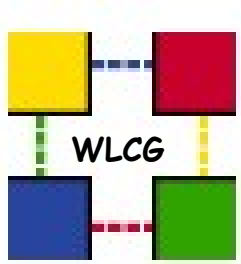
The LHC Computing Environment



Challenges in Building up the Full Production Environment

[Formerly known as the LCG Service Challenges]

Jamie.Shiers@cern.ch

Deploying the Worldwide LCG Computing Service

WLCG Pilot: May 2006 on

WLCG Production: October 2006 on

Ready for data-taking: April 2nd 2007

Introduction

- The (W)LCG Service Challenges are about **preparing**, **hardening** and **delivering** the production Worldwide LHC Computing Environment (WLCG)
- The date for delivery of the production LHC Computing Environment is **30 September 2006**
- Production Services are required as from **1 September 2005** and
  - service phase of Service Challenge 3

  1 May 2006
  - service phase of Service Challenge 4
- This is not a drill.

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 involves **all Tier1** sites, several Tier2s and all Offline Use Cases except (officially) Analysis
- Building up Production Services Requires significant effort – and time
  - Neither of which are in abundance
- Urgent to understand analysis-oriented Use Cases – and the corresponding Services / VO / Site

➢ **Component Services for SC4 (the pilot Worldwide LCG Service) need to be in place end January 2006**

Status of Services

- All services required for SC4 (and hence WLCG pilot) now deployed at all Tier1s and participating Tier2s
- There are no new services required for SC4
- Expect to prototype (end-user) analysis services in parallel with a few key Analysis Facilities
- Bringing a new service to full production level takes about 1 year (or more)
- Much needs to be done in terms of Monitoring, Reporting, Problem Tracking & Logging and User Support

Major Challenges Ahead

1. Get data rates at all Tier1s up to MoU Values
   - Stable, reliable, rock-solid services

2. (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'…

pp / AA data rates (equal split) - TDR

Centre	ALICE	ATLAS	CMS	LHCb	Rate into T1 (pp)	Rate into T1 (AA)
ASGC, Taipei	0	1	1	0	**118.7**	**28.2**
CNAF, Italy	1	1	1	1	**205.0**	**97.2**
PIC, Spain	0	1	1	1	**179.0**	**28.2**
IN2P3, Lyon	1	1	1	1	**205.0**	**97.2**
GridKA, Germany	1	1	1	1	**205.0**	**97.2**
RAL, UK	1	1	1	1	**205.0**	**97.2**
BNL, USA	0	1	0	0	**152.2 (all ESD)**	**11.3**
FNAL, USA	0	0	1	0	**46.5 (expect more)**	**16.9**
TRIUMF, Canada	0	1	0	0	**72.2**	**11.3**
NIKHEF/SARA, NL	1	1	0	1	**158.5**	**80.3**
Nordic Data Grid Facility	1	1	0	0	**98.2**	**80.3**
Totals	6	10	7	6		

N.B. these calculations assume equal split as in Computing Model documents. It is clear that this is not the 'final' answer…

pp data rates – 'weighted' – MoU

Centre	ALICE	ATLAS	CMS	LHCb	Rate into T1 (pp)
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**

Full AOD & TAG to all T1s (probably not in early days)

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 rates since, but need to rerun tests…

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
- ➢ **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?)

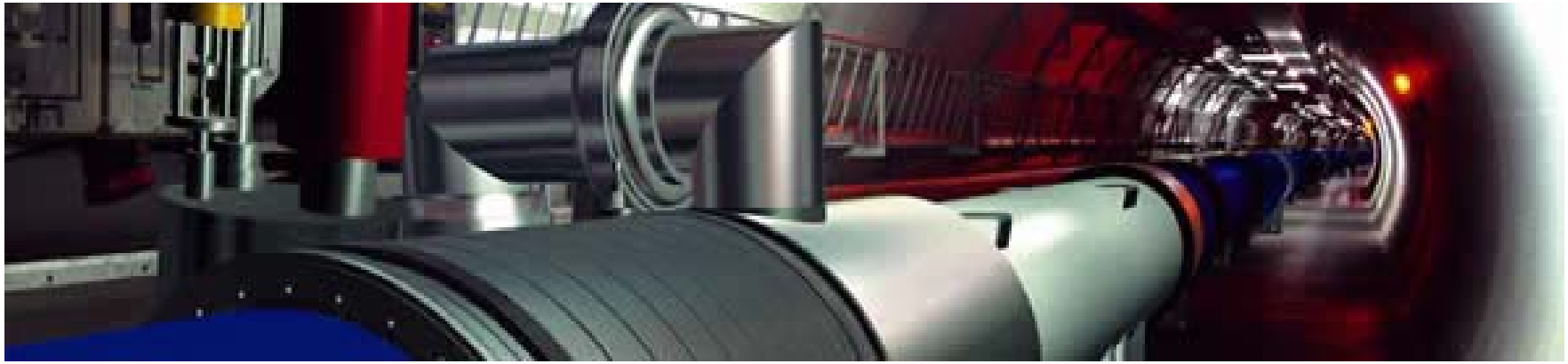
LHC Running Parameters

- Standard assumption:
  - pp running = 10^7 seconds (real time 1.8 x 10^7)
  - Heavy ion = 10^6 seconds (real time 2.6 x 10^6)
  - Data rate independent of luminosity
  - 50 day run in 2007
  - Data volumes per Tier1: 200MB/s x 10^7 = 2PB
  - (Plus results of reprocessing, MC from T2s etc.)

LHC Operation Schedule

- During a normal year…
  - 1 day machine setup with beam
  - 20 days physics
  - 4 days machine development
  - 3 days technical stop

Repeated 7 times



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?**

Services at CERN

- Building on **'standard service model'**

1. First level support: operations team
   - Box-level monitoring, reboot, alarms, procedures etc

2. Second level support team: Grid Deployment group
   - Alerted by operators and/or alarms (and/or production managers…)
   - Follow 'smoke-tests' for applications
   - Identify appropriate 3rd level support team to call
   - Responsible for maintaining and improving procedures
   - 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?)**

Service Challenge 4 – SC4

- SC4 starts April 2006
- SC4 ends with the deployment of the **FULL PRODUCTION SERVICE**

➢ **Deadline for component (production) delivery: end January 2006**

- Adds further complexity over SC3 – 'extra dimensions'
  - Additional components and services, e.g. COOL and other DB-related applications
  - Analysis Use Cases
  - SRM 2.1 features required by LHC experiments **← have to monitor progress!**
  - MostTier2s, all Tier1s at full service level
  - Anything that dropped off list for SC3…
  - ➢ **Services oriented at analysis & end-user**
  - What implications for the sites?

- Analysis farms:
  - Batch-like analysis at some sites (no major impact on sites)
  - Large-scale parallel interactive analysis farms and major sites
  - (100 PCs + 10TB storage) x N

- User community:
  - No longer small (<5) team of production users
  - 20-30 work groups of 15-25 people
  - Large (100s – 1000s) numbers of users worldwide

Analysis Use Cases (HEPCAL II)

- Production Analysis (PA)
  - **Goals in Context** *Create AOD/TAG data from input for physics analysis groups*
  - **Actors** *Experiment production manager*
  - **Triggers** *Need input for "individual" analysis*
- (Sub-)Group Level Analysis (GLA)
  - **Goals in Context** *Refine AOD/TAG data from a previous analysis step*
  - **Actors** *Analysis-group production manager*
  - **Triggers** *Need input for refined "individual" analysis*
- End User Analysis (EA)
  - **Goals in Context** *Find "the" physics signal*
  - **Actors** *End User*
  - **Triggers** *Publish data and get the Nobel Prize :-)*

SC4 Use Cases (?)

Not covered so far in Service Challenges:

- T0 recording to tape (and then out)
- Reprocessing at T1s
- Calibrations & distribution of calibration data
- HEPCAL II Use Cases
- Individual (mini-) productions (if / as allowed)

Additional services to be included:

- Full VOMS integration
- COOL, other AA services, experiment-specific services (e.g. ATLAS HVS)
- **PROOF? xrootd? (analysis services in general…)**
- Testing of next generation IBM and STK tape drives

SC4 Timeline

- Now: clarification of SC4 Use Cases, components, requirements, services etc.
- October 2005: SRM 2.1 testing starts; FTS/MySQL; target for post-SC3 services
- January 31st 2006: basic components delivered and in place
  - This is not the date the s/w is released – it is the date production services are ready
- February / March: integration testing
- February: SC4 planning workshop at CHEP (w/e before)
- March 31st 2006: integration testing successfully completed
- April 2006: throughput tests
- May 1st 2006: Service Phase **(WLCG Pilot)** starts (note compressed schedule!)
- September 30th 2006: Initial LHC Service **(WLCG Production)** in stable operation
- April 2007: LHC Computing Service Commissioned
- Summer 2007: first LHC event data

Remaining Challenges

- Bring core services up to robust 24 x 7 standard required
- Bring remaining Tier2 centres into the process
- Identify the additional Use Cases and functionality for SC4
- Build a cohesive service out of distributed community
- Clarity; simplicity; ease-of-use; functionality

➢ **Getting the (stable) data rates up to the target**

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'…

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;

Centre	*ALICE*	*ATLAS*	*CMS*	*LHCb*	*Rate into T1 (pp)*
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.**

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 outwith 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.

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

Example Services & Service Levels

Service	Service Level	Runs Where
Resource Broker	Critical	Main sites
Compute Element	High	All
MyProxy	Critical	
BDII		
R-GMA		
LFC	High	All sites (ATLAS, ALICE) CERN (LHCb)
FTS	High	T0, T1s (except FNAL)
SRM	Critical	All sites

- This list needs to be completed and verified
- Then plans / timescales for achieving the necessary service levels need to be agreed
  - sharing solutions where-ever possible / appropriate

Tier0 Services

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

Tier1 Services

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

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

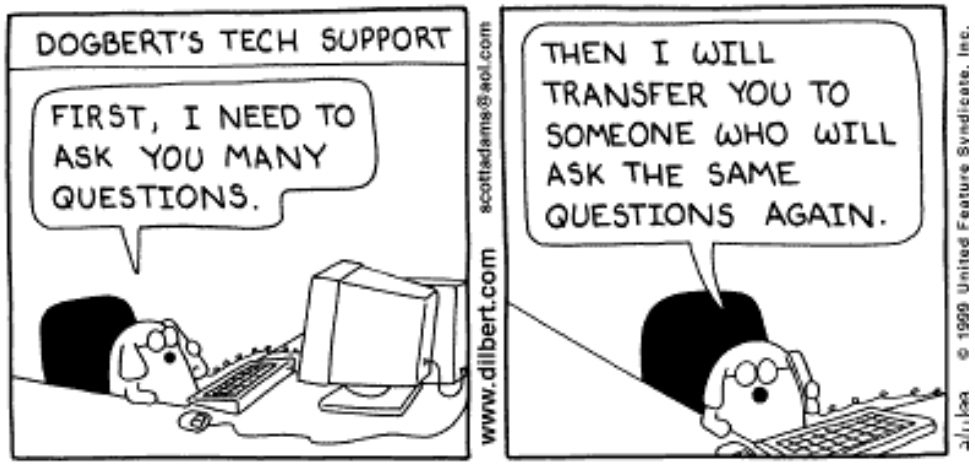
There are also some optional services and some for CIC/ROC and other such sites

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 techiques / 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: Requirements on Database Services

- WLCG measures Services through Site Functional Tests
- MoU targets aggregate individual services into higher level deliverables / responsibilities
  - e.g. 'acceptance of raw data from T0'
- Unclear if MoU targets are realistic
- They will be regularly reviewed by the GDB / MB
- See next slide for details of WLCG Service Coordination

WLCG Service Coordination

- Bi-weekly Service Coordination meetings held at CERN
- Weekly con-calls will probably split into two:
  - Focus on experiment usage of WLCG Services
    - Coupled with weekly "Task Force" meetings
  - Focus on setting up and running WLCG Services
- Quarterly WLCG Service Coordination Meetings
  - All Tier1s, main Tier2s, … minutes, agenda etc
- Bi-annual Service workshops
  - One at CERN (May?), one outside (September – October)
- Thematic workshops, site visits as required
  - Each Tier1 visited once per quarter
  - Regular 1-1 Video Meetings

Service Challenge 4 and the Production Service

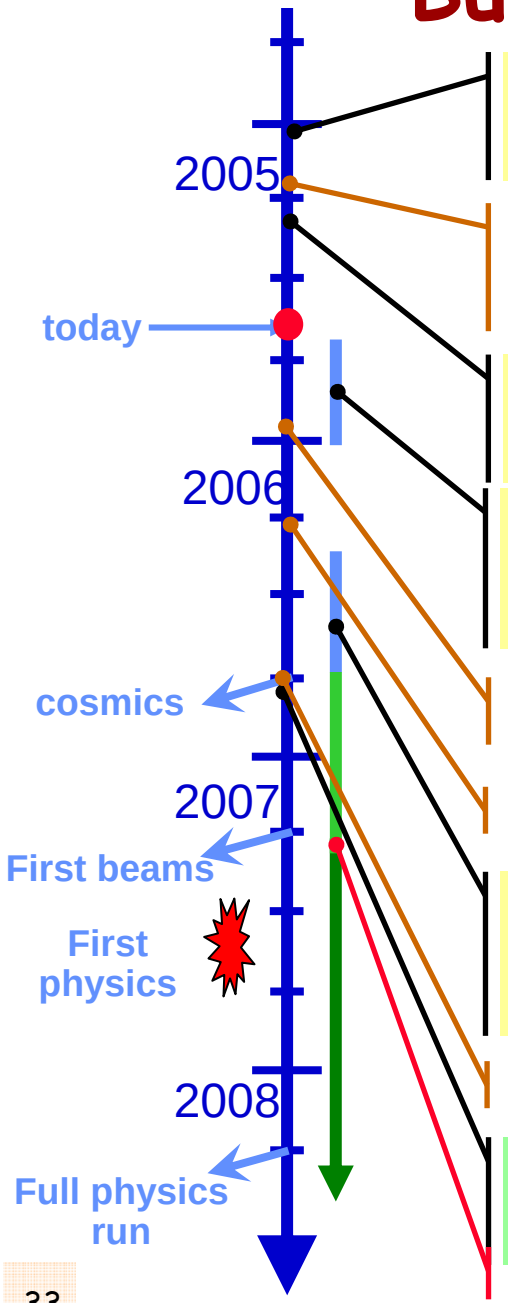
- The Service Challenge 4 setup is the Production Service
- All (LCG) Production is run in this environment
- There is no other…

So we decided to call it:

The Worldwide LCG Pilot Service (May on)

The Worldwide LCG Production Service (October on)

Building the WLCG Service



SC1 - *Nov04-Jan05* - data transfer between CERN and three Tier-1s (FNAL, NIKHEF, FZK)

DRC1 – *Mar05* - data recording at CERN sustained at 450 MB/sec for one week

SC2 – *Apr05* - data distribution from CERN to 7 Tier-1s – 600 MB/sec sustained for 10 days (one third of final nominal rate)

SC3 – *Sep-Dec05* - demonstrate reliable basic service – most Tier-1s, some Tier-2s; push up Tier-1 data rates to 150 MB/sec (60 MB/sec to tape)

DRC2 – *Dec05* - data recording sustained at 750 MB/sec for one week

DRC3 – *Apr06* - data recording sustained at 1 GB/sec for one week

SC4 – *May-Aug06* - demonstrate full service – all Tier-1s, major Tier-2s; full set of baseline services; data distribution and recording at nominal LHC rate (1.6 GB/sec)

DRC4 – *Sep06* - data recording sustained at 1.6 GB/sec

LHC Service in operation – *Sep06* – over following six months ramp up to full operational capacity & performance

LHC service commissioned – *Apr07*

First data in less than 2 years

- CERN + Tier-1s must provide an ***integrated*** and ***reliable*** service for the bulk data from first beams

➢ ***NOT an option to get things going later***

- Priority must be to concentrate on getting the basic service going
  - modest goals
  - pragmatic solutions
  - collaboration