configuration management tools.

Core Software / 268

## Developments of Mathematical Software Libraries for the LHC experiments

**Author:** L. Moneta<sup>1</sup>

**Co-authors:** A Zsenei<sup>1</sup>; F. James<sup>1</sup>; M. Hatlo<sup>1</sup>; M. Winkler<sup>1</sup>; P. Mato Vila<sup>1</sup>

<sup>1</sup> CERN

**Corresponding Author:** lorenzo.moneta@cern.ch

The main objective of the MathLib project is to give expertise and support to the LHC experiments on mathematical and statistical computational methods. The aim is to provide a coherent set of mathematical libraries. Users of this set of libraries are developers of experiment reconstruction and simulation software, of analysis tools frameworks, such as ROOT, and physicists performing data analysis.

After having performed a detailed evaluation of the existing functionality present in GSL, a general purpose mathematical library, and in more HEP specific libraries such as CLHEP, CERNLIB and ROOT, a new object oriented library has been started to be developed. The new library incorporates or uses most of the functions and algorithms of the already existing libraries. Examples of these functions and algorithms are mathematical special functions, linear algebra, minimization and any other required numerical algorithms. Wrappers to these are written in C++ and integrated in a coherent object oriented framework. Interfaces to the Python interactive environment are as well provided.

An overview of the project activities will be presented, describing in detail the current functionality of the library and its design. Furthermore, the object oriented implementation of Minuit, a fitting and minimization framework, will be covered in the presentation.

**Poster Session 3 - Board: 11 / 269**

## ”RecPack”, a general reconstruction tool-kit

**Authors:** A. CERVERA VILLANUEVA<sup>1</sup>; J A. HERNANDO MORATA<sup>2</sup>; J J. GOMEZ CADENAS<sup>3</sup>

<sup>1</sup> *University of Geneva*

<sup>2</sup> *CERN and University of Santiago de Compostela*

<sup>3</sup> *University of Valencia*

**Corresponding Author:** anselmo.cervera.villanueva@cern.ch

We have developed a c++ software package, called “RecPack”, which allows the reconstruction of dynamic trajectories in any experimental setup. The basic utility of the package is the fitting of trajectories in the presence of random and systematic perturbations to the system (multiple scattering, energy loss, inhomogeneous magnetic fields, etc) via a Kalman Filter fit. It also includes an analytical navigator which allows: extrapolation of the trajectory parameters (and their covariance matrix) to any surface, path length computations, matching functions (trajectory-trajectory, trajectory-measurement, etc) and much more. The RecPack tool-kit also includes the algorithms for vertex fitting via Kalman Filter, and the necessary tools for easily coding pattern recognition algorithms. In summary, “RecPack” provides all the necessary tools and algorithms that are common to any reconstruction program. In addition, a toy simulator is provided. This is very useful to debug new reconstruction algorithms and also to perform simple physics analysis. The modularity of the package allows extensions in any direction: new propagation models, measurement types, volume and surface types, fitting algorithms, etc.

I

**Poster Session 3 - Board: 15 / 270**

## A Pattern-based Continuous Integration Framework for Distributed EGEE Grid Middleware Development

**Author:** A. Di Meglio<sup>1</sup>

**Co-authors:** E. Ronchieri<sup>2</sup>; J. Flammer<sup>1</sup>; M. Zurek<sup>1</sup>; R. Harakaly<sup>1</sup>

<sup>1</sup> *CERN*

<sup>2</sup> *INFN*

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

Software Configuration Management (SCM) Patterns and the Continuous Integration method are recent and powerful techniques to enforce a common software engineering process across large, heterogeneous, rapidly changing development projects where a rapid release lifecycle is required. In particular the Continuous Integration method allows tracking and addressing problems in the software



components integration as early as possible in the release cycle. Since new incremental code builds are done several times per day, only small amounts of new code is built and integrated at relatively short intervals. Developers are immediately notified of arising problems and integrators can pinpoint configuration and build problems to the level of single files within any given software component. This paper presents the implementation and the initial results of the application of such techniques in the SCM and Integration of the EGEE Grid Middleware software. The software is based on a Service Oriented Architecture model where services are developed in different programming languages by development groups in several European locations under stringent quality requirements. A number of basic SCM patterns, such as the Workspace, the Active Line, the Repository, are introduced and the Continuous Integration tools used in the project are presented with a discussion of the advantages and disadvantages of using the method.

Poster Session 2 - Board: 3 / 271

## Job Interactivity using a Steering Service in an Interactive Grid Analysis Environment

**Authors:** A. Ali<sup>1</sup>; A. Anjum<sup>1</sup>; C. Steenberg<sup>2</sup>; F. van Lingen<sup>2</sup>; H. Newman<sup>3</sup>; I. Willers<sup>4</sup>; J. Bunn<sup>5</sup>; M. A. Zafar<sup>1</sup>; M. Thomas<sup>2</sup>; R. Cavanaugh<sup>6</sup>; R. McClatchey<sup>7</sup>

<sup>1</sup> NIIT

<sup>2</sup> CALIFORNIA INSTITUTE OF TECHNOLOGY

<sup>3</sup> Caltech

<sup>4</sup> CERN

<sup>5</sup> CALTECH

<sup>6</sup> University of Florida

<sup>7</sup> UWE

**Corresponding Author:** [ashiq.anjum@niit.edu.pk](mailto:ashiq.anjum@niit.edu.pk)

In the context of Interactive Grid-Enabled Analysis Environment (GAE), physicists desire bi-directional interaction with the job they submitted. In one direction, monitoring information about the job and hence a “progress bar” should be provided to them. On other direction, physicist should be able to control their jobs. Before submission, they may direct the job to some specified resource or computing element. Before execution, its parameter may be changed or it may be moved to another location. During execution, its intermediate results should be fetched or it may be moved to another location. Also, physicists should be able to kill, restart, hold and resume their jobs.

Interactive job execution requires that at each step, the user must make choices between alternative application components, files, or locations. So a dead end may be reached where no solution can be found, which would require backtracking to undo some previous choice. Another desire is reliable and optimal execution of the job. Grid should take some decisions regarding the job execution to help in reliable and optimal execution of the job. Reliability can be achieved using the job recovery mechanism. When a job on grid fails, the recovery mechanism should resubmit the job on either the same resource or on different resource. Check-pointing the job will make resource utilization low when recovering the job from failure.

In this paper the architecture and design of an autonomous grid service is described that fulfills the above stated requirements for interactivity in Grid-enabled data analysis.

Poster Session 2 - Board: 1 / 272

## Job Monitoring in Interactive Grid Analysis Environment

**Authors:** A. Ali<sup>1</sup>; A. Anjum<sup>1</sup>; C. Steenberg<sup>2</sup>; F. van Lingen<sup>2</sup>; H. Newman<sup>3</sup>; I. Willers<sup>4</sup>; J. Bunn<sup>3</sup>; M. Thomas<sup>3</sup>; R. Cavanaugh<sup>5</sup>; R. Mcclatchey<sup>6</sup>; W. Rehman<sup>1</sup>

<sup>1</sup> NIIT

<sup>2</sup> CALIFORNIA INSTITUTE OF TECHNOLOGY

<sup>3</sup> Caltech

<sup>4</sup> CERN

<sup>5</sup> University of South Florida

<sup>6</sup> UWE

**Corresponding Author:** ashiq.anjum@niit.edu.pk

Grid is emerging as a great computational resource but its dynamic behaviour makes the Grid environment unpredictable. System failure or network failure can occur or the system performance can degrade. So once the job has been submitted monitoring becomes very essential for user to ensure that the job is completed in an efficient way. In current environments once user submits a job he loses direct control over the job, system behaves like a batch system, user submits the job and gets the result back. Only information a user can obtain about a job is whether it is scheduled, running, cancelled or finished. This information is enough from the Grid management point of view but not from the point of view of a user. User wants interactive environment in which he can check the progress of the job, obtain intermediate results, terminate the job based on the progress of job or intermediate results, steer the job other nodes to achieve better performance and check the resources consumed by the job. So a mechanism is needed that can provide user with secure access to information about different attributes of a job. In this paper we describe a monitoring service, a java based web service that will provide secure access to different attributes of a job once a job has been submitted to Interactive Grid Analysis Environment.

Poster Session 2 - Board: 2 / 273

## Predicting Resource Requirements of a Job Submission

**Authors:** A. Ali<sup>1</sup>; A. Anjum<sup>1</sup>; A. Mehmood<sup>1</sup>; H. Newman<sup>2</sup>; I. Willers<sup>2</sup>; J. Bunn<sup>2</sup>; R. Meclatchey<sup>3</sup>

<sup>1</sup> NIIT

<sup>2</sup> Caltech

<sup>3</sup> UWE

**Corresponding Author:** ashiq.anjum@niit.edu.pk

Grid computing provides key infrastructure for distributed problem solving in dynamic virtual organizations. However, Grids are still the domain of a few highly trained programmers with expertise in networking, high-performance computing, and operating systems.

One of the big issues in the full-scale usage of a grid is the matching of the

resource requirements of a job submission to available resources. In order for resource brokers/job schedulers to ensure efficient use of grid resources, an initial estimate of the likely resource usage of a submission must be made. In the context of the Grid Enabled Analysis Environment (GAE), physicists want the ability to discover, acquire, and reliably manage computational resources dynamically, in the course of their everyday activities. They do not want to be bothered with the location of these resources, the mechanisms that are required to use them, keeping track of the status of computational tasks operating on these resources, or with reacting to failure. They do care about how long their tasks are likely to run and how much these tasks will cost.

So the grid scheduler must have the capability to estimate before job submission, how much time and resources the job will consume on execution site. Our proposed module, Prediction engine will be part of scheduler and it will provide estimates of resource use along with the duration of use. This will enable scheduler to choose the optimum site for job execution.

This paper presents the survey of existing grid schedulers and then based on this survey states the need for resource usage estimation. Also the architecture and design of “grid prediction engine” that predicts the resource requirements of a job submission is discussed.

**Distributed Computing Services / 274**

## **Design and Implementation of a Notification Model for Grid Monitoring Events**

**Authors:** G. Rubini<sup>1</sup>; G. Tortone<sup>2</sup>; N. De Bortoli<sup>2</sup>; S. Andreozzi<sup>1</sup>; S. Fantinel<sup>3</sup>

<sup>1</sup> *INFN-CNAF BOLOGNA (ITALY)*

<sup>2</sup> *INFN - NAPLES (ITALY)*

<sup>3</sup> *INFN - LEGNARO (ITALY)*

**Corresponding Author:** [nataschia.debortoli@na.infn.it](mailto:nataschia.debortoli@na.infn.it)

GridICE is a monitoring service for the Grid, it measures significant Grid related resources parameters in order to analyze usage, behavior and performance of the Grid and/or to detect and notify fault situations, contract violations, and user-defined events. In its first implementation, the notification service relies on a simple model based on a pre-defined set of events.

The growing interest for more flexible and scalable notification capabilities from several LHC experiments has led us to study a more suitable solution satisfying their requirements. With this paper we present both model and design of a notification service which main functionalities are: filtering, transformation, and routing of data. It basically collects a large number of incoming streams of data items from monitored resources (events), filters them according to user profiles or queries describing users information preferences (subscriptions) and finally, after a customization of matched data items, notifies users whose interests are satisfied (event consumers).

Our proposal significantly improves the notification capabilities in current Grid systems by providing flexible means for specifying both topic and content based subscriptions, moreover it provides an efficient matchmaking engine. The new component has been developed and integrated in the GridICE service based on users expressed interests.

**Event Processing / 275**

## Offline Software for the ATLAS Combined Test Beam

**Authors:** A. FARILLA<sup>1</sup>; M. GALLAS<sup>2</sup>

<sup>1</sup> *INFN. ROMA3*

<sup>2</sup> *CERN*

**Corresponding Author:** ada.farilla@roma3.infn.it

A full slice of the barrel detector of the ATLAS experiment at the LHC is being tested this year with beams of pions, muons, electrons and photons in the energy range 1-300 GeV in the H8 area of the CERN SPS. It is a challenging exercise since, for the first time, the complete software suite developed for the full ATLAS experiment has been extended for use with real detector data, including simulation, reconstruction, online and offline conditions databases, detector and physics monitoring, and distributed analysis. Important integration issues like combined simulation, combined reconstruction, connection with the online services and management of many different types of conditions data are being addressed for the first time, with the goal of both achieving experience on such integration aspects and of performing physics studies requiring the combined analysis of simultaneous data coming from different subdetectors. It is a unique opportunity to test, with real data, new algorithms for pattern recognition, particle tracking and identification and High Level Trigger strategies. A relevant outcome of this combined test beam will be a detailed comparison of Monte Carlo - based on Geant4 - with real data. In the talk the main components of the software suite are described, together with some preliminary results obtained both with simulated and real data.

**Poster Session 2 - Board: 27 / 276**

## Distributed Testing Infrastructure and Processes for the EGEE Grid Middleware

**Authors:** D. Bosio<sup>1</sup>; D. Collados<sup>1</sup>; D. Groep<sup>2</sup>; D. Salomoni<sup>2</sup>; J. Templon<sup>2</sup>; L. Guy<sup>1</sup>; M. Reale<sup>1</sup>; M. Theile<sup>1</sup>; S. Traylen<sup>3</sup>

<sup>1</sup> *CERN*

<sup>2</sup> *NIKHEF*

<sup>3</sup> *RAL*

**Corresponding Author:** leanne.guy@cern.ch

Extensive and thorough testing of the EGEE middleware is essential to ensure that a production quality Grid can be deployed on a large scale as well as across the broad range of heterogeneous resources that make up the hundreds of Grid computing centres both in Europe and worldwide.

Testing of the EGEE middleware encompasses the tasks of both verification and validation. In addition we test the integrated middleware for stability, platform independence, stress resilience, scalability and performance.

The EGEE testing infrastructure is distributed across three major EGEE grid centres in three countries: CERN, NIKHEF and RAL. As much as is possible the testing procedures are automated and integrated with the EGEE build system. This allows for continuous testing together with the incremental daily code builds, fast and early feedback to developers of bug, and for the easy inclusion of regression tests.

This paper will report on the initial results of the testing procedures, frameworks and automation techniques adopted by the EGEE project, the advantages and disadvantages of test automation and the issues involved in testing a complex distributed middleware system in a distributed environment.

**Poster Session 3 - Board: 46 / 277**

## Managing third-party software for the LCG

**Author:** E. Poinson<sup>1</sup>

**Co-authors:** Andreas Pfeiffer<sup>1</sup>; Ioannis Papadopoulos<sup>1</sup>; Pere Mato Vila<sup>1</sup>; Saul Youssef<sup>2</sup>; Shaun Ashby<sup>1</sup>; Yannick Patois<sup>3</sup>

<sup>1</sup> *CERN*

<sup>2</sup> *BU*

<sup>3</sup> *EGEE*

**Corresponding Author:** eric.poinson@cern.ch

The External Software Service of the LCG SPI project provides open source and public domain packages required by the LCG projects and experiments. Presently, more than 50 libraries and tools are provided for a set of platforms decided by the architect forum. All packages are installed following a standard procedure and are documented on the web.

A set of scripts has been developed to ease new installations.

In addition to providing these packages, a software configuration management “toolbox” is provided, containing a coherent set of package-version combinations for each release of a project, as well as a distribution script which manages the dependencies of the LCG projects such that users can easily download and install a release of a project including its depended packages. Emphasis here has been put on ease of use for the end-user.

**Distributed Computing Services / 278**

## Evolution of LCG-2 Data Management

**Authors:** J-P. Baud<sup>1</sup>; J. Casey<sup>1</sup>

<sup>1</sup> *CERN*

**Corresponding Author:** james.casey@cern.ch

LCG-2 is the collective name for the set of middleware released for use on the LHC Computing Grid in December 2003. This middleware,

based on LCG-1, had already several improvements in the Data Management area. These included the introduction of the Grid File Access Library(GFAL), a POSIX-like I/O Interface, along with MSS integration via the Storage Resource Manager(SRM)interface.

LCG-2 was used in the Spring 2004 data challenges by all four LHC experiments. This produced the first useful feedback on scalability and functionality problems in the middleware, especially with regards to data management.

One of the key goals for the Data Challenges in 2004 is to show that the LCG can handle the data for the LHC, even if the computing model is still quite simple. In light of the feedback from the data challenges, and in conjunction with the LHC experiments, a strategy for the improvements required in the data management area was developed. The aim of these improvements was to allow both easier interaction and better performance from the experiment frameworks and other middleware such as POOL.

In this talk, we will first introduce the design of the current data management solution in LCG-2. We will cover the problems and issues highlighted by the data challenges, as well as the strategy for the required improvements to allow LCG-2 to handle effectively data management at LCG volumes. In particular, we will highlight the new APIs provided, and the integration of GFAL and the EDG Replica Manager functionality with ROOT.

## Event Processing / 279

### The GeoModel Toolkit for Detector Description

**Authors:** J. Boudreau<sup>1</sup>; V. Tsulaia<sup>1</sup>

<sup>1</sup> UNIVERSITY OF PITTSBURGH

**Corresponding Author:** vakhtang.tsulaia@cern.ch

The GeoModel toolkit is a library of geometrical primitives that can be used to describe detector geometries. The toolkit is designed as a data layer, and especially optimized in order to be able to describe large and complex detector systems with minimum memory consumption. Some of the techniques used to minimize the memory consumption are: shared instancing with reference counting, compressed representations of Euclidean transformations, special nodes which encode the naming of volumes without storing name-strings, and, especially, parameterization through embedded symbolic expressions of transformation fields. A faithful representation of a GeoModel description can be transferred to Geant4, and, we predict, to other engines that simulate the interaction of particles with matter. GeoModel comes with native capabilities for geometry clash detection and for material integration. Its only external dependencies are upon CLHEP.

This talk describes this toolkit for the first time in a public forum.

## Event Processing / 280

### Update On the Status of the FLUKA Monte Carlo Transport Code

**Authors:** A. Empl<sup>1</sup>; A. Fasso<sup>2</sup>; A. Ferrari<sup>3</sup>; A. Ottolenghi<sup>4</sup>; D. Scannicchio<sup>4</sup>; E. Gadioli<sup>5</sup>; F. Ballarini<sup>4</sup>; F. Cerutti<sup>5</sup>; G. Battistoni<sup>5</sup>; G. SMIRNOV<sup>6</sup>; J. Ranft<sup>7</sup>; K. LEE<sup>6</sup>; K. TSOULOU<sup>8</sup>; L. Pinsky<sup>9</sup>; M V. Garzelli<sup>10</sup>; M. CAMPANELLA<sup>11</sup>; M. CARBONI<sup>12</sup>; M. PELLICCIONI<sup>12</sup>; N. Zapp<sup>1</sup>; P. Sala<sup>13</sup>; S. MURARO<sup>11</sup>; S. Roesler<sup>14</sup>; T. RANCATI<sup>11</sup>; V. ANDERSON<sup>6</sup>; V. PARINI<sup>15</sup>; V. Vlachoudis<sup>14</sup>

<sup>1</sup> *University of Houston*

<sup>2</sup> *SLAC*

<sup>3</sup> *CERN & Universita' and INFN Milano*

<sup>4</sup> *Universita' and INFN Pavia*

<sup>5</sup> *Universita' and INFN Milano*

<sup>6</sup> *University of Houston, USA*

<sup>7</sup> *University of Leipzig*

<sup>8</sup> *CERN, Geneva, Switzerland*

<sup>9</sup> *UNIVERSITY OF HOUSTON*

<sup>10</sup> *Univeristy and INFN, Milano, Italy*

<sup>11</sup> *INFN Milano, Italy*

<sup>12</sup> *INFN Frascati, Italy*

<sup>13</sup> *ETH Zurich & INFN Milan*

<sup>14</sup> *CERN*

<sup>15</sup> *INFN Pavia, Italy*

**Corresponding Author:** pinsky@uh.edu

The FLUKA Monte Carlo transport code is a well-known simulation tool in High Energy Physics. FLUKA is a dynamic tool in the sense that it is being continually updated and improved by the authors. Here we review the progresses achieved in the last year on the physics models. From the point of view of hadronic physics, most of the effort is still in the field of nucleus–nucleus interactions. The currently available version of FLUKA already includes the internal capability to simulate inelastic nuclear interactions beginning with lab kinetic energies of 100 MeV/A up to the highest accessible energies by means of the DPMJET-II.5 event generator to handle the interactions for >5 GeV/A and rQMD for energies below that. The new developments concern, at high energy, the embedding of the DPMJET-III generator, which represent a major change with respect to the DPMJET-II structure. This will also allow to achieve a better consistency between the nucleus-nucleus section with the original FLUKA model for hadron-nucleus collisions. Work is also in progress to implement a third event generator model based on the Master Boltzmann Equation approach, in order to extend the energy capability from 100 MeV/A down to the threshold for these reactions. In addition to these extended physics capabilities, structural changes to the programs input and scoring capabilities are continually being upgraded. In particular we want to mention the upgrades in the geometry packages, now capable of reaching higher levels of abstraction. Work is also proceeding to provide direct import into ROOT of the FLUKA output files for analysis and to deploy a user-friendly GUI input interface.

**Online Computing / 281**

## Migrating PHENIX databases from object to relational model

**Authors:** D. Morrison<sup>1</sup>; I. Sourikova<sup>1</sup>

<sup>1</sup> *BROOKHAVEN NATIONAL LABORATORY*

**Corresponding Author:** irina@bnl.gov

To benefit from substantial advancements in Open Source database technology and ease deployment and development concerns with

Objectivity/DB, the Phenix experiment at RHIC is migrating its principal databases from Objectivity to a relational database management system (RDBMS). The challenge of designing a relational DB schema to store a wide variety of calibration classes was solved by using ROOT I/O and storing each calibration object opaquely as a BLOB ( Binary Large Object ). Calibration metadata is stored as built-in types to allow fast index-based database search. To avoid a database back-end dependency the application was made ODBC-compliant (Open DataBase Connectivity is a standard database interface). An existent well-designed calibration DB API allowed users to be shielded from the underlying database technology change. Design choices and experience with transferring a large amount of Objectivity data into relational DB will be presented.

**Poster Session 3 - Board: 6 / 282**

## INDICO - the software behind CHEP 2004

**Authors:** H. Sanchez<sup>1</sup>; J B. Gonzalez Lopez<sup>1</sup>; J-Y. Le Meur<sup>1</sup>; T. Baron<sup>1</sup>; V Turney<sup>1</sup>

<sup>1</sup> CERN

**Corresponding Authors:** thomas.baron@cern.ch, jean-yves.le.meur@cern.ch

CHEP 2004 conference is using the Integrated Digital Conferencing product to manage part of its web site and processes to run the conference.

This software has been built in the framework of InDiCo European Project. It is designed to be generic and extensible with the goal of providing help for single seminars as well as large conferences management. Partly developed at CERN within the Document Server (CDS) team, it focuses on supporting future events in HEP domain and it will be distributed with an open source license.

The presentation will explain the main application features before going into the details of its object oriented development. It will be demoed how the InDiCo can be used as a platform for scheduling events in a large institution like CERN and how it can give to conference chairperson a solid basis to set up, run and archive content of meetings.

**Poster Session 3 - Board: 52 / 283**

## The Description of the Atlas Detector

**Authors:** A. NAIRZ<sup>1</sup>; A. VANIACHINE<sup>2</sup>; A. ZALITE<sup>3</sup>; B. SELIGMAN<sup>4</sup>; C. BOURDARIOS<sup>5</sup>; D. COSTANZO<sup>6</sup>; D. POMAREDE<sup>7</sup>; G. GORFINE<sup>8</sup>; G. UNAL<sup>5</sup>; I. TRIGGER<sup>1</sup>; J. BOUDREAU<sup>9</sup>; J. TOTH<sup>10</sup>; K.A. ASSAMAGAN<sup>11</sup>; M. LELCHUK<sup>4</sup>; P. Strizenc<sup>12</sup>; S. BARANOV<sup>13</sup>; S. Spagnolo<sup>14</sup>; T. KITTELMANN<sup>15</sup>; V. TSULAIA<sup>9</sup>; Y. HASEGAWA<sup>16</sup>

<sup>1</sup> CERN, Geneva, Switzerland

<sup>2</sup> ANL, USA

<sup>3</sup> Petersburg Nuclear Physics Institute, Russia

<sup>4</sup> Nevis Laboratory, USA

<sup>5</sup> LAL-Orsay, France

<sup>6</sup> LBNL, USA



<sup>7</sup> CEA/DAPNIA, Saclay, France

<sup>8</sup> NIKHEF, The Netherlands

<sup>9</sup> Pittsburgh University, USA

<sup>10</sup> KFKI-RMKI, Hungary

<sup>11</sup> BNL, USA

<sup>12</sup> IEP SAS Slovakia

<sup>13</sup> Albert-Ludwigs University, Germany

<sup>14</sup> DIP. FISICA E INFN, LECCE

<sup>15</sup> Niels Bohr Institute, Denmark

<sup>16</sup> Shinshu University, Japan

**Corresponding Author:** stefania.spagnolo@le.infn.it

The ATLAS Detector consists of several major subsystems: an inner detector composed of pixels, microstrip detectors and a transition radiation tracker; electromagnetic and hadronic calorimetry, and a muon spectrometer. Over the last year, these systems have been described in terms of a set of geometrical primitives known as GeoModel. Software components for detector description interpret structured data from a relational database and build from that a complete description of the detector. This description is now used in the Geant-4 based simulation program and also for reconstruction. Detector-specific services that are not handled in a generic way (e.g strip pitches and calorimetric tower boundaries) are added as an additional layer which is synched to the raw geometry. Detector misalignments may also be fed through the model to both simulation and reconstruction. Visualization of the detector geometry is accomplished through Open Inventor and its HEPVis extensions. The ATLAS geometry system in the last year has undergone extensive visual debugging, and experience with the new system has been gained not only through the data challenge but also through the combined test beam. This talk gives an overview of the ATLAS detector description and discusses operational experience with the system in the data challenges and combined test beam.

**Wide Area Networking / 284**

## On-demand Layer VPN Support for Grid Applications

**Authors:** E. Ronchieri<sup>1</sup>; S. Androzzi<sup>1</sup>; T. Ferrari<sup>1</sup>

<sup>1</sup> INFN CNAF

**Corresponding Authors:** elisabetta.ronchieri@cnaif.infn.it, tiziana.ferrari@cnaif.infn.it

The problem of finding the best match between jobs and computing resources is critical for an efficient work load distribution in Grids. Very often jobs are preferably run on the Computing Elements (CEs) that can retrieve a copy of the input files from a local Storage Element (SE). This requires that multiple file copies are generated and managed by a data replication system.

We propose the use of scheduled on-demand Layer 2 Virtual Private Networks (L2 VPNs) for an alternative data access model based on the possibility to connect to the same virtual LAN both CEs and SEs from remote Grid domains. The L2 VPN members are “close” to each other. In this way a CE can be selected by a Resource Broker without requiring the presence of a local file replica. This simplifies the data management and allows a more efficient use of the network resources on the links connecting the Grid to its main data sources.

In this paper we detail how L2 VPNs are dynamically provisioned through the Grid Network Agreement Service. We propose a hierarchical network resource abstraction, the Path, and we show how it can be integrated in the Grid Information Service to perform network resource discovery and matchmaking. We then describe the User Interface through which the Path negotiation terms are specified by the user and we propose a path management approach that integrates different network technologies, namely MPLS and the Differentiated Services architecture. The implementation details of a system prototype are described together with some initial experimental results.

**Distributed Computing Systems and Experiences / 287**

## Super scaling PROOF to very large clusters

**Authors:** F. Rademakers<sup>1</sup>; G. Roland<sup>2</sup>; K. Gulbrandsen<sup>2</sup>; M. Ballintijn<sup>2</sup>; P. Canal<sup>3</sup>; R. Brun<sup>1</sup>

<sup>1</sup> *CERN*

<sup>2</sup> *MIT*

<sup>3</sup> *Fermilab*

**Corresponding Author:** maartenb@mit.edu

The Parallel ROOT Facility, PROOF, enables a physicist to analyze and understand very large data sets on an interactive time scale. It makes use of the inherent parallelism in event data and implements an architecture that optimizes I/O and CPU utilization in heterogeneous clusters with distributed storage. Scaling to many hundreds of servers is essential to process tens or hundreds of gigabytes of data interactively. This is supported by the industry trend to pack more CPU's into single systems and to create bigger clusters by increasing the number of systems per rack. We will describe the latest developments in PROOF and the development of a standardized benchmark for PROOF clusters. The benchmark is self contained and measures the network, the I/O and the processing characteristics of a cluster. We will present the comprehensive results of the benchmark for several clusters, demonstrating the performance and scalability of PROOF on very large clusters.

**Poster Session 3 - Board: 50 / 288**

## Parallel compilation of CMS software

**Authors:** G. Eulisse<sup>1</sup>; L. Tuura<sup>1</sup>; S. Ashby<sup>2</sup>; S. Schmid<sup>3</sup>

<sup>1</sup> *Northeastern University, Boston, MA, USA*

<sup>2</sup> *CERN*

<sup>3</sup> *ETH Zurich*

**Corresponding Author:** lassi.tuura@cern.ch

LHC experiments have large amounts of software to build. CMS has studied ways to shorten project build times using parallel and distributed builds as well as improved ways to decide what to rebuild. We have experimented with making idle desktop and server machines easily available as a virtual build cluster using distcc and zeroconf.

We have also tested variations of ccache and more traditional make dependency analysis.  
We report on our test results, with analysis of the factors that most improve or limit build performance.

**Poster Session 2 - Board: 64 / 289**

## LHC data files meet mass storage and networks: going after the lost performance

**Authors:** L. Tuura<sup>1</sup>; T. Barrass<sup>2</sup>; V. Innocente<sup>3</sup>

<sup>1</sup> *NORTHEASTERN UNIVERSITY, BOSTON, MA, USA*

<sup>2</sup> *Bristol University, UK*

<sup>3</sup> *CERN, PH/SFT*

**Corresponding Author:** lassituura@cern.ch

Experiments frequently produce many small data files for reasons beyond their control, such as output splitting into physics data streams, parallel processing on large farms, database technology incapable of concurrent writes into a single file, and constraints from running farms reliably. Resulting data file size is often far from ideal for network transfer and mass storage performance. Provided that time to analysis does not significantly deteriorate, files arriving from a farm could easily be merged into larger logical chunks, for example by physics stream and file type within a configurable time and size window.

Uncompressed zip archives seem an attractive candidate for such file merging and are currently tested by the CMS experiment. We describe the main components now in use: the merging tools, tools to read and write zip files directly from C++, plug-ins to the database system, mass-storage access optimisation, consistent handling of application and replica metadata, and integration with catalogues and other grid tools. We report on the file size ratio obtained in the CMS 2004 data challenge and observations and analysis on changes to data access as well as estimated impact on network usage.

**Core Software / 291**

## Automatic Procedures as Generated Analysis Tool

**Authors:** G. Asova<sup>1</sup>; K. Abrahamyan<sup>1</sup>; M. Winde<sup>1</sup>; P. Castro-Garcia<sup>2</sup>; S. Weisse<sup>1</sup>

<sup>1</sup> *DESY ZEUTHEN*

<sup>2</sup> *DESY*

**Corresponding Author:** winde@ifh.de

The photo injector test facility at DESY Zeuthen (PITZ) was built to develop, operate and optimize photo injectors for future free electron lasers and linear colliders. In PITZ we use a DAQ system

that stores data as a collection of ROOT files, forming our database for offline analysis. Consequently, the offline analysis will be performed by a ROOT application, written at least partly by the user (a physicist). To help the user to develop safe filters and data visualisation (graphs, histograms) with minimal effort in an existing ROOT framework application, we provide a GUI that generates C++ source files, compiles and links them to the rest of the application. We call these C++ routines “Automatic Procedures” (AP). Standard filter conditions and data visualisation can be generated by click or drag- and-drop, while more complex tasks may be expressed as small pieces of C++ code. Once compiled by ACLiC (ROOTs Automatic Compiler Linker), an Automatic Procedure may be reused without repeated compilation. E. g. the injector shift crew will run a number of ROOT applications, controlled by APs in regular intervals. Alternatively every AP can be read in and loaded to the GUI for further improvement. A number of APs can run in a logical sequence, parameters can be transferred from one AP to another. They can be selected by picking a point from a graph. The GUI was constructed with Qt, because that offers a comprehensive GUI programming toolkit.

The photo injector test facility at DESY Zeuthen (PITZ) was built to develop, operate and optimize photo injectors for future free electron lasers and linear colliders. In PITZ we use a DAQ system that stores data as a collection of ROOT files, forming our database for offline analysis. Consequently, the offline analysis will be performed by a ROOT application, written at least partly by the user (a physicist). To help the user to develop safe filters and data visualisation (graphs, histograms) with minimal effort in an existing ROOT framework application, we provide a GUI that generates C++ source files, compiles and links them to the rest of the application. We call these C++ routines “Automatic Procedures” (AP). Standard filter conditions and data visualisation can be generated by click or drag- and-drop, while more complex tasks may be expressed as small pieces of C++ code. Once compiled by ACLiC (ROOTs Automatic Compiler Linker), an Automatic Procedure may be reused without repeated compilation. E. g. the injector shift crew will run a number of ROOT applications, controlled by APs in regular intervals. Alternatively every AP can be read in and loaded to the GUI for further improvement. A number of APs can run in a logical sequence, parameters can be transferred from one AP to another. They can be selected by picking a point from a graph. The GUI was constructed with Qt, because that offers a comprehensive GUI programming toolkit.

Keywords: Automatic Procedure, ROOT, ACLiC, Data Analysis, Data Visualisation, GUI, Qt

**Core Software / 292**

## **IceTray: a Software Framework for IceCube**

**Author:** T. DeYoung<sup>1</sup>

<sup>1</sup> UNIVERSITY OF MARYLAND

**Corresponding Author:** deyoung@pa.msu.edu

IceCube is a cubic kilometer-scale neutrino telescope under construction at the South

Pole. The minimalistic nature of the instrument poses several challenges for the software framework. Events occur at random times, and frequently overlap, requiring some modifications of the standard event-based processing paradigm. Computational requirements related to modeling the detector medium necessitate the ability for software components to defer processing events. With minimal information from the detector, events must be reconstructed many times with different hypotheses or methods, and the results compared. The appropriate series of software components required to process an event varies considerably, and can be determined only at run time. Finally, reconstruction algorithms are constantly evolving, with development taking place throughout the collaboration, so it is essential that conversion of private analysis code to online production software be simple and, given the inaccessibility of the experimental site, robust. The IceCube collaboration has developed the IceTray framework, which meets these needs by blending aspects of push- and pull-based architectures to produce a highly modular system which nevertheless allows each software component a significant degree of control over the execution flow.

Poster Session 2 - Board: 12 / 293

## JIM Deployment for the CDF Experiment

**Authors:** A. Baranovski<sup>1</sup>; A. Kreymer<sup>2</sup>; A. Lyon<sup>2</sup>; A. Sill<sup>3</sup>; F. Ratnikov<sup>4</sup>; G. Garzoglio<sup>2</sup>; I. Terekhov<sup>1</sup>; J. Trumbo<sup>1</sup>; L. Loebel Carpenter<sup>1</sup>; L. Lueking<sup>2</sup>; M. Burgon-Lyon<sup>5</sup>; M. Leslie<sup>6</sup>; R. Herber<sup>2</sup>; R. Illingworth<sup>2</sup>; R. Kennedy<sup>2</sup>; R. St.Denis<sup>7</sup>; S. Belforte<sup>8</sup>; S. Stonjek<sup>7</sup>; S. Veseli<sup>1</sup>; S. White<sup>1</sup>; U. Kerzel<sup>9</sup>; V. Bartsch<sup>6</sup>; W. Merritt<sup>2</sup>

<sup>1</sup> *Fermilab*

<sup>2</sup> *FERMI NATIONAL ACCELERATOR LABORATORY*

<sup>3</sup> *Texas Tech. University*

<sup>4</sup> *Rutgers University*

<sup>5</sup> *UNIVERSITY OF GLASGOW*

<sup>6</sup> *Oxford University*

<sup>7</sup> *Glasgow University*

<sup>8</sup> *INFN, Trieste*

<sup>9</sup> *Karlsruhe University*

**Corresponding Author:** m.burgon-lyon@physics.gla.ac.uk

JIM (Job and Information Management) is a grid extension to the mature data handling system called SAM (Sequential Access via Metadata) used by the CDF, DZero and Minos Experiments based at Fermilab. JIM uses a thin client to allow job submissions from any computer with Internet access, provided the user has a valid certificate or kerberos ticket. On completion the job output can be downloaded using a web interface. The JIM execution site software can be installed on shared resources, such as ScotGRID, as it may be configured for any batch system and does not require exclusive control of the hardware. Resources that do not belong entirely to CDF and thus cannot run DCAF (Decentralised CDF Analysis Farm), may therefore be accessed using JIM. We will report on the initial deployment of JIM for CDF and the steps taken to integrate JIM with DCAF.

Poster Session 2 - Board: 33 / 294

## Data management services of NorduGrid

**Authors:** A. Konstantinov<sup>1</sup>; A. Waananen<sup>2</sup>; B. Konya<sup>3</sup>; F. Ould-Saada<sup>4</sup>; J L. Nielsen<sup>2</sup>; J R. Hansen<sup>2</sup>; M. Ellert<sup>5</sup>; O. Smirnova<sup>3</sup>; P. Eerola<sup>3</sup>; T. Ekelof<sup>6</sup>

<sup>1</sup> Vilnius University, Lithuania

<sup>2</sup> Niels Bohr Institute, Copenhagen, Denmark

<sup>3</sup> Lund University, Sweden

<sup>4</sup> University of Oslo, Norway

<sup>5</sup> CERN

<sup>6</sup> Uppsala University, Sweden

In common grid installations, services responsible for storing big data chunks, replication of those data and indexing their availability are usually completely decoupled. And a task of synchronizing data is passed to either user-level tools or separate services (like spiders) which are subject to failure and usually cannot perform properly if one of underlying services fails too.

The NorduGrid Smart Storage Element (SSE) was designed to try to overcome those problems by combining the most desirable features into one service. It uses HTTPS/G for secure data transfer, Web Services for control (through same HTTPS/G channel) and can provide information to indexing services used in middlewares based on the Globus Toolkit (TM). At the moment, those are the Replica Catalog and the Replica Location Service. The modular internal design of the SSE and the power of C++ object programming allows to add support for other indexing services in an easy way.

There are plans to complement it with a Smart Indexing Service capable of resolving inconsistencies hence creating a robust distributed data storage system.

## Wide Area Networking / 295

### Realization of a stable network flow with high performance communication in high bandwidth-delay product network

**Author:** Y. Kodama<sup>1</sup>

**Co-authors:** O. Tatebe<sup>2</sup>; S. Sekiguchi<sup>2</sup>; T. Kudoh<sup>2</sup>

<sup>1</sup> NATIONAL INSTITUTE OF ADVANCED INDUSTRIAL SCIENCE AND TECHNOLOGY (AIST)

<sup>2</sup> AIST

**Corresponding Author:** y-kodama@aist.go.jp

It is important that the total bandwidth of the multiple streams should not exceed the network bandwidth in order to achieve a stable network flow with high performance in high bandwidth-delay product networks. Software control of bandwidth for each stream sometimes exceed the specified bandwidth. We proposed the hardware control technique for total bandwidth of multiple streams with high accuracy.

GNET-1 is the hardware gigabit network testbed that we developed. It provides functions such as wide area network emulation, network instrumentation, and traffic generation at gigabit Ethernet wire speeds. GNET-1 is a powerful tool for developing network-aware grid software. It can control the total bandwidth of the multiple streams with high accuracy by adjusting the interframe gap (IFG).

To see the effect of the highly accurate bandwidth control by GNET-1, the file exchange of large-scale data was done on a Trans-pacific Grid Datafarm testbed between Japan-U.S.. We used three trans-pacific networks, APAN/TransPAC Los Angeles line and its Chicago line and SuperSINET New York line. Its total bandwidth that can be used was 3.9

Gbps. In this feasible study, GNET-1 controlled five gigabit Ethernet ports, and achieved the total bandwidth of 3.78 Gbps in stable for about one hour. The bandwidth was 97 % of the peak bandwidth of used networks.

**Poster Session 2** - Board: 51 / 296

## A Multidimensional Approach to the Analysis of Grid Monitoring Data

**Authors:** G. Rubini<sup>1</sup>; G. Tortone<sup>2</sup>; N. De Bortoli<sup>2</sup>; S. Andreozzi<sup>1</sup>; S. Fantinel<sup>3</sup>

<sup>1</sup> *INFN-CNAF*

<sup>2</sup> *INFN-Napoli*

<sup>3</sup> *INFN-LNL*

**Corresponding Author:** gianluca.rubini@cnaif.infn.it

Analyzing Grid monitoring data requires the capability of dealing with multidimensional concepts intrinsic to Grid systems. The meaningful dimensions identified in recent works are the physical dimension referring to geographical location of resources, the Virtual Organization (VO) dimension, the time dimension and the monitoring metrics dimension. In this paper, we discuss the application of On-Line Analytical Processing (OLAP), an approach to the fast analysis of shared multidimensional information, to the mentioned problem. OLAP relies on structures called 'OLAP cubes', that are created by a reorganization of data contained inside a relational database, thus transforming operational data into dimensional data.

Our OLAP model is a four-dimension cuboid based on time, geographic, Virtual Organization (VO), and monitoring metric. Time and geographic dimensions have total order relation and form two concept hierarchies, respectively hours<days<weeks<months and services<hosts<sites<countries. This model is applied to a set of Grid monitoring data generated by a monitoring tool called GridICE.

**Poster Session 3** - Board: 32 / 297

## New specific solids definitions in the Geant4 geometry modeller

**Author:** O. Link<sup>1</sup>

<sup>1</sup> *CERN, PH/SFT*

**Corresponding Author:** oliver.link@cern.ch

Twisted trapezoids are important components in the LAr end cap calorimeter of the Atlas detector. A similar solid, the so-called twisted tube consists of two end planes, inner and outer hyperboloidal surfaces, and twisted surfaces, and is an indispensable component for cylindrical drift chambers (see K. Hoshina et al, Computer Physics Communications 153 (2003) 373-391). In Geant3 exists a general version of a twisted trapezoid, however the implementation puts very strong restrictions on its use.

In the Geant4 toolkit no solids have been available to date to describe twisted objects. The design and realisation of new twisted solids within the framework of

Geant4 will be presented together with the algorithmic details, followed by a discussion of the performance and accuracy test results.

**Poster Session 3 - Board: 4 / 298**

## **ROOT : detector visualization**

**Authors:** A. Gheata<sup>1</sup>; M. Gheata<sup>2</sup>; O. Couet<sup>1</sup>; R. Brun<sup>1</sup>

<sup>1</sup> *CERN*

<sup>2</sup> *Institute of Space Science - Bucharest*

**Corresponding Author:** olivier.couet@cern.ch

The ROOT geometry package is a tool designed for building, browsing, tracking and visualizing a detector geometry. The code is independent from other external MC for simulation, therefore it does not contain any constraints related to physics. However, the package defines a number of hooks for tracking, such as media, materials, magnetic field or track state flags, in order to allow interfacing to tracking MC's. The final goal is to be able to use the same geometry for several purposes, such as tracking, reconstruction or visualization, taking advantage of the ROOT features related to bookkeeping, I/O, histogramming, browsing and GUI's.

In this poster, we will show the various graphics tools to render complex geometries, from ray tracing tools that have the advantage to test the real geometry like when tracking particles, to sophisticated 3-D dynamic graphics with the OpenGL, X3D, Coin3D or OpenInventor viewers. An abstract interface has been defined and it is common to all the viewers.

**Distributed Computing Systems and Experiences / 300**

## **The ARDA Prototypes**

**Authors:** A. Demichev<sup>1</sup>; A. Maier<sup>2</sup>; A. Peters<sup>2</sup>; B. Koblitz<sup>2</sup>; D. Feichtinger<sup>3</sup>; D. Liko<sup>2</sup>; F. Orellana<sup>2</sup>; J. Andreeva<sup>2</sup>; J. Herralá<sup>2</sup>; J. MOSCICKI<sup>2</sup>; J. Moscicki<sup>2</sup>; M. Lamanna<sup>2</sup>; T. Chen<sup>4</sup>; V. Pose<sup>5</sup>; W. Ueng<sup>4</sup>

<sup>1</sup> *Moscow State University*

<sup>2</sup> *CERN*

<sup>3</sup> *Swiss Institute of Particle Physics*

<sup>4</sup> *Academica Sinica, Nanking*

<sup>5</sup> *JINR, Dubna*

**Corresponding Author:** julia.andreeva@cern.ch

The ARDA project was started in April 2004 to support the four LHC experiments (ALICE, ATLAS, CMS and LHCb) in the implementation of individual production and analysis environments based on the EGEE middleware.

The main goal of the project is to allow a fast feedback between the experiment and the middleware development teams via the construction and the usage of end-to-end prototypes



allowing users to perform analyses out of the present data sets from recent montecarlo productions.

We present the status of the integration of the EGEE prototype Grid middleware into the analysis environment of the four LHC experiments. First an overview is given on the individual architectures of the four experiments' prototypes with a strong focus on how the EGEE middleware is incorporated into the framework. We outline common points in the usage of the middleware and try to point out differences in the decisions taken by the experiments on the inclusion of different parts of the EGEE software. We will conclude by presenting the first feedback from the usage of these analysis environments.

## Wide Area Networking / 301

# HEPBook - A Personal Collaborative HEP notebook

**Authors:** G. Roediger<sup>1</sup>; P. POMATTO<sup>2</sup>

<sup>1</sup> CORPORATE COMPUTER SERVICES INC. - FERMILAB

<sup>2</sup> FNAL, Batavia, USA

**Corresponding Author:** roediger@fnal.gov

A High Energy Physics experiment has between 200 and 1000 collaborating physicists from nations spanning the entire globe. Each collaborator brings a unique combination of interests, and each has to search through the same huge heap of messages, research results, and other communication to find what is useful.

Too much scientific information is as useless as too little. It is time consuming, tedious, and difficult to sift and search for the pertinent bits. Often, the exact words to search for are unknown, or the information is badly organized, and the pertinent bits are not found. The search is abandoned, the time is lost, and valuable information is never communicated as it was intended.

Much of a collaboration's information is in the individual physicists paper logbooks. The physicists record important and pertinent information for their research. They save the log books to refer to it later, copy pages, and distribute them to their collaborators who share their interest and research.

Electronic Logbooks are now used in the control room of large detectors during the acquisition phase. They have proven useful for communicating the status of the detector and to keep the history of lab sessions in a format that can be queried and retrieved quickly. It has enabled remote monitoring of the detector and remote emergency help.

We have implemented an electronic Control Room Logbook, called CRL. It is used in the D0 experiment's detector control room for the Run II acquisition. As of mid April 2004 there are over 305,000 entries in the D0 logbook, all viewable and able to be annotated from the web. Other experiments such as CMS, MiniBoone, and Minos have also adapted the CRL. These experiments all have very different needs, so they all configured and customized the CRL in many different ways. The HEPBook will move the logbook from the control room to the personal and collaborative HEP notebook. In this paper we will review the HEPBook technology and capabilities and discuss the new HEPBook architecture. Among the topics discussed will be the use of Java reflection to recursively produce an XML representation of an entry, the ability to save personal entries as well as share entries among a collaboration through multiple repositories which incorporate software agent technology, interface with the GRID, and implement multiple security models. The HEPBook runs on all Java

platforms including Apple, Win32, and Linux. A brief demo will be given of the HEPBook.

**Online Computing / 302**

## Control in the ATLAS TDAQ System

**Authors:** A. AMORIM<sup>1</sup>; A. Kazarov<sup>2</sup>; D. BURCKHART-CHROMEK<sup>3</sup>; D. KLOSE<sup>1</sup>; D. Liko<sup>4</sup>; E. BADESCU<sup>5</sup>; I. ALEXANDROV<sup>6</sup>; I. SOLOVIEV<sup>7</sup>; J. FLAMMER<sup>3</sup>; L. MAPELLI<sup>3</sup>; L. PEDRO<sup>1</sup>; M. CAPRINI<sup>5</sup>; M. DOBSON<sup>3</sup>; M. Mineev<sup>8</sup>; N. FIUZA DE BARROS<sup>1</sup>; R. JONES<sup>3</sup>; S. KOLOS<sup>4</sup>; S. KOROBOV<sup>6</sup>; V. Kotov<sup>8</sup>; Y. RYABOV<sup>7</sup>

<sup>1</sup> *Universidade de Lisboa, Portugal*

<sup>2</sup> *Petersburg Nuclear Physics Institute (PNPI)*

<sup>3</sup> *CERN, Geneva, Switzerland*

<sup>4</sup> *CERN*

<sup>5</sup> *National Institute for Nuclear Physics and Engineering, Bucharest, Romania*

<sup>6</sup> *JINR, Dubna, Russia*

<sup>7</sup> *PNPI, St-Petersburg, Russia*

<sup>8</sup> *Joint Institute for Nuclear Research (JINR)*

**Corresponding Author:** dietrich.liko@cern.ch

The unprecedented size and complexity of the ATLAS TDAQ system requires a comprehensive and flexible control system. Its role ranges from the so-called run-control, e.g. starting and stopping the data-taking, to error handling and fault tolerance. It also includes initialisation and verification of the overall system. Following the traditional approach a hierarchical system of customizable controllers has been proposed. For the final system all functionality would be therefore available in a distributed manner, with the possibility of local customisation.

After a technology survey the open source expert system CLIPS has been chosen as a basis for the implementation of the supervision and the verification system. The CLIPS interpreter has been extended to provide a general control framework. Other ATLAS Online software components have been integrated as plugins and provide the mechanism for configuration and communication.

Several components have been implemented, that share this technology. The dynamic behaviour of the individual component is fully described by the rules, while the framework is based on a common implementation. During these year these components have been the subject of scalability tests up to the full system size. Encouraging results are presented and validate the technology choice.

**Poster Session 3 - Board: 10 / 303**

## The ROOT Linear Algebra

**Authors:** E. Offermann<sup>1</sup>; F. Rademakers<sup>1</sup>; R. Brun<sup>1</sup>

<sup>1</sup> *CERN*

**Corresponding Author:** rene.brun@cern.ch

The ROOT linear algebra package has been invigorated . The hierarchical structure has been improved allowing different flavors of matrices, like dense and symmetric . A fairly complete set of matrix decompositions has been added to support matrix inversions and solving linear equations.

The package has been extensively compared to other algorithms for its accuracy and performance.

In this poster we will describe the structure of the package and several benchmarks obtained with typical linear algebra applications.

**Poster Session 1 - Board: 17 / 304**

## **The Design, Installation and Management of a Tera-Scale High Throughput Cluster for Particle Physics Research**

**Author:** A. Martin<sup>1</sup>

<sup>1</sup> *QUEEN MARY, UNIVERSITY OF LONDON*

**Corresponding Author:** a.j.martin@qmul.ac.uk

We describe our experience in building a cost efficient High Throughput Cluster (HTC) using commodity hardware and free software within a university environment. Our HTC has a modular system architecture and is designed to be upgradable. The current, second phase configuration, consists of 344 processors and 20 Tbyte of RAID storage.

In order to rapidly install and upgrade software, we have developed automatic remote system installation and configuration tools to deploy standard software configurations on individual machines. To efficiently manage machines we have written a custom cluster configuration database. This database is used to track all hardware components in the cluster, the network and power distribution and the software configuration. Access to this database and the cluster performance and monitoring systems is provided by a web portal, which allows efficient remote management in our low-manpower environment.

We describe the performance of our system under a mixed load of scalar and parallel tasks and discuss future possible improvements.

**Distributed Computing Services / 305**

## **Experiences with the gLite Grid Middleware**

**Authors:** A. Demichev<sup>1</sup>; A. Maier<sup>2</sup>; A. Peters<sup>2</sup>; B. Koblitz<sup>2</sup>; D. Feichtinger<sup>3</sup>; D. Liko<sup>2</sup>; F. Orellana<sup>2</sup>; J. Herrala<sup>2</sup>; J. Moscicki<sup>2</sup>; M. Lamanna<sup>4</sup>; T. Chen<sup>5</sup>; V. Pose<sup>6</sup>; W. Ueng<sup>5</sup>

<sup>1</sup> *Moscow State University*

<sup>2</sup> *CERN*

<sup>3</sup> *Swiss Institute of Particle Physics*

<sup>4</sup> *CERN-IT*

<sup>5</sup> *Academica Sinica, Nanking*

<sup>6</sup> *JINR, Dubna*

The ARDA project was started in April 2004 to support the four LHC experiments (ALICE, ATLAS, CMS and LHCb) in the implementation of individual production and analysis environments based on the EGEE middleware.

The main goal of the project is to allow a fast feedback between the experiment and the middleware development teams via the construction and the usage of end-to-end prototypes allowing users to perform analyses out of the present data sets from recent montecarlo productions.

The LCG ARDA project is contributing to the development of the new EGEE Grid middleware by exercising it with realistic analysis systems developed within the four LHC experiments. We will present our experiences in using the EGEE middleware in first prototypes developed by the experiments together with the ARDA project. We will cover aspects such as the usability of individual components of the middleware and give an overview on which components are used by which experiments.

**Distributed Computing Systems and Experiences / 307**

## **Use of Condor and GLOW for CMS Simulation Production**

**Authors:** D. Bradley<sup>1</sup>; M. Livny<sup>1</sup>; S. Dasu<sup>1</sup>; S. Rader<sup>1</sup>; V. Puttabuddhi<sup>1</sup>; W. Smith<sup>1</sup>

<sup>1</sup> UNIVERSITY OF WISCONSIN

**Corresponding Author:** dasu@hep.wisc.edu

The University of Wisconsin distributed computing research groups developed a software system called Condor for high throughput computing using commodity hardware. An adaptation of this software, Condor-G, is part of Globus grid computing toolkit. However, original Condor has additional features that allows building of an enterprise level grid. Several UW departments have Condor computing pools that are integrated in such a way as to flock jobs from one pool to another as resources become available. An interdisciplinary team of UW researchers recently built a new distributed computing facility, the Grid Laboratory of Wisconsin (GLOW). In total Condor pools in the UW have about 2000 Intel CPUs (P-III and Xeon) which are available for scientific computation. By exploiting special features of Condor such as checkpointing and remote IO we have generated over 10 million fully simulated CMS events. We were able to harness about 260 CPU-days per day for a period of 2 months when we were operational late fall. We have scaled to using 500 CPUs concurrently when opportunity to exploit unused resources in laboratories on our campus. We have built a scalable job submission and tracking system called Jug using Python and MySQL which enabled us to scale to run hundreds of jobs simultaneously. Jug also ensured that the data generated is transferred to US Tier-I center at Fermilab. We have also built a portal to our resources and participated in Grid2003 project. We are currently adapting our environment for providing analysis resources. In this paper we will discuss our experience and observations regarding the use of opportunistic resources, and generalize them to wider grid computing context.

**Poster Session 2 - Board: 61 / 308**

## ARDA Project Status Report

**Authors:** M. Lamanna<sup>1</sup>; The ARDA Team<sup>2</sup>

<sup>1</sup> *CERN-IT*

<sup>2</sup> *CERN*

The ARDA project was started in April 2004 to support the four LHC experiments (ALICE, ATLAS, CMS and LHCb) in the implementation of individual production and analysis environments based on the EGEE middleware.

The main goal of the project is to allow a fast feedback between the experiment and the middleware development teams via the construction and the usage of end-to-end prototypes allowing users to perform analyses out of the present data sets from recent montecarlo productions.

In this talk the project is presented with highlights of the first results and lessons learnt so far. The relations of the project with similar initiatives within and outside the High Energy Physics community are reviewed (notably in the EGEE application identification and support).

**Distributed Computing Systems and Experiences / 309**

## A GRID approach for Gravitational Waves Signal Analysis with a Multi-Standard Farm Prototype

**Authors:** A. Eleuteri<sup>1</sup>; F. Acernese<sup>1</sup>; F. Barone<sup>2</sup>; F.M. Taurino<sup>3</sup>; G. Russo<sup>4</sup>; G. Tortone<sup>3</sup>; L. Giordano<sup>5</sup>; L. Milano<sup>4</sup>; P. Mastroserio<sup>6</sup>; R. De Rosa<sup>4</sup>; R. Esposito<sup>6</sup>; S. Pardi<sup>7</sup>

<sup>1</sup> *Dipartimento di Scienze Fisiche - Univeristà di Napoli Federico II*

<sup>2</sup> *Dipartimento di Scienze Farmaceutiche - Università di Salerno*

<sup>3</sup> *INFN, Napoli*

<sup>4</sup> *Dipartimento di Scienze Fisiche - Università di Napoli Federico II*

<sup>5</sup> *INFN and University of Napoli*

<sup>6</sup> *INFN - sez.Napoli*

<sup>7</sup> *DIPARTIMENTO DI MATEMATICA ED APPLICAZIONI "R.CACCIOPPOLI"*

**Corresponding Author:** spardi@na.infn.it

The standard procedures for the extraction of gravitational wave signals coming from coalescing binaries provided by the output signal of an interferometric antenna may require computing powers generally not available in a single computing centre or laboratory. A way to overcome this problem consists in using the computing power available in different places as a single geographically distributed computing system. This solution is now effective within the GRID environment, that allows distributing the required computing effort for specific data analysis procedure among different sites according to the available computing power.

Within this environment we developed a system prototype with application software for the experimental tests of a geographically distributed computing system for the analysis of gravitational wave signal from coalescing binary systems. The facility has been developed as a general purpose system that uses only standard hardware and software components, so that it can be easily upgraded and configured. In fact, it can be partially or totally configured as a GRID farm, as MOSIX farm or as MPI farm. All these three configurations may coexist since the facility can be split

into configuration subsets. A full description of this farm is reported, together with the results of the performance tests and planned developments.

**Poster Session 3 - Board: 40 / 310**

## Dynamic Matched Filters for Gravitational Waves Detection

**Authors:** A. ELEUTERI<sup>1</sup>; F. ACERNESE<sup>1</sup>; F. BARONE<sup>2</sup>; G. RUSSO<sup>3</sup>; L. MILANO<sup>4</sup>; R. DE ROSA<sup>1</sup>; S. Pardi<sup>5</sup>

<sup>1</sup> *Napoli University and INFN Napoli, Italy*

<sup>2</sup> *INFN Napoli and Salerno University, Italy*

<sup>3</sup> *Napoli University, Italy*

<sup>4</sup> *Dipartimento di Scienze Fisiche - Università di Napoli Federico II*

<sup>5</sup> *DIPARTIMENTO DI MATEMATICA ED APPLICAZIONI "R.CACCIOPPOLI"*

**Corresponding Author:** spardi@na.infn.it

The algorithms for the detection of gravitational waves are usually very complex due to the low signal to noise ratio. In particular the search for signals coming from coalescing binary systems can be very demanding in terms of computing power, like in the case of the classical Standard Matched Filter Technique. To overcome this problem, we tested a Dynamic Matched Filter Technique, still based on Matched Filters, whose main advantage is the requirement of a lower computing power. In this work this technique is described, together with its possible application as a pre-data analysis algorithm. Also the results on simulated data are reported.

**Core Software / 311**

## OpenScientist. Status of the project.

**Author:** G B. Barrand<sup>1</sup>

<sup>1</sup> *CNRS / IN2P3 / LAL*

**Corresponding Author:** barrand@lal.in2p3.fr

We want to present the status of this project.  
After quickly remembering the basic choices around GUI, visualization and scripting we would like to develop what had been done in order to have an AIDA-3.2.1 compliant system, to visualize Geant4 data (G4Lab module), to visualize ROOT data (Mangrove module), to have an hippodraw module and what had been done in order to run on MacOSX by using the native NextStep (Cocoa) environment.

**Core Software / 312**

## CLHEP Infrastructure Improvements

**Author:** L. Garren<sup>1</sup>

<sup>1</sup> *F.N.A.L.*

**Corresponding Author:** garren@fnal.gov

CLHEP is a set of HEP-specific foundation and utility classes such as random number generators, physics vectors, and particle data tables. Although CLHEP has traditionally been distributed as one large library, the user community has long wanted to build and use CLHEP packages separately.

With the release of CLHEP 1.9, CLHEP has been reorganized and enhanced to enable building and using CLHEP packages individually as well as collectively. The revised build strategy employs all the components of the standard autotools suite: automake, autoconf, and libtool. In combination with the reorganization, the use of these components makes it easy not only to rebuild any single package (e.g., when that package changes), but also to add new packages.

This presentation will discuss the new CLHEP structure, illustrate the role and use of the autotools, and describe how other packages with similar organization can be seamlessly integrated with the CLHEP libraries.

**Poster Session 1 - Board:** 26 / 313

## Remote Shifting at the CLEO Experiment

**Authors:** H. Schwarthoff<sup>1</sup>; T. Wilksen<sup>2</sup>

<sup>1</sup> *Cornell University*

<sup>2</sup> *CORNELL UNIVERSITY*

**Corresponding Author:** wilksen@mail.lns.cornell.edu

The CLEO III data acquisition was from the beginning in the late 90's designed to allow remote operations and monitoring of the experiment. Since changes in the coordination and operation of the CLEO experiment two years ago enabled us to separate tasks of the shift crew into an operational and a physics task, existing remote capabilities have been revisited. In 2002/03 CLEO started to deploy its remote monitoring tasks for performing remote shifts and evaluated various communication tools e.g. video conferencing and remote desktop sharing. Remote, collaborating institutions were allowed to perform the physicist shift part from their home institutions keeping only the professional operator of the CLEO experiment on site. After a one year long testing and evaluation phase the remote shifting for physicists is now in production mode.

This talk reports on experiences made when evaluating and deploying various options and technologies used for remote control, operation and monitoring e.g. CORBA's IIOP, X11 and VNC in the CLEO experiment. Furthermore some aspects of the usage of video conferencing tools by distributed shift crews are being discussed.

**Event Processing / 314**

## Geant4: status and recent developments

**Authors:** H.P. wellisch<sup>1</sup>; J. Apostolakis<sup>1</sup>; M. Asai<sup>2</sup>

<sup>1</sup> CERN

<sup>2</sup> SLAC

**Corresponding Author:** john.apostolakis@cern.ch

Geant4 is relied upon in production for increasing number of HEP experiments and for applications in several other fields. Its capabilities continue to be extended, as its performance and modelling are enhanced.

This presentation will give an overview of recent developments in diverse areas of the toolkit. These will include, amongst others, the optimisation for complex setups using different production thresholds, improvements in the propagation in fields, and highlights from the physics processes and event biasing.

In addition it will note the physics validation effort undertaken in collaboration with a number of experiments, groups and users.

**Event Processing / 315**

## Synergia: A Modern Tool for Accelerator Physics Simulation

**Authors:** J. Amundson<sup>1</sup>; P. Spentzouris<sup>2</sup>

<sup>1</sup> Fermi National Accelerator Laboratory

<sup>2</sup> FERMI NATIONAL ACCELERATOR LABORATORY

**Corresponding Author:** spentz@fnal.gov

Computer simulations play a crucial role in both the design and operation of particle accelerators. General tools for modeling single-particle accelerator dynamics have been in wide use for many years. Multi-particle dynamics are much more computationally demanding than single-particle dynamics, requiring supercomputers or parallel clusters of PCs. Because of this, simulations of multi-particle dynamics have been much more specialized. Although several multi-particle simulation tools are now available, they tend to cover a narrow range of topics. Most also present difficulties for the end user ranging from platform portability to arcane interfaces.

In this presentation, we discuss Synergia, a multi-particle accelerator simulation tool developed at Fermilab, funded by the DOE SciDAC program. Synergia was designed to cover a variety of physics processes while presenting a flexible and humane interface to the end user. It is a hybrid application, primarily based on the existing packages mxyzptlk/beamline and Impact. Our presentation covers Synergia's physics capabilities and human interface. We focus on the computational problems we encountered and solved in the process of building an application out of codes written in Fortran 90, C++, and wrapped with a Python front-end. We also discuss some approaches we have used in the visualization of the high-dimensional data that comes out of a particle accelerator simulations, especially our work with OpenDX.



**Core Software / 316****CDB - Distributed Conditions Database of BaBar Experiment****Authors:** D. Brown<sup>1</sup>; I. Gaponenko<sup>2</sup><sup>1</sup> *Lawrence Berkeley National Laboratory (LBNL)*<sup>2</sup> *LAWRENCE BERKELEY NATIONAL LABORATORY***Corresponding Author:** gapon@slac.stanford.edu

A new, completely redesigned Condition/DB was deployed in BaBar in October 2002. It replaced the old database software used through the first three and half years of data taking.

The new software aims at performance and scalability limitations of the original database. However this major redesign brought in a new model of the metadata, brand new technology- and implementation- independent API, flexible configurability and extended functionality.

One of the greatest strength of new CDB is that it's been designed to be a distributed kind database from the ground up to facilitate propagation and exchange of conditions (calibrations, detector alignments, etc.) in the realm of the international HEP collaboration.

The first implementation of CDB uses Objectivity/DB as its underlying persistent technology. There is an ongoing study to understand how to implement CDB on top of other persistent technologies.

The talk will cover the whole spectrum of topics ranging from the basic conceptual model of the new database through the way CDB is currently exploited in BaBar to the directions of further developments.

**Poster Session 1 - Board: 25 / 317****Linux for the CLEO-c Online system****Authors:** H. Schwarthoff<sup>1</sup>; T. Wilksen<sup>1</sup><sup>1</sup> *CORNELL UNIVERSITY***Corresponding Author:** hubert@mail.lns.cornell.edu

The CLEO collaboration at the Cornell electron positron storage ring CESR has completed its transition to the CLEO-c experiment. This new program contains a wide array of Physics studies of  $e^+e^-$  collisions at center of mass energies between 3 GeV and 5 GeV.

New challenges await the CLEO-c Online computing system, as the trigger rates are expected to rise from < 100 Hz to around 300 Hz at the J/Psi production threshold, with a moderate increase in data throughput requirements. While the current Solaris and VxWorks based readout system will perform adequately under those conditions, there is a desire to improve the performance of the central components to extend monitoring capabilities and provide larger safety margins. The solution, as in most modern particle detector systems, is to deploy Linux on Intel architecture computers for the performance critical applications.

For reasons of hardware and software availability, the existing CLEO Online and Offline computing environment has been ported to the Linux platform. This development allows the described challenge to be met.

In this presentation, we will report on our experiences adapting the CLEO Online computing system for operation under Linux. Issues regarding third party software and code portability will be addressed. Performance measurements will be presented.

#### Distributed Computing Services / 318

### The STAR Unifid Meta-Scheduler project, a front end around evolving technologies for user analysis and data production.

**Authors:** D. Olson<sup>1</sup>; E. Efstathiadis<sup>2</sup>; G. CARCASSI<sup>2</sup>; I. Sakrejda<sup>3</sup>; Jerome LAURET<sup>2</sup>; L. Didenko<sup>2</sup>; V. Fine<sup>2</sup>

<sup>1</sup> *Lawrence Berkeley Laboratory*

<sup>2</sup> *BROOKHAVEN NATIONAL LABORATORY*

<sup>3</sup> *NERSC*

**Corresponding Author:** jlauret@bnl.gov

While many success stories can be told as a product of the Grid middleware developments, most of the existing systems relying on workflow and job execution are based on integration of self-contained production systems interfacing with a given scheduling component or portal, or directly uses the base component of the Grid middleware (globus-job-run, globus-job-submit). However, such systems usually do not take advantage of the presence of Resource Manager System (RMS); they hardly allow for a mix of local RMS and are either Grid or non-grid enabled. We intend to present an approach taking advantage of both worlds.

The STAR Unified Meta-Scheduler (SUMS) project provides users a way to submit jobs on a farm, at a site (multiple pools or farms) or on the Grid without the need to know or adapt to the diversity of technologies and knowledge involved while using multiple LRMS and their specificities. The strategy was adopted in 2002 to shield the users against changes in technologies inherent to the emerging Grid infrastructure and developments.

Java based and taking as input a simple user job description language (U-JDL), SUMS allows connection with multiple (overlapping or not) LRMS and Grid job submission (Condor-G, grid-job-submit, ...) without the need for changing the U-JDL. Fully integrated with the STAR File and Replica Catalog, information providers (load and queue information), SUMS provides a single point of reference for users to migrate from a traditional to a distributed computing environment. Results and the evolutionist architecture of the SUMS will be presented and its future, improvements and evolution will be discussed.

#### Distributed Computing Systems and Experiences / 319

### Grid Collector: Using an Event Catalog to Speed up User Analysis in Distributed Environment

**Authors:** A. Shoshani<sup>1</sup>; K. Wu<sup>2</sup>; V. Perevoztchikov<sup>3</sup>; W-M. Zhang<sup>4</sup>

<sup>1</sup> *Lawrence Berkeley National Lab*

<sup>2</sup> *LAWRENCE BERKELEY NATIONAL LAB*

<sup>3</sup> *Brookhaven National Lab*

<sup>4</sup> *Kent State University*

**Corresponding Author:** kwu@lbl.gov

Nuclear and High Energy Physics experiments such as STAR at BNL are generating millions of files with PetaBytes of data each year. In most cases, analysis programs have to read all events in a file in order to find the interesting ones.

Since most analyses are only interested in some subsets of events in a number of files, a significant portion of the computer time is wasted on reading the unwanted events. To address this issue, we developed a software system called the Grid Collector. The core of the Grid Collector is an “Event Catalog”.

This catalog can be efficiently searched with compressed bitmap indices. Tests show that it can index and search STAR event data much faster than database systems.

It is fully integrated with an existing analysis framework so that a minimal effort is required to use the Grid Collector in an analysis program. In addition, by taking advantage of existing file catalogs, Storage Resource Managers (SRMs) and GridFTP, the Grid Collector automatically downloads the needed files anywhere on the Grid without user intervention.

The Grid Collector can significantly improve user productivity. The improvement in productivity is more significant as users converge toward searching for rare events, because only the rare events are read into memory and the necessary files are automatically located and downloaded through the best available route. For a user that typically performs computation on 50% of the events, using the Grid Collector could reduce the turn around time by a half.

**Event Processing / 320**

## Overview and new developments in Geant4 electromagnetic physics

**Authors:** H. BURKHARDT<sup>1</sup>; L. Urban<sup>2</sup>; M. Maire<sup>3</sup>; P. GUMPLINGER<sup>4</sup>; R.P. KOKOULIN<sup>5</sup>; V. Grichine<sup>2</sup>; V. Ivantchenko<sup>6</sup>

<sup>1</sup> CERN, Geneva, Switzerland

<sup>2</sup> CERN

<sup>3</sup> LAPP

<sup>4</sup> TRIUMF, Vancouver, Canada

<sup>5</sup> MEPhI, Moscow, Russia

<sup>6</sup> CERN, ESA

**Corresponding Author:** vladimir.ivantchenko@cern.ch

We will summarize the recent and current activities of the Geant4 working group responsible of the standard package of electromagnetic physics. The major recent activities include an design iteration in energy loss and multiple scattering domain providing “process versus models” approach, and development of the following physics models: multiple scattering, ultra relativistic muon physics, photo-absorption-ionisation model, ion ionisation, optical processes. An automatic acceptance suite of validation of physics is under development. Also we will comment on evolution of the concept of physics list.

**Distributed Computing Services / 321****SPHINX: A Scheduling Middleware for Data Intensive Applications on a Grid****Authors:** J.U. In<sup>1</sup>; L. Chitnis<sup>2</sup>; M. Kulkarni<sup>3</sup>; P. Avery<sup>3</sup>; R. Cavanaugh<sup>3</sup>; S. Ranka<sup>3</sup><sup>1</sup> UNIVERSITY OF FLORIDA<sup>2</sup> University of Florida<sup>3</sup> UNIVERSITY OF FLORIDA**Corresponding Author:** richard.cavanaugh@cern.ch

A grid consists of high-end computational, storage, and network resources that, while known a priori, are dynamic with respect to activity and availability. Efficient co-scheduling of requests to use grid resources must adapt to this dynamic environment while meeting administrative policies. We discuss the necessary requirements of such a scheduler and introduce a distributed framework called SPHINX that schedules complex, data intensive High Energy Physics and Data Mining applications in a grid environment, respecting local and global policies along with a specified level of quality of service. The SPHINX design allows for a number of functional modules and/or distributed services to flexibly schedule workflows representing multiple applications on grids. We present experimental results for SPHINX that effectively utilize existing grid middleware such as monitoring and workflow management/execution systems. These results demonstrate that SPHINX can successfully schedule work across a large number of grid sites that are owned by multiple units in a virtual organization.

**Poster Session 3 - Board: 65 / 322****Extending EGS with SVG for Track Visualization****Authors:** B. White<sup>1</sup>; K. Shrivastava<sup>2</sup>; R. Cowan<sup>3</sup><sup>1</sup> STANFORD LINEAR ACCELERATOR CENTER (SLAC)<sup>2</sup> Stanford Linear Accelerator Center/Stanford University<sup>3</sup> Stanford Linear Accelerator Center**Corresponding Author:** bebo@slac.stanford.edu

The Electron Gamma Shower (EGS) Code System at SLAC is designed to simulate the flow of electrons, positrons and photons through matter at a wide range of energies. It has a large user base among the high-energy physics community and is often used as a teaching tool through a Web interface that allows program input and output. Our work aims to improve the user interaction and shower visualization model of the EGS Web interface. Currently, manipulation of the graphical output (a GIF file) is limited to simple operations like panning and zooming, and each such operation requires server-side calculations. We use SVG (Scalable Vector Graphics) to allow a much richer set of operations, letting users select a track and visualize it with the aid of 3-D rotations, adjustable particle display intensities, and interactive display of the interactions happening over time. A considerable advantage of our method is that once a track is selected for visualization, all further manipulations on that track can be done client-side without requiring server-side calculations. We hence combine the advantages of the SVG format (powerful interaction models over the Web) with those of conventional image formats (file size independent of scene complexity) to allow a composite set of operations for users, and enhance the value of EGS as a pedagogical tool.

**Event Processing / 323****From Geant 3 to Virtual Monte Carlo: Approach and Experience**

**Authors:** Jerome LAURET<sup>1</sup>; M. POTEKHIN<sup>1</sup>; V. Perevoztchikov<sup>1</sup>; Y. FISYAK<sup>1</sup>

<sup>1</sup> *BROOKHAVEN NATIONAL LABORATORY*

**Corresponding Author:** jlauret@bnl.gov

The STAR Collaboration is currently using simulation software based on Geant 3. The emergence of the new Monte Carlo simulation packages, coupled with evolution of both STAR detector and its software, requires a drastic change of the simulation framework.

We see the Virtual Monte Carlo (VMC) approach as providing a layer of abstraction that facilitates such transition.

The VMC platform is a candidate to replace the present legacy software, and help avoid its certain shortcomings, such as the use of a particular algorithmic language to describe the detector geometry. It will also allow us to introduce a more flexible in-memory representation of the geometry.

The Virtual Monte Carlo concept includes a platform-neutral kernel of the application, to the highest degree possible. This kernel is then equipped with interfaces to the modules responsible for simulating the physics of particle propagation, and tracking.

We consider the geometry description classes in the ROOT system (in its latest form known as TGeo classes) as a good choice for the in-memory geometry representation.

We present an application design based on the Virtual Monte Carlo, along with the results of testing, benchmarking and comparison to Geant 3. Internal event representation and IO model will be also discussed.

**Plenary / 324****Improving Standard C++ for the Physics Community**

**Authors:** M. Paterno<sup>1</sup>; W. Brown<sup>1</sup>

<sup>1</sup> *FERMILAB*

**Corresponding Authors:** paterno@fnal.gov, wb@fnal.gov

As Fermilab's representatives to the C++ standardization effort, we have been promoting directions of special interest to the physics community. We here report on selected recent developments toward the next revision of the C++ Standard. Topics will include standardization of random number and special function libraries, as well as core language issues promoting improved run-time performance.

The random number library provides an extensible framework for random number generators. It includes a handful of widely-used and high-quality

random number engines, as well as some of the most widely-used random number distributions. The modular design makes it easy for users to add their own engines, and perhaps more importantly their own distributions, on an equal footing with those in the library.

The special functions library contains many of the commonly-used functions of mathematical physics. These include a variety of cylindrical and spherical Bessel functions, Legendre and associated Legendre functions, hypergeometric and confluent hypergeometric functions, among others.

We also report on an ongoing analysis, and proposal for core language additions, with the goal of improved run-time performance. Current compilers routinely perform inter-procedural flow analysis within a compilation unit. These additions would allow compilers to perform comparable analysis between compilation units, and to optimize code based on their findings.

**Computer Fabrics / 325**

## **Disk storage technology for the LHC T0/T1 centre at CERN**

**Authors:** A. Horvath<sup>1</sup>; H. Meinhard<sup>1</sup>; P. Kelemen<sup>1</sup>

<sup>1</sup> *CERN-IT*

**Corresponding Author:** helge.meinhard@cern.ch

By 2008, the T0/T1 centre for the LHC at CERN is estimated to use about 5000 TB of disk storage. This is a very significant increase over the about 250 TB running now. In order to be affordable, the chosen technology must provide the required performance and at the same time be cost-effective and easy to operate and use.

We will present an analysis of the cost (both in terms of material and personnel) of the current implementation (network-attached storage), and then describe detailed performance studies with hardware currently in use at CERN in different configurations of filesystems on software or hardware RAID arrays over disks. Alternative technologies that have been evaluated by CERN in varying depth (such as arrays of SATA disks with a Fiber Channel uplink, distributed disk storage across worker nodes, iSCSI solutions, SANFS, ...) will be discussed. We will conclude with an outlook of the next steps to be taken at CERN towards defining the future disk storage model.

**Poster Session 2 - Board: 18 / 327**

## **An intelligent resource selection system based on neural network for optimal application performance in a grid environment**

**Authors:** G. PISCITELLI<sup>None</sup>; M. Castellano<sup>1</sup>; T. Coviello<sup>1</sup>

<sup>1</sup> *DEE - POLITECNICO DI BARI, V. ORABONA, 4, 70125 - BARI, ITALY*

**Corresponding Author:** castellano@deemail.poliba.it

Grid computing is a large scale geographically distributed and heterogeneous system that provides a common platform for running different grid enabled applications. As each application has different characteristics and requirements, it is a difficult task to develop a scheduling strategy able to achieve optimal performance because application-specific and dynamic system status have to be taken into account.

Moreover it may be possible to obtain optimal performance for multiple application simultaneously using a single scheduler. Hence in a lot of cases the application scheduling strategy is assigned to an expert application user who provides a ranking criterion for selecting the best computational element on a set of available resources. Such criteria are based on user perception of system capabilities and knowledge about the features and requirements of his application.

In this paper an intelligent mechanism has been both implemented and evaluated to select the best computational resource in a grid environment from the application viewpoint.

A neural network based system has been used to capture automatically the knowledge of a grid application expert user. The system scalability problem is also tackled and a preliminary solution based on sorting algorithm is discussed. The aim is to allow a common grid application user to benefit of this expertise.

**Distributed Computing Services / 328**

## **The Next Generation Root File Server**

**Authors:** A. Dorigo<sup>1</sup>; A. Hanushevsky<sup>2</sup>; F. Furano<sup>1</sup>

<sup>1</sup> *INFN, Padova*

<sup>2</sup> *SLAC*

As the BaBar experiment shifted its computing model to a ROOT-based framework, we undertook the development of a high-performance file server as the basis for a fault-tolerant storage environment whose ultimate goal was to minimize job failures due to server failures. Capitalizing on our five years of experience with extending Objectivity's Advanced Multithreaded Server (AMS), elements were added to remove as many obstacles to server performance and fault-tolerance as possible. The final outcome was *xrootd*, upwardly and downwardly compatible with the current file server, *rootd*. This paper describes the essential protocol elements that make high performance and fault-tolerance possible; including asynchronous parallel requests, stream multiplexing, data pre-fetch, automatic data segmenting, and the framework for a structured peer-to-peer storage model that allows massive server scaling and client recovery from multiple failures. The internal architecture of the server is also described to explain how high performance was maintained and full compatibility was achieved. Now in production at Stanford Linear Accelerator Center, Rutherford Appleton Laboratory (RAL), INFN, and IN2P3; *xrootd* has shown that our design provides what we set out to achieve. The *xrootd* server is now part of the standard ROOT distribution so that other experiments can benefit from this data serving model within a standard HEP event analysis framework.

**Online Computing / 329**

## The PHENIX Event Builder

**Author:** D. Winter<sup>1</sup>

<sup>1</sup> COLUMBIA UNIVERSITY

**Corresponding Author:** winter@nevis.columbia.edu

The PHENIX detector consists of 14 detector subsystems. It is designed such that individual subsystems can be read out independently in parallel as well as a single unit. The DAQ used to read the detector is a highly-pipelined parallel system. Because PHENIX is interested in rare physics events, the DAQ is required to have a fast trigger, deep buffering, and very high bandwidth.

The PHENIX Event Builder is a critical part of the back-end of the PHENIX DAQ. It is responsible for assembling event fragments from each subsystem into complete events ready for archiving. It allows subsystems to be read out either in parallel or simultaneously and supports a high rate of archiving. In addition, it implements an environment where Level-2 trigger algorithms may be optionally executed, providing the ability to tag and/or filter rare physics events.

The Event Builder is a set of three Windows NT/2000 multithreaded executables that run on a farm of over 100 dual-cpu 1U servers. All control and data messaging is transported over a Foundry Layer2/3 Gigabit switch. Capable of recording a wide range of event sizes from central Au-Au to p-p interactions, data archiving rates of over 400 MB/s at 2 KHz event rates have been achieved in the recent Run 4 at RHIC. Further improvements in performance are expected from migrating to Linux for Run 5.

The PHENIX Event Builder design and implementation, as well as performance and plans for future development, will be discussed.

**Computer Fabrics / 330**

## A Regional Analysis Center at the University of Florida

**Authors:** C. Prescott<sup>1</sup>; D. Bourilkov<sup>1</sup>; J. Rodriguez<sup>1</sup>; P. Avery<sup>1</sup>; R. Cavanaugh<sup>1</sup>; Y. Fu<sup>1</sup>

<sup>1</sup> UNIVERSITY OF FLORIDA

**Corresponding Author:** richard.cavanaugh@cern.ch

The High Energy Physics Group at the University of Florida is involved in a variety of projects ranging from High Energy Experiments at hadron and electron positron colliders to cutting edge computer science experiments focused on grid computing. In support of these activities members of the Florida group have developed and deployed a local computational facility which consists of several service nodes, computational clusters and disk storage services. The resources contribute collectively or individually to a variety of production and development activities such as the UFlorida Tier2 center for the CMS experiment at the Large Hadron Collider (LHC), Monte Carlo production for the CDF experiment at Fermi Lab, the CLEO experiment, and research on grid computing for the GriPhyN and iVDGL projects. The entire collection of servers, clusters and storage services is managed as a single facility using the ROCKS cluster management system. Managing the facility as a single centrally managed system enhances our ability to relocate and reconfigure the resources as necessary in support of both research and production activities. In this paper we describe the architecture deployed, including details on our local implementation of the ROCKS systems, how this simplifies the maintenance and administration of the facility and finally the advantages and disadvantages of using such a scheme to manage a modest size facility.



Online Computing / 331

## Integration of ATLAS Software in the Combined Beam Test

Author: M. Dobson<sup>1</sup>

<sup>1</sup> CERN

Corresponding Author: marc.dobson@cern.ch

The ATLAS collaboration had a Combined Beam Test from May until October 2004. Collection and analysis of data required integration of several software systems that are developed as prototypes for the ATLAS experiment, due to start in 2007. Eleven different detector technologies were integrated with the Data Acquisition system and were taking data synchronously. The DAQ was integrated with the High Level Trigger software, which will perform online selection of ATLAS events. The data quality was monitored at various stages of the Trigger and DAQ chain. The data was stored in a format foreseen for ATLAS and was analyzed using a prototype of the experiments' offline software, using the Athena framework. Parameters recorded by the Detector Control System were recorded in a prototype of the ATLAS Conditions Data Base and were made available for the offline analysis of the collected event data. The combined beam test provided a unique opportunity to integrate and to test the prototype of ATLAS online and offline software in its complete functionality.

Poster Session 2 - Board: 30 / 332

## Production data export and archiving system for new data format of the BaBar experiment.

Authors: A. Hasan<sup>1</sup>; A. Trunov<sup>1</sup>; T. Azemoon<sup>1</sup>; W. Kroeger<sup>1</sup>

<sup>1</sup> SLAC

Corresponding Author: artem@slac.stanford.edu

For The BaBar Computing Group

BaBar has recently moved away from using Objectivity/DB for its event store towards a ROOT-based event store. Data in the new format is produced at about 20 institutions worldwide as well as at SLAC. Among new challenges are the organization of data export from remote institutions, archival at SLAC and making the data visible to users for analysis and import to their own institutions.

The new system is designed to be scalable, easily configurable on the client and server side and adaptive to server load. It's integrated to work with SLAC's mass storage system (HPSS) and with the xrootd service. Design, implementation and experience with new system, as well as future development is discussed in this article.

Event Processing / 333

## A statistical toolkit for data analysis

**Authors:** A. Pfeiffer<sup>1</sup>; A. Ribon<sup>1</sup>; B. Mascialino<sup>2</sup>; M.G. Pia<sup>3</sup>; P. VIARENGO<sup>4</sup>; S. Donadio<sup>2</sup>; S. Guatelli<sup>5</sup>

<sup>1</sup> CERN

<sup>2</sup> INFN Genova

<sup>3</sup> INFN GENOVA

<sup>4</sup> IST, Genova, Italy

<sup>5</sup> INFN Genova, Italy

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

Statistical methods play a significant role throughout the life-cycle of HEP experiments, being an essential component of physics analysis. We present a project in progress for the development of an object-oriented software toolkit for statistical data analysis. More in particular, the Statistical Comparison component of the toolkit provides algorithms for the comparison of data distributions in a variety of use cases typical of HEP experiments, as regression testing (in various phases of the software life-cycle), validation of simulation through comparison to experimental data, comparison of expected versus reconstructed distributions, comparison of data from different sources - such as different sets of experimental data, or experimental with respect to theoretical distributions. The toolkit contains a variety of goodness-of-fit tests, from chi-squared to Kolmogorov-Smirnov, to less known, but generally much more powerful tests such as Anderson-Darling, Cramer-von Mises, Kuiper, Tiku etc.

Thanks to the component-based design and the usage of the standard AIDA interfaces, this tool can be used by other data analysis systems or integrated in experimental software frameworks. We present the architecture of the system, the statistics methods implemented and some results of its applications to the comparison of Geant4 simulations with respect to experiment.

**Distributed Computing Services / 334**

## **Production mode Data-Replication framework in STAR using the HRM Grid**

**Authors:** A. Shoshani<sup>1</sup>; A. Sim<sup>1</sup>; D. Olson<sup>1</sup>; E. Hjort<sup>2</sup>; Jerome LAURET<sup>3</sup>

<sup>1</sup> Lawrence Berkeley Laboratory

<sup>2</sup> LAWRENCE BERKELEY LABORATORY

<sup>3</sup> BROOKHAVEN NATIONAL LABORATORY

**Corresponding Author:** elhjort@lbl.gov

The STAR experiment utilizes two major computing facilities for its data processing needs - the RCF at Brookhaven and the PDSF at LBNL/NERSC. The sharing of data between these facilities utilizes data grid services for file replication, and the deployment of these services was accomplished in conjunction with the Particle Physics Data Grid (PPDG). For STAR's 2004 run it will be necessary to replicate ~100 TB. The file replication is based on Hierarchical Resource Managers (HRMs) along with Globus tools for security (GSI) and data transport (GridFTP). HRMs are grid middleware developed by the Scientific Data Management group at LBNL, and STAR file replication consists of an HRM interfaced to HPSS at each site with GridFTP transfers between the HRMs. Each site also has its own installation of the STAR file and metadata catalog, which is implemented in MySQL. Queries to the catalogs are used to generate file transfer requests. Single requests typically consist of

many thousands of files with a volume of hundreds of GBs. The HRMs implement a plugin to a Replica Registration Service (or RRS) which is utilized for automatic registration of new files as they are successfully transferred across sites. This allows STAR users immediate use of the distributed data. Data transfer statistics and system architecture will be presented.

Poster Session 2 - Board: 13 / 335

## Tools for GRID deployment of CDF offline and SAM data handling systems for Summer 2004 computing.

**Authors:** A. Baranovski<sup>1</sup>; A. Kreymer<sup>2</sup>; A. Lyon<sup>3</sup>; A. Sill<sup>4</sup>; F. Ratnikov<sup>5</sup>; G. Garzoglio<sup>3</sup>; I. Terekhov<sup>1</sup>; J. Trumbo<sup>1</sup>; L. Loebel Carpenter<sup>1</sup>; M. Burgon-Lyon<sup>6</sup>; M. Leslie<sup>7</sup>; R. Herber<sup>3</sup>; R. Illingworth<sup>3</sup>; R. Kennedy<sup>3</sup>; R. St.Denis<sup>6</sup>; S. Belforte<sup>8</sup>; S. Stonjek<sup>1</sup>; S. Veseli<sup>1</sup>; S. White<sup>1</sup>; U. Kerzel<sup>9</sup>; V. Bartsch<sup>7</sup>; W. Merritt<sup>3</sup>

<sup>1</sup> *Fermilab*

<sup>2</sup> *FERMILAB*

<sup>3</sup> *FERMI NATIONAL ACCELERATOR LABORATORY*

<sup>4</sup> *Texas Tech. University*

<sup>5</sup> *Rutgers University*

<sup>6</sup> *Glasgow University*

<sup>7</sup> *Oxford University*

<sup>8</sup> *INFN*

<sup>9</sup> *Karlsruhe University*

**Corresponding Author:** kreymer@fnal.gov

The Fermilab CDF Run-II experiment is now providing official support for remote computing, expanding this to about 1/4 of the total CDF computing during the Summer of 2004.

I will discuss in detail the extensions to CDF software distribution and configuration tools and procedures, in support of CDF GRID/DCAF computing for Summer 2004. We face the challenge of unreliable networks, time differences, and remote managers with little experience with this particular software.

We have made the first deployment of the SAM data handling system outside its original home in the D0 experiment.

We have deployed to about 20 remote CDF sites.

We have created light weight testing and monitoring tools to assure that these sites are in fact functional when installed.

We are distributing and configuring both client code within CDF code releases, and the SAM servers to which the clients connect.

Procedures which once took days are now performed in minutes.

These tools can be used to install SAM servers for D0 and other experiments.

Networks permitting, we will give a live SAM installation demonstration.

We have separated the data handling components from the main CDF offline code releases by means of shared libraries, permitting live upgrades to otherwise frozen code.

We now use a special 'development lite' release to ensure that all sites have the latest tools available.

We have put substantial effort into revision control, so that essentially all active CDF sites are running exactly the same code.

Poster Session 3 - Board: 43 / 336

## Comparative study of the power of goodness-of-fit algorithms

**Authors:** A. Pfeiffer<sup>1</sup>; A. Ribon<sup>1</sup>; B. Mascialino<sup>2</sup>; M.G. Pia<sup>3</sup>; S. Donadio<sup>2</sup>; S. Guatelli<sup>4</sup>

<sup>1</sup> CERN

<sup>2</sup> INFN Genova

<sup>3</sup> INFN GENOVA

<sup>4</sup> INFN Genova, Italy

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

A Toolkit for Statistical Data Analysis has been recently released. Thanks to this novel software system, for the first time an ample set of sophisticated algorithms for the comparison of data distributions (goodness of fit tests) is made available to the High Energy Physics community in an open source product. The statistical algorithms implemented belong to two sets, for the comparison of binned and unbinned distributions respectively; they include the Chi-squared Test, the Kolmogorov-Smirnov Test, the Kuiper Test, the Goodman Test, the Anderson-Darling Test, the Fisz-Cramer-von Mises test, the Tiku Test.

Since the Toolkit provides the user a wide choice of algorithms, it is important to evaluate them comparatively and to estimate their power, to provide guidance to the users about the selection of the most appropriate algorithm for a given use case.

We present a study of the power of a variety of mathematical algorithms implemented in the Toolkit. The study is performed by evaluating the behaviour of the various tests in a set of well identified use cases relevant to data analysis applications. To our knowledge, such a comparative study of the power of goodness of fit algorithms has never been performed previously.

Poster Session 2 - Board: 31 / 337

## Alibaba: A heterogeneous grid-based job submission system used by the BaBar experiment

**Authors:** A. Forti<sup>1</sup>; A. McNab<sup>2</sup>; M. Jones<sup>1</sup>; R. Barlow<sup>1</sup>

<sup>1</sup> Manchester University

<sup>2</sup> UNIVERSITY OF MANCHESTER

**Corresponding Authors:** mike.jones@man.ac.uk, roger.barlow@man.ac.uk

The BaBar experiment has accumulated many terabytes of data on particle physics reactions, accessed by a community of hundreds of users.

Typical analysis tasks are C++ programs, individually written by the user, using shared templates and libraries. The resources have outgrown a single platform and a distributed computing model is needed. The grid provides the natural toolset. However, in contrast to the LHC experiments, BaBar has an existing user community with an existing non-Grid usage pattern, and providing users with an acceptable evolution presents a challenge.

The 'Alibaba' system, developed as part of the UK GridPP project, provides the user with a familiar command line environment. It draws on the existing global file systems employed and understood by the current user base. The main difference is that they submit jobs with a 'gsub' command that looks and feels like the familiar 'qsub'. However it enables them to submit jobs to computer systems at different institutions, with minimal requirements on the remote sites. Web based job monitoring is also provided. The problems and features (the input and output sandboxes, authentication, data location) and their solutions are described.

**Distributed Computing Services / 338**

## **BaBar Bookkeeping - a distributed meta-data catalog of the BaBar event store.**

**Authors:** A. Ceseracci<sup>1</sup>; A. Forti<sup>2</sup>; D. Bukin<sup>3</sup>; D. Hutchcroft<sup>4</sup>; D. Kovalskyi<sup>5</sup>; D. Smith<sup>6</sup>; G. Dubois-Felsmann<sup>7</sup>; P. Jackson<sup>8</sup>; T. Adye<sup>9</sup>; W. Roethel<sup>10</sup>

<sup>1</sup> INFN SEZIONE DI PADOVA

<sup>2</sup> UNIVERSITY OF MANCHESTER

<sup>3</sup> BUDKER

<sup>4</sup> UNIVERSITY OF LIVERPOOL

<sup>5</sup> UNIVERSITY OF MARYLAND

<sup>6</sup> STANFORD LINEAR ACCELERATOR CENTER

<sup>7</sup> Caltech

<sup>8</sup> UNIVERSITY OF LONDON, ROYAL HOLLOWAY

<sup>9</sup> RUTHERFORD APPLETON LABORATORY

<sup>10</sup> UC, IRVINE

**Corresponding Author:** douglas@slac.stanford.edu

The BaBar experiment has migrated its event store from an objectivity based system to a system using ROOT-files, and along with this has developed a new bookkeeping design. This bookkeeping now combines data production, quality control, event store inventory, distribution of BaBar data to sites and user analysis in one central place, and is based on collections of data stored as ROOT-files. These collections are grouped into pre-determined datasets, which define subsets of BaBar data to be used in analysis. Datasets are updated automatically to contain at any times the most up-to-date BaBar data. Local mirrors of the bookkeeping database can be used with the data distribution features to import collections and maintain local event stores containing subsets of the available BaBar data. The bookkeeping system is scalable and supports sites containing all available data and hundreds of users down to the single user with a laptop. Oracle and MySQL relational databases are supported in its use, and sites can choose which to support. Database mirrors in the bookkeeping system can be accessed over network, which allows to browse local inventories from remote sites. This book keeping system has been in active use in BaBar since early this year, and the scope of its use along with technologies developed to keep it working will be presented.

**Distributed Computing Systems and Experiences / 339****BaBar simulation production - A millennium of work in under a year**

**Authors:** C. Bozzi<sup>1</sup>; D. Smith<sup>2</sup>; F. Blanc<sup>3</sup>

<sup>1</sup> *INFN SEZIONE DI FERRARA*

<sup>2</sup> *STANFORD LINEAR ACCELERATOR CENTER*

<sup>3</sup> *UNIVERSITY OF COLORADO*

**Corresponding Author:** douglas@slac.stanford.edu

for the BaBar Computing Group.

The analysis of the BaBar experiment requires many times the measured data to be produced in simulation. This requirement has resulted in one of the largest distributed computing projects ever completed.

The latest round of simulation for BaBar started in early 2003, and completed in early 2004, and encompassed over 1 million jobs, and over 2.2 billion events. By the end of the production cycle over 2 dozen different computing centers and nearly 1.5 thousand cpus were in constant use in North America and Europe. The whole effort was managed from a central database at SLAC, with real-time updates of the status of all jobs. Utilities were developed to tie together production with many different batch systems, and with different needs for security.

The produced data was automatically transferred to SLAC for use and distribution to analysis sites. The system developed to manage this effort was a combination of web and database applications, and command line utilities. The technologies used to complete this effort along with its complete scope will be presented.

**Core Software / 340****A Dynamically Reconfigurable Data Stream Processing System**

**Authors:** J. Nogiec<sup>1</sup>; K Trombly-Freytag<sup>2</sup>

<sup>1</sup> *FERMI NATIONAL ACCELERATOR LABORATORY*

<sup>2</sup> *Fermilab*

**Corresponding Author:** nogiec@fnal.gov

The paper describes a component-based framework for data stream processing that allows for configuration, tailoring, and run-time system reconfiguration. The system's architecture is based on a pipes and filters pattern, where data is passed through routes between components. Components process data and add, substitute, and/or remove named data items from a data stream. They can also manipulate data streams by buffering data, compressing/decompressing individual streams, and combining, splitting, or synchronizing multiple data streams. Configurable general-purpose filters for manipulating streams, visualizing data, persisting data, and reading data from various standard data sources are supplemented with many application specific filters, such as DSP, scripting, or instrumentation-specific components. A network of pipes and filters can be dynamically reconfigured at run-time, in response to a preplanned sequence of processing steps, operator intervention, or a change in one or more data streams. Four distinctive methods supporting reconfiguration are provided by the framework: modification of data

routes, management of components' activity states, triggering processing based on the content of the data, or the use of source addressing in components. The framework can be used to build static data stream processing applications such as monitoring or data acquisition systems as well as self-adjusting systems that would adapt their processing algorithm, presentation layer, or data persistency layer in response to changes in input data streams.

**Poster Session 2 - Board: 42 / 341**

## **Patriot: Physics Archives and Tools required to Investigate Our Theories**

**Author:** S. Mrenna<sup>1</sup>

<sup>1</sup> *FERMILAB*

**Corresponding Author:** mrenna@fnal.gov

PATRIOT is a project that aims to provide better predictions of physics events for the high-Pt physics program of Run2 at the Tevatron collider.

Central to Patriot is an enstore or mass storage repository for files describing the high-Pt physics predictions. These are typically stored as StdHep files which can be handled by CDF and D0 and run through detector and triggering simulations. The definition of these datasets in the CDF and D0 data handling system SAM is under way.

Patriot relies heavily on a new generation of Monte Carlo tools (such as MadEvent, Alpgen, Grappa, CompHEP, etc.) to calculate the hard structure of high-Pt events and the more venerable event generators (Pythia and Herwig) to make particle level predictions.

An early informational database, describing the types of data files stored in Patriot, already exists. A new database is under development.

In parallel with PATRIOT, we wish to develop the QCD tools that describe the detailed properties of high-Pt events. Some of the essential features of particle-level events must be described by non-perturbative functions, whose form is often constrained by theory, but which must be ultimately tuned to data.

**Poster Session 3 - Board: 44 / 342**

## **Precision validation of Geant4 electromagnetic physics**

**Author:** M.G. Pia<sup>1</sup>

**Co-authors:** A. Mantero <sup>2</sup>; B. Mascialino <sup>2</sup>; L. Pandola <sup>3</sup>; M. Maire <sup>4</sup>; P. Cirrone <sup>5</sup>; S. Donadio <sup>2</sup>; S. Guatelli <sup>2</sup>; S. Parlati <sup>3</sup>; V. Ivanchenko <sup>6</sup>

<sup>1</sup> *INFN GENOVA*

<sup>2</sup> *INFN Genova*

<sup>3</sup> *INFN LNGS*

<sup>4</sup> *LAPP*

<sup>5</sup> *INFN LNS*

<sup>6</sup> *Budker Institute*

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

The Geant4 Toolkit provides an ample set of alternative and complementary physics models to handle the electromagnetic interactions of leptons, photons, charged hadrons and ions.

Because of the critical role often played by simulation in the experimental design and physics analysis, an accurate validation of the physics models implemented in Geant4 is essential, down to the quantitative understanding of the accuracy of their microscopic features.

Results from a series of detailed tests with respect to well established reference data sources and experiments are presented, focusing in particular

on the precision validation of the microscopic components of Geant4 physics, such as cross sections and angular distributions, provided in the various alternative physics models of Geant4 electromagnetic packages.

The validation of Geant4 physics is performed by means of quantitative evaluations of the comparison of Geant4 models to reference data are presented, making use of statistical analysis algorithms to estimate the compatibility of simulated and experimental distributions.

Such precision tests are especially relevant for critical applications of simulation models, such as tracking detectors, neutrino and other astroparticle experiments, medical physics

**Computer Fabrics / 343**

## **Chimera - a new, fast, extensible and Grid enabled namespace service**

**Authors:** M. Gasthuber<sup>1</sup>; P. Fuhrmann<sup>1</sup>; T. Mkrтчhyan<sup>1</sup>

<sup>1</sup> *DESY*

After successful implementation and deployment of the dCache system over the last years, one of the additional required services, the namespace service, is faced additional and completely new requirements. Most of these are caused by scaling the system, the integration with Grid services and the need for redundant (high availability) configurations. The existing system, having only an NFSv2 access path, is easy to understand and well accepted by the users. This single 'access path' limits data management task to make use of classical tools like 'find', 'ls' and others. This is intuitiv for most users, but failed while dealing with millions of entries (files) and more sophisticated organizational schemes (metadata). The new system should support a native programmable interface (deep coupled, but fast), the 'classical' NFS path (now version 3 at least),

a dCache native access and the SQL path allowing any type of metadata to be used in complex queries. Extensions with other 'access paths' will be possible. Based on the experience with the current system we highlight on the following requirements:

- large file support (64 Bit) + large number of files (> 10<sup>8</sup>)
- fast
- Platform independents (runtime + persistent objects)
- Grid name service integration
- custom dCache integration



- redundant, high available runtime configurations (concurrent backup etc.)
- user usable metadata (store and query)
- ACL support
- pluggable authentication (e.g. GSSAPI)
- external processes can register for namespace events (e.g. removal/creation of files)

The presentation will show a detailed analysis of the requirements, the chosen design and selection of existing components. The current schedule should allow to show the first prototype results.

## Distributed Computing Services / 345

### Storage Resource Managers at Brookhaven

**Authors:** O. Rind<sup>1</sup>; R. Popescu<sup>1</sup>; S. O'Hare<sup>1</sup>; Z. Liu<sup>1</sup>

<sup>1</sup> *BROOKHAVEN NATIONAL LABORATORY*

**Corresponding Author:** rind@bnl.gov

Providing Grid applications with effective access to large volumes of data residing on a multitude of storage systems with very different characteristics prompted the introduction of storage resource managers (SRM). Their purpose is to provide consistent and efficient wide-area access to storage resources unconstrained by their particular implementation (tape, large disk arrays, dispersed small disks). To assess their viability in the context of the US Atlas Tier 1 facility at Brookhaven, two implementations of SRM were tested: dCache (FNAL/DESY joint project) and HRM/DRM (NERSC Berkeley). Both systems included a connection to the local HPSS mass data store providing Grid access to the main tape repository. In addition, dCache offered storage aggregation of dispersed small disks (local drives on computing farm nodes). An overview of our experience with both systems will be presented, including details about configurations, performance, inter-site transfers, interoperability and limitations.

## Computer Fabrics / 346

### Lattice QCD Clusters at Fermilab

**Authors:** A. Singh<sup>1</sup>; D. Holmgren<sup>2</sup>; J. Simone<sup>1</sup>; P. Mackenzie<sup>1</sup>

<sup>1</sup> *Fermilab*

<sup>2</sup> *FERMILAB*

**Corresponding Author:** djholm@fnal.gov

As part of the DOE SciDAC "National Infrastructure for Lattice Gauge Computing" project, Fermilab builds and operates production clusters for lattice QCD simulations. We currently operate three clusters: a 128-node dual Xeon Myrinet cluster, a 128-node Pentium 4E Myrinet cluster, and a 32-node dual Xeon Infiniband cluster. We will discuss the operation of these systems and examine their performance in detail. We will describe the uniform user runtime environment emerging from the SciDAC collaboration.

The design of lattice QCD clusters requires careful attention towards balancing memory bandwidth, floating point throughput, and network performance. We will discuss our investigations of various commodity processors, including Pentium 4E, Xeon, Itanium2, Opteron, and PPC970, in terms of their suitability for building balanced QCD clusters. We will also discuss our early experiences with the emerging Infiniband and PCI Express architectures. Finally, we will examine historical trends in price to performance ratios of lattice QCD clusters, and we will present our predictions and plans for future clusters.

**Event Processing / 347**

## The new BaBar Analysis Model

**Authors:** D. Brown<sup>1</sup>; D. Lange<sup>2</sup>; E. Charles<sup>1</sup>; G. De Nardo<sup>3</sup>; G. Finocchiaro<sup>4</sup>; L. Wilden<sup>5</sup>; P. Elmer<sup>6</sup>

<sup>1</sup> *LAWRENCE BERKELEY NATIONAL LAB*

<sup>2</sup> *Lawrence Livermore National Lab*

<sup>3</sup> *INFN Naples*

<sup>4</sup> *INFN Frascati*

<sup>5</sup> *NIKHEV Amsterdam*

<sup>6</sup> *Princeton University*

**Corresponding Author:** dave\_brown@lbl.gov

This talk will describe the new analysis computing model deployed by BaBar over the past year. The new model was designed to better support the current and future needs of physicists analyzing data, and to improve BaBar's analysis computing efficiency.

The use of RootIO in the new model is described in other talks.

Babar's new analysis data content format contains both high and low level information, allowing physicists to pick a tradeoff between speed and precision/flexibility appropriate to their analysis.

The new format is customizable, allowing physicists to create analysis-specific content using simple and familiar tools.

Babar's new analysis processing model involves selecting events according to their physics content, and writing them together with analysis-customized content to dedicated output streams.

Currently 120 such 'skims' are written as part of a periodic central processing cycle. Skims can be further reduced and customized as well as queried interactively in root-based applications. Skims and subskims retain links back to the original full event information.

This processing model eliminates the need for large tuple production efforts by physicists.

The entire BaBar data sample is available in the new format, and the new model has been used to produce physics results presented at the summer HEP conferences.

We will also present reactions from the BaBar analysis community, and describe the issues that arose in deploying the new model.

**Wide Area Networking / 348**

## Wide Area Network Monitoring system for HEP experiments at Fermilab

**Authors:** C. Logg<sup>1</sup>; L. Cottrell<sup>1</sup>; M. Grigoriev<sup>2</sup>

<sup>1</sup> *SLAC, USA*

<sup>2</sup> *FERMILAB, USA*

**Corresponding Author:** maxim@fnal.gov

Large, distributed HEP collaborations, such as D0, CDF and US-CMS, depend on stable and robust network paths between major world research centers. The evolving emphasis on data and compute Grids increases the reliance on network performance.

FermiLab's experimental groups and network support personnel identified a critical need for WAN monitoring to ensure the quality and efficient utilization of such network paths. This has led to the development of the Network Monitoring system we will present in this paper.

The system evolved from the IEPM-BW project, started at SLAC two years ago.

At Fermilab it has developed into a fully functional infrastructure with bi-directional active network probes and path characterizations.

It is based on the Iperf achievable throughput tool, Ping and Synack to test ICMP/TCP connectivity, Pipechar and Traceroute to test, compare and report hop-by-hop network path characterization, and real file transfer performance by BBFTP and GridFTP. The Monitoring system has an extensive web-interface and all the data is available through standalone SOAP web services or by a MonaLISA client.

Also in this paper we will present a case study of network path asymmetry and abnormal performance between FNAL and SDSC which was discovered and resolved by utilizing the Network Monitoring system.

**Poster Session 2 - Board: 46 / 349**

## **Distributed computing and oncological radiotherapy: technology transfer from HEP and experience with prototype systems**

**Authors:** F. Foppiano<sup>1</sup>; J. Moscicki<sup>2</sup>; M.G. Pia<sup>3</sup>; S. Guatelli<sup>4</sup>

<sup>1</sup> *IST Genova*

<sup>2</sup> *CERN*

<sup>3</sup> *INFN GENOVA*

<sup>4</sup> *INFN Genova, Italy*

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

We show how nowadays it is possible to achieve the goal of accuracy and fast computation response in radiotherapeutic dosimetry using Monte Carlo methods, together with a distributed computing model.

Monte Carlo methods have never been used in clinical practice because, even if they are more accurate than available commercial software, the calculation time needed to accumulate sufficient statistics is too long for a realistic use in radiotherapeutic treatment.

We present a complete, fully functional prototype dosimetric system for radiotherapy, integrating various components based on HEP software systems: a Geant4-based simulation, an AIDA-based dosimetric analysis, a web-based user interface, and distributed processing either on a local computing farm or on geographically spread nodes.

The performance of the dosimetric system has been studied in three execution modes: sequential on a single dedicated machine, parallel on a dedicated computing farm, parallel on a grid test-bed. An intermediate software layer, the DIANE

system, makes the three execution modes completely transparent to the user, allowing to use the same code in any of the three configurations. Thanks to the integration in a grid environment, any hospital, even small ones or in less wealthy countries, that could not afford the high costs of commercial treatment planning software, may get the chance of using advanced software tools for oncological therapy, by accessing distributed computing resources, shared with other hospitals and institutes belonging to the same virtual organization

**Poster Session 2 - Board: 56 / 350**

## The BABAR Analysis Task Manager

**Author:** W. Roethel<sup>1</sup>

**Co-authors:** Alessandra Forti <sup>2</sup>; Antonio Ceseracciu <sup>3</sup>; David E. Hutchcroft <sup>4</sup>; Dmitry Bukin <sup>5</sup>; Dmytro Kovalskyi <sup>6</sup>; Douglas Smith <sup>7</sup>; Gregory Dubois-Felsmann <sup>8</sup>; Paul Stephen Jackson <sup>9</sup>; Tim Adye <sup>10</sup>

<sup>1</sup> *University of California, Irvine*

<sup>2</sup> *University of Manchester*

<sup>3</sup> *Universita' di Padova*

<sup>4</sup> *University of Liverpool*

<sup>5</sup> *Budker Institute of Nuclear Physics*

<sup>6</sup> *University of Maryland*

<sup>7</sup> *Stanford Linear Accelerator Center*

<sup>8</sup> *California Institute of Technology*

<sup>9</sup> *University of London, Royal Holloway and Bedford New College*

<sup>10</sup> *Rutherford Appleton Laboratory*

**Corresponding Authors:** douglas@slac.stanford.edu, roethel@slac.stanford.edu

The new BaBar bookkeeping system comes with tools to directly support data analysis tasks. This Task Manager system acts as an interface between datasets defined in the bookkeeping system, which are used as input to analyzes, and the offline analysis framework. The Task Manager organizes the processing of the data by creating specific jobs to be either submitted to a batch system, or run in the background on a local desktop, or laptop. The current system has been designed to support pbs and lsf batch systems. Changes to defined datasets due production is directly supported by the Task Manager, where new collections that add to a dataset or replace other collections are automatically detected, allowing an analysis at any time to be up-to-date with the latest available data. The output of tasks, whether new data collections, ntuple/hbook files, or text files, can be put back into a collections bookkeeping system or stored in the private Task Manager database. Currently MySQL and Oracle relational databases are supported. The BABAR Task Manager has been in use for data production since January this year, and the schema of the working system will be presented.

**Event Processing / 351**

## Precision electromagnetic physics in Geant4: the atomic relaxation models

**Authors:** A. Mantero<sup>1</sup>; M.G. Pia<sup>2</sup>; S. Guatelli<sup>3</sup>; S. Saliceti<sup>1</sup>

<sup>1</sup> *INFN Genova*

<sup>2</sup> *INFN GENOVA*

<sup>3</sup> *INFN Genova, Italy*

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

Various experimental configurations - such as, for instance, some gaseous detectors, require a high precision simulation of electromagnetic physics processes, accounting not only for the primary interactions of particles with matter, but also capable of describing the secondary effects deriving from the de-excitation of atoms, where primary collisions may have created vacancies. The Geant4 Simulation Toolkit encompasses a set of models to handle the atomic relaxation induced by the photoelectric effect, Compton scattering and ionization, with the production of X-ray fluorescence and of Auger electrons.

We describe the physics models implemented in Geant4 to handle the atomic relaxation, the object-oriented design of the software and the validation of the models with respect to test beam data.

In particular, we present a novel development of an original model for particle induced X-ray emission, to be released for the first time in the summer of 2004.

We illustrate applications of Geant4 atomic relaxation models for physics reach studies in a real-life experimental context

**Poster Session 3 - Board: 45 / 352**

## Experience with the Unified Process

**Author:** M.G. Pia<sup>1</sup>

<sup>1</sup> *INFN GENOVA*

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

The adoption of a rigorous software process is well known to represent a key factor for the quality of the software product and the most effective usage of the human resources available to a software project.

The Unified Process, in particular its commercial packaging known as the RUP (Rational Unified Process) has been one of the most widely used software process models in the software industry for a number of years.

We present the application of the Unified Process and of the RUP to a variety of software projects in the High Energy Physics environment. We illustrate how the UP/RUP provide a flexible process framework, that can be tailored to the different needs of individual software projects. We describe the experience with different approaches (top-down and bottom-up) to the implementation of the process in software organizations.

We document a critical analysis of the effects of the adoption of the UP/RUP, and discuss the relative benefits of the public (UP) and commercial (RUP) versions of the process.

Finally, we discuss the curious results of the effects of applying the RUP to a software development environment that is not aware of adopting it

**Core Software / 353**

## The Offline Framework of the Pierre Auger Observatory

**Authors:** L. Nellen<sup>1</sup>; L. Prado Jr<sup>2</sup>; S. Argiro<sup>3</sup>; T. Paul<sup>4</sup>; T. Porter<sup>5</sup>

<sup>1</sup> *I. DE CIENCIAS NUCLEARES, UNAM*

<sup>2</sup> *State University of Campinas*

<sup>3</sup> *University of Torino*

<sup>4</sup> *Northeastern University*

<sup>5</sup> *Louisiana State University*

The Pierre Auger Observatory is designed to unveil the nature and the origin of the highest energy cosmic rays. Two sites, one currently under construction in Argentina, and another pending in the Northern hemisphere, will observe extensive air showers using a hybrid detector comprising a ground array of 1600 water Cerenkov tanks overlooked by four atmospheric fluorescence detectors. Though the computing demands of the experiment are less severe than those of traditional high energy physics experiments in terms of data volume and detector complexity, the large geographically dispersed collaboration and the heterogeneous set of simulation and reconstruction requirements confronts the offline software with some special challenges.

We have designed and implemented a framework to allow collaborators to contribute algorithms and sequencing instructions to build up the variety of applications they require. The framework includes machinery to manage these user codes, to organize the abundance of user-contributed configuration files, to facilitate multi-format file handling, and to provide access to event and time-dependent detector information which can reside in various data sources. A number of utilities are also provided, including a novel geometry package which allows manipulation of abstract geometrical objects independent of coordinate system choice. The framework is implemented in C++, follows an object oriented paradigm, and takes advantage of some of the more widespread tools that the open source community offers, while keeping the user-side simple enough for C++ non-experts to learn in a reasonable time. The distribution system includes unit and acceptance testing in order to support rapid development of both the core framework and contributed user code. Great attention has been paid to the ease of installation.

**Poster Session 3 - Board: 36 / 354**

## The Geometry Package for the Pierre Auger Observatory

**Authors:** L. Nellen<sup>1</sup>; L. Prado Jr<sup>2</sup>; S. Argiro<sup>3</sup>; T. Paul<sup>4</sup>; T. Porter<sup>5</sup>

<sup>1</sup> *I. DE CIENCIAS NUCLEARES, UNAM*

<sup>2</sup> *State University of Campinas*

<sup>3</sup> *University of Torino*

<sup>4</sup> *Northeastern University*

<sup>5</sup> *Louisiana State University*

The Pierre Auger Observatory consists of two sites with several semi-autonomous detection systems. Each component, and in some cases each event, provides a preferred coordinate system for simulation and analysis. To avoid a proliferation of coordinate systems in the offline software of the Pierre Auger Observatory, we have developed a

geometry package that allows the treatment of fundamental geometrical objects in a coordinate-independent way. This package makes transformations between coordinate systems transparent to the user, without taking the control about the internal representation completely from the user.

The geometry package allows easy combination of the results from different sub-detectors, at the same time as ensuring that effects like the earth curvature, which is non-negligible on the scale of a single Auger site, are dealt with properly.

The internal representations used are Cartesian. For interfacing, including I/O, the package includes support for Cartesian coordinates, geodetic (latitude/longitude and UTM), and astrophysical coordinate systems.

## Core Software / 356

### The ZOOM Minimization Package

**Authors:** D. Sachs<sup>1</sup>; M. Fischler<sup>2</sup>

<sup>1</sup> *Fermilab*

<sup>2</sup> *FERMILAB*

**Corresponding Author:** mf@fnal.gov

A new object-oriented Minimization package is available via the ZOOM cvs repository. This package, designed for use in HEP applications, has all the capabilities of Minuit, but is a re-write from scratch, adhering to modern C++ design principles.

A primary goal of this package is extensibility in several directions, so that its capabilities can be kept fresh with as little maintenance effort as possible. These flexibility goals have been met, as demonstrated by extensions of the package to add new types of termination conditions, new domains and restrictions on the solution space, and a new minimization algorithm. Each of these extensions was straightforward to implement.

The object-oriented design style also has several advantages at the user level. One such advantage is the ability to consider several problems simultaneously within a single program. Another is that we can make the Minimizer objects easy to coordinate the Minimizer with the use of other products. To verify that this goal is met, we demonstrate examples of using the Minimizer in the context of a Root application, and in an application using the “R” statistical analysis environment and language.

We compare and contrast this package with other free C++ Minimization packages suitable for HEP (most of which have origins in Minuit). Following Minuit overly precisely makes it difficult to design an object oriented package without undue distortions. This package is distinguished by the priority that was assigned to C++ design issues, and the focus on producing an extensible system that will resist becoming obsolete.

## The GANGA user interface for physics analysis on distributed resources

**Authors:** A. Maier<sup>1</sup>; A. Soroko<sup>2</sup>; C.L. Tan<sup>3</sup>; D. Adams<sup>4</sup>; G.N. Patrick<sup>5</sup>; J. Martyniak<sup>6</sup>; J. Moscicki<sup>1</sup>; K. Harrison<sup>7</sup>; P. Charpentier<sup>1</sup>; P. Mato<sup>1</sup>; R. Jones<sup>8</sup>; U. Egede<sup>9</sup>

<sup>1</sup> CERN

<sup>2</sup> Oxford University

<sup>3</sup> Birmingham University

<sup>4</sup> BNL

<sup>5</sup> RAL

<sup>6</sup> Imperial College London

<sup>7</sup> Cambridge University

<sup>8</sup> Lancaster University

<sup>9</sup> IMPERIAL COLLEGE LONDON

**Corresponding Author:** u.egede@imperial.ac.uk

Any physicist who will analyse data from the LHC experiments will have to deal with data and computing resources which are distributed across multiple locations and with different access methods. GANGA helps the end user by tying in specifically to the solutions for a given experiment ranging from specification of data to retrieval and post-processing of produced output. For LHCb and ATLAS the main goal is to assist in running jobs based on the Gaudi/Athena C++ framework. GANGA is written in Python and presents the user with a single GUI rather than a set of different applications. It interacts with external resources like experiments bookkeeping databases, job configuration, and Grid submission systems through plug-able modules. The user is upon start-up presented with a list of templates for common analysis tasks and GANGA persists information about ongoing tasks between invocations. GANGA can also be used through a command line interface that has a tight connection to the GUI to ease the transition from one to the other. Examples will be presented that demonstrates the integration into the distributed analysis systems of the LHCb and ATLAS experiments as used during their 2004 data challenges.

**Poster Session 2 - Board:** 55 / 358

## SRM AND GFAL TESTING FOR LCG2

**Authors:** E. Slabospitskaya<sup>1</sup>; M. Sapunov<sup>1</sup>; V. Petoukhov<sup>1</sup>

<sup>1</sup> Institute for High Energy Physics, Protvino, Russia

**Corresponding Author:** elena.slabospitskaya@ihep.ru

Storage Resource Manager (SRM) and Grid File Access Library (GFAL) are GRID middleware components used for transparent access to Storage Elements. SRM provides a common interface (WEB service) to backend systems giving dynamic space allocation and file management. GFAL provides a mechanism whereby an application software can access a file at a site without having to know which transport mechanism to use or at which site it is running.

Two separate Test Suites have been developed for testing of SRM interface v 1.1 and testing against the GFAL file system. Test Suites are written in C and Perl languages. SRM test suite: a script in Perl generates files and their replicas. These files are copied to the local SE and registered (published). Replicas of files are made to the specified SRM site. All replicas are used by the C-program. The SRM functions, such as get, put, pin, unPin etc. are tested using a program written in C. As SRMs do not perform file movement operations, the C-program transfers files using "globus-url-copy". It then compares the data files before and after transfer.



GFAL test suite: as GFAL allows users to access a file in a Storage Element directly (read and write) without copying it locally, a C-program tests the implementation of POSIX I/O functions such as open/seek/read/write. A Perl script executes almost all Unix based commands: dd, cat, cp, mkdir and so on. Also the Perl script launches a stress test, creating many small files (~5000), nested directories and huge files. The investigation of interactions between the Replica Manager, the SRM and the file access mechanism will help making the Data Management software better.

**Wide Area Networking / 359**

## **LambdaStation: A forwarding and admission control service to interface production network facilities with advanced research network paths**

**Authors:** D. Petravick<sup>1</sup>; P. DeMar<sup>1</sup>

<sup>1</sup> *FERMILAB*

**Corresponding Author:** demar@fnal.gov

Advanced optical-based networks have the capacity and capability to meet the extremely large data movement requirements of particle physics collaborations. To date, research efforts in the advanced network area have been primarily been focused on provisioning, dynamically configuring, and monitoring the wide area optical network infrastructure itself. Application use of these facilities has been largely limited to demonstrations using prototype high performance computing systems. Fermilab has initiated a project to enable our production network facilities to exploit these advanced research networks. Our objective is to selectively forward designated data transfers, on a per-flow basis, between capacious production-use storage systems on local campus networks, using a dynamically provisioned alternate path on a wide area advanced research network. To accomplish this, it is necessary to develop the capability to dynamically reconfigure forwarding of specific flows within our local production-use routers, provide an interface that enables applications to utilize the service, and dynamically implement appropriate access control on the alternate network path. Our project involves developing that infrastructure. We call it LambdaStation. If one envisions wide area optical network paths as high bandwidth data railways, then LambdaStation would functionally be the railroad terminal that regulates which flows at the local site get directed onto the high bandwidth data railways. LambdaStation is in a very early stage of development. Our paper will discuss its design, early deployment experiences, and future directions for the project.

**Poster Session 3 - Board: 9 / 360**

## **Muon Event Filter Software for the ATLAS Experiment at LHC**

**Author:** M. Biglietti<sup>1</sup>

<sup>1</sup> *UNIVERSITY OF MICHIGAN*

**Corresponding Author:** michela.biglietti@cern.ch

At LHC the 40 MHz bunch crossing rate dictates a high selectivity of the ATLAS Trigger system, which has to keep the full physics potential of the experiment in spite of a limited storage capability.

The level-1 trigger, implemented in a custom hardware, will reduce the

initial rate to 75 kHz and is followed by the software based level-2 and Event Filter, usually referred as High Level Triggers (HLT), which further reduce the rate to about 100 Hz.

In this paper an overview of the implementation of the offline muon reconstruction algorithms MOORE (Muon Object Oriented REconstruction) and MuId (Muon Identification) as Event Filter in the Atlas online framework is given.

The MOORE algorithm performs the reconstruction inside the Muon Spectrometer providing a precise measurement of the muon track parameters outside the calorimeters; MuId combines the measurements of all ATLAS sub-detectors in order to identify muons and provides the best estimate of their momentum at the production vertex.

In the HLT implementation the muon reconstruction can be executed in the “full scan mode”, performing pattern recognition in the whole muon spectrometer, or in the “seeded mode”, taking advantage of the results of the earlier trigger levels.

An estimate of the execution time will be presented along with the performances in terms of efficiency, momentum resolution and rejection power for muons coming from hadron decays and for fake muon tracks, due to accidental hit correlations in the high background environment of the experiment.

## Core Software / 361

### Quality Assurance and Testing in LCG

**Authors:** A. Aimar<sup>1</sup>; J. Moscicki<sup>1</sup>; L. MANCERA<sup>1</sup>; M. GALLAS<sup>1</sup>; M. Lamanna<sup>1</sup>

<sup>1</sup> CERN

Software Quality Assurance is an integral part of the software development process of the LCG Project and includes several activities such as automatic testing, test coverage reports, static software metrics reports, bug tracker, usage statistics and compliance to build, code and release policies.

As a part of QA activity all levels of the sw-testing should be run as a part of automatic process, the SPI project delivers a general test-framework solution based on open source software together with test document templates and software testing policies. The test-framework solution is built on QMtest, Oval and the X-Unit family (CppUnit, PyUnit, JUnit). The specific languages testing features are covered at the unit-testing level with the X-Unit family, the validation testing activity can be done through Oval. And Qmtest offers a way to integrate all the tests and write custom python tests, having a nice web interface for running and browse the test results.

Test coverage reports allow to understand to which extent software products are tested and they are based on the approach used by Linux Testing Project. Code size and development effort of the software is estimated using sloccount utility based on standard development models. Statistics are automatically extracted from the savannah bug tracker which enables to analyze the evolution of the quality, amount of feedback from the users etc. Finally the compliance with the standard LCG policies is verified. It includes the build and CVS repository structure and the standard release procedure.

## Data Rereprocessing on Worldwide Distributed Systems

**Authors:** D. Wicke<sup>1</sup>; M. Diesburg<sup>2</sup>

<sup>1</sup> *Fermilab*

<sup>2</sup> *Fermi National Accelerator Laboratory*

**Abstract:**

The D0 experiment faces many challenges enabling access to large datasets for physicists on 4 continents. The strategy of solving these problems on worldwide distributed computing clusters is followed.

Already since the begin of TEvatron RunII (March 2001) all Monte-Carlo simulations are produced outside of Fermilab at remote systems. For analyses as system of regional analysis centers (RACs) was established which supply the associated institutes with the data. This structure which is similar the the Tier structure foreseen for LHC was used in autumn 2003 to rereprocess all D0-data with the uptodate and much improved reconstruction software.

As the first running experiment D0 has implemented and operated all important computing dask of a high energy physics experiment on worldwide distributed systems.

The experiences gained in D0 can be applied to judge the LHC computing model.

**Poster Session 2 - Board: 70 / 363**

## On the Management of Certification Authority in Large Scale GRID Infrastructure

**Author:** E. Berdnikov<sup>1</sup>

<sup>1</sup> *INSTITUTE FOR HIGH ENERGY PHYSICS, PROTVINO, RUSSIA*

**Corresponding Author:** [evgueni.berdnikov@cern.ch](mailto:evgueni.berdnikov@cern.ch)

The scope of this work is the study of scalability limits of the Certification Authority (CA), running for large scale GRID environments.

The operation of Certification Authority is analyzed from the view of the rate of incoming requests, complexity of authentication procedures, LCG security restrictions and other limiting factors. It is shown, that standard CA operational model has some native "bottlenecks", which can be resolved with proper management and technical tools.

The central point is the discussion of "decentralized" scheme with single CA and multiple authentication agents, called Registration Authorities (RA). Single CA retains a role for technical center, responsible for support of GRID security infrastructure, while general role of RAs is verification of requests from end-users.

Practical implementation of this scheme (including the development and installation of end-user software) have been done in CERN in 2002 (<http://service-grid-ca.web.cern.ch/service-grid-ca/help/RA.html>). Second implementation of the same ideas was the GRID project of the Russia Ministry of Atomic Energy, 2003 (<http://grid.ihep.su/MAG/>).

These two implementations are compared in aspects of security and functionality.

**Poster Session 2 - Board: 49 / 364**

## **LEXOR, the LCG-2 Executor for the ATLAS DC2 Production System**

**Authors:** A. De Salvo<sup>1</sup>; C. Negri<sup>2</sup>; D. Rebatto<sup>3</sup>; L. Vaccarossa<sup>4</sup>

<sup>1</sup> *INFN Roma*

<sup>2</sup> *CNAF Bologna*

<sup>3</sup> *INFN - MILANO*

<sup>4</sup> *INFN Milano*

**Corresponding Author:** david.rebatto@mi.infn.it

In this paper we present an overview of the implementation of the LCG interface for the ATLAS production system. In order to take profit of the features provided by DataGRID software, on which LCG is based, we implemented a Python module, seamless integrated into the Workload Management System, which can be used as an object-oriented API to the submission services. On top of it we implemented Lexor, an executor component conforming to the pull/push model designed by the DC2 production system team. It pulls job descriptions from the supervisor component and uses them to create job objects, which in turn are submitted to the Grid. All the typical Grid operations (match-making with respect to input data location, registration of output data in the replica catalog, workload balancing) are performed by the underlying middleware, while interactions with ATLAS metadata catalog and the production database are granted by the integration with the Data Management System (Don Quijote) client module and via XML messages to the production supervisor (Windmill).

**Poster Session 2 - Board: 25 / 365**

## **DIRAC Workload Management System**

**Authors:** A. Tsaregorodtsev<sup>1</sup>; I. Stokes-Rees<sup>2</sup>; V. garonne<sup>1</sup>

<sup>1</sup> *CPPM-IN2P3 MARSEILLE*

<sup>2</sup> *oxford*

**Corresponding Author:** garonne@cppm.in2p3.fr

The Workload Management System (WMS) is the core component of the DIRAC distributed MC production and analysis grid of the LHCb experiment. It uses a central Task database which is accessed via a set of central Services with Agents running on each of the LHCb sites. DIRAC uses a 'pull' paradigm where Agents request tasks whenever they detect their local resources are available.

The collaborating central Services allow new components to be plugged in easily. These Services can perform functions such as scheduling optimization, task prioritization, job splitting and merging, to name a few. They provide also job status information for various

monitoring clients. We will discuss the services deployment and operation with particular emphasis on the robustness and scalability issues.

The distributed Agents have modular design which allows easy functionality extensions to adapt to the needs of a particular site. The Agent installation have only basic pre-requisites which makes it easy for new sites to be incorporated. An Agent can be deployed on a gatekeeper of a large cluster or just on a single worker node of the LCG grid. PBS,LSF,BQS, Condor,LCG,Globus can be used as the DIRAC computing resources.

The WMS components use XML-RPC and instant messaging Jabber protocols for communication which increases the overall reliability of the system. The jobs handled by the WMS are described using Classad library which facilitates the interoperability with other grids.

## Distributed Computing Systems and Experiences / 366

### Globally Distributed User Analysis Computing at CDF

**Authors:** A. Fella<sup>1</sup>; A. Kreymer<sup>2</sup>; A. Sill<sup>3</sup>; D. Han<sup>4</sup>; E. Lipeles<sup>5</sup>; F. Ratnikov<sup>6</sup>; F. Wuerthwein<sup>7</sup>; H-T. Lung<sup>8</sup>; H. Matsunaga<sup>9</sup>; H. Park<sup>10</sup>; I. Sfiligoi<sup>11</sup>; K. Cho<sup>4</sup>; L. Groer<sup>12</sup>; M. Burgon-Lyon<sup>13</sup>; M. Neubauer<sup>7</sup>; R. St.Denis<sup>13</sup>; R. Tafirout<sup>12</sup>; S. Belforte<sup>14</sup>; S. Hou<sup>8</sup>; S. Stonjek<sup>15</sup>; V. Bartsch<sup>15</sup>

<sup>1</sup> INFN, Pisa

<sup>2</sup> Fermi National Accelerator Laboratory

<sup>3</sup> TEXAS TECH UNIVERSITY

<sup>4</sup> Kyungpook National University

<sup>5</sup> UNIVERSITY OF CALIFORNIA SAN DIEGO

<sup>6</sup> Rutgers University

<sup>7</sup> University of California, San Diego

<sup>8</sup> Academia Sinica, Taiwan

<sup>9</sup> University of Tsukuba

<sup>10</sup> Kyungpook National University/KISTI

<sup>11</sup> INFN, Frascati

<sup>12</sup> University of Toronto

<sup>13</sup> University of Glasgow

<sup>14</sup> INFN, Trieste

<sup>15</sup> University of Oxford

To maximize the physics potential of the data currently being taken, the CDF collaboration at Fermi National Accelerator Laboratory has started to deploy user analysis computing facilities at several locations throughout the world. Over 600 users are signed up and able to submit their physics analysis and simulation applications directly from their desktop or laptop computers to these facilities. These resources consist of a mix of customized computing centers and a decentralized version of our Central Analysis Facility (CAF) initially used at Fermilab, which we have designated Decentralized CDF Analysis Facilities (DCAFs).

We report on experience gained during the initial deployment and use of these resources for the summer conference season 2004. During this period, we allowed MC generation as well as data analysis of selected data samples at several globally distributed centers. In addition, we discuss a migration path from

this first generation distributed computing infrastructure towards a more open implementation that will be interoperable with LCG, OSG and other general-purpose grid installations at the participating sites.

Core Software / 367

## Supporting the Development Process of the DataGrid Workload Management System Software with GNU autotools, CVS and RPM

**Authors:** E. Ronchieri<sup>1</sup>; F. Prelz<sup>2</sup>; R. Peluso<sup>3</sup>

**Co-authors:** A. Gianelle<sup>3</sup>; A. Guarise<sup>4</sup>; A. Krenek<sup>5</sup>; A. Maraschini<sup>6</sup>; A. Werbrouck<sup>4</sup>; B. Cantalupo<sup>7</sup>; D. Colling<sup>8</sup>; D. Kouril<sup>5</sup>; F. Giacomini<sup>1</sup>; F. Pacini<sup>6</sup>; G. Avellino<sup>6</sup>; J. Pospisil<sup>5</sup>; J. Sitera<sup>5</sup>; L. Matyska<sup>5</sup>; M. Mezzadri<sup>2</sup>; M. Mulac<sup>5</sup>; M. Pappalardo<sup>9</sup>; M. Ruda<sup>5</sup>; M. Sgaravatto<sup>3</sup>; M. Vocu<sup>5</sup>; R. Piro<sup>4</sup>; S. Beco<sup>7</sup>; S. Monforte<sup>9</sup>; Z. Kabelac<sup>5</sup>; Z. Salvat<sup>5</sup>

<sup>1</sup> INFN CNAF

<sup>2</sup> INFN Milano

<sup>3</sup> INFN Padova

<sup>4</sup> INFN Torino

<sup>5</sup> CESNET

<sup>6</sup> Datamat Ingegneria dei Sistemi S.p.A

<sup>7</sup> Datamat Ingegneria dei Sistemi S.p.A.

<sup>8</sup> Imperial College, London

<sup>9</sup> INFN Catania

**Corresponding Author:** elisabetta.ronchieri@cnafe.infn.it

We described the process for handling software builds and releases for the Workload Management package of the DataGrid project. The software development in the project was shared among nine contractual partners, in seven different countries, and was organized in work-packages covering different areas.

In this paper, we discuss how a combination of Concurrent Version System, GNU autotools and other tools and practices was organised to allow the development, build, test and distribution of the DataGrid Workload Management System. This is not only characterised by a rather high internal geographic and administrative dispersion (four institutions with developers at nine different locations in three countries), but by the fact we had to integrate and interface to a dozen of third-party code packages coming from different sources, and to the software products coming from other three development work-packages internal to the project.

A high level of central co-ordination needed to be maintained for project-wide steering, and this had also to be reflected in the software development infrastructure, while maintaining ease-of-use for distributed developers and automated procedures wherever possible.

Poster Session 2 - Board: 57 / 368

## DIRAC Lightweight information and monitoring services using XML-RPC and Instant Messaging

**Authors:** A. Tsaregorodtsev<sup>1</sup>; I. Stokes-Rees<sup>2</sup>; V. GARONNE<sup>3</sup>

<sup>1</sup> CPPM, IN2P3

<sup>2</sup> UNIVERSITY OF OXFORD PARTICLE PHYSICS

<sup>3</sup> CNRS-IN2P3-CPPM, MARSEILLE

**Corresponding Author:** ian.stokes-rees@cern.ch

The DIRAC system developed for the CERN LHCb experiment is a grid infrastructure for managing generic simulation and analysis jobs. It enables jobs to be distributed across a variety of computing resources, such as PBS, LSF, BQS, Condor, Globus, LCG, and individual workstations.

A key challenge of distributed service architectures is that there is no single point of control over all components. DIRAC addresses this via two complementary features: a distributed Information System, and an XMPP (Extensible Messaging and Presence Protocol) Instant Messaging framework.

The Information System provides a concept of local and remote information sources.

Any information which is not found locally will be fetched from remote sources. This allows a component to define its own state, while fetching the state of other components directly from those components, or via a central Information Service. We will present the architecture, features, and performance of this system.

XMPP has provided DIRAC with numerous advantages. As an authenticated, robust, lightweight, and scalable asynchronous message passing system, XMPP is used, in addition to XML-RPC, for inter-Service communication, making DIRAC very fault-tolerant, a critical feature when using Service Oriented Architectures. XMPP is also used for monitoring real-time behaviour of the various DIRAC components.

Finally, XMPP provides XML-RPC like facilities which are being developed to provide control channels direct to Services, Agents, and Jobs. We will describe our novel use of Instant Messaging in DIRAC and discuss directions for the future.

**Poster Session 1 - Board: 2 / 369**

## WAN Emulation Development and Testing at Fermilab

**Authors:** A. Bobyshev<sup>1</sup>; M. Ernst<sup>2</sup>; P. Demar<sup>1</sup>; R. Rechenmacher<sup>1</sup>

<sup>1</sup> FERMILAB

<sup>2</sup> DESY

**Corresponding Author:** bobyshev@fnal.gov

The Compact Muon Solenoid (CMS) experiment at CERN's Large Hadron Collider (LHC) is scheduled to come on-line in 2007. Fermilab will act as the CMS Tier-1 center for the US and make experiment data available to more than 400 researchers in the US participating in the CMS experiment. The US CMS Users Facility group, based at Fermilab, has initiated a project to develop a model for optimizing movement of CMS experiment data between CERN and the various tiers of US CMS data centers. Fermilab has initiated a project to design a WAN emulation facility which will enable controlled testing of unmodified or modified CMS applications and TCP implementations locally under conditions that emulate WAN connectivity. The WAN emulator facility is configurable for latency, jitter, and packet loss. The initial implementation is based on the NISTnet software product. In this paper we will

describe the status of this project to date, the results of validation and comparison of performance measurements obtained in emulated and real environment for different applications including multistreams GridFTP. We also will introduce future short term and intermediate term plans, as well as outstanding problems and issues.

**Poster Session 1 - Board: 3 / 370**

## **Methodologies and techniques for analysis of network flow data**

**Authors:** A. Bobyshev<sup>1</sup>; M. Grigoriev<sup>2</sup>

<sup>1</sup> *FERMILAB*

<sup>2</sup> *Fermilab*

**Corresponding Author:** bobyshev@fnal.gov

Network flow data gathered on border routers and core network switch/routers is used at Fermilab for statistical analysis of traffic patterns, passive network monitoring, and estimation of network performance characteristics. Flow data is also a critical tool in the investigation of computer security incidents. Development and enhancement of flow-based tools is on-going effort. The current state of flow analysis is based on the open source Flow-Tools package. This paper describes the most recent developments in flow analysis at Fermilab. Our goal is to provide a multidimensional view of network traffic patterns, with a detailed breakdown based on site, experiment, domain, subnet, hosts, protocol, or application. The latest analysis tool provides a descriptive and graphical representation of network traffic broken down by combinations of experiment and DNS domain. The tool can be utilized in real-time mode, as well as to provide a historical view. Another tool analyzes flow data to provide performance characteristics of completed multistream GridFTP data transfers. The current prototype provides a web interface for dynamic administration of the flow reports. We will describe and discuss the new features that we plan on developing in future enhancements to our flow analysis tool set.

**Poster Session 2 - Board: 71 / 372**

## **The D0 Virtual Center and planning for large scale computing**

**Authors:** A. Boehnlein<sup>1</sup>; D0 Computing Planning Board for the<sup>2</sup>

<sup>1</sup> *FERMI NATIONAL ACCELERATOR LABORATORY*

<sup>2</sup> *D0 Collaboration*

**Corresponding Author:** cope@fnal.gov

The D0 experiment relies on large scale computing systems to achieve her physics goals. As the experiment lifetime spans, multiple generations of computing hardware, it is fundamental to make projective models in to use available resources to meet the anticipated needs. In addition, computing resources can be supplied as in-kind contributions by collaborating institutions and countries, however, such resources typically require scheduling, thus adding another dimension for planning. In addition, to avoid over-subscription of the resources, the experiment has to be educated on the limitations and trade-offs for various computing activities to enable the management to prioritize. We present the metrics and mechanisms used for



planning and discuss the uncertainties and unknowns, as well as some of the mechanisms for communicating the resource load to the stakeholders.

In order to correctly account for in-kind contributions of remote computing, D0 uses the concept of a Virtual Center, in which all of the costs are estimated as if the computing were located at solely at FNAL. In contrast to other such models in common use, D0 accounts for contributions based on computer usage rather than strictly on money spend on hardware. This gives incentive to achieve the maximum efficiency of the systems as well as encouraging active participation in the computing model by collaborating institutions. This method of operation leverages a common tool and infrastructure base for all production-type activities.

Plenary / 373

## Run II computing

**Author:** A. Boehnlein<sup>1</sup>

<sup>1</sup> *FERMI NATIONAL ACCELERATOR LABORATORY*

**Corresponding Author:** cope@fnal.gov

In support of the Tevatron physics program, the Run II experiments have developed computing models and hardware facilities to support data sets at the petabyte scale, currently corresponding to 500 pb<sup>-1</sup> of data and over 2 years of production operations. The systems are complete from online data collection to user analysis, and make extensive use of central services and common solutions developed with the FNAL CD and experiment collaborating institutions, and make use of global facilities to meet the computing needs. We describe the similarities and differences between computing on CDF and D0 while describing solutions for database and database servers, data handling, movement and storage and job submission mechanisms. The facilities for production computing and analysis and the use of commodity file servers will also be described. Much of the knowledge gained from providing computing at this scale can be abstracted and applied to design and planning for future experiments with large scale computing.

Computer Fabrics / 374

## Operation of the CERN Managed Storage environment; current status and future directions

**Authors:** C. Curran<sup>1</sup>; D. Hughes<sup>1</sup>; G. Lee<sup>1</sup>; H. Cacote<sup>1</sup>; J. van Eldik<sup>1</sup>; T. Osborne<sup>1</sup>; T. Smith<sup>1</sup>; V. Bahyl<sup>1</sup>

<sup>1</sup> *CERN*

**Corresponding Author:** tim.smith@cern.ch

This paper discusses the challenges in maintaining a stable Managed Storage Service for users built upon dynamic underlying disk and tape layers.

Early in 2004 the tools and techniques used to manage disk, tape, and stage servers were refreshed in adopting the QUATTOR tool set. This has markedly increased the coherency and efficiency of the configuration of data servers. The LEMON monitoring suite was deployed to raise alarms and gather performance metrics. Exploiting this foundation, higher level service displays are being added, giving

comprehensive and near-real-time views of operations. The scope of our monitoring has been broadened to include low-level machine sensors such as thermometer, IPMI and SMART readings, improving our ability to detect impending hardware failure.

In terms of operations, widespread disk reliability problems which were manpower intensive to chase, were overcome by exchanging a bad batch of 1200 disks. Recent LHC data challenges have ventured into new operating domains for the CASTOR system, with massive disk resident file catalogues requiring special handling. The tape layer has focused on STK 9940 drives for bulk recording capacity: a large scale data migration to this media permitted old drive technologies to be retired. Repacking 9940A data to 9940B high density media allows us to recycle tapes, giving substantial savings by avoiding acquisition of new media.

In addition to more robust software, hardware developments are required for LHC era services. We are moving from EIDE to SATA based disk storage and envisage a tape drive technology refresh. Details will be provided of our investigations in these areas.

Poster Session 3 - Board: 20 / 376

## Geant4 as Simulation Toolkit addressed to interplanetary manned missions studies: required developments and improvements

**Authors:** M.G. Pia<sup>1</sup>; S. Guatelli<sup>1</sup>

<sup>1</sup> INFN Genova, Italy

**Corresponding Author:** guatelli@ge.infn.it

The study of the effects of space radiation on astronauts is an important concern of space missions for the exploration of the Solar System. The radiation hazard to crew is critical to the feasibility of interplanetary manned missions.

To protect the crew, shielding must be designed, the environment must be anticipated and monitored, and a warning system must be put in place.

A Geant4 simulation has been developed for a preliminary quantitative study of vehicle concepts and Moon surface habitat designs, and the radiation exposure of crews therein. This project is defined in the context of the European AURORA programme, whose primary object is to study solutions for the robotic and human exploration of the Solar System, with Mars, the Moon and the asteroids as the most likely objects.

This study intends to evaluate whether the energy range typical of the radiation environment of interplanetary missions is adequately treated in Geant4 physics packages, for all the major types of particles involved, identifying the availability of appropriate electromagnetic and hadronic physics models and verifying the status of their validation. Recommendations for further Geant4 developments or improvements and of further validation tests, necessary for the interplanetary manned missions, are issued as a result of this study.

Distributed Computing Services / 377

## DIRAC - The Distributed MC Production and Analysis for LHCb

**Authors:** A. Pickford<sup>1</sup>; A. Soroko<sup>2</sup>; A. TSAREGORODTSEV<sup>3</sup>; G. Kuznetsov<sup>4</sup>; G.N. Patrick<sup>4</sup>; I. Korolko<sup>5</sup>; I. Stokes-Rees<sup>6</sup>; J. Blouw<sup>7</sup>; J. Clozier<sup>8</sup>; J. Saborido-Silva<sup>9</sup>; M. Frank<sup>8</sup>; M. Gandelman<sup>10</sup>; M. Sanchez-Garcia<sup>9</sup>; M. Tobin<sup>11</sup>; M. Witek<sup>12</sup>; N. Brook<sup>13</sup>; P. Charpentier<sup>8</sup>; R. Bernet<sup>14</sup>; R. Graciani-Diaz<sup>15</sup>; U. Egede<sup>16</sup>; V. GARONNE<sup>3</sup>; V. Romanovski<sup>17</sup>; V. Vagnoni<sup>18</sup>

<sup>1</sup> *University of Glasgow*

<sup>2</sup> *Oxford*

<sup>3</sup> *CNRS-IN2P3-CPPM, MARSEILLE*

<sup>4</sup> *Rutherford Appleton Lab, Chilton*

<sup>5</sup> *ITEP, Moscow*

<sup>6</sup> *Particle Physics, Oxford; CPPM, Marseille*

<sup>7</sup> *GRIDKA, Karlsruhe*

<sup>8</sup> *CERN*

<sup>9</sup> *University of Santiago de Compostela*

<sup>10</sup> *University of Rio de Janeiro*

<sup>11</sup> *University of Liverpool*

<sup>12</sup> *University of Cracow*

<sup>13</sup> *H.H. Wills Physics Laboratory, Bristol*

<sup>14</sup> *University of Zurich*

<sup>15</sup> *University of Barcelona*

<sup>16</sup> *Imperial College, London*

<sup>17</sup> *IHEP, Protvino*

<sup>18</sup> *INFN, Bologna*

**Corresponding Author:** atsareg@in2p3.fr

DIRAC is the LHCb distributed computing grid infrastructure for MC production and analysis. Its architecture is based on a set of distributed collaborating services. The service decomposition broadly follows the ARDA project proposal, allowing for the possibility of interchanging the EGEE/ARDA and DIRAC components in the future. Some components developed outside the DIRAC project are already in use as services, for example the File Catalog developed by the AliEn project.

An overview of the DIRAC architecture will be given, in particular the recent developments to support user analysis. The main design choices will be presented. One of the main design goals of DIRAC is the simplicity of installation, configuring and operation of various services. This allows all the DIRAC resources to be easily managed by a single Production Manager.

The modular design of the DIRAC components allows its functionality to be easily extended to include new computing and storage elements or to handle new tasks. The DIRAC system already uses different types of computing resources - from single PC's to a variety of batch system to the Grid environment. In particular, the use of the LCG2 environment will be presented. Different ways to utilise LCG2 resources will be examined as well as the issue of interoperability between the LCG2 and DIRAC sites.

**Poster Session 3 - Board: 14 / 378**

## **Pixel Reconstruction in the CMS High-Level Trigger**

**Authors:** D. Kotlinski<sup>1</sup>; M. Konecki<sup>2</sup>; S. Cucciarelli<sup>3</sup>; T. Todorov<sup>4</sup>

<sup>1</sup> *Paul Scherrer Institut (PSI)*

<sup>2</sup> *University of Basel*

<sup>3</sup> *CERN*

<sup>4</sup> *Universite Louis Pasteur*

**Corresponding Author:** susanna.cucciarelli@cern.ch

The Pixel Detector is the innermost one in the tracking system of the Compact Muon Solenoid (CMS) experiment. It provides the most precise measurements not only supporting the full track reconstruction but also allowing the standalone reconstruction useful especially for the online event selection at High-Level Trigger (HLT). The performance of the Pixel Detector is given. The HLT algorithms using Pixel Detector are presented, including pixel track reconstruction, primary vertex finding, tau identification, isolation and track seeding.

**Event Processing / 379**

## Python-based physics analysis environment for LHCb

**Authors:** A. TSAREGORODTSEV<sup>1</sup>; E. DE OLIVEIRA<sup>2</sup>; G. B. BARRAND<sup>3</sup>; I. BELYAEV<sup>4</sup>; M. FRANK<sup>5</sup>; P. MATO<sup>5</sup>

<sup>1</sup> *Uni Marseil II/Marseil*

<sup>2</sup> *CBPF/Rio de Janeiro*

<sup>3</sup> *LAL/Orsay*

<sup>4</sup> *ITEP/Moscow*

<sup>5</sup> *CERN*

**Corresponding Authors:** pere.mato@cern.ch, ivan.belyaev@itep.ru

Bender, the Python based physics analysis application for LHCb combines the best features of underlying Gaudi C++ software architecture with the flexibility of Python scripting language and provides end-users with friendly physics analysis oriented environment. It is based in one hand, on the generic Python bindings for the Gaudi framework, called GaudiPython, and in the other hand on an efficient C++ physics analysis toolkit called LoKi. Bender and LoKi use the tools from the physics analysis framework, called DaVinci. Bender achieves a clear separation between the technical details and the physical contents of end-user physicist's code. The usage of Python, AIDA abstract interfaces and standard LCG reflection techniques allows an easy integration of Bender's analysis environment with third party products like the interactive event display and visualization tools like Panoramix/LaJoconde, ROOT and HippoDraw. We'll present the overall design and capabilities of the system, its status and prospects.

**Poster Session 2 - Board: 9 / 380**

## GILDA: a Grid for dissemination activities

**Authors:** A. Carrieri<sup>1</sup>; A. Falzone<sup>2</sup>; A. Italiano<sup>3</sup>; E. Ferro<sup>4</sup>; E. Giorgio<sup>1</sup>; G. Andronico<sup>1</sup>; G. La Rocca<sup>1</sup>; G. Platania<sup>1</sup>; M. Pappalardo<sup>1</sup>; M. Verlato<sup>4</sup>; R. Barbera<sup>5</sup>; R. Catania<sup>1</sup>; S. Monforte<sup>1</sup>; V. Ardizzone<sup>1</sup>

<sup>1</sup> *INFN Catania*

<sup>2</sup> *NICE s.r.l.*

<sup>3</sup> *INFN CNAF*

<sup>4</sup> *INFN Padova*

<sup>5</sup> *Univ. Catania and INFN Catania*

**Corresponding Author:** roberto.barbera@ct.infn.it

Computational and data grids are now entering a more mature phase where experimental test-beds are turned into production quality infrastructures operating around the clock. All this is becoming true both at national level, where an example is the Italian INFN production grid (<http://grid-it.cnaf.infn.it>), and at the continental level, where the most striking example is the European Union EGEE Project Infrastructure (<http://www.eu-egee.org>).

However, the impact of grid technologies on the next future way of doing e-science and research in Europe will be proportional to the capability of National and European Grid Infrastructures to attract and serve many diverse scientific and industrial communities through serious and detailed dissemination and tutoring programs.

In this contribution we present GILDA, the Grid Infn Laboratory for Dissemination Activities (<http://gilda.ct.infn.it>). GILDA is a complete suite of grid elements (Certification Authority, Virtual Organization, Distributed Test-bed, Grid Demonstrator, etc.) completely devoted to dissemination activities. GILDA can also act as a fast-prototyping test-bed where to start the porting/interfacing of new applications with the grid middle-ware. The use and exploitation of GILDA in the context of the Network Activities of the EGEE Project will be discussed.

**Core Software / 381**

## POOL Development Status and Plans

**Authors:** A. Vaniachine<sup>1</sup>; C. Cioffi<sup>2</sup>; D. Duellmann<sup>3</sup>; D. Malon<sup>4</sup>; G. Govi<sup>5</sup>; H. Schmuecker<sup>5</sup>; I. Papadopoulos<sup>5</sup>; J. Moscicki<sup>5</sup>; K. Karr<sup>1</sup>; M. Frank<sup>5</sup>; M. Gironé<sup>5</sup>; R. Chytracék<sup>5</sup>; T. Barrass<sup>6</sup>; W. Tannenbaum<sup>7</sup>; Z. Xie<sup>8</sup>

<sup>1</sup> *Argonne National Lab*

<sup>2</sup> *Oxford University*

<sup>3</sup> *CERN IT/DB & LCG POOL PROJECT*

<sup>4</sup> *ANL*

<sup>5</sup> *CERN*

<sup>6</sup> *Bristol University*

<sup>7</sup> *Fermilab*

<sup>8</sup> *Princeton University*

**Corresponding Author:** dirk.duellmann@cern.ch

The LCG POOL project is now entering the third year of active development. The basic functionality of the project is provided but some functional extensions will move into the POOL system this year. This presentation will give a summary of the main functionality provided by POOL, which used in physics productions today. We will then present the design and implementation of the main new interfaces and components planned such as the POOL RDBMS abstraction layer and the RDBMS based Storage Manager back-end.

**Core Software / 382**

## POOL Integration into three Experiment Software Frameworks

**Authors:** A. Schaffer<sup>1</sup>; A. Vaniachine<sup>2</sup>; C. Cioffi<sup>3</sup>; D. Duellmann<sup>4</sup>; D. Malon<sup>5</sup>; G. Govi<sup>4</sup>; H. Schmuecker<sup>4</sup>; I. Papadopoulos<sup>4</sup>; J. Moscicki<sup>4</sup>; K. Karr<sup>2</sup>; M. Frank<sup>4</sup>; M. Girone<sup>4</sup>; P van Gemmeren<sup>6</sup>; P. MATO<sup>4</sup>; R. Chytracsek<sup>4</sup>; T. Barrass<sup>7</sup>; V. Innocente<sup>8</sup>; W. Tannenbaum<sup>9</sup>; Z. Xie<sup>10</sup>

<sup>1</sup> Orsay

<sup>2</sup> Argonne National Lab

<sup>3</sup> Oxford University

<sup>4</sup> CERN

<sup>5</sup> ANL

<sup>6</sup> Brookhaven National Lab

<sup>7</sup> Bristol University

<sup>8</sup> CERN, PH/SFT

<sup>9</sup> Fermilab

<sup>10</sup> Princeton University

**Corresponding Author:** dirk.duellmann@cern.ch

The POOL software package has been successfully integrated with the three large experiment software frameworks of ATLAS, CMS and LHCb. This presentation will summarise the experience gained during these integration efforts and will try to highlight the commonalities and the main differences between the integration approaches. In particular we'll discuss the role of the POOL object cache, the choice of the main storage technology in ROOT (tree or named objects) and approaches to collection and catalogue integration.

**Distributed Computing Services / 383**

## Experience with POOL from the LCG Data Challenges of the three LHC experiments

**Authors:** A. Vaniachine<sup>1</sup>; C. Cioffi<sup>2</sup>; C. Grandi<sup>3</sup>; D. Duellmann<sup>4</sup>; D. Malon<sup>1</sup>; G. Govi<sup>4</sup>; H. Schmuecker<sup>4</sup>; I. Papadopoulos<sup>4</sup>; J. Moscicki<sup>4</sup>; K. Karr<sup>1</sup>; L. GOOSSENS<sup>4</sup>; L. Tuura<sup>5</sup>; M. Branco<sup>4</sup>; M. Frank<sup>4</sup>; M. Girone<sup>4</sup>; R. Chytracsek<sup>4</sup>; T. Barrass<sup>6</sup>; V. Innocente<sup>7</sup>; W. Tannenbaum<sup>8</sup>; Z. Xie<sup>9</sup>

<sup>1</sup> Argonne National Lab

<sup>2</sup> University of Oxford

<sup>3</sup> INFN-Bologna

<sup>4</sup> CERN

<sup>5</sup> Northeastern University

<sup>6</sup> University of Bristol

<sup>7</sup> CERN, PH/SFT

<sup>8</sup> Fermilab

<sup>9</sup> Princeton University

**Corresponding Author:** dirk.duellmann@cern.ch

This presentation will summarise the deployment experience gained with POOL during the first larger LHC experiments data challenges performed. In particular we discuss the storage access performance and optimisations, the integration issues with grid middleware services such as the LCG Replica Location Service (RLS) and the LCG Replica Manager and experience with the POOL

proposed way of exchanging meta data (such as File Catalog catalogue entries) in a de-coupled production system.

**Poster Session 3** - Board: 41 / 384

## The LCG Savannah software development portal

**Author:** Y. Perrin<sup>1</sup>

**Co-authors:** D. Feichtinger<sup>2</sup>; F. Orellana<sup>3</sup>; M. Roy<sup>4</sup>

<sup>1</sup> CERN

<sup>2</sup> PSI

<sup>3</sup> Univ. of Geneva

<sup>4</sup> FSF France

A web portal has been developed, in the context of the LCG/SPI project, in order to coordinate workflow and manage information in large software projects. It is a development of the GNU Savannah package and offers a range of services to every hosted project: Bug / support / patch trackers, a simple task planning system, news threads, and a download area for software releases. Features and functionality can be fine-tuned on a per project basis and the system displays content and grants permissions according to the user's status (project member, other Savannah user, or visitor). A highly configurable notification system is able to channel tracker submissions to developers in charge of specific project modules.

The portal is based on the GNU Savannah package which is now developed as 'Savane' by the Free Software Foundation of France. It is a descendant of the well known SourceForge-2.0 software. The original trackers were contributed to the open source community by XEROX, which uses a similar system for their internal software development. Several features and extensions were introduced in a collaboration of LCG/SPI with the current main developer of Savannah to adapt the software for use at CERN and the results were given back to the open source. Cern Savannah currently provides services to more than 600 users in 90 projects.

**Poster Session 2** - Board: 36 / 385

## Cross Experiment Workflow Management: The Runjob Project

**Authors:** D. Evans<sup>1</sup>; G. Graham<sup>1</sup>; I. Bertram<sup>2</sup>; P. Love<sup>2</sup>

<sup>1</sup> Computing Division, Fermi National Accelerator Laboratory

<sup>2</sup> Lancaster University

**Corresponding Authors:** peter.love@cern.ch, i.bertram@lancaster.ac.uk

Building on several years of success with the MCRunjob projects at DZero and CMS, the fermilab sponsored joint Runjob project aims to provide a Workflow description language common to three experiments: DZero, CMS and CDF. This project will encapsulate the remote processing experiences of the three experiments in an extensible software architecture using web services as a

communication medium. The core of the Runjob project will be the Shahkar software packages that provide services for describing jobs and targeting them at different execution environments. A common interface to multiple storage and compute grid elements will be provided, allowing the three experiments to share hardware resources in a transparent manner. Several tools provided by Shahkar are discussed including FileMetaBrokers, which provide a uniform way to handle files and metadata over a distributed cluster, the ShREEK runtime execution environment that allows executable jobs to provide a real time monitoring and control interface to any system, the scriptObject generic task encapsulation objects and XMLProcessor object persistency tool.

**Event Processing / 386**

## **Self-Filling Histograms: An object-oriented analysis framework**

**Authors:** B. List<sup>1</sup>; J. LIST<sup>2</sup>

<sup>1</sup> *ETH Zurich*

<sup>2</sup> *University of Wuppertal*

**Corresponding Author:** jenny.boehme@desy.de

Analyses in high-energy physics often involve the filling of large amounts of histograms from n-tuple like data structures, e.g. RooT trees. Even when using an object-oriented framework like RooT, the user code often follows a functional programming approach, where booking, application of cuts, calculation of weights and histogrammed quantities and finally the filling of the histogram is performed separately in different places of the program.

We will present a set of RooT based histogram classes that allow to define the histogrammed quantity, its weight and the cuts to be applied at the time of booking.

We use lightweight function object classes to define plotted quantities and cut conditions; the “self-filling” histograms hold references to these objects, and evaluate them in a fill method that thus needs no parameters. The use of function objects rather than strings to define plotted quantities and cuts permits error detection at compile rather than run time, and allows the implementation of caching mechanisms if costly computations are to be performed. Arithmetic and logical expressions are implemented by operator overloading. Histograms can be grouped in collections. We apply the visitor pattern to perform operations like filling, writing, fitting or attribute setting on such a group, without having to extend the collection class each time a new functionality is needed.

Although developed within the object oriented analysis framework of the H1 experiment, this toolkit can be used on any RooT tree.

**Event Processing / 387**

## **The Virtual Geometry Model**



**Author:** I. Hrivnacova<sup>1</sup>

<sup>1</sup> *IPN, ORSAY, FRANCE*

**Corresponding Author:** ivana.hrivnacova@cern.ch

In order for physicist to easily benefit from the different existing geometry tools used within the community, the Virtual Geometry Model (VGM) has been designed. In the VGM we introduce the abstract interfaces to geometry objects and an abstract factory for geometry construction, import and export. The interfaces to geometry objects were defined to be suitable to describe “geant-like” geometries with a hierarchical volume structure.

The implementation of the VGM for a concrete geometry model represents a small layer between the VGM and the particular native geometry.

At the present time this implementation is provided for the Geant4 and the Root TGeo geometry models.

Using the VGM factory, geometry can first be defined independently from a concrete geometry model, and then built by choosing a concrete instantiation of it. Alternatively, the import function of the VGM factory makes it possible to use VGM directly with native geometries (Geant4, TGeo). The export functions provide conversion into other native geometries or the XML format.

In this way, the VGM surpasses one-directional geometry converters within Geant4 VMC (Virtual Monte Carlo): roottog4 and g4toxml, and automatically provides missing directions: g4toroot, roottoxml. To port a third geometry model, then providing the VGM layer for it is sufficient to obtain all the converters between this third geometry and already ported geometries (Geant4, Root).

The design and implementation of the VGM classes, the status of existing implementations for Geant4 and TGeo, and simple examples of usage will be presented.

**Distributed Computing Services / 388**

## **A Lightweight Monitoring and Accounting System for LHCb DC’04 Production**

**Authors:** J. Saborido-Silva<sup>1</sup>; M. Sanchez-Garcia<sup>2</sup>; R. Graciani-Diaz<sup>3</sup>; R. Vizcaya Carrillo<sup>3</sup>; V. Garonne<sup>4</sup>

<sup>1</sup> *University of Santiago de Compostela*

<sup>2</sup> *UNIVERSITY OF SANTIAGO DE COMPOSTELA*

<sup>3</sup> *University of Barcelona*

<sup>4</sup> *IN2P3*

**Corresponding Author:** manusan@usc.es

The LHCb Data Challenge 04 includes the simulation of over 200 M simulated events using distributed computing resources on N sites and extending along 3 months. To achieve this goal a dedicated Production grid (DIRAC) has been deployed. We will present the Job Monitoring and Accounting services developed to follow the status of the production along its way and to evaluate the results at the end of the Data Challenge.

The end user connects with a web browser to WEB-SERVER applications showing dynamic reports for a whole set of possible queries. These applications in turn interrogate the Job Monitoring

Service of the DIRAC Workload Management system and Accounting Database service by means of dedicated XML-RPC interfaces, querying for the information requested by the user. The reports provide an uniform view of the usage of the computing resources available. All the system components are implemented as a set of cooperating python classes following the design choice of LHCb. The different services are distributed over a number of independent machines. This allows to achieve the scalability level of multiple thousands of concurrent jobs monitored by the system

**Distributed Computing Systems and Experiences / 389**

## Deploying and operating LHC Computing Grid 2 (LCG2) During Data Challenges

**Authors:** CERN-IT-GD-GIS section<sup>1</sup>; E. Leonardi<sup>2</sup>; M. Schulz<sup>1</sup>

<sup>1</sup> CERN

<sup>2</sup> INFN Rome

**Corresponding Author:** markus.schulz@cern.ch

LCG2 is a large scale production grid formed by more than 40 worldwide distributed sites. The aggregated number of CPUs exceeds 3000 several MSS systems are integrated in the system. Almost all sites form an independent administrative domain. On most of the larger sites the local computing resources have been integrated into the grid.

The system has been used for large scale production by LHC experiments for several month.

During the operation the software went through several versions and had to be upgraded including non backward compatible upgrades.

We report on the experience gained setting up the service, integrating sites and operating it under the load of the production.

**Poster Session 1 - Board: 28 / 390**

## The Condor based CDF CAF

**Authors:** E. Lipeles<sup>1</sup>; F. Wuerthwein<sup>2</sup>; I. Sfiligoi<sup>3</sup>; M. Neubauer<sup>4</sup>

<sup>1</sup> UNIVERSITY OF CALIFORNIA SAN DIEGO

<sup>2</sup> UCSD

<sup>3</sup> INFN Frascati

<sup>4</sup> University of California, San Diego

**Corresponding Author:** fkw@fnal.gov

The CDF Analysis Facility (CAF) has been in use since April 2002 and has successfully served 100s of users on 1000s of CPUs. The original CAF used FBSNG as a batch manager.

In the current trend toward multisite deployment, FBSNG was found to be a limiting factor, so the CAF has been reimplemented to use Condor instead. Condor is a more widely used batch system and is well integrated with the emerging grid tools. One of the most useful being the ability to run seamlessly on top of other batch systems. The transition has brought us a lot of additional benefits, such as ease of installation, fault tolerance and increased manageability of the cluster. The CAF infrastructure has also been simplified a lot since Condor implements a number of features we had to implement ourselves with FBSNG. In addition, our users have found that Condor's fair share mechanism provides a more equitable and predictable distribution of resources. In this talk the Condor based CAF will be presented, with particular emphasis on the changes needed to run with Condor, the problems found during and the advantages gained by the transition. Some background and the plans for the future, as well as results from Condor scalability tests will also be presented.

**Poster Session 3 - Board: 57 / 391**

## **The FEDRA - Framework for Emulsion Data Reconstruction and Analysis in OPERA experiment.**

**Author:** V. Tioukov<sup>1</sup>

<sup>1</sup> *INFN NAPOLI*

**Corresponding Author:** [valeri@na.infn.it](mailto:valeri@na.infn.it)

OPERA is a massive lead/emulsion target for a long-baseline neutrino oscillation search. More than 90% of the useful experimental data in OPERA will be produced by the scanning of emulsion plates with the automatic microscopes. The main goal of the data processing in OPERA will be the search, analysis and identification of primary and secondary vertexes produced by neutrino in lead-emulsion target.

The volume of middle and high-level data to be analysed and stored is expected to be of the order of several Gb per event. The storage, calibration, reconstruction, analysis and visualization of this data is the task of FEDRA - system written in C++ and based on ROOT framework. The system is now actively used for processing of test beams and simulation data. Several interesting algorithmic solutions permits us to make very effective code for fast pattern recognition in heavy signal/noise conditions. The system consists of the storage part, intercalibration and segments linking part, track finding and fitting, vertex finding and fitting and kinematical analysis parts. Kalman Filtering technique is used for tracks&vertex fitting. ROOT-based event display is used for interactive analysis of the special events.

**Distributed Computing Services / 392**

## **File-Metadata Management System for the LHCb Experiment**

**Authors:** C. CIOFFI<sup>1</sup>; F. LOVERRE<sup>2</sup>; J. CLOSIER<sup>2</sup>; M. FRANK<sup>2</sup>; S. PONCE<sup>2</sup>

<sup>1</sup> *Oxford University*

<sup>2</sup> *CERN*

**Corresponding Author:** c.cioffi1@physics.ox.ac.uk

The LHCb experiment needs to store all the information about the datasets and their processing history of recorded data resulting from particle collisions at the LHC collider at CERN as well as of simulated data.

To achieve this functionality a design based on data warehousing techniques was chosen, where several user-services can be implemented and optimized individually without losing functionality nor performance. This approach results in an experiment-independent and flexible system. It allows fast access to the catalogue of available data, to detailed history information and to the catalogue of data replicas. Queries can be made based on these three sets of information. A flexible underlying database schema allows the implementation and evolution of these services without the need to change the basic database schema. The consequent implementation of interfaces based on XML-RPC allows to access and to modify the stored information using a well defined encapsulating API.

**Distributed Computing Services / 393**

## **Development and use of MonALISA high level monitoring services for the star unified Meta-Scheduler**

**Authors:** E. Efstathiadis<sup>1</sup>; I. Legrand<sup>2</sup>; L Hajdu<sup>3</sup>

<sup>1</sup> *BROOKHAVEN NATIONAL LABORATORY*

<sup>2</sup> *California Institute of Technology*

<sup>3</sup> *BNL*

**Corresponding Author:** stratos@bnl.gov

As a PPDG cross-team joint project, we proposed to study, develop, implement and evaluate a set of tools that allow Meta-Schedulers to take advantage of consistent information (such as information needed for complex decision making mechanisms) across both local and/or Grid Resource Management Systems (RMS).

We will present and define the requirements and schema by which one can consistently provide queue attributes for the most common batch systems (PBS, LSF, Condor, SGE, etc). We evaluate the best scalable and lightweight approach to access the monitored parameters from a client perspective and, in particular, the feasibility of accessing real-time and aggregate information using the MonALISA monitoring framework. Client programs are envisioned to function in a non-centralized, fault tolerant fashion. Inherent delays as well as scalability issues of each approach (implementing it at a large number of sites) will be discussed.

The MonALISA monitoring framework, being an ensemble of autonomous multi-threaded, agent based systems which are registered as dynamic services and are able to collaborate and cooperate in performing a wide range of monitoring tasks in a large scale distributed applications, is a natural choice for such a project. MonALISA is designed to easily integrate existing monitoring tools and procedures

and provide information in a dynamic self-describing way to any other service or client. We intend to demonstrate the usefulness of this consistent approach for queue monitoring by implementing a monitoring agent within the STAR Unified Meta-Scheduler (SUMS) framework.

We believe that such developments could highly benefit Grid laboratory efforts such as the Grid3+ and the OpenScience Grid (OSG).

**Poster Session 1 - Board: 31 / 394**

## Using HEP Systems to Provide Storage for Biologists

**Author:** M. Calef<sup>1</sup>

<sup>1</sup> *VANDERBILT UNIVERSITY*

**Corresponding Author:** matthew.t.calef@vanderbilt.edu

Protein analysis, imaging, and DNA sequencing are some of the branches of biology where growth has been enabled by the availability of computational resources. With this growth, biologists face an associated need for reliable, flexible storage systems. For decades the HEP community has been driving the development of such storage systems to meet their own needs. Two of these systems - the dCache disk caching system and the Enstore hierarchical storage manager - are viable candidates for addressing the storage needs of biologists. Both incorporate considerable experience from the HEP community.

While biologists have much to gain from the HEP community's experience with storage systems, they face several issues that are unique to the biological sciences. There is a wider diversity in experiments, in number and size of datafiles, and in client operating systems in biology than there is in HEP. Patient information must be kept confidential. Disparate IT departments set up firewalls that separate client systems and the storage system.

Vanderbilt University is developing a storage system with the goal of meeting biologists' needs. This system will use Enstore for its robustness and reliability, and will use the flexible door-based architecture of dCache to provide storage services to biologists via web-portal, the dCache copy command, and custom applications. This system will be deployed using an automated tape library, several secure central servers, and nodes placed near biologists' existing compute infrastructure to ensure locality of caches and secure data channels between researchers and the central servers.

**Event Processing / 395**

## WIRED 4 - A generic Event Display plugin for JAS 3

**Author:** M. Donszelmann<sup>1</sup>

<sup>1</sup> *SLAC*

**Corresponding Author:** mark.donszelmann@slac.stanford.edu

WIRED 4 is a experiment independent event display plugin module for JAS 3 (Java Analysis Studio) generic analysis framework. Both WIRED and JAS are written in Java.

WIRED, which uses HepRep (HEP Representables for Event Display) as its input format, supports viewing of events using either conventional 3D projections as well as specialized projections such as a fish-eye or a rho-Z projection. Projections allow the user to scale, rotate, position or change parameters on the plot as he wishes. All interactions are handled as separate edits which can be undone and/or redone, so the user can try out things and get back to a previous situation. All edits are scriptable by any of the scripting languages supported by JAS, such as pnuts, jython or java. Hits and tracks can be picked to display physics information and cuts can be made on physics parameters to allow the user to filter the number of objects drawn into the plot. Multiple event display plots can be laid out on pages combined with histograms and other plots, available from JAS itself or from other plugin modules. Configuration information on the state of all plots can be saved and restored allowing the user to save his session, share it with others or later continue where he left off.

This version of WIRED is written to be easily extensible by the user/developer. Projections, representations, interaction handlers and edits are all services and new ones can be added by writing additional plugins.

Both JAS 3 and WIRED 4 are built on top of the FreeHEP Java Libraries, which support a multitude of vector graphics output formats, such as PostScript, PDF, SVG, SWF and EMF, allowing document quality output of event display plots and histograms.

References:

<http://wired.freehep.org>  
<http://jas.freehep.org/jas3>  
<http://java.freehep.org>

**Poster Session 2 - Board: 20 / 396**

## **Job-monitoring over the Grid with GridIce infrastructure.**

**Authors:** A. Pierro<sup>1</sup>; G. Donvito<sup>2</sup>; G. Maggi<sup>1</sup>; G. Tortone<sup>3</sup>; M. Maggi<sup>1</sup>; N. De Filippis<sup>1</sup>; T. Coviello<sup>1</sup>

<sup>1</sup> *INFN Bari*

<sup>2</sup> *UNIVERSITÀ DEGLI STUDI DI BARI*

<sup>3</sup> *INFN Napoli*

**Corresponding Author:** [giacinto.donvito@ba.infn.it](mailto:giacinto.donvito@ba.infn.it)

In a wide-area distributed and heterogeneous grid environment, monitoring represents an important and crucial task. It includes system status checking, performance tuning, bottlenecks detecting, troubleshooting, fault notifying. In particular a good monitoring infrastructure must provide the information to track down the current status of a job in order to locate any problems. Job monitoring requires interoperation between the monitoring system and other grid services.

Currently development and deployment LCG testbeds integrate GridICE monitoring system which measures and publishes the state of a grid resource at a particular point in time. In this paper we present the efforts to integrate in the current GridICE infrastructure, additional useful information about job status, e.g. the name of job, the virtual organization to which it belongs, eventually real and

mapped user who has submitted the job, the effective CPU time consumed and its exit status.

**Core Software / 397**

## **AIDA, JAIDA and AIDAJNI: Data Analysis using interfaces**

**Authors:** M. Donszelmann<sup>1</sup>; M. Turri<sup>1</sup>; T. Johnson<sup>1</sup>; V. Serbo<sup>1</sup>

<sup>1</sup> SLAC

**Corresponding Authors:** mark.donszelmann@slac.stanford.edu, serbo@slac.stanford.edu

AIDA, Abstract Interfaces for Data Analysis, is a set of abstract interfaces for data analysis components: Histograms, Ntuples, Functions, Fitter, Plotter and other typical analysis categories. The interfaces are currently defined in Java, C++ and Python and implementations exist in the form of libraries and tools using C++ (Anaphe/Lizard, OpenScientist), Java (Java Analysis Studio) and Python (PAIDA).

JAIDA is the full implementation of AIDA in Java. It is used internally by JAS3 as its analysis core but it can also be used independently for either batch or interactive processing, or for web applications to access data, make plots and simple data analysis through a browser. Some of the JAIDA features are the ability to open AIDA, ROOT and PAW files and the support of an extensible set of fit methods (chi-square, least squares, binned/unbinned likelihood, etc) to be matched with an extensible set of optimizers including Minuit and Uncmin.

AIDAJNI is glue code between C++ and Java that allows any C++ code to access any Java implementation of the AIDA interfaces. For example AIDAJNI is used with Geant4 to access the JAIDA implementation of AIDA.

This paper gives an update on the AIDA 3.2.1 interfaces and its corresponding JAIDA implementation. Examples will be provided on how to use JAIDA within JAS3, as a standalone library and from C++ using AIDAJNI.

References:

<http://aida.freehep.org/>  
<http://java.freehep.org/jaida>  
<http://java.freehep.org/aidajni>  
<http://jas.freehep.org/jas3>

**Event Processing / 399**

## **JASSimApp plugin for JAS3: Interactive Geant4 GUI**

**Authors:** E. Suzuki<sup>1</sup>; H Yoshidah<sup>1</sup>; K. Minamimoto<sup>1</sup>; K. Murakami<sup>2</sup>; M. Donszelmann<sup>3</sup>; M. Nagamatsu<sup>1</sup>; M. Turri<sup>3</sup>; T. Johnson<sup>3</sup>; V. Serbo<sup>3</sup>

<sup>1</sup> *Naruto University of Education*

<sup>2</sup> *KEK*

<sup>3</sup> *SLAC*

**Corresponding Authors:** serbo@slac.stanford.edu, mark.donszelmann@slac.stanford.edu

JASSimApp is joint project of SLAC, KEK, and Naruto University to create integrated GUI for Geant4, based on JAS3 framework, with ability to interactively:

- Edit Geant4 geometry, materials, and physics processes
- Control Geant4 execution, local and remote: pass commands and receive output, control event loop
- Access AIDA histograms defined in Geant4
- Show generated Geant4 events in integrated event display

JAS3 is the latest development of JAS, a general-purpose data analysis tool. It employs a highly modular component-based framework and allows flexible and powerful customized plugin modules. JASSimApp is a concrete implementation of its design concept, Geant4 as the problem domain. It is composed of existing interactive tools like : GAG, Gain, Momo, WIRED etc.. A new C++ class of the Geant4 interfaces category was developed to exploit multi-threaded control over Geant4 execution. The plugin modules of JAS3 reused existing classes with little modification.

References:

<http://erpc1.naruto-u.ac.jp/~geant4/index.html> ; <http://jas.freehep.org/jas3> ;  
<http://wired.freehep.org> ; <http://wwwasd.web.cern.ch/wwwasd/geant4/geant4.html>

**Core Software / 400**

## The FreeHEP Java Library Root IO package

**Authors:** M. Donszelmann<sup>1</sup>; M. Turri<sup>1</sup>; T. Johnson<sup>1</sup>; V. Serbo<sup>1</sup>

<sup>1</sup> SLAC

The FreeHEP Java library contains a complete implementation of Root IO for Java. The library uses the “Streamer Info” embedded in files created by Root 3.x to dynamically create high performance Java proxies for Root objects, making it possible to read any Root file, including files with user defined objects. In this presentation we will discuss the status of this code, explain its implementation and demonstrate performance using benchmark comparisons to standard Root IO. We will also describe recently added support for reading files remotely using rootd and xrootd protocols.

We will also show some uses of this library, including using JAS3 to analyze Root data, using the WIRED event display to visualize data from Root files and using rootd and Java servlet technology to make live plots web accessible - with examples from GLAST and BaBar. We will also explain how you can trivially make your own root data web-accessible using the AIDA Tag Library and Jakarta Tomcat.

**Distributed Computing Systems and Experiences / 401**

## Interactive Data Analysis on the Grid using Globus 3 and JAS3

**Authors:** B. Ananthan<sup>1</sup>; D. Alexander<sup>1</sup>; M. Turri<sup>2</sup>; T. Johnson<sup>2</sup>; V. Serbo<sup>2</sup>



<sup>1</sup> *Tech-X corporation*

<sup>2</sup> *SLAC*

The aim of the service is to allow fully distributed analysis of large volumes of data while maintaining true (sub-second) interactivity. All the Grid related components are based on OGSA style Grid services, and to the maximum extent uses existing Globus Toolkit 3.0 (GT3) services. All transactions are authenticated and authorized using GSI (Grid Security Infrastructure) mechanism - part of GT3. JAS3, and experiment independent data analysis tool is used as the interactive analysis client.

The system consists of three main service components:

**Dataset Catalog Service:**

The Dataset Catalog supports browsing for an interesting dataset, or searching for data using a query language which operates on metadata stored in the catalog. The catalog makes few assumptions about the metadata stored in the catalog, except that the metadata consists of key-value pairs, stored in a hierarchical tree. The Dataset Catalog Service is designed to allow easy interfacing to existing data catalog back-ends.

**Dataset Analysis Grid Service:**

This service is responsible for resolving the dataset id from the catalog service, and transferring chunks of data to worker nodes for analysis processing. This service also manages the worker nodes, distributes analysis code to the worked nodes and retrieves intermediate results from the worker nodes before sending merged results back to the analysis client.

**Worker Execution Services:**

This service runs on each worker node and is responsible for processing analysis requests.

In this presentation we will demonstrate the current system, and will describe some of the choices made in architecting the system, in particular the challenges of obtaining interactive response times from GT3.

## Event Processing / 402

### Writing Extension Modules (Plug-ins) for JAS3

**Author:** T. Johnson<sup>1</sup>

**Co-authors:** Mark Donszelmann<sup>1</sup>; Massimiliano Turri<sup>1</sup>; Victor Serbo<sup>1</sup>

<sup>1</sup> *SLAC*

**Corresponding Author:** [tonyj@slac.stanford.edu](mailto:tonyj@slac.stanford.edu)

JAS3 is a general purpose, experiment independent, open-source, data analysis tool. JAS3 includes a variety of features, including histogramming, plotting, fitting, data access, tuple analysis, spreadsheet and event display capabilities. More complex analysis can be performed using several scripting languages (pnuts, jython, etc.), or by writing Java analysis classes. All of these features are provided by loosely coupled “plug-in” modules which are installed into the JAS3 base application framework.

In this presentation we will describe the JAS3 plug-in architecture, and explain how different plug-ins can interact via service interfaces and event dispatch mechanisms. We will demonstrate how this architecture makes it possible for individual plug-ins to be added, removed or upgraded to customize the application.

We will then give an overview of how to design new experiment or domain specific plug-ins to extend the functionality of JAS3 for your own requirements, or to provide general purpose components for use by others.

**Distributed Computing Systems and Experiences / 403**

## Results of the LHCb experiment Data Challenge 2004

**Authors:** A. Pickford<sup>1</sup>; A. Saroka<sup>2</sup>; A. Tsaregorodtsev<sup>3</sup>; G. Kuznetsov<sup>4</sup>; G.N. Patrick<sup>5</sup>; I. Korolko<sup>6</sup>; I. Stokes-Rees<sup>2</sup>; J. Blouw<sup>7</sup>; J. Closier<sup>8</sup>; J. Saborido-Silva<sup>9</sup>; M. Frank<sup>8</sup>; M. Gandelman<sup>10</sup>; M. Sanchez-Garcia<sup>9</sup>; M. Tobin<sup>11</sup>; M. Witek<sup>12</sup>; N. Brook<sup>13</sup>; P. Charpentier<sup>8</sup>; R. Bernet<sup>14</sup>; R. Graciani-Diaz<sup>15</sup>; U. Egede<sup>16</sup>; V. Garonne<sup>3</sup>; V. Romanovski<sup>17</sup>; V. Vagnoni<sup>18</sup>

<sup>1</sup> *University of Glasgow*

<sup>2</sup> *University of Oxford*

<sup>3</sup> *Université d'Aix - Marseille II*

<sup>4</sup> *Rutherford Appleton Lab, Chilton*

<sup>5</sup> *Rutherford Appleton Laboratory*

<sup>6</sup> *ITEP, Moscow*

<sup>7</sup> *Max-Planck-Institut für Kernphysik (MPI)*

<sup>8</sup> *CERN*

<sup>9</sup> *Universidad de Santiago de Compostela*

<sup>10</sup> *Universidade Federal do Rio de Janeiro (UFRJ)*

<sup>11</sup> *University of Liverpool*

<sup>12</sup> *High Energy Department Krakow*

<sup>13</sup> *H.H. Wills Physics Laboratory*

<sup>14</sup> *Universität Zürich*

<sup>15</sup> *University of Barcelona*

<sup>16</sup> *Imperial College*

<sup>17</sup> *State Res.Center of Russian Feder. Inst.f.High Energy Phys. (IFVE)*

<sup>18</sup> *INFN*

**Corresponding Author:** joel.closier@cern.ch

The LHCb experiment performed its latest Data Challenge (DC) in May-July 2004. The main goal was to demonstrate the ability of the LHCb grid system to carry out massive production and efficient distributed analysis of the simulation data.

The LHCb production system called DIRAC provided all the necessary services for the DC: Production and Bookkeeping Databases, File catalogs, Workload and Data Management systems, Monitoring and Accounting tools. It allowed to combine in a consistent way resources of more than 20 LHCb production sites as well as the LCG2 grid resources. 200M events constituting 90 TB of data were produced and stored in 6 Tier 1 centers. The subsequent analysis was carried out at CERN as well as in all the Tier 1 centers to where preselected datasets were distributed. The GANGA User Interface was used to assist users in preparation of their analysis jobs and running them on the local and remote computing resources.

We will present the DC results, the experience gained utilising DIRAC and LCG2 grids as well as further developments necessary to achieve the scalability level of the real running LHCb experiment.

Poster Session 3 - Board: 27 / 404

## Visualisation of the ROOT geometries (TVolume and TGeoVolume) with Coin-based model.

**Authors:** A. Kulikov<sup>1</sup>; D. Turkin<sup>1</sup>; M. Panebrattsev<sup>1</sup>; V. Fine<sup>2</sup>

<sup>1</sup> *Joint Institute for Nuclear Research, Dubna, Russia.*

<sup>2</sup> *Brookhaven National Laboratory, Upton, USA*

**Corresponding Author:** shurik@sunhe.jinr.ru

Using the modern 3D visualization software and hardware to represent the object models of the HEP detectors would create the impressive pictures of events and the detail views of the detectors facilitating the design, simulation and data analysis and representation the huge amount of the information flooding the modern HEP experiments. In this paper we represent the work made by members of STAR collaboration from Laboratory of High Energy Physics JINR Dubna. This work devoted to visualisation of the STAR detector geometry. Initially the detector geometry is described by means of specific AGE - geometry specification language and it can be converted to either TVolume or TGeoVolume type object of ROOT environment using of specially developed software. We created class library for conversion of the ROOT OO model of the detector from ROOT environment to the text "iv" file. Our class library assumes the conversion of ROOT OO models to Coin-based C++ representation and the Coin-based 3d Viewer with cutting / highlighting / selecting pieces of the image features. Since the class library implementation is free of the STAR experiment specific it can be used to visualize any detector geometry represented in the ROOT environment. The results of our work can be downloaded from the LHE web server.

Event Processing / 405

## Full Event Reconstruction in Java

**Author:** N. Graf<sup>1</sup>

<sup>1</sup> *SLAC*

**Corresponding Author:** norman.graf@slac.stanford.edu

We describe a Java toolkit for full event reconstruction and analysis. The toolkit is currently being used for detector design and physics analysis for a future linear e+ e- linear collider. The components are fully modular and are available for tasks from digitization of tracking detector signals through to cluster finding, pattern recognition, fitting, jetfinding, and analysis. We discuss the architecture as well as the implementation for several candidate detector designs.

Poster Session 3 - Board: 18 / 406

## Monte Carlo Event Generation in a Multilanguage, Multiplatform Environment

**Authors:** A. Johnson<sup>1</sup>; N. Graf<sup>1</sup>

<sup>1</sup> SLAC

**Corresponding Author:** norman.graf@slac.stanford.edu

We discuss techniques used to access legacy event generators from modern simulation environments. Examples will be given of our experience within the linear collider community accessing various FORTRAN-based generators from within a Java environment. Coding to a standard interface and use of shared object libraries enables runtime selection of generators, and allows for extension of the suite of available generators without having to rewrite core code.

**Event Processing / 409**

## LCG Generator

**Authors:** A. RIBON<sup>1</sup>; A. SHERSTNEV<sup>2</sup>; A. VOLOGDIN<sup>2</sup>; F. AMBROGLINI<sup>3</sup>; F. MOORTGAT<sup>1</sup>; H. NAVES<sup>4</sup>; J. CUEVAS MAESTRO<sup>4</sup>; L. DUDKO<sup>2</sup>; M. KIRSANOV<sup>5</sup>; P. Bartalini<sup>6</sup>; S. BELOV<sup>7</sup>; S. KOROBV<sup>7</sup>; S. MAKARYCHEV<sup>8</sup>; V. UZHINSKY<sup>7</sup>

<sup>1</sup> CERN, Geneva, Switzerland

<sup>2</sup> SINP MSU, Moscow, Russia

<sup>3</sup> Perugia University, Italy

<sup>4</sup> Cantabria University, Santander, Spain

<sup>5</sup> INR RAS Troitsk, Moscow, Russia

<sup>6</sup> CERN

<sup>7</sup> JINR, Dubna, Russia

<sup>8</sup> ITEP, Moscow, Russia

**Corresponding Author:** paolo.bartalini@cern.ch

In the framework of the LCG Simulation Project, we present the Generator Services Sub-project, launched in 2003 under the oversight of the LHC Monte Carlo steering group (MC4LHC). The goal of the Generator Services Subproject is to guarantee the physics generator support for the LHC experiments. Work is divided into four work packages: Generator library; Storage, event interfaces and particle services; Public event files and event database; Validation and tuning. The current status and the future plans in the four different work packages are presented. Some emphasis is put on the Monte Carlo Generator Library (GENSER) and on the Monte Carlo Generator Database (MCDB).

GENSER is the central code repository for Monte Carlo generators and generator tools. It was the first CVS repository in the LCG Simulation project and it is currently distributed in AFS. GENSER comprises release and building tools for librarians and end users. GENSER is going to gradually replace the obsolete CERN library in Monte Carlo generators support.

MCDB is a public database for the configuration, book-keeping and storage of the generator level event files. The generator events often need to be prepared and documented by Monte Carlo experts. MCDB aims at facilitating the communication between Monte-Carlo experts and end-users. Its use can be optionally extended to the official event production of the LHC experiments.

**Poster Session 1 - Board: 10 / 410**

## Status of the alignment calibrations in the ATLAS-Muon experiment

**Authors:** B. DUBOUE<sup>1</sup>; C. GUYOT<sup>1</sup>; J.C BARRIERE<sup>1</sup>; J.P. SCHULLER<sup>1</sup>; O. CLOUE<sup>1</sup>; P. PERRIN<sup>1</sup>; P. PONSOT<sup>1</sup>; P. SCHUNE<sup>1</sup>; V. GAUTARD<sup>2</sup>; Y. REINERT<sup>1</sup>

<sup>1</sup> DAPNIA, CEA-Saclay, France

<sup>2</sup> CEA-SACLAY

**Corresponding Author:** gautard@cea.fr

ATLAS is a particle detector which will be built at CERN in Geneva. The muon detection system is made up among other things, of 600 chambers measuring 2 to 6 m<sup>2</sup> and 30 cm thick. The chambers' position must be known with an accuracy of +/- 30  $\mu$ m for translations and +/-100  $\mu$ rad for rotations for a range of +/- 5mm and +/-5mrad. In order to fulfill these requirements, we have designed different optical sensors.

Due to (i) the very high accuracy required, (ii) the number of sensors (over 1000) and (iii) the different type of sensors, we developed one user interface which manages among other things several control command software. Each of this software is associated with an accurate calibration bench. In this conference, we will present only the most complex one which combines command control, an analysis module, real time processing and database access. These softwares are now currently used for sensors calibration.

Online Computing / 411

## The High Level Filter of the H1 Experiment at HERA

**Authors:** A. Campbell<sup>1</sup>; M. Vorobiev<sup>2</sup>; S. Levonian<sup>1</sup>

<sup>1</sup> DESY

<sup>2</sup> DESY,IITEP

We present the scheme in use for online high level filtering, event reconstruction and classification in the H1 experiment at HERA since 2001.

The Data Flow framework ( presented at CHEP2001 ) will be reviewed. This is based on CORBA for all data transfer, multi-threaded C++ code to handle the data flow and synchronisation and fortran code for reconstruction and event selection. A controller written in python provides setup,initialisation and process management. Specialised java programs provide run control and online access to and display of histograms. A C++ logger program provides central logging of standard printout from all processes.

We show how the system handles online preparation and update of detector calibration and beam parameter data. Newer features are the selection of rare events for the online event display and the extension to multiple input sources and output channels.

We discuss how the system design provides automatic recovery from various failures and show the overall and long term performance.

In addition we present the framework of event selection and classification and the features it provides.

**Core Software / 412**

## XML I/O in ROOT

**Authors:** H. Essel<sup>1</sup>; R. Brun<sup>2</sup>; S. Linev<sup>3</sup>

<sup>1</sup> *GSI, Darmstadt*

<sup>2</sup> *CERN*

<sup>3</sup> *GSI*

**Corresponding Authors:** s.linev@gsi.de, h.essel@gsi.de

Till now, ROOT objects can be stored only in a binary ROOT specific file format. Without the ROOT environment the data stored in such files are not directly accessible. Storing objects in XML format makes it easy to view and edit (with some restriction) the object data directly. It is also plausible to use XML as exchange format with other applications. Therefore XML streaming has been implemented in ROOT. Any object which is in the ROOT dictionary can be stored/retrieved in XML format. Two layouts of object representation in XML are supported: class-dependent and generic. In the first case all XML tag names are derived from class and member names. To avoid name intersections, XML namespaces for each class are used. A Document Type Definition (DTD) file is automatically generated for each class (or set of classes). It can be used to validate the structure of the XML document. The generic layout of XML files includes tag names like "Object", "Member", "Item" and so on. In this case the DTD is common for all produced XML files. Further development is required to provide tools for accessing created XML files from other applications like: pure C++ code without ROOT libraries and dictionaries, Java and so on.

**Distributed Computing Services / 414**

## Data Management in EGEE

**Authors:** A. Frohner<sup>1</sup>; G. McCance<sup>1</sup>; K. Nienartowicz<sup>1</sup>; P. Badino<sup>1</sup>; P. Kunszt<sup>1</sup>; R. Da Rocha<sup>1</sup>

<sup>1</sup> *CERN*

**Corresponding Authors:** krzysztof.nienartowicz@cern.ch, gavin.mccance@cern.ch

Data management is one of the cornerstones in the distributed production computing environment that the EGEE project aims to provide for a European e-Science infrastructure. We have designed a set of services based on previous experience in other Grid projects, trying to address the requirements of our user communities.

In this paper we summarize the most fundamental requirements and constraints as well as the security, reliability, stability and robustness considerations that have driven the architecture and the particular choice for service decomposition in our service-oriented architecture. We discuss the interaction of our services with each other, their deployment models and how failures are being managed.

The three service groups for data management services are the Storage Element, the Data Scheduling and the Catalog services. The Storage Element exposes interfaces to Grid managed storage, with the appropriate semantics in the Grid distributed

environment. The Catalog services contain all the metadata related to data: The File Catalog maintains a file-system-like view of the files in the Grid in a logical user namespace, the Replica Catalog keeps track of identical copies of the files distributed in different Storage Elements and the Metadata Catalog keeps application specific information about the files. The Data Scheduling services take care of controlled data transfer and keep the information in the Catalog services consistent with what is actually available in the Storage Elements, acting as the binding between the two.

We conclude with first experiences and examples of use-cases for High Energy Physics applications.

**Poster Session 3 - Board: 56 / 416**

## Software Management in the HARP experiment

**Author:** E. Tcherniaev<sup>1</sup>

<sup>1</sup> *CERN*

**Corresponding Author:** evgueni.tcherniaev@cern.ch

This paper discusses some key points in the organization of the HARP software. In particular it describes the configuration of the packages, data and code management, testing and release procedures.

Development of the HARP software is based on incremental releases with strict respect of the design structure.

This poses serious challenges to the software management, which has gone through essential evolution during the life of the experiment.

A progressively better understanding of the organizational issues, like the environment settings, package versioning, release procedures, etc., was achieved.

Mastering of the CVS and CMT tools, plus an essential reduction of the manual work allowed to reach the situation where one software development iteration (compilation from scratch, full testing, installation in the official area) takes only a few hours.

**Distributed Computing Services / 417**

## Information and Monitoring Services within a Grid Environment

**Authors:** A. Cooke<sup>1</sup>; A. Djaoui<sup>2</sup>; A. Gray<sup>1</sup>; A. Wilson<sup>2</sup>; B. Coghlan<sup>3</sup>; D. O'Callaghan<sup>3</sup>; J. Leake<sup>4</sup>; J. Magowan<sup>5</sup>; J. Walk<sup>2</sup>; L. Cornwall<sup>2</sup>; M. Craig<sup>2</sup>; N. Podhorszki<sup>6</sup>; O. Lyttleton<sup>3</sup>; P. Taylor<sup>5</sup>; R. Byrom<sup>2</sup>; R. Cordenonsi<sup>7</sup>; R. Middleton<sup>2</sup>; S. Fisher<sup>8</sup>; S. Hicks<sup>2</sup>; S. Kenny<sup>3</sup>; W. Nutt<sup>1</sup>

<sup>1</sup> *Heriot Watt*

<sup>2</sup> *Rutherford Appleton Laboratory*

<sup>3</sup> *Trinity College Dublin*

<sup>4</sup> *Objective Engineering Ltd.*

<sup>5</sup> *IBM*

<sup>6</sup> *SZTAKI*

<sup>7</sup> *Queen Mary, University of London*

<sup>8</sup> *RAL*

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

The R-GMA (Relational Grid Monitoring Architecture) was developed within the EU DataGrid project, to bring the power of SQL to an information and monitoring system for the grid. It provides producer and consumer services to both publish and retrieve information from anywhere within a grid environment. Users within a Virtual Organization may define their own tables dynamically into which to publish data.

Within the DataGrid project R-GMA was used for the information system, making details about grid resources available for use by other middleware components. R-GMA has also been used for monitoring grid jobs by members of the CMS and D0 collaborations where information about jobs is published from within a job wrapper, transported across the grid by R-GMA and made available to users. An accounting package for processing PBS logging data and sending it to one or more Grid Operation Centres using R-GMA has been written and is being deployed within LCG. There are many other existing and potential applications.

R-GMA is currently being re-engineered to fit into a Web Service environment as part of the EU EGEE project. Improvements being developed include fine grained authorization, an improved user interface and measures to ensure superior scaling behaviour.

**Core Software / 418**

## Go4 analysis design

**Author:** H. Essel<sup>1</sup>

**Co-authors:** D. Bertini<sup>1</sup>; J. Adamczewski<sup>1</sup>; M. Al-Turany<sup>1</sup>; S. Linev<sup>1</sup>

<sup>1</sup> *GSI*

**Corresponding Author:** h.essel@gsi.de

The GSI online-offline analysis system Go4 is a ROOT based framework for medium energy ion- and nuclear physics experiments. Its main features are a multithreaded online mode with a non-blocking Qt GUI, and abstract user interface classes to set up the analysis process itself which is organised as a list of subsequent analysis steps. Each step has its own event objects and a processor instance. It can handle its event i/o independently. It can be set up by macros or by generic a GUI. With respect to the more complex experiments planned at GSI, a configurable network of steps is required. Multiple IO channels per step and multiple references to steps can be set up by macros or via generic GUI. The required mechanisms are provided by an upgrade of the Go4 analysis step manager using the new ROOT TTasks. Support for IO configuration and references across the task tree is provided.

**Poster Session 2 - Board: 40 / 419**

## Distributed Tracking, Storage, and Re-use of Job State Information on the Grid

**Authors:** A. GUARISE<sup>1</sup>; A. Gianelle<sup>2</sup>; A. Krenek<sup>3</sup>; A. MARASCHINI<sup>4</sup>; A. Terracina<sup>5</sup>; D. Kouril<sup>3</sup>; D. Rebatto<sup>6</sup>; E. Ronchieri<sup>7</sup>; F. GIACOMINI<sup>8</sup>; F. Pacini<sup>5</sup>; F. Prezl<sup>6</sup>; G. Avellino<sup>5</sup>; G. Patania<sup>9</sup>; J. Pospisil<sup>3</sup>; J. Sitera<sup>3</sup>; J. Skrabal<sup>3</sup>; L.



Matyska<sup>3</sup>; L. Zangrando<sup>2</sup>; M. MARCHI<sup>10</sup>; M. MEZZADRI<sup>10</sup>; M. MORDACCHINI<sup>11</sup>; M. MULAC<sup>12</sup>; M. Pappalardo<sup>13</sup>; M. Ruda<sup>3</sup>; M. Sgaravatto<sup>2</sup>; M. Vocu<sup>3</sup>; P. Andreetto<sup>2</sup>; S. Andreozzi<sup>7</sup>; S. Beco<sup>5</sup>; S. Borgia<sup>2</sup>; S. MONFORTE<sup>14</sup>; V. Ciaschini<sup>7</sup>; Z. Salvet<sup>3</sup>

<sup>1</sup> INFN Torino

<sup>2</sup> INFN Padova, Italy

<sup>3</sup> CESNET, CZECH REPUBLIC

<sup>4</sup> DATAMAT

<sup>5</sup> DATAMAT, Italy

<sup>6</sup> INFN Milano, Italy

<sup>7</sup> INFN Cnaf, Italy

<sup>8</sup> INFN Cnaf

<sup>9</sup> INFN Torino, Italy

<sup>10</sup> INFN Milano

<sup>11</sup> INFN Padova

<sup>12</sup> CESNET

<sup>13</sup> INFN Catania, Italy

<sup>14</sup> INFN Catania

**Corresponding Author:** ljocha@ics.muni.cz

The Logging and Bookkeeping service tracks job passing through the Grid. It collects important events generated by both the grid middleware components and applications, and processes them at a chosen L&B server to provide the job state. The events are transported through secure reliable channels. Job tracking is fully distributed and does not depend on a single information source, the robustness is achieved through speculative job state computation in case of reordered, delayed or lost events. The state computation is easily adaptable to modified job control flow.

The events are also passed to the related Job Provenance service. Its purpose is a long-term storage of information on job execution, environment, and the executable and input sandbox files. The data can be used for debugging, post-mortem analysis, or re-running jobs. The data are kept by the job-provenance storage service in a compressed format, accessible on per-job basis. A complementary index service is able to find particular jobs according to configurable criteria, e.g. submission time or “tags” assigned by the user. A user client to support job re-execution is planned.

Both the L&B and Job Provenance index server provide web-service interfaces for querying. Those interfaces comply with the On-demand producer specification of the R-GMA infrastructure. Hence R-GMA capabilities can be utilized to perform complex distributed queries across multiple servers. Also, aggregate information about job collections can be easily provided.

The L&B service was deployed in the EU DataGrid and Cern LCG projects, the Job Provenance will be deployed in the EGEE project.

**Distributed Computing Services / 420**

## Practical approaches to Grid workload and resource management in the EGEE project

**Authors:** A. GUARISE<sup>1</sup>; A. Gianelle<sup>2</sup>; A. KRENEK<sup>3</sup>; A. MARASCHINI<sup>4</sup>; A. TERRACINA<sup>4</sup>; D. KOURIL<sup>3</sup>; D. REBATO<sup>5</sup>; E. RONCHIERI<sup>6</sup>; F. GIACOMINI<sup>6</sup>; F. PACINI<sup>4</sup>; F. PRELZ<sup>5</sup>; G. AVELLINO<sup>4</sup>; G. PATANIA<sup>1</sup>; J. POSPISIL<sup>3</sup>; J. SITERA<sup>3</sup>; J. SKRABAL<sup>3</sup>; L. MATYSKA<sup>3</sup>; L. ZANGRANDO<sup>7</sup>; M. MARCHI<sup>5</sup>; M. MEZZADRI<sup>5</sup>; M. MORDACCHINI<sup>7</sup>; M. MULAC<sup>3</sup>;

M. PAPPALARDO<sup>8</sup>; M. RUDA<sup>3</sup>; M. Sgaravatto<sup>7</sup>; M. VUCU<sup>3</sup>; P. ANDREETTO<sup>7</sup>; S. ANDREOZZI<sup>6</sup>; S. BECO<sup>4</sup>; S. BORGIA<sup>7</sup>; S. MONFORTE<sup>8</sup>; V. CIASCHINI<sup>6</sup>; Z. SALVET<sup>3</sup>

<sup>1</sup> INFN Torino

<sup>2</sup> INFN Padova, Italy

<sup>3</sup> CESNET

<sup>4</sup> DATAMAT

<sup>5</sup> INFN Milano

<sup>6</sup> INFN Cnaf

<sup>7</sup> INFN Padova

<sup>8</sup> INFN Catania

**Corresponding Author:** massimo.sgaravatto@pd.infn.it

Resource management and scheduling of distributed, data-driven applications in a Grid environment are challenging problems. Although significant results were achieved in the past few years, the development and the proper deployment of generic, reliable, standard components present issues that still need to be completely solved. Interested domains include workload management, resource discovery, resource matchmaking and brokering, accounting, authorization policies, resource access, reliability and dependability. The evolution towards a service-oriented architecture, supported by emerging standards, is another activity that will demand attention.

All these issues are being tackled within the EU-funded EGEE project (Enabling Grids for E-science in Europe), whose primary goals are the provision of robust middleware components and the creation of a reliable and dependable Grid infrastructure to support e-Science applications.

In this paper we present the plans and the preliminary activities aiming at providing adequate workload and resource management components, suitable to be deployed in a production-quality Grid.

**Event Processing / 421**

## The ZEUS Global Tracking Trigger Barrel Algorithm

**Author:** M. Sutton<sup>1</sup>

<sup>1</sup> UNIVERSITY COLLEGE LONDON

**Corresponding Author:** sutt@cern.ch

The current design, implementation and performance of the ZEUS global tracking trigger barrel algorithm are described. The ZEUS global tracking trigger integrates track information from the ZEUS central tracking chamber (CTD) and micro vertex detector (MVD) to obtain a global picture of the track topology in the ZEUS detector at the second level trigger stage. Algorithm processing is performed on a farm of Linux PCs and, to avoid unacceptable deadtime in the ZEUS readout system, must be completed within the strict requirements of the ZEUS trigger system. The GTT plays a vital role in the selection of good physics events and the rejection of non-physics background within the very harsh trigger environment provided by the upgraded HERA collider. The GTT barrel algorithm greatly improves the vertex resolution and the track finding efficiency of the ZEUS second level trigger while the mean event processing latency and throughput are

well within the trigger requirements. Recent running experience with HERA production luminosity is briefly discussed.

**Online Computing / 422**

## The Architecture of the ZEUS Second Level Global Tracking Trigger

**Authors:** M. Sutton<sup>1</sup>; ZEUS GTT group<sup>None</sup>

<sup>1</sup> UNIVERSITY COLLEGE LONDON

**Corresponding Author:** [sutt@cern.ch](mailto:sutt@cern.ch)

The architecture and performance of the ZEUS Global Track Trigger (GTT) are described. Data from the ZEUS silicon Micro Vertex detector's HELIX readout chips, corresponding to 200k channels, are digitized by 3 crates of ADCs and PowerPC VME board computers push cluster data for second level trigger processing and strip data for event building via Fast and GigaEthernet network connections. Additional tracking information from the central tracking chamber and forward straw tube tracker are interfaced into the 12 dual CPU PC farm of the global track trigger where track and vertex finding is performed by separately threaded algorithms. The system is data driven at the ZEUS first level trigger rates <500Hz, generating trigger results after a mean time of 10ms. The GTT integration into the ZEUS second level trigger and recent performance are reviewed.

**Poster Session 3 - Board: 55 / 423**

## The ZEUS Global Tracking Trigger

**Author:** M. Sutton<sup>1</sup>

<sup>1</sup> UNIVERSITY COLLEGE LONDON

**Corresponding Author:** [sutt@cern.ch](mailto:sutt@cern.ch)

The design, implementation and performance of the ZEUS Global Tracking Trigger (GTT) Forward Algorithm is described. The ZEUS GTT Forward Algorithm integrates track information from the ZEUS Micro Vertex Detector (MVD) and forward Straw Tube Tracker (STT) to provide a picture of the event topology in the forward direction ( $1.5 < \eta < 3$ ) of the ZEUS detector. This region is particularly challenging because of inhomogeneities in the solenoid magnetic field, and the high occupancies in the forward direction from beam-gas interactions and secondary scatters with the ZEUS beampipe. The forward algorithm is distinct from the GTT barrel algorithm, but will run in parallel on the GTT CPU farm. To avoid unacceptable downtime in the ZEUS readout system, the forward algorithm processing must be compliant with the strict requirements of the ZEUS trigger system. The current status of the integration with the ZEUS DAQ and trigger systems is also reviewed.

Core Software / 424

## The AliRoot framework, status and perspectives

**Authors:** A. Morsch<sup>1</sup>; F. Carminati<sup>1</sup>; F. Rademakers<sup>1</sup>; K. Safarik<sup>1</sup>; P. Buncic<sup>1</sup>; P. Hristov<sup>1</sup>

<sup>1</sup> CERN

**Corresponding Author:** federico.carminati@cern.ch

The ALICE collaboration at the LHC is developing since 1998 an OO offline framework, written entirely in C++.

In 2001 a GRID system (AliEn - ALICE Environment) has been added and successfully integrated with ROOT and the offline. The resulting combination allows ALICE to do most of the design of the detector and test the validity of its computing model by performing large scale Data Challenges, using OO technology in a distributed framework. The early migration of all ALICE users to C++ and the adoption of advanced software development techniques are two of the strong points of the ALICE offline strategy. The offline framework is heavily based on virtual interfaces, which allows the use of different generators and even different Monte-Carlo transport codes with no change in the framework or the scoring, reconstruction and analysis code. This talk presents a review of the development path, current status and future perspectives of the ALICE Offline environment.

Poster Session 3 - Board: 21 / 425

## GraXML

**Author:** J. Hrivnac<sup>1</sup>

<sup>1</sup> LAL

**Corresponding Author:** julius.hrivnac@cern.ch

GraXML is the framework for manipulation and visualization of 3D geometrical objects in space. The full framework consists of the GraXML toolkit, libraries implementing Generic and Geometric Models and end-user interactive front-ends. GraXML Toolkit provides a foundation for operations on 3D objects (both detector elements and events). Each external source of 3D data is automatically translated into Generic Model which is then analyzed and translated into Geometric Model using GraXML modules. The construction of this Geometric Model is parametrised by several parameters (optimization level, quality level, ...) so that it can be used in applications with different requirements (graphical or not). Two visualization applications are provided in the GraXML framework: GraXML Interactive Display and GraXML Converter into various 3D geometry formats. Other applications can be easily developed.

The presentation will concentrate on GraXML graphical capabilities and relation with geometric data providers. The difference between specific GraXML features and properties of other similar tools will be highlighted. The questions of different visualization needs and possibilities for different kinds of geometrical data will be also explained.

**Core Software / 426**

## Aspects

**Author:** J. Hrivnac<sup>1</sup><sup>1</sup> LAL**Corresponding Author:** julius.hrivnac@cern.ch

Aspect-Oriented Programming (AOP) is a new paradigm promising to allow further modularization of large software frameworks, like those developed in HEP. Such frameworks often manifest several orthogonal axes of contracts (Crosscutting Concerns - CC) leading to complex multidependencies. Currently used programming languages and development methodologies don't allow to easily identify and encapsulate such CC. AOP offers ways to solve CC problems by identifying places where they appear (Joint Points) and specifying actions to be applied at those places (Advices). While Aspects can be added in principle to any programming paradigm, they are mostly used in Object-Oriented environments. Thanks to wide acceptance and rich object model, most Aspect-Oriented toolkits have been developed for Java language. Probably the most used AOP language is AspectJ.

The presentation will demonstrate using AspectJ language to solve several common HEP Crosscutting Concerns from simple cases (like logging or debugging) to complex ones (like object persistency, data analysis or graphics).

**Poster Session 3 - Board: 22 / 427**

## Persistence for Analysis Objects

**Author:** J. Hrivnac<sup>1</sup><sup>1</sup> LAL**Corresponding Author:** julius.hrivnac@cern.ch

There are two kinds of analysis objects with respect to their persistent requirements:

\* Objects, which need direct access to the persistency service only for their IO operations (read/write/update/...): histograms, clouds, profiles, ...

All Persistency requirements for those objects can be implemented by standard Transient-Persistent Separation techniques like JDO, Serialization, etc.

\* Objects, which need direct access to the persistency service for some of their standard operations: NTuples, Tags,.... It is not feasible to completely separate Transient and Persistent form of those objects. Their Persistency should be tightly interfaced with their transient form. One possibility is to directly implement a persistent extension of those objects for each persistency mechanism.

The SQLTuple has been developed to deliver efficient SQL persistency for AIDA standard NTuple objects. The implementation is based on FreeHEP AIDA implementation and is completely inter-operable with other FreeHEP components as well as with other AIDA implementations. SQLTuple dependency on SQL database implementation is handled at run-time by textual configuration. In principle all mainstream SQL databases are supported. The default mapping layer can be customized so that, for example, LCG Pool Tag databases can be transparently supported. This customization is used to implement

higher level management utilities for Pool Tag databases - package ColMan. ColMan utilities are accessible also from the C++ environment and via standard Web Service.

The representation will cover both SQLTuple and ColMan packages and their inter-operability with other tools. Performance assessment of various available technologies will be covered as well.

#### Distributed Computing Services / 430

### Global Distributed Parallel Analysis using PROOF and AliEn

**Authors:** A. Peters<sup>1</sup>; F. Carminati<sup>1</sup>; F. Rademakers<sup>1</sup>; M. Ballintijn<sup>2</sup>; P. BUNCIC<sup>1</sup>; P. Canal<sup>3</sup>; R. Brun<sup>1</sup>

<sup>1</sup> CERN

<sup>2</sup> MIT

<sup>3</sup> FNAL

**Corresponding Author:** federico.carminati@cern.ch

The ALICE experiment and the ROOT team have developed a Grid-enabled version of PROOF that allows efficient parallel processing of large and distributed data samples. This system has been integrated with the ALICE-developed AliEn middleware. Parallelism is implemented at the level of each local cluster for efficient processing and at the Grid level, for optimal workload management of distributed resources. This system allows harnessing large Computing on Demand capacity during an interactive session. Remote parallel computations are spawned close to the data, minimising network traffic. If several copies of the data are available, a workload management system decides automatically where to send the task. Results are automatically merged and displayed at the user workstation. The talk will describe the different components of the system (PROOF, the parallel ROOT engine, and the AliEn middleware), the present status and future plans for the development and deployment and the consequences for the ALICE computing model.

#### Poster Session 2 - Board: 17 / 431

### Usage of ALICE Grid middleware for medical applications

**Authors:** E. Lopez Torres<sup>1</sup>; F. Carminati<sup>2</sup>; F. Rademakers<sup>2</sup>; P. BUNCIC<sup>2</sup>; P. Cerello<sup>3</sup>; P. Hristov<sup>2</sup>; S. Bagnasco<sup>3</sup>; S. Cheran<sup>3</sup>

<sup>1</sup> CEADEN - Cuba

<sup>2</sup> CERN

<sup>3</sup> INFN Torino

**Corresponding Authors:** cerello@to.infn.it, federico.carminati@cern.ch

Breast cancer screening programs require managing and accessing a huge amount of data, intrinsically distributed, as they are collected in different Hospitals. The development of an application based on Computer Assisted Detection algorithms for the analysis of digitised

mammograms in a distributed environment is a typical GRID use case. In particular, AliEn (ALICE Environment) services, whose development was carried on by the ALICE Collaboration, were used to configure a dedicated Virtual Organisation; a PERL-based interface to AliEn commands allows the registration of new patients and mammograms in the AliEn Data Catalogue as well as queries to retrieve images associated to selected patients. The analysis of selected mammograms can be performed interactively, making use of PROOF services, or taking advantage of the AliEn capabilities to generate “sub-jobs”; each of them analyzes the fraction of the selected sample stored on a site, and the results are merged. All the required functionality is available: by the end of 2004 a working prototype is foreseen, with an AliEn Client installed in each of the Hospitals participating to the INFN-funded MAGIC-5 project.

The same approach will be applied in the near future in two other application areas:

- Lung cancer screening, equivalent to the mammographic screening from the middleware point of view, where Computer Assisted Detection algorithms are being developed;
- Diagnosis of the Alzheimer disease, where the application is intrinsically distributed: it should, in fact, compare the PET-generated image to a set of reference images which are scattered on many sites and merge the results.

**Online Computing / 432**

## **Fault Tolerance and Fault Adaption for High Performance Large Scale Embedded Systems**

**Author:** P. Sheldon<sup>1</sup>

<sup>1</sup> VANDERBILT UNIVERSITY

**Corresponding Author:** paul.sheldon@vanderbilt.edu

The BTeV experiment, a proton/antiproton collider experiment at the Fermi National Accelerator Laboratory, will have a trigger that will perform complex computations (to reconstruct vertices, for example) on every collision (as opposed to the more traditional approach of employing a first level hardware based trigger). This trigger requires large-scale fault adaptive embedded software: with thousands of processors involved in performing event filtering in the trigger farm fault conditions must be given proper treatment. Without fault mitigation, it is conceivable that the trigger system will experience failures at a high enough rate to have an unacceptable negative impact on BTeV's physics goals. The RTES (Real Time Embedded Systems) collaboration is a group of physicists, engineers, and computer scientists working to address the problem of reliability in large-scale clusters with real-time constraints such as this. Resulting infrastructure must be highly scalable, verifiable, extensible by users, and dynamically changeable. An initial prototype has been built to test design ideas and methods for the final system, and a larger scale and more ambitious prototype is currently under construction. I will discuss the lessons learned from these prototypes as well as the overall design and deliverables for the BTeV experiment.

**Event Processing / 433**

## **The Virtual MonteCarlo : status and applications**

**Authors:** A. Fasso<sup>1</sup>; A. Gheata<sup>2</sup>; A. Morsch<sup>2</sup>; E. Futo<sup>2</sup>; F. Carminati<sup>2</sup>; I. Hrivnacova<sup>3</sup>; M. Gheata<sup>4</sup>; P. Hristov<sup>2</sup>; R. Brun<sup>2</sup>

<sup>1</sup> SLAC

<sup>2</sup> CERN

<sup>3</sup> IPN Orsay

<sup>4</sup> Institute of Space Science - Bucharest

**Corresponding Authors:** federico.carminati@cern.ch, andreas.morsch@cern.ch

The current major detector simulation programs, i.e. GEANT3, GEANT4 and FLUKA have largely incompatible environments. This forces the physicists willing to make comparisons between the different transport Monte Carlos to develop entirely different programs. Moreover, migration from one program to the other is usually very expensive, in manpower and time, for an experiment offline environment, as it implies substantial changes in the simulation code. To solve this problem, the ALICE Offline project has developed a virtual interface to these three programs allowing their seamless use without any change in the framework, the geometry description or the scoring code. Moreover a new geometrical modeller has been developed in collaboration with the ROOT team, and successfully interfaced to the three programs. This allows the use of one description of the geometry, which can be used also during reconstruction and visualisation. The talk will describe the present status and future plans for the Virtual Monte Carlo. It will also present the capabilities and performance of the geometrical modeller.

**Online Computing / 434**

## The ALICE High Level Trigger

**Authors:** A. Vestbo<sup>1</sup>; A. Wiebalck<sup>2</sup>; B. Skaali<sup>3</sup>; C. Loizides<sup>4</sup>; D. Rohrich<sup>1</sup>; H. Helstrup<sup>5</sup>; H. Tilsner<sup>2</sup>; K. Ullaland<sup>1</sup>; M. Richter<sup>1</sup>; R. Stock<sup>4</sup>; T. Alt<sup>2</sup>; T. Vik<sup>3</sup>; T.M. Steinbeck<sup>2</sup>; V. Lindenstruth<sup>2</sup>

<sup>1</sup> Department of Physics and Technology, University of Bergen, Norway

<sup>2</sup> Kirchhoff Institute for Physics, University of Heidelberg, Germany

<sup>3</sup> Department of Physics, University of Oslo, Norway

<sup>4</sup> Institute for Nuclear Physics, University of Frankfurt, Germany

<sup>5</sup> Faculty of Engineering, Bergen University College, Norway

**Corresponding Author:** matthias.richter@ift.uib.no

The ALICE experiment at LHC will implement a High Level Trigger System, where the information from all major detectors are combined, including the TPC, TRD, DIMUON, ITS etc. The largest computing challenge is imposed by the TPC, requiring realtime pattern recognition. The main task is to reconstruct the tracks in the TPC, and in a final stage combine the tracking information from all detectors. Based on the physics observables selective readout is done by generation of a software trigger (High Level Trigger), capable of selecting interesting (sub)events from the input data stream. Depending on the physics program various processing options are currently being developed, including region of interest processing, rejecting events based on software trigger and data compression schemes. Examples of such triggers are verification of candidates for high-pt dielectron heavy-quarkonium decays, momentum filter to enhance the open-charm signal, high-pt jets selection etc.



Technically the HLT system entails a very large scale processing farm with about 1000 active processors. The input data stream is designed for 25 GB/sec. The system nodes will be interfaced to the local data concentrators of the DAQ system via optical fibers receiving a copy of the raw data.

The optical fibers will be connected to the PCI-bus of HLT nodes using a custom PCI card. These cards provide a co-processor functionality for the first steps of the pattern recognition.

The talk will give an overview of the HLT project and will focus on the latest results regarding efficient data compression and trigger performance.

## Event Processing / 435

### Track reconstruction in high density environment

**Authors:** I. Belikov<sup>1</sup>; K. Safarik<sup>1</sup>; M. Ivanov<sup>1</sup>; P. Hristov<sup>1</sup>; T. Kuhr<sup>1</sup>

<sup>1</sup> CERN

**Corresponding Author:** marian.ivanov@cern.ch

Tracks finding and fitting algorithm in ALICE Time projection chamber (TPC) and Inner Tracking System (ITS)

based on the Kalman-filtering are presented. The filtering algorithm is able to cope with non-Gaussian noise

and ambiguous measurements in high-density environments. The tracking algorithm consists of two parts:

one for the TPC and one for the prolongation into the ITS. The occupancy in the TPC can reach up to 40 %.

Usually, due to the overlaps, a number of points along the track are lost or significantly displaced. At first the clusters are found and the space points are reconstructed. The shape of a cluster provides information about the overlap. An unfolding algorithm is applied for points with distorted shapes. Then, the

expected space point error is estimated using information about the cluster shape and track parameters.

Further, the available information about local track overlap is used.

In the TPC-ITS matching, the distance between the TPC and the ITS sensitive volume is rather large and the

track density inside the ITS is so high that the straightforward continuation of the tracking procedure is

ineffective. Using only chi<sup>2</sup> minimisation there is a high probability of assigning a wrong hit to the track.

Therefore for each TPC track a candidate tree of the possible track prolongation in the ITS is build. Finally the

most probable track candidates are chosen.

The approach have been implemented within the ALICE simulation/reconstruction framework (ALI-ROOT), and

algorithm's efficiency have been estimated using the ALIROOT Monte Carlo data.

## Poster Session 2 - Board: 8 / 436

### The LCG-AliEn interface, a realization of a MetaGrid system

**Authors:** A. Peters<sup>1</sup>; P. Buncic<sup>1</sup>; P. Cerello<sup>2</sup>; P. Saiz<sup>1</sup>; S. Bagnasco<sup>2</sup>

<sup>1</sup> CERN

<sup>2</sup> INFN Torino

AliEn (ALICE Environment) is a GRID middleware developed and used in the context of ALICE, the CERN LHC heavy-ion experiment. In order to run Data Challenges exploiting both AliEn “native” resources and any infrastructure based on EDG-derived middleware (such as the LCG and the Italian GRID.IT), an interface system was designed and implemented; some details of a prototype were already presented at CHEP2003. In the spring of 2004 an ALICE Data Challenge began with the simulated data production on this multiple infrastructure, thus qualifying as the first large production carried out transparently making use of very different middleware system. This system is a practical realisation of the “federated” or “meta-” grid concept, and it has been successfully tested in a very large production. This talk reports about new developments of the interface system, the successful DC running experience, the advantages and limitations of this concept, the plans for the future and some lessons learned.

**Online Computing / 437**

## Simulations and Prototyping of the LHCb L1 and HLT Triggers

**Authors:** A. Barczyk<sup>1</sup>; A. SMOKER<sup>2</sup>; B. Jost<sup>1</sup>; J-P. Dufey<sup>1</sup>; J. ANDERSON<sup>2</sup>; N. Neufeld<sup>1</sup>; R. McNulty<sup>3</sup>; T. Kechadi<sup>4</sup>; T. Shears<sup>5</sup>

<sup>1</sup> CERN

<sup>2</sup> University College Dublin, Ireland

<sup>3</sup> UNIVERSITY COLLEGE DUBLIN

<sup>4</sup> University College Dublin

<sup>5</sup> University of Liverpool

**Corresponding Author:** [ronan.mcnulty@cern.ch](mailto:ronan.mcnulty@cern.ch)

The Level 1 and High Level triggers for the LHCb experiment are software triggers which will be implemented on a farm of about 1800 CPUs, connected to the detector read-out system by a large Gigabit Ethernet LAN with a capacity of 8 Gigabyte/s and some 500 Gigabit Ethernet links. The architecture of the readout network must be designed to maximise data throughput, control data flow, allow load balancing between the nodes and be proven to perform at scale. Issues of stability, robustness and fault tolerance are vital to the effective operation of the trigger. We report on the development and results of two independent software simulations which allow us to evaluate the performance of various network configurations and to specify the switch parameters. In order to validate the results of the simulation and to experimentally test the performance of the readout network in conditions similar to those expected at the LHC, we have constructed a hardware prototype of the LHCb Level 1 and High Level triggers. This prototype allows a scaled evaluation of our design, soak-testing, and an evaluation of the overall system response to the deliberate introduction of faults. The performance of this test-bed is described and the results compared to simulation.

Poster Session 2 - Board: 54 / 438

## Participation of Russian sites in the Data Challenge of ALICE experiment in 2004

**Authors:** A. Selivanov<sup>1</sup>; A. Zarochencev<sup>2</sup>; E. Ryabinkin<sup>3</sup>; G. Shabratova<sup>4</sup>; V. Dobretsov<sup>3</sup>; V. Kolosov<sup>5</sup>; V. Korenkov<sup>4</sup>; V. Mitcin<sup>4</sup>; V. Zolotarev<sup>2</sup>; Y. Bugaenko<sup>4</sup>; Y. Kharlov<sup>6</sup>; Y. Lyublev<sup>7</sup>

<sup>1</sup> ITEP, Moscow

<sup>2</sup> SPbSU

<sup>3</sup> KIAE, Moscow

<sup>4</sup> Joint Institute for Nuclear Research (JINR)

<sup>5</sup> ITEP

<sup>6</sup> State Res.Center of Russian Feder. HEP Inst (IFVE)

<sup>7</sup> Institute for Theoretical and Experimental Physics (ITEP)

The report presents an analysis of the Alice Data Challenge 2004. This Data Challenge has been performed on two different distributed computing environments. The first one is the Alice Environment for distributed computing (AliEn) used standalone. Presently this environment allows ALICE physicists to obtain results on simulation, reconstruction and analysis of data in ESD format for AA and pp collisions at LHC energies. The second environment is the LCG-2 middleware accessed via AliEn with the help of an interface, developed at INFN. Three Russian sites have been configured as AliEn nodes for the Data Challenge. These sites (IHEP at Protvino, ITEP in Moscow and JINR at Dubna) could run a maximal of 86 jobs. The initial analysis shows that the architecture of one site was not adequate for distributed computing. Another farm had nodes with insufficient RAM for efficient job processing. All these problems have been cured subsequent DC phases.

Actions have also been taken to reduce the downtime due to wrong site configuration. The local AliEn server installed at the JINR site has been used as a standard configuration for the other Russian sites. The total number of jobs processed in Russia constitute ~2% of total run in the ALICE DC 2004.

Grid Security / 439

## Building Global HEP Systems on Kerberos

**Author:** M. Crawford<sup>1</sup>

<sup>1</sup> FERMILAB

**Corresponding Author:** matt.crawford@cern.ch

As an underpinning of AFS and Windows 2000, and as a formally proven security protocol in its own right, Kerberos is ubiquitous among HEP sites. Fermilab and users from other sites have taken advantage of this and built a diversity of distributed applications over Kerberos v5. We present several projects in which this security infrastructure has been leveraged to meet the requirements of far-flung collaborations. These range from straightforward “Kerberization” of applications such as database and batch services, to quick tricks like simulating a user-authenticated web service with AFS and the “file:” schema, to more

complex systems. Examples of the latter include experiment control room operations and the Central Analysis Farm (CAF).

We present several use cases and their security models, and examine how they attempt to address some of the outstanding problems of secure distributed computing: delegation of the least necessary privilege; establishment of trust between a user and a remote processing facility; credentials for long-queued or long-running processes, and automated processes running without any user's instigation; security of remotely-stored credentials; and ability to scale to the numbers of sites, machines and users expected in the collaborations of the coming decade.

## Distributed Computing Systems and Experiences / 440

### The architecture of the AliEn system

**Authors:** A. Peters<sup>1</sup>; J.F. Grosse-Oetringhaus<sup>1</sup>; P. Buncic<sup>1</sup>; P. Saiz<sup>1</sup>

<sup>1</sup> CERN

**Corresponding Author:** predrag.buncic@cern.ch

AliEn (ALICE Environment) is a Grid framework developed by the Alice Collaboration and used in production for almost 3 years. From the beginning, the system was constructed using Web Services and standard network protocols and Open Source components. The main thrust of the development was on the design and implementation of an open and modular architecture. A large part of the component came from state-of-the-art modules available in the Open Source domain. Thus, in a very short time, the ALICE experiment had a prototype Grid that, while constantly evolving, has allowed large distributed simulation and reconstruction vital to the design of the experiment hardware and software to be performed with very limited manpower. This proved to be the correct path to which many Grid project and initiatives are now converging. The architecture of AliEn inspired the ARDA report and subsequently AliEn provided the foundation of components for the first EGEE prototype. This talk presents the architecture of the original AliEn system, describes its evolution. A critical review of the major technology choices, their implementation and the development process is also presented.

## Poster Session 2 - Board: 28 / 441

### D0 data processing within EDG/LCG

**Authors:** D. Groep<sup>1</sup>; J. Templon<sup>1</sup>; K. Bos<sup>1</sup>; P. Maettig<sup>2</sup>; R. Byrom<sup>3</sup>; S. Fisher<sup>3</sup>; T. Harenberg<sup>4</sup>; W. Leeuwen<sup>1</sup>

<sup>1</sup> NIKHEF

<sup>2</sup> University of Wuppertal

<sup>3</sup> Rutherford Appleton Laboratory

<sup>4</sup> UNIVERSITY OF WUPPERTAL

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

The D0 experiment at the Tevatron is collecting some 100 Terabytes of data each year and has a very high need of computing resources for the various parts of the physics program. D0 meets these demands by establishing a world - increasingly based on GRID technologies.

Distributed resources are used for D0 MC production and data reprocessing of 1 billion events, requiring 250 TB to be transported over WANs. While in 2003 most of this computing at remote sites was distributed manually, some data reprocessing was performed with the EDG. In 2004 GRID tools are increasingly and successfully employed.

We will report on performing MC production and data reprocessing using EDG and LCG. We will explain how the D0 computing environment was linked to these GRID platforms, and will discuss some lessons learned (for both Grid computing and preparing applications for distributed operation) from the D0 reprocessing on EDG, subjecting a generic Grid infrastructure to real data for the first time.

An outlook on plans for applying LCG within D0 is given.

## Distributed Computing Systems and Experiences / 443

### AliEn Web Portal

**Authors:** A. Peters<sup>1</sup>; L. Betev<sup>1</sup>; M. Meoni<sup>2</sup>; P. E. Tissot-Daguette<sup>1</sup>; P. Buncic<sup>1</sup>; P. Saiz<sup>1</sup>; V. Pinto Morais<sup>1</sup>

<sup>1</sup> CERN

<sup>2</sup> CERN and INFN Firenze

**Corresponding Author:** predrag.buncic@cern.ch

The AliEn system, an implementation of the Grid paradigm developed by the ALICE Offline Project, is currently being used to produce and analyse Monte Carlo data at over 30 sites on four continents. The AliEn Web Portal is built around Open Source components with a backend based on Grid Services and compliant with the OGSA model. An easy and intuitive presentation layer gives the opportunity to the user to access information from multiple sources in a transparent and convenient way. Users can browse job provenance and access monitoring information from MonaLisa repository.

The presentation layer is separated from the content layer which is implemented via Grid and Web Services serving one or more users or Virtual Organizations.

Security and authentication of the portal are based on the Globus Grid Security infrastructure, OGSI::Lite and MyProxy online credentials repository.

In this presentation the architecture and functionality of AliEn Portal implementation will be presented.

## Event Processing / 444

### The High Level Trigger software for the CMS experiment

**Authors:** C. Delaere<sup>1</sup>; O. van der Aa<sup>1</sup>

<sup>1</sup> *INSTITUT DE PHYSIQUE NUCLEAIRE, UNIVERSITE CATHOLIQUE DE LOUVAIN*

**Corresponding Author:** olivier.van.der.aa@cern.ch

The observation of Higgs bosons predicted in supersymmetric theories will be a challenging task for the CMS experiment at the LHC, in particular for its High Level trigger (HLT). A prototype of the High Level Trigger software to be used in the filter farm of the CMS experiment and for the filtering of monte carlo samples will be presented. The implemented prototype heavily uses recursive processing of a HLT tree and allows dynamic trigger definition. Firstly the general architecture and design choices as well as the timing performance of the system will be reviewed in the light of the DAQ constraints. Secondly, specific trigger implementations in the context of the object-oriented Reconstruction for CMS Analysis (ORCA) software will be detailed. Finally, the analysis for the selection of a CP even Higgs decaying in tau pairs will be presented. The Aforementioned analysis will illustrate the importance of the trigger strategies required to achieve the various physics analysis in CMS.

**Event Processing / 445**

## **Bayesian Approach for Combined Particle Identification in ALICE Experiment at LHC**

**Authors:** I. Belikov<sup>1</sup>; K. Safarik<sup>1</sup>; M. Ivanov<sup>1</sup>; P. Hristov<sup>1</sup>; T. Kuhr<sup>1</sup>

<sup>1</sup> *CERN*

**Corresponding Author:** jouri.belikov@cern.ch

One of the main features of the ALICE detector at LHC is the capability to identify particles in a very broad momentum range from 0.1 GeV/c up to 10 GeV/c. This can be achieved only by combining, within a common setup, several detecting systems that are efficient in some narrower and complementary momentum sub-ranges. The situation is further complicated by the amount of data to be processed (about  $10^7$  events with about  $10^4$  tracks in each). Thus, the particle identification (PID) procedure should satisfy the following requirements:

- 1) It should be as much as possible automatic.
- 2) It should be able to combine PID signals of different nature (e.g.  $dE/dx$  and TOF measurements).
- 3) When several detectors contribute to the PID, the procedure must profit from this situation by providing an improved PID.
- 4) When only some detectors identify a particle, the signals from the other detectors must not affect the combined PID.
- 5) It should take into account the fact that the PID depends, due to different track selection, on the kind of analysis.

In this report we will demonstrate how combining the single detector PID signals in the Bayesian way satisfies these requirements. We will also discuss how one can obtain the needed probability distribution functions and a priory probability from the experimental data. The approach has been implemented within the ALICE offline framework, and the algorithm efficiency and PID contamination have been estimated

using the  
ALICE simulation.

**Wide Area Networking / 446**

## ALICE Multi-site Data Transfer Tests on a Wide Area Network

**Authors:** A. Fritz<sup>1</sup>; D. DiBari<sup>2</sup>; D. Mura<sup>3</sup>; E. Fragiaco<sup>4</sup>; F. Carminati<sup>5</sup>; F. Minafra<sup>6</sup>; G. Donvito<sup>6</sup>; G. Lo Re<sup>7</sup>; J. Svec<sup>8</sup>; M. Maserà<sup>9</sup>; M. Sitta<sup>10</sup>; P. Cerello<sup>10</sup>; R. Barbera<sup>11</sup>; R. Turrisi<sup>12</sup>; S. Bagnasco<sup>10</sup>

<sup>1</sup> *University of Houston*

<sup>2</sup> *Bari University*

<sup>3</sup> *INFN Cagliari*

<sup>4</sup> *INFN Trieste*

<sup>5</sup> *CERN*

<sup>6</sup> *INFN Bari*

<sup>7</sup> *INFN & CNAF Bologna*

<sup>8</sup> *ASCR Prague*

<sup>9</sup> *Torino University*

<sup>10</sup> *INFN Torino*

<sup>11</sup> *Catania University and INFN*

<sup>12</sup> *INFN Padova*

**Corresponding Authors:** giuseppe.lore@cnafe.infn.it, federico.carminati@cern.ch

Next generation high energy physics experiments planned at the CERN Large Hadron Collider is so demanding in terms of both computing power and mass storage that data and CPU's can not be concentrated in a single site and will be distributed on a computational Grid according to a "multi-tier".

LHC experiments are made of several thousands of people from a few hundreds of institutes spread out all over the world. These people, according to their collaborations on specific physics analysis topics, can constitute highly dynamic Virtual Organizations rapidly changing as a function of both time and topology. The impact of future experiments on Wide Area Networks (WAN) will be non negligible especially for what concerns the capillarity of bandwidths (down to the "last mile"), quality of service, adaptivity and configurability.

In this paper we report on a series of multi-site data transfer tests performed within the ALICE Experiment on a wide area network test-bed in order to spot possible bottlenecks and pin down critical elements and parameters of actual research networks.

In order to make the tests as realistic as possible, reflecting the real use cases foreseen in the next future, we have taken into account all the aspects of the elements involved in the transfer of a file:

- Local disk Input/Output (I/O) performance;
- I/O block size;
- TCP parameters and number of parallel streams;
- Bandwidth Delay Product (BDP) expressed as the product of the Bandwidth (BW) times the Round Trip Time (RTT).

Core Software / 447

## LCG Conditions Database Project Overview

**Authors:** A. Amorim<sup>1</sup>; A. Valassi<sup>2</sup>; D. Duellmann<sup>2</sup>; D. Klose<sup>3</sup>; L. Pedro<sup>1</sup>; N. Barros<sup>1</sup>; S.A. Schmidt<sup>4</sup>; T. Franco<sup>1</sup>; V. Tsulaia<sup>5</sup>

<sup>1</sup> *University of Lisboa*

<sup>2</sup> *CERN*

<sup>3</sup> *Universidade de Lisboa, Portugal*

<sup>4</sup> *University of Mainz*

<sup>5</sup> *University of Pittsburg*

**Corresponding Author:** andrea.valassi@cern.ch

The Conditions Database project has been launched to implement a common persistency solution for experiment conditions data in the context of the LHC Computing Grid (LCG) Persistency Framework. Conditions data, such as calibration, alignment or slow control data, are non-event experiment data characterized by the fact that they vary in time and may have different versions. The LCG project draws on preexisting projects which have led to the definition of a generic C++ API for condition data access and its implementation using different storage technologies, such as Objectivity, MySQL or Oracle. The project is assigned the task to deliver a production release of the software including implementation libraries for several technologies and high level tools for data management. The presentation will review the current status of the LCG common project at the time of the conference and the plans for its evolution.

Poster Session 3 - Board: 37 / 448

## Harp data and software migration from Objectivity to Oracle

**Authors:** A. Valassi<sup>1</sup>; D Geppert<sup>1</sup>; D Kolev<sup>2</sup>; E. Tcherniaev<sup>1</sup>; K. Nienartowicz<sup>1</sup>; M Lubeck<sup>1</sup>; M Nowak<sup>1</sup>

<sup>1</sup> *CERN*

<sup>2</sup> *University of Sofia*

**Corresponding Author:** andrea.valassi@cern.ch

The migration of the Harp data and software from an Objectivity-based to an Oracle-based data storage solution is reviewed in this presentation. The project, which was successfully completed in January 2004, involved three distinct phases. In the first phase, which profited significantly from the previous COMPASS data migration project, 30 TB of Harp raw event data were migrated in two weeks to a hybrid persistency solution, storing raw event records in standard "flat" files and the corresponding metadata in Oracle as relational tables. In the second phase, the longest to achieve in spite of the relatively limited data volume to migrate, the complex data model of Harp event collections was reimplemented for the Oracle-based solution. The relational schema design and the implementation of read-only navigational access to event collections in the Harp software framework using Oracle are reviewed in detail in the presentation.



The third phase was the easiest, as it involved the migration of conditions data (time-varying non-event data) from the Objectivity to the Oracle implementation of a same C++ API, which acted as a screening layer between the data model and its implementation.

**Online Computing / 449**

## **DZERO Data Acquisition Monitoring and History Gathering**

**Author:** G. Watts<sup>1</sup>

**Co-author:** Andy Haas<sup>2</sup>

<sup>1</sup> UNIVERSITY OF WASHINGTON

<sup>2</sup> University of Washington

**Corresponding Author:** gwatts@uw.edu

The DZERO Collider Experiment logs many of its Data Acquisition Monitoring Information in long term storage. This information is most frequently used to understand shift history and efficiency. Approximately two kilobytes of information is stored every 15 second. We describe this system and the web interface provided. The current system is distributed, running on Linux for the back end and Windows for the web interface front end and data logging. We also discuss the development path we have taken for the database backend, from use of root, to Oracle, and back to root.

**Wide Area Networking / 450**

## **Bringing High-Performance Networking to HEP Users**

**Authors:** N. Pezzi<sup>1</sup>; R. Hughes-Jones<sup>2</sup>; S. Dallison<sup>2</sup>; Y-T. Li<sup>1</sup>

<sup>1</sup> University College London

<sup>2</sup> THE UNIVERSITY OF MANCHESTER

**Corresponding Author:** steved@hep.man.ac.uk

How do we get High Throughput data transport to real users? The MB-NG project is a major collaboration which brings together expertise from users, industry, equipment providers and leading edge e-science application developers. Major successes in the areas of Quality of Service (QoS) and managed bandwidth have provided a leading edge U.K. Diffserv enabled network running at 2.5 Gbit/s. One of the central aims of MB-NG is the investigation of high performance data transport mechanisms for Grid data transfer across heterogeneous networks.

New transport stacks implement sender side modifications to the TCP algorithm which enable increased bandwidth utilisation in long-delay high-bandwidth environments. This allows a single stream of a modified TCP stack to transmit at rates that would otherwise require multiple streams of standard RENO TCP. This paper reports on investigations of the performance of these TCP stacks and their use with data transfer applications such as GridFTP, BBFTP, BSCP and APACHE. End-host performance behaviour was also examined in order to determine effects of the Network Interface, PCI bus performance, and disk and RAID sub-systems.

In a Collaboration between the BaBar experiment and MB-NG we demonstrated high performance data transport using these new TCP/IP transport protocol stacks and QoS

provisioning. We report on the benefits of this introduction of high speed networks and advanced TCP stacks together with various levels of QoS to the BaBar computing environment. The benefits achieved are contrasted with network behaviour and application performance using today's "production" network.

Poster Session 2 - Board: 38 / 451

## SAMGrid Monitoring Service and its Integration with MonALisa

**Authors:** A. Baranovski<sup>1</sup>; A. Kreymer<sup>1</sup>; A. Lyon<sup>1</sup>; A. Sill<sup>2</sup>; F. Ratnikov<sup>3</sup>; G. Garzoglio<sup>1</sup>; I. Terekhov<sup>1</sup>; J. Trumbo<sup>1</sup>; L. Loebel Carpenter<sup>4</sup>; L. Lueking<sup>1</sup>; M. Burgon-Lyon<sup>5</sup>; M. Leslie<sup>6</sup>; R. Herber<sup>1</sup>; R. Illingworth<sup>1</sup>; R. Kennedy<sup>1</sup>; R. St.Denis<sup>5</sup>; S. Belforte<sup>7</sup>; S. Stonjek<sup>8</sup>; S. Veseli<sup>1</sup>; S. White<sup>1</sup>; U. Kerzel<sup>9</sup>; V. Bartsch<sup>6</sup>; W. Merritt<sup>1</sup>

<sup>1</sup> *FERMI NATIONAL ACCELERATOR LABORATORY*

<sup>2</sup> *Texas Tech University*

<sup>3</sup> *Rutgers University*

<sup>4</sup> *Fermilab*

<sup>5</sup> *Glasgow University*

<sup>6</sup> *Oxford University*

<sup>7</sup> *INFN/Trieste*

<sup>8</sup> *FERMI NATIONAL ACCELERATOR LABORATORY/Oxford University*

<sup>9</sup> *Karlsruhe University*

**Corresponding Author:** lyon@fnal.gov

The SAMGrid team is in the process of implementing a monitoring and information service, which fulfills several important roles in the operation of the SAMGrid system, and will replace the first generation of monitoring tools in the current deployments. The first generation tools are in general based on text logfiles and represent solutions which are not scalable or maintainable. The roles of the monitoring and information service are: 1) providing diagnostics for troubleshooting the operation of SAMGrid services; 2) providing support for monitoring at the level of user jobs; 3) providing runtime support for local configuration and other information currently which currently must be stored centrally (thus moving the system toward greater autonomy for the SAM station services, which include cache management and job management services); 4) providing intelligent collection of statistics in order to enable performance monitoring and tuning. The architecture of this service is quite flexible, permitting input from any instrumented SAM application or service. It will allow multiple backend storage for archiving of (possibly) filtered monitoring events, as well as real time information displays and active notification service for alarm conditions. This service will be able to export, in a configurable manner, information to higher level Grid monitoring services, such as MonALisa. We describe our experience to date with using a prototype version together with MonALisa.

Distributed Computing Systems and Experiences / 455

## Application of the SAMGrid Test-Harness for Performance Evaluation and Tuning of a Distributed Cluster Implementation of Data Handling Services

**Authors:** A. Baranovski<sup>1</sup>; A. Kreymer<sup>1</sup>; A. Lyon<sup>1</sup>; A. Sill<sup>2</sup>; F. Ratnikov<sup>3</sup>; G. Garzoglio<sup>1</sup>; I. Terekhov<sup>1</sup>; J. Trumbo<sup>1</sup>; L. Loebel Carpenter<sup>4</sup>; L. Lueking<sup>1</sup>; M. Burgon-Lyon<sup>5</sup>; M. Leslie<sup>6</sup>; R. Herber<sup>1</sup>; R. Illingworth<sup>1</sup>; R. Kennedy<sup>1</sup>; R. St.Denis<sup>5</sup>; S. Belforte<sup>7</sup>; S. Stonjek<sup>8</sup>; S. White<sup>1</sup>; U. Kerzel<sup>9</sup>; V. Bartsch<sup>6</sup>; W. Merritt<sup>1</sup>

<sup>1</sup> *FERMI NATIONAL ACCELERATOR LABORATORY*

<sup>2</sup> *Texas Tech University*

<sup>3</sup> *Rutgers University*

<sup>4</sup> *Fermilab*

<sup>5</sup> *Glasgow University*

<sup>6</sup> *Oxford University*

<sup>7</sup> *INFN/Trieste*

<sup>8</sup> *FERMI NATIONAL ACCELERATOR LABORATORY/Oxford University*

<sup>9</sup> *Karlsruhe University*

**Corresponding Author:** lyon@fnal.gov

The SAMGrid team has recently refactored its test harness suite for greater flexibility and easier configuration. This makes possible more interesting applications of the test harness, for component tests, integration tests, and stress tests. We report on the architecture of the test harness and its recent application to stress tests of a new analysis cluster at Fermilab, to explore the extremes of analysis use cases and the relevant parameters for tuning in the SAMGrid station services. This reimplementaion of the test harness is a python framework which usesXML for configuration and small plug-in python modules for specific test purposes. One current testing application is running on a 128-CPU analysis cluster with access to 6 TB distributed cache and also to a 2 TB centralized cache, permitting studies of different cache strategies. We have studied the service parameters which affect the performance of retrieving data from tape storage as well. The use cases studied vary from those which will require rapid file delivery with short processing time per file, to the opposite extreme of long processing time per file. We also show how the same harness can be used to run regular unit tests on a production system to aid early fault detection and diagnosis. These results are interesting for their implications with regard to Grid operations, and illustrate the type of monitoring and test facilities required to accomplish such performance tuning.

**Poster Session 2 - Board: 21 / 456**

## **K5 @ INFN.IT: an infrastructure for the INFN cross REALM & AFS cell authentication.**

**Author:** E.M.V. Fasanelli<sup>1</sup>

<sup>1</sup> *INFN.*

**Corresponding Author:** enrico.m.v.fasanelli@le.infn.it

The infn.it AFS cell has been providing a useful single file-space and authentication mechanism for the whole INFN, but the lack of a distributed management system, has lead several INFN sections and LABs to setup local AFS cells. The hierarchical transitive cross-realm authentication introduced in the Kerberos 5 protocol and the new versions of the OpenAFS and MIT implementation of Kerberos 5, make possible to setup an

AFS cross cell authentication in a transparent way, using the Kerberos 5 cross-realm one. The goal of the K5 @ INFN.IT project is to provide a Kerberos 5 authentication infrastructure for the INFN and cross-realm authentication to be used for the cross cell AFS authentication. In this work we describe the scenario, the results of various tests performed, the solution chosen and the status of the K5 @ INFN.IT project.

**Computer Fabrics / 458**

## **Gfarm v2: A Grid file system that supports high-performance distributed and parallel data computing**

**Authors:** N. Soda<sup>1</sup>; O. Tatebe<sup>2</sup>; S. Matsuoka<sup>3</sup>; S. Sekiguchi<sup>4</sup>; Y. Morita<sup>5</sup>

<sup>1</sup> *SRA*

<sup>2</sup> *GRID TECHNOLOGY RESEARCH CENTER, AIST*

<sup>3</sup> *Tokyo Institute of Technology/National Institute of Informatics*

<sup>4</sup> *Grid Technology Research Center, AIST*

<sup>5</sup> *KEK*

**Corresponding Author:** o.tatebe@aist.go.jp

Gfarm v2 is designed for facilitating reliable file sharing and high-performance distributed and parallel data computing in a Grid across administrative domains by providing a Grid file system. A Grid file system is a virtual file system that federates multiple file systems. It is possible to share files or data by mounting the virtual file system. This paper discusses the design and implementation of secure, robust, scalable and high-performance Grid file system.

The most time-consuming, but also the most typical, task in data computing such as high energy physics, astronomy, space exploration, human genome analysis, is to process a set of files in the same way. Such a process can be typically performed independently on every file in parallel, or at least have good locality. Gfarm v2 supports high-performance distributed and parallel computing for such a process by introducing a “Gfarm file”, a new “file-affinity” process scheduling based on file locations, and new parallel file access semantics. An arbitrary group of files possibly dispersed across administrative domains can be managed as a single Gfarm file. Each member file will be accessed in parallel in a new file view called “local file view” by a parallel process possibly allocated by file-affinity scheduling based on replica locations of the member files. File-affinity scheduling and new file view enable the owner “computes” strategy, or move the computation to data” approach for parallel and distributed data computing of member files of a Gfarm file in a single system image.

**Distributed Computing Services / 460**

## **SAMGrid Integration of SRMs**

**Authors:** A. Baranovski<sup>1</sup>; A. Kreymer<sup>2</sup>; A. Kumar<sup>3</sup>; A. Lyon<sup>2</sup>; A. Sill<sup>4</sup>; F. Ratnikov<sup>5</sup>; G. Garzoglio<sup>2</sup>; I. Terekhov<sup>3</sup>; J. Trumbo<sup>3</sup>; L. Loebel Carpenter<sup>3</sup>; L. Lueking<sup>2</sup>; M. Burgon-Lyon<sup>6</sup>; M. Leslie<sup>7</sup>; R. Illingworth<sup>2</sup>; R. Kennedy<sup>2</sup>; R. St.Denis<sup>6</sup>; S. Belforte<sup>8</sup>; S. Stonjek<sup>7</sup>; S. Veseli<sup>3</sup>; S. White<sup>3</sup>; U. Kerzel<sup>9</sup>; V. Bartsch<sup>7</sup>; W. Merritt<sup>2</sup>

<sup>1</sup> *Fermi National Accelerator Laboratory*

<sup>2</sup> *FERMI NATIONAL ACCELERATOR LABORATORY*

<sup>3</sup> *Fermilab*

<sup>4</sup> *Texas Tech. University*

<sup>5</sup> *Rutgers University*

<sup>6</sup> *Glasgow University*

<sup>7</sup> *Oxford University*

<sup>8</sup> *INFN, Trieste*

<sup>9</sup> *Karlsruhe University*

**Corresponding Author:** kennedy@fnal.gov

SAMGrid is the shared data handling framework of the two large Fermilab Run II collider experiments: DZero and CDF. In production since 1999 at D0, and since mid-2004 at CDF, the SAMGrid framework has been adapted over time to accommodate a variety of storage solutions and configurations, as well as the differing data processing models of these two experiments. This has been very successful for both experiments. Backed by primary data repositories of approximately 1 PB in size for each experiment, the SAMGrid framework delivers over 100 TB/day to DZero and CDF analyses at Fermilab and around the world. Each of the storage systems used with SAMGrid, however, has distinct interfaces, protocols, and behaviors. This led to different levels of integration of the various storage devices into the framework, which complicated the exploitation of their functionality and limited in some cases SAMGrid expansion across the experiments' Grid.

In an effort to simplify the SAMGrid storage interfaces, SAMGrid has

adopted the Storage Resource Manager (SRM) concept as the universal interface to all storage devices. This has simplified the SAMGrid framework, especially the implementation of storage device interactions. It prepares the SAMGrid framework for future storage solutions equipped with SRM interfaces, without the need for long and risky software integration projects. In principle, any storage device with an SRM interface can be used now with the SAMGrid framework. The integration of SRMs is an important further step towards evolving the SAMGrid framework into a co-operating collection of distinct, modular grid-oriented services. To date, SRMs for Enstore, dCache, local caches, and permanent disk locations are tested and in production use. This report outlines how the SRMs were integrated into the existing SAMGrid framework without disturbing on-going operations, and describes our operational experience with SAMGrid and SRMs in the field.

**Computer Fabrics / 461**

## StoRM: grid middleware for disk resource management

**Authors:** A. Ghiselli<sup>1</sup>; F. Donno<sup>2</sup>; L. Magnoni<sup>1</sup>; R. Zappi<sup>1</sup>

<sup>1</sup> *INFN-CNAF*

<sup>2</sup> *INFN*

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

Within a Grid the possibility of managing storage space is fundamental, in particular, before and during application execution. On the other hand, the increasing availability of highly performant computing resources raises the need for fast and efficient I/O operations and drives the development of parallel distributed file systems able to satisfy these needs granting access to distributed storage. The demand of POSIX compliant access to storage and the need to have a uniform interface for both Grid integrated and pure vanilla applications stimulate developers to investigate the possibility to integrate already existing filesystems into a Grid infrastructure, allowing users to take advantage of storage resources without being forced to change their applications.

This paper describes the design and implementation of StoRM, a storage resource manager (SRM) for disk only. Through StoRM an application can reserve and manage space on disk storage systems. It can then access the space either in a Grid environment or locally in a transparent way via classic POSIX calls.

The StoRM architecture is based on a pluggable model in order to easily add new functionalities. The StoRM implementation uses now filesystems such as GPFS or LUSTRE. The StoRM prototype includes space reservation functionalities that complement SRM space reservation to allow applications to directly access/use the managed space through POSIX calls. Moreover, StoRM includes quota management and a space guard. StoRM will serve as policy enforcement point (PEP) for the Grid Policy Management System over disk resources. The experimental results obtained are promising.

**Computer Fabrics / 462**

## The SAMGrid Database Server Component: Its Upgraded Infrastructure and Future Development Path

**Authors:** A. Baranovski<sup>1</sup>; A. Kreymer<sup>2</sup>; A. Lyon<sup>2</sup>; A. Sill<sup>3</sup>; F. Ratnikov<sup>4</sup>; G. Garzoglio<sup>2</sup>; I. Terekhov<sup>1</sup>; J. Trumbo<sup>1</sup>; L. Loebel Carpenter<sup>1</sup>; L. Lueking<sup>2</sup>; M. Burgon-Lyon<sup>5</sup>; M. Leslie<sup>6</sup>; R. Herber<sup>2</sup>; R. Illingworth<sup>2</sup>; R. Kennedy<sup>2</sup>; R. St.Denis<sup>5</sup>; S. Belforte<sup>7</sup>; S. Stonjek<sup>6</sup>; S. Veseli<sup>1</sup>; S. White<sup>1</sup>; U. Kerzel<sup>8</sup>; V. Bartsch<sup>6</sup>; W. Merritt<sup>2</sup>

<sup>1</sup> *Fermilab*

<sup>2</sup> *FERMI NATIONAL ACCELERATOR LABORATORY*

<sup>3</sup> *Texas Tech. University*

<sup>4</sup> *Rutgers University*

<sup>5</sup> *Glasgow University*

<sup>6</sup> *Oxford University*

<sup>7</sup> *INFN Trieste*

<sup>8</sup> *Karlsruhe University*

**Corresponding Author:** veseli@fnal.gov

The SAMGrid Database Server encapsulates several important services, such as accessing file metadata and replica catalog, keeping track of the processing information, as well as providing the runtime support for SAMGrid station services. Recent deployment of the SAMGrid system for CDF has resulted in unification of the database schema used by CDF and D0, and the complexity of changes required for the unified metadata catalog has warranted a complete redesign of the DB Server.

We describe here the architecture and features of the new server. In particular, we discuss the new CORBA infrastructure that utilizes python wrapper classes around IDL structs and exceptions. Such infrastructure allows us to use the same code on both server and client sides, which in turn results in significantly improved code maintainability and easier development.

We also discuss future integration of the new server with an SBIR II project which is directed toward allowing the dserver to access distributed

databases, implemented in different DB systems and possibly using different schema.

**Grid Security / 463**

## Mis-use Cases for the Grid

**Author:** D. Skow<sup>1</sup>

<sup>1</sup> *FERMILAB*

**Corresponding Author:** dane@fnal.gov

There have been a number of efforts to develop use cases for the Grid to guide development and useability testing. This talk examines the value of “mis-use cases” for guiding the development of operational controls and error handling. A couple of the more common current network attack patterns will be extrapolated to a global Grid environment. The talk will walk through the various activities necessary for incident response and recovery and strive to be technology neutral.

Gedanken incident response exercises are being discussed among the HEP PKI infrastructure specialists, but a systems-wide approach to the issues and necessary tools is needed. Determining scope of incidents, performing forensics and containing the spread requires a much more distributed approach than our previous experiences. A new set of tools and communication patterns are likely to be needed. This talk will be aimed at applications and middleware developers as well as operations teams for grids.

As time allows, the talk will survey current grid testbed middleware, identify the current control points and responsibilities and suggest places where extensions or modifications would be beneficial.

**Computer Fabrics / 464**

## The status of Fermilab Enstore Data Storage System

**Authors:** A. Moibenko<sup>1</sup>; C-H. Huang<sup>1</sup>; D. Petravick<sup>1</sup>; E. Berman<sup>1</sup>; J. Bakken<sup>1</sup>; M. Zalokar<sup>1</sup>

<sup>1</sup> *FERMI NATIONAL ACCELERATOR LABORATORY, USA*

**Corresponding Author:** moibenko@fnal.gov

Fermilab has developed and successively uses Enstore Data Storage System. It is a primary data store for the Run II Collider Experiments, as well as for the others. It provides data storage in robotic tape libraries according to requirements of the experiments. High fault tolerance and availability, as well as multilevel priority based request processing allows experiments to effectively store and access data stored in the Enstore, including storing raw data from data acquisition systems. The distributed structure and modularity of Enstore allow to scale the system and add more storage equipment as the requirements grow. Currently Fermilab Data Storage System storage system Enstore includes 5 robotic tape libraries, 96 tape drives of different type. Amount of data

stored in the system is ~1.7 PetaBytes. Users access Enstore directly using a special command. They also can use ftp, grid ftp, SRM interfaces to dCache system, that uses Enstore as its lower layer storage.

**Computer Fabrics / 465**

## **Network Information and Management Infrastructure Project**

**Authors:** D. LaMore<sup>1</sup>; D. Petravick<sup>1</sup>; D. Skow<sup>1</sup>; E. Berman<sup>1</sup>; I. Mandrichenko<sup>1</sup>; M. Crawford<sup>1</sup>; M. Grigoriev<sup>1</sup>; P. DeMar<sup>1</sup>; R. Reitz<sup>1</sup>; V. Podstavkov<sup>1</sup>

<sup>1</sup> *FNAL*

**Corresponding Author:** [ivm@fnal.gov](mailto:ivm@fnal.gov)

Management of large site network such as FNAL LAN presents many technical and organizational challenges. This highly dynamic network consists of around 10 thousand network nodes. The nature of the activities FNAL is involved in and its computing policy require that the network remains as open as reasonably possible both in terms of connectivity to the outside networks and in with respect to procedural simplicity of joining the network by temporary network participants such as visitors notebook computers. The goal of the Network Information and Management Infrastructure project at FNAL is to build software infrastructure which would help network management and computer security teams organize monitoring and management of the network, simplify communication between these entities and users, integrate network management into FNAL computer center management infrastructure.

Primary authors: Phil DeMar (FNAL), Igor Mandrichenko (FNAL), Don Petravick(FNAL), Dane Skow (FNAL)

**Poster Session 1 - Board: 20 / 466**

## **Online Monitoring and online calibration/reconstruction for the PHENIX experiment**

**Author:** C. Pinkenburg<sup>1</sup>

<sup>1</sup> *BROOKHAVEN NATIONAL LABORATORY*

**Corresponding Author:** [pinkenburg@bnl.gov](mailto:pinkenburg@bnl.gov)

The PHENIX experiment consists of many different detectors and detector types, each one with its own needs concerning the monitoring of the data quality and the calibration. To ease the task for the shift crew to monitor the performance and status of each subsystem in PHENIX we developed a general client server based framework which delivers events at a rate in excess of 100Hz.

This model was chosen to minimize the possibility of accidental interference with the monitoring tasks themselves. The user only interacts with the client which can be restarted any time without loss or alteration of information on the server side. It also enables multiple people to check simultaneously the same detector -



if need be even from remote locations. The information is transferred in form of histograms which are processed by the client. These histograms are saved for each run and some html output is generated which is used later on to remove problematic runs from the offline analysis. An additional interface to a data base is provide to enable the display of long term trends.

This framework was augmented to perform an immediate calibration pass and a quick reconstruction of rare signals in the counting house. This is achieved by filtering out interesting triggers and processing them on a local Linux cluster. That enabled PHENIX to e.g. keep track of the number of J/Psi's which could be expected while still taking data.

**Poster Session 1 - Board: 21 / 467**

## **Boosting the data logging rates in run 4 of the PHENIX experiment**

**Author:** C. Pinkenburg<sup>1</sup>

<sup>1</sup> *BROOKHAVEN NATIONAL LABORATORY*

**Corresponding Author:** pinkenburg@bnl.gov

With the improvements in CPU and disk speed over the past years, we were able to exceed the original design data logging rate of 40MB/s by a factor of 3 already for the Run 3 in 2002. For the Run 4 in 2003, we increased the raw disk logging capacity further to about 400MB/s.

Another major improvement was the implementation of compressed data logging. The PHENIX raw data, after application of the standard data reduction techniques, were found to be further compressible by utilities like *gzip* by almost a factor of 2, and we defined a PHENIX standard of a compressed raw data format. The buffers that make up a raw data file consist of buffers that would get compressed and the resulting smaller data volume written out to disk. For a long time, this proved to be much too slow to be usable in the DAQ, until we could shift the compression to the event builder machines and so distributed the load over many fast CPU's. We also selected a different compression algorithm, LZO, which is about a factor of 4 faster than the "compress2" algorithm used internally in *gzip*. With the compression, the raw data volume shrinks to about 60% of the original size, boosting the original data rate before compression to more than 700MB/s.

We will the present the techniques and architecture, and the impact this has had on the data taking in Run 4.

**Poster Session 2 - Board: 14 / 468**

## **Deployment of SAM for the CDF Experiment**

**Authors:** A. Baranovski<sup>1</sup>; A. Kreymer<sup>1</sup>; A. Lyon<sup>1</sup>; A. Sill<sup>2</sup>; F. Ratnikov<sup>3</sup>; G. Garzoglio<sup>1</sup>; I. Terekhov<sup>1</sup>; J. Trumbo<sup>1</sup>; L. Loebel Carpenter<sup>4</sup>; L. Lueking<sup>1</sup>; M. Burgon-Lyon<sup>5</sup>; M. Leslie<sup>6</sup>; R. Herber<sup>7</sup>; R. Illingworth<sup>7</sup>; R. Kennedy<sup>1</sup>; R. St.Denis<sup>5</sup>; S. Belforte<sup>8</sup>; S. Stonjek<sup>9</sup>; S. Veseli<sup>1</sup>; S. White<sup>1</sup>; U. Kerzel<sup>10</sup>; V. Bartsch<sup>6</sup>; W. Merritt<sup>1</sup>

- <sup>1</sup> *Fermi National Accelerator Laboratory*  
<sup>2</sup> *Texas Tech University*  
<sup>3</sup> *Rutgers University*  
<sup>4</sup> *Fermilab*  
<sup>5</sup> *Glasgow University*  
<sup>6</sup> *Oxford University*  
<sup>7</sup> *FERMI NATIONAL ACCELERATOR LABORATORY*  
<sup>8</sup> *INFN / Trieste*  
<sup>9</sup> *Fermi National Accelerator Laboratory / University of Oxford*  
<sup>10</sup> *Karlsruhe University*

CDF is an experiment at the Tevatron at Fermilab. One dominating factor of the experiments' computing model is the high volume of raw, reconstructed and generated data. The distributed data handling services within SAM move these data to physics analysis applications. The SAM system was already in use at the D-Zero experiment. Due to difference in the computing model of the two experiments some aspects of the SAM system had to be adapted. We will present experiences from the adaptation and the deployment phase. This includes the behavior of the SAM system on batch systems of very different sizes and type as well as the interaction between the datahandling and the storage systems, ranging from disk pools to tape systems. In particular we will cover the problems faced on large scale compute farms. To accommodate the needs of Grid computing, CDF deployed installations consisting of SAM for datahandling and CAF for high throughput batch processing. The CDF experiment already had experiences with the CAF system. We will report on the deployment of the combined system.

## Event Processing / 469

### A New STAR Event Reconstruction Chain

**Authors:** A. Rose<sup>1</sup>; B. Hippolyte<sup>2</sup>; C. Pruneau<sup>1</sup>; M. Calderon<sup>3</sup>

- <sup>1</sup> *WAYNE STATE UNIVERSITY*  
<sup>2</sup> *Yale University*  
<sup>3</sup> *Brookhaven National Lab*

**Corresponding Author:** pruneau@physics.wayne.edu

We present the design and performance analysis of a new event reconstruction chain deployed for analysis of STAR data acquired during the 2004 run and beyond. The creation of this new chain involved the elimination of obsolete FORTRAN components, and the development of equivalent or superior modules written in C++. The new reconstruction chain features a new and fast TPC cluster finder, a new track reconstruction software (ITTF discussed at CHEP2003), which seamlessly integrate all detector components of the experiment, a new vertex finder, and various post-tracking analysis modules including a V0 finder, and a track kink finder. The new chain is the culmination of a large software development effort involving in excess of ten FTEs.

**Poster Session 2** - Board: 15 / 471

## Monitoring a Petabyte Scale Storage System

**Authors:** A. Moibenko<sup>1</sup>; C-H. Huang<sup>1</sup>; D. Petravick<sup>1</sup>; E. Berman<sup>1</sup>; J. Bakken<sup>1</sup>; M. Zalokar<sup>1</sup>

<sup>1</sup> FERMILAB

**Corresponding Author:** berman@fnal.gov

Fermilab operates a petabyte scale storage system, Enstore, which is the primary data store for experiments' large data sets. The Enstore system regularly transfers greater than 15 Terabytes of data each day. It is designed using a client-server architecture providing sufficient modularity to allow easy addition and replacement of hardware and software components. Monitoring of this system is essential to insure the integrity of the data that is stored in it and to maintain the high volume access that this system supports.

The monitoring of this distributed system is accomplished using a variety of tools and techniques that present information for use by a variety of roles (operator, storage system administrator, storage software developer, user).

All elements of the system are monitored: performance, hardware, firmware, software, network, data integrity.

We will present details of the deployed monitoring tools with an emphasis on the different techniques that have proved useful to each role. Experience with the monitoring tools and techniques, what worked and what did not will be presented.

**Poster Session 1** - Board: 22 / 473

## Designing a Useful Email System

**Authors:** A. Lilianstrom<sup>1</sup>; J. Schmidt<sup>1</sup>; K. Hill<sup>1</sup>; R. Pasetes<sup>1</sup>

<sup>1</sup> Fermilab

Email is an essential part of daily work. The FNAL gateways process in excess of 700,000 messages per week. Among those messages are many containing viruses and unwanted spam. This paper outlines the FNAL email system configuration. We will discuss how we have defined our systems to provide optimum uptime as well as protection against viruses, spam and unauthorized users.

**Poster Session 1** - Board: 8 / 474

## Experiences Building a Distributed Monitoring System

**Authors:** J. Fromm<sup>1</sup>; K. Genser<sup>1</sup>; M. Mengel<sup>1</sup>; T. Levshina<sup>2</sup>; V. Podstavkov<sup>1</sup>

<sup>1</sup> Fermilab

<sup>2</sup> FERMILAB

The NGOP Monitoring Project at FNAL has developed a package which has demonstrated the capability to efficiently monitor tens of thousands of entities on thousands of hosts, and has been in operation for over 4 years. The project has met the majority

of its initial requirements, and also the majority of the requirements discovered along the way. This paper will describe what worked, and what did not, in the first 4 years of the NGOP Project at Fermilab; and we hope will provide valuable lessons for others considering undertaking even larger (GRID-scale) monitoring projects.

**Poster Session 1 - Board: 23 / 475**

## **Patching PCs**

**Authors:** C. Sieh<sup>1</sup>; D. Skow<sup>1</sup>; J. Schmidt<sup>1</sup>; M. Crawford<sup>1</sup>; T. Dawson<sup>1</sup>; T. Doody<sup>1</sup>

<sup>1</sup> *Fermilab*

FNAL has over 5000 PCs running either Linux or Windows software. Protecting these systems efficiently against the latest vulnerabilities that arise has prompted FNAL to take a more central approach to patching systems. We outline the lab support structure for each OS and how we have provided a central solution that works within existing support boundaries. The paper will cover how we identify what patches are considered crucial for a system on the FNAL network and how we verify that systems are appropriately patched.

**Computer Fabrics / 476**

## **CHOS, a method for concurrently supporting multiple operating system**

**Authors:** C.L. Whitney<sup>1</sup>; S. Canon<sup>1</sup>

<sup>1</sup> *NATIONAL ENERGY RESEARCH SCIENTIFIC COMPUTING CENTER*

**Corresponding Author:** canon@nersc.gov

Supporting multiple large collaborations on shared compute farms has typically resulted in divergent requirements from the users on the configuration of these farms. As the frameworks used by these collaborations are adapted to use Grids, this issue will likely have a significant impact on the effectiveness of Grids. To address these issues, a method was developed at Lawrence Berkeley National Lab and is being used in production on the PDSF cluster. This method, termed CHOS, uses a combination of a Linux kernel module, the change root system call, and several utilities to provide access to multiple Linux distributions and versions concurrently on a single system. This method will be presented, along with an explanation on how it is integrated into the login process, grid services, and batch scheduler systems. We will also describe how a distribution is installed and configured to run in this environment and explore some common problems that arise. Finally, we will relate our experience in deploying this framework on a production cluster used by several high energy and nuclear physics collaborations.

**Online Computing / 477**

## **The DZERO Run II Level 3 Trigger and Data Acquisition System**

**Authors:** A Haas<sup>1</sup>; D Chapin<sup>2</sup>; D Cutts<sup>2</sup>; G Brooijmans<sup>3</sup>; G. Watts<sup>4</sup>; R Rechenmacher<sup>5</sup>; R. HAUSER<sup>6</sup>; S Snyder<sup>7</sup>

<sup>1</sup> *University of Washington*

<sup>2</sup> *Brown University*

<sup>3</sup> *Columbia University*

<sup>4</sup> *UNIVERSITY OF WASHINGTON*

<sup>5</sup> *Fermi National Accelerator Laboratory*

<sup>6</sup> *Michigan State University, USA*

<sup>7</sup> *Brookhaven National Laboratory*

**Corresponding Authors:** dpchapin@fnal.gov, gwatts@uw.edu

The DZERO Level 3 Trigger and Data Acquisition (L3DAQ) system has been running continuously since Spring 2002. DZERO is located at one of the two interaction points in the Fermilab Tevatron Collider. The L3DAQ moves front-end readout data from VME crates to a trigger processor farm. It is built upon a Cisco 6509 Ethernet switch, standard PCs, and commodity VME single board computers. We will report on operating experience, performance, and upgrades. In particular, issues related to hardware quality, networking and security, and an expansion of the trigger farm will be discussed.

**Poster Session 2 - Board: 23 / 480**

## Generic logging layer for the distributed computing

**Authors:** G. Van Buren<sup>1</sup>; Jerome LAURET<sup>1</sup>; V. Fine<sup>1</sup>

<sup>1</sup> *BROOKHAVEN NATIONAL LABORATORY*

**Corresponding Author:** jlauret@bnl.gov

Most HENP experiment software includes a logging or tracing API allowing for displaying in a particular format important feedback coming from the core application. However, inserting log statements into the code is a low-tech method for tracing the program execution flow and often leads to a flood of messages in which the relevant ones are occluded. In a distributed computing environment, accessing the information via a log-file is no longer applicable and the approach fails to provide runtime tracing.

Running a job involves a chain of events where many components are involved often written in diverse languages and not offering a consistent and easily adaptable interface for logging important events.

We will present an approach based on a new generic layer built on top of a logger family derived from the Jakarta log4j project that includes log4cxx, log4c, log4perl packages. This provides consistency across packages and framework.

Additionally, the power of using log4j, is the possibility to enable logging (or features) at runtime without modifying the application binary or the wrapper layers. We provide a C++ abstract class library that serves as a proxy between the application framework and the distributed environment. The approach is designed so that the debugging statements can remain in shipped code without incurring a heavy performance cost. Logging equips the developer with as detailed context as necessary for application failures, from testing, quality assurance to a production mode limited amount of information. We will explain and show its implementation in the STAR production environment.

Poster Session 2 - Board: 60 / 481

## SAMGrid Experiences with the Condor Technology in Run II Computing

**Authors:** A. Baranovski<sup>1</sup>; A. Kreymer<sup>2</sup>; A. Kumar<sup>1</sup>; A. Lyon<sup>2</sup>; A. Nishandar<sup>3</sup>; G. Garzoglio<sup>2</sup>; I. Terekhov<sup>2</sup>; J. Trumbo<sup>1</sup>; L. Loebel Carpenter<sup>1</sup>; L. Lueking<sup>2</sup>; R. Herber<sup>2</sup>; R. Illingworth<sup>2</sup>; R. Kennedy<sup>2</sup>; R. St.Denis<sup>4</sup>; S. Jain<sup>3</sup>; S. Veseli<sup>1</sup>; S. White<sup>1</sup>; W. Merritt<sup>2</sup>

<sup>1</sup> *Fermilab*

<sup>2</sup> *FERMI NATIONAL ACCELERATOR LABORATORY*

<sup>3</sup> *University of Texas, Arlington*

<sup>4</sup> *Glasgow University*

**Corresponding Author:** terekhov@fnal.gov

SAMGrid is a globally distributed system for data handling and job management, developed at Fermilab for the D0 and CDF experiments in Run II. The Condor system is being developed at the University of Wisconsin for management of distributed resources, computational and otherwise. We briefly review the SAMGrid architecture and its interaction with Condor, which was presented earlier. We then present our experiences using the system in production, which have two distinct aspects.

At the global level, we deployed Condor-G, the Grid-extended Condor, for the resource brokering and global scheduling of our jobs. At the heart of the system is Condor's Matchmaking Service. As a more recent work at the computing element level, we have been benefitting from the large computing cluster at the University of Wisconsin campus. The architecture of the computing facility and the philosophy of Condor's resource management have prompted us to improve the application infrastructure for D0 and CDF, in aspects such as parting with the shared file system or reliance on resources being dedicated. As a result, we have increased productivity and made our applications more portable and Grid-ready. We include some statistics gathered from our experience. Our fruitful collaboration with the Condor team has been made possible by the Particle Physics Data Grid.

Online Computing / 482

## A Level-2 trigger algorithm for the identification of muons in the Atlas Muon Spectrometer

**Author:** A. Di Mattia<sup>1</sup>

<sup>1</sup> *INFN*

The Atlas Level-2 trigger provides a software-based event selection after the initial Level-1 hardware trigger. For the muon events, the selection is decomposed in a number of broad steps: first, the Muon Spectrometer data are processed to give physics quantities associated to the muon track (standalone features extraction) then, other detector data are used to refine the extracted features. The "muFast" algorithm performs the standalone feature extraction, providing a first reduction of the muon event rate from Level-1. It confirms muon track candidates with a precise measurement of the muon momentum. The algorithm is designed to be both conceptually simple and fast so as to be readily implemented in the demanding online environment in which the Level-2 selection code will run.

Never-the-less its physics performance approaches, in some cases, those of the offline reconstruction algorithms. This paper describes the implemented algorithm together with the software techniques employed to increase its timing performance.

**Poster Session 3 - Board: 19 / 483**

## A high-level language for specifying detector coincidences

**Author:** F. Gray<sup>1</sup>

<sup>1</sup> UNIVERSITY OF CALIFORNIA, BERKELEY

**Corresponding Author:** fgray@berkeley.edu

The muCap experiment at the Paul Scherrer Institut (PSI) will measure the rate of muon capture on the proton to a precision of 1% by comparing the apparent lifetimes of positive and negative muons in hydrogen. This rate may be related to the induced pseudoscalar weak form factor of the proton.

Superficially, the muCap apparatus looks something like a miniature model of a collider detector. Muons pass through several beam counters before reaching a hydrogen-filled time projection chamber (TPC) at its core, which acts as both a stopping target and the primary muon detector. It is surrounded by cylindrical wire chambers and a scintillator hodoscope to observe the Michel electrons that emerge from muon decay. The first key step in the analysis of our data is the proper definition of coincidence events across these many detector layers, maximizing the signal significance by suppressing accidental and pileup backgrounds. Part of our analysis software is written in a special-purpose high-level language, called “muon query language” (MQL), in which these coincidences may be specified cleanly. It uses a variant of the relational model, representing the data as a set of tables upon which selection and join operations may be performed. ROOT histograms and trees are defined based on the contents of tables. A preprocessor generates optimized C++ code that implements the operations described in the MQL file, which is suitable for incorporation into our analyzer framework. This talk will describe the MQL approach and our collaboration’s experience with it.

**Poster Session 1 - Board: 27 / 484**

## Monitoring the CDF distributed computing farms

**Authors:** A. Eskola<sup>1</sup>; E. Lipeles<sup>2</sup>; F. Wuerthwein<sup>3</sup>; I. Sfiligoi<sup>4</sup>; M. Neubauer<sup>5</sup>

<sup>1</sup> Rovaniemi Polytechnic, Finland

<sup>2</sup> UNIVERSITY OF CALIFORNIA SAN DIEGO

<sup>3</sup> University of California San Diego

<sup>4</sup> INFN Frascati

<sup>5</sup> University of California, San Diego

**Corresponding Author:** lipeles@cern.ch

CDF is deploying a version of its analysis facility (CAF) at several globally distributed sites. On top of the hardware at each of these sites is either an FBSNG or Condor batch manager and a SAM data handling system which in some cases also makes use of dCache.

The jobs which run at these sites also make use of a central database located at

Fermilab. Each of these systems has its own monitoring. In order to maintain and effectively use the distributed system, it is important that both the administrators and the users can get a complete global view of the system. We will present a system which integrates the monitoring of all of these services into one globally accessible system based on the Monalisa product. This system is intended for administrators to monitor the system status and service level and for users to better locate resources and monitor job progress. In addition, it is meant to satisfy the request by the CDF International Finance Committee that global computing resource usage by CDF can be audited.

**Plenary / 485**

## **Enterasys - Networks that Know**

**Author:** J. ROESE<sup>None</sup>

Today and in the future businesses need an intelligent network. And Enterasys has the smarter solution. Our active network uses a combination of context-based and embedded security technologies - as well as the industry's first automated response capability - so it can manage who is using your network. Our solution also protects the entire enterprise - from the edge, through the distribution layer, and into the core of the network. Threats are recognized and isolated at the user level, rather than taking your entire network down. It even has the ability to coexist with and enhance your legacy data networking infrastructure and existing security appliances - regardless of the vendor. By continually offering a context-based analysis of network traffic, our solution allows you to see not only what the problem is, but also where it is, and who caused it. And, with the industry's most advanced controls, we're the first solution that's able to resolve threats across the entire network - dynamically or on demand.

**Distributed Computing Systems and Experiences / 486**

## **A Data Grid for the Analysis of Data from the Belle Experiment**

**Authors:** G R. Moloney<sup>None</sup>; J. Smillie<sup>None</sup>; L J. Winton<sup>None</sup>; S. McMahon<sup>None</sup>

**Corresponding Author:** glenn@physics.unimelb.edu.au

We have developed and deployed a data grid for the processing of data from the Belle experiment, and for the production of simulated Belle data. The Belle Analysis Data Grid brings together compute and storage resources across five separate partners in Australia, and the Computing Research Centre at the KEK laboratory in Tsukuba, Japan.

The data processing resources are general purpose, shared use, compute clusters at the Universities of Melbourne and Sydney, the Australian Partnership for Advanced Computing (APAC), the Victorian Partnership for Advanced Computing (VPAC) and the Australian Centre for Advanced Computing and Communications (AC3).

This system is in use for the Australian contribution to the production of simulated data for the Belle experiment, and for physics



analyses.

The Storage Resource Broker (SRB), from the San Diego Supercomputing Centre, is used to provide a robust underlying data repository. A federation of SRB servers has been established to share and manage Belle data between the KEK laboratory, the mass data store at the Australian National University (ANU) and satellite storage at each of the compute clusters.

The globus toolkit is the underlying technology for the management of the computing resources, and the despatching of jobs. A network aware job scheduler has been developed. The scheduler queries the SRB servers for location of data replicas, and arranges scheduling of processing and production jobs on the compute resources according to a static model of the network connectivity and dynamic assessment of the relative system loads.

**Computer Fabrics / 487**

## **InfiniBand for High Energy Physics**

**Authors:** A. Heiss<sup>1</sup>; G. GANIS<sup>2</sup>; U. Schwickerath<sup>3</sup>

<sup>1</sup> FORSCHUNGSZENTRUM KARLSRUHE

<sup>2</sup> CERN

<sup>3</sup> Forschungszentrum Karlsruhe

**Corresponding Author:** andreas.heiss@iwr.fzk.de

Distributed physics analysis techniques as provided by the rootd and proofd concepts require a fast and efficient interconnect between the nodes. Apart from the required bandwidth the latency of message transfers is important, in particular in environments with many nodes. Ethernet is known to have large latencies, between 30 and 60 micro seconds for the common Giga-bit Ethernet.

The InfiniBand architecture is a relatively new, open industry standard. It defines a switched high-speed, low-latency fabric designed to connect compute nodes

and I/O nodes with copper or fibre cables. The theoretical bandwidth is up to 30 Gbit/s. The Institute for Scientific Computing (IWR) at the Forschungszentrum Karlsruhe is testing InfiniBand technology since begin of 2003, and has a cluster of

dual Xeon nodes using the 4X (10 Gbit/s) version of the interconnect.

Bringing the RFIO protocol - which is part of the CERN CASTOR facilities for sequential file transfers - to InfiniBand has been a big success, allowing significant reduction of CPU consumption and increase of file transfer speed.

A first prototype of a direct interface to InfiniBand for the root toolkit has been designed and implemented.

Experiences with hard- and software, in particular MPI performance results, will be reported.

The methods and first performance results on rfio and root will be shown and compared to other fabric technologies like Ethernet.

**Poster Session 2 - Board: 66 / 488**

## CERN Modular Physics Screensaver or Using spare CPU cycles of CERN's Desktop PCs

**Authors:** A. Wagner<sup>1</sup>; E.M. McIntosh<sup>1</sup>

<sup>1</sup> CERN

**Corresponding Authors:** andreas.wagner@cern.ch, eric.mcintosh@cern.ch

CERN has about 5500 Desktop PCs. These computers offer a large pool of resources that can be used for physics calculations outside office hours.

The paper describes a project to make use of the spare CPU cycles of these PCs for LHC tracking studies. The client server application is implemented as a lightweight, modular screensaver and a Web Application containing the physics job repository. The information exchange between client and server is done using the HTTP protocol. The design and implementation is presented together with results of performance and scalability studies. A typical LHC tracking study involves some 1500 jobs, each over 100,000 turns, requiring about 1 hour of CPU on a modern PC. A reliable and easy to use Linux interface to the CPSS Web application has been provided. It has been used for a production run of 15,000 jobs, using some 50 desktop Windows PCs, which uncovered a numerical incompatibility between Windows 2000 and XP. It is expected to make available up to two orders of magnitude more computing power for these studies at zero cost.

**Computer Fabrics / 489**

## Current Status of Fabric Management at CERN

**Authors:** D. Front<sup>1</sup>; D. Waldron<sup>2</sup>; G. Cancio<sup>2</sup>; H. Meinhard<sup>3</sup>; J. van Eldik<sup>2</sup>; M. Siket<sup>2</sup>; M. Stepniewski<sup>2</sup>; P. Poznanski<sup>2</sup>; S. Chapeland<sup>2</sup>; T. Kleinwort<sup>2</sup>; T. Smith<sup>2</sup>; V. Bahyl<sup>2</sup>; V. Lefebure<sup>2</sup>; W. Tomlin<sup>2</sup>

<sup>1</sup> Weizman Institute of Science, Israel

<sup>2</sup> CERN

<sup>3</sup> CERN-IT

This paper describes the evolution of fabric management at CERN's T0/T1 Computing Center, from the selection and adoption of prototypes produced by the European DataGrid (EDG) project[1] to enhancements made to them.

In the last year of the EDG project, developers and service managers have been working to understand and solve operational and scalability issues.

CERN has adopted and strengthened Quattor[2], EDG's installation and configuration management toolsuite, for managing all Linux clusters and servers in the Computing Center, replacing existing legacy management systems. Enhancements to the original prototype include a redundant and scalable server architecture using proxy technology and producing plug-in components for configuring system and LHC computing services.

CERN now coordinates the maintenance of Quattor, making it available to other sites.

Lemon[3], the EDG fabric monitoring framework, has been progressively deployed onto all managed Linux nodes. We have developed sensors to instrument fabric nodes to provide us with complete performance and exception monitoring information. Performance visualization displays and interfaces to the existing alarm system have also been provided.

LEAF[4], the LHC-Era Automated Fabric toolset, comprises the State Management System, a tool to enable high-level configuration commands to be issued to sets of nodes during both hardware and service management Use Cases, and the Hardware

Management System, a tool for administering hardware workflows and for visualizing and locating equipment.

Finally, we will describe issues currently being addressed and planned future developments.

**Distributed Computing Services / 490**

## **Grid2003 Monitoring, Metrics, and Grid Cataloging System**

**Authors:** A. Zahn<sup>1</sup>; B K. Kim<sup>2</sup>; C. Prescott<sup>2</sup>; I. Fisk<sup>3</sup>; I. Legrand<sup>4</sup>; J. Hicks<sup>5</sup>; J. Rodriguez<sup>2</sup>; J. Weigand<sup>3</sup>; L. Grundhoefer<sup>5</sup>; M. Green<sup>6</sup>; M. Mambelli<sup>1</sup>; P. Avery<sup>2</sup>; R. Gardner<sup>1</sup>; R. Quick<sup>5</sup>

<sup>1</sup> *University of Chicago*

<sup>2</sup> *UNIVERSITY OF FLORIDA*

<sup>3</sup> *FERMI NATIONAL LABORATORY*

<sup>4</sup> *CERN*

<sup>5</sup> *Indiana University*

<sup>6</sup> *University of Buffalo*

**Corresponding Authors:** marco@hep.uchicago.edu, bockjoo@phys.ufl.edu

Grid computing involves the close coordination of many different sites which offer distinct computational and storage resources to the Grid user community. The resources at each site need to be monitored continuously. Static and dynamic site information need to be presented to the user community in a simple and efficient manner.

This paper will present both the design and implementation of the Grid3 monitoring infrastructure and the design details and the functionalities of a new application called the Gridcat.

The Grid3 monitoring architecture follows a user-oriented design that specifies standard metrics and uses different underlying monitoring tools to collect them and build a very diversified framework. In the monitoring framework we integrated existing tools, extended their functionality and developed original new tools. The main tools used include ACDC Job Monitoring from University of Buffalo, Ganglia, a preliminary version of Gridcat, Globus MDS, the University of Chicago Grid telemetry MDViewer, and US CMS MonALISA. From the collected data is extracted information of interest for the VOs participating in the Grid, for example resources provided and used by all VOs and jobs submitted by each VO.

The Gridcat shows site status using a web interface that is simple and powerful enough to determine the site's readiness to accept grid applications by collecting and storing dynamic site information to a database. The status information displayed by the prototype Gridcat was used extensively by the Grid2003 project as a coordination point for the grid operations center.

**Plenary / 491**

## **The BIRN Project: Distributed Information Infrastructure and Multi-scale Imaging of the Nervous System (BIRN = Biomedical Informatics Research Network)**

**Author:** M. Ellisman<sup>1</sup>

<sup>1</sup> *National Center for Microscopy and Imaging Research of the Center for Research in Biological Systems - The Department of Neurosciences, University of California San Diego School of Medicine - La Jolla, California - USA*

The grand goal in neuroscience research is to understand how the interplay of structural, chemical and electrical signals in nervous tissue gives rise to behavior. Experimental advances of the past decades have given the individual neuroscientist an increasingly powerful arsenal for obtaining data, from the level of molecules to nervous systems. Scientists have begun the arduous and challenging process of adapting and assembling neuroscience data at all scales of resolution and across disciplines into computerized databases and other easily accessed sources. These databases will complement the vast structural and sequence databases created to catalogue, organize and analyze gene sequences and protein products. The general premise of the neuroscience goal is simple; namely that with “complete” knowledge of the genome and protein structures accruing rapidly we next need to assemble an infrastructure that will facilitate acquisition of an understanding for how functional complexes operate in their cell and tissue contexts. Our U.C. San Diego-based group is leading several interdisciplinary projects around this grand challenge. We are evolving a shared infrastructure that allows for mapping molecular and cellular brain anatomy in the context of a shared multi-scale mouse brain atlas system, the Cell-Centered Database (CCDB). Complementary to these neuroinformatics activities at the National Center for Microscopy and Imaging Research in San Diego (NCMIR) we have developed new molecular labeling methods compatible with advanced ultra-wide field laser-scanning light microscopy and multi-resolution 3 dimensional electron microscopy. These new labeling and imaging methods are being used to populate the CCDB, using as a driver mouse models of neurological and neuropsychiatric disorders. The informatics framework is facilitating cooperative work by distributed teams of scientists engaged in focused collaborations aimed to deliver new fundamental understanding of structures on the scale of 1 nm<sup>3</sup> to 10<sup>3</sup>’s of μm<sup>3</sup>, a dimensional range that encompasses macromolecular complexes, organelles, and multi-component structures like synapses and the cellular interactions in the context of the complex organization of the entire nervous system. This is a unique and pioneering effort that links new neuroscience techniques and revolutionary advances in information technology. Database federation tools are critical to the scalability of these efforts and future development plans will be described in the context of the NIH-supported project to create a new framework for collaboration and data integration in the Biomedical Informatics Research Network (BIRN). BIRN is the leading example of a virtual database effort that is using the challenge of federating multi-scale distributed data about the nervous systems to help guide the evolution of an International Cyberinfrastructure serving all science disciplines, including biomedicine.

Plenary / 492

## Computing for Belle

Author: N. KATAYAMA<sup>1</sup>

<sup>1</sup> KEK

The Belle experiment operates at the KEKB accelerator, a high luminosity asymmetric energy e<sup>+</sup> e<sup>-</sup> machine. KEKB has achieved the world highest luminosity of 1.39 times 10<sup>34</sup> cm<sup>-2</sup>s<sup>-1</sup>. Belle accumulates more than 1 million B Bbar pairs in one good day. This corresponds to about 1.2 TB of raw data per day. The amount of the raw and processed data accumulated so far exceeds 1.4 PB. Belle’s computing model has been a traditional one and very successful so far. The computing has been managed by minimal number of people using cost effective solutions. Looking at the future, KEKB/Belle plans to improve the luminosity to a few times 10<sup>35</sup> cm<sup>-2</sup>s<sup>-1</sup>, 10 times as much as we obtain now. This presentation describes Belle’s efficient computing operations, struggles to

manage large amount of raw and physics data, and plans for Belle computing for Super KEKB/Belle.

**Event Processing / 493**

## Physics validation of the simulation packages in a LHC-wide effort

**Author:** A. Ribon<sup>1</sup>

<sup>1</sup> CERN

**Corresponding Author:** alberto.ribon@cern.ch

In the framework of the LCG Simulation Physics Validation Project, we present comparison studies between the GEANT4 and FLUKA shower packages and LHC sub-detector test-beam data. Emphasis is given to the response of LHC calorimeters to electrons, photons, muons and pions. Results of “simple-benchmark” studies, where the above simulation packages are compared to data from nuclear facilities, are also shown.

**Core Software / 494**

## Reflection-Based Python-C++ Bindings

**Authors:** J. Generowicz<sup>1</sup>; M. Marino<sup>2</sup>; P. Mato<sup>1</sup>; W. LAVRIJSEN<sup>2</sup>

<sup>1</sup> CERN

<sup>2</sup> LBNL

**Corresponding Authors:** wim.lavrijsen@cern.ch, pere.mato@cern.ch

Python is a flexible, powerful, high-level language with excellent interactive and introspective capabilities and a very clean syntax. As such it can be a very effective tool for driving physics analysis.

Python is designed to be extensible in low-level C-like languages, and its use as a scientific steering language has become quite widespread. To this end, existing and custom-written C or C++ libraries are bound to the Python environment as so-called extension modules. A number of tools for easing the process of creating such bindings exist, such as SWIG or Boost.Python. Yet, the the process still requires a considerable amount of effort and expertise.

The C++ language has little built-in introspective capabilities, but tools such as LCGDict and CINT add this by providing so-called dictionaries: libraries that contain information about the names, entry points, argument types, etc. of other libraries.

The reflection information from these dictionaries can be used for the creation of bindings and so the process can be fully automated, as dictionaries are already provided for many end-user libraries for other purposes, such as object persistency.

PyLCGDict is a Python extension module that uses LCG dictionaries, as PyROOT uses CINT reflection information, to allow Python users to access C++ libraries with essentially no preparation on the users’

behalf. In addition, and in a similar way, PyROOT gives ROOT users access to Python libraries.

**Poster Session 1** - Board: 24 / 495

## The Project CampusGrid

**Author:** O. Schneider<sup>1</sup>

<sup>1</sup> FZK

**Corresponding Author:** olaf.schneider@iwr.fzk.de

A central idea of Grid Computing is the virtualization of heterogeneous resources. To meet this challenge the Institute for Scientific Computing, IWR, has started the project CampusGrid. Its medium term goal is to provide a seamless IT environment supporting the on-site research activities in physics, bioinformatics, nanotechnology and meteorology. The environment will include all kinds of HPC resources: vector computers, shared memory SMP servers and clusters of commodity components as well as a shared high-performance storage solution. After introducing the general ideas the talk will inform about the current project status and scheduled development tasks. This is associated with reports on other activities in the fields of Grid computing and high performance computing at IWR.

**Computer Fabrics** / 496

## Developing & Managing a large Linux farm - the Brookhaven Experience

**Authors:** A. Withers<sup>1</sup>; C. Hollowell<sup>2</sup>; J. Smith<sup>1</sup>; O. Rind<sup>2</sup>; R. Hogue<sup>2</sup>; S. Misawa<sup>2</sup>; T. Chan<sup>1</sup>; Tomasz WLODEK<sup>3</sup>

<sup>1</sup> BROOKHAVEN NATIONAL LAB

<sup>2</sup> Brookhaven National Lab

<sup>3</sup> BNL

**Corresponding Author:** tony@bnl.gov

This presentation describes the experiences and the lessons learned by the RHIC/ATLAS Computing Facility (RACF) in building and managing its 2,700+ CPU (and growing) Linux Farm over the past 6+ years. We describe how hardware cost, end-user needs, infrastructure, footprint, hardware configuration, vendor selection, software support and other considerations have played a role in the process of steering the growth of the RACF Linux Farm, and how they help shape our future hardware purchase decisions. As well as a detailed description of the challenges encountered and of the solutions used in managing and configuring a large, heterogenous Linux Farm (2700+ CPU's) in the midst of an ongoing transition from being a generally local resource to a global,

Grid-aware resource within a larger, distributed computing environment is provided.

**Distributed Computing Systems and Experiences / 497**

## Distributed Computing Grid Experiences in CMS DC04

**Authors:** A. Anzar<sup>1</sup>; A. Fanfani<sup>2</sup>; A. Heavey<sup>1</sup>; A. Sciaba<sup>3</sup>; A. Zahn<sup>4</sup>; B. MacEvoy<sup>5</sup>; B. Martelli<sup>6</sup>; C. Grandi<sup>7</sup>; C. Prescott<sup>4</sup>; D. Bonacorsi<sup>6</sup>; D. Bradley<sup>8</sup>; D. Colling<sup>9</sup>; F. Donno<sup>3</sup>; F. Fanzago<sup>10</sup>; G. Donvito<sup>11</sup>; G. Graham<sup>1</sup>; G. Maggi<sup>11</sup>; H. Tallini<sup>9</sup>; I. Fisk<sup>1</sup>; I. Legrand<sup>12</sup>; J. Hernandez<sup>13</sup>; J. Kaiser<sup>1</sup>; J. Letts<sup>14</sup>; J. Rodriguez<sup>4</sup>; J. Weigand<sup>1</sup>; K. Bockjoo<sup>4</sup>; L. Wakefield<sup>9</sup>; L. Dell'Agnello<sup>6</sup>; L. Giacchetti<sup>1</sup>; L. Tuura<sup>15</sup>; M. Biasotto<sup>16</sup>; M. Corvo<sup>10</sup>; M. Ernst<sup>17</sup>; M. Mazzucato<sup>10</sup>; N. De Filippis<sup>11</sup>; N. Kuropatine<sup>1</sup>; N. Ratnikova<sup>1</sup>; N. Sinanis<sup>18</sup>; P. Capiluppi<sup>7</sup>; P. GARCIA-ABIA<sup>19</sup>; R. Pordes<sup>1</sup>; S. Campana<sup>3</sup>; S. Fantinel<sup>16</sup>; S. Suresh<sup>12</sup>; T. Martin<sup>14</sup>; T. Perelmutov<sup>1</sup>; W. Jank<sup>18</sup>; Y. Wu<sup>1</sup>

<sup>1</sup> FERMILAB

<sup>2</sup> INFN-BOLOGNA (ITALY)

<sup>3</sup> EIS CERN

<sup>4</sup> UFL

<sup>5</sup> IMPERIAL COLLEGE LONDON

<sup>6</sup> INFN-CNAF

<sup>7</sup> INFN-Bologna

<sup>8</sup> University of Wisconsin

<sup>9</sup> Imperial College London

<sup>10</sup> INFN-Padova

<sup>11</sup> INFN-Bari

<sup>12</sup> Caltech

<sup>13</sup> CIEMAT Madrid

<sup>14</sup> UCSD

<sup>15</sup> Northeastern University, Boston

<sup>16</sup> INFN-Legnaro

<sup>17</sup> DESY

<sup>18</sup> CERN

<sup>19</sup> CIEMAT

**Corresponding Authors:** alessandra.fanfani@cern.ch, ian.fisk@cern.ch

In March-April 2004 the CMS experiment undertook a Data Challenge(DC04). During the previous 8 months CMS undertook a large simulated event production. The goal of the challenge was to run CMS reconstruction for sustained period at 25Hz input rate, distribute the data to the CMS Tier-1 centers and analyze them at remote sites. Grid environments developed in Europe by the LHC Computing Grid (LCG) in Europe and in the US with Grid2003 were utilized to complete the aspects of the challenge.

During the simulation phase, US-CMS utilized Grid2003 to simulate and process approximately 17 million events. Simultaneous usage of CPU resources peaked at 1200 CPUs, controlled by a single FTE. Using Grid3 was a milestone for CMS computing in reaching a new magnitude in the number of autonomously cooperating computing sites for production. The use of Grid-based job execution resulted in reducing the overall support effort required to submit and monitor jobs by a factor of two.

During the challenge itself, the CMS groups from Italy and Spain used the LCG Grid Environment to satisfy challenge requirements . The LCG Replica

Manager was used to transfer the data. The CERN RLS provided the needed replica catalogue functionality. The LCG submission system based on the Resource Broker was used to submit analysis jobs to the sites hosting the data. A CMS dedicated GridICE monitoring was activated to monitor both services and resources.

A description of the experiences, successes and lessons learned from both experiences with grid infrastructure is presented.

## Distributed Computing Systems and Experiences / 498

### Role of Tier-0, Tier-1 and Tier-2 Regional Centres in CMS DC04

**Authors:** A. Anzar<sup>1</sup>; A. Chierici<sup>2</sup>; A. Fanfani<sup>3</sup>; A. Nowack<sup>4</sup>; A. Pacheco<sup>5</sup>; B. Martelli<sup>2</sup>; C. Charlot<sup>6</sup>; C. Grandi<sup>7</sup>; C. Shepherd-Themistocleous<sup>8</sup>; D. Newbold<sup>8</sup>; D. Stickland<sup>9</sup>; F. Fanzago<sup>10</sup>; F. Martinez<sup>5</sup>; F. Rosso<sup>2</sup>; F. Ruggieri<sup>2</sup>; F.J. Rodriguez Calonge<sup>11</sup>; G Donivito<sup>12</sup>; G. Lo Re<sup>13</sup>; G. Maggi<sup>12</sup>; G. Merino<sup>5</sup>; G. Quast<sup>14</sup>; I. Fisk<sup>1</sup>; I. Semenouk<sup>6</sup>; J. Andreeva<sup>15</sup>; J. Hernandez<sup>16</sup>; J. Rehn<sup>14</sup>; K. Rabbertz<sup>14</sup>; L. Bauerdick<sup>1</sup>; L. Dell'Agnello<sup>2</sup>; L. Tuura<sup>17</sup>; M. Biasotto<sup>18</sup>; M. Corvo<sup>10</sup>; M. Delfino<sup>5</sup>; M. Ernst<sup>19</sup>; M. Mazzucato<sup>20</sup>; M. Rodriguez<sup>5</sup>; N. Colino<sup>11</sup>; N. De Filippis<sup>12</sup>; N. Sinanis<sup>15</sup>; P. Capiluppi<sup>3</sup>; P. Garcia-Abia<sup>11</sup>; P. Mine<sup>6</sup>; P. Ricci<sup>2</sup>; R. Harris<sup>1</sup>; S. Fantinel<sup>18</sup>; S. Metson<sup>21</sup>; T. Barrass<sup>21</sup>; T. Wildish<sup>9</sup>; W. Jank<sup>15</sup>; Y. Wu<sup>1</sup>

<sup>1</sup> *Fermilab*

<sup>2</sup> *INFN-CNAF*

<sup>3</sup> *INFN Bologna*

<sup>4</sup> *RWTH Aachen*

<sup>5</sup> *PIC Barcelona*

<sup>6</sup> *Ecole Poly*

<sup>7</sup> *INFN-Bologna*

<sup>8</sup> *RAL*

<sup>9</sup> *Princeton University*

<sup>10</sup> *INFN Padova*

<sup>11</sup> *CIEMAT*

<sup>12</sup> *INFN Bari*

<sup>13</sup> *INFN & CNAF Bologna*

<sup>14</sup> *Karlsruhe University*

<sup>15</sup> *CERN*

<sup>16</sup> *CIEMAT Madrid*

<sup>17</sup> *Northeastern University*

<sup>18</sup> *INFN Legnaro*

<sup>19</sup> *DESY*

<sup>20</sup> *INFN-Padova*

<sup>21</sup> *Bristol University*

The CMS 2004 Data Challenge (DC04) was devised to test several key aspects of the CMS Computing Model in three ways: by trying to sustain a 25 Hz reconstruction rate at the Tier-0; by distributing the reconstructed data to six Tier-1 Regional Centers (FNAL in US, FZK in Germany, Lyon in France, CNAF in Italy, PIC in Spain, RAL in UK) and handling catalogue issues; by redistributing data to Tier-2 centers for analysis. Simulated events, up to the digitization step, were produced prior to the DC as input for the reconstruction in the Pre-Challenge Production (PCP04).



In this paper, the model of the Tier-0 implementation used in DC04 is described, as well as the experience gained in using the newly developed data distribution management layer, which allowed CMS to successfully direct the distribution of data from Tier-0 to Tier-1 sites by loosely integrating a number of available Grid components. While developing and testing this system, CMS explored the overall functionality and limits of each component, in any of the different implementations which were deployed within DC04.

The role of Tier-1's is presented and discussed, from the import of reconstructed data from Tier-0, to the archiving on to the local mass storage system and the data distribution management to Tier-2's for analysis. Participating Tier-1's differed in available resources, set-up and configuration: a critical evaluation of the results and performances achieved adopting different strategies in the organization and management of each Tier-1 center to support CMS DC04 is presented.

#### Distributed Computing Systems and Experiences / 499

### Performance of the NorduGrid ARC and the Dulcinea Executor in ATLAS Data Challenge 2

**Authors:** A. Filipcic<sup>1</sup>; A. Konstantinov<sup>2</sup>; A. Read<sup>3</sup>; A. Sandgren<sup>4</sup>; A. Teras<sup>5</sup>; A. Waananen<sup>6</sup>; B. Konya<sup>7</sup>; C. Haeberli<sup>8</sup>; D. Kalici<sup>9</sup>; F. Ould-Saada<sup>10</sup>; G. Duckeck<sup>11</sup>; H. Heller<sup>12</sup>; H. Jensen<sup>13</sup>; H. Riiser<sup>3</sup>; J.R. Hansen<sup>6</sup>; J. Kennedy<sup>11</sup>; J. Kleist<sup>13</sup>; J. Lindemann<sup>14</sup>; J.F. Mycklebust<sup>15</sup>; J.L. Nielsen<sup>6</sup>; K. Pajchel<sup>3</sup>; L. Nixon<sup>16</sup>; L. malinowsky<sup>17</sup>; M. Ellert<sup>18</sup>; M. Hanshaugen<sup>19</sup>; M. Tadel<sup>1</sup>; M. Wadenstein<sup>4</sup>; N. Smeds<sup>17</sup>; O. Smirnova<sup>20</sup>; P. Eerola<sup>7</sup>; R. Sturrock<sup>21</sup>; S. Sunde<sup>19</sup>; T. Ekelof<sup>22</sup>

<sup>1</sup> JSI Ljubljana

<sup>2</sup> Vilnius University, Lithuania

<sup>3</sup> University of Oslo

<sup>4</sup> Umea University

<sup>5</sup> Helsinki

<sup>6</sup> Niels Bohr Institute, Copenhagen, Denmark

<sup>7</sup> Lund University, Sweden

<sup>8</sup> University of Bern

<sup>9</sup> Odense

<sup>10</sup> University of Oslo, Norway

<sup>11</sup> LMU Munich

<sup>12</sup> LR, Munich

<sup>13</sup> Alborg University

<sup>14</sup> Lund

<sup>15</sup> Bergen

<sup>16</sup> Linkoping University

<sup>17</sup> KTH Stockholm

<sup>18</sup> CERN

<sup>19</sup> USIT Oslo

<sup>20</sup> LUND UNIVERSITY

<sup>21</sup> University of Melbourne

<sup>22</sup> Uppsala University, Sweden

**Corresponding Author:** oxana.smirnova@hep.lu.se

This talk describes the various stages of ATLAS Data Challenge 2 (DC2) in what concerns usage of resources deployed via NorduGrid's Advanced Resource Connector (ARC). It also describes the integration of these resources with the ATLAS production system using the Dulcinea executor.

ATLAS Data Challenge 2 (DC2), run in 2004, was designed to be a step forward in the distributed data processing. In particular, much coordination of task assignment to resources was planned to be delegated to Grid in its different flavours. An automatic production management system was designed, to direct the tasks to Grids and conventional resources.

The Dulcinea executor is a part of this system that provides interface to the information system and resource brokering capabilities of the ARC middleware. The executor translates the job definitions received from the supervisor to the extended resource specification language (XRSL) used by the ARC middleware. It also takes advantage of the ARC middleware's built-in support for the Globus Replica Location Server (RLS) for file registration and lookup.

NorduGrid's ARC has been deployed on many ATLAS-dedicated resources across the world in order to enable effective participation in ATLAS DC2. This was the first attempt to harness large amounts of strongly heterogeneous resources in various countries for a single collaborative exercise using Grid tools. This talk addresses various issues that arose during different stages of DC2 in this environment: preparation, such as ATLAS software installation; deployment of the middleware; and processing. The results and lessons are summarized as well.

**Distributed Computing Services / 500**

## Housing Metadata for the Common Physicist Using a Relational Database

**Authors:** A. Baranovski<sup>1</sup>; A. Kreymer<sup>2</sup>; A. Lyon<sup>2</sup>; A. Sill<sup>3</sup>; C. Cioffi<sup>4</sup>; C. Nicholson<sup>5</sup>; F. Ratnikov<sup>6</sup>; G. Garzoglio<sup>2</sup>; G. McCance<sup>7</sup>; I. Terekhov<sup>1</sup>; J. Fulachier<sup>8</sup>; J. Trumbo<sup>9</sup>; K. Nienartowicz<sup>7</sup>; L. Loebel Carpenter<sup>1</sup>; L. Lueking<sup>2</sup>; M. Burgon-Lyon<sup>10</sup>; M. Leslie<sup>4</sup>; M. Zimmler<sup>1</sup>; P. Vokac<sup>1</sup>; P. Kunszt<sup>7</sup>; P. Millar<sup>5</sup>; R. Herber<sup>2</sup>; R. Illingworth<sup>2</sup>; R. Kennedy<sup>2</sup>; R. Rocha<sup>7</sup>; R. St.Denis<sup>11</sup>; S. Albrand<sup>12</sup>; S. Belforte<sup>13</sup>; S. Hanlon<sup>10</sup>; S. Stonjek<sup>1</sup>; S. Veseli<sup>1</sup>; S. White<sup>1</sup>; T. Barrass<sup>14</sup>; T. Doyle<sup>15</sup>; U. Kerzel<sup>16</sup>; V. Bartsch<sup>4</sup>; W. Merritt<sup>2</sup>

<sup>1</sup> *Fermilab*

<sup>2</sup> *FERMI NATIONAL ACCELERATOR LABORATORY*

<sup>3</sup> *Texas Tech University*

<sup>4</sup> *Oxford University*

<sup>5</sup> *UNIVERSITY OF GLASGOW*

<sup>6</sup> *Rutgers University*

<sup>7</sup> *CERN*

<sup>8</sup> *LPSC, Grenoble*

<sup>9</sup> *FERMILAB*

<sup>10</sup> *Glasgow University*

<sup>11</sup> *University of Glasgow*

<sup>12</sup> *Grenoble*

<sup>13</sup> *INFN/Trieste*

<sup>14</sup> *University of Bristol*

<sup>15</sup> *Glasgow University*

<sup>16</sup> *Karlsruhe University*

**Corresponding Author:** wyatt@fnal.gov

SAM was developed as a data handling system for Run II at Fermilab. SAM is a collection of services, each described by metadata. The metadata are modeled on a relational database, and implemented in ORACLE. SAM, originally deployed in production for the D0 Run II experiment, has now been also deployed at CDF and is being commissioned at MINOS. This illustrates that the metadata decomposition of its services has a broader applicability than just one experiment. A joint working group on metadata with representatives from ATLAS, BaBar, CDF, CMS, D0, and LHCb in cooperation with EGEE has examined this metadata decomposition in the light of general HEP user requirements.

Greater understanding of the required services of a performant data handling system has emerged from Run II experience. This experience is being merged with the understanding being developed in the course of LHC experience with data challenges and user case discussions. We describe the SAM schema and the commonalities of function and service support between this schema and proposals for the LHC experiments. We describe the support structure required for SAM schema updates, the use of development, integration, and production instances. We are also looking at the LHC proposals for the evolution of schema using keyword-value pairs that are then transformed into a normalized, performant database schema.

**Distributed Computing Systems and Experiences / 501**

## **ATLAS Production System in ATLAS Data Challenge 2**

**Authors:** K. DE<sup>1</sup>; L. GOOSSENS<sup>2</sup>

<sup>1</sup> *UNIVERSITY OF TEXAS AT ARLINGTON*

<sup>2</sup> *CERN*

**Corresponding Author:** luc.goossens@cern.ch

In order to validate the Offline Computing Model and the complete software suite, ATLAS is running a series of Data Challenges (DC). The main goals of DC1 (July 2002 to April 2003) were the preparation and the deployment of the software required for the production of large event samples, and the production of those samples as a worldwide distributed activity.

DC2 (May 2004 until October 2004) is divided into three phases: (i) Monte Carlo data are produced using GEANT4 on three different Grids, LCG, Grid3 and NorduGrid; (ii) simulate the first pass reconstruction of data expected in 2007, also called Tier0 exercise, using the MC sample; and (iii) test the Distributed Analysis model.

A new automated data production system has been developed for DC2. The major design objectives are minimal human involvement, maximal robustness, and interoperability with several grid flavors and legacy systems. A central component of the production system is the production database holding information about all jobs. Multiple instances of a 'supervisor' component pick up unprocessed jobs from this database, distribute them to 'executor'

processes, and verify them after execution. The 'executor' components interface to a particular grid or legacy flavour. The job distribution model is a combination of push and pull. A data management system keeps track of all produced data and allows for file transfers.

The basic elements of the production system are described. Experience with the use of the system in world-wide DC2 production of ten million events will be presented. We also present how the three Grid flavors are operated and monitored. Finally we discuss the first attempts on using the Distributed Analysis system.

**Plenary / 502**

## **BaBar computing - From collisions to physics results**

**Author:** P. ELMER<sup>1</sup>

<sup>1</sup> *Princeton University*

**Corresponding Author:** peter.elmer@cern.ch

The BaBar experiment at SLAC studies B-physics at the Upsilon(4S) resonance using the high-luminosity e+e- collider PEP-II at the Stanford Linear Accelerator Center (SLAC). Taking, processing and analyzing the very large data samples is a significant computing challenge.

This presentation will describe the entire BaBar computing chain and illustrate the solutions chosen as well as their evolution with the ever higher luminosity being delivered by PEP-II. This will include data acquisition and software triggering in a high availability, low-deadtime online environment, a prompt, automated calibration pass through the data SLAC and then the full reconstruction of the data that takes place at INFN-Padova within 24 hours. Monte Carlo production takes place in a highly automated fashion in 25+ sites. The resulting real and simulated data is distributed and made available at SLAC and other computing centers.

For analysis a much more sophisticated skimming pass has been introduced in the past year, along with a reworked eventstore. This allows 120 highly customized analysis-specific skims to be produced for direct use by the analysis groups. This skim data format is the same eventstore data as that produced directly by the data and monte carlo productions and can be handled and distributed in the same way.

The total data volume in BaBar is about 1.5PB.

**Distributed Computing Systems and Experiences / 503**

## **ATLAS Data Challenge Production on Grid3**

**Authors:** A. Vaniachine<sup>1</sup>; E. May<sup>1</sup>; G. Gieraltowski<sup>1</sup>; H. Severini<sup>2</sup>; K. De<sup>3</sup>; M. Mambelli<sup>4</sup>; M. Sosebee<sup>3</sup>; N. Ozturk<sup>3</sup>; P. McGuigan<sup>3</sup>; P. Nevski<sup>3</sup>; R. Baker<sup>6</sup>; R. Gardner<sup>4</sup>; W. Deng<sup>6</sup>; X. Zhao<sup>4</sup>; Y. Smirnov<sup>4</sup>

<sup>1</sup> *ARGONNE NATIONAL LABORATORY*

<sup>2</sup> *Oklahoma University*

<sup>3</sup> *University of Texas at Arlington*

<sup>4</sup> *UNIVERSITY OF CHICAGO*

<sup>5</sup> BROOKHAVEN NATIONAL LABORATORY

<sup>6</sup> Brookhaven National Laboratory

**Corresponding Authors:** marco@hep.uchicago.edu, robert.w.gardner@cern.ch

We describe the design and operational experience of the ATLAS production system as implemented for execution on Grid3 resources. The execution environment consisted of a number of grid-based tools: Pacman for installation of VDT-based Grid3 services and ATLAS software releases, the Capone execution service built from the Chimera/Pegasus virtual data system for directed acyclic graph (DAG) generation, DAGMan/Condor-G for job submission and management, and the Windmill production supervisor which provides the messaging system for distributing production tasks to Capone. Produced datasets were registered into a distributed replica location service (Globus RLS) that was integrated with the Don Quixote proxy service for interoperability with other Grids used by ATLAS. We discuss performance, scalability, and fault handling during the first phase of ATLAS Data Challenge 2.

**Poster Session 2 - Board: 6 / 504**

## Global Grid User Support for LCG

**Authors:** G. GREIN<sup>None</sup>; H. DRES<sup>None</sup>; T. ANTONI<sup>1</sup>; W. THOENE<sup>None</sup>

<sup>1</sup> GGUS

For very large projects like the LHC Computing Grid Project (LCG) involving 8,000 scientists from all around the world, it is an indispensable requirement to have a well organized user support. The Institute for Scientific Computing at the Forschungszentrum Karlsruhe started implementing a Global Grid User Support (GGUS) after official assignment of the Grid Deployment Board in March 2003. For this purpose a web portal and a helpdesk application have been developed. As a single entry point for all Grid related issues and problems GGUS follows the objectives of providing news, documentation and status information about Grid resources. The user will find forms to submit and track service requests. GGUS collaborates with different support teams in the Grid environment like the Grid Operations Center and the Experiment Specific Support. They can access the helpdesk system via web interface. GGUS stores all the incoming trouble tickets and outgoing solutions in a central database and plans to build up a knowledge base where all the information can be offered in a structured manner.

As a prototype GGUS started operation at the Forschungszentrum Karlsruhe in October 2003 and supported local user groups of the German Tier 1 Computing Center, called GridKa. 4 month later the GGUS system was opened for the LCG community. The GGUS system will be explained and demonstrated. The present status of GGUS within the LCG environment will be discussed.

**Plenary / 505**

## Concepts and technologies used in contemporary DAQ systems

**Author:** M. Purschke<sup>1</sup>

<sup>1</sup> Brookhaven National Laboratory

**Corresponding Author:** purschke@bnl.gov

The concepts and technologies applied in data acquisition systems have changed dramatically over the past 15 years. Generic DAQ components and standards such as CAMAC and VME have largely been replaced by dedicated FPGA and ASIC boards, and dedicated real-time operation systems like OS9 or VxWorks have given way to Linux-based trigger processor and event building farms. We have also seen a shift from standard or proprietary bus systems used in event building to GigaBit networks and commodity components, such as PCs. With the advances in processing power, network throughput, and storage technologies, today's data rates in large experiments routinely reach hundreds of MegaBytes/s.

We will present examples of contemporary DAQ systems from different experiments, try to identify or categorize new approaches, and will compare the performance and throughput of existing DAQ systems with the projected data rates of the LHC experiments to see how close we have come to accomplish these goals. We will also try to look beyond the field of High-Energy Physics and see if there are trends and technologies out there which are worth keeping an eye on.

#### Distributed Computing Systems and Experiences / 506

### The ALICE Data Challenge 2004 and the ALICE distributed analysis prototype

**Authors:** A. Morsch<sup>1</sup>; A. Peters<sup>2</sup>; F. Carminati<sup>1</sup>; F. Rademakers<sup>1</sup>; L. Betev<sup>1</sup>; P. Buncic<sup>1</sup>; P. Hristov<sup>1</sup>

<sup>1</sup> CERN

<sup>2</sup> ce

**Corresponding Authors:** andreas.joachim.peters@cern.ch, federico.carminati@cern.ch

During the first half of 2004 the ALICE experiment has performed a large distributed computing exercise with two major objectives: to test the ALICE computing model, included distributed analysis, and to provide data sample for a refinement of the ALICE Jet physics Monte-Carlo studies. Simulation reconstruction and analysis of several hundred thousand events were performed, using the heterogeneous resources of tens of computer centres worldwide. These resources belong to different GRID systems and were steered by the AliEn (ALICE Environment) framework, acting as a meta-GRID. This has been a very thorough test of the middleware of AliEn and LCG (LCG-2 and grid.it resources) and their compatibility. During the Data Challenge more than 1,500 jobs run in parallel for several weeks. More than 50 TB of data have been produced and analysed worldwide in one of the major exercises of this kind run to date. ALICE has developed an analysis system based on AliEn and ROOT. This system starts with a metadata selection in the AliEn file catalogue, followed by a computation phase. Analysis jobs are sent where the data is, thus minimising data movement. The control is performed by an intelligent workload management system. The analysis can be done either via batch or interactive jobs. The latter are "spawned" on remote systems and report the results back to the user workstation. The talk will describe the ALICE experience with this large-scale use of the Grid, the major lessons learned and the consequences for the ALICE computing model.

#### Plenary / 509

### The Evolving Wide Area Network Infrastructure in the LHC era

**Author:** P. CLARKE<sup>None</sup>

The global network is more than ever taking its role as the great “enabler” for many branches of science and research. Foremost amongst such science drivers is of course the LHC/LCG programme, although there are several other sectors with growing demands of the network.

Common to all of these is the realisation that a straightforward over provisioned best efforts wide area IP service is probably not enough for the future.

This talk will summarise the needs of several science sectors, and the advances being made to exploit the current best efforts infrastructure. It will then describe current projects aimed as provisioning “better than best efforts” services (such bandwidth on demand), the global optical R&D testbeds and the strategy of the research network providers to move towards hybrid multi-service networks for the next generation of the global wide area production network.

**Plenary / 510**

## **Physics Validation of the LHC Software**

**Author:** F. GIANOTTI<sup>None</sup>

The LHC Software will be confronted to unprecedented challenges as soon as the LHC will turn on.

We summarize the main Software requirements coming from the LHC detectors, triggers and physics, and we discuss several examples of Software components developed by the experiments and the LCG project (simulation, reconstruction, etc.), their validation, and their adequacy for LHC physics.

**Plenary / 511**

## **Summary - Online Computing**

**Author:** P. VANDE VYVRE<sup>None</sup>

**Plenary / 512**

## **Summary - Event Processing**

**Author:** S. GOWDY<sup>None</sup>

**Plenary / 513**

## **Summary - Core Software**

**Author:** P. CANAL<sup>None</sup>

**Plenary / 514**

## **Summary - Distributed Computing Services**

**Author:** M. Lamanna<sup>1</sup>

<sup>1</sup> *CERN*

**Plenary / 515**

## **Summary - Distributed Computing Systems and Experiences**

**Author:** D. Olson<sup>1</sup>

<sup>1</sup> *Lawrence Berkeley Laboratory*

**Plenary / 516**

## **Summary - Computer Fabrics**

**Author:** T. SMITH<sup>None</sup>

**Plenary / 517**

## **Summary - Wide Area Networking**

**Author:** P. CLARKE<sup>None</sup>

**Plenary / 518**

## **Conference Conclusions**

**Author:** L. BAUERDICK<sup>1</sup>

<sup>1</sup> *FNAL*

**Plenary / 519**

## **Conference Closeout**

**Author:** W. VON RUEDEN<sup>None</sup>

**Plenary / 520**



## Welcome Address

**Author:** W. VON RUEDEN<sup>None</sup>

Plenary / 521

## 50 years of Computing at CERN

**Author:** D. WILLIAMS<sup>None</sup>

“Where are your Wares”

Computing in the broadest sense has a long history, and Babbage (1791-1871), Hollerith (1860-1929) Zuse (1910-1995), many other early pioneers, and the wartime code breakers, all made important breakthroughs. CERN was founded as the first valve-based digital computers were coming onto the market.

I will consider 50 years of Computing at CERN from the following viewpoints:-

Where did we come from? What happened? Who was involved? Which wares (hardware, software, netware, peopeware and now middleware) were important? Where did computers (not) end up in a physics lab? What has been the impact of computing on particle physics? What about the impact of particle physics computing on other sciences? And the impact of our computing outside the scientific realm?

I hope to conclude by looking at where we are going, and by reflecting on why computing is likely to remain challenging for a long time yet.

The topic is so vast that my remarks are likely to be either prejudiced or trivial, or both.

Plenary / 522

## The LCG Project - Preparing for Startup

**Author:** L. ROBERTSON<sup>None</sup>

The talk will cover briefly the current status of the LHC Computing Grid project and will discuss the main challenges facing us as we prepare for the startup of LHC.

Plenary / 523

## The impact of e-science

**Author:** K. PEACH<sup>None</sup>

Just as the development of the World Wide Web has had its greatest impact outside particle physics, so it will be with the development of the Grid.

E-science, of which the Grid is just a small part, is already making a big impact upon many scientific disciplines, and facilitating new scientific discoveries that would be difficult to achieve in any other way. Key to this is the definition and use of metadata.

Plenary / 524

## The role of scientific middleware in the future of HEP computing

Author: M. Livny<sup>1</sup>

<sup>1</sup> UNIVERSITY OF WISCONSIN

In the 18 months since the CHEP03 meeting in San Diego, the HEP community deployed the current generation of grid technologies in a veracity of settings. Legacy software as well as recently developed applications was interfaced with middleware tools to deliver end-to-end capabilities to HEP experiments in different stages of their life cycles. In a series of data challenges, reprocessing efforts and data distribution activities the community demonstrated the benefits distributed computing can offer and the power a range of middleware tools can deliver. After running millions of jobs, moving tera-bytes of data, creating millions of files and resolving hundreds of bug reports, the community also exposed the limitations of these middleware tools. As we move to the next level of challenges, requirements and expectations, we must also examine the methods and procedures we employ to develop, implement and maintain our common suite of middleware tools. The talk will focus on the role common middleware developed by the scientific community can and should play in the software stack of current and future HEP experiments.

Plenary / 525

## The Evolution of Computing: Slowing down? Not Yet!

Author: A. SUTHERLAND<sup>None</sup>

Dr Sutherland will review the evolution of computing over the past decade, focusing particularly on the development of the database and middleware from client server to Internet computing.

But what are the next steps from the perspective of a software company? Dr Sutherland will discuss the development of Grid as well as the future applications revolving around collaborative working, which are appearing as the next wave of computing applications.

Plenary / 526

## Grand Challenges facing Storage Systems

Author: J. MENON<sup>None</sup>

In this talk, we will discuss the future of storage systems. In particular, we will focus on several big challenges which we are facing in storage, such as being able to build, manage and backup really massive storage systems, being able to find information of interest, being able to do long-term archival of data, and so on. We also present ideas and research being done to address these challenges, and provide a perspective on how we expect these challenges to be resolved as we go forward.

Plenary / 527

## Evolution and Revolution in the Design of Computers Based on Nanoelectronics

**Author:** S. WILLIAMS<sup>None</sup>

Today's computers are roughly a factor of one billion less efficient at doing their job than the laws of fundamental physics state that they could be. How much of this efficiency gain will we actually be able to harvest? What are the biggest obstacles to achieving many orders of magnitude improvement in our computing hardware, rather than the roughly factor of two we are used to seeing with each new generation of chip? Shrinking components to the nanoscale offers both potential advantages and severe challenges. The transition from classical mechanics to quantum mechanics is a major issue. Others are the problems of defect and fault tolerance: defects are manufacturing mistakes or components that irreversibly break over time and faults are transient interruptions that occur during operation. Both of these issues become bigger problems as component sizes shrink and the number of components scales up massively. In 1955, John von Neumann showed that a completely general approach to building a reliable machine from unreliable components would require a redundancy overhead of at least 10,000 - this would completely negate any advantages of building at the nanoscale. We have been examining a variety of defect and fault tolerant techniques that are specific to particular structures or functions, and are vastly more efficient for their particular task than the general approach of von Neumann. Our strategy is to layer these techniques on top of each other to achieve high system reliability even with component reliability of no more than 97% or so, and a total redundancy of less than 3. This strategy preserves the advantages of nanoscale electronics with a relatively modest overhead.

**Plenary / 528**

## The IBM Research Global Technology Outlook

**Author:** D. MCQUEENEY<sup>None</sup>

The Global Technology Outlook (GTO) is IBM Research's projection of the future for information technology (IT). The GTO identifies progress and trends in key indicators such as raw computing speed, bandwidth, storage, software technology, and business modeling. These new technologies have the potential to radically transform the performance and utility of tomorrow's information processing systems and devices, ultimately creating new levels of business value.

**Plenary / 530**

## Griding The Nordic Supercomputing Infrastructure

**Author:** B A. YNNERMAN<sup>None</sup>

This talk gives a brief overview of recent development of high performance computing and Grid initiatives in the Nordic region. Emphasis will be placed on the technology and policy demands posed by the integration of general purpose supercomputing centers into Grid environments. Some of the early experiences of bridging national eBorders in the Nordic region will also be presented.

Rather than giving an exhaustive presentation of all projects in the Nordic countries the presentation uses selected examples of Grid projects to show the potential as well as some of the current limitations of Grids.

Plans for a common Nordic Grid Core Facility are currently being made. The presentation gives an overview of these plans and the status of the project. It will also cover a few examples of Nordic Grid initiatives in more detail such as the recently launched SweGrid test bed for production.

**Plenary / 531**

## **Grid Security**

**Author:** D. KELSEY<sup>None</sup>

The aim of Grid computing is to enable the easy and open sharing of resources between large and highly distributed communities of scientists and institutes across many independent administrative domains. Convincing site security officers and computer centre managers to allow this to happen in view of today's ever-increasing Internet security problems is a major challenge. Convincing users and application developers to take security seriously is equally difficult. This paper will describe the main Grid security issues, both in terms of technology and policy, that have been tackled over recent years in LCG and related Grid projects. Achievements to date will be described and opportunities for future improvements will be addressed.

**Plenary / 532**

## **Computing Models and Data Challenges of the LHC experiments**

**Author:** D. STICKLAND<sup>None</sup>

The LHC experiments are undertaking various data-challenges in the run-up to completion of their computing models and the submission of the experiment and of the LHC Computing Grid (LCG), Technical Design Reports(TDR) in 2005. In this talk we summarize the current working models of the LHC Computing Models, identifying their similarities and differences. We summarize the results and status of the data challenges and identify critical areas still to be tested

**BOF : Semantic Web applications in HEP / 533**

## **Defining a Semantic Web Initiative for High Energy Physics**

**Author:** B. White<sup>1</sup>

<sup>1</sup> SLAC

**Corresponding Author:** bebo@slac.stanford.edu

During a recent visit to SLAC, Tim Berners-Lee challenged the High Energy Physics community to identify and implement HEP resources to which Semantic Web technologies could be applied. This challenge comes at a time when a number of other scientific disciplines (for example, bioinformatics and chemistry) have taken a strong initiative in making information resources compatible with Semantic

Web technologies and in the development of associated tools and applications.

The CHEP conference series has a strong history of identifying and encouraging adoption of new technologies. The most notable of these technologies include the Web itself and Grid computing. The Semantic Web could have a similar potential.

Topics of discussion in this BoF include (but are not limited to):  
Definition of the Semantic Web; Semantic Web component technologies; Review of current Semantic Web-related efforts in HEP; Semantic Web resources that are publicly available; What needs to be done next.

**INTAS / 534**

## **INTAS Discussion**

**Author:** INTAS<sup>None</sup>

INTAS (<http://www.intas.be>): International Association for the promotion of co-operation with scientists from the New Independent States of the former Soviet Union (NIS). INTAS encourages joint activities between its INTAS Members and the NIS in all exact and natural sciences, economics, human and social sciences.

INTAS supports a number of NIS participants to attend the 2004 Computing in High Energy Physics Conference (CHEP'04). During CHEP'04, this discussion has been organised so that NIS delegates can meet specifically with their physicists and computer scientists counterparts to discuss topics of mutual interest.

Provisional agenda

- Welcome message from Wolfgang Von Rueden
- Present status about East-West science cooperation inside the High Energy Physics Environment
- Main directions to go
- New developments and technologies needed to achieve the goals, especially in the network infrastructure area
- Time and financial schedules
- Needed man power
- Financial scenarios including funding scenarios (INTAS will need this as they plan for their future activities to focus on selected thematic)

**Plenary / 535**

## **EU Grid Research - Projects and Vision**

**Author:** M. Lemke<sup>None</sup>

The European Grid Research vision as set out in the Information Society Technologies Work Programmes of the EU's Sixth Research Framework Programme is to advance, consolidate and mature Grid technologies for widespread e-science, industrial, business and societal use. A batch of Grid research projects with 52 Million EUR EU support was launched during the European Grid Technology Days 15 - 17 September 2004. The portfolio of projects has the potential for turning Europe's strong competence and critical mass in Grid Research

into competitive advantages. In this presentation, the Grid research vision of the programme and the new project portfolio will be introduced. More information: [www.cordis.lu/ist/grids](http://www.cordis.lu/ist/grids).

**Distributed Computing Services / 536**

## **Huge Memory systems for data-intensive science**

**Author:** R. MOUNT<sup>1</sup>

<sup>1</sup> *SLAC*