WLCG – Worldwide LHC Computing Grid

Where we are now & the Challenges of Real Data

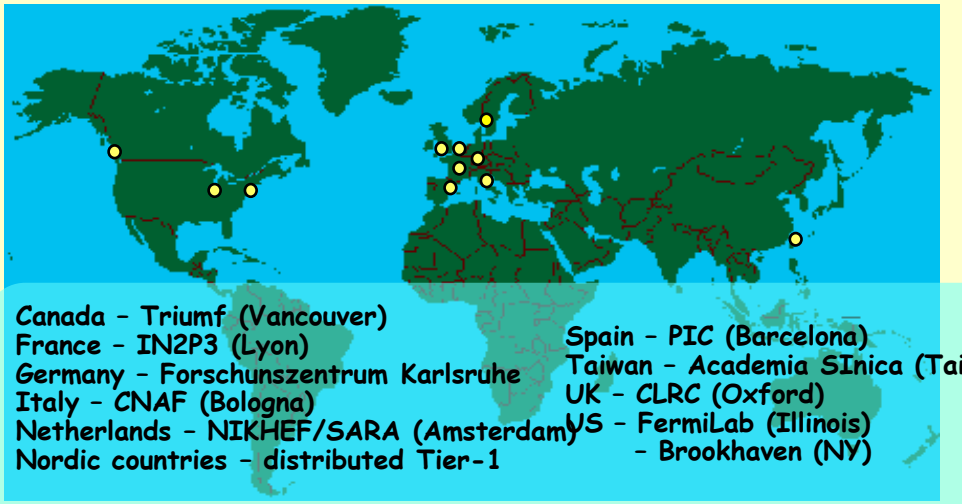
CHEP 2007
Victoria BC
3 September 2007

Les Robertson
WLCG Project Leader

WLCG – The Collaboration – 4 Experiments +

Tier-0 – the accelerator centre

- Data acquisition & initial processing
- Long-term data curation
- Distribution of data → Tier-1 centres



Canada – Triumf (Vancouver)
France – IN2P3 (Lyon)
Germany – Forschunszentrum Karlsruhe
Italy – CNAF (Bologna)
Netherlands – NIKHEF/SARA (Amsterdam)
Nordic countries – distributed Tier-1
Spain – PIC (Barcelona)
Taiwan – Academia SInica (Taipei)
UK – CLRC (Oxford)
US – FermiLab (Illinois)
– Brookhaven (NY)

11 Tier-1 Centres – "online" to the data acquisition process → high availability

- Managed Mass Storage – → grid-enabled data service
- Data-heavy analysis
- National, regional support

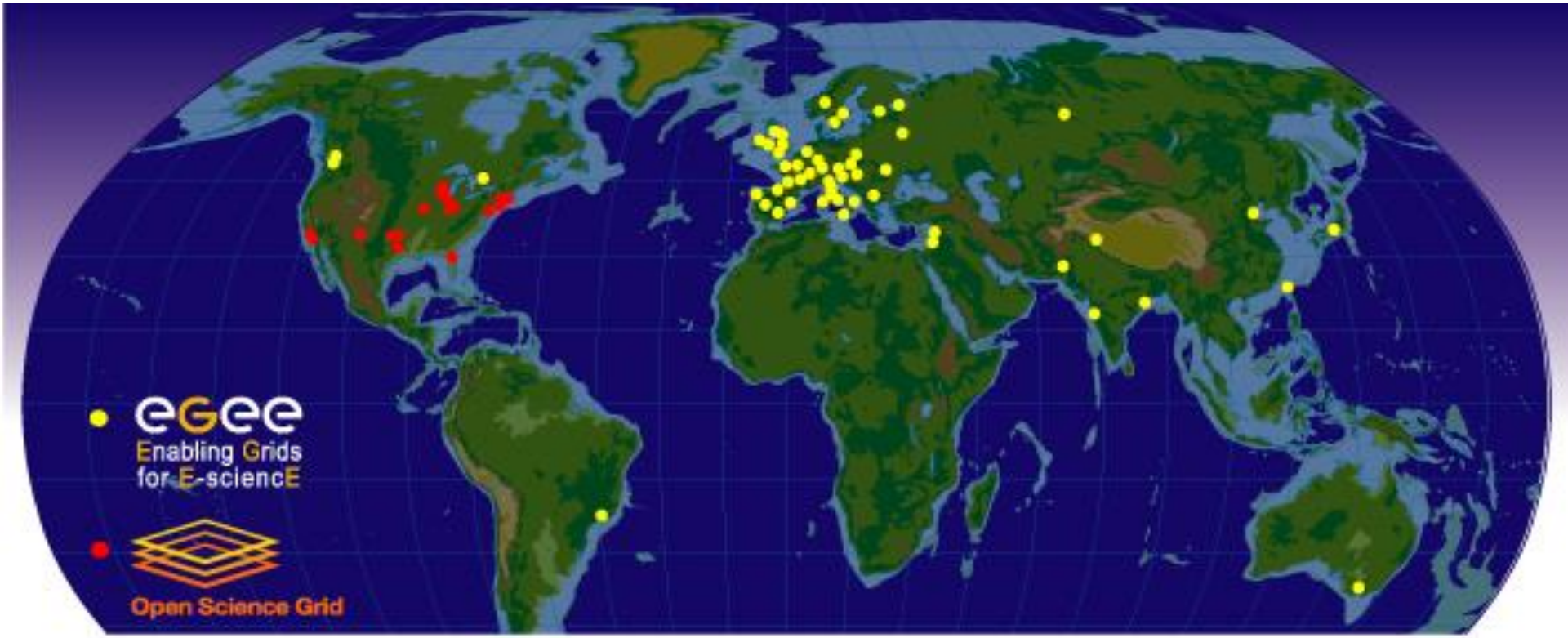
Tier-2 – 112 Centres in 53 Federations in 26 countries

- **End-user (physicist, research group) analysis** – where the discoveries are made
- Simulation

WLCG depends on two major science grid infrastructures ….

EGEE - Enabling Grids for E-Science

OSG - US Open Science Grid

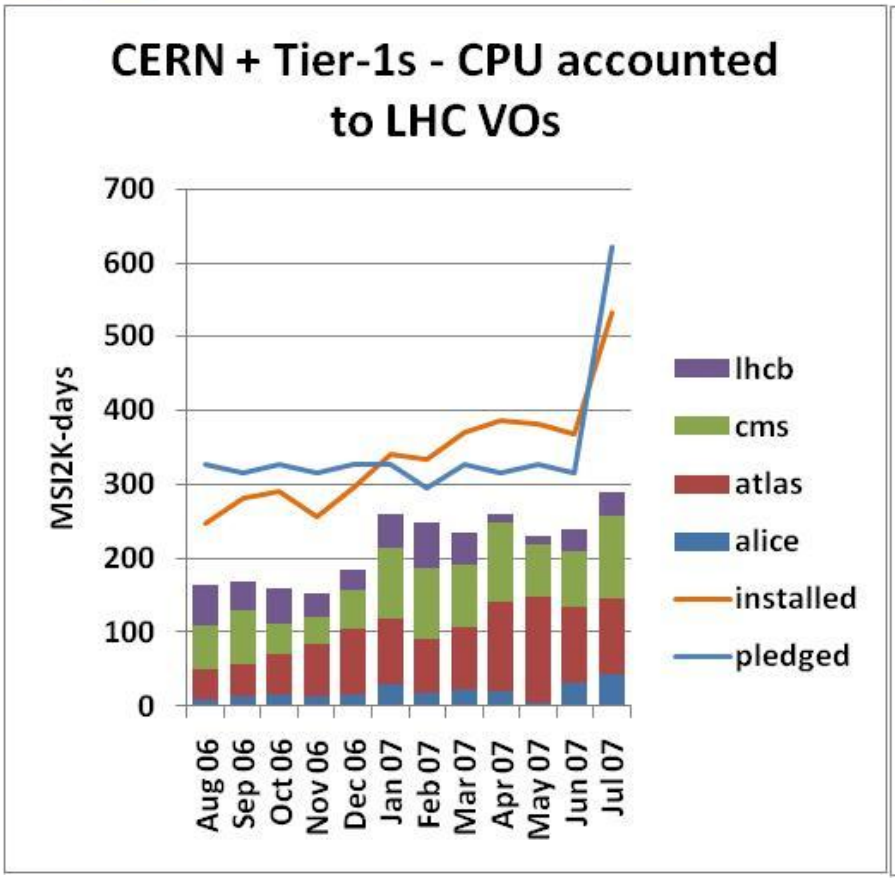


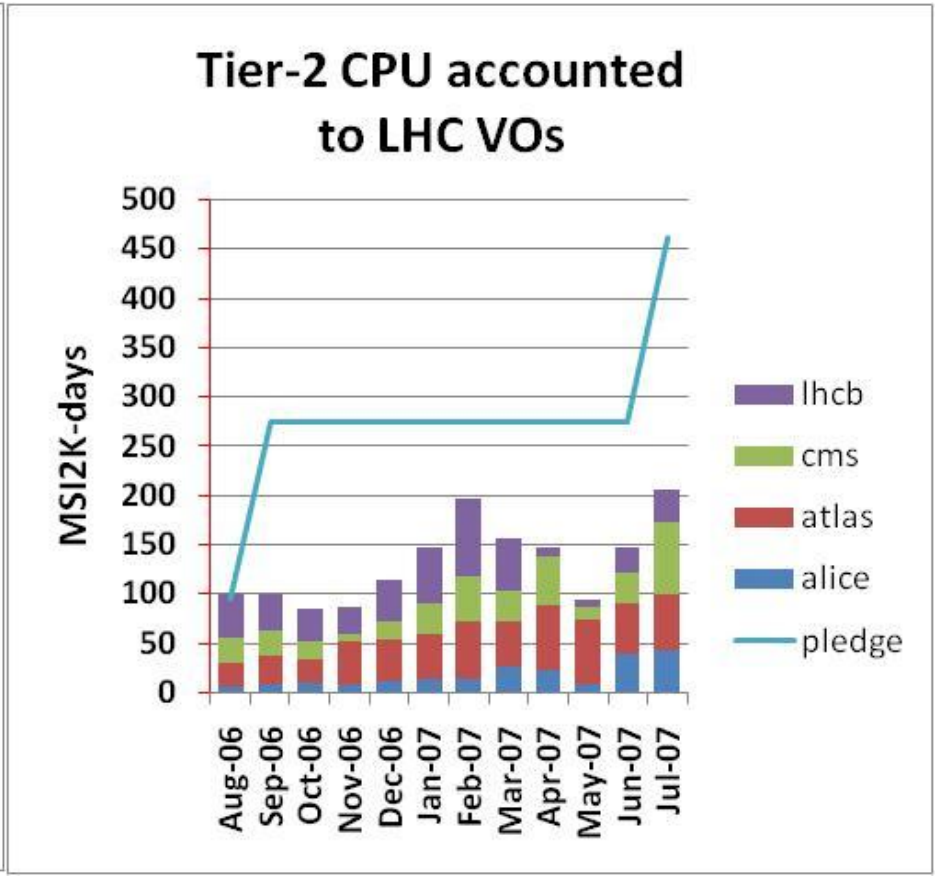
A map of the worldwide LCG infrastructure operated by EGEE and OSG.

The Worldwide LHC Computing Grid

Does it work?

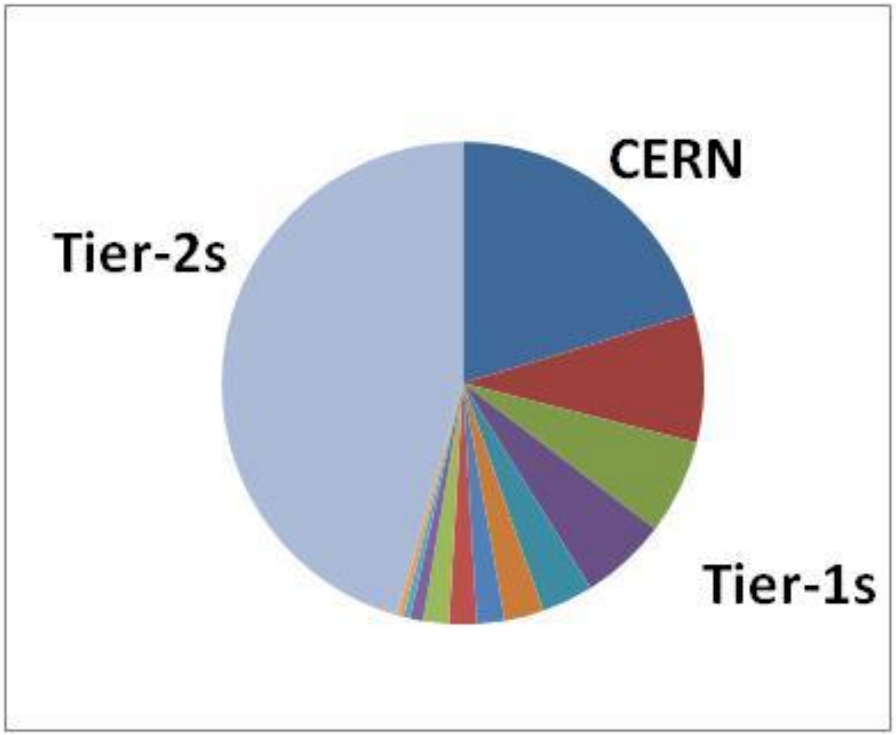
CPU Usage *accounted* to LHC Experiments





accounted/pledged over past 12 months	:	62%	48%
Ramp-up needed over next 8 months	:	6 X	4 X

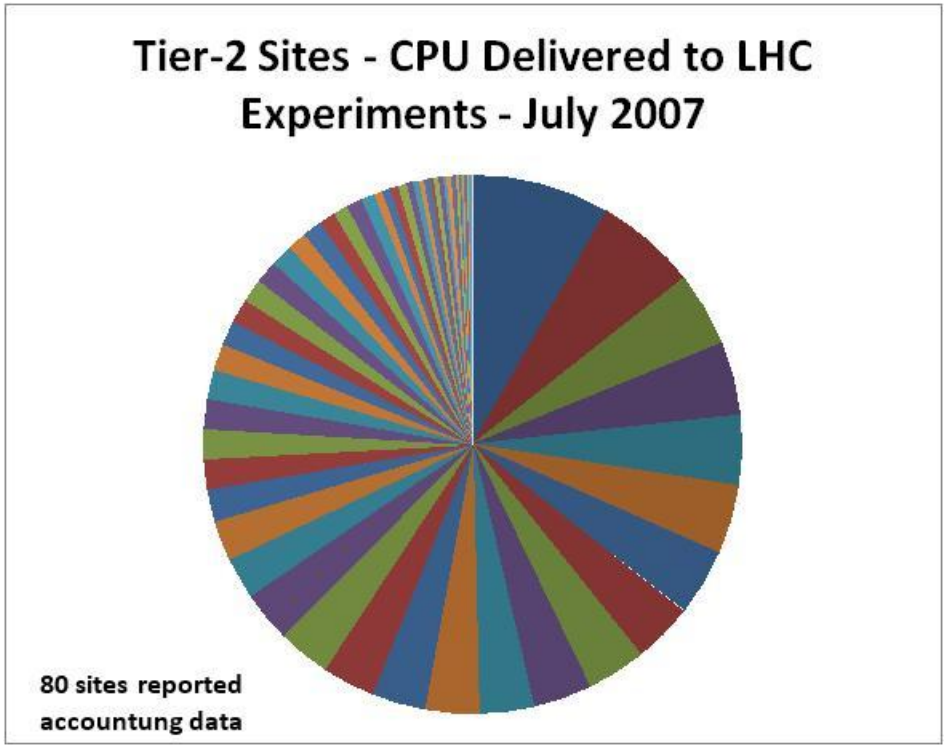
CPU Usage accounted to LHC Experiments July 2007



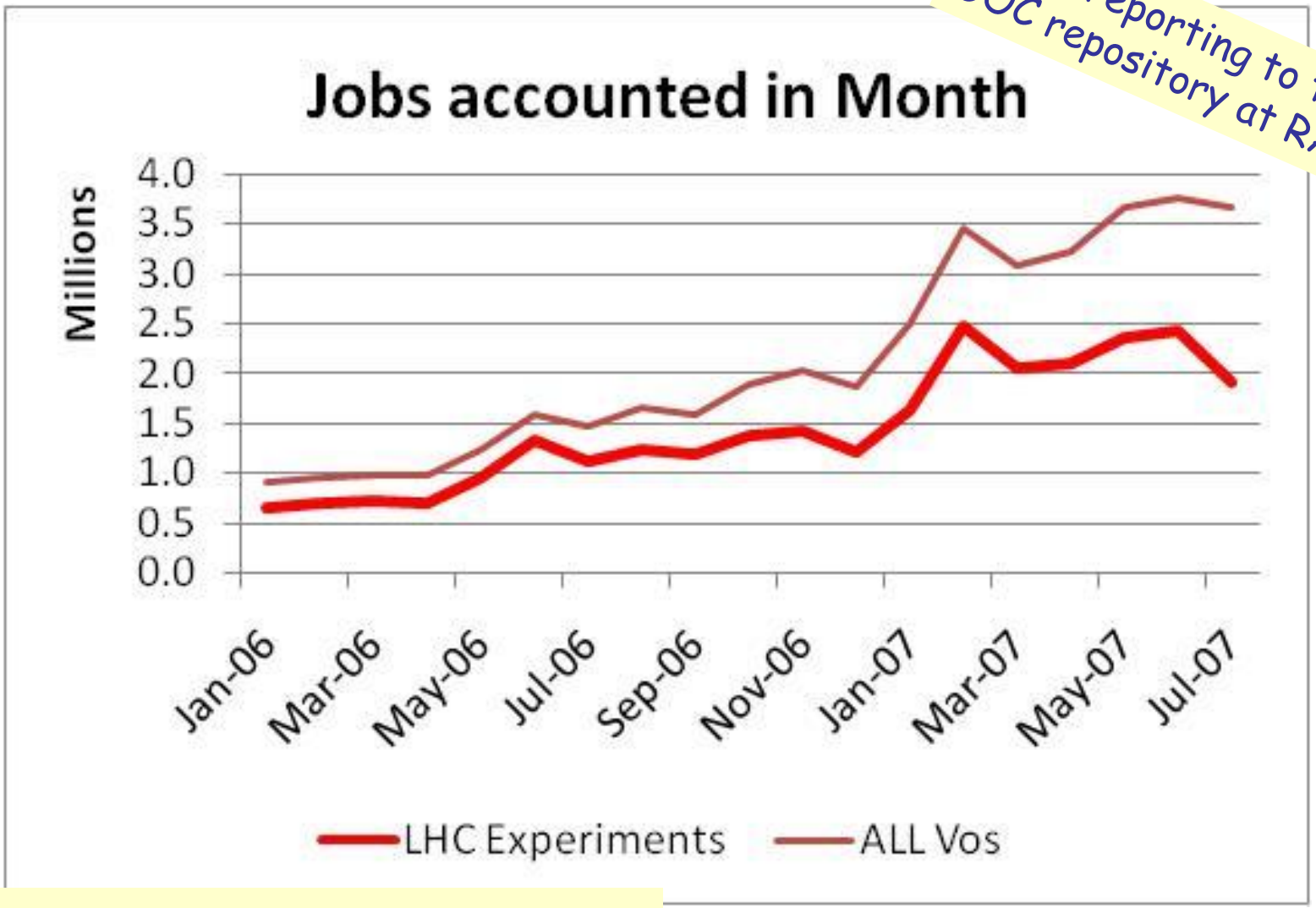
80 Tier-2s 45%
11 Tier-1s 35%
CERN 20%

530M SI2K-days/month (CPU)

9 PB disk at CERN + Tier-1s



Sites reporting to the GOC repository at RAL

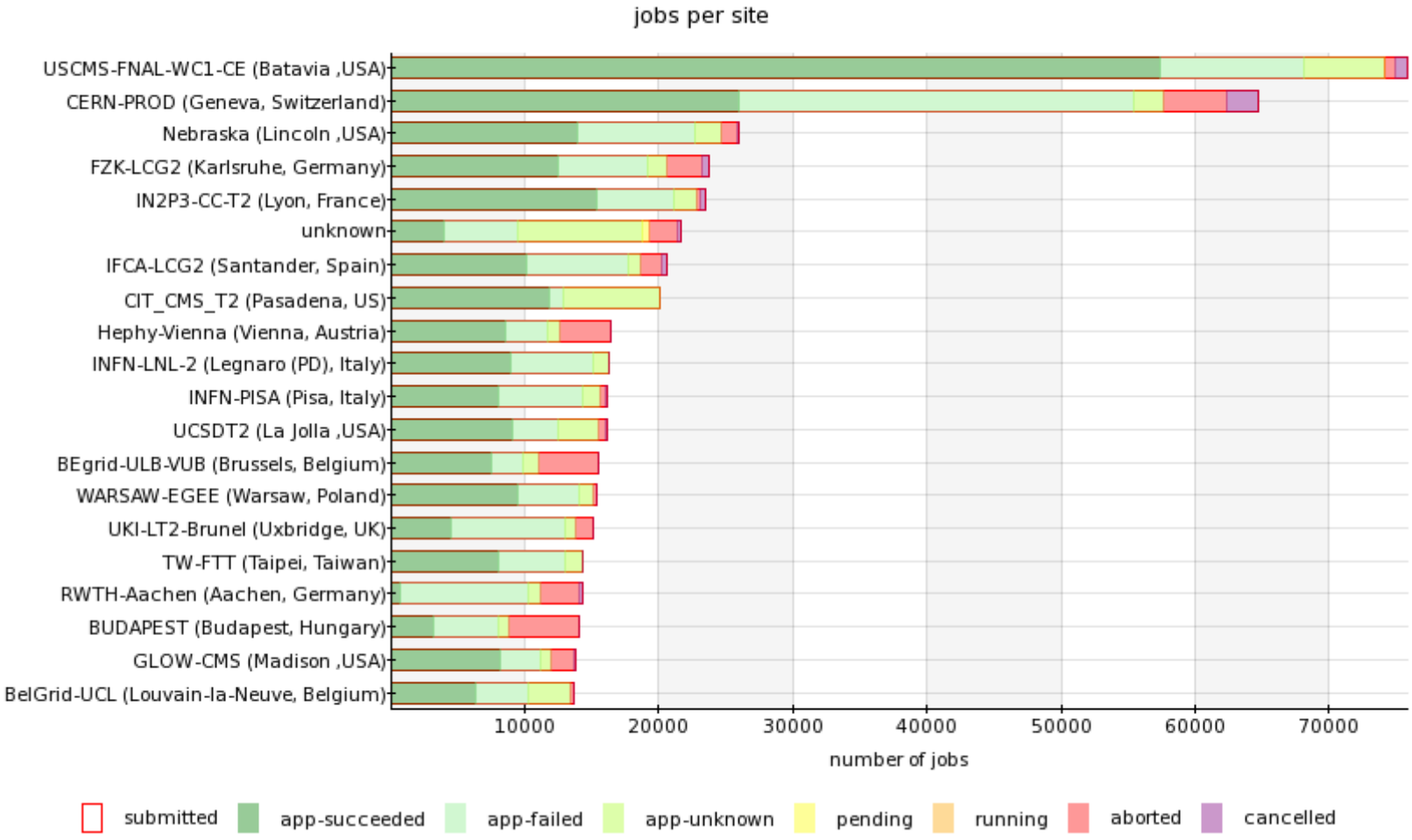


ATLAS+CMS Targets (jobs/month)

End 2007	Mid 2008
3 M	9M

CMS Dashboard - Crab Analysis Jobs

Top 20 of 88 sites running at least one job



Mid-July → mid-August 2007

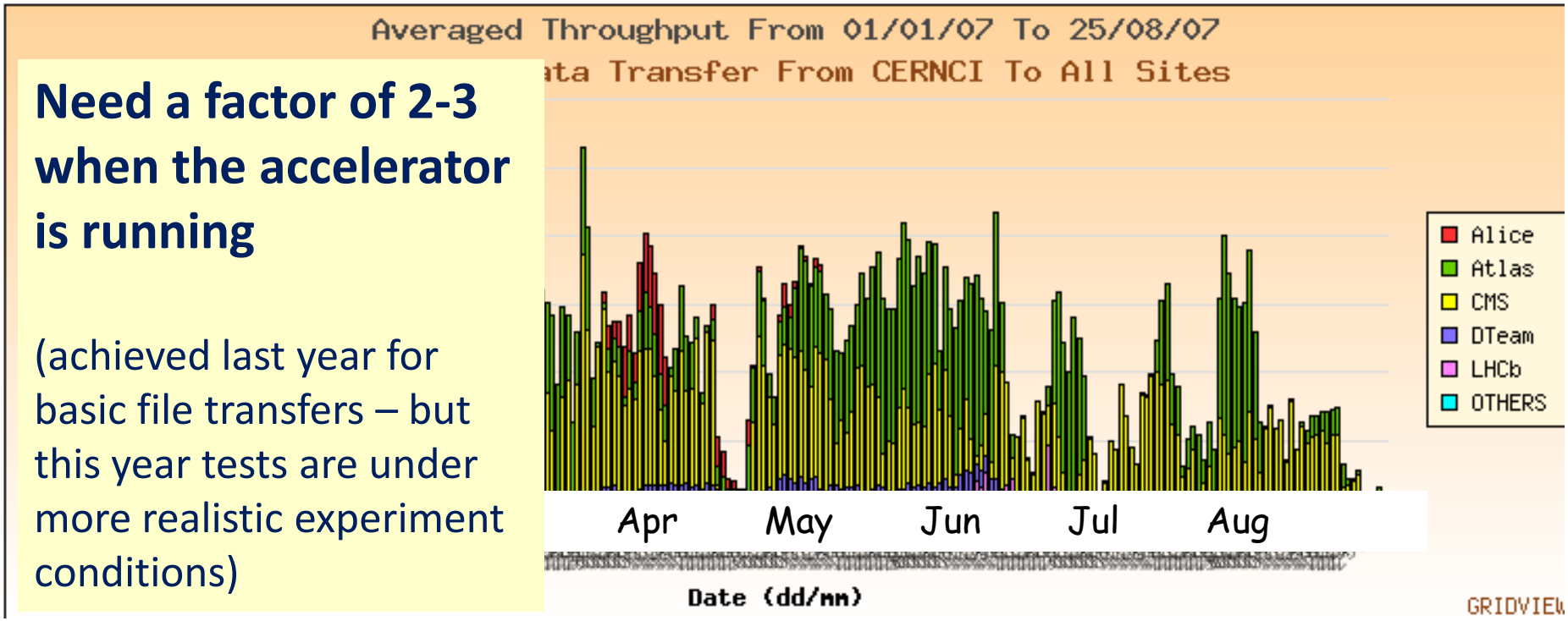
- **645K jobs (20K jobs/day)**
- **89% grid success rate**

2007 – CERN →Tier-1 Data Distribution

Daily Report

(VO-wise Data Transfer From CERNCI To All Sites)

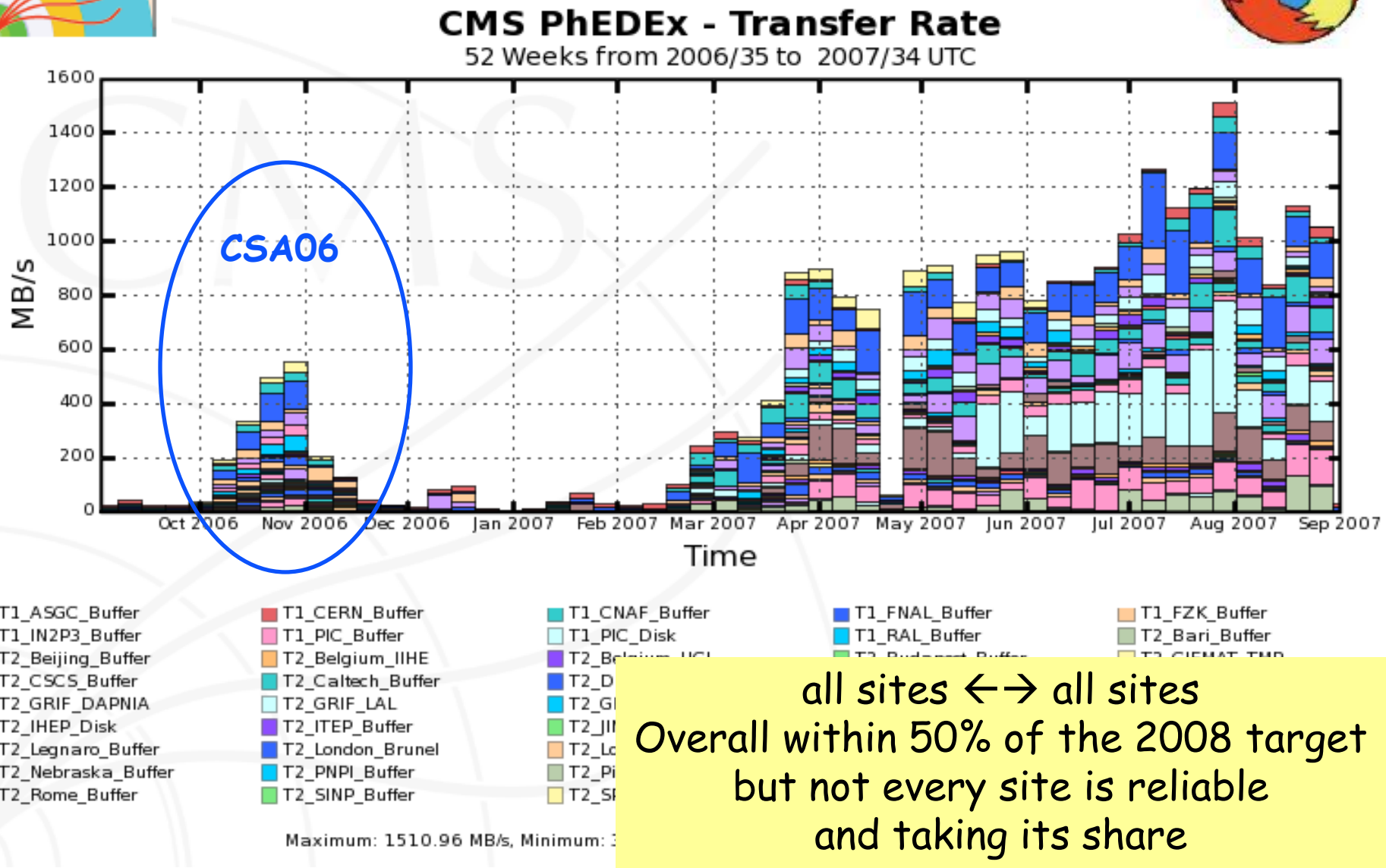
Revert Source/Dest Site(s)



Need a factor of 2-3 when the accelerator is running

(achieved last year for basic file transfers – but this year tests are under more realistic experiment conditions)

Average data rate per day by experiment (Mbytes/sec)



all sites ←→ all sites
Overall within 50% of the 2008 target
but not every site is reliable
and taking its share

Baseline Services

The *Basic* Baseline Services – from the TDR (2005)

- Storage Element
  - Castor, dCache, DPM *(with SRM 1.1)*
  - Storm added in 2007
  - SRM 2.2 – spec. agreed in May 2006 -- being deployed now
- Basic transfer tools – Gridftp, ..
- File Transfer Service (FTS)
- LCG File Catalog (LFC)
- LCG data mgt tools - lcg-utils
- Posix I/O –
  - Grid File Access Library (GFAL)
- Synchronised databases T0←→T1s
  - 3D project

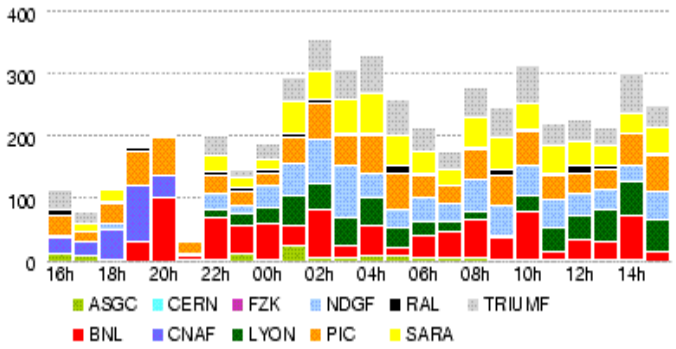
- [illegible]
- [illegible]

… continuing evolution reliability, performance, functionality, requirements

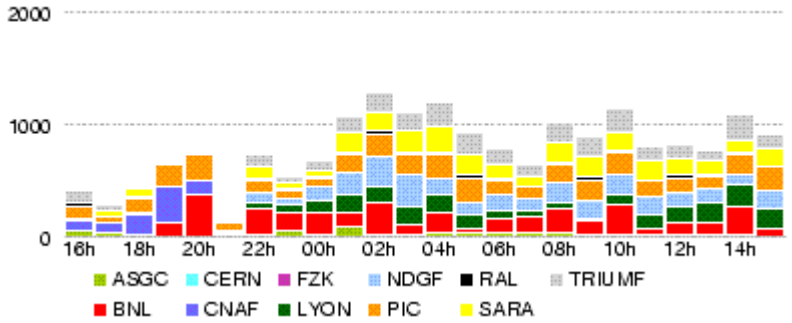
- VO Manage[illegible] (VOMS)
- VO Boxes
- Application software installation
- Job Monitoring Tools

M4 data taking August 31

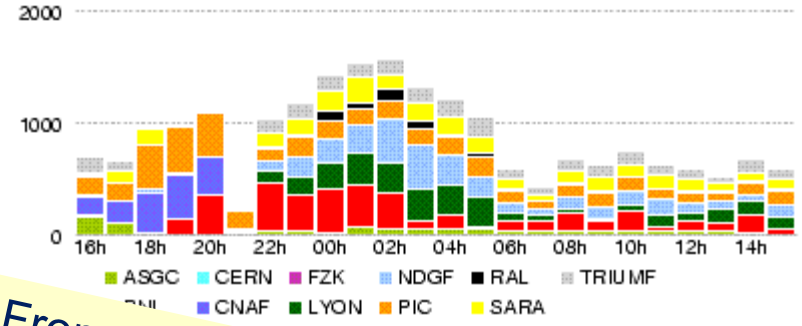
Throughput MB/s



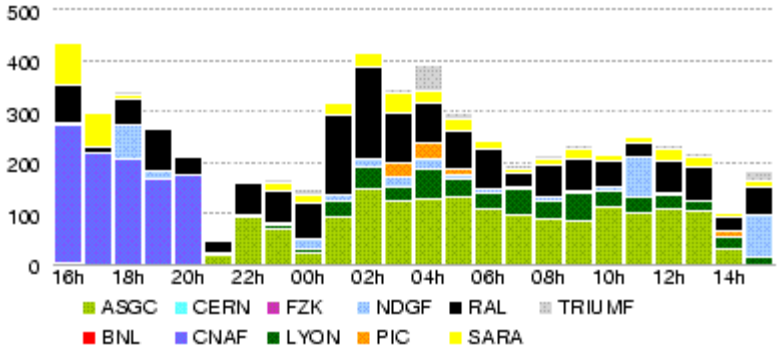
Data transferred GB



Completed filetransfers



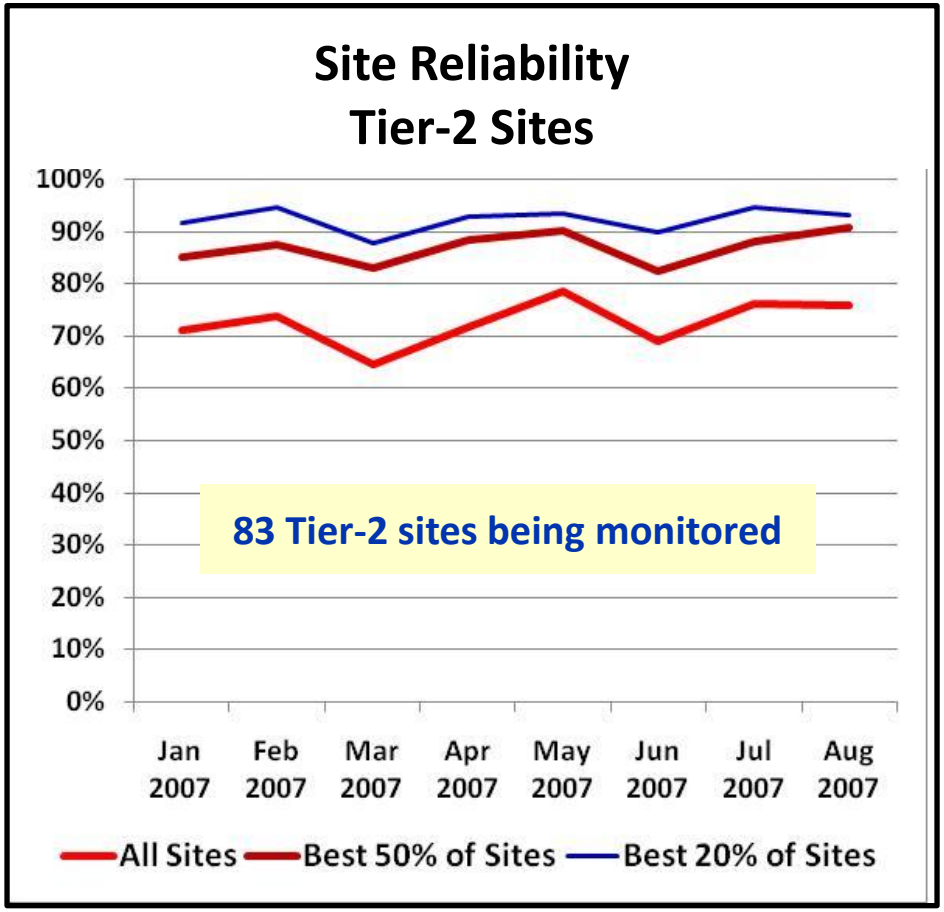
Total number of errors

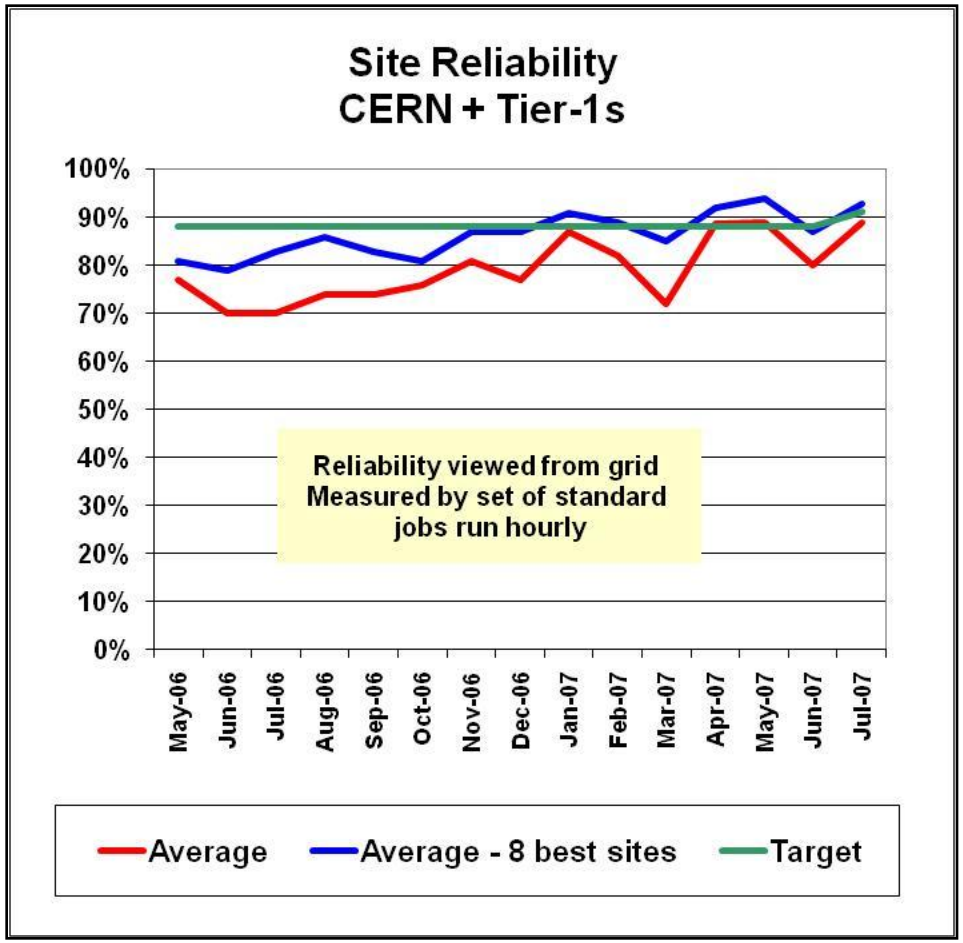


From the detector to analysis at Tier2s

Reliability?

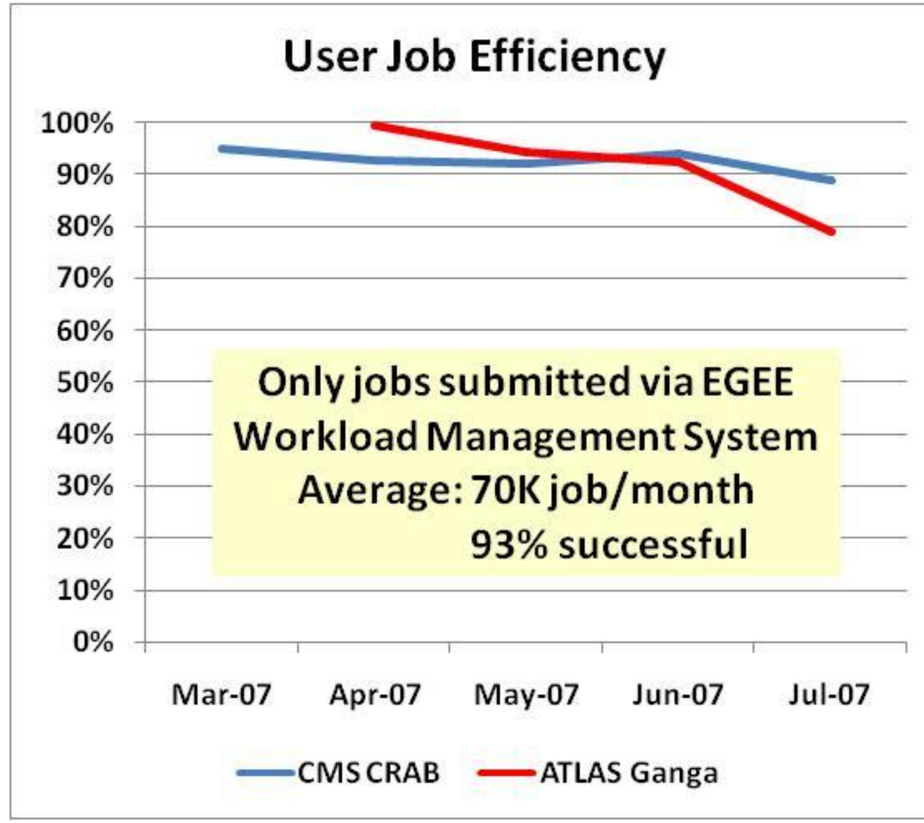
SAM "critical" system tests

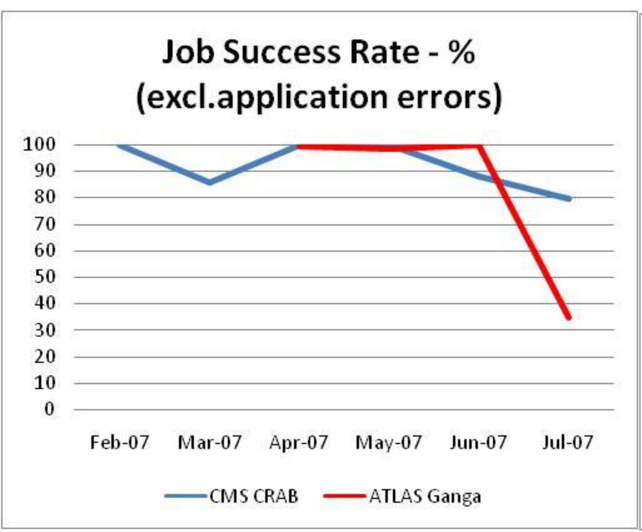


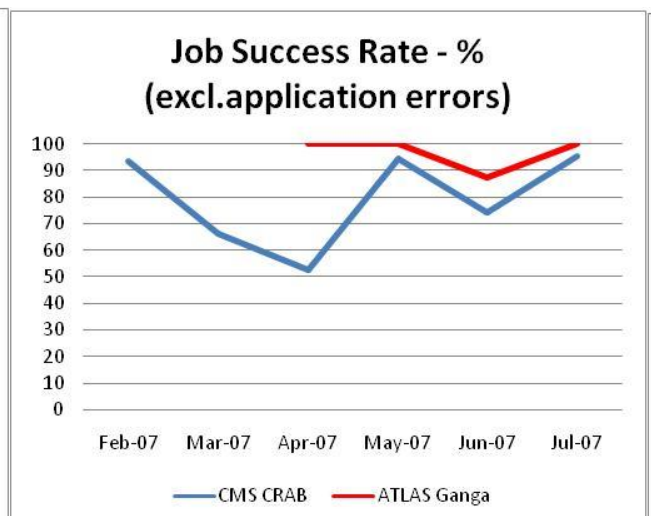


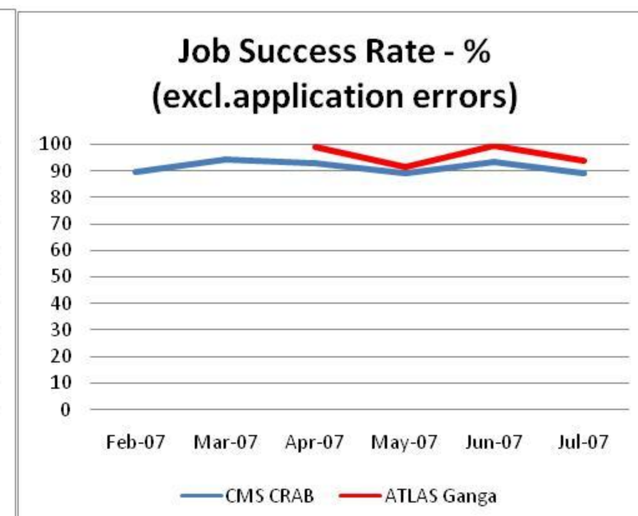
User Job Efficiency

- Job success rate – excluding application errors
- Measured by job log analysis
- At present only for jobs submitted via the EGEE workload management system









Reliability

- Operational complexity is now the weakest link
  - Inconsistent error reporting -- confused by many layers of software – local system, grid middleware, application layers
  - Communications difficult –
    -- sites supporting several (not just LHC) experiments and sometimes other sciences
    -- experiments coordinating across a large number of sites
    -- multiple sites, services implicated in difficult data problems
  - Sites have different histories, different procedures, different priorities

→ A major effort now on monitoring**
  - Integrating grid monitoring with site operations
  - Experiment specific dashboards for experiment operations and end-users

.. and on standard metrics – comparing sites, experiments

***Session on monitoring - Grid Middleware and Tools – Wednesday afternoon*

Reminder – one of the conclusions from the plenary talk at CHEP'04 by Fabiola Gianotti

My 2 main worries today (as an LHC physicist and end-user):

- End-users not yet exposed to massive use/navigation of database and of GRID
  → what will happen when $O(10^3)$ physicists will simultaneously access these systems ?
- Software and Computing Model developed for steady-state LHC operation (≥ 2009 ?)

But : at the beginning they will be confronted with most atypical (and stressful) situations, for which a lot of flexibility will be needed:

-- staged, non-perfect, non-calibrated, non-aligned detectors with all sorts of problems
-- cosmic and beam-halo muons used to calibrate detectors during machine commissioning
-- machine backgrounds ; higher-than-expected trigger rates
-- fast/frequent reprocessing of part of data (e.g. special calibration streams)
-- $O(10^3)$ physicists in panic-mode using and modifying the Software and accessing the database, GRID …

⇒ it is time for the Software/Computing to address the early phase of LHC operation, not to hinder the fast delivery of physics results (and a possible early discovery …)

Are we approaching the Plateau of Productivity?

Gartner Group

The Technology Hype Cycle

HEP Grid on the CHEP timeline

Expectations

Beijing

San Diego

Padova

Interlaken

Mumbai

Victoria?

Trigger | Peak of Expectations | Trough of Disillusionment | Slope of Enlightenment | Plateau of Productivity | Time

Middleware & services:

- Initial goals over-ambitious – but we now have basic functionality, tools, services
- SRM 2.2 is late – and storage management is hard
- Experiments have to live with the functionality that we have now

Usage:

- Experiments are running large numbers of jobs – despite their (justified) complaints
- And transferring large amounts of data – though not always to where they want it
- ATLAS has taken cosmic data from the detector to analysis at Tier-2s
- End-users beginning to run analysis jobs – but sites need to understand much better how analysis will be done during the first couple of years → and what the implications are for data

Scalability:

- 5-6 X needed for resource capacity, number of jobs
- 2-3 X needed for data transfer

Reliability:

- Not yet good enough
- Data Transfer is still the most worrying - despite many years of planning and testing
  - Many errors → complicated recovery procedures
  - Many sources of error – storage systems, site operations, experiment data management systems, databases, grid middleware and services, networks, ...

  Hard to get to the roots of the problems

Are we getting there? Slowly!

Need continuous testing from now until first beams

- Driven by experiments with realistic scenarios, good monitoring and measurements
- and the pro-active participation of sites, developers, storage experts

After so many years ---
the beams are now on the horizon & we can all focus on the contribution that we can make to extracting the physics