The Grid Observatory

www.grid-observatory.org

www.eu-egee.org

e-infrastructure

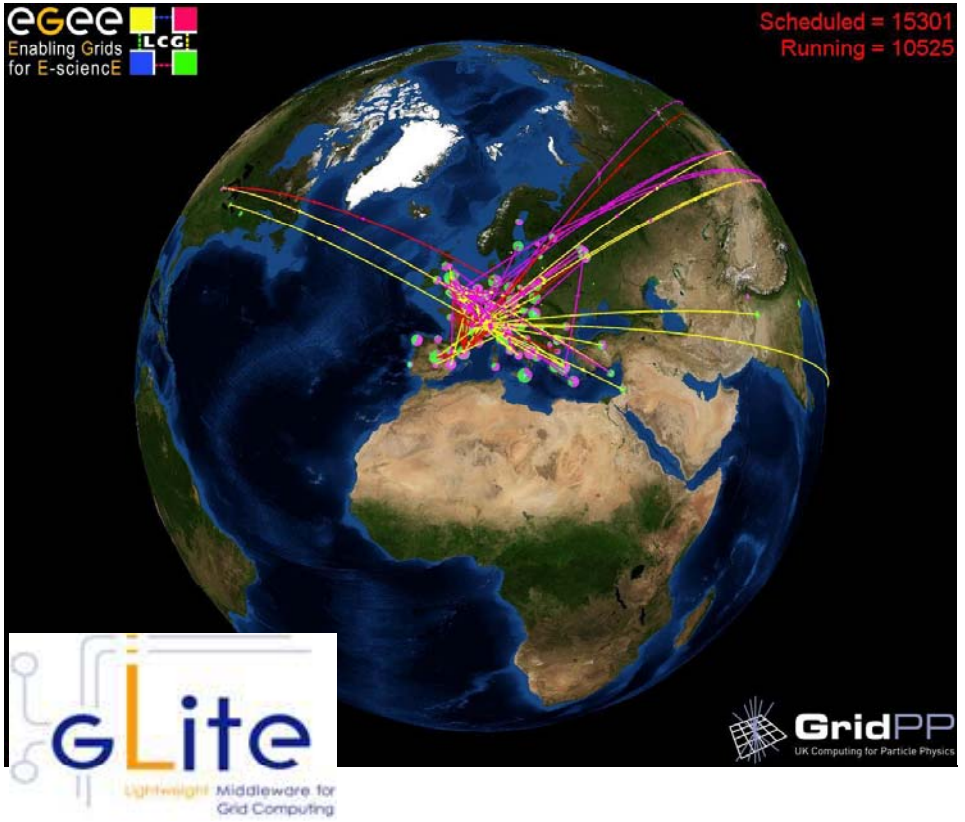
Grid Observatory Project

- ***Integrate the collection of data on the behaviour of the EGEE grid and users with the development of models and of an ontology for the domain knowledge***
- **Goals**
  - Data collection and publication
    - Public access to *a posteriori* traces of EGEE activity
  - Analysis
    - Propose realistic methods to answer real operational issues
  - Contribute to interactions with computer science research
- **EGEE Scientific Cluster (NA4)**
  - LRI (CNRS), Università Piemonte Orientale, London Imperial College, ASGC
  - With collaboration from Digiteo Labs, CNRS, LAL, MIS

EGEE Grid Infrastructure

Flagship grid infrastructure project funded by the European Commission

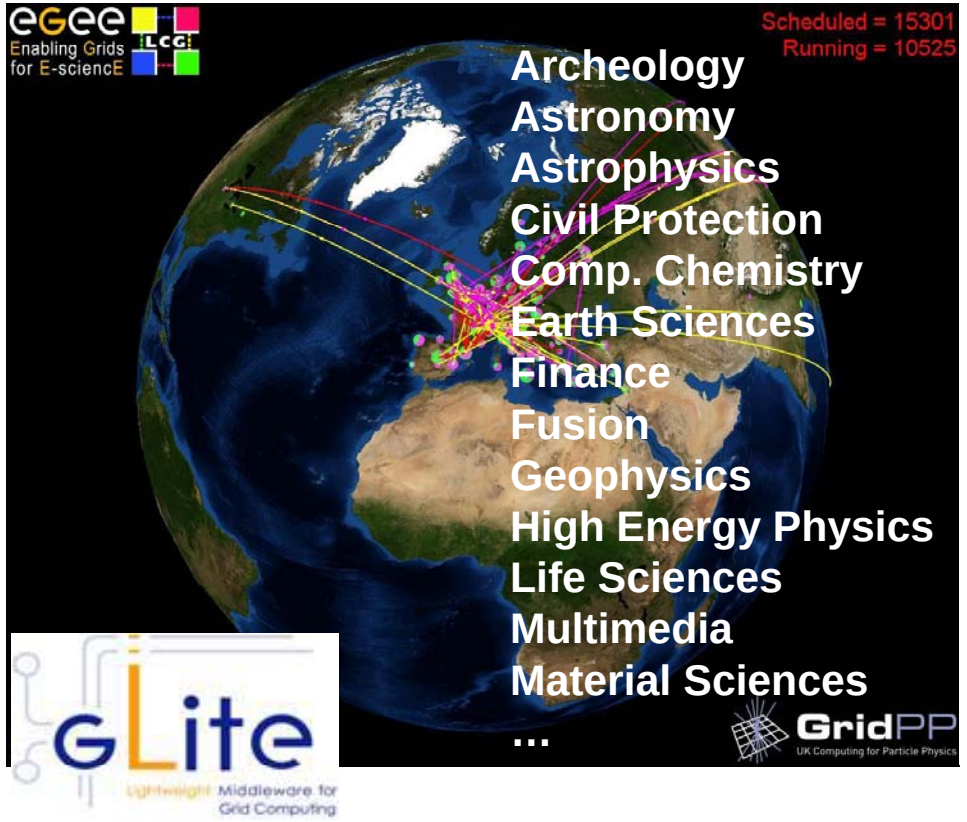
- **Operate a large-scale, production-quality grid**
- **300 sites, 140 partners, 50 countries**
- **80,000 cores, 5PB**
- **10,000 users**
- **300,000 jobs/day**
- **FP6 & FP7**



EGEE Grid Infrastructure

Flagship grid infrastructure project funded by the European Commission
The best approximation of the current needs of e-science

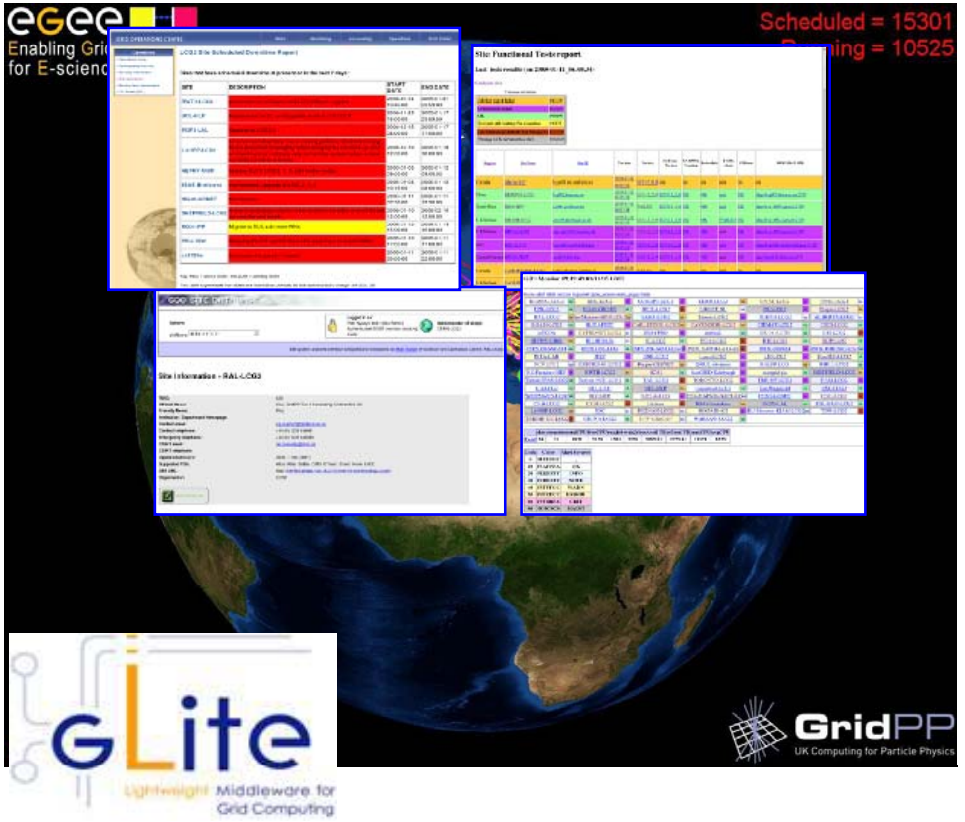
- **Operate a large-scale, production-quality grid**
- **300 sites, 140 partners, 50 countries**
- **80,000 cores, 5PB**
- **10,000 users**
- **300,000 jobs/day**
- **FP6 & FP7**



EGEE Grid Infrastructure

Flagship grid infrastructure project funded by the European Commission
The best approximation of the current needs of e-science
Extensive monitoring facilities

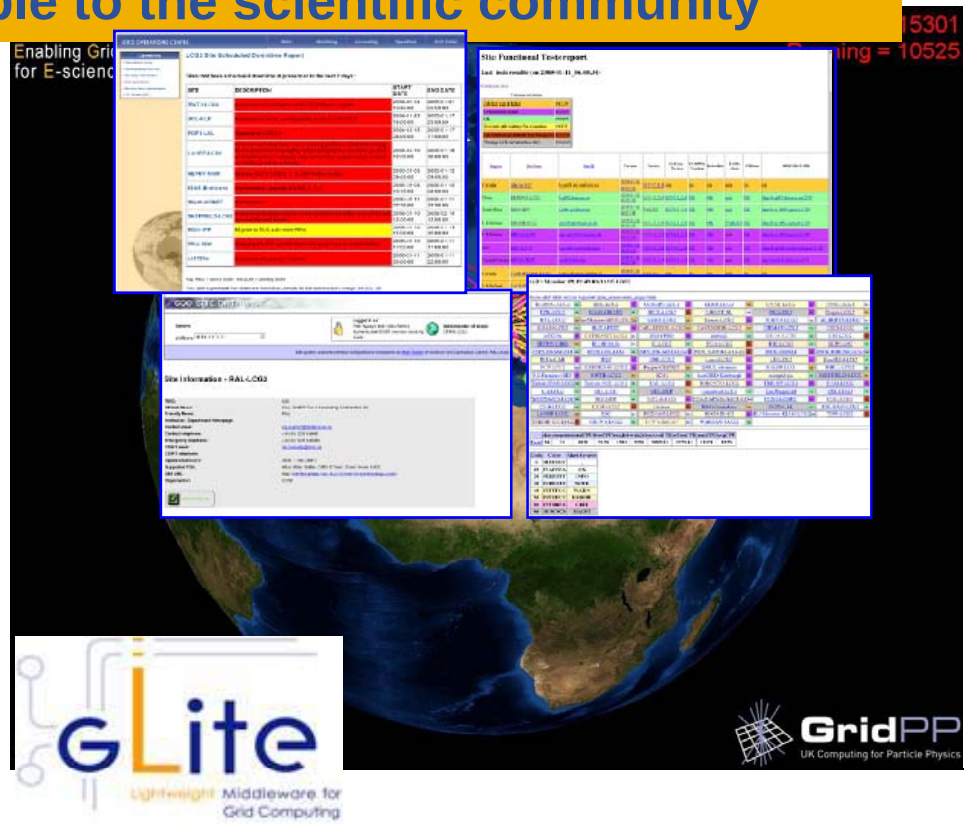
- CIC tools
  - GOCDB, SAM, SFT, …
- Core gLite
  - L&B, BDII, …
- Site Services
  - Maui/PBS logs
- gLite integrators
  - R-GMA, Job Provenance
- Experience integrators
  - DashBoard
- External software
  - MonaLisa



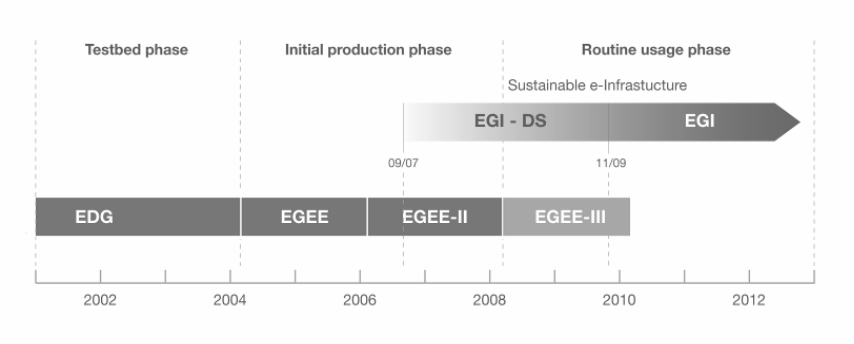
EGEE Grid Infrastructure

Flagship grid infrastructure project funded by the European Commission
The best approximation of the current needs of e-science
Extensive monitoring facilities
Data were discarded after operational usage,
and in any case not available to the scientific community

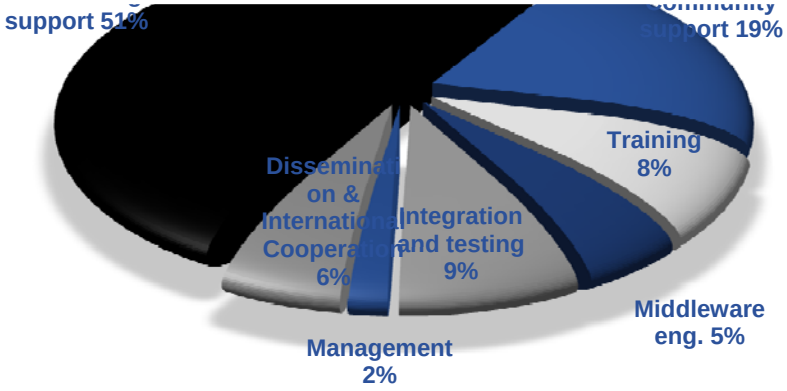
Image copyright © 2005-2008 David Opie



A grand challenge for production



How can we reduce the effort required to operate this expanding infrastructure?
Bob Jones' talk at EGEE'08



EC co-funding: 32 Million €

Some immediate questions

- **Resource allocation**
  - Performance of the gLite scheduling hierarchy
  - Responsive grids – Everybody's grid
  - Common goods – prevent abuse of the grid resources
- **Dimensioning**
  - Patterns and trends in requests and usage
  - Scalability of the information system
- **Dependability**
  - Detection: black holes
  - Diagnosis: disappearing jobs
  - Performance of a probe-free approach

From and to scientific research

- **Autonomic Computing: *"Computing systems that manage themselves in accordance with high-level objectives from humans".*** Kephart & Chess A vision of Autonomic Computing, IEEE Computer 2003
  - Self-*: configuration, optimization, healing, protection
  - On open non steady state systems
- **Statistical analysis, machine learning, data mining**
- **Promote a quantitative approach**
  - Optimize for the most frequent case
  - Compare solutions on defined datasets—benchmarks

Data Collection

- **Acquisition, consolidation, long-term conservation of traces of EGEE activities**
- **GO portal available since Oct. 2008**
  - www.grid-observatory.org
  - Heap of data “as is”
    - Get understanding of the complexities of the data
    - Start analysis and bootstrap interactions
  - Next step (September 09): filtered data

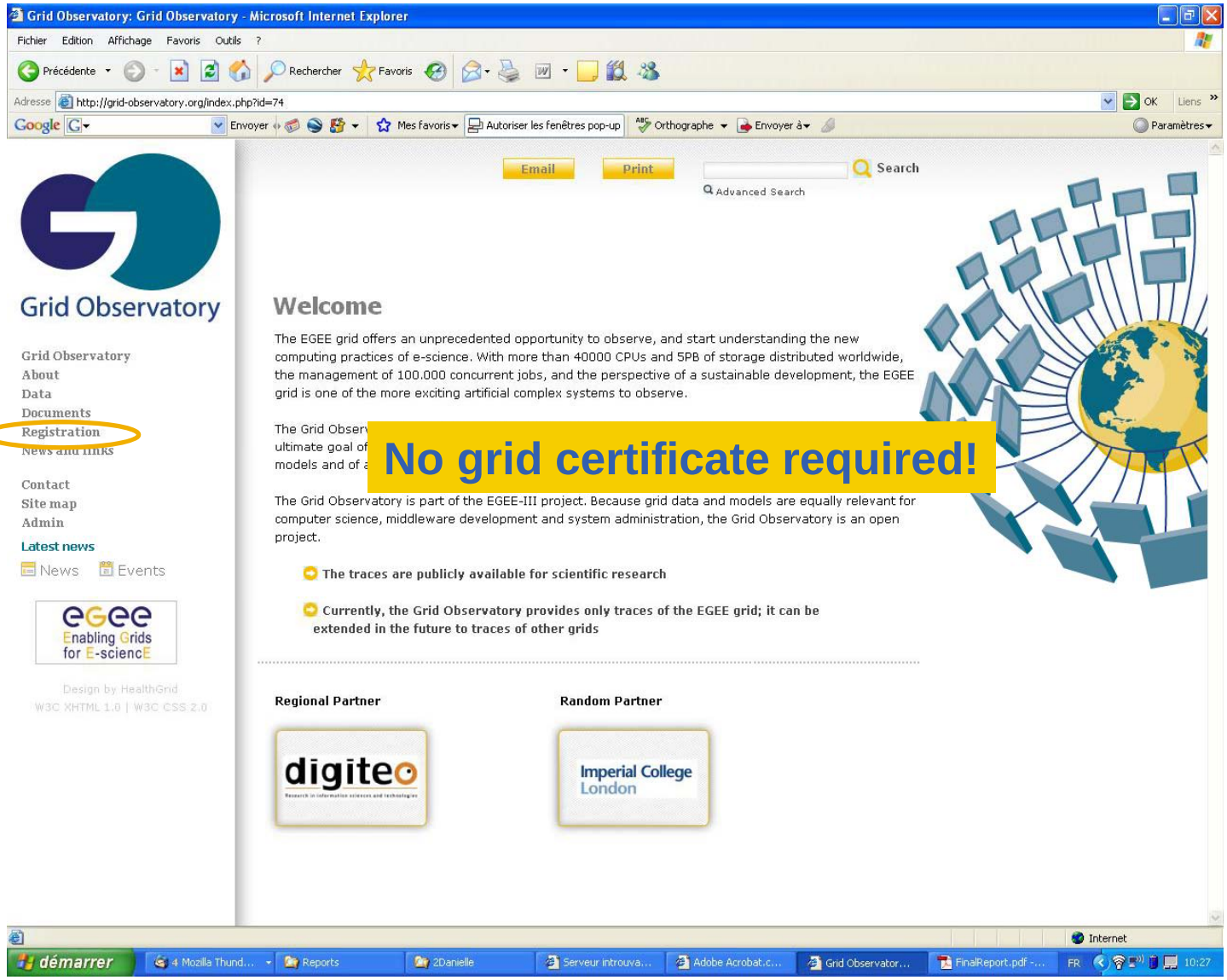
- **Long-term goal: permanent storage of reliable, exhaustive, filtered information**
  - Reliable: data curation and provenance
  - Exhaustive: Added value in snapshots of the inputs and grid state e.g. workload and available services during a relevant time range
  - Filtered: remove redundancy

Data

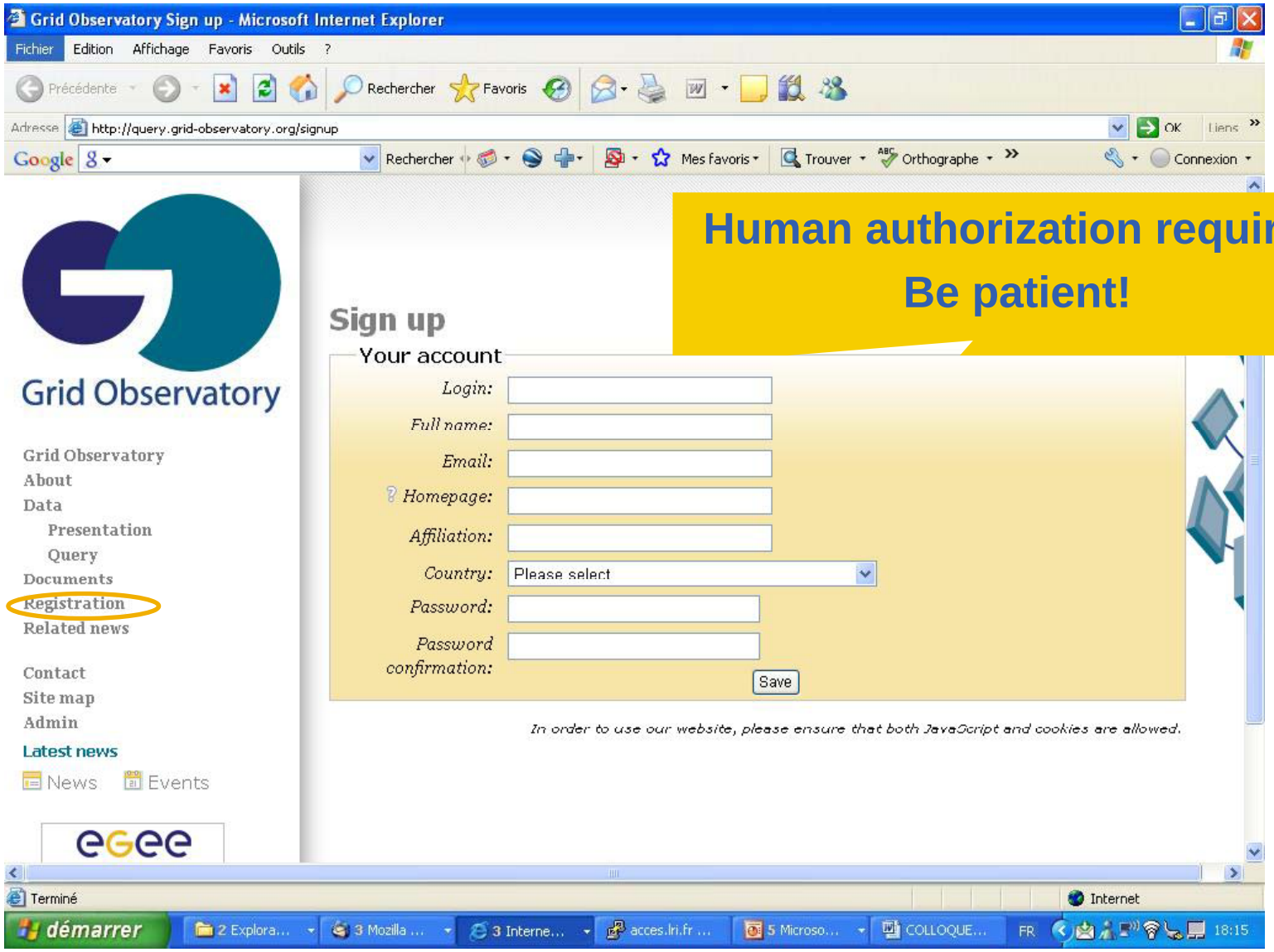
- **Information System (BDII)**
  - LDIF format, CE and SE information, limited information about services
  - One master file each day, plus diff each 15 minutes
- **GRIF (Grille de Recherche Ile de France)/LAL Site**
  - Logging and Bookkeeping service
    - ascii dump of the SQL tables events, short_fields and long_fields: all the events in the lifcycle of a job, tagged
  - Batch system(s)
    - Job controller traces allow to recover the grid identifier
  - WMS internals (condorG, wm_proxy, etc.)
    - Possible application: disappearing jobs
- **Real Time Monitor**
  - Summary of the lifecycle of jobs from the Real Time Monitor project

- **Next: Storage traffic – open issue: how to relate them to user files**
- **Outside the scope: external traffic on shared resources**

How to get an account?



How to get an account?



Grid Observatory Search for Traces - Windows Internet Explorer

http://query.grid-observatory.org/

Grid Observatory Search for Traces

Grid Observatory

- Grid Observatory
- About
- Data
  - Presentation
  - Query
- Documents
- Registration
- Related news

- Contact
- Site map
- Admin

Latest news

News Events

eGee Enabling Grids for E-sciencE

Design by HealthGrid
W3C XHTML 1.0 | W3C CSS 2.0

My Profile | Logout | Search

Search for Traces

Filters Hide

Site: GRIF-LAL

Service: Information System

From... To...: 2008-06-23 2008-09-09

Search

Documentation about the traces Display

Query results Hide

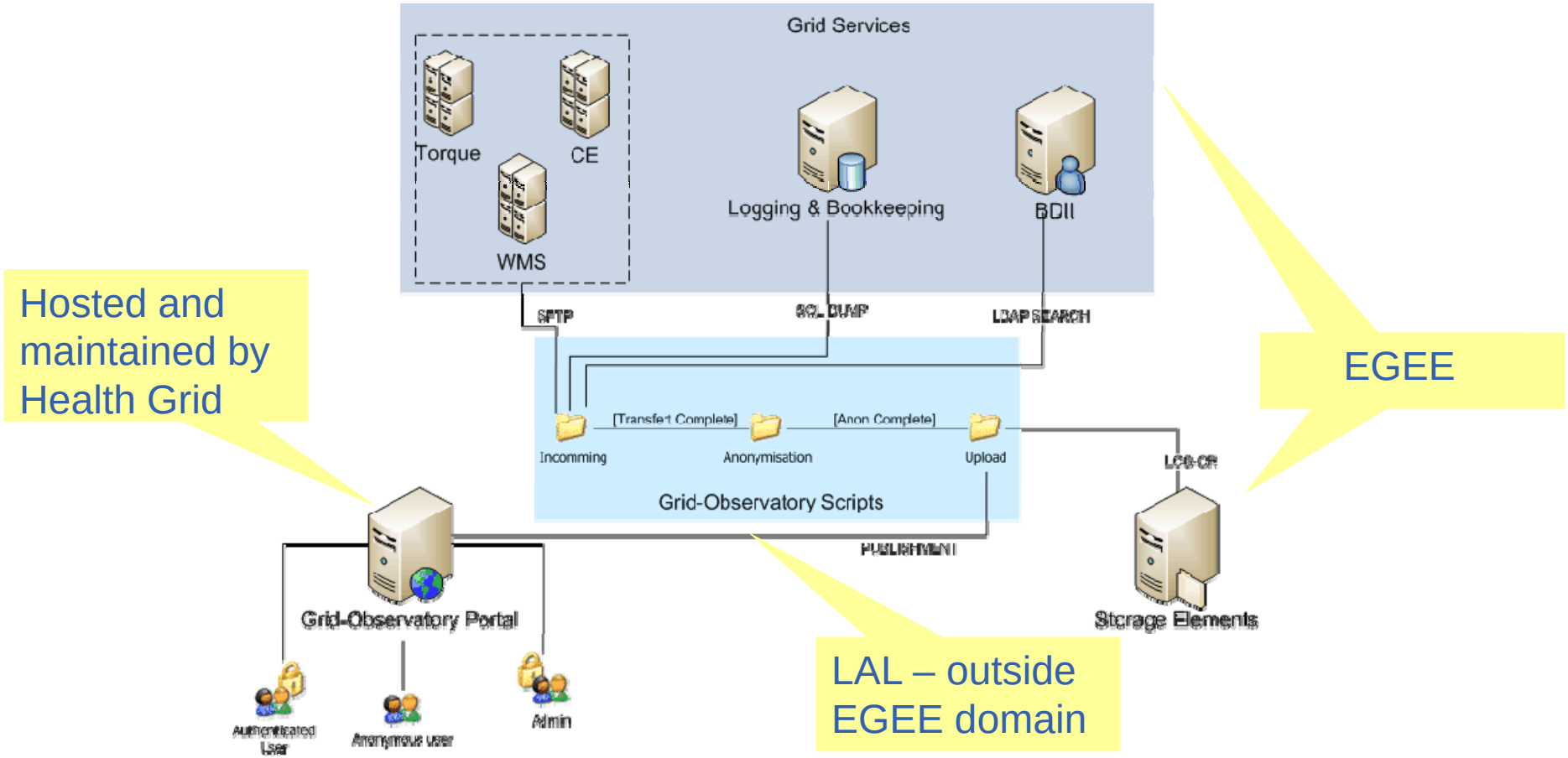
Affiliation	Service	Period	Download?
GRIF-LAL	Information System	2008-07-28 2008-08-03	Download (288 MB)
GRIF-LAL	Information System	2008-08-11 2008-08-17	Download (198 MB)
GRIF-LAL	Information System	2008-08-18 2008-08-24	Download (295 MB)
GRIF-LAL	Information System	2008-08-25 2008-08-31	Download (306 MB)
GRIF-LAL	Information System	2008-09-01 2008-09-07	Download (408 MB)
GRIF-LAL	Information System	2008-09-08 2008-09-14	Download (299 MB)

Displaying **all 6** traces

In order to use our website, please ensure that both JavaScript and cookies are allowed.

démarrer | Gestionnaire des tâch... | Connexion réseau sa... | Logos | Mulberry (Connected) | Grid Observatory sta... | Grid Observatory Sea... | FR | 22:56

Architecture

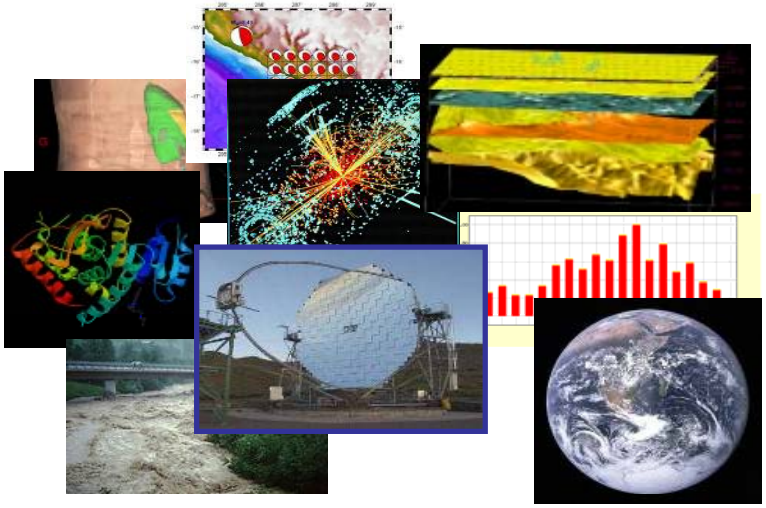


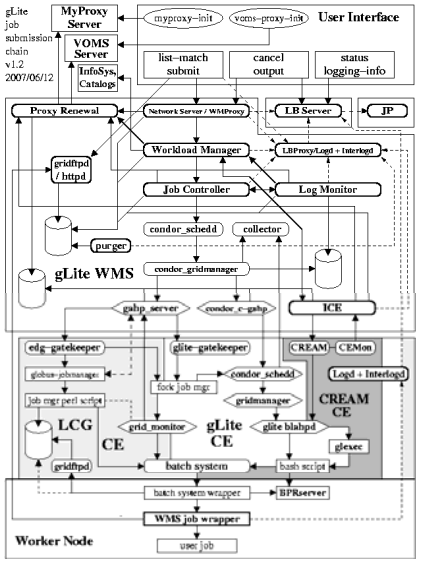
Towards a knowledge base

- **Execution trace formats (à la Grid Workload Format)**
- **Event trace models (à la IBM Common Event Base)**
  - Automatic ontology construction tools may help for dealing with undocumented logs e.g. automatic conversion to IBM CBE of the cryptic logs of the WMS
- **Ontology**
  - The Glue Information Model is an ontology of the resources
  - More concepts
    - Concepts for the grid dynamics e.g. job lifecycle or users relations
    - Expert concepts as prior knowledge of non-trivial correlations: workflows, failure modes,…
    - Concepts for elementary analysis
  - Validated models will become concepts
  - More semantics: relations between concepts and typing to allow logical inferences. Generic application: Data curation and consolidation, e.g. RTM + BDII

A complex system

- **Coupled usage: Virtual Organization**
  - Community software, community activity
  - Access rights
- **Feedback loops in the middleware**
  - Job dispatch
- **Emerging policies**
  - As the result of sites and stakeholders decisions
- **Inference of models for middleware components and applications, users and usage profiles, users interactions**





Time-series analysis - CE level

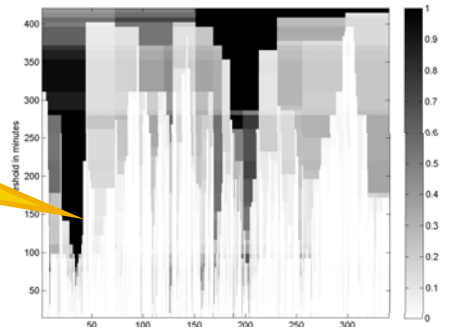
- **Evidence of self-similarity and heavy-tailedness in the arrival process**
- **Simple methods for realistic workload prediction**
  - Linear from the load history
  - As the mean of the past executions x number of jobs in the queue

  are defeated by very long jobs
- **Correlated by good fits of GARCH distributions**

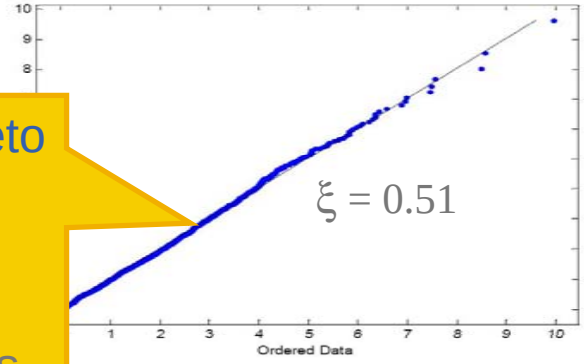
More: http://indico.lal.in2p3.fr/conferenceDisplay.py?confId=443

- **Ongoing work**
  - Systematic AR segmentation by Minimal Description Length
  - Evaluation of the gLite Expected Response Time computation

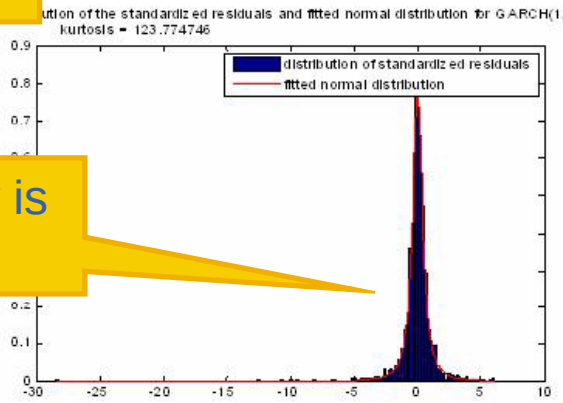
Burst of arrivals, at all scales



Generalized Pareto for inter-arrival times, $\xi > 0$ indicates stong heavy-tailedness

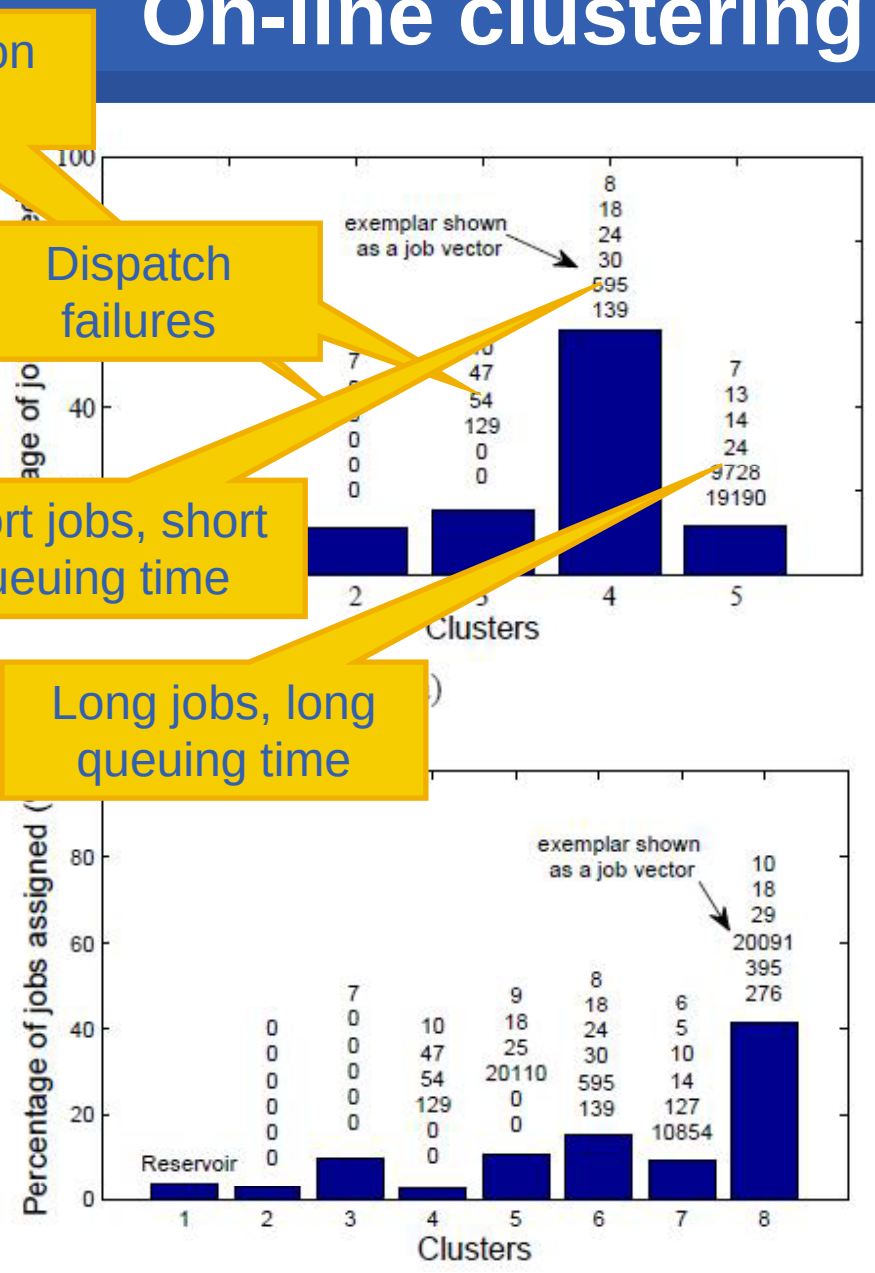


Only the volatility is predictable



On-line clustering

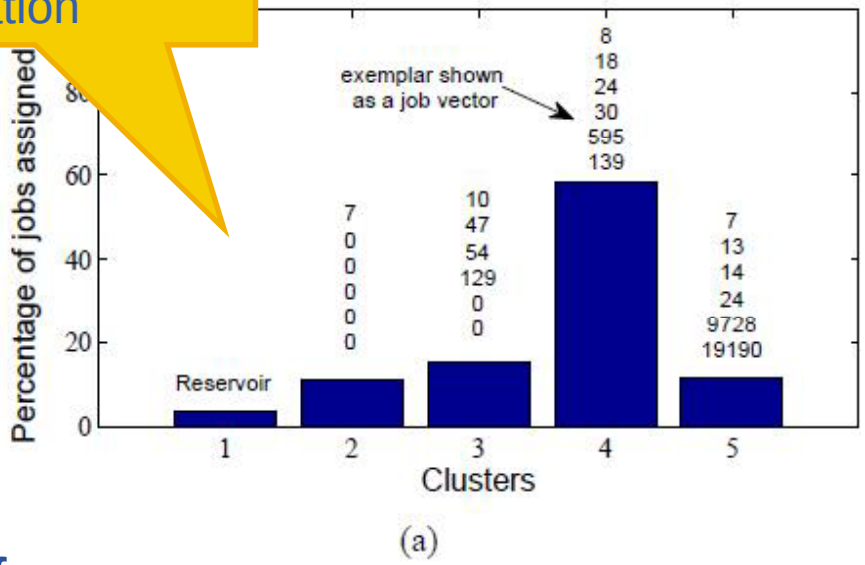
- **Clustering the jobs timestamps by StrAP, an on-line version of the Affinity Propagation algorithm**
- **Clusters described by real jobs, the exemplars**
- **Complexity compatible with line analysis**
- **Provides a concise description of the system**



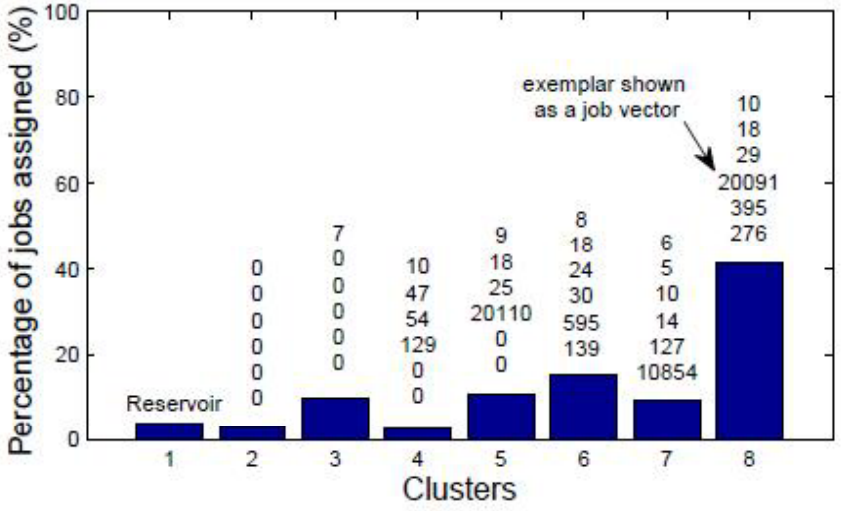
On-line clustering

Overall, a normal situation

- Clustering the jobs tim… by StrAP, an on-line version of the Affinity Propagation algorithm
- Clusters described by real jobs, the exemplars
- Complexity compatible with on-line analysis
- Provides a concise description of the system

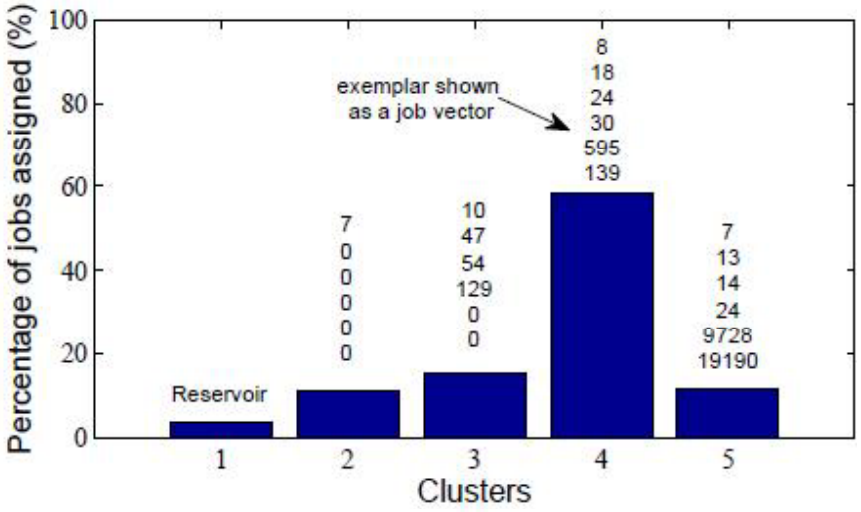


(a)

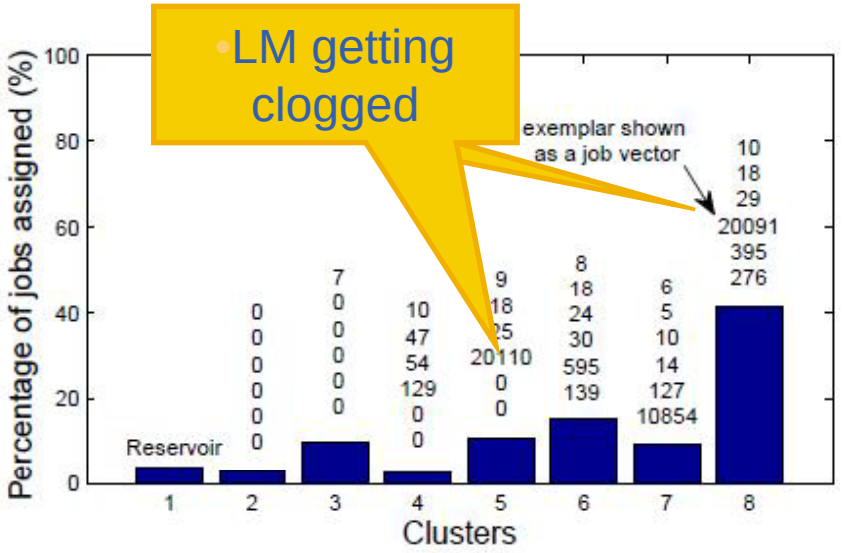


On-line clustering

- **Clustering the jobs timestamps by StrAP, an on-line version of the Affinity Propagation algorithm**
- **Clusters described by real jobs, the exemplars**
- **Complexity compatible with on-line analysis**
- **Provides a concise description of the system, and exhibits dangerous evolutions**
- **More:**
  - Poster # 36
  - X.Zhang, M. Sebag, C. Germain-Renaud, "Multi-scale Realtime Grid Monitoring with Job Stream Mining". To appear (CCGrid'09),



(a)

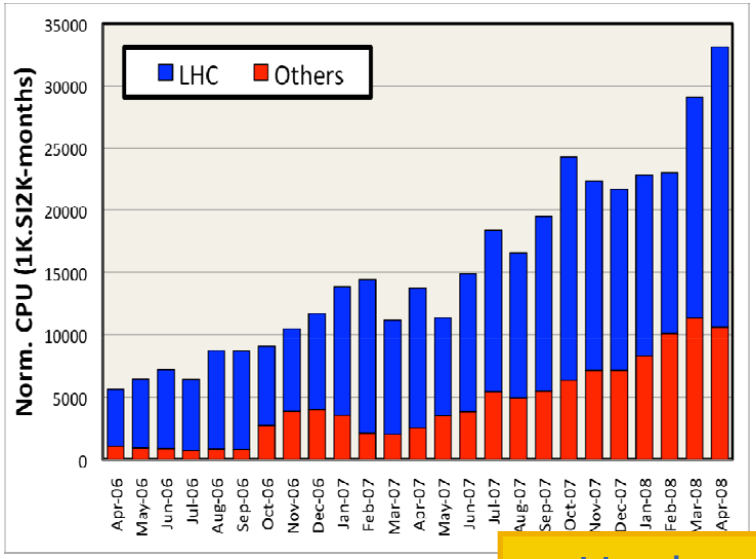


And more…

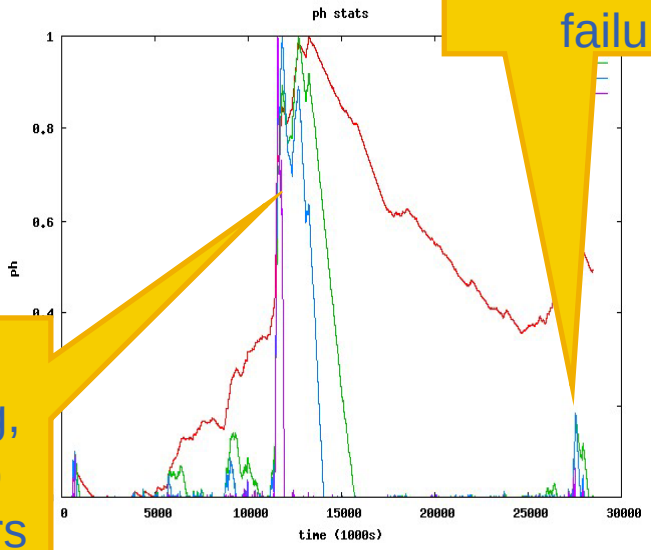
- **Correlation of Estimated Response Time, from VOViews objects in the BDII, with actual response time, from job traces**
  - https://savannah.cern.ch/task/?8589
  - Already: detection of anomalous behaviour - incredibly high ERTs
- **Elastic responsive computing through Reinforcement Learning [CCGRID'08, ICAC'08, GMAC'09?]**
  - Principled dynamic prioritization of urgent jobs
  - Integrates fair-share (VOs) and utilization
  - Possible applications: local scheduler, systems of pilot jobs
- **On-line black hole detection**
  - Page-Hinkley Statistics

Issues

- **Curse of dimensionality: representation and computational complexity**
  - e.g. we needed to devise a new version of the Affinity Propagation method
- **Sparse data: largely unexplored state/action space**
  - Each problem requires to build a specific simulator
- **No steady-state: still in expanding phase**
  - Quantitative and qualitative (e.g. extension of pilot jobs) trends
- **Data acquisition at full grid scale**
  - Deployment of the acquisition suite will be a long process
- **Validation: lack of expert interpretation**
  - What is a blackhole?



Hardware failure



Experience software bug, irrelevant to site managers

Computer Science Collaboration

- **GMAC'09: Grids Meets Autonomic Computing**
  - Barcelona, Spain, 15 June 2009
  - Associated with the 6th International Conference on Autonomic Computing (ICAC'09)
  - http://www.frombarcelona.org/GRIDmeetsAC/

- **Workshop Focus**
  - Identify key scientific challenges related to the management and evolution of grids as a specific category of complex large-scale systems. The goal of the workshop is to promote community wide discussion of, and collaboration on potentially high-impact ideas that will influence and foster continued research in improving the manageability and reliability of grids.