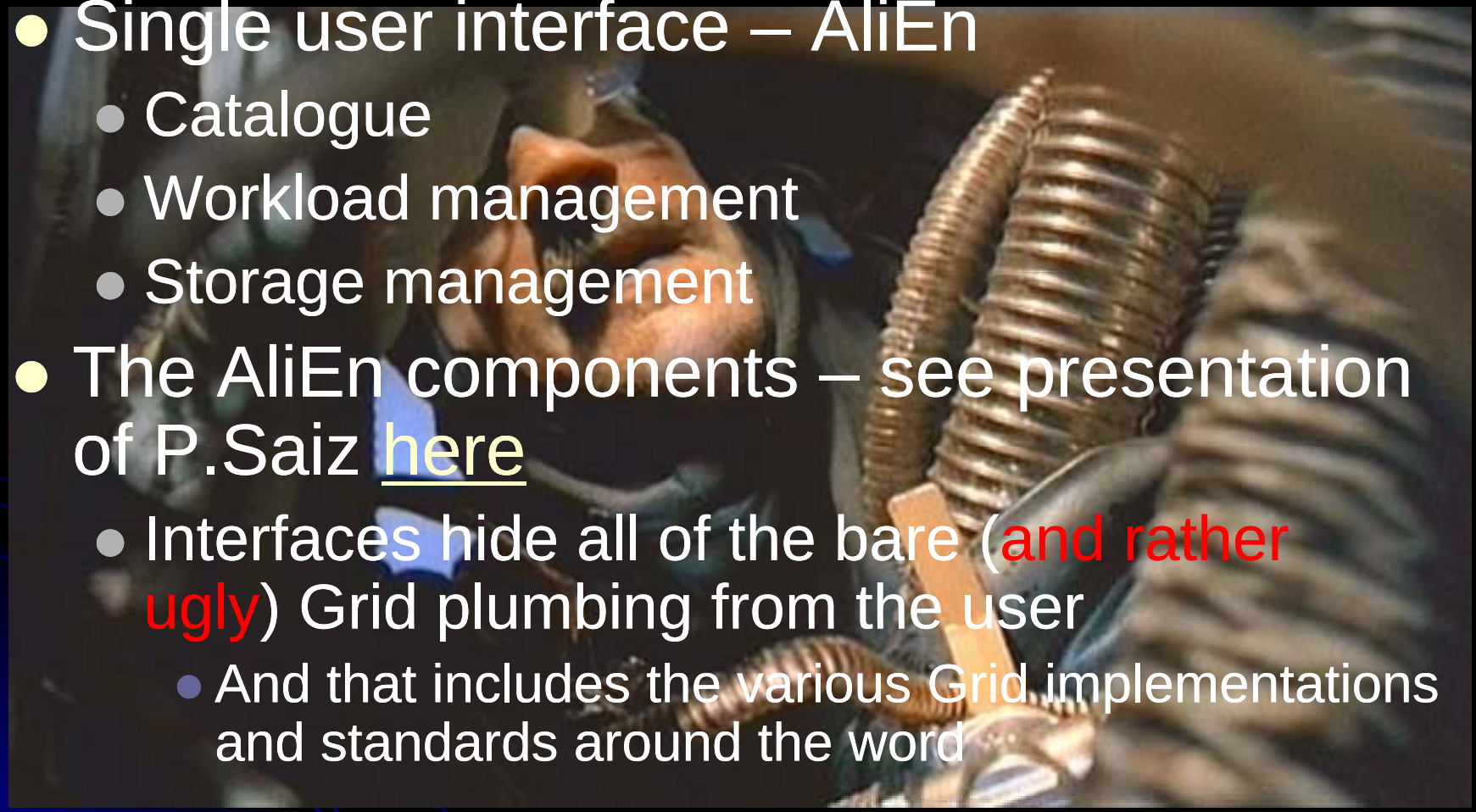
# The ALICE Grid

ALICE-FAIR Computing meeting

29 April 2008

Latchezar Betev

# The ALICE Grid basics



- Single user interface – AliEn
  - Catalogue
  - Workload management
  - Storage management
- The AliEn components – see presentation of P.Saiz here
  - Interfaces hide all of the bare (and rather ugly) Grid plumbing from the user
    - And that includes the various Grid implementations and standards around the word

# The ALICE Grid in numbers

- 65 participating sites
  - 1 T0 (CERN/Switzerland)
  - 6 T1s (France, Germany, Italy, The Netherlands, Nordic DataGrid Facility, UK)
  - 58 T2s spread over 4 continents
    - T2s in Germany - GSI and Muenster
- As of today  the ALICE share is some 7000 (out of ~30000 total) CPUs and 1.5 PB of distributed storage
- In ½ year ~15K CPUs, x2 storage

# The ALICE Grid history

- First AliEn prototype in 2002
  - Vertical (full) Grid implementation
  - Some 15 sites, MC production, storage at a single site
- 2003-2005 – development of various Grid interfaces for AliEn
  - Horizontal (central services + site services) implementation
  - Some 30 sites, MC production, storage (still) at a single site
    - There were interfaces to many (raw) storage systems, but no single client library support

# The ALICE Grid history (2)

- 2006-2008 – refinement of services, Grid sites buildup, AliEn Catalogue updates, xrootd as a single supported I/O protocol, user analysis
  - Majority of sites integrated (4-6 more expected)
  - Standard high-volume MC production
  - Central services in full production regime
  - Rapid deployment (and use) of storage with xrootd support
    - Standard LCG SEs (DPM, CASTOR2, dCache) and xrootd as pool manager
  - User analysis on the Grid
    - Not as bad as everyone expected ☺

# The ALICE Grid Map

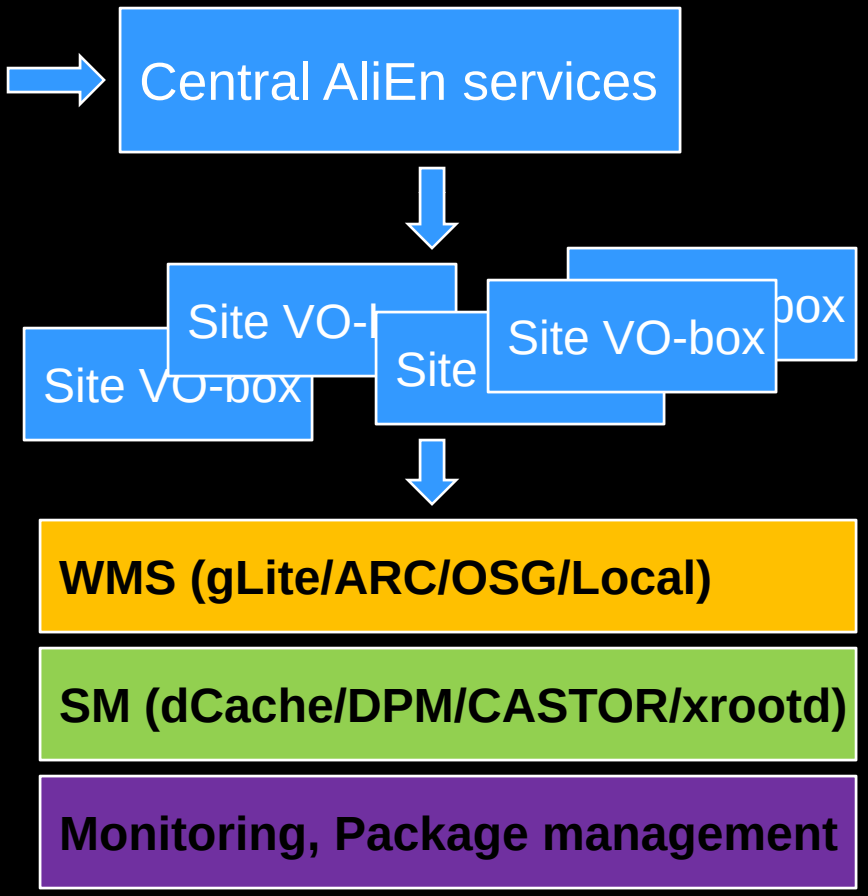

Here is the live picture

# Operation

- All sites provide resources through the WLCG gateways …or directly
- And software: gLite (EGEE), OSG (US), ARC (NDGF), local batch systems
- A (very short) hint of the existing complexity (only for workload management)
  - gLite: edg-job-submit, glite-wms-job-submit (…RB, CE, cancel, check, etc…)
  - ARC: ngsub (…cluster, type, etc…)
  - Local: bsub, qsub (well known to many)
  - OSG: globus-job-run (…cluster, proxy type, etc…)
  - All of the above is replaced by AliEn 'submit'

# Operation (2)

- Services schema


The VO-box system (very controversial in the beginning)
- Has been extensively tested
- Allows for site services scaling
- Is a simple isolation layer for the VO in case of troubles

# Operation – central/site support

- Central services support (2 FTEs equivalent)
  - There are no experts which do exclusively support – there are 7 highly-qualified  experts doing development/support
- Site services support - handled by ‘regional experts’ (one per country) in collaboration with local cluster administrators
  - ***Extremely important part of the system***
  - In normal operation ~0.2FTEs/regions
- Regular discussions with everybody and active all-activities mailing lists

# Operation – critical elements (2)

- Central services, VO-boxes, storage servers – capable of running 24/7, maintenance free
  - ***Get the best hardware money can buy***
- Multiple service instances in failover configuration
  - AliEn services are enabled for this
  - Use of ‘load balanced’ DNS aliases for service endpoints
    - Load-balancer is external to the system

# Operation – critical elements (3)

- Monitoring
  - Fast and detailed – AliEn command line interface
  - History – MonALISA
  - There is never enough monitoring, but if not careful, it can saturate the system (and the expert)
- Automatic tools for production and data management
  - Lightweight Production Manager – for MC and RAW data production
  - Lightweight Transfer Manager – for data replication (sparse storage resources)

# Central services setup

Linux 32, 64 bit build servers

*MacOS build server*

*MonALISA repository*

*3TB xrootd disk servers*
*Application software*
*Conditions data*
*User files*

*AliEn services*
*Proxy, Authen, JobBroker, JobOptimizer, TransferOptimizer, etc..*

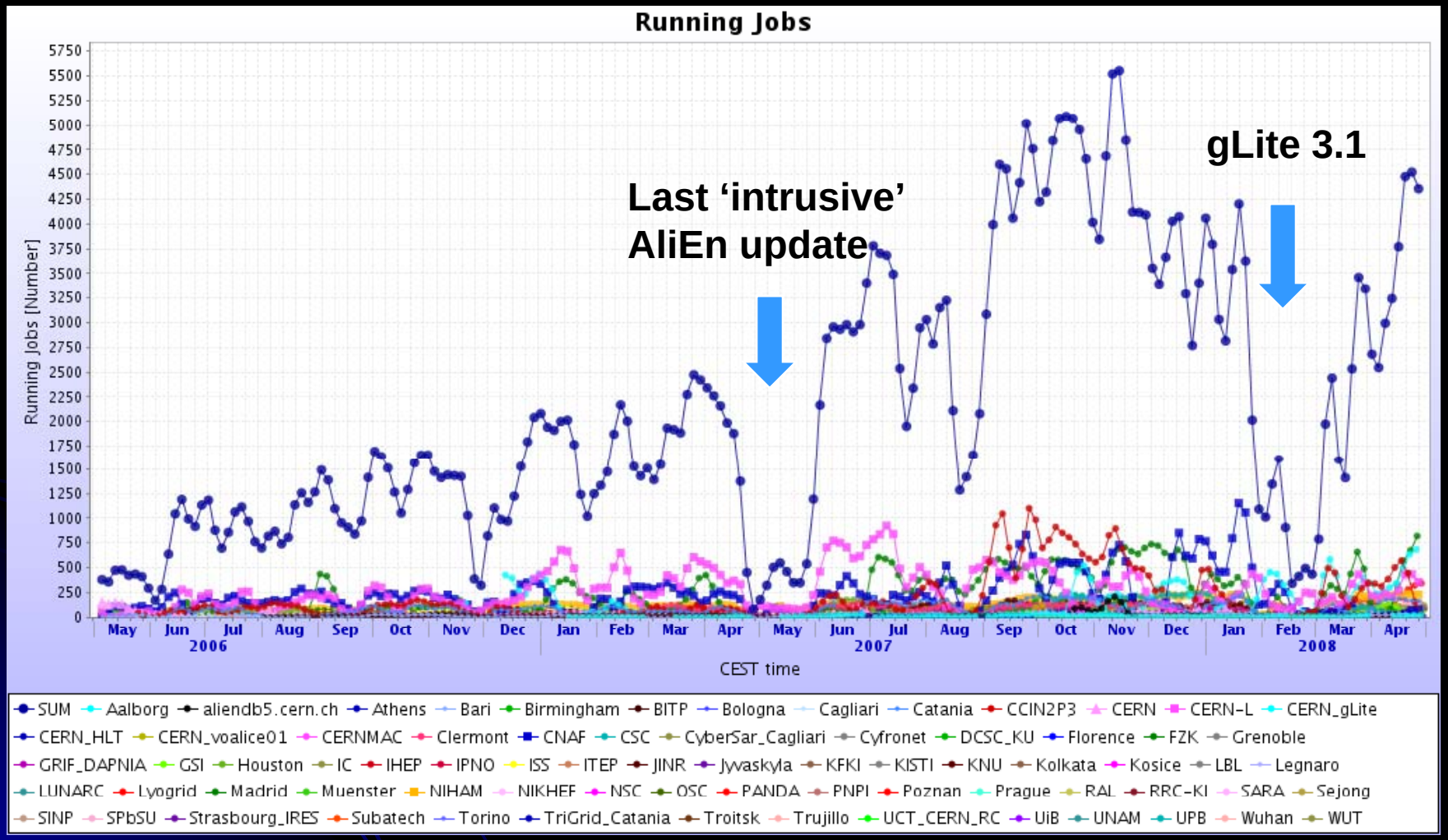
*MySQL DB (replicated) for AliEn Catalogue Task Queue*

*APIServers*

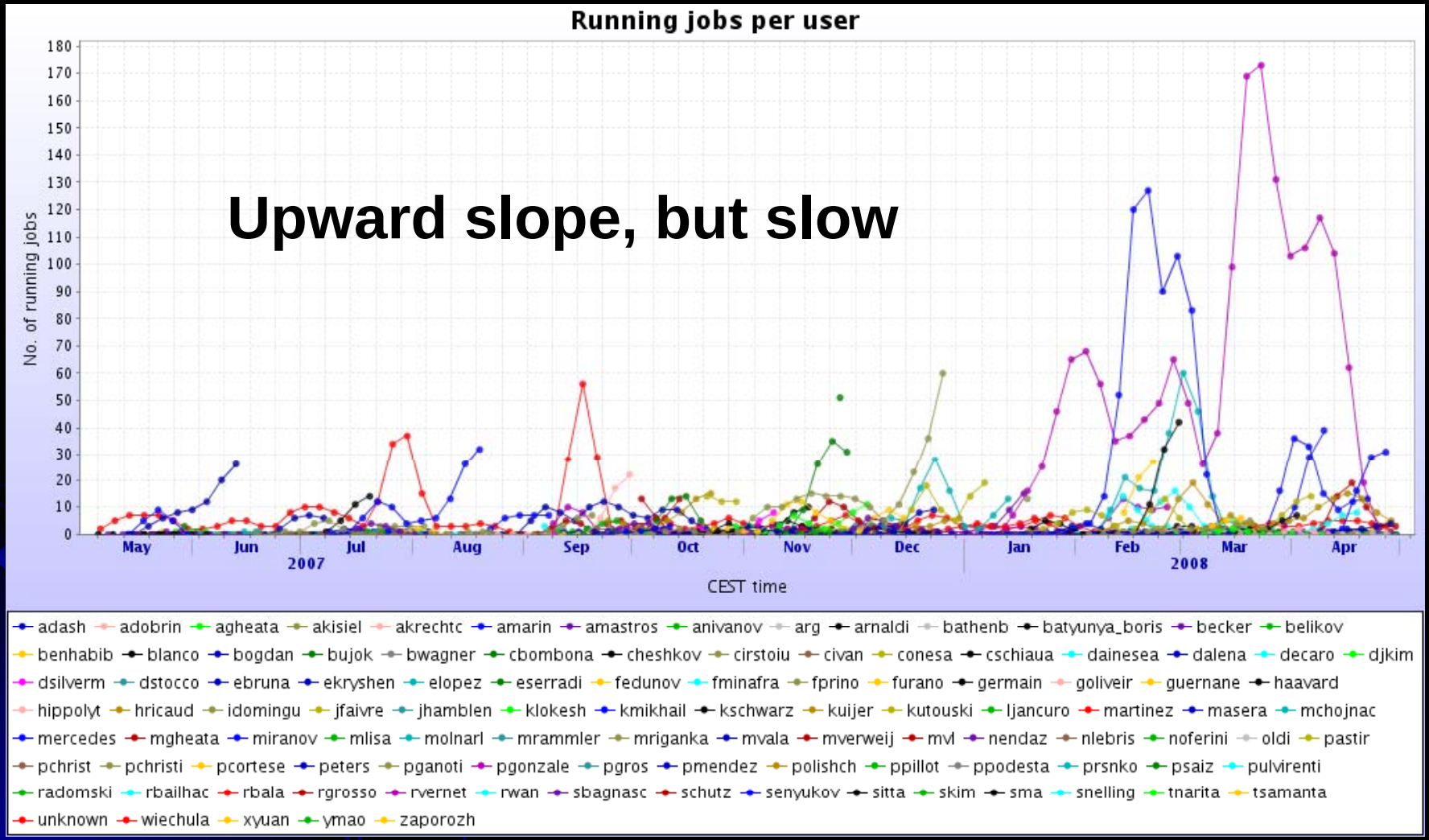
*Alien.cern.ch*

# Central services upgrade

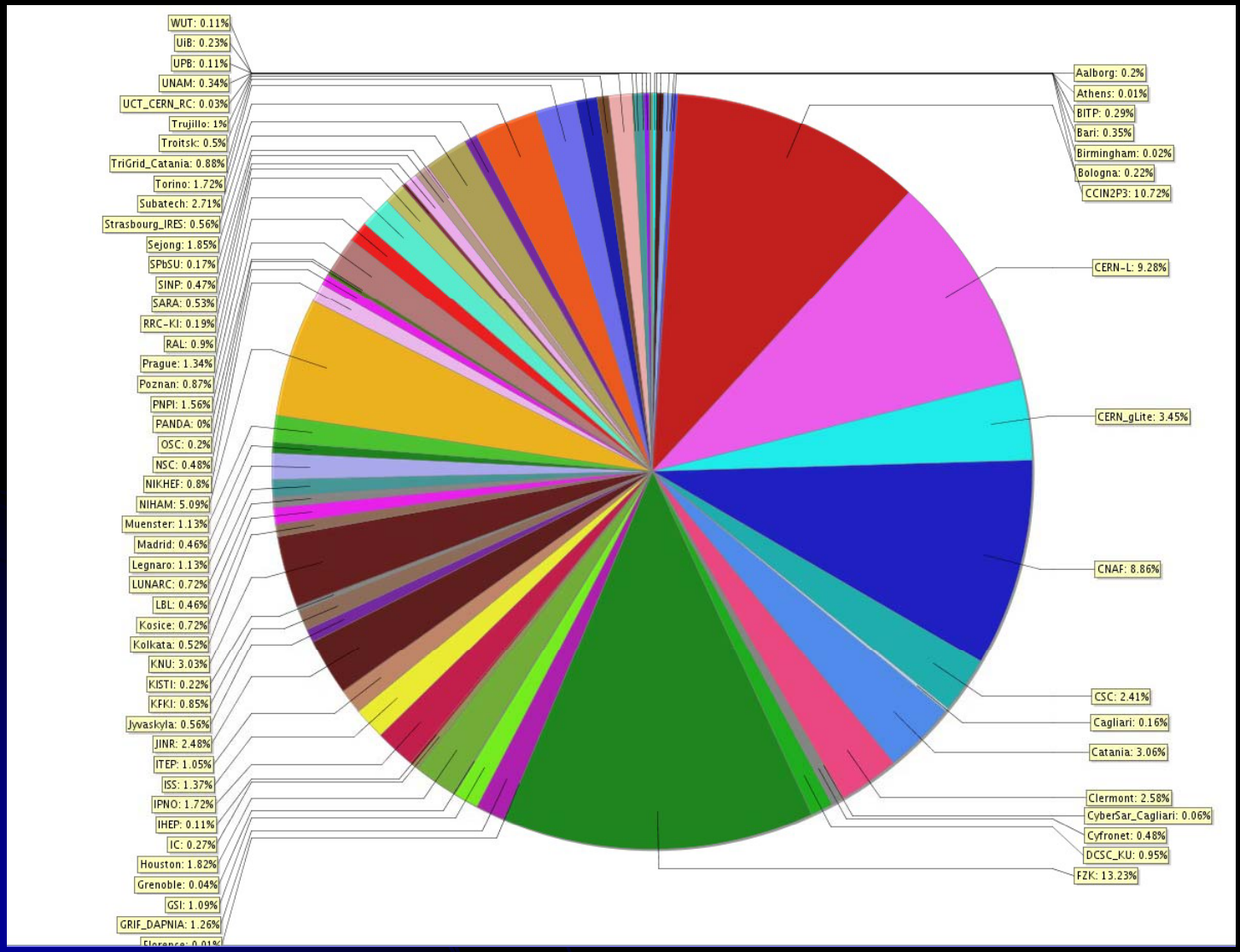
# Running profile – all



# Running profile – users



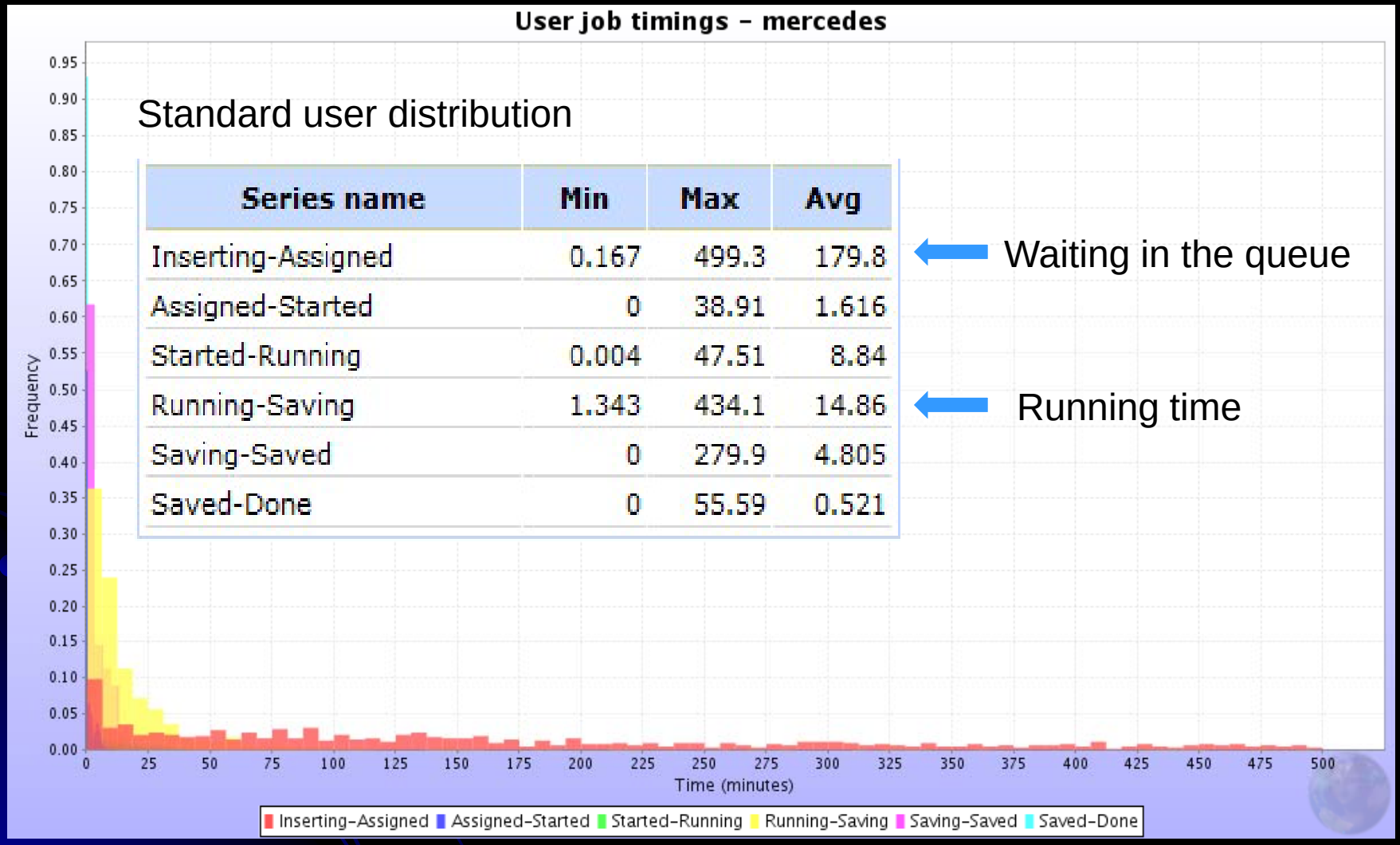
# Sites contribution




**50% resources contribution from T2s!**

**Harnessing the power of small computing centres is a must**

# GRID waiting times




| Series name | Min | Max | Avg |
|---|---|---|---|
| Inserting-Assigned | 0.167 | 499.3 | 179.8 |
| Assigned-Started | 0 | 38.91 | 1.616 |
| Started-Running | 0.004 | 47.51 | 8.84 |
| Running-Saving | 1.343 | 434.1 | 14.86 |
| Saving-Saved | 0 | 279.9 | 4.805 |
| Saved-Done | 0 | 55.59 | 0.521 |

← Waiting in the queue

← Running time

# Current activity in ALICE

- Substantial part of the resources is used to produce Monte-Carlo data for physics and detector studies
  - Increasing number of physicists are using the Grid in their daily work
- Since December 2007 the ALICE detector is being commissioned with cosmic ray trigger
  - Reconstruction and analysis of this data is ongoing
- Ramping up of CPU and storage capacity in preparation for the LHC startup – expected in summer 2008

# Summary

- The ALICE Grid is getting ready for the LHC data production and analysis
  - It took 6 ‘short’ years to get there
- The main system components have been battle-hardened
  - Development and simultaneous heavy use
- The ‘single interface to the Grid’ is a must
  - Otherwise the Grid will be limited to ‘selected few’
- Integration of computing resources into a coherent working system takes a lot of time an effort
  - This is a combined effort between the site and Grid experts, which has to be repeated n times (n=number of sites)
  - A trust relation depends not only on the high quality of the Grid software

# Summary (2)

- It is never too early to start user analysis on the Grid
  - It takes time to ‘convert’ user from local to global reality
    - And the conversion is not without pain
  - The experts should have time to learn how to do Grid user support