



# WLCG Status Report

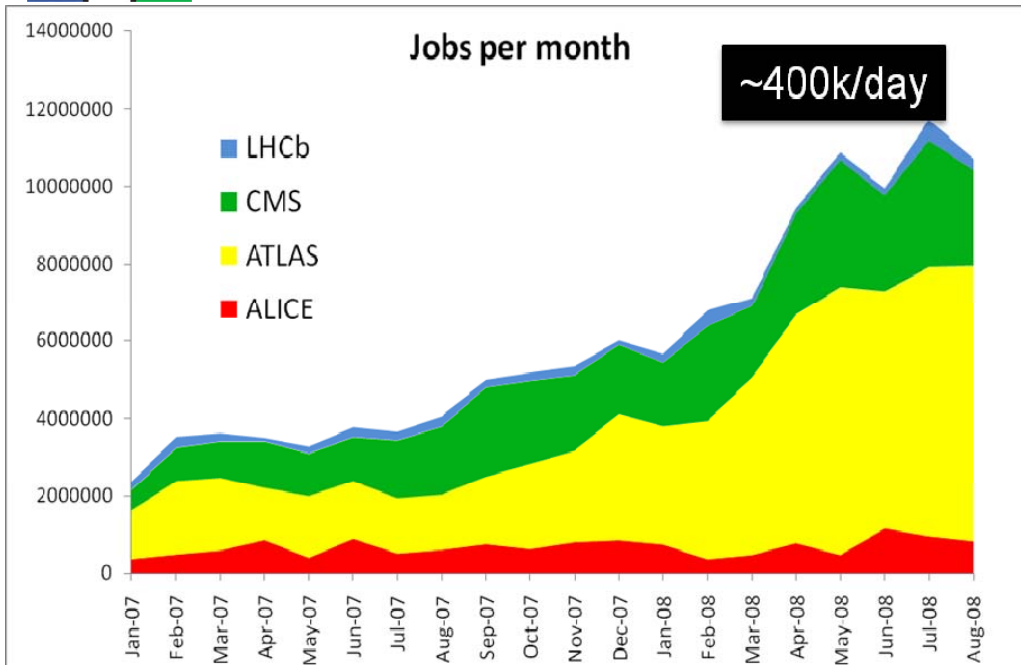
27<sup>th</sup> October, 2008  
Project Overview  
Board

Ian Bird  
LCG Project Leader

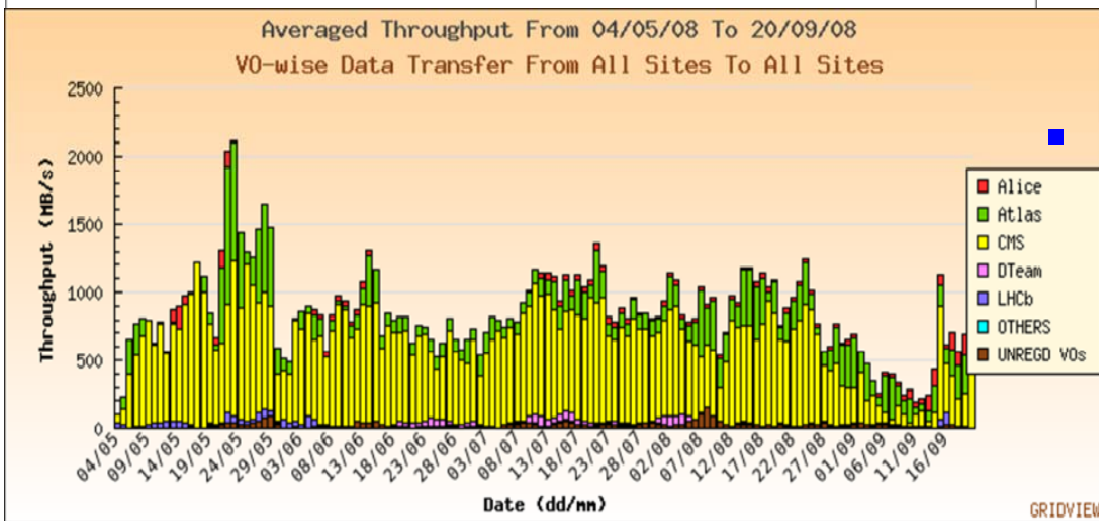




# LCG Service



- Continued to run after CCRC'08
- Loads continued to increase
- Maintained data transfers at more or less significant levels
- Experiments run simulations, cosmic data, etc.
- Service incidents → Jamie's talk



- Support levels – concern for lack of available staff during vacations – will this be an issue in future
  - Esp. Storage systems



# Procurement

ID	Date	Milestone	ASGC	CC IN2P3	CERN	DE-KIT	INFN CNAF	NDGF	PIC	RAL	SARA NIKHE F	TRIUM F	BNL	FNAL
WLCG-07-17	1 Apr 2008	MoU 2008 Pledges Installed To fulfill the agreement that all sites procure they MoU pledged by April of every year	Sept 2008	CPU OK May Disk Sep 08	Jul 2008	Apr 2008	CPU Jul 08 Disk Sept 08	CPU OK May Disk Sep 08	CPU OK May Disk	Apr 2008	Nov 2008	Apr 2008	CPU OK Disk Nov 08	CPU OK May Disk
WLCG-08-04	Sep 2008	Sites Report on the Status of the MoU 2009 Procurement Reporting whether is on track with the MoU pledges by April. Or which is the date when the pledges will be fulfilled.		Tender Sept Jan Install May	Tender Sept Dec Install Apr	Tender Sept Oct Install Apr	Tender Sept Install May	Tender Sept Install Apr	Tender Oct Install Apr	Tender CPU Sep Disk Oct	Tender Sept Install TBD	Tender CPU Disk Oct	Tender CPU Sep Disk Oct	Tender Sep Install Apr
WLCG-08-05	1 Apr 2009	MoU 2009 Pledges Installed To fulfill the agreement that all sites procure they MoU pledged by April of every year												

## ■ CPU:

- ASGC: 72% installed, expect to install the remainder in October
- CNAF: 57% installed, installation of remainder ongoing
- NL-T1: 88% installed

## ■ Disk:

- ASGC: 300 TB missing (20%)
- BNL: 1 PB missing, November with new machine room
- IN2P3: 700 TB missing, ongoing, together with 50% of 2009 capacity
- CNAF: 750 TB missing (60%), delivery complete and installation ongoing
- NDGF: 200 TB missing, procurement is complete, installation ongoing
- NL-T1: 1400 TB missing (56%): lack of available power and cooling; no new estimate yet, but not before 2009.

# Outstanding Milestones

17-Oct-08		WLCG High Level Milestones - 2007													
ID	Date	Milestone	Done (green)				Late < 1 month (orange)				Late > 1 month (red)				
			ASGC	CC IN2P3	CERN	DE-KIT	INFN CNAF	NDGF	PIC	RAL	SARA NIKHE F	TRIUM F	BNL	FNAL	
<b>24x7 Support</b>															
WLCG-07-01	Feb 2007	24x7 Support Definition Definition of the levels of support and rules to follow,													
WLCG-07-02	Apr 2007					Apr 2008	June 2008								
WLCG-07-03	Jul 2007	The sites provides 24x7 support to users as standard operations				July 2008	June 2008		Apr 2008		July 2008				
<b>VOBoxes Support</b>															
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	Aug 2008	Aug 2008					Aug 2008						
WLCG-07-05	May 2007	VOBoxes SLA Implemented VOBoxes service implemented at the site according to the SLA	Aug 2008	Aug 2008				Mar 2008	Aug 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

After-hours support now tested in practice – still some tuning to be done

## VOBox SLAs –

- mainly awaiting experiment sign-off
- NDGF + NL still have work to do to complete this



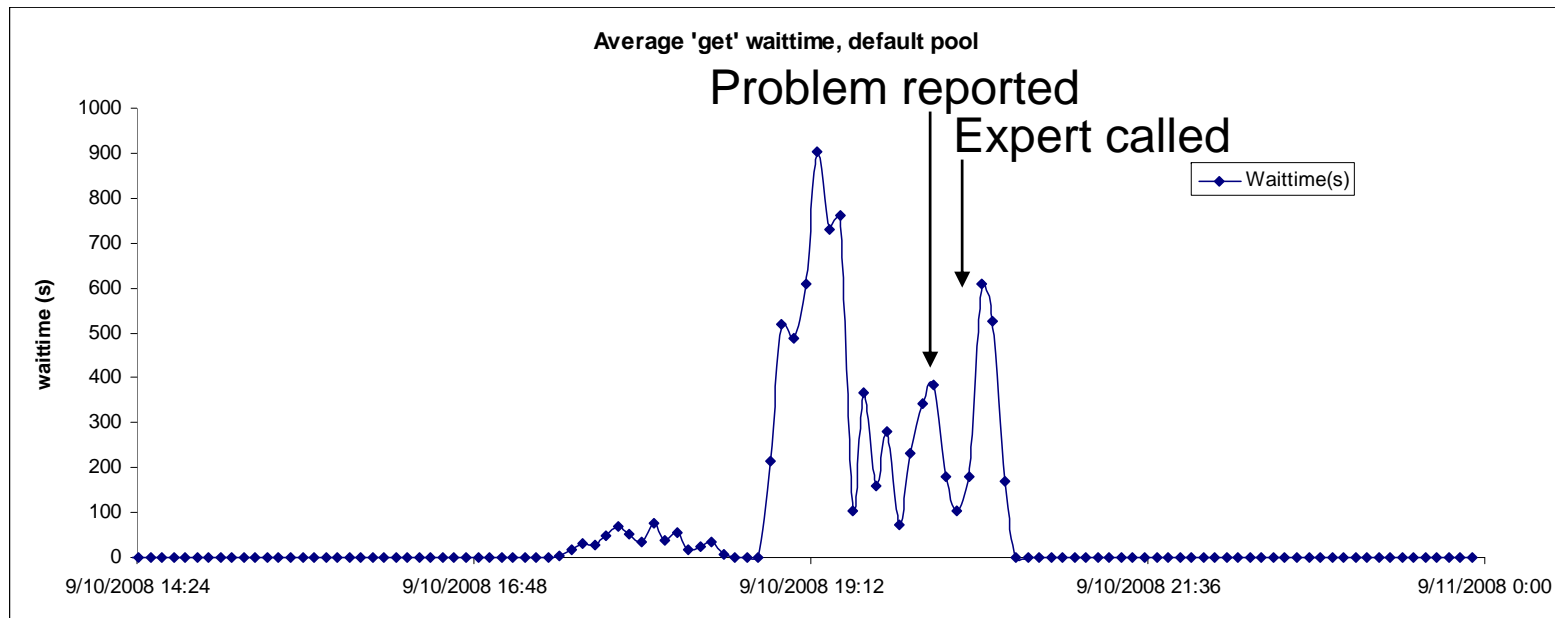
# Tier 0 status -1

- 2008:
  - CPU: 100% available since May
  - Disk: 100% available in August
    - Delayed due to late delivery and minor technical problems with replacement hardware ordered after the supplier bankruptcy in April
  - Tape: 100% available since April
- 2009:
  - Tendering complete
  - CPU: 1<sup>st</sup> part December/Jan; 2<sup>nd</sup> part March/April
  - Disk: 1<sup>st</sup> part January; 2<sup>nd</sup> part March/April



# Tier 0: First LHC data

- No CASTOR issues but ...
  - ATLAS: file(s) unavailable for transfer



- LHCb: raw data file(s) 'disappeared'
  - Files were accidentally deleted from CASTOR by LHCb



# Tier 0: Performance metrics

- Metrics have been implemented and deployed on preproduction cluster
  - Data collected in lemon
  - RRD graphs not yet implemented
- Production deployment delayed for several reasons
  - New metrics imply several changes to exception/alarms and automated actions used in production
  - An unexpected technical dependency on the late SRM 2.7 version
    - Ongoing work to back-port the implementation



# Tier 1 status - Overview

- Per Site:
  - Deployment of 2008 Resources
  - Forward look to 2009 procurements
  - Status of 24 by 7 Services Support – the sites provide 24x7 support to users as standard operations: **completed by all sites**
  - Status of VO-boxes Support (**note all ATLAS VOboxes are at CERN except for BNL**)
- Status as of end September →





## Tier 1 Status (1/4)

- **ASGC: CPU at 70%, disk at 80%, tape at 60%**
  - Will meet 2008 pledges November (and exceed in CPU).
  - 2009 CPU (most of) already in Nov 2008, storage on April target.
  - VObox SLA defined and in place and needs to be signed-off by the experiments.
- **CC-IN2P3: CPU at 100%, disk at 60%, tape at 100%**
  - Expecting rest of 2008 disk and half of 2009 disk to be delivered last week.
  - 2009 launching the purchasing framework. Part will be available by the end of this year, remaining orders to be made when budget is known.
  - VObox SLA ready to go to experiments for approval before implementation. Will be completed in next few weeks.
- **CERN: All 2008 pledges available**
  - 2009 pledges on track for April 2009 availability.
  - VObox support is defacto available but needs to be officially approved.



## Tier 1 Status (2/4)

- **DE-KIT: CPU at 80%, disk at 80%, tape at 70%**
  - Remainder of 2008 pledge in October as synchronised with experiments (mostly ALICE) requirements.
  - 2009 CPU orders sent and disk order to be made soon. Tape libraries big enough for 2009 – will order media and drives. Expect to meet April target.
  - VObox support complete.
- **INFN/CNAF: CPU at 60%, disk at 40%, tape at 40%**
  - Rest of 2008 pledges in place and starting to be installed.
  - 2009 tenders start end September for end February delivery so will be tight to have all ready for April deadline.
  - VObox support complete.
- **NDGF: CPU at 120%, disk at 80%, tape at 35%**
  - Have all the storage and should be installed this month.
  - 2009 will have 2 more sites and expect to be on target.
  - Only ALICE have VOboxes in NDGF. These need detailed description of their many functions for the SLA.

# Tier 1 Status (3/4)



- PIC: All 2008 pledges available
  - 2009 planning awaited
  - VObox SLA in place and the CMS contact persons need to be defined.
- RAL: All 2008 pledges available
  - 2009 Disk tender closing now and CPU tender in October. New building available September so on target for April 2009.
  - VObox support complete.
- NL-T1: CPU at 60%, disk at 20%, tape at 10%
  - 2008 disk finally passed acceptance and being configured. All 2008 pledges by November (tape media quickly added as/when needed).
  - 2009 framework tenders in place. Possible to be ready by April but needs additional space/power/cooling currently under acquisition to be on time.
  - SLA document not finalised yet. Needs both NIKHEF and SARA to approve.

# Tier 1 Status (4/4)



- TRIUMF: All 2008 pledges available
  - 2009 procurements to be made early October for February delivery so should be on target for April 2009.
  - No local VOboxes.
- US-ATLAS: CPU at 100%, disk at 70%, tape at 100%
  - New power and cooling nearly ready so should fulfill 2008 disk pledge in October/November (adding 1 PB).
  - 2009 specifications are ready. Next comes discussions with US-ATLAS funding agency.
  - VObox support complete.
- US-CMS: All 2008 pledges available.
  - Need to verify 2009 disk order will be in place then will be on target for April 2009.
  - VObox support complete.



# ATLAS progress

WLCG MB - 14 October 2008

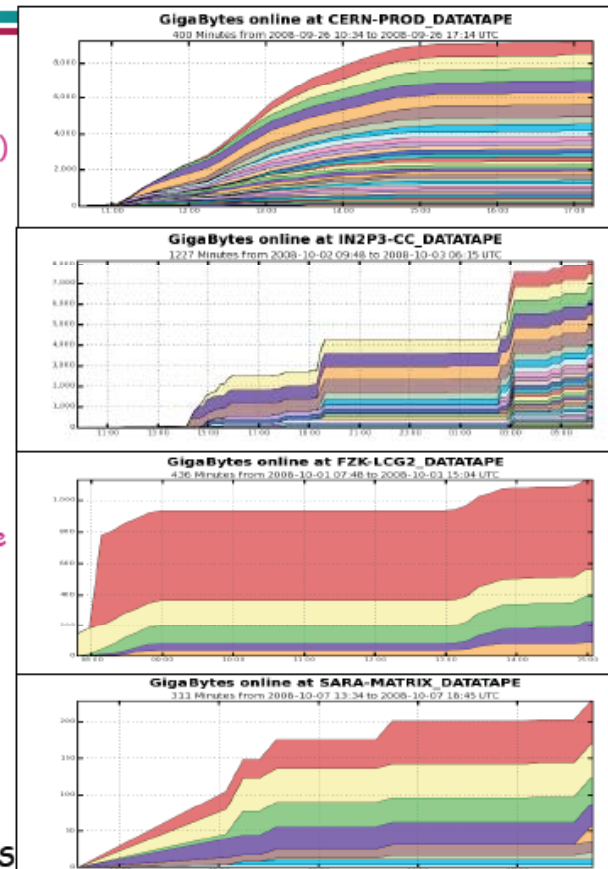
WLCG MB - 14 October 2008

## Prestaging tests

- We are ray dat until 3r
- Wi det
- The Ti data r
- A f sor wit
- In Nov commis
- Bu on

- We started during the summer prestaging tests at all Tier-1s
  - Recalling whole datasets at a time (up to 10 TB)
- Performance varies a lot as tape back-ends are different at each site
  - After a few tries, most sites are mostly OK
- Outstanding (different) problems at PIC, FZK and SARA
  - This exercise also showed that the number of available tape drives varies a lot from site to site
  - There is no point in having 1000s of processing cores if they cannot be fed at the correct rate with data
  - Example:
    - Our reprocessing tasks consume 1.6 MB of raw data every ~7 real seconds
    - One needs a total read rate from tape of 400-500 MB/s to keep 1000 cores busy
      - Including x2 contingency

Dario Barberis: ATLAS







- **Commissioning of DIRAC3**
  - **Fully reengineered system**
  - **Main features:**
    - ↳ **Single framework for services, clients and agents**
    - ↳ **Fully integrated Workload and Data Management Systems**
    - ↳ **Supports production and user analysis activities**
      - ↳ Allow to apply VO policy: priorities, quotas...
    - ↳ **Uses pilot jobs as DIRAC2**
      - ↳ Ready for using generic pilot jobs (not switched on yet)
      - ↳ Full scale test with generic pilots will take place in the coming weeks
    - ↳ **New bookkeeping system (also integrated)**
- **Production activities**
  - **Complete simulation and stripping of MC data (so-called DC06 as was launched in 2006)**
  - **CCRC-like activity at low rate (10%)**
  - **Start 2008 simulation**
    - ↳ **Mainly for alignment and calibration studies**
    - ↳ **Wait for first data for tuning generators and detector response**



## Recent issues encountered

- **Storage**
  - **Instability of SEs, in particular dCache**
    - ↳ Very good response from sites and dCache developers
    - ↳ Permanent struggle due to various causes:
      - ↳ Software issues (addressed with sites and developers)
      - ↳ Sub-optimal hardware configuration at some Tier1's
      - ↳ Unavailability of files: are in the namespace at site but cannot be accessed or even get a tURL
        - Damaged tapes, unavailable servers...
  - **Transfers are OK (low throughput needed: 70 MB/s)**
- **Workload Management**
  - **Three severe issues with WMS**
    - ↳ Mixing up credentials of jobs submitted by the same user with different roles
    - ↳ Limitation in proxy handling (too few delegations allowed) preventing some users to run jobs (e.g. from French CA)
    - ↳ Misbehavior of WMS after some idle time: cannot find suitable sites even for a job without requirements!
- **Issues with local shared software repository at sites**
  - **Stability and access rights** (being addressed)



# ALICE

- Monte Carlo
  - Continuous production in prevision of 2009 data taking: production in T1/T2, data and end user analysis in T2
- Analysis
  - CAF (fast), analysis train (organized) and end user analysis (chaotic) operational
- Raw data
  - Online production of condition parameters, first pass processing @ T0, replication in T1s, N pass processing @ T1s operational
- Software
  - Stable release ready for data taking; code evaluation and some refactoring to be done before LHC start...
- Services
  - New AliEn version deployed routinely with effectively no downtime
  - Job management in all its form: RB (phased out but still widely used), WMS, CREAM (very promising initial stability and scaling tests)
  - xrootd-enabled SE continuous deployment (T2s are a concern)
- Accounting
  - Used 40% of allocated CPU and 53% of required
  - 27% of pledged storage is operational and 64% of that is used
- Resources
  - Requirements for 2008/2009 had been reevaluated (before LHC incident)
  - New requirements for 2009, depending on LHC scedule; expect larger requirements with respect to C-TDR (CPU, disk)



# User Analysis - new working group

- Clarify roles of Tier 1, 2, 3 for user analysis for each experiment
  - Batch Analysis – the assumption is that this is grid based
  - End-user Analysis – what is this? – laptops, local analyses, ...?
  - → what are the distinctions?
- The presentations to the GDB on 8/10/08 are a good summary of the analysis models:
  - <http://indico.cern.ch/conferenceDisplay.py?confId=20234>
- It is important to understand what is missing to allow the experiments to implement and manage these models. This includes correct configuration of shares in batch systems, appropriate set ups of disk pools, etc. and a summary of which tools may be needed to implement the models. What is the set of services that are needed (in addition to what exists now) to support these analysis models?
  - NOT an open door for new service development requirements!
- What is a “standard model” for a Tier 2 supporting analysis?
  - E.g. File systems (Lustre, NFS4, gpfs, AFS, xrootd) with what interface, etc.?
  - How many users? Size per user, etc. How to scale the resources?

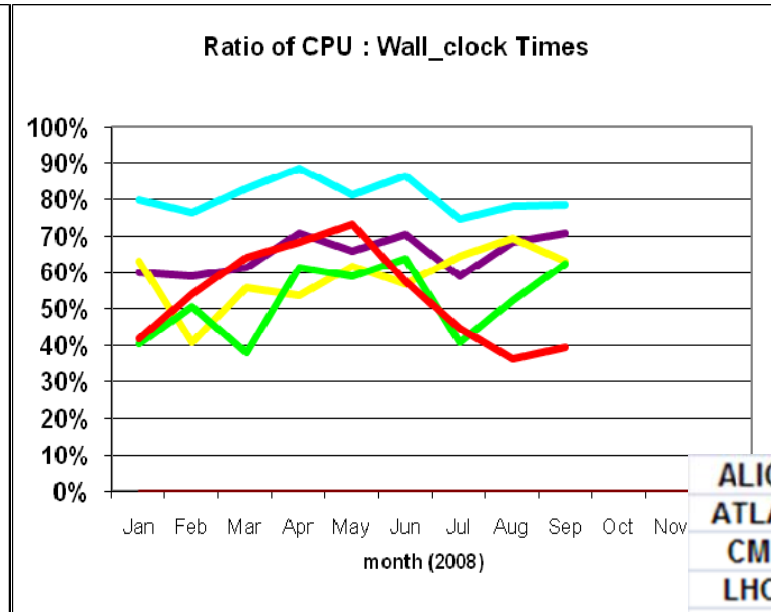
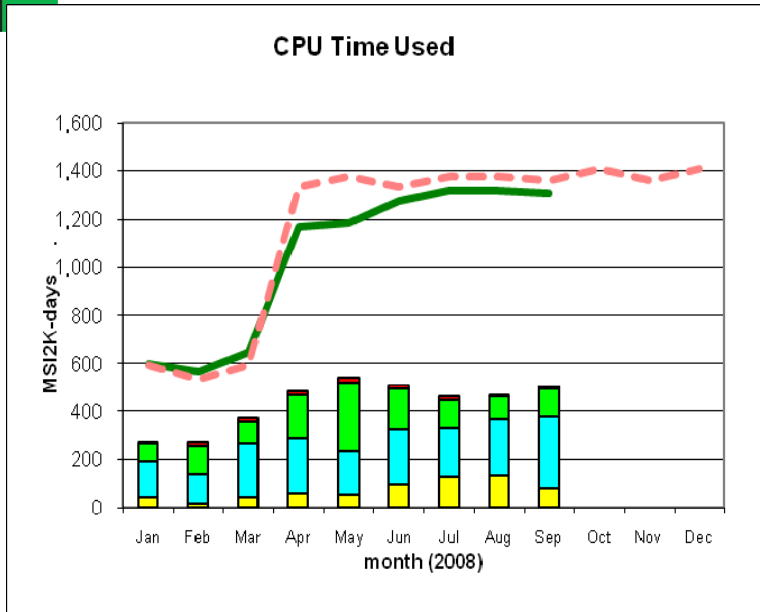


# Mandate

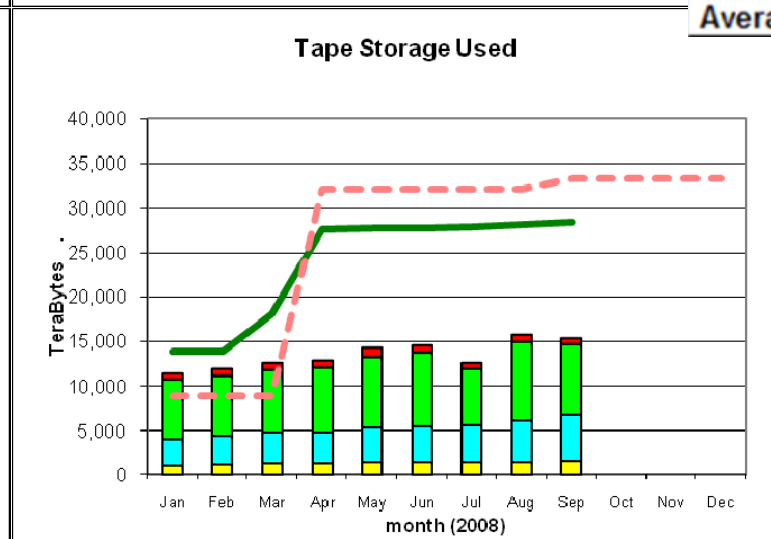
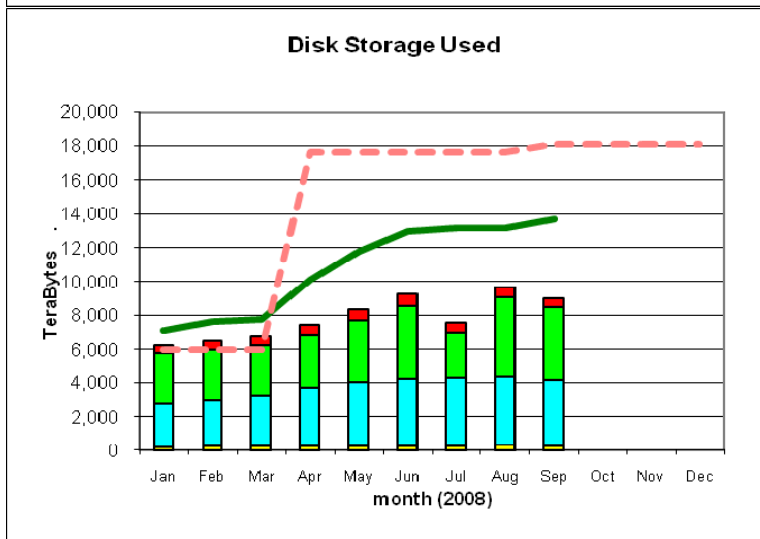
- Clarify and confirm the roles of the Tier 1, 2, 3 sites for analysis activities for each experiment; categorising the various types of analysis: batch or interactive, centrally organised or individual users.
- List the minimal set of services, their characteristics (availability, scale, concurrent usage, users, etc.), or changes to existing services, essential to implement those models. These must be prioritised clearly between “essential” and “desirable”. The expectation is that this is addressing configuration of services, recommendations for service implementation, or new tools to assist managing or using the services rather than requests for new services. Such developments must be clearly justified.
- Deliverables:
  - Documented analysis models
  - Report on requirements for services and developments with priorities and timescales.



# Resources: CERN + Tier1s



ALICE	Yellow
ATLAS	Cyan
CMS	Green
LHCb	Red
Average	Purple





# Pledges/requirements

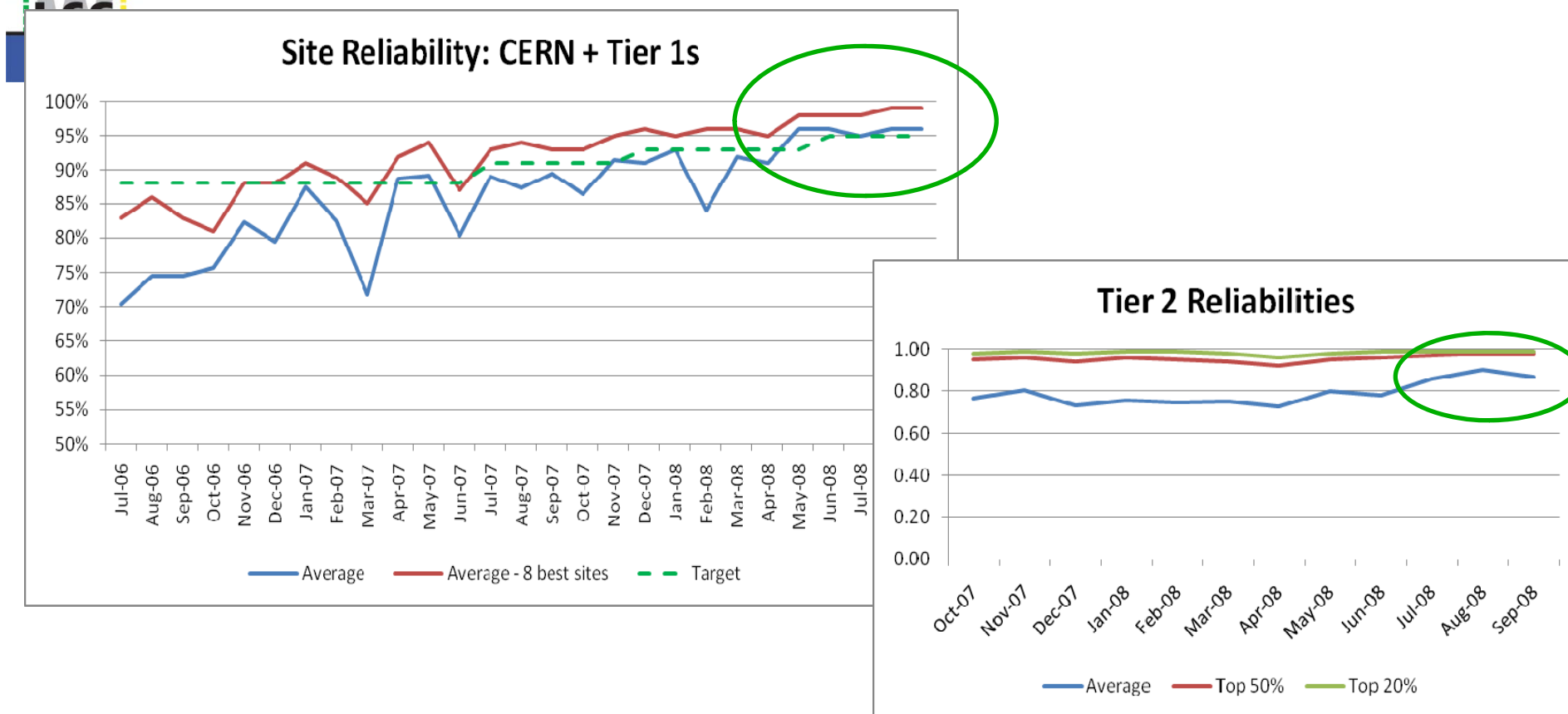
- To be raised at C-RRB:
  - 5 year outlook is unrealistic for experiment requirements and for funding pledges
    - See this in lack of responses for plans (previous year's numbers)
  - Propose to move to a 3 year outlook
- Pledge cycle is too late to affect procurements
  - Today Autumn RRB confirms pledges for next year
  - Procurements should be well in hand by this time – can we move this forward to Spring RRB? Or do we just live with uncertainty?
  - Scrutiny process needs to look +1 year in advance; i.e. Ideally we should have 2010 needs scrutinised by early next year in order to have any effect on procurements
- First discussion in MB on splitting (disk) procurements so that delivery/installation is done in 2 pieces (April + late summer)
  - Less load on testing/installation staff; better cost optimisation (?)



# Benchmarks

- HEPiX group has provided results and guidance
- In MB have agreed, to use the SPEC2006 C++ benchmark suite
  - Mix of int + fp close to HEP use
  - Can be run simply – but in 6 hrs instead of 24
  - No published results for this test – so vendors will be forced to run it
  - Sites must buy a licence
- MB has small team led by Gonzalo Merino working to:
  - Define exact recipe for implementing the benchmark
  - Define the conversion from SI2K to the new units for pledges and requirements

# CERN+Tier 1 reliabilities



Improvement during CCRC and later is encouraging

- Tests do not show full picture – e.g. Hide experiment-specific issues,
- “OR” of service instances probably too simplistic

- a) publish VO-specific tests regularly;
- b) rethink algorithm for combining service instances

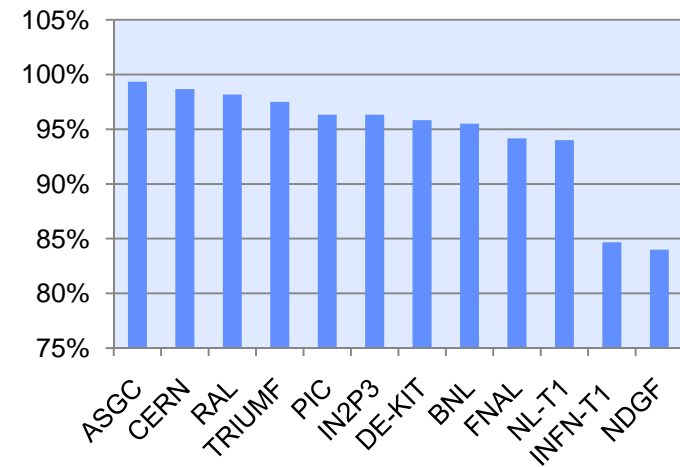


# Tier 1 reliabilities

Detailed Monthly Site Reliability

Site	Apr 08	May 08	Jun 08	Jul 08	Aug 08	Sept 08
CA-TRIUMF	96	98	98	98	99	96
CERN	95	100	98	99	100	100
DE-KIT (FZK)	95	97	98	96	99	90
ES-PIC	94	99	99	99	99	95
FR-CCIN2P3	98	97	96	94	95	98
IT-INFN-CNAF	76	88	86	79	99	82
NDGF	84	96	96	88	43	97
NL-T1(NIKHEF)	90	95	98	91	96	94
TW-ASGC	97	99	100	100	100	100
UK-T1-RAL	93	98	99	99	100	100
US-FNAL-CMS	92	96	93	100	99	100
US-T1-BNL	93	94	95	96	95	100
Target	93	93	95	95	95	95
Above Target (+ > 90% Target)	7 +3	11 +1	10 +1	8 +2	11 +1	9 +2

Last 6 months: average



Colors: Green > Target, Orange > 90% Target, Red > 90% Target





# VO-specific testing ...



VO-wise Availability and Reliability for WLCG Tier-1s + CERN

August 2008

$$\text{Reliability} = \text{time\_site\_is\_available} / (\text{total\_time} - \text{time\_site\_is\_scheduled\_down})$$

$$\text{Availability} = \text{time\_site\_is\_available} / \text{total\_time}$$

## Reliability

Site	ALICE	ATLAS	CMS	LHCb	OPS
CA-TRIUMF	N/A	95 %	N/A	N/A	99 %
CERN	99 %	99 %	100 %	30 %	100 %
DE-KIT	100 %	98 %	100 %	68 %	99 %
ES-PIC	N/A	98 %	99 %	52 %	99 %
FR-CCIN2P3	94 %	81 %	99 %	88 %	95 %
IT-INFN-CNAF	100 %	100 %	100 %	80 %	99 %
NDGF	100 %	86 %	N/A	N/A	43 %
NL-T1	99 %	87 %	N/A	0 %	96 %
TW-ASGC	N/A	99 %	98 %	N/A	100 %
UK-T1-RAL	100 %	63 %	97 %	88 %	100 %
US-FNAL-CMS	N/A	N/A	100 %	N/A	99 %
US-T1-BNL	N/A	N/A	N/A	N/A	95 %

Draft



# Consequences of LHC shutdown

- The present shutdown of the LHC has a number of consequences for the planning of WLCG:
  - Capacities and Procurements for 2009
  - Software and service upgrades during the shutdown
  - (Re-)Validation of services for 2009 following changes



# Capacities and procurements

- The WLCG MB has agreed that with the information currently available to us and the present understanding of the accelerator schedule for 2009:
  - The amount of data gathered in 2009 is likely to be at least at the level originally planned, with pressure to run for as long a period as possible this may be close to or exceed the amount originally anticipated in 2008 + 2009 together
  - The original planning meant that the capacity to be installed in 2009 was still close to x2 with respect to 2008 as part of the initial ramp up of WLCG capacity
  - Many procurement and acceptance problems arose in 2008 which meant that the 2008 capacities were very late in being installed; there is a grave concern that such problems will continue with the 2009 procurements
  - The 2009 procurement processes should have been well advanced by the time of the LHC problem in September
- The WLCG MB thus does not regard the present situation as a reason to delay the 2009 procurements, and we urge the sites and funding agencies to proceed as planned. It is essential that adequate resources are available to support the first years of LHC data taking.



# Upgrade plans

- Since several software upgrades were postponed in anticipation of LHC start-up. We propose that the following changes are addressed in the coming months:
  - See the following →

# Middleware planning



## • FTS/SL4

- This was postponed and will now be deployed. Has been tested extensively.

## • WN/SL5

- Already have a 1<sup>st</sup> installation at CERN, to be tested by experiments.
- Target - available on the infrastructure in parallel to SL4

## • glxexec/SCAS

- Target - enabling of multi-user pilot jobs via glxexec. SCAS currently in testing. Essential for analysis use cases with pilot jobs.

## • CREAM

- Here we should be more aggressive: LCG-CE problematic for analysis with many users)
- If the use case is direct submission with no proxy renewal, CREAM is basically ready. Proxy renewal should be fixed in the simplest possible way (reproduce the lcg-CE solution)
- WMS submission will come with ICE, timescale months
- Target - maximum availability in parallel with lcg-CE

# Middleware Planning



## • WMS

- Status: Patched WMS ( fixing issues with proxy delegation) to be deployed now
- ICE to submit to CREAM
  - Not required for certification of CREAM
  - ICE will be added in a subsequent update (but better before Feb. 2009)

## • Multiple parallel versions of middleware available on the WN

- Status - at the moment it is not easy to install or use multiple parallel versions of the middleware at a site. While the multi middleware versions and multi compiler support are not disruptive, they require some changes on the packaging side and a small adaptation on the user side.
- Target - it seems advisable to introduce this relatively shortly after the bare bone WN on SL5.



# Middleware Planning

- Other anticipated upgrades:
  - Glue2 - deploy in parallel - provides better description of resources
  - CE publishing
    - Better description of heterogeneous clusters
    - gridftp2 patches
      - These are being back ported to VDT1.6 ; Important for dCache and FTS
- Sites to install 64-bit OS by default with compatibility libraries
- SRM changes/updates: agreed programme



## Re-validation of the service

- All experiments are continually running simulations, cosmics, specific tests (and have been since CCRC'08) at high workload levels – this will continue
- A full CCRC'09 in the same mode as 2008 is not regarded as useful
- But, we will perform specific tests/validations:
  - Service validation if software is changed/upgraded
  - Specific tests (e.g. throughput) to ensure that no problems have been introduced
  - Tests of functions not yet tested (e.g. Reprocessing/data recall at Tier 1s)
- Details of the test programme will be discussed and agreed in the workshop already planned for November





# Future Infrastructure support

- A second draft of the EGI blueprint has been produced
- There are still some serious shortcomings in the process and in the blueprint:
  - It is not clear exactly what is being proposed in terms of the roles and functions of the National and central organisations;
  - There is no representation of the user communities, and no description of how those communities interact with the infrastructures; (they own many resources)
  - It is not clear how the present operational infrastructure upon which WLCG depends will evolve and appear in the future;
    - Insufficient resources for central operations
    - Risk of discontinuation of ROCs
    - User support is being challenged
  - Very few of the NGIs are as yet established, and so how they can support the WLCG sites is not clear, in particular during a transition period;
  - Given the state of the current blueprint, it seems unlikely that there will be an organisation in place in time to take over the European grid infrastructure from EGEE in early 2010 with a managed transition process during the preceding year.



# WLCG position

- The Tier 1 and Tier 2 sites in Europe will rely on National Grid Infrastructures being in place to provide the support and functions today provided by EGEE
  - Important for operations and middleware support (maintenance and distribution)
  - See position statements provided earlier this year and recent updates
  - Still important that OB/CB members raise this to the NGI and national funding agencies
- The Tier 0 is probably in a good position – the CERN planning does not rely on external funding; but the capability will be strictly limited to core WLCG Tier 0/CAF tasks
- Even if there will be a full EGI/NGI it seems optimistic that this will be in place by the end of EGEE-3, and likely that not all NGIs will be in existence when it starts
- WLCG must have a concrete plan to operate without relying on the European level support, either for an interim period or indefinitely



# Planning for EGI

- Will now document how each of the core functions that we today rely on will be managed in future
  - Start with the position statements sent by Tier 1 sites
  - Consider each of the major functions:
    - GGUS, operational support, monitoring tools, middleware support, certification, deployment support, service management, etc.
    - Work with EGEE to understand the expected status of each of these in mid-2010
    - Negotiate who will manage /contribute to each function if there is no EGI



# LHC Grid Fest

- Very successful event
- Upcoming article in CERN Courier
- Video links emphasised world-wide collaboration
- Excellent statements from openlab partners
- Globe now situated in CC entrance

