

# Computing Technology for Physics Research

Jerome Lauret

Axel Naumann

# Presentations

- 15 parallel contributions
  - 50% distributed computing, little online / control...
- Many well-designed posters
- 7 plenaries
  - wide range of innovative topics
  - thus vividly visited
  - thus no need to repeat here!
- 3 ~~coffee~~ tea breaks

# Panels

- Multicore panel
  - Mohammad Al-Turany, Sverre Jarpe, Alfio Lazzaro, Rama Malladi (Intel)
- Data management panel
  - Rene Brun, Andrew Hanushevsky, Tony Cass, Beob Kyun Kim, Alberto Pace
- Vivid discussion with audience
- Issues brought up will define future of our computing

# PARALLEL PRESENTATIONS

# Computing at Belle II

Thomas Kuhr:

- Belle II going Grid
- Demand peaks might move to cloud
- Bookkeeping, metadata,...
- “nearest Grid site”

# Distributed parallel processing analysis: Belle II, Hyper Suprime-Cam

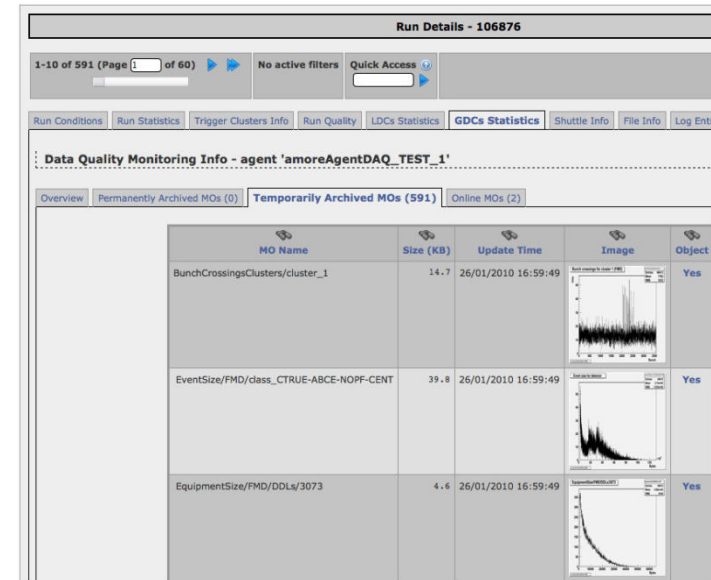
Sogo Mineo:

- Distributed parallel analysis framework ROOBASF
  - ROOT embedded, controls modular-structured workflow
- With MPI, workflow program- or data-parallelized
- Not merely event-parallel, also algorithm-parallel
- Quick development and quick analysis with boost.python:
  - analysis code in Python
  - analysis code in shared libraries called from Python

# The ALICE Online Data Quality Monitoring

Barthelemy Von Haller:

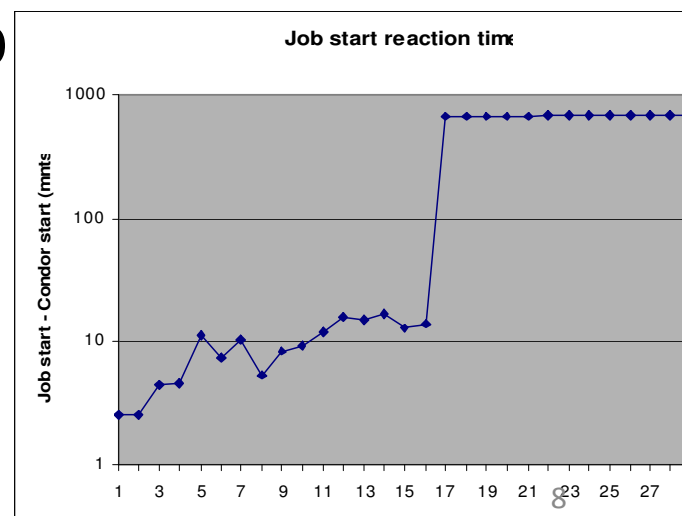
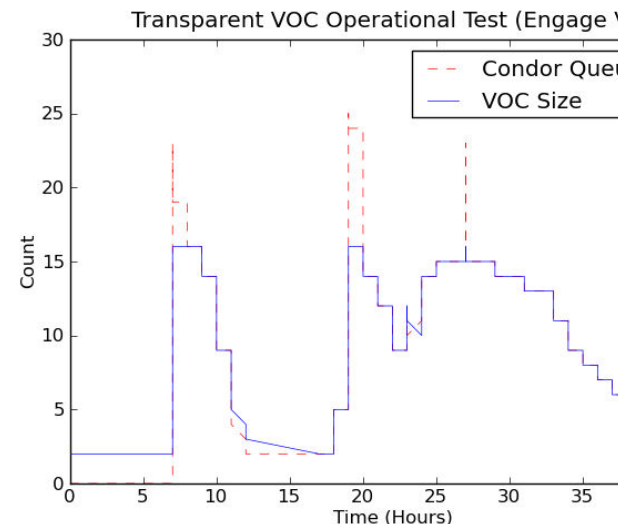
- AMORE (Automatic Monitoring Environment) based on ROOT
- Main repository for all monitoring data
- Multi-core: dedicated threads for time consuming operations + parallelization at event-level
- LHC restart: 35 monitoring agents publishing more than 3400 objects per second



# Contextualization in Practice: The Clemson Experience

Jerome Lauret:

- Grid has differing sites, idea: send environment with it!
  - portable (contextualization), convenient?
- Virtual Organization Clusters VOCs
  - Read-only VO VM on shared storage, writes overlaid to /tmp
  - Managed lifetime (queue,...)
- Promising test results, low overhead





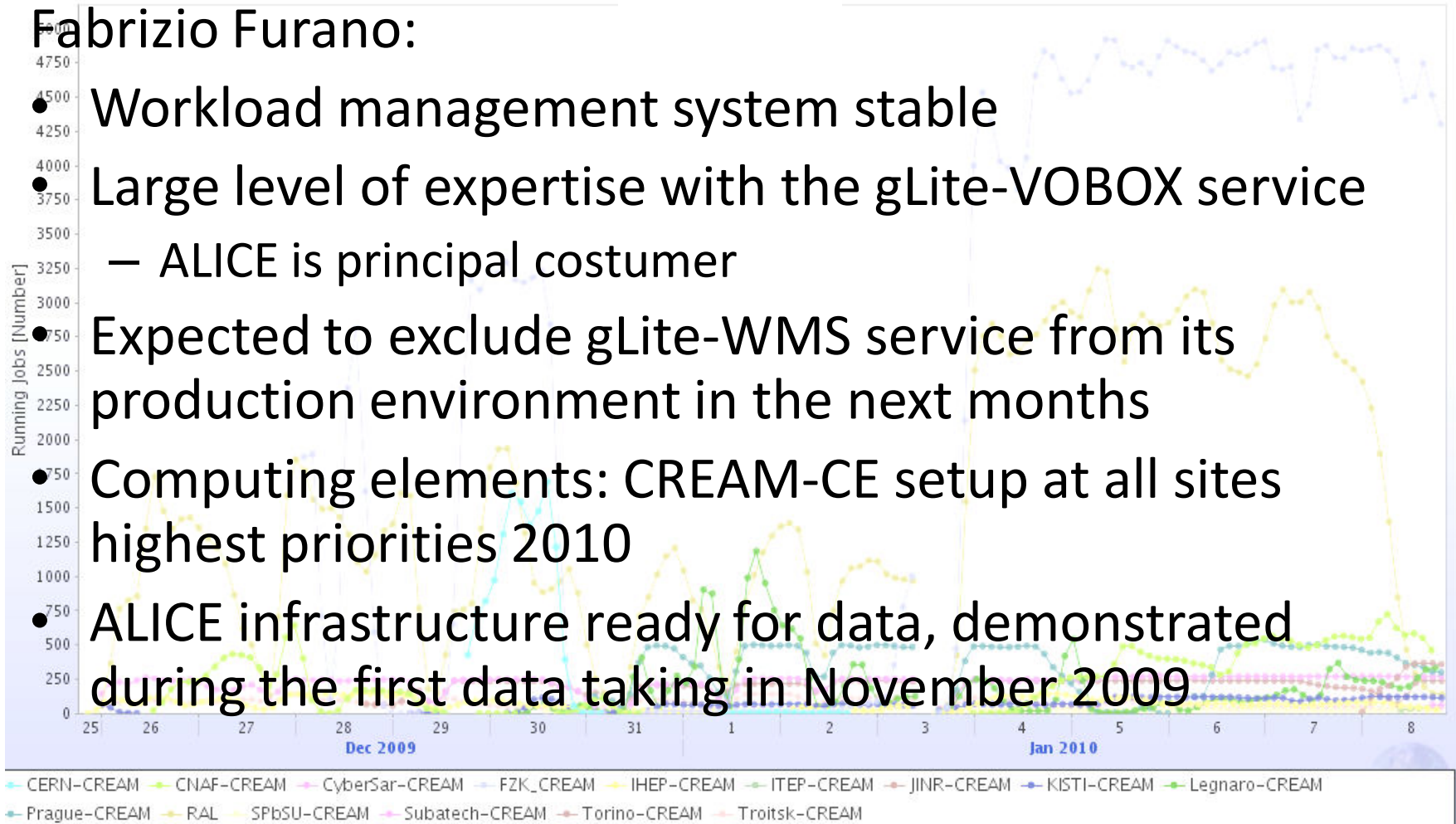
# EU-IndiaGrid2 - Sustainable e-infrastructures across Europe and India

- New high-bandwidth connection to India
- Connecting EU and India, also via Internet
- bandwidth there, how will it be used?

# New WLCG services and ALICE's AliEn Computing model\*

Fabrizio Furano:

- Workload management system stable
- Large level of expertise with the gLite-VOBOX service
  - ALICE is principal customer
- Expected to exclude gLite-WMS service from its production environment in the next months
- Computing elements: CREAM-CE setup at all sites highest priorities 2010
- ALICE infrastructure ready for data, demonstrated during the first data taking in November 2009



# Tools to use heterogeneous Grid schedulers and storage system

Mattia Cinquilli:

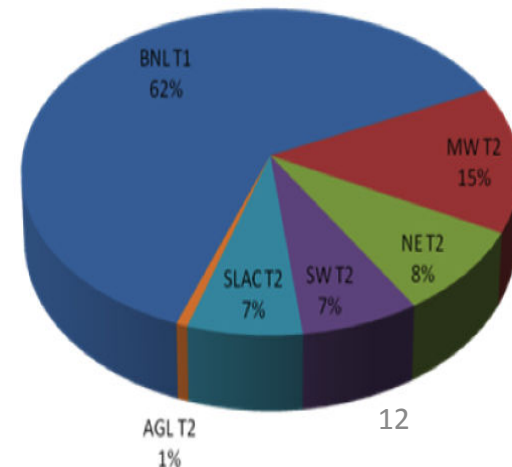
- CMS uses EGEE, OSG and NorduGrid, batch systems, CAFs; sites choose computational and storage solutions: inhomogeneous
- Need a standard interface
- BossLite handles interaction with different middleware/batch systems + logging facilities
- SEAPI: higher level API to storage protocols
- Results for of job efficiency, number of jobs and users in the case of CMS analysis applications

# BNL Batch and DataCarousel systems



Jerome Lauret:

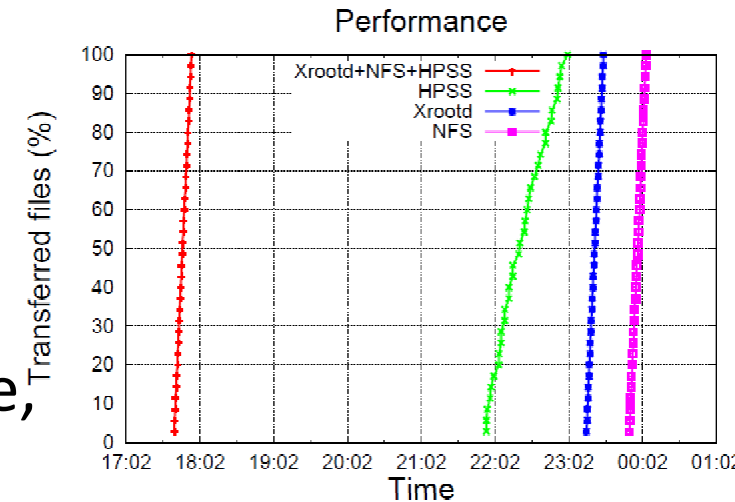
- Problem: data taken in time sequence, access to tape often stochastic
  - Potential for chaos, random and excessive overhead
- Solution:
  - ERADAT – Efficient Retrieval and Access to Data Archived on Tape (File retrieval scheduler)
  - DataCarousel – policy based for resource sharing
- Results
  - RHIC/STAR: 6 MB/sec → 46.2 MB/sec (LTO3)
  - US-Atlas ESD: 1 MB/sec → 72 MB/sec (LTO4)
  - DataCarousel: 1.21 mount / tape only
- Conclusions
  - Tape access optimization is crucial
  - Practical tools in use for RHIC and US-Atlas production
  - Code and knowledge shared with IN2P3



# Building Efficient Data Planner for Peta-scale Science

Michal Zerola:

- Functional Planner, database, web interface
  - extensive studies of performance, simulation
  - Installed in STAR, currently running tests
- Perspectives:
  - multi-site transfers, similar benefits expected
- Controlled and efficient data movement: higher efficiency, coordination, load-balancing



# Optimization of Grid Resources Utilization

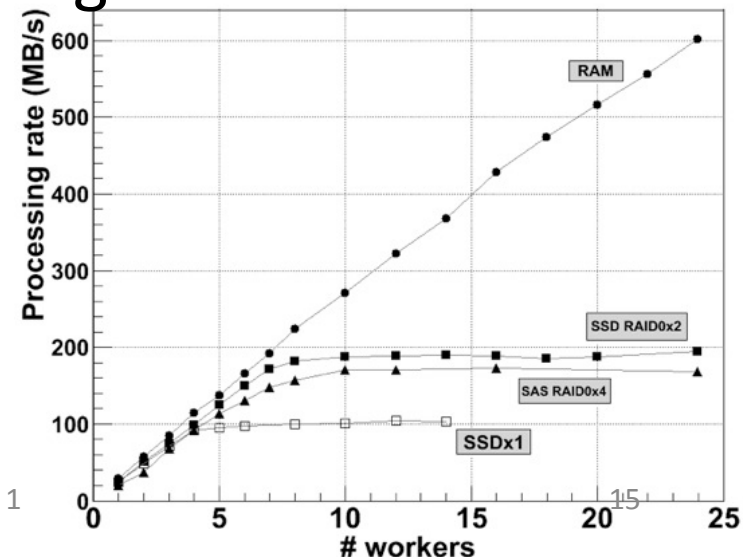
Costin Grigorias:

- Automation of storage operations by feeding back monitoring information in the decision-taking components of AliEn
- Flexible storage configuration
  - Addressing the storage elements by QoS tag only
  - Parallel storing of multiple replicas of the output
- Reliable and efficient file access
  - No more failed jobs due to auto discovery and failover in case of temporary problems
  - Use working storages closest to application

# PROOF - Best Practices

Fons Rademakers:

- Always fighting bottlenecks
  - file layout: much improved in ROOT 5.26
  - merge: multiple mergers
  - disks, network: comparing storage solutions
  - RAM
  - CPUs: PROOF lite
- Feeding 24 cores not trivial



# Optimizing CMS software to the CPU

- Settling on x86\_64, using move for performance review with good tools
- Memory cost of 64bit under control
- Allocation locality is an issue
- 64bit math has surprises
- Multi-core
  - this year: fork for shared data
  - then: “more fine grained”
  - deployment?





# "NoSQL" databases in CMS Data and Workflow Management

Andrew Melo:

- SQL schemes seen as too rigid
- Non-SQL DBs (CouchDB, Hadoop, ...) suitable for selected applications
- Commonly use map/reduce for query
- Successful implementation for logs etc

# Teaching a Compiler your Coding Rules

Axel Naumann:

- C++ 0x might require new generation of tools
- Need to control C+ 0x features
- LLVM as compiler toolset allows trivial implementations for some tools
- Example: coding rule checker

Clouds, GridLite, Cores, Tools, Data

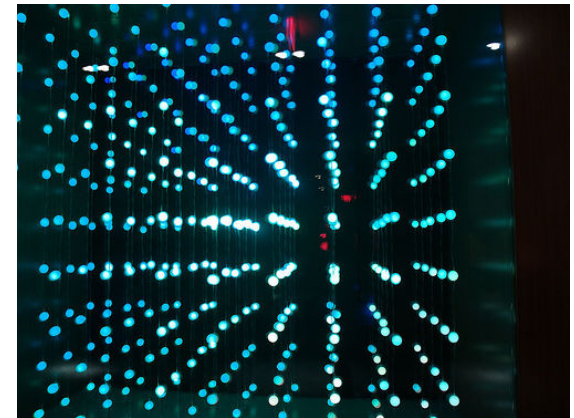
# **THEMES**

# Clouds

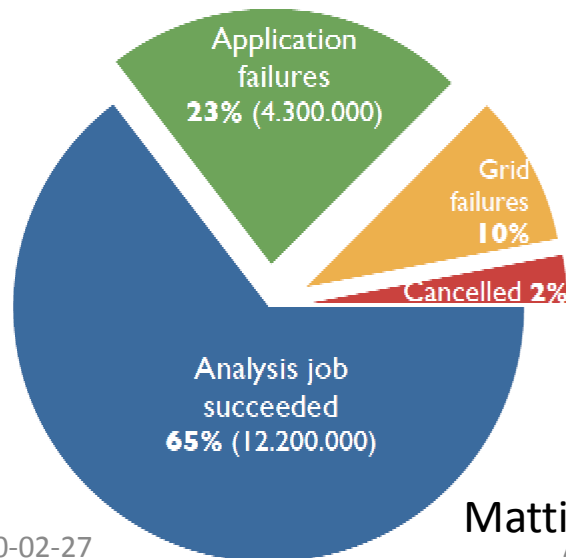


- From theory to practice to integration
  - Use cases still investigated
  - How to justify local batch vs. paying cloud
- Why Clouds == outsourcing?
  - Burst resource on peak demands?
  - Opportunistic resources truly at reach?
- Understand virtualization. Still open:
  - is it needed? do we want it?
  - is it transparent enough for end-users?
  - integrity, reproducibility?
  - image management, networking?

# GridLite

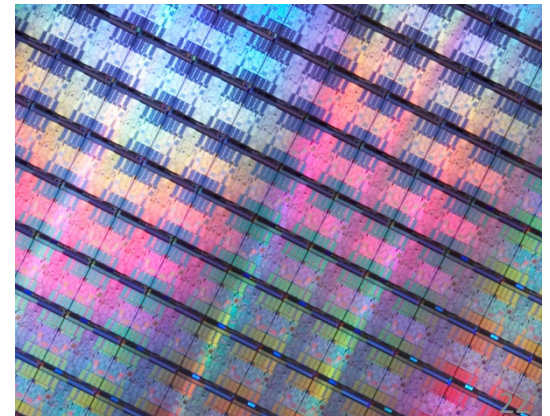


- Usability
- Management
- Fights are over, reality arrived
- 1/3 of jobs lost: inherent property of the grid?



# Multi- & Other-Cores

- x86\_64 understood, optimizing
- An occasion to optimize / re-think our code
- Finally again competition across chips designs / architectures!
- What will be the surviving architecture?
  - Lots of dynamic ...

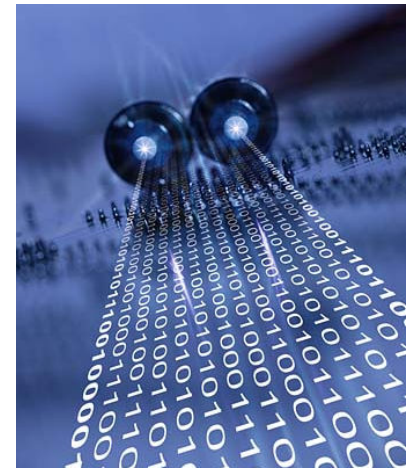


# Tools & Languages



- Many tools, only some mastered. Do we need new non-expert tools?
  - New compilers for parallelization
  - New tools for detecting language constructs / best practice / code analysis
  - Library / extension helping novice to develop parallel applications (OpenCL)
- Languages vs. optimizations / architectures: do we sell our souls? Standards!
- C++ 0x not a topic yet, unlike 64bit CPUs when they came out, Larrabee,...

# Data



- Equilibrium money - CPU - I/O: alchemy
- CPU focus swinging to I/O: did we overestimate relevance of CPU (interpreter in event loop)?
- Cache hierarchy: tape, disks, **SSDs**, RAM
  - All of the above?
  - Throw a WAN in?
  - Need algorithms for data lifetime across tiers
- Latency vs. cooperation / concurrency, theory!
- 24 core (“wow!”) but no data coming through!



# Conclusion

- We're done.
- NO: the race goes on. Complexity is fought over generalization
- Handling and analyzing data will remain challenging in every respect

# Thanks!

- Thanks to IAC for insight
- Sudhir for everlasting help
- All the chairs for creating wonderful workshop atmosphere
- Students etc for flawless technical support
- Speakers for a mosaic of present and future  
– and their help with this summary
- Let's see what the future brings to Brunel!