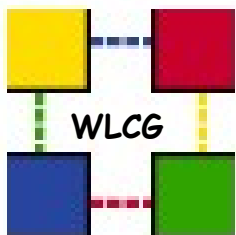




The Worldwide LCG Service Challenges: POW 2006

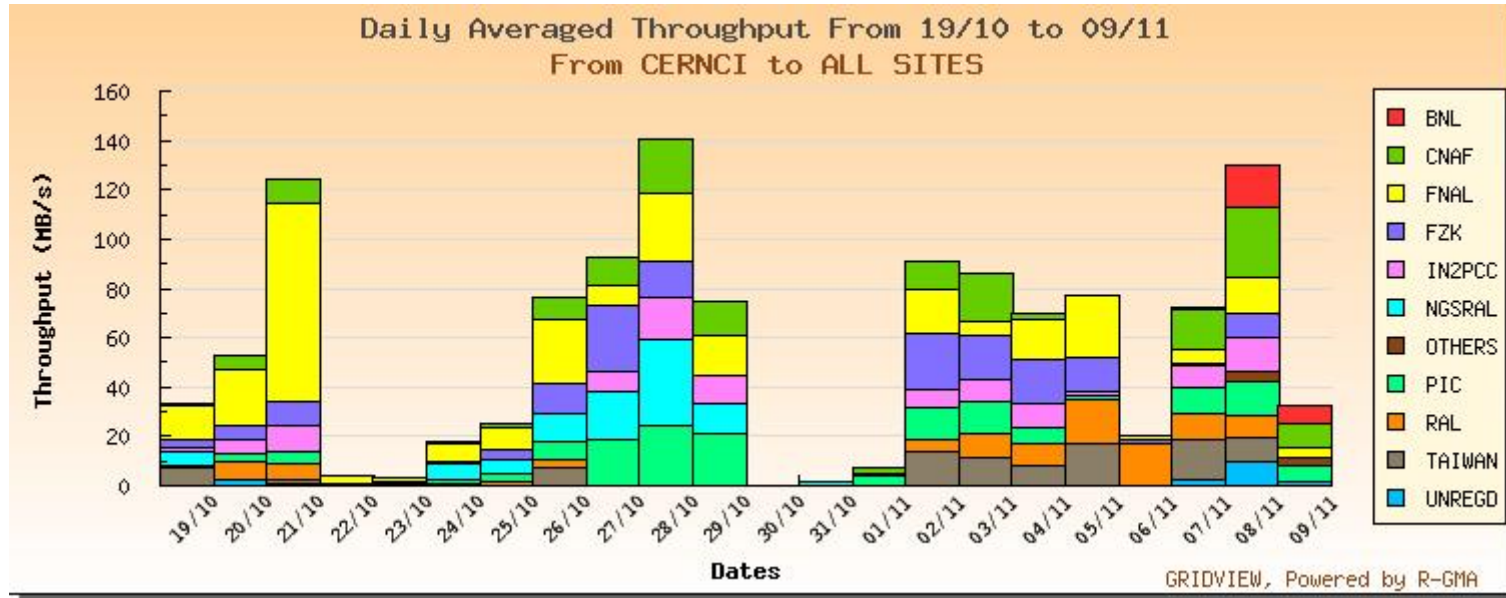
Jamie Shiers, CERN
IT Department POW Retreat
Les Rousses, November 2005
(Updated for November 9th GDB)



Antarctica

Agenda

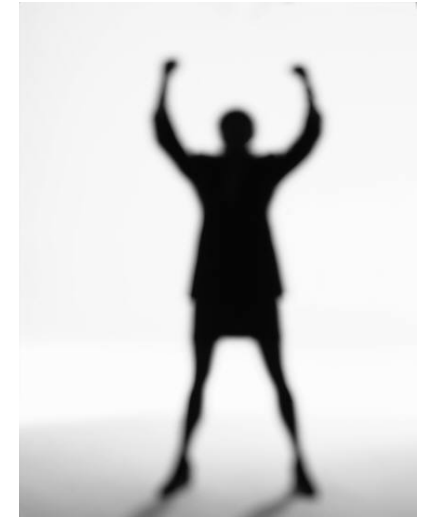
- The main technical problems and how we plan to address them
- [Many slides are hidden - focus on key issues!]



Agenda

- The Service Challenge Programme - a year in Retrospect...
- The Worldwide LCG Collaboration: Sites & Roles
- The "Service Challenge" Programme - Ensuring We Are Ready!
- Measuring our State of Readiness... and Success...

Introduction



- Neither SC1 nor SC2 **fully** met their goals
 - ☺ SC2 **exceeded** its throughput goals
 - ☹ **But not its service goals...**

- Multiple threads started early 2005 to address:
 - Bringing experiments into loop (SC3+)
 - Bringing T2s into loop (ditto)
 - Preparing for full production services
 - Addressing problems beyond 'throughput goals'
 - e.g. site / experiment goals, additional services etc

- ☺ All Tier1s are now involved! Many Tier2s! New s/w successfully deployed!

- ☺ Successful workshops, tutorials (April, May, June, October) and site visits!

- Throughput tests gradually approaching target (more later)

- ☺ Need to understand the problems areas and address them

- ☺ Acknowledge progress / successes / hard-work of many!

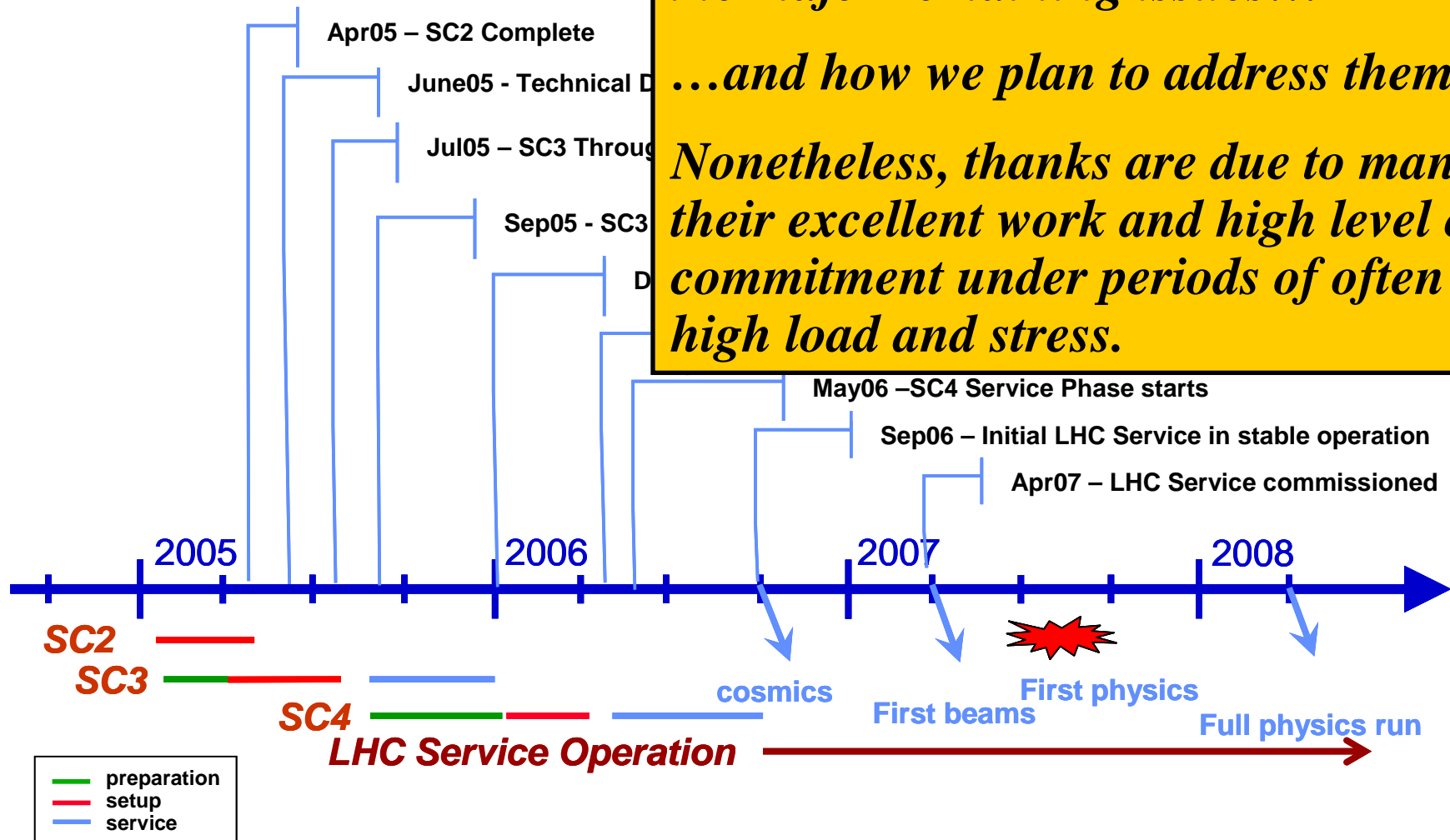
LCG Service De

Unfortunately, I do not have time to report on any of the major successes.

The rest of the presentation will focus on the major remaining issues...

...and how we plan to address them.

Nonetheless, thanks are due to many for their excellent work and high level of commitment under periods of often very high load and stress.



WLCG - Major Challenges Ahead

1. Get data rates at all Tier1s up to MoU Values
 - Stable, reliable, rock-solid services
 - We are currently about 1/2 the target level, without including tape
 2. (Re-)implement Required Services at Sites so that they can meet MoU Targets
 - Measured, delivered Availability, maximum intervention time etc.
 - Ensure that the services delivered match the experiments' requirements
- T0 and T1 services are tightly coupled!
- Particularly during accelerator operation
 - Need to build strong collaborative spirit to be able to deliver required level of services
 - And survive the inevitable 'crises'...

How do we measure success?

- By measuring the service we deliver against the MoU targets
 - Data transfer rates
 - Service availability and time to resolve problems
- By the “challenge” established at CHEP 2004:
 - [The service] *“should not limit ability of physicist to exploit performance of detectors nor LHC’s physics potential”*
 - *“...whilst being stable, reliable and easy to use”*
- Preferably both...
 - Actually I have a 3rd metric but I’m saving that for CHEP

WLCG - Major Challenges Ahead

1. Get data rates at all Tier1s up to MoU Values

- Stable, reliable, rock-solid services
- We are currently about 1/2 the target level, without including tape

2. (Re-)implement Required Services at Sites so that they can meet MoU Targets

- Measured, delivered Availability, maximum intervention time etc.
- Ensure that the services delivered match the experiments' requirements

➤ T0 and T1 services are tightly coupled!

- Particularly during accelerator operation
- Need to build strong collaborative spirit to be able to deliver required level of services
 - And survive the inevitable 'crises'...

Site Components - Updated

- Each T1 to provide 10Gb network link to CERN
- Each site to provide SRM 1.1 interface to managed storage
 - All sites involved in SC3: T0, T1s, T2s.
- T0 to provide File Transfer Service; also at named T1s for T2-T1 transfer tests
 - Named Tier1s: BNL, CNAF, FZK, RAL; Others also setting up FTS
 - CMS T2s being supported by a number of T1s using PhEDEx
- LCG File Catalog - not involved in Throughput but **needed for Service**
 - ALICE / ATLAS: site local catalog
 - LHCb: central catalog with >1 R/O 'copies' (on ~October timescale)
 - IN2P3 to host one copy; CNAF? Taiwan? RAL?
 - CMS: evaluating different catalogs
 - FNAL: Globus RLS, T0+other T1s: LFC; T2s: POOL MySQL, GRLS, ...
- T2s - many more than foreseen
 - Running DPM or dCache, depending on T1 / local preferences / support
 - [Support load at CERN through DPM / LFC / FTS client]
- Work still needed to have these consistently available as services

Services & Service Levels

- List of services that need to be provided by each site is now clear
 - Including any VO-specific variations...
- For SC4 / pilot WLCG none of these services are new
 - Expect to see some analysis-oriented services coming later...
 - Maybe prototyped at some 'volunteer' T2s, e.g. DESY, CALTECH, Padua, .. ?
- The service list at CERN has been classified based on impact of service degradation / unavailability
 - Draft classification for Tier1s and Tier2s also exists & sent to GDB (August)
- A check-list has been produced and the Critical Services are being reimplemented target end-2005
 - Must provide operator procedures, support contacts etc etc
- We will measure service availability at all sites and report regularly
 - Results visible through Web used for daily operations purposes

Service Level Definitions

Class	Description	Downtime	Reduced	Degraded	Availability
C	Critical	1 hour	1 hour	4 hours	99%
H	High	4 hours	6 hours	6 hours	99%
M	Medium	6 hours	6 hours	12 hours	99%
L	Low	12 hours	24 hours	48 hours	98%
U	Unmanaged	None	None	None	None

- Downtime defines the time between the start of the problem and restoration of service at minimal capacity (i.e. basic function but capacity < 50%)
- Reduced defines the time between the start of the problem and the restoration of a reduced capacity service (i.e. >50%)
- Degraded defines the time between the start of the problem and the restoration of a degraded capacity service (i.e. >80%)
- Availability defines the sum of the time that the service is down compared with the total time during the calendar period for the service. Site wide failures are not considered as part of the availability calculations. 99% means a service can be down up to 3.6 days a year in total. 98% means up to a week in total.
- None means the service is running unattended

Tier0 Services

Service	VOs	Class
SRM 2.1	All VOs	C
LFC	LHCb	C
LFC	ALICE, ATLAS	H
FTS	ALICE, ATLAS, LHCb, (CMS)	C
CE	All VOs	C
RB		C
Global BDII		C
Site BDII		H
Myproxy		C
VOMS		H→C
R-GMA		H

Services at CERN

- Building on 'standard service model'
 1. First level support: operations team
 - Box-level monitoring, reboot, alarm
 2. Second level support team: Grid D
 - Alerted by operators and/or alarm
 - Follow 'smoke-tests' for applications
 - Identify appropriate 3rd level support
 - Responsible for maintaining and improving services
 - Two people per week: complementary to Service Manager on Duty
 - Provide daily report to SC meeting (09:00); interact with experiments
 - Members: IT-GD-EIS, IT-GD-SC
 - Phone numbers: 164111; 164222
 3. Third level support teams: by service
 - Notified by 2nd level and / or through operators (by agreement)
 - Should be called (very) rarely... **(Definition of a service?)**

Big on-going effort in this area:

- *Services being reimplemented*
- *Merge of daily OPS meetings*
- *Service Coordination meetings*
- *Con-calls with sites*
- *Workshops*
- *etc.*
- *Goal is all Critical Services ready by Christmas*
- *(This means essentially all...)*

Tier0 Service Dashboard

An evaluation for each product within the four primary task areas:

- Requirements - covers the infrastructure requirements with regard to machines, disks, network;
- Development - covers from software creation and documentation to certification and delivery to the installation teams;
- Hardware - covers the procurement, delivery, burn in, physical installation and base operating systems;
- Operations - covers the administration, monitoring, configuration and backup of the service to the levels requested.

Operations Checklist

- 2nd level support organisation defined (who to call when there is a problem with the application or middleware)
- Mechanism to contact 2nd level organisation
- Response time for 2nd level organisation
- List of machines where service is running defined
- List of configuration parameters and their values for the software components
- List of processes to monitor
- List of file systems and their emergency thresholds for alarms
- Application status check script requirements defined
- Definition of scheduled processes (e.g. cron)
- Test environment defined and available
- Problem determination procedures including how to determine application vs middleware vs database issues
- Procedures for start/stop/drain/check status defined
- Automatic monitoring of the application in place
- Backup procedures defined and tested

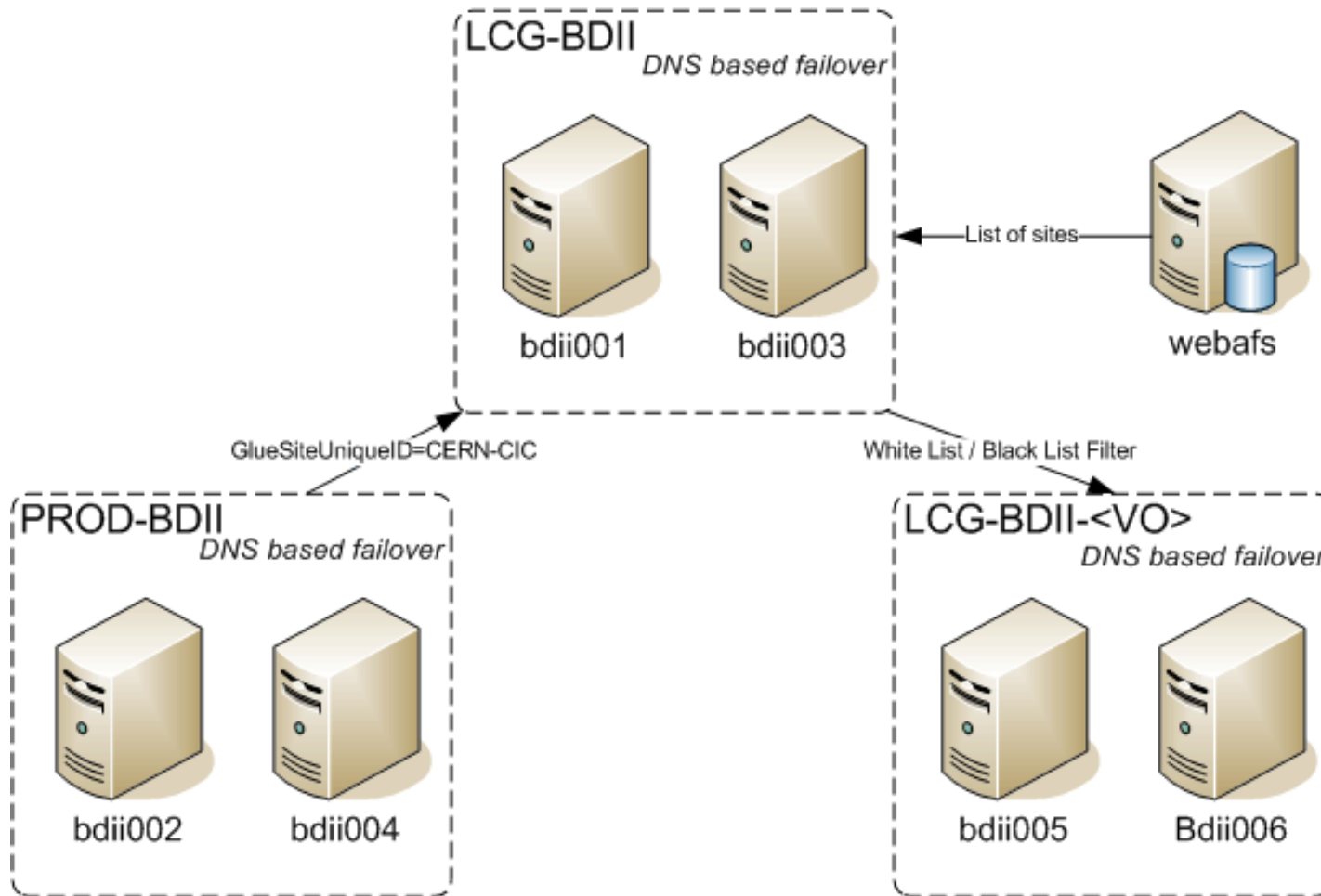
Tier0 Service Coordination

- Progress on re-implementing services monitored at fortnightly LCG Service Coordination Meeting
 - <http://agenda.cern.ch/displayLevel.php?fid=654>
- Area updates provided by area coordinators on Wiki prior to meeting
- Meeting remains crisp, focussed and short
 - Typically less than one hour...
- Target is to get all Critical services re-implemented by year-end

Tier0 Services

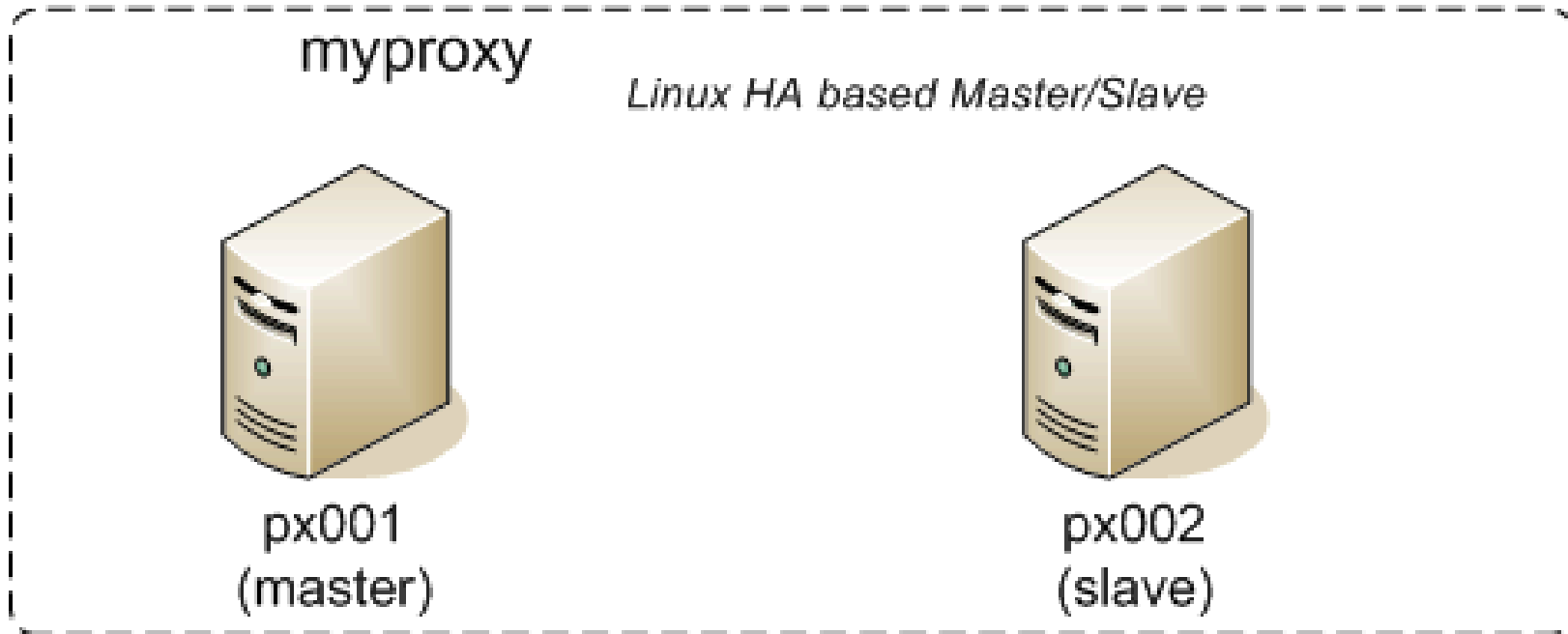
Service	VOs	Class
SRM 2.1	All VOs	C
LFC	LHCb	C
LFC	ALICE, ATLAS	H
FTS	ALICE, ATLAS, LHCb, (CMS)	C
CE	All VOs	C
RB		C
Global BDII		C
Site BDII		H
Myproxy		C
VOMS		H→C
R-GMA		H

CERN BDII Production Deployment Layout



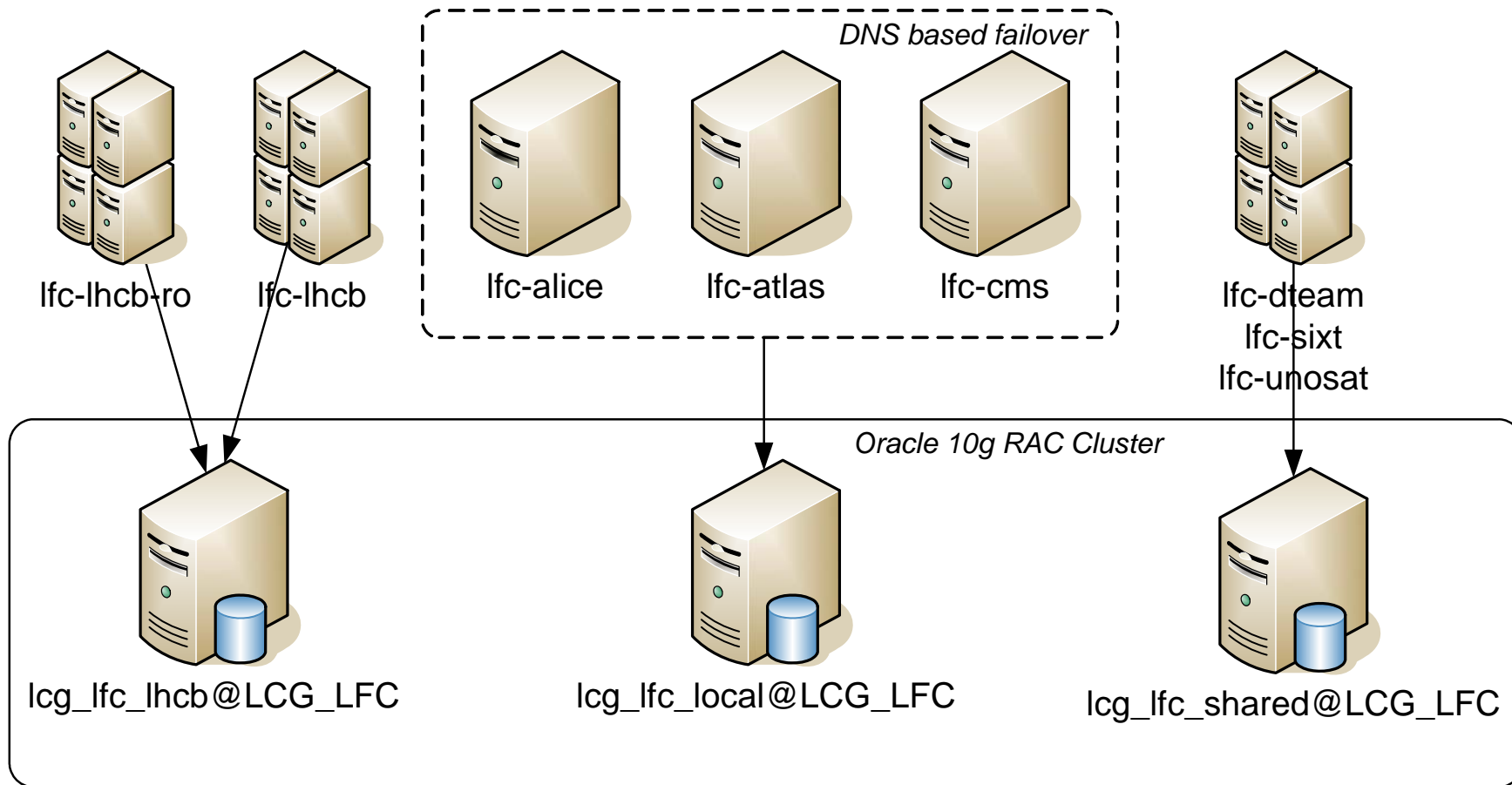
- The Global **BDII** which provides a world wide view of the **BDII** data on the grid
- The site **GIIS** which provides a consolidated view of the various **GRIS** servers on the **CE** and **SE**.
- A **vo-specific BDII** which is a view on the Global **BDII** with the inclusion of the **VO** white and black listing of sites

CERN PX Production Deployment Layout



- Master/Slave set up using Linux-HA and shared IP service address
- Master stores data in /var/proxy and replicates using myproxy_replicate to slave in /var/proxy.slave
- Master rsync's data from /var/proxy to the slave /var/proxy directory
- The slave myproxy server is started in slave mode to read from /var/proxy.slave (i.e. read-only mode)
- In the event of master failure as detected by Linux-HA, the daemon is stopped on the slave and then restarted with the read-write copy from /var/proxy

LFC Production Deployment Layout



27th October 2005

- Failover both at middle and database tiers

WLCG and Database Services

- Many 'middleware' components require a database:
 - dCache - PostgreSQL (CNAF porting to Oracle?)
 - CASTOR / DPM / FTS* / LFC / VOMS - Oracle or MySQL
 - **Some MySQL only: RB, R-GMA#, SFT#**
- Most of these fall into the 'Critical' or 'High' category at Tier0
 - See definitions below; T0 = C/H, T1 = H/M, T2 = M/L
- Implicit requirement for 'high-ish service level'
 - (to avoid using a phrase such as H/A...)
- At this level, no current need beyond site-local+ services
 - Which may include RAC and / or DataGuard
 - [TBD together with service provider]
 - **Expected at AA & VO levels**

*gLite 1.4 end October

#Oracle version foreseen

+R/O copies of LHCb FC?

Required Tier1 Services

Service	VOs	Class
SRM 2.1	All VOs	H/M
LFC	ALICE, ATLAS	H/M
FTS	ALICE, ATLAS, LHCb, (CMS)	H/M
CE		H/M
Site BDII		H/M
R-GMA		H/M

Many also run e.g. an RB etc. Current status for ALICE (hidden)

ALICE RBs in SC3 Production (for ex.)

- **CERN:**
 - gdrb01.cern.ch:7772
 - gdrb02.cern.ch:7772
 - gdrb03.cern.ch:7772
 - gdrb07.cern.ch:7772
 - gdrb08.cern.ch:7772
 - gdrb11.cern.ch:7772
 - lxn1177.cern.ch:7772
 - lxn1186.cern.ch:7772
 - lxn1188.cern.ch:7772
- **SARA:**
 - mu3.matrix.sara.nl:7772
- **NIKHEF:**
 - bosheks.nikhef.nl:7772
- **GridKA:**
 - a01-004-127.gridka.de:7772
- **RAL:**
 - lcgrb01.gridpp.rl.ac.uk:7772
- **CNAF:**
 - egee-rb-01.cnaf.infn.it:7772
 - gridit-rb-01.cnaf.infn.it:7772
- **SINICA:**
 - lcg00124.grid.sinica.edu.tw:7772

Tier1 MoU Availability Targets

Service	Maximum delay in responding to operational problems			Average availability measured on an annual basis	
	Service interruption	Degradation of the capacity of the service by more than 50%	Degradation of the capacity of the service by more than 20%	During accelerator operation	At all other times
Acceptance of data from the Tier-0 Centre during accelerator operation	12 hours	12 hours	24 hours	99%	n/a
Networking service to the Tier-0 Centre during accelerator operation	12 hours	24 hours	48 hours	98%	n/a
Data-intensive analysis services, including networking to Tier-0, Tier-1 Centres outside accelerator operation	24 hours	48 hours	48 hours	n/a	98%
All other services – prime service hours ^[1]	2 hour	2 hour	4 hours	98%	98%
All other services – outside prime service hours	24 hours	48 hours	48 hours	97%	97%

^[1] Prime service hours for Tier1 Centres: 08:00-18:00 in the time zone of the Tier1 Centre, during the working week of the centre, except public holidays and other scheduled centre closures.

Required Tier2 Services

Service	VOs	Class
SRM 2.1	All VOs	M/L
LFC	ATLAS, ALICE	M/L
CE		M/L
Site BDII		M/L
R-GMA		M/L

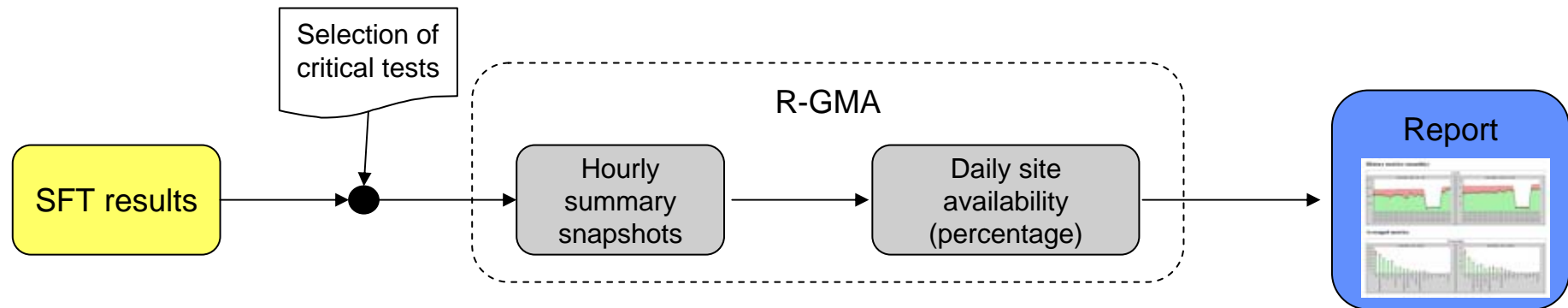
There are also some optional services and some for CIC/ROC and other such sites (this applies also / more to Tier1s...)

Measuring Service Availability

- Will be measured using standard tests run from the Site Functional Test framework
- Will start by regular tests, frequency matched to Service Class
 - i.e. Critical components will be tested every hour
 - High every 4 hours etc.
- This means that interruptions shorter than sampling frequency may be missed
 - But will be supplemented by logs and other information...
- More complex jobs, including VO-specific ones, can / will be added
 - e.g. transfer of data from Tier0 - Tier1 is higher-level function closer to MoU responsibilities

Measuring computing resources availability - status

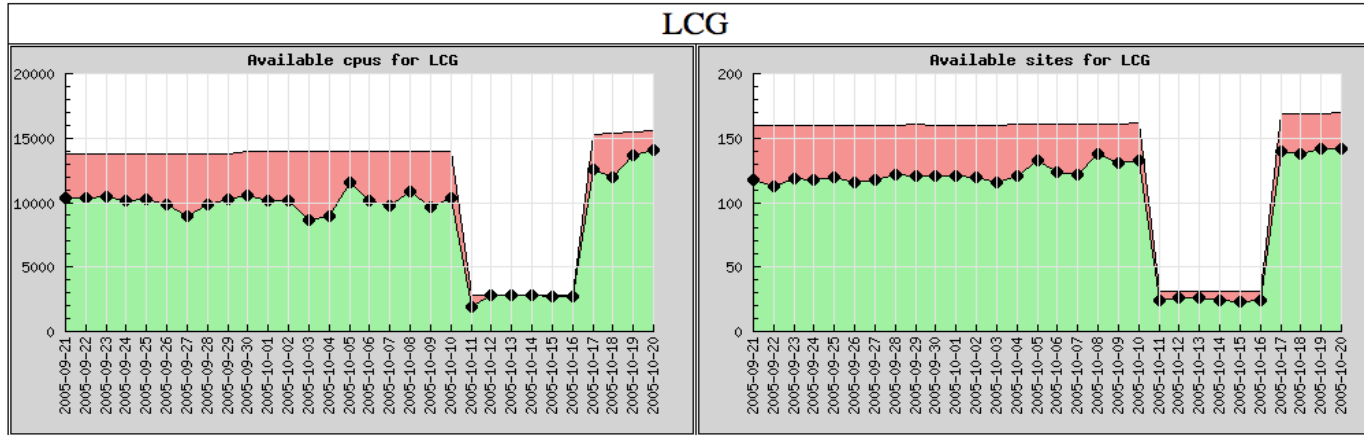
- Based on SFT jobs sent to all sites at least once per 3 hours
 - More frequent submissions if needed



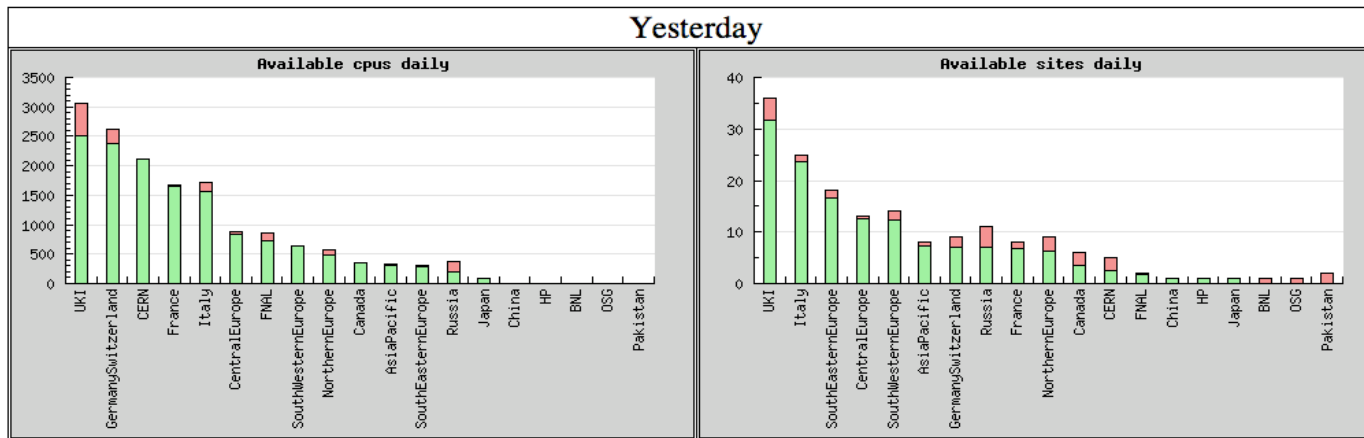
- Measurements stored and archived in R-GMA
 - Currently MySQL but Oracle foreseen
- Aggregated by region (ROC) and for the whole grid
- Current report shows only regional aggregation but "per site" view will be available soon
 - Data is already there
- Additional metric: availability multiplied by published amount of CPUs
 - "Good" resources vs. potential resources
- No direct testing of storage resources
 - Indirect testing - replica management tests

Measuring computing resources availability - graphs

History metrics (monthly)



Averaged metrics



Tier0 Services - Status of Monitoring

Service	Responsible	Class
SRM 2.1	Dave Kant	C
LFC	LFC support	C
LFC	LFC support	H
FTS	FTS support	C
CE	Monitored by SFT today	C
RB	Dave Kant (partially done)	C
Global BDII	Tbd (Gstat) Min Tsai	C
Site BDII	Done (Gstat) Min Tsai	H
Myproxy	Maarten Litmaath	C
VOMS	Valerio Venturi	H→C
R-GMA	Lawrence Field	H

WLCG - Major Challenges Ahead

1. Get data rates at all Tier1s up to MoU Values

- Stable, reliable, rock-solid services
- We are currently about 1/2 the target level, without including tape

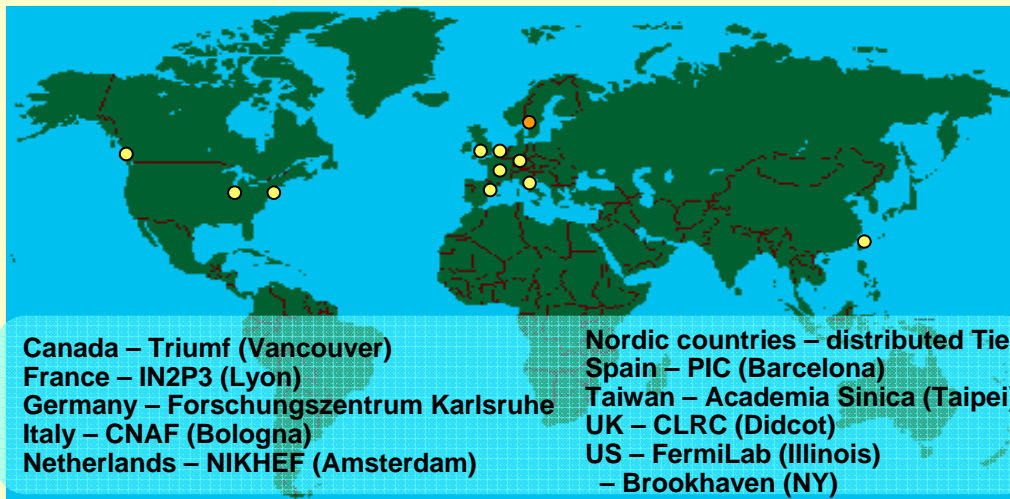
2. (Re-)implement Required Services at Sites so that they can meet MoU Targets

- Measured, delivered Availability, maximum intervention time etc.
 - Ensure that the services delivered match the experiments' requirements
- T0 and T1 services are tightly coupled!
- Particularly during accelerator operation
- Need to build strong collaborative spirit to be able to deliver required level of services
 - And survive the inevitable 'crises'...

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 intensive analysis
- National, regional support
- Continual reprocessing activity

Tier-2 - ~100 centres in ~40 countries

- Simulation
- End-user analysis – batch and interactive

Overview of pp running

Experiment	SIM	SIMESD	RAW	Trigger	RECO	AOD	TAG
ALICE	400KB	40KB	1MB	100Hz	200KB	50KB	10KB
ATLAS	2MB	500KB	1.6MB	200Hz	500KB	100KB	1KB
CMS	2MB	400KB	1.5MB	150Hz	250KB	50KB	10KB
LHCb		400KB	25KB	2KHz	75KB	25KB	1KB

Nominal pp data rates - MoU

Centre	ALICE	ATLAS	CMS	LHCb	Rate into T1 (pp) MB/s
ASGC, Taipei	-	8%	10%	-	100
CNAF, Italy	7%	7%	13%	11%	200
PIC, Spain	-	5%	5%	6.5%	100
IN2P3, Lyon	9%	13%	10%	27%	200
GridKA, Germany	20%	10%	8%	10%	200
RAL, UK	-	7%	3%	15%	150
BNL, USA	-	22%	-	-	200
FNAL, USA	-	-	28%	-	200
TRIUMF, Canada	-	4%	-	-	50
NIKHEF/SARA, NL	3%	13%	-	23%	150
Nordic Data Grid Facility	6%	6%	-	-	50
Totals	-	-	-	-	1,600

These rates must be sustained to tape 24 hours a day, 100 days a year.

Extra capacity is required to cater for backlogs / peaks.

This is currently our biggest data management challenge.

Tier1 Responsibilities - Rates to Tape

- i. acceptance of an agreed share of raw data from the Tier0 Centre, keeping up with data acquisition;
- ii. acceptance of an agreed share of first-pass reconstructed data from the Tier0 Centre;

<i>Centre</i>	<i>ALICE</i>	<i>ATLAS</i>	<i>CMS</i>	<i>LHCb</i>	<i>Rate into T1 (pp)</i>
ASGC, Taipei	-	8%	10%	-	100
CNAF, Italy	7%	7%	13%	11%	200
PIC, Spain	-	5%	5%	6.5%	100
IN2P3, Lyon	9%	13%	10%	27%	200
GridKA, Germany	20%	10%	8%	10%	200
RAL, UK	-	7%	3%	15%	150
BNL, USA	-	22%	-	-	200
FNAL, USA	-	-	28%	-	200
TRIUMF, Canada	-	4%	-	-	50
NIKHEF/SARA, NL	3%	13%	-	23%	150
Nordic Data Grid Facility	6%	6%	-	-	50
Totals	-	-	-	-	1,600

Tier1 Responsibilities - cont.

- iii. acceptance of processed and simulated data from other centres of the WLCG;
- iv. recording and archival storage of the accepted share of raw data (distributed back-up);
- v. recording and maintenance of processed and simulated data on permanent mass storage;
- vi. provision of managed disk storage providing permanent and temporary data storage for files and databases;
- vii. provision of access to the stored data by other centres of the WLCG and by named AF's;
- viii. operation of a data-intensive analysis facility;
- ix. provision of other services according to agreed Experiment requirements;
- x. ensure high-capacity network bandwidth and services for data exchange with the Tier0 Centre, as part of an overall plan agreed amongst the Experiments, Tier1 and Tier0 Centres;
- xi. ensure network bandwidth and services for data exchange with Tier1 and Tier2 Centres, as part of an overall plan agreed amongst the Experiments, Tier1 and Tier2 Centres;
- xii. administration of databases required by Experiments at Tier1 Centres.

Results of SC3 in terms of Transfers

- Target data rates 50% higher than during SC2
- All T1s (most supporting T2s) participated in this challenge
- Transfers between SRMs (not the case in SC1/2)
- Important step to gain experience with the services before SC4

Site	MoU Target (Tape)	Daily average MB/s (Disk)
ASGC	100	10
BNL	200	107
FNAL	200	185
GridKa	200	42
CC-IN2P3	200	40
CNAF	200	50
NDGF	50	129
PIC	100	54
RAL	150	52
SARA/NIKHEF	150	111
TRIUMF	50	34

Rates during July throughput tests. Better single-site rates since, but need to rerun tests...

For this we need dCache 1.6.6(+) to be released/deployed, latest FTS (now), network upgrades etc.

January?? (<CHEP)

Nominal	These are the raw figures produced by multiplying e.g. event size x trigger rate.
Headroom	A factor of 1.5 that is applied to cater for peak rates.
Efficiency	A factor of 2 to ensure networks run at less than 50% load.
Recovery	A factor of 2 to ensure that backlogs can be cleared within 24 - 48 hours and to allow the load from a failed Tier1 to be switched over to others.
Total Requirement	Dedicated 10Gbit/second network links are being provisioned to all Tier1 centres. These will be used to provide a Reliable File Transfer Service for bulk data exchange. (Tier0 -> Tier1s for RAW and 1st pass reconstructed data, Tier1 -> Tier0 & other Tier1s for reprocessed data.)

Service Challenges

- Purpose
 - Understand what it takes to operate a real grid service - run for days/weeks at a time (outside of experiment Data Challenges)
 - Trigger/encourage the Tier1 & large Tier-2 planning - move towards real resource planning - based on realistic usage patterns
 - Get the essential grid services ramped up to target levels of reliability, availability, scalability, end-to-end performance
 - Set out milestones needed to achieve goals during the service challenges
- NB: This is focussed on Tier 0 - Tier 1/large Tier 2
 - Data management, batch production and analysis
- Short term goal - **by end 2004** - have in place a robust and reliable data management service and support infrastructure and robust batch job submission

From early proposal, May 2004

Where do we stand today?

- Main focus of first two Service Challenges was building up service infrastructure to handle production data flows
 - Distribution of RAW + reconstructed data during machine run
 - 👉 No experiment s/w involved, just basic infrastructure
- Current challenge (3/4) involves all Tier1 sites, several Tier2s and all Offline Use Cases except (officially) Analysis
- Roles of each site (tier) and services offered / required, including minor VO-specific variations, now well understood
- 💣 **Building up Production Services requires significant effort - and time**
 - Neither of which are in abundance

Data Transfer Rates

- 2 years before data taking *can* transfer from SRM at CERN to DPM SRM at T1 at ~target data rate
- Stably, reliably, days on end
 - ☺ Great, so we got a fallback(?)
- Need to do this to all T1s at target data rates to tape to all supported SRM implementations (dCache, CASTOR + b/e MSS)
- Plus factor 2 for backlogs / peaks
- Need to have fully debugged recovery procedures
- Data flows from re-processing need to be discussed
 - New ESD copied back to CERN (and to another T1 for ATLAS)
 - AOD and TAG copied to other T1s, T0, T2s (subset for AOD?)

Data Transfer Plans



- A lot of debugging work was done over the summer
- Many improvements have been implemented .. and now released...
- **The failure to meet (without effort) the target data rates is probably good in the long run**
 - Reality check - this stuff ain't easy
- The Services have improved by leaps and bounds
- At least we can now see the Wood - last year it was just Trees
- But the existing plan (from 18 months ago) needs to be revised...

Transfer Plans #0

- Transfer plans = Tier1 plans
 - For the time being, no collaboration-accepted Use Case for high T2 related traffic
 - But Harvey's got plenty of plausible arguments...
 - T1-T1 and Tx-Ty traffic to be discussed prior to CHEP W/S
 - Continual reprocessing activity (not continuous...)
- All Tier1s are (very) different
 - And we're now at the level that we have to address this..
 - Some have simple / complex Tier2 situation
 - Big variation in VOs supported etc
 - Significant differences in implementation
- Customised plans, converging on a common goal?

SRM / MSS by Tier1

<i>Centre</i>	<i>SRM</i>	<i>MSS</i>	<i>Tape H/W</i>
<i>Canada, TRIUMF</i>	dCache	TSM	
<i>France, CC-IN2P3</i>	dCache	HPSS	STK
<i>Germany, GridKA</i>	dCache	TSM	LTO3
<i>Italy, CNAF</i>	CASTOR	CASTOR	STK 9940B
<i>Netherlands, NIKHEF/SARA</i>	dCache	DMF	STK
<i>Nordic Data Grid Facility</i>	DPM	N/A	N/A
<i>Spain, PIC Barcelona</i>	CASTOR	CASTOR	STK
<i>Taipei, ASGC</i>	CASTOR	CASTOR	STK
<i>UK, RAL</i>	dCache	ADS→ CASTOR(?)	STK
<i>USA, BNL</i>	dCache	HPSS	STK
<i>USA, FNAL</i>	dCache	ENSTOR	STK

Transfer Plans: #1

- Identify the key sites
 - In Europe: CNAF, FZK, IN2P3, (nikhef/sara, ral)
 - In the US: BNL, FNAL
- Find a 'flag-ship' site
 - ?
- Get them up to speed
 - Others in the wake
- Move on to next level
- So far we didn't find a flag-ship site...

<i>Centre</i>	<i>Rate into T1</i>
ASGC, Taipei	100
CNAF, IT	200
PIC, ES	100
IN2P3, FR	200
GridKA, DE	200
RAL, UK	150
BNL, USA	200
FNAL, USA	200
TRIUMF, CA	50
N/S, NL	150
NDGF	50
Totals	1,600

Transfer plans #2

- Split out export of data out of CERN from Tier1 writing to tape
 - NDGF have proved its possible to support ~target data rate today
 - Tape layer involves 'untimely' purchases
 - Its mandatory to test this layer, but:
 - Scale targets to current h/w
 - Clearly emphasise need for full planning for data taking
- **Demonstration of 'nominal data rates' out of CERN to SRMs at all Tier1s would already be a major milestone!**
- Plan January 2006 with dCache 1.6.6 etc at relevant sites

Pre-Requisites for Re-Run of Throughput Tests

- Deployment of gLite FTS 1.4 (srmcp support)
 - ✓ Done at CERN in recent intervention
- dCache 1.6.6 (or later) release and deployed at all dCache sites.
 - ✓ Pre-release in test at a few key sites
- CASTOR2 clients and CASTORSRM version 2.2.8 (or later) at all CASTOR sites (ASGC, CNAF, PIC).
- Upgrade to CERN internal network infrastructure.
 - Partly done - remainder at Christmas shutdown?
- 10Gbit/s network connections at operational at the following sites:
 - IN2P3, GridKA, CNAF, NIKHEF/SARA, BNL, FNAL

dCache - the Upgrade (CHEP 2006)

- *For the last two years, the dCache/SRM Storage Element has been successfully integrated into the LCG framework and is in heavy production at several dozens of sites, spanning a range from single host installations up to those with some hundreds of TB of disk space, delivering more than 50 TB per day to clients. Based on the permanent feedback from our users and the detailed reports given by representatives of large dCache sites during our workshop at DESY end of August 2005, the dCache team has been identified important areas of improvement.*
- *This includes a more sophisticated handling of the various supported tape back-ends, the introduction of multiple I/O queues per pool with different properties to account for the diverse behaviours of the different I/O protocols and the possibility to have one dCache instance spread over more than one physical site.*
- *... changes in the name-space management as short and long term perspective to keep up with future requirements.*
- *... initiative to make dCache a widely scalable storage element by introducing dCache, the Book, plans for improved packaging and more convenient source code license terms.*
- *Finally I would like to cover the dCache part of the German e-science project, d-Grid, which will allow for improved scheduling of tape to disk restore operations as well as advanced job scheduling by providing extended information exchange between storage elements and Job Scheduler.*

Disk - Disk Rates (SC3 Repeat)

<i>Centre</i>	<i>ALICE</i>	<i>ATLAS</i>	<i>CMS</i>	<i>LHCb</i>	<i>Target Data Rate MBytes/sec</i>
<i>Canada, TRIUMF</i>		X			50
<i>France, CC-IN2P3</i>	X	X	X	X	150
<i>Germany, GridKA</i>	X	X	X	X	150
<i>Italy, CNAF</i>	X	X	X	X	150
<i>Netherlands, NIKHEF/SARA</i>	X	X		X	150
<i>Nordic Data Grid Facility</i>	X	X	X		50
<i>Spain, PIC Barcelona</i>		X	X	X	100
<i>Taipei, ASGC</i>		X	X		100
<i>UK, RAL</i>	X	X	X	X	150
<i>USA, BNL</i>		X			150
<i>USA, FNAL</i>			X		150
<i>Target data rate at CERN</i>					1,000

January 2006

Disk - Disk Rates (SC4 part 1)

<i>Centre</i>	<i>ALICE</i>	<i>ATLAS</i>	<i>CMS</i>	<i>LHCb</i>	<i>Rate into T1 (pp) MB/s</i>
ASGC, Taipei	-	8%	10%	-	100
CNAF, Italy	7%	7%	13%	11%	200
PIC, Spain	-	5%	5%	6.5%	100
IN2P3, Lyon	9%	13%	10%	27%	200
GridKA, Germany	20%	10%	8%	10%	200
RAL, UK	-	7%	3%	15%	150
BNL, USA	-	22%	-	-	200
FNAL, USA	-	-	28%	-	200
TRIUMF, Canada	-	4%	-	-	50
NIKHEF/SARA, NL	3%	13%	-	23%	150
Nordic Data Grid Facility	6%	6%	-	-	50
Totals	-	-	-	-	1,600

Disk - Tape Rates (SC4 part 1)

<i>Centre</i>	<i>ALICE</i>	<i>ATLAS</i>	<i>CMS</i>	<i>LHCb</i>	<i>Target Data Rate MB/s</i>
<i>Canada, TRIUMF</i>		X			50
<i>France, CC-IN2P3</i>	X	X	X	X	75
<i>Germany, GridKA</i>	X	X	X	X	75
<i>Italy, CNAF</i>	X	X	X	X	75
<i>Netherlands, NIKHEF/SARA</i>	X	X		X	75
<i>Nordic Data Grid Facility</i>	X	X	X		50
<i>Spain, PIC Barcelona</i>		X	X	X	75
<i>Taipei, ASGC</i>		X	X		75
<i>UK, RAL</i>	X	X	X	X	75
<i>USA, BNL</i>		X			75
<i>USA, FNAL</i>			X		75

- Still using SRM 1.1 & Current Tape Technology?

April 2006

Disk - Tape Rates (SC4 part 2)

<i>Centre</i>	<i>ALICE</i>	<i>ATLAS</i>	<i>CMS</i>	<i>LHCb</i>	<i>Rate into T1 (pp) MB/s</i>
ASGC, Taipei	-	8%	10%	-	100
CNAF, Italy	7%	7%	13%	11%	200
PIC, Spain	-	5%	5%	6.5%	100
IN2P3, Lyon	9%	13%	10%	27%	200
GridKA, Germany	20%	10%	8%	10%	200
RAL, UK	-	7%	3%	15%	150
BNL, USA	-	22%	-	-	200
FNAL, USA	-	-	28%	-	200
TRIUMF, Canada	-	4%	-	-	50
NIKHEF/SARA, NL	3%	13%	-	23%	150
Nordic Data Grid Facility	6%	6%	-	-	50
Totals	-	-	-	-	1,600

Tier2 Sites - Target is 20 (April) / 40 (July)

Site	ALICE	ATLAS	CMS	LHCb
Bari	X		X	
Catania	X			
Bologna			x	
Legnaro			x	
Pisa			X	
Rome			X	
Catania	X			
GSI	X			
Torino	X			
DESY			X	
CIEMAT+IFCA			X	
jlnr	x			
itep	x		x	x
sinp			x	
mano			x	x
TAIWAN NCU			X	
IC			X	
Caltech			x	
Florida			x	
Nebraska			X	
Purdue			x	
UCSD			X	
Wisconsin			X	

This is not an official list!

We should easily(?) meeting April target! But need to measure service delivered!

Transfer plans #3

- Using new tap test
- Target is TV
- Setup worksh
 - Not just the infrastruc
- Work with si schedule 'FUL
 - i.e. July 20
- DECLARE THE OFFICIALLY



T0-T1 transfer

T1 (i.e. 400MB/s)

results

/ network

production plans to end of SC4

SERVICE

Major Challenges (Reminder)

- Get data rates at all Tier1s up to MoU Values
 - Stable, reliable, rock-solid services
- (Re-)implement Required Services at Sites so that they can meet MoU Targets
 - Measured, delivered Availability, maximum intervention time etc.
- **T0 and T1 services are tightly coupled!**
 - Particularly during accelerator operation
- Need to build strong collaborative spirit to be able to deliver required level of services
 - And survive the inevitable 'crises'...

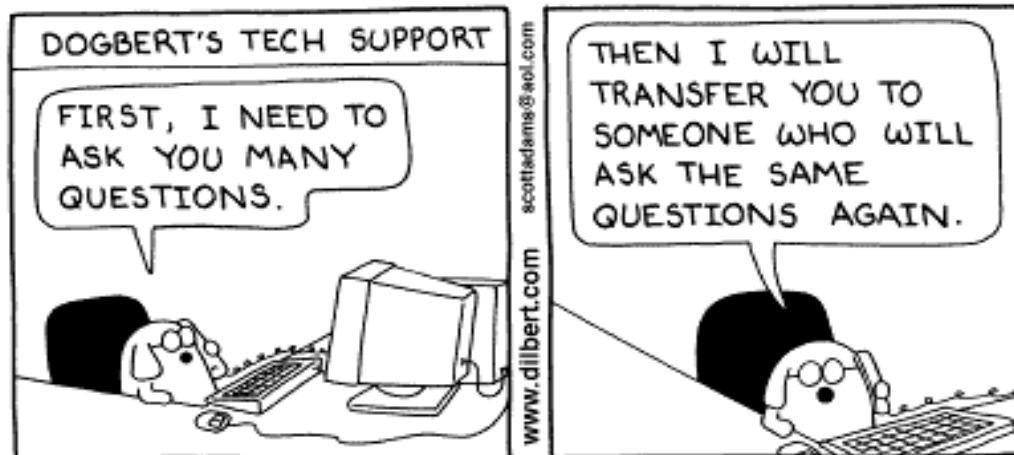


Operations Goals

- Already understand what core services need to run at which site (and VO variations...)
- Goal: MoU targets automatically monitored using Site Functional Tests prior to end-2005
- Tier0 services being re-architected / implemented to meet MoU targets
- Will share techniques / procedures etc with other sites
- This will provide required basis on which to build Grid User Support

User Support Goals

- As services become well understood and debugged, progressively hand-over first Operations, then User Support, to agreed Grid bodies
- Target: all core services will prior to end-September 2006 milestone for the Production WLCG Service
- This will require a significant amount of effort in parallel to goals regarding Reliable Transfer Rates etc.



WLCG Service Coordination

- Fortnightly Service Coordination meetings held at CERN
 - Almost all work prepared beforehand
- Weekly con-calls will possibly be split into two (but seem to work well):
 1. Focus on experiment usage of WLCG Services
 2. Focus on setting up and running WLCG Services
- Quarterly WLCG Service Coordination Meetings
 - All Tier1s, main Tier2s, ... minutes, agenda etc, material circulated in advance...
- Bi-annual Service workshops
 - One at CERN (April / May?), one outside (September - October?)
 - Easter 2006 is April 14 - 17
- Thematic workshops, site visits as required
 - Each Tier1 visited once per quarter(?)
 - (Combined with other events where appropriate)
 - Regular 1-1 Video Meetings

The WLCG Team

- A small "A-Team" to:
 - a. Parachute in and fix problems;
 - b. Identify problem areas but act as "catalysts" ;
 - c. Present a high-level focused overview - evangelise;
 - d. All of the above and more?
 - e. ... What ever it takes ...

- Snow Bandit
- Blond Bond
- Monster Head
- Zed Phatal
- Seth Vicious

WLCG Services (- DM)

- Tier0
 - Strongly motivated people driving the effort
- Tier1 / Tier2
 - Now well understood what has to be provided
 - Experiments are providing a big push
 - Agreement on SFT monitoring
 - This will go just fine

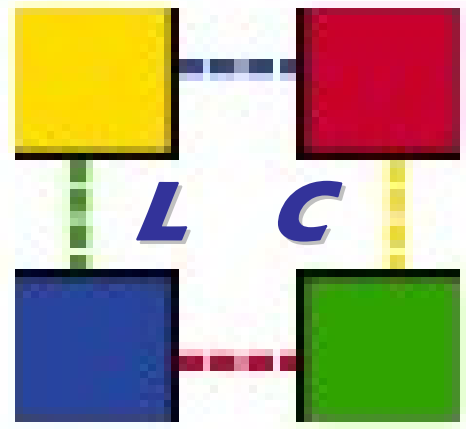
How do we measure success?

- By measuring the service we deliver against the MoU targets
 - Data transfer rates
 - Service availability and time to resolve problems
- By the “challenge” established at CHEP 2004:
 - [The service] *“should not limit ability of physicist to exploit performance of detectors nor LHC’s physics potential”*
 - *“...whilst being stable, reliable and easy to use”*
- Preferably both...

WLCG - Major Challenges Ahead

1. **Get data rates at all Tier1s up to MoU Values**
 - This is currently our biggest challenge - by far
 - Plan is to work with a few key sites and gradually expand
 - (Focus on highest-data rate sites initially...)
2. **(Re-)implement Required Services at Sites so that they can meet MoU Targets**
 - Tier0 will have all services re-implemented prior to SC4 Service Phase (WLCG Pilot)
 - Plans are being shared with Tier1s and Tier2s, as will be experience
 - LCG Service Coordination team will be proactive in driving this forward
 - A lot of work, but no major show-stopper foreseen





**The Worldwide
LHC Computing Grid**

