



WLCG Status Report

7th May, 2008
LHCC Open Session

Ian Bird
LCG Project Leader





Online details - cern.ch/LCG

- ▶ Project Structure
 - Boards
 - CRRB
 - MD
 - CB
 - OB
 - GDB
 - Committees
 - LHCC
 - SC2
 - Architects Forum
- ▶ Project Planning
 - Documents
 - Dissemination
 - Related Projects
 - ▶ LCG Bulletin
 - ▶ Press & Media
 - ▶ Jobs

The Large Hadron Collider (LHC), currently being built at CERN near Geneva, is the largest scientific instrument on the planet. When it begins operations in 2007, it will produce roughly 15 Petabytes (15 million Gigabytes) of data annually, which thousands of scientists around the world will access and analyse.

The mission of the LHC Computing Project (LCG) is to build and maintain a data storage and analysis infrastructure for the entire high energy physics community that will use the LHC.

▶ Project Overview



Worldwide LHC Computing Grid

Distributed Production Environment for Physics data Processing

Activities

- ▶ Distributed Analysis (ARDA)
- ▶ Grid Deployment
- ▶ Security
 - ▶ CCRC'08 (wiki)
 - ▶ Physics Application Software
- ▶ LCG Optical Private Network
- ▶ Technical Design Report (TDR)
- ▶ Status of WLCG (presentation at IEEE NSS Conference 06)

LCG Users

New Users	Registered Users
■ User Registration	■ User Support
	■ Experiment Integration Support

LCG Sites

■ Getting Started	■ Site Guides and FAQ
■ Software Releases	■ Site Security

LCG Operations

■ Monitoring	■ Security Incidents
■ Core Infrastructure Center	

LCG Bulletin





WLCG Collaboration

- The Collaboration
 - 4 LHC experiments
 - ~130 computing centres
 - 12 large centres (Tier-0, Tier-1)
 - 56 federations of smaller "Tier-2" centres – 121 sites
 - Growing to ~40 countries
 - Grids: EGEE, OSG, Nordugrid
- Technical Design Reports
 - WLCG, 4 Experiments: June 2005
- Memorandum of Understanding
 - Agreed in October 2005
- Resources
 - 5-year forward look

MoU Signing Status

Tier 1 – all have now signed

Tier 2:

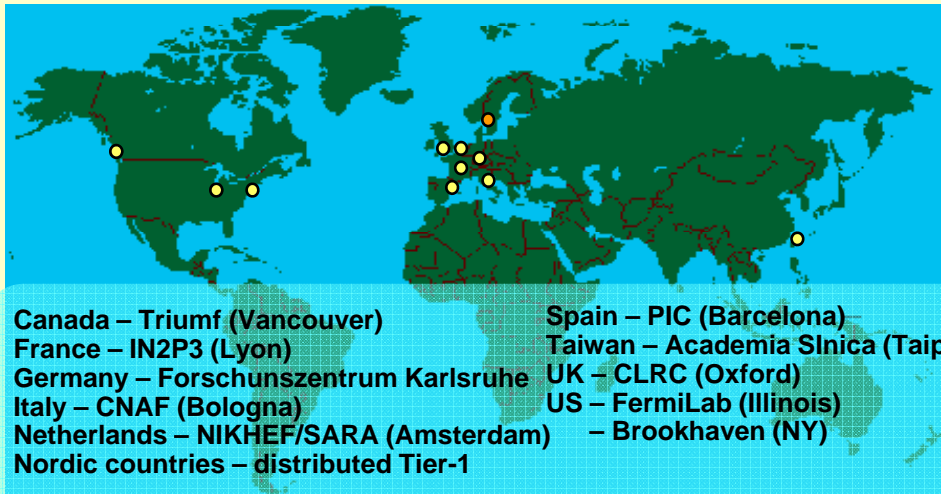
Australia	Netherlands	<u>Still to sign:</u>
Belgium	Norway *	Austria
Canada *	Pakistan	Brazil (under discussion)
China	Poland	
Czech Rep. *	Portugal	
Denmark	Romania	* Recent additions
Estonia	Russia	
Finland	Slovenia	
France	Spain	
Germany (*)	Sweden *	
Hungary *	Switzerland	
Italy	Taipei	
India	Turkey *	
Israel	UK	
Japan	Ukraine	
JINR	USA	
Korea		



LCG Service Hierarchy

Tier-0: the accelerator centre

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



Tier-1: “online” to the data acquisition process → high availability

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

Tier-2: ~120 centres in ~33 countries

- Simulation
- End-user analysis – batch and interactive

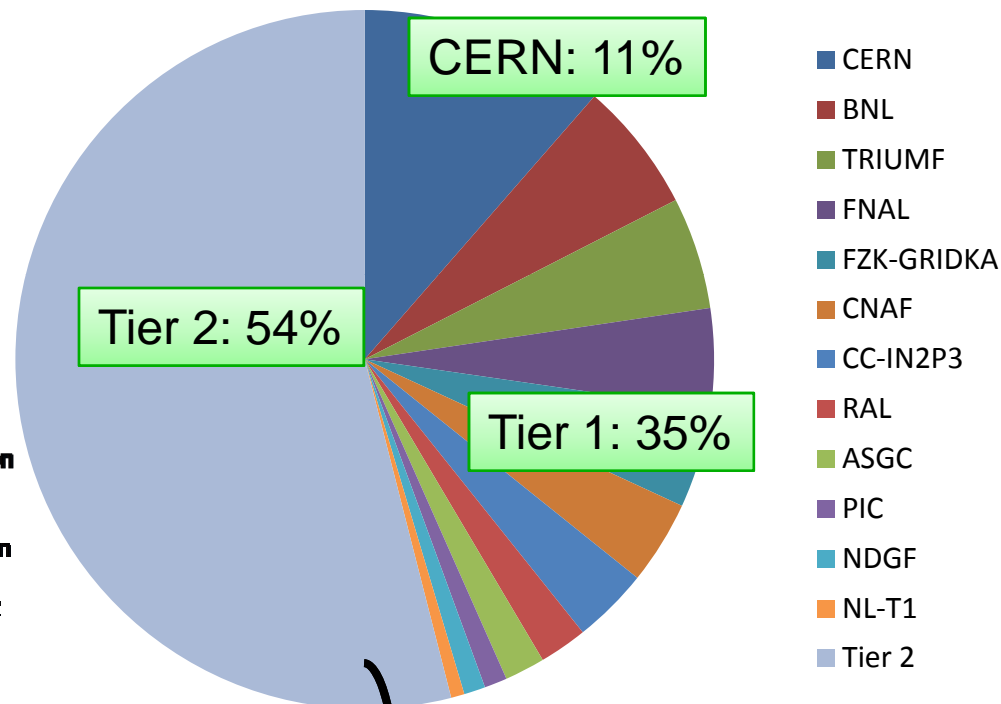




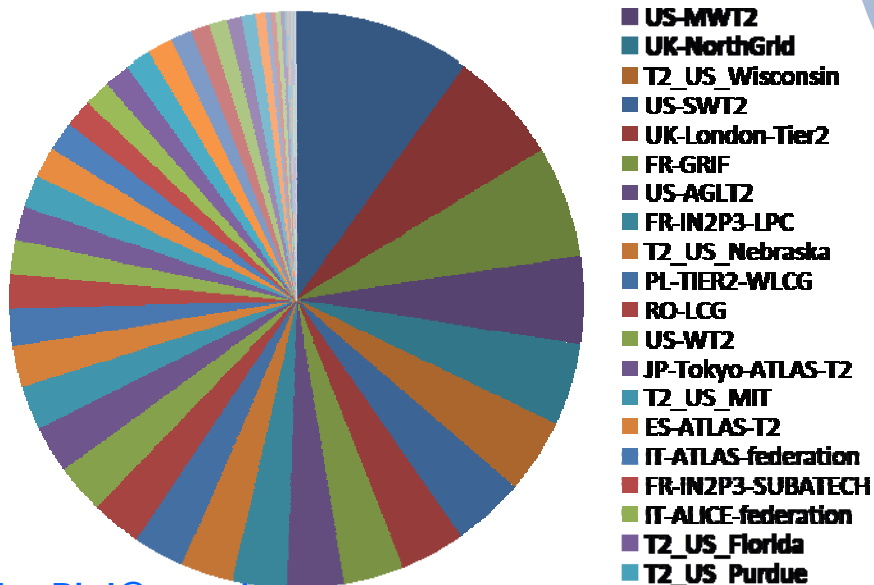
Recent grid use

- Across all grid infrastructures
- Preparation for, and execution of CCRC'08 phase 1
 - Move of simulations to Tier 2s

CPU Usage Jan-Feb 2008

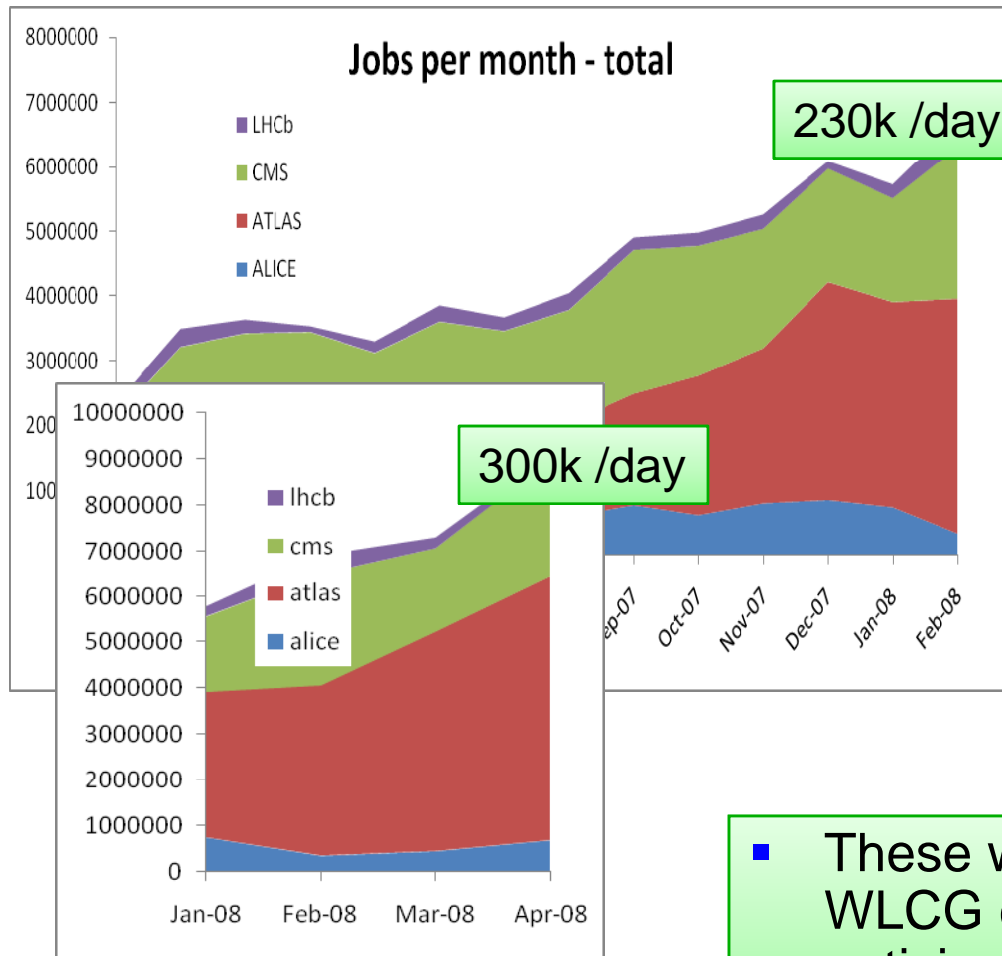


Tier 2 federation use





Recent grid activity



- WLCG ran ~ 44 M jobs in 2007 – workload has continued to increase
 - 29M in 2008 – now at ~ >300k jobs/day
- Distribution of work across Tier0/Tier1/Tier 2 really illustrates the importance of the grid system
 - Tier 2 contribution is around 50%; > 85% is external to CERN

- These workloads (reported across all WLCG centres) are at the level anticipated for 2008 data taking



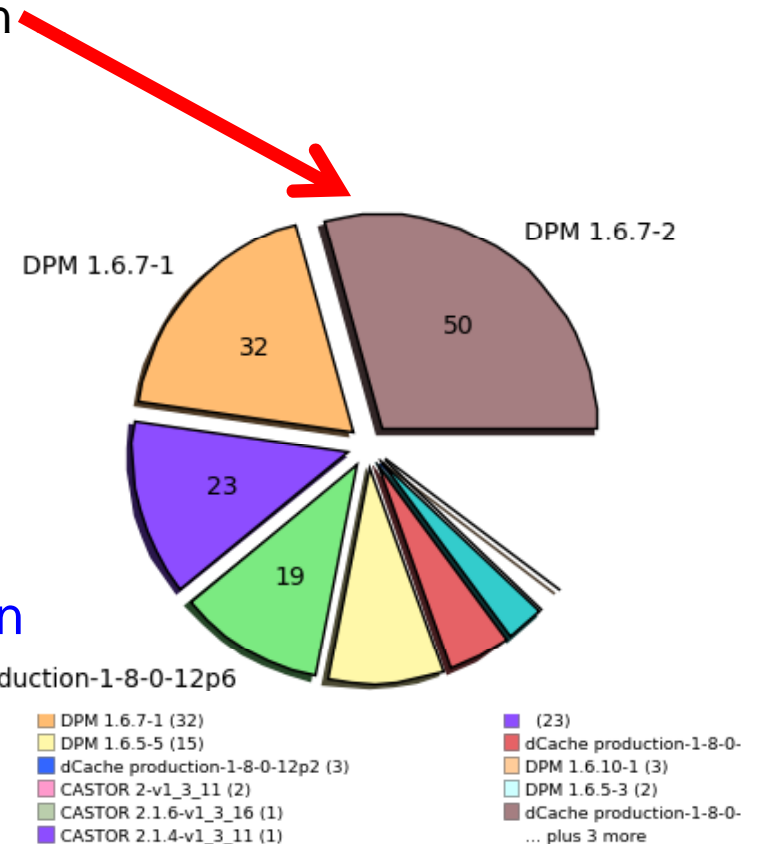
Combined Computing Readiness Challenge - CCRC'08

- Objective was to show that we can run together (4 experiments, all sites) at 2008 production scale:
 - All functions, from DAQ \Rightarrow Tier 0 \Leftrightarrow Tier 1s \Leftrightarrow Tier 2s
- Two challenge phases were foreseen:
 1. **Feb:** not all 2008 resources in place – still adapting to new versions of some services (e.g. SRM) & experiment s/w
 2. **May:** all 2008 resources in place – full 2008 workload, all aspects of experiments' production chains
- Agreed on specific targets and metrics – helped integrate different aspects of the service
 - ❑ Explicit “**scaling factors**” set by the experiments for each functional block (e.g. data rates, # jobs, etc.)
 - ❑ Targets for “**critical services**” defined by experiments – essential for production, with analysis of impact of service degradation / interruption
 - ❑ WLCG “**MoU targets**” – services to be provided by sites, target availability, time to intervene / resolve problems ...



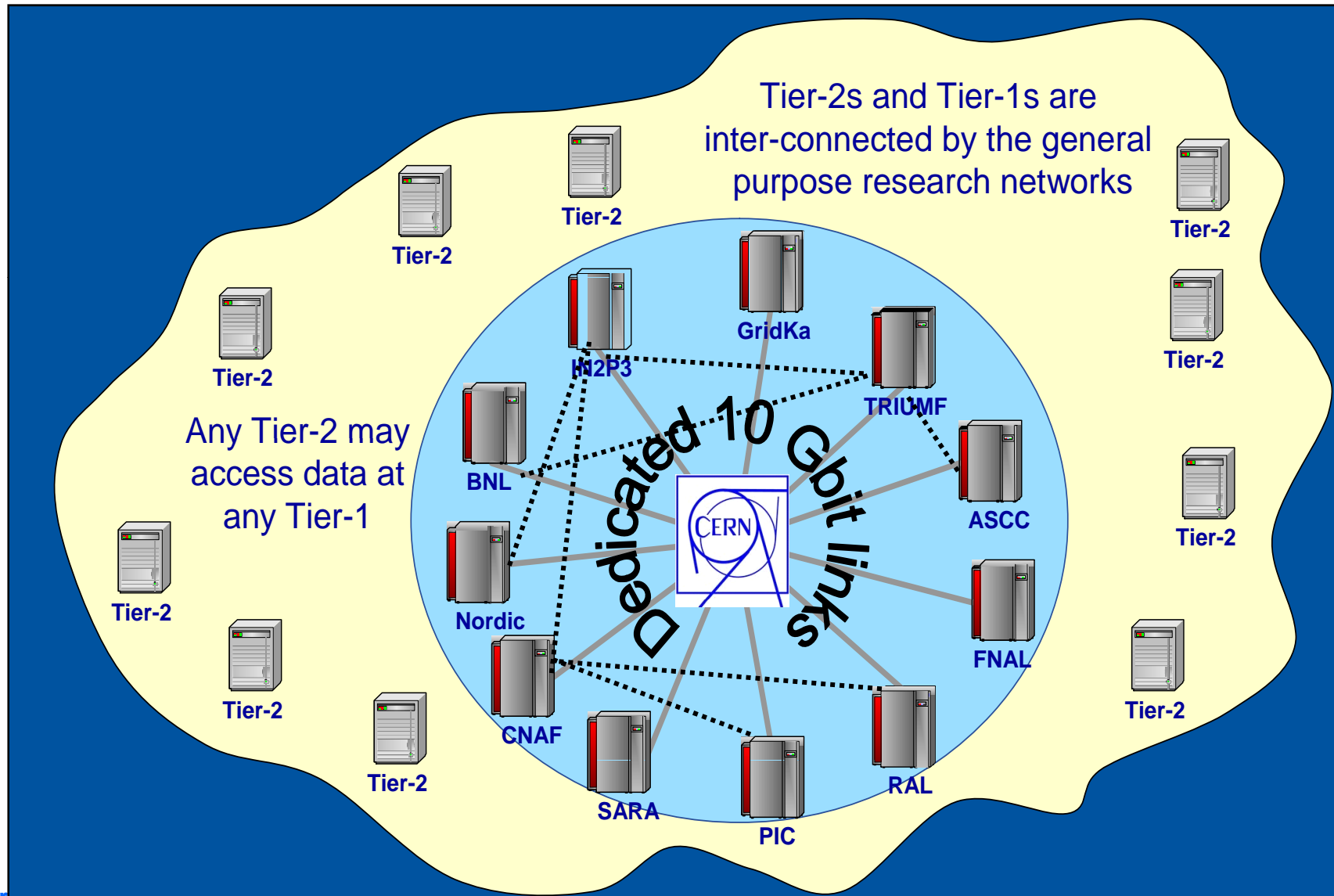
SRM v2.2 Deployment

- Deployment plan was defined and agreed last September, but schedule was very tight
- Deployment of dCache 1.8.x and Castor with srm v2.2 was achieved at all Tier0/Tier 1 by December
 - Today 174 srm v2 endpoints are in production
- During February phase of CCRC'08 relatively few problems were found:
 - Short list of SRM v2 issues highlighted, 2 are high priority
 - Will be addressed with fixes or workarounds for May
 - Effort in testing was vital
- Still effort needed in site configurations of MSS – iterative process with experience in Feb & May





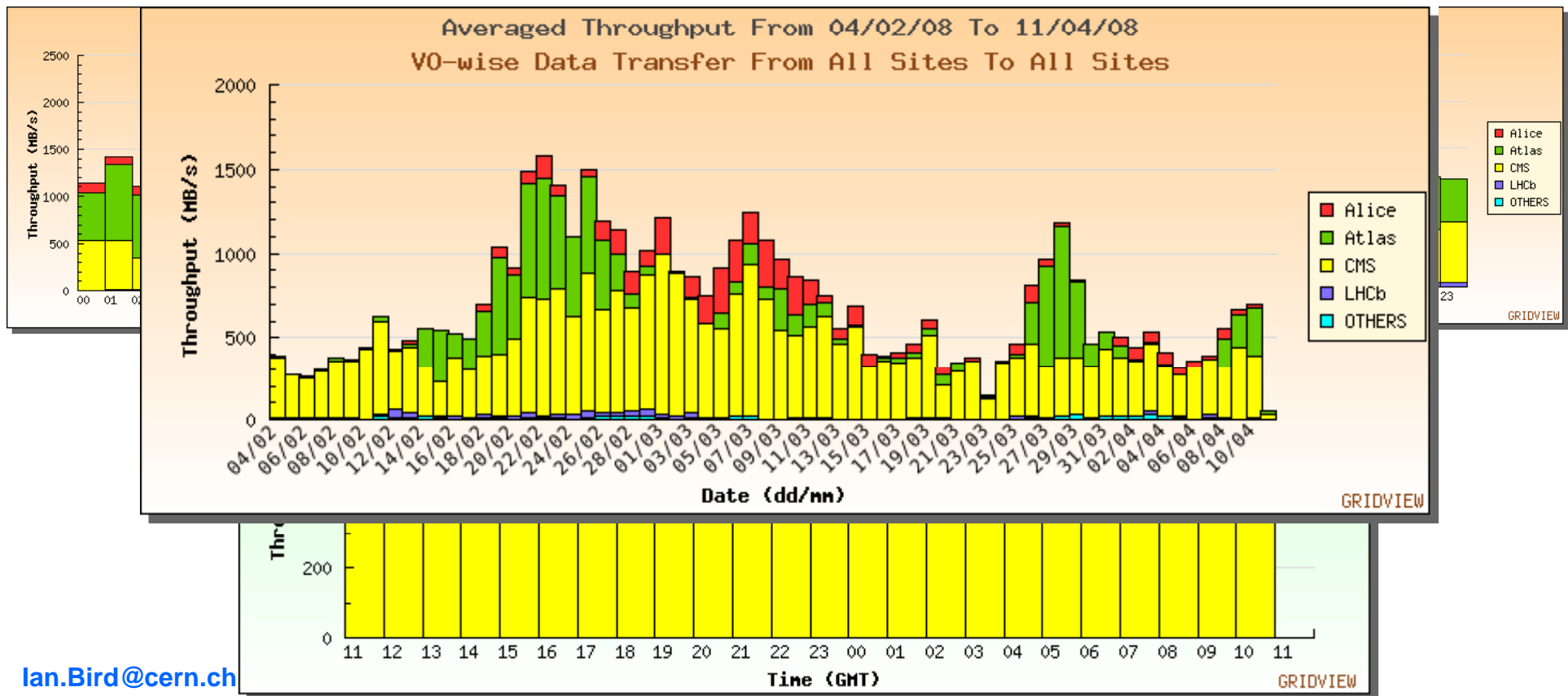
LHC OPN





Data transfer

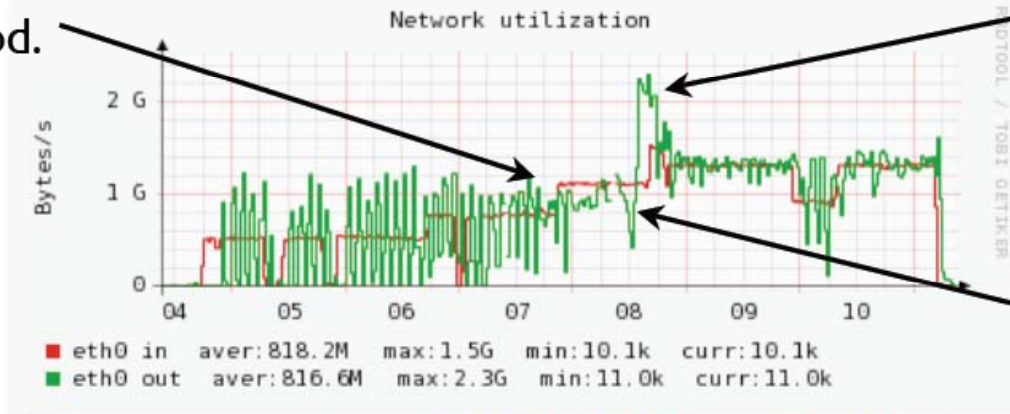
- Data distribution from CERN to Tier-1 sites
 - The target rate was achieved in 2006 under test conditions
 - Autumn 2007 & CCRC'08 under more realistic experiment testing, reaching & sustaining target rate with ATLAS and CMS active





Castor performance - Tier 0

Migration
policy mod.



t0export: in from WN, out to tape

Recovered

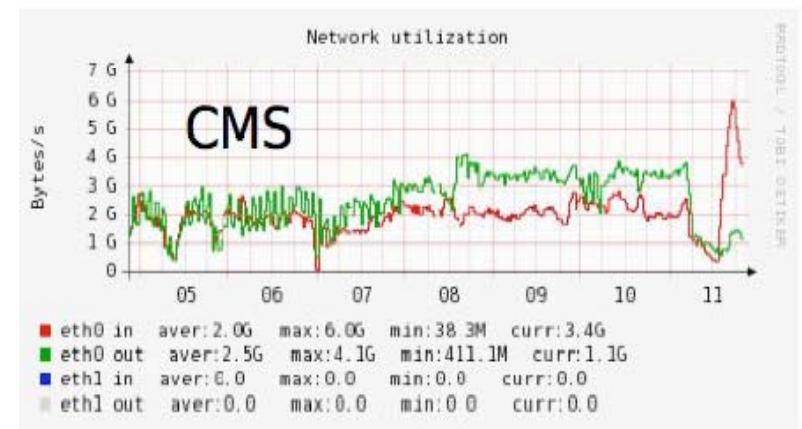
Ran out of
tapes

■ CMS:

- Aggregate rates in/out of castor of 3-4 GB/s
- Sustained rate to tape 1.3 GB/s with peaks > 2 GB/s

■ May:

- Need to see this with all experiments





CCRC'08 Results (Feb)

- Preparation:
 - SRM deployment (Castor, dCache, DPM) completed
 - Relatively few problems seen in production
 - Middleware: agreed functional needs in place
- Production:
 - Data rates demonstrated to all Tier 1s; experiments all achieved in excess of their targets for sustained periods
 - Castor (Tier 0) shown to support in excess of rates needed – total throughput and to tape
 - Still need to show this for all experiments together
 - Grid workloads (>300k jobs/day) in excess of 2008/2009 needs
 - Demonstrated that this is sustainable with existing staffing levels
- Preparation for May:
 - Pragmatic solutions for few remaining middleware issues
 - Middleware updates in place;
 - Resource installation – delays at many Tier 1s

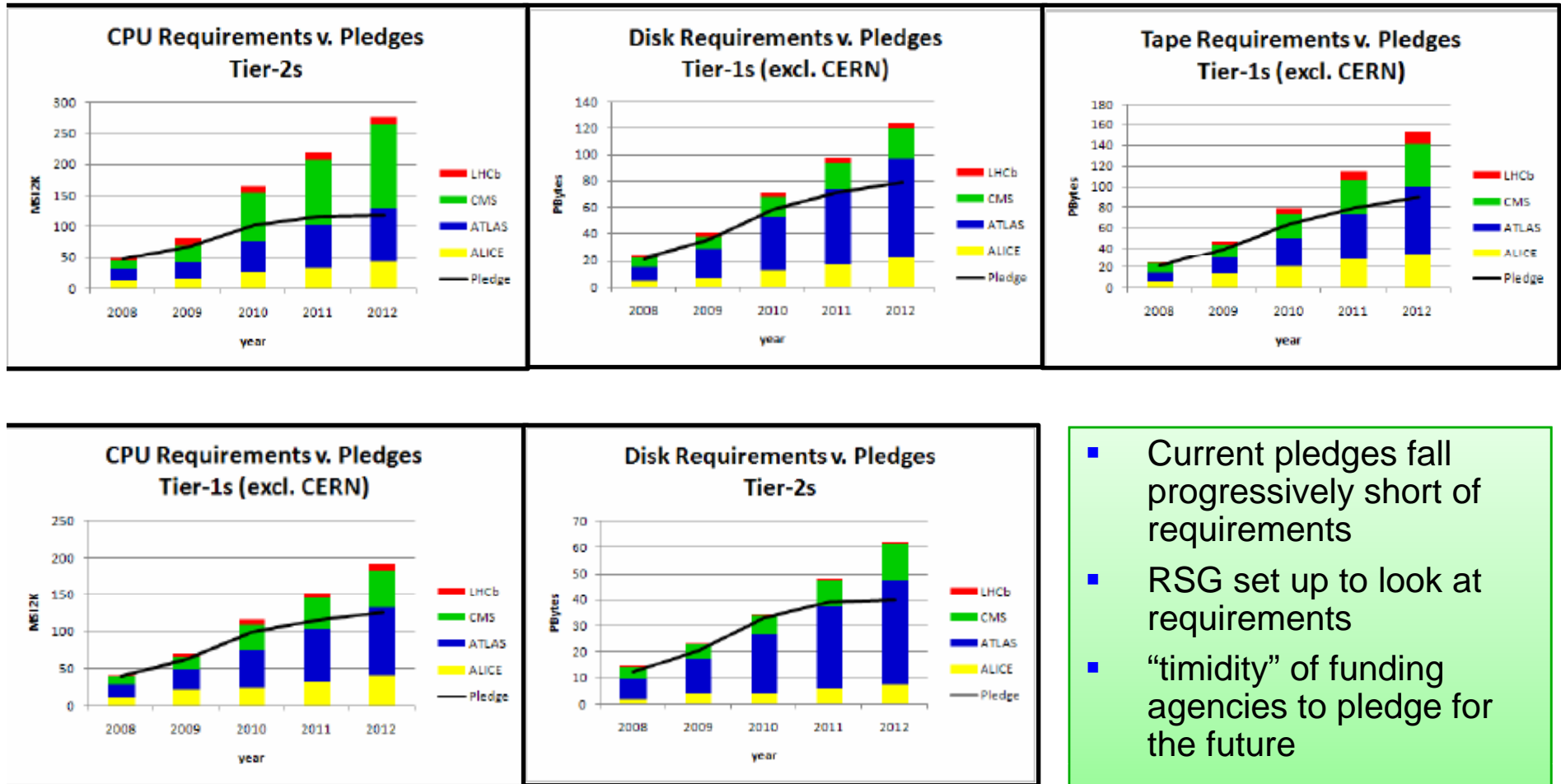


Resource ramp up for 2008

- CPU:
 - Most Tier 1 sites will have full 2008 pledges in place in May
 - Largest missing is +2500 at NL-T1 due Nov.
- Disk and tape
 - Many sites will catch up later in the year as need expands:
 - 2008 disk requirements are 23 PB with 15.5 PB expected by 1 May
 - 2008 tape requirements are 24 PB with 15 PB expected by 1 May.
- May run of CCRC'08 at 55% only requires +1PB of disk and +1.5PB of tape (mostly reusable) so should have no resource problems.
- Full status of resource installation was reported at C-RRB in April.
- Many sites had problems with procurement process/ vendor delivery/ faulty equipment
 - These issues must be taken into account in future – the process is long, but yearly deadlines are important



Resource pledges vs requirements

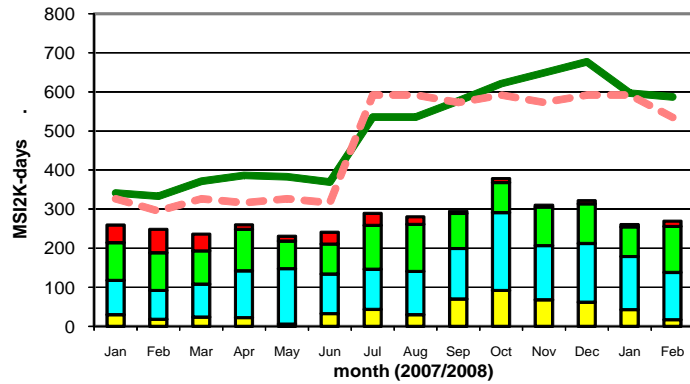


- Current pledges fall progressively short of requirements
- RSG set up to look at requirements
- “timidity” of funding agencies to pledge for the future

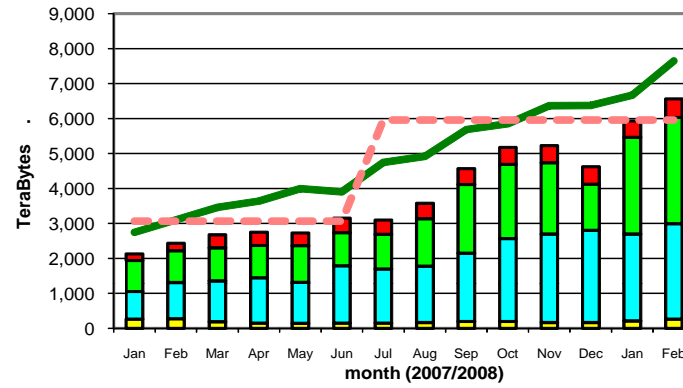


Resource utilization Tier 0+Tier 1

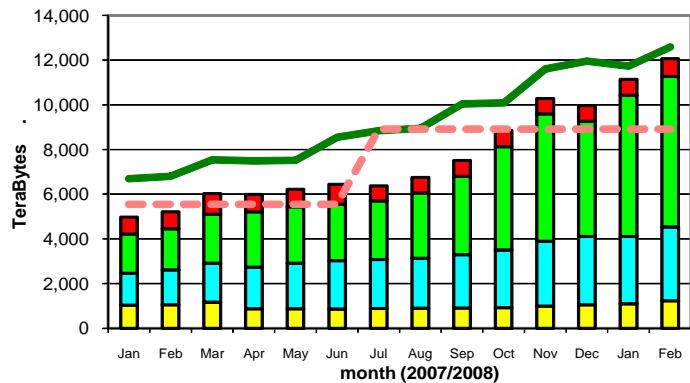
CPU Time Delivered



Disk Storage Used



Tape Storage Used



- Installed capacity exceeds 2007 pledges
 - But – ramp up to 2008 – due in April

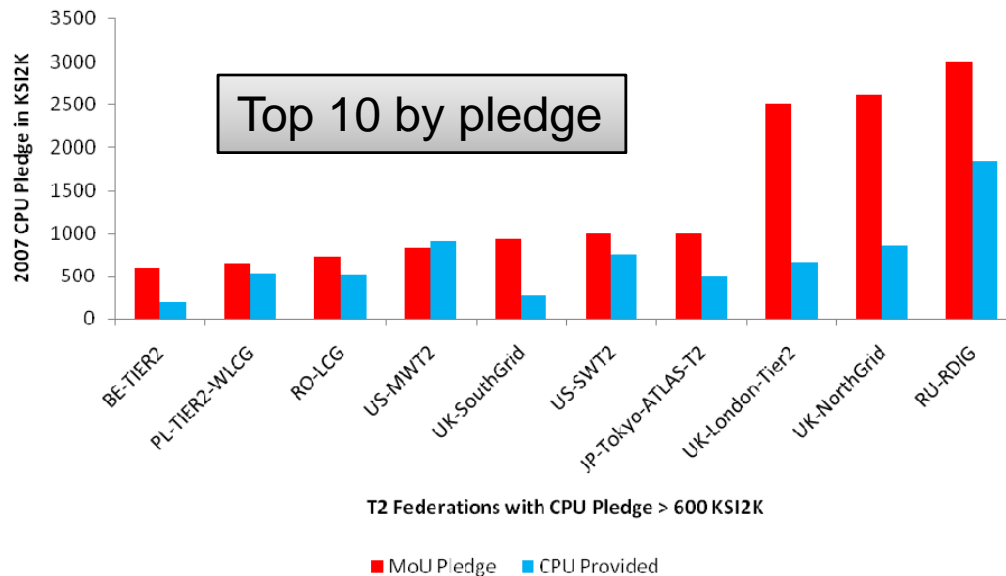
ALICE ■ ATLAS ■ LHCb ■ CMS ■
 installed capacity (inc. efficiency factor) —
 MoU commitment (inc. efficiency factor) - - -



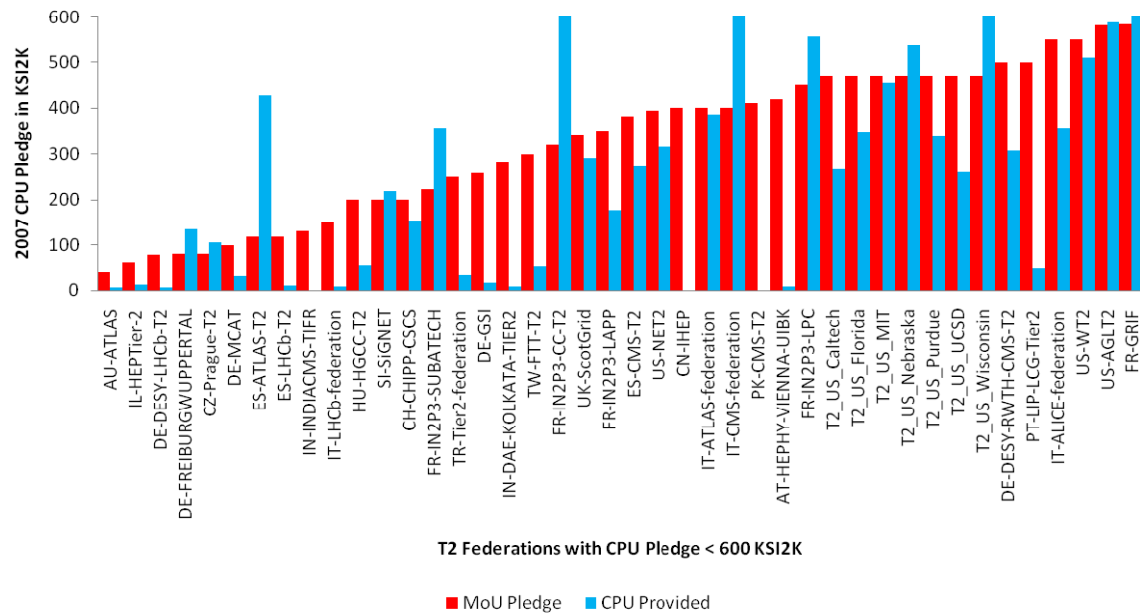
Resource Utilization Tier 2

- 52 of 57 federations are reporting
- 114 identified Tier 2 sites
- 107 publish accounting data

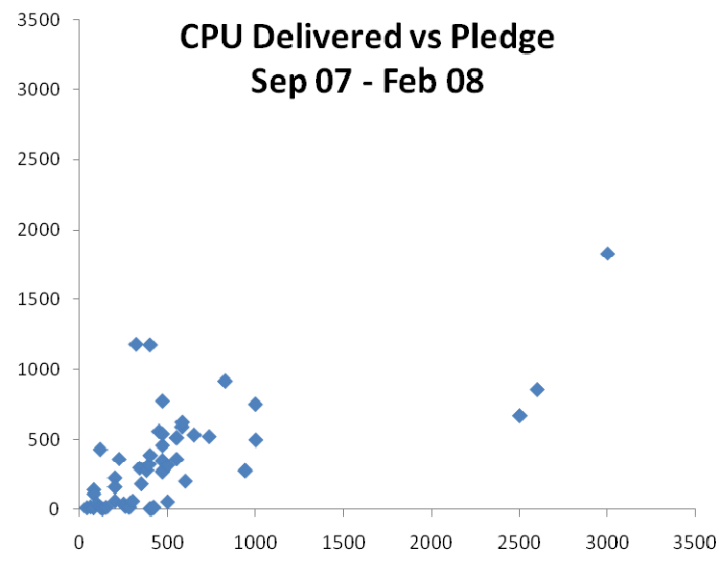
T2 Accounting September 2007-February 2008



T2 Accounting September 2007-February 2008



CPU Delivered vs Pledge
Sep 07 - Feb 08





Milestones - 24x7; VOBoxes

ID	Date	Milestone	ASGC	CC IN2P3	CERN	FZK GridKa	INFN CNAF	NDGF	PIC	RAL	SARA NIKHEF	TRIUMF	BNL	FNAL
24x7 Support														
WLCG-07-01	Feb 2007	24x7 Support Definition Definition of the levels of support and rules to follow, depending on the issue/alarm												
WLCG-07-02	Apr 2007	24x7 Support Tested Support and operation scenarios tested via realistic alarms and situations				Apr 2008	Apr 2008							
WLCG-07-03	Jun 2007	24x7 Support in Operations The sites provides 24x7 support to users as standard operations				Apr 2008	Apr 2008		Mar 2008		Apr 2008			
VOBoxes Support														
WLCG-07-04	Apr 2007	VOBoxes SLA Defined Sites propose and agree with the VO the level of support (upgrade, backup, restore, etc) of VOBoxes	Mar 2008	Apr 2008					Mar 2008					
WLCG-07-05	May 2007	VOBoxes SLA Implemented VOBoxes service implemented at the site according to the SLA	Apr 2008	Apr 2008	Mar 2008			Mar 2008	Mar 2008		Apr 2008			
WLCG-07-05b	Jul 2007	VOBoxes Support Accepted by the Experiments VOBoxes support level agreed by the experiments	ALICE	n/a					n/a			n/a	n/a	n/a
			ATLAS					n/a	n/a					n/a
			CMS					n/a			n/a	n/a	n/a	
			LHCb	n/a				n/a				n/a	n/a	n/a

Only 10 sites have tested their 24 X 7 support, and only 8 have put the support into operation

→ Understood after February experience – to be in place for May

Ian.Bird@cern.ch

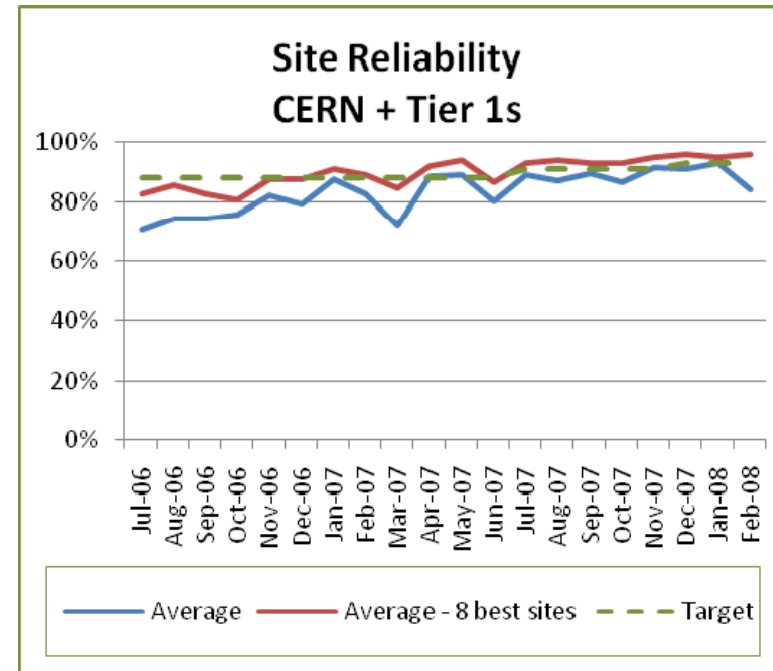
Only 6 sites have completed the set of VO BOX milestones

- some difficulty initially to specify an SLA –
- good models now exist



Tier 0/Tier 1 Site reliability

- Target:
 - Sites 91% & 93% from December
 - 8 best: 93% and 95% from December
- See QR for full status



	Sep 07	Oct 07	Nov 07	Dec 07	Jan 08	Feb 08
All	89%	86%	92%	87%	89%	84%
8 best	93%	93%	95%	95%	95%	96%
Above target (+>90% target)	7 + 2	5 + 4	9 + 2	6 + 4	7 + 3	7 + 3

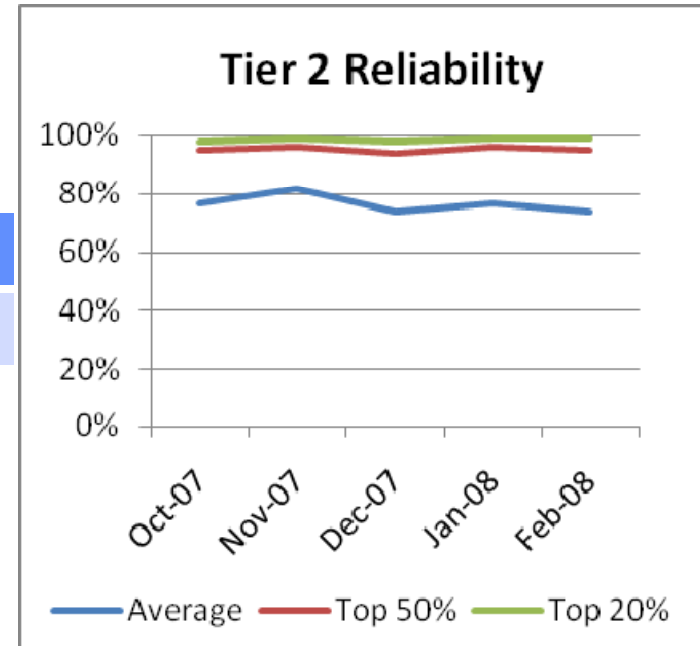
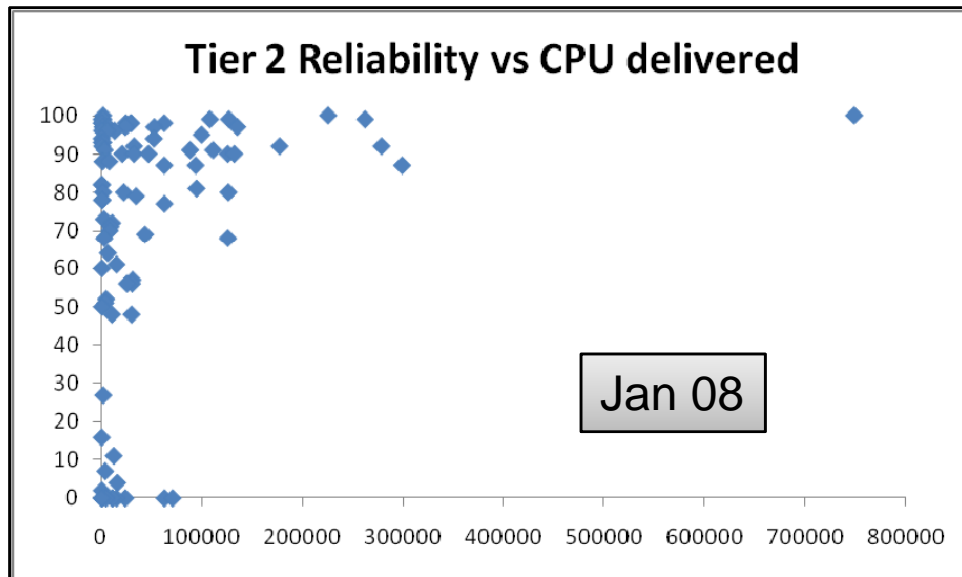


Tier 2 Reliabilities

- Reliabilities published regularly since October

Overall	Top 50%	Top 20%	Sites
76%	95%	99%	89→100

- In February 47 sites had > 90% reliability



- For the Tier 2 sites reporting:

Sites	Top 50%	Top 20%	Sites > 90%
%CPU	72%	40%	70%

- For Tier 2 sites not reporting, 12 are in top 20 for CPU delivered



Improving Reliability

- Monitoring
- Metrics
- Workshops
- Data challenges
- Experience
- Systematic problem analysis
- Priority from software developers



WLCG - Sites Reliability and Job Efficiency

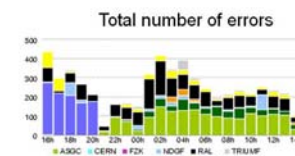
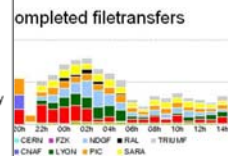
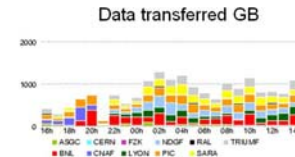
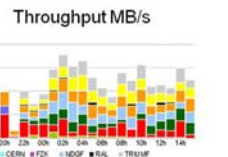
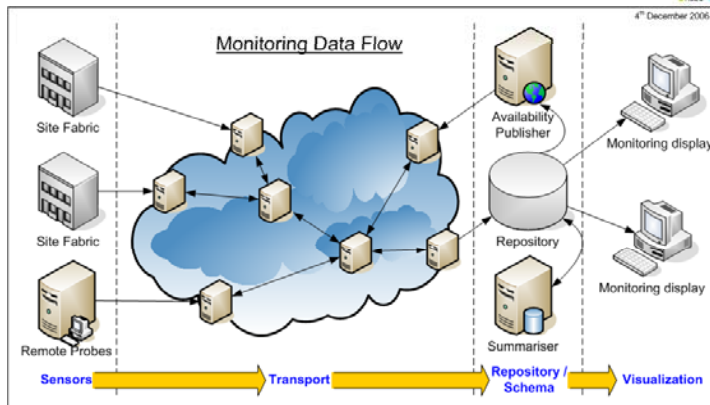
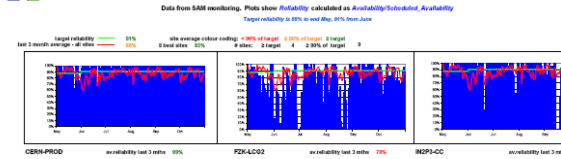
Site	ALICE			ATLAS			CMS		LHCb	
	SAM	SAM	AGENT	SAM	GANGA	PROD	SAM	CRAB	SAM	PI
ASGC	93%	-	-	98%	22%	82%	95%	90%	-	-
BNL	91%	-	-	72%	0%	0%	-	-	-	-
CERN	100%	97%	99%	100%	50%	92%	100%	76%	96%	S
CNAF	80%	97%	53%	85%	52%	74%	100%	97%	66%	S
FNAL	89%	-	-	-	-	-	38%	99%	-	-
FZK	91%	95%	96%	62%	73%	93%	99%	96%	91%	S
IN2P3	70%	45%	89%	26%	77%	79%	8%	99%	97%	S
NDGF	97%	0%	0%	76%	0%	84%	0%	0%	-	-
NIKHEF	92%	96%	100%	92%	45%	84%	53%	-	90%	19%
PIC	93%	-	-	100%	7%	61%	100%	100%	93%	88%
RAL	90%	96%	99%	100%	15%	93%	100%	90%	97%	90%
TRIUMF	95%	-	-	98%	4%	94%	-	-	-	-

>=91% >=82% <82%



Reliability of WLCG Tier-1 Sites + CERN

May 2007 - October 2007

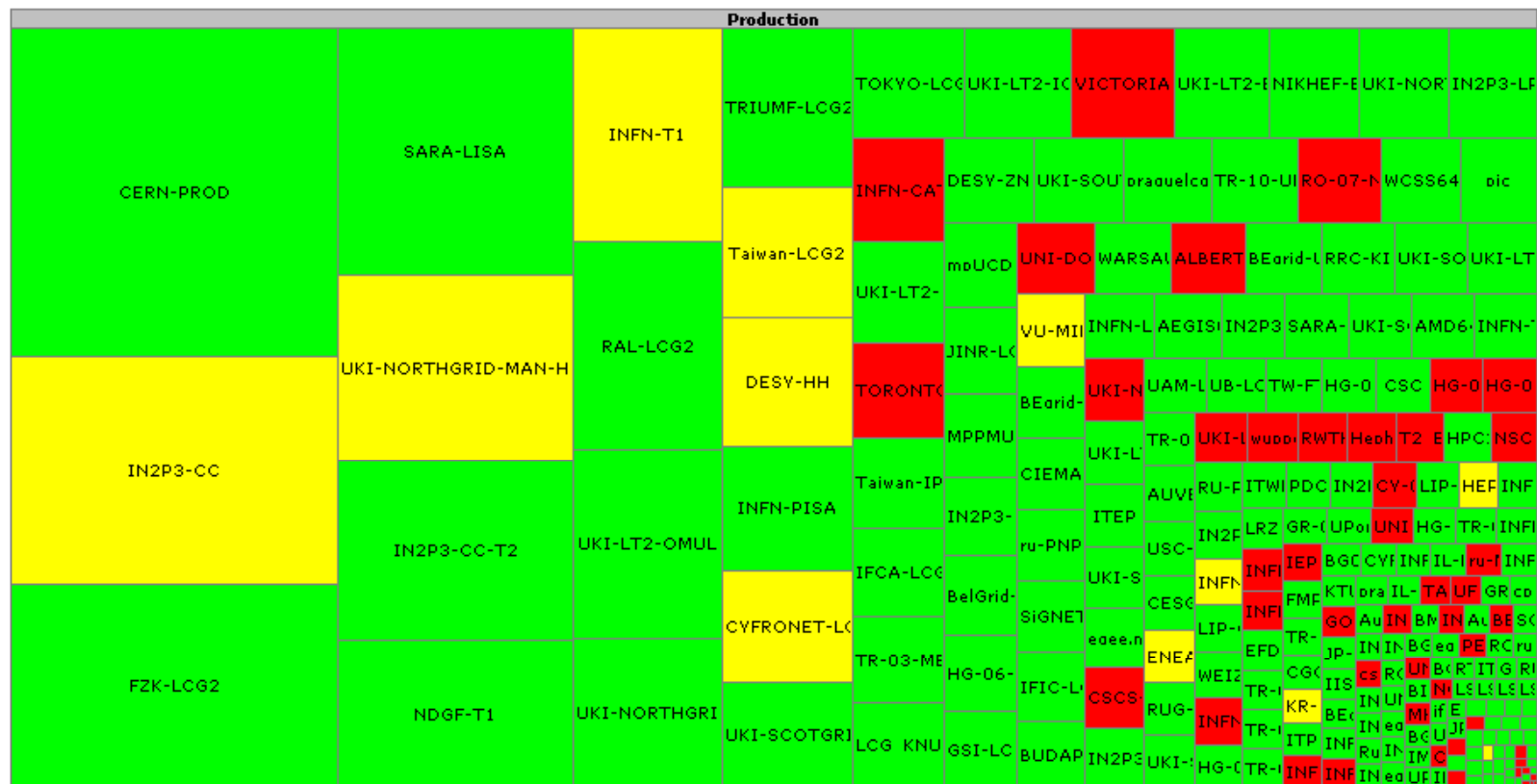


ATLAS M4 Data Monitoring - August 31



Gridmap

GridMap Prototype – Visualizing the "State" of the Grid



© CERN openlab / EDS

Site Status, for 'OPS' VO, 11 Oct 2007 18:00-18:59 GMT.
Size of site rectangles is number of CPUs from BDII.
All sites known by GStat having data in BDII.





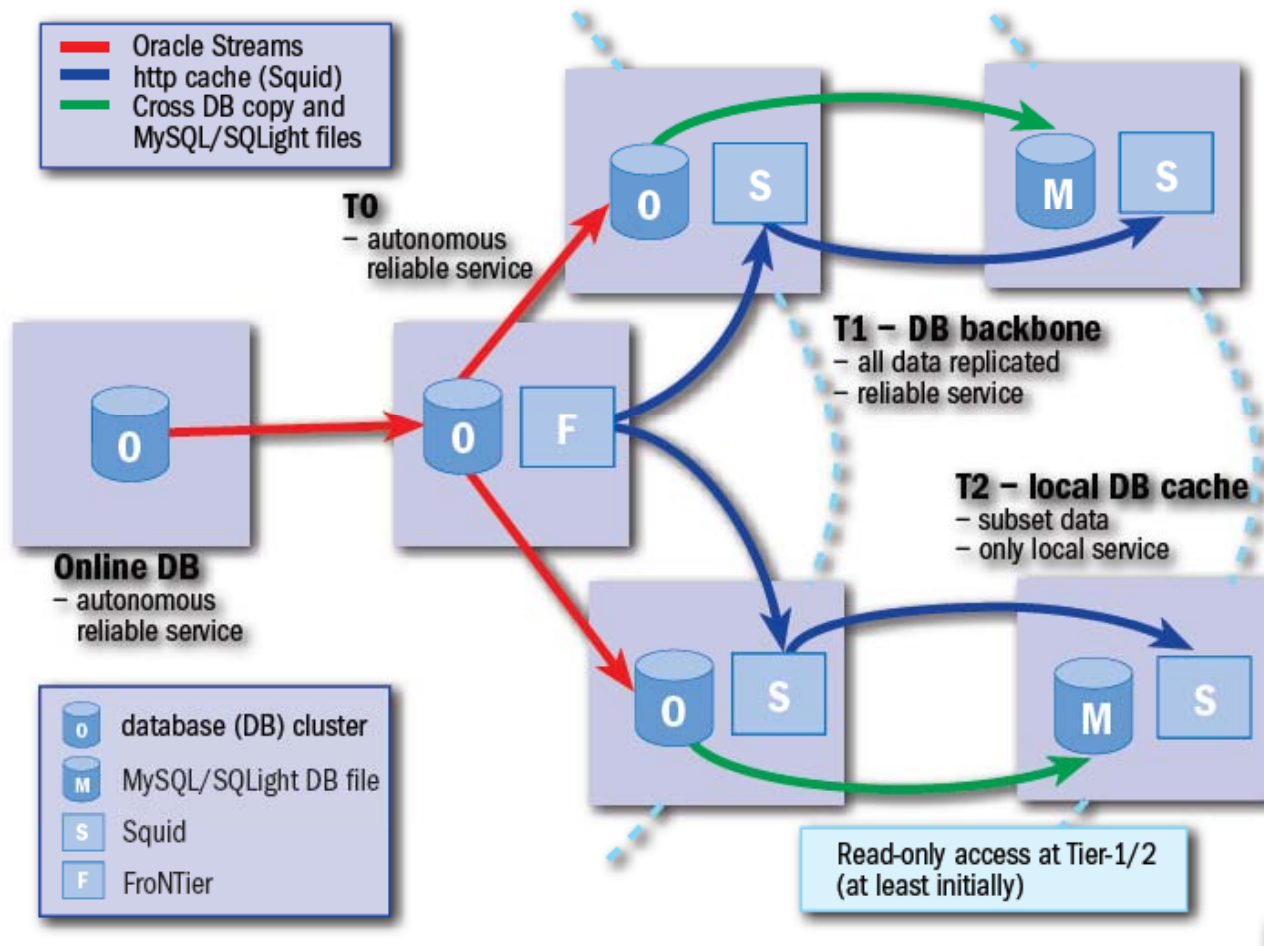
Middleware: Baseline Services

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

- Storage Element
 - Castor, dCache, DPM
 - Storm added in 2007
 - SRM 2.2 – de production –
- Basic transfer tool
- File Transfer Service
- LCG File Catalog (LFC)
- LCG data mgt tools – lcg-utils
- Posix I/O –
 - Grid File Access Library (GFAL)
- Synchronised databases T0 \leftrightarrow T1s
 - 3D project
- Information System
 - Scalability improvements
- Compute Elements
 - s/Condor-C –
 - vements to LCG-CE for reliability
 - services (CREAM)
 - rt for multi-user pilot (lexec, SCAS)
 - ad Management
 - in production
- VO Management System (VOMS)
- VO Boxes
- Application software installation
- Job Monitoring Tools

Focus now on continuing evolution of reliability, performance, functionality, requirements

Database replication



- In full production
 - Several GB/day user data can be sustained to all Tier 1s
- ~100 DB nodes at CERN and several 10's of nodes at Tier 1 sites
 - Very large distributed database deployment
- Used for several applications
 - Experiment calibration data; replicating (central, read-only) file catalogues



Applications Area

- Recent focus has been on major releases to be used for 2008 data taking:
 - QA process and nightly build system to improve release process
- Geant4 9.1 released in December
- ROOT 5.18 release in January
- Two data analysis simulation and computing projects in the PH R&D proposal (July 2007) (Whitepaper)
 - WP8-1 - Parallelization of software frameworks to exploit multi-core processors
 - Adaptation of experiment software to new generations of multi-core processors – essential for efficient utilisation of resources
 - WP9-1 - Portable analysis environment using virtualization technology
 - Study how to simplify the deployment of the complex software environments to distributed (grid) resources



Progress in EGEE-III

- EGEE-III now approved
 - Starts 1st May, 24 months duration (EGEE-II extended 1 month)
 - Objectives:
 - Support and expansion of production infrastructure
 - Preparation and planning for transition to EGI/NGI
 - Many WLCG partners benefit from EGEE funding, especially for grid operations: **effective staffing level is 20-25% less**
 - Many tools: accounting, reliability, operations management funded via EGEE
 - Important to plan on long term evolution of this
 - Funding for middleware development significantly reduced
 - Funding for specific application support (inc HEP) reduced
- Important for WLCG that we are able to rely on EGEE priorities on operations, management, scalability, reliability



Operations evolution

- Existing model of “central” management – while essential in getting to the point we are at now – is unsustainable in the long run
- Devolve the responsibility for operational oversight to the regions (regional, national operations teams):
 - We now begin to have the understanding and tools to facilitate this
 - Local (site) fabric monitoring should now get grid as well as local alarms – sites can respond directly without needing a central operator to spot a problem and open a ticket
 - Define critical tests (generic and VO-specific) that can generate alarms at a site
 - Tools and monitoring “architecture” can now start to support this
- Central project management tasks will simplify to gathering data relevant to the MoU
 - Accounting, reliability, responsiveness, etc.



Middleware in EGEE-III

- In EGEE-III the focus on middleware is the support of the foundation services
 - These map almost directly to the services WLCG relies on
 - Should include addressing the issues with these services exposed with large scale production
 - Should also address still missing services (SCAS, glexec, etc)
 - Should also address the issues of portability, interoperability, manageability, scalability, etc.
 - Little effort is available for new developments
 - (NB tools like SAM, accounting, monitoring etc. are part of Operations and not middleware)

European Grid Initiative

Goal:

- Long-term sustainability of grid infrastructures in Europe

Approach:

- Establishment of a new federated model bringing together NGIs to build the EGI Organisation

EGI Organisation:

- Coordination and operation of a common multi-national, multi-disciplinary Grid infrastructure
 - To enable and support international Grid-based collaboration
 - To provide support and added value to NGIs
 - To liaise with corresponding infrastructures outside Europe

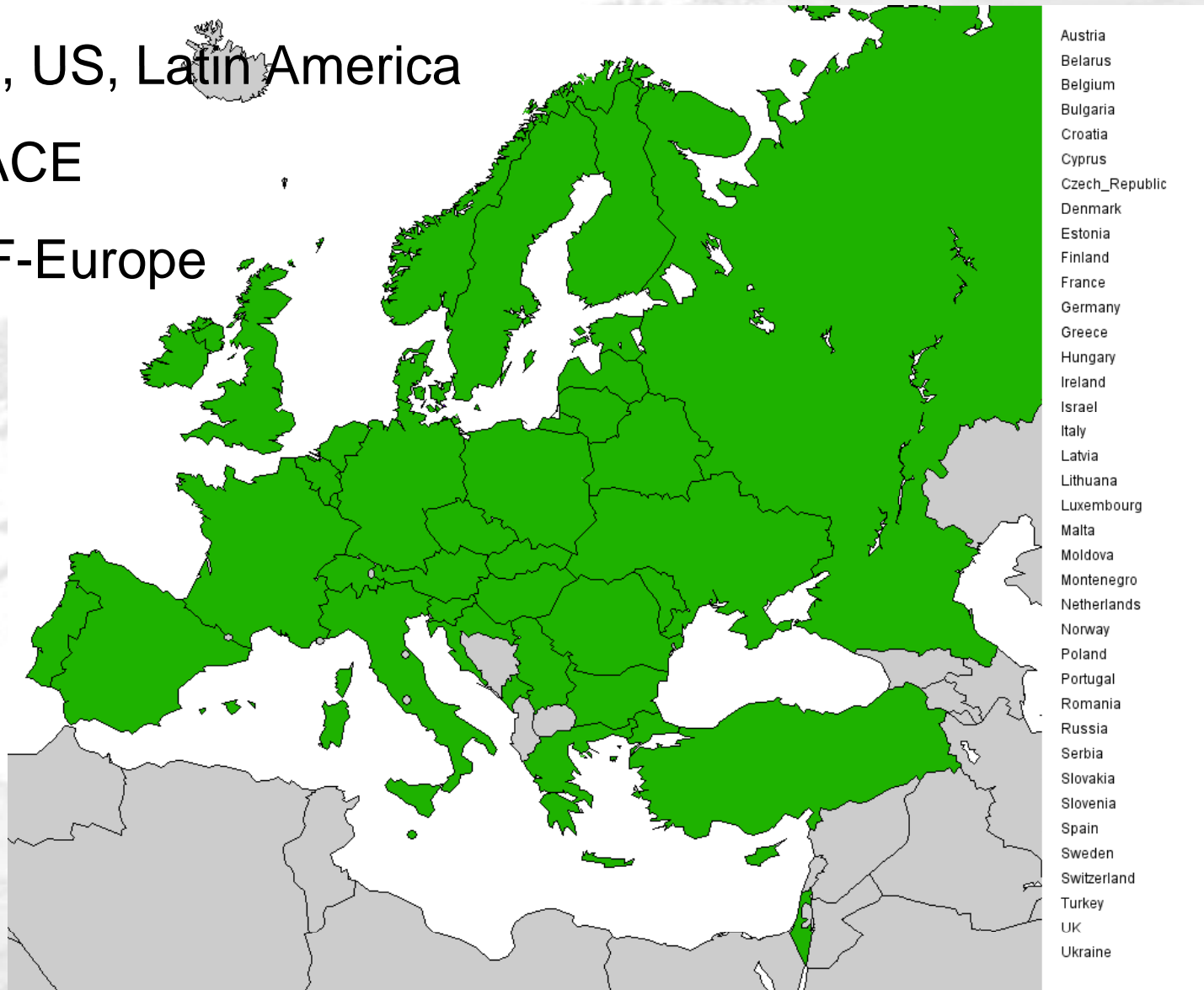
38 European NGIs

+ Asia, US, Latin America

+ PRACE

+ OGF-Europe

+ ...





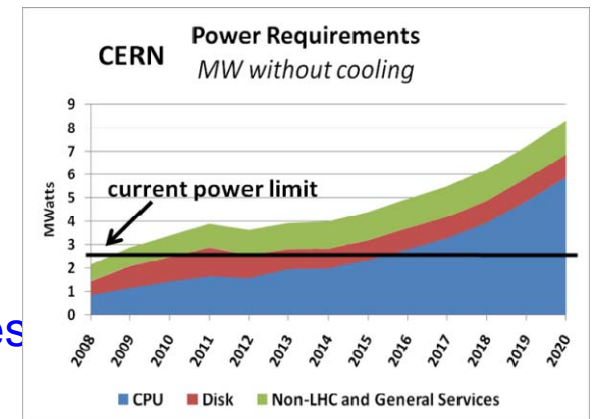
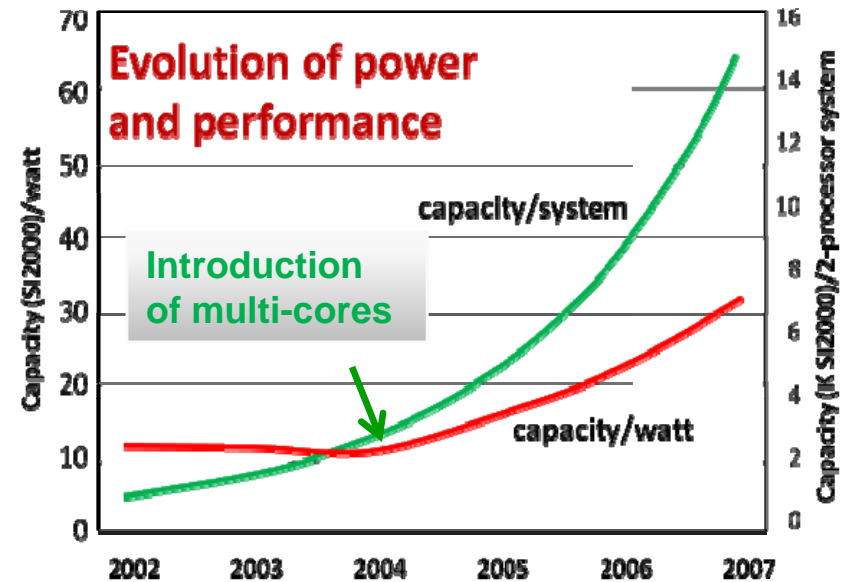
Comments on EGI design study

- Goal is to have a fairly complete blueprint in June
 - Main functions presented to NGIs in Rome workshop in March
- Essential for WLCG that EGI/NGI continue to provide support for the production infrastructure after EGEE-III
 - We need to see a clear transition and assurance of appropriate levels of support; Transition will be 2009-2010
 - Exactly the time that LHC services should not be disrupted
- Concerns:
 - NGIs agreed that a large European production-quality infrastructure is a goal
 - Not clear that there is agreement on the scope
 - Reluctance to accept level of functionality required
 - Tier 1 sites (and existing EGEE expertise) not well represented by many NGIs
- WLCG representatives must approach their NGI reps and ensure that EGI/NGIs provide the support we need



Power and infrastructure

- Expect power requirements to grow with capacity of CPU
 - This is not a smooth process: depends on new approaches and market-driven strategies (hard to predict) e.g. improvement in cores/chip is slowing; power supplies etc. already >90% efficient
 - No expectation to get back to earlier capacity/power growth rate
- e.g. Existing CERN Computer Centre will run out of power in 2010
 - Current usable capacity is 2.5MW
 - Given the present situation Tier 0 capacity will stagnate in 2010
- Major investments are needed for new Computer Centre infrastructure at CERN and major Tier 1 centres
 - IN2P3, RAL, FNAL, BNL, SLAC already have plans
 - IHEPCCC report to ICFA at DESY in Feb '08





Summary

- CCRC'08 phase 1 has been a success – open items have been addressed for May; for data taking preparation
 - work is still needed on configuration of the MSS systems together with experiments
 - Tuning of tape access with real use patterns – may require experiments to reconsider analysis patterns
- Resource ramp-up: based on experiences and problems with 2008 procurements
 - Must ensure in future years that allowance is made for delays and problems
 - Important that the yearly April schedules are met – to be ready for accelerator start ups
- WLCG – should influence the directions of the EGI Design study
 - Must ensure that we see a clear and appropriate strategy emerging that is fully supported by the NGIs
 - Must engage the NGI representatives in this



Summary

- We have an operating production quality grid infrastructure that:
 - Is in continuous use by all 4 experiments (and many other applications);
 - Is still growing in size – sites, resources (and still to finish ramp up for LHC start-up);
 - Demonstrates interoperability (and interoperation!) between 3 different grid infrastructures (EGEE, OSG, Nordugrid);
 - Is becoming more and more reliable;
 - **Is ready for LHC start up**
- For the future we must:
 - Learn how to reduce the effort required for operation;
 - Tackle upcoming issues of infrastructure (e.g. Power, cooling);
 - Manage migration of underlying infrastructures to longer term models;
 - Be ready to adapt the WLCG service to new ways of doing distributed computing.