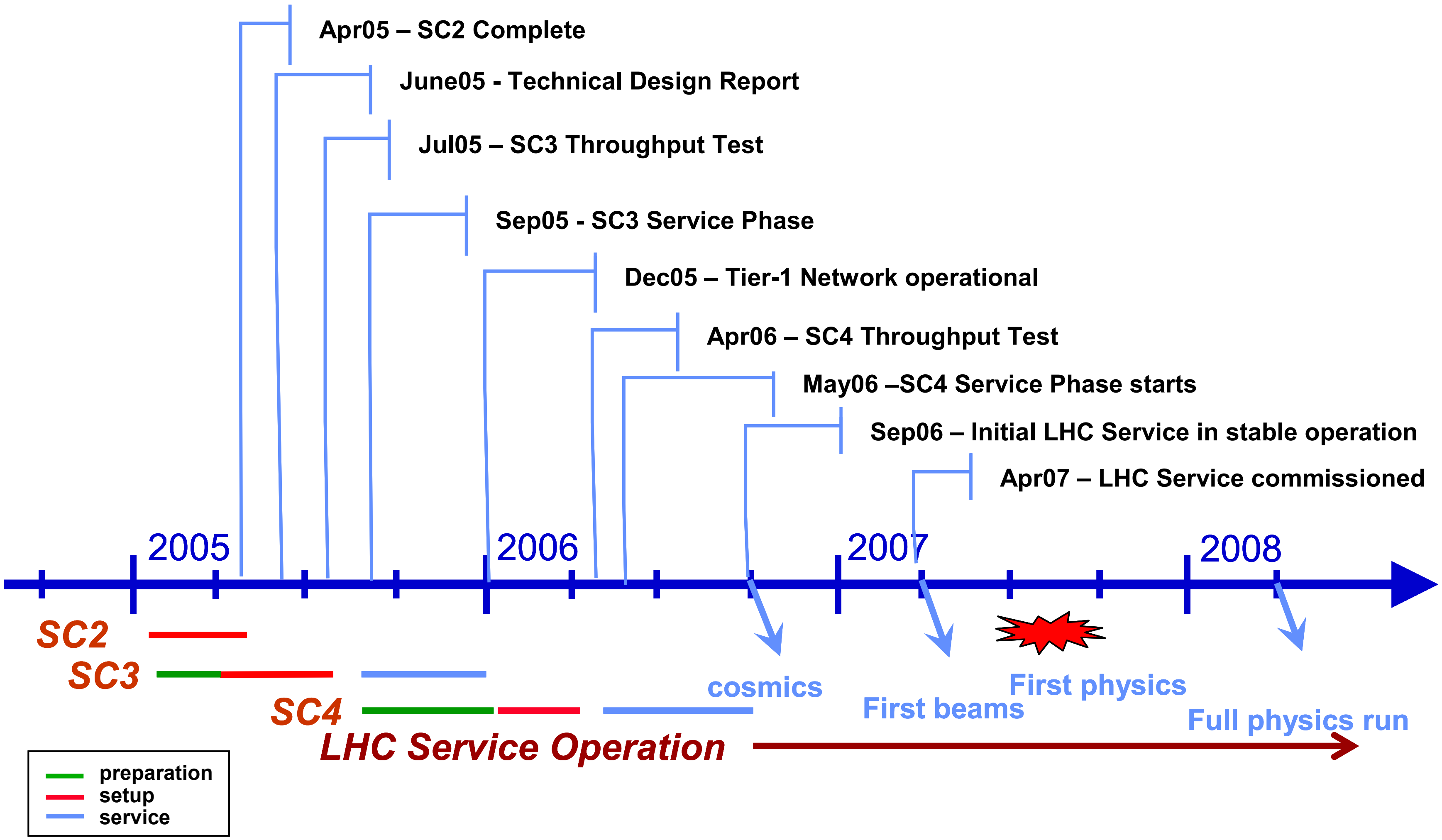
LCG Storage Management Workshop - Goals and Timeline of SC3

Jamie Shiers, CERN-IT-GD

April 2005

LCG Deployment Schedule



Apr05 – SC2 Complete

June05 - Technical Design Report

Jul05 – SC3 Throughput Test

Sep05 - SC3 Service Phase

Dec05 – Tier-1 Network operational

Apr06 – SC4 Throughput Test

May06 –SC4 Service Phase starts

Sep06 – Initial LHC Service in stable operation

Apr07 – LHC Service commissioned

2005

2006

2007

2008

SC2

SC3

SC4

LHC Service Operation

cosmics

First beams

First physics

Full physics run

preparation

setup

service

LCG Service Challenges - Overview

- LHC will enter production (physics) in April 2007
  - Will generate an enormous volume of data
  - Will require huge amount of processing power
- LCG 'solution' is a world-wide Grid
  - Many components understood, deployed, tested..
- But…
  - Unprecedented scale
  - Humungous challenge of getting large numbers of institutes and individuals, all with existing, sometimes conflicting commitments, to work together
- LCG must be ready at full production capacity, functionality and reliability in less than 2 years from now
  - Issues include h/w acquisition, personnel hiring and training, vendor rollout schedules etc.

➢ Should not limit ability of physicist to exploit performance of detectors nor LHC's physics potential

  - Whilst being stable, reliable and easy to use

LCG Project, Grid Deployment Group, CERN

Why Service Challenges?

To test Tier-0 ←→ Tier-1 ←→Tier-2 services

- Network service
  - Sufficient bandwidth: ~10 Gbit/sec
  - Backup path
  - Quality of service: security, help desk, error reporting, bug fixing, ..
- Robust file transfer service
  - File servers
  - File Transfer Software (GridFTP)
  - Data Management software (SRM, DCache)
  - Archiving service: tapeservers,taperobots, tapes, tapedrives, ..
- Sustainability
  - Weeks in a row un-interrupted 24/7 operation
  - Manpower implications: ~7 fte/site
  - Quality of service: helpdesk, error reporting, bug fixing, ..

➢ **Towards a stable production environment for experiments**

Whither Service Challenges?

- First discussions: GDB May - June 2004
  - May 18 - **Lessons from Data Challenges and planning for the next steps (+ Discussion)** (1h10') (transparencies)
  - June 15 - **Progress with the service plan team** (10') (document)
- Other discussions: PEB June 2004
  - June 8 - **Service challenges - proposal** (40') (transparencies)
  - June 29 - **Service challenges - status and further reactions** (30') (transparencies)
- May 2004 HEPiX
  - LCG Service Challenges Slides from Ian Bird (CERN)
- My involvement: from January 2005
  - Current Milestones: http://lcg.web.cern.ch/LCG/PEB/Planning/deployment/Grid%20Deployment%20Schedule.htm

Key Principles

- Service challenges results in a series of services that exist in parallel with baseline production service

- Rapidly and successively approach production needs of LHC

- Initial focus: core (data management) services

- Swiftly expand out to cover full spectrum of production and analysis chain

- Must be as realistic as possible, including end-end testing of key experiment use-cases over extended periods with recovery from glitches and longer-term outages

➢ Necessary resources and commitment pre-requisite to success!

- Effort should not be under-estimated!

SC1 Review

- ☹ SC1 did not successfully complete its goals
  - Dec04 - Service Challenge I complete
    - mass store (disk) - mass store (disk)
    - 3 T1s (Lyon, Amsterdam, Chicago) (others also participated…)
    - 500 MB/sec (individually and aggregate)
    - 2 weeks sustained
    - Software; GridFTP plus some scripts
- ➢ **We did not meet the milestone of 500MB/s for 2 weeks**
  - We need to do these challenges to see what actually goes wrong
    - A lot of things do, and did, go wrong
  - We need better test plans for validating the infrastructure before the challenges (network throughput, disk speeds, etc…)

SC1/2 - Conclusions

- Setting up the infrastructure and achieving reliable transfers, even at much lower data rates than needed for LHC, is complex and requires a lot of technical work + coordination
- Even within one site – people are working very hard & are stressed. Stressed people do not work at their best. Far from clear how this scales to SC3/SC4, let alone to LHC production phase
- Compound this with the multi-site / multi-partner issue, together with time zones etc and you have a large "non-technical" component to an already tough problem (example of technical problem follows…)
- But… the end point is fixed (time + functionality)
- We should be careful not to over-complicate the problem or potential solutions
- And not forget there is still a humungous amount to do…
- (much much more than we've done…)

Service Challenge 3 - Phases

High level view:

- Throughput phase
  - 2 weeks sustained in July 2005
    - "Obvious target" – GDB of July 20th
  - Primary goals:
    - 150MB/s disk – disk to Tier1s;
    - 60MB/s disk (T0) – tape (T1s)
  - Secondary goals:
    - Include a few named T2 sites (T2 -> T1 transfers)
    - **Encourage remaining T1s to start disk – disk transfers**
- Service phase
  - September – end 2005
    - Start with ALICE & CMS, add ATLAS and LHCb October/November
    - All offline use cases except for analysis
    - More components: WMS, VOMS, catalogs, experiment-specific solutions
  - Implies production setup (CE, SE, …)

SC3 – Will We Succeed?

- Throughput goals will almost certainly be achieved
- But at what cost in manpower and hardware?
- Are we really converging on goal of **production services?**
  - Monitoring, alarms, procedures, all working 24x7?
  - If this was a plane, would **you** fly in it?
- The test – let's try with some of the key people **on vacation** and set what happens…
  - Well OK, they can 'pretend' to be on vacation…

SC3 – Production Services

- SC3 is a relatively small step wrt SC2 (throughput!)
- We know we can do it technology-wise, but do we have a solution that will scale?
- **Let's make it a priority for the coming months to streamline our operations**
- And not just throw resources at the problem…
  - which we don't have…
- Whilst not forgetting 'real' goals of SC3…

SC3 – Service Phase

- It sounds easy:
  "all offline Use Cases except for analysis"
- And it some senses it is:
  these are well understood and tested
- So its clear what we have to do:
  - Work with the experiments to understand and agree on the experiment-specific solutions that need to be deployed
  - Agree on a realistic and achievable work-plan that is consistent with overall goals / constraints
- Either that or send a 'droid looking for Obi-Wan Kenobi…

Service Phase - Priorities

- Experiments have repeatedly told us to focus on reliability and functionality
- This we need to demonstrate as a first step…

➢ **But cannot lose sight of need to pump up data rates – whilst maintaining production service – to pretty impressive "DC" figures**

SC3 on

- SC3 is <u>**significantly**</u> more complex than previous challenges
- It includes experiments s/w, additional m/w, Tier2s etc
  - Proving we can transfer dummy files from A-B proves nothing
  - Obviously need to show that basic infrastructure works…

- Preparation for SC3 includes:
  - Understanding experiments' Computing Models
  - Agreeing involvement of experiments' production teams
  - Visiting all (involved) Tier1s (multiple times)
  - Preparing for the involvement of 50-100 Tier2s

- Short of resources at all levels:
  - "Managerial" – discussing with experiments and Tier1s (visiting)
  - "Organizational" – milestones, meetings, workshops, …
  - "Technical" – preparing challenges and running CERN end – 24 x 7 ???

LCG Project, Grid Deployment Group, CERN

2005 Q1 - SC3 preparation

Prepare for the next service challenge (SC3)
-- in parallel with SC2 (reliable file transfer) –

Build up 1 GByte/s *challenge* facility at CERN

- **The current 500 MByte/s facility used for SC2 will become the *testbed* from April onwards (10 ftp servers, 10 disk servers, network equipment)**

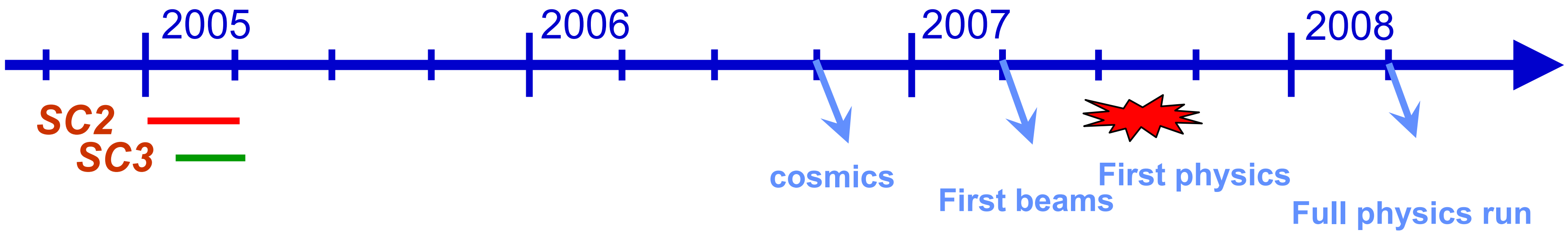
Build up infrastructure at each external centre

- **Average *capability* ~150 MB/sec at a Tier-1 (to be agreed with each T-1)**

Further develop reliable transfer framework software

- **Include catalogues, include VO's**

disk-network-disk bandwidths



LCG Project, Grid Deployment Group, CERN

2005 Q2-3 - SC3 challenge

SC3 - 50% service infrastructure

- Same T1s as in SC2 (Fermi, NIKHEF/SARA, GridKa, RAL, CNAF, CCIN2P3)
- Add at least two T2s
- "50%" means approximately 50% of the nominal rate of ATLAS+CMS

Using the 1 GByte/s *challenge* facility at CERN -

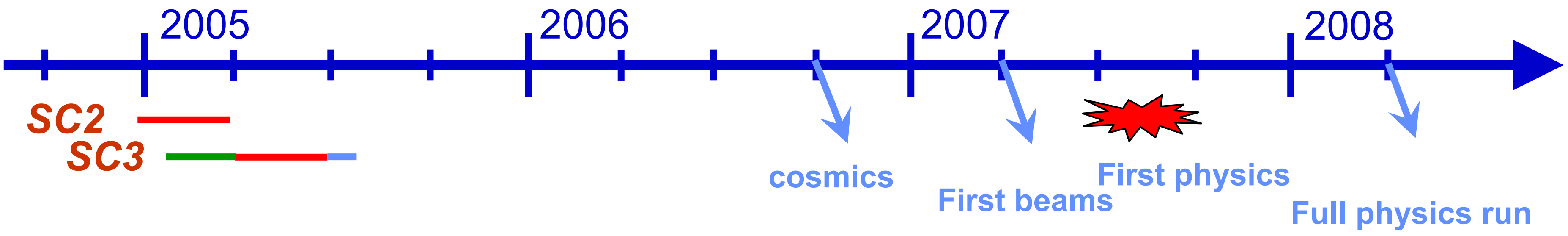
- Disk at T0 to tape at all T1 sites at 60 Mbyte/s
- Data recording at T0 from same disk buffers
- Moderate traffic disk-disk between T1s and T2s

tape-network-disk bandwidths

Use ATLAS and CMS files, reconstruction, ESD skimming codes
(numbers to be worked out when the models are published)

Goal - 1 month sustained service in July

- 500 MBytes/s aggregate at CERN, 60 MBytes/s at each T1
- ➔ end-to-end data flow peaks at least a factor of two at T1s
- → network bandwidth peaks ??



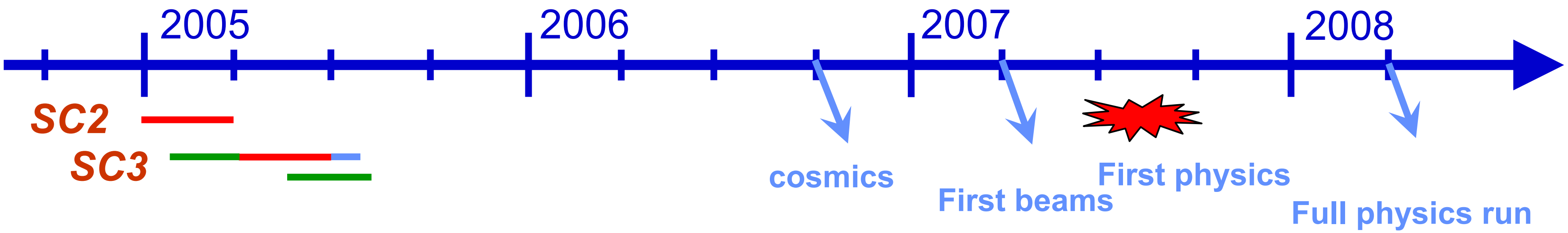
2005 Q2-3 - SC3 additional centres

In parallel with SC3 prepare additional centres using the 500 MByte/s test facility

- Test Taipei, Vancouver, Brookhaven, additional Tier-2s

Further develop framework software

- Catalogues, VO's, use experiment specific solutions



2005 Sep-Dec - SC3 Service

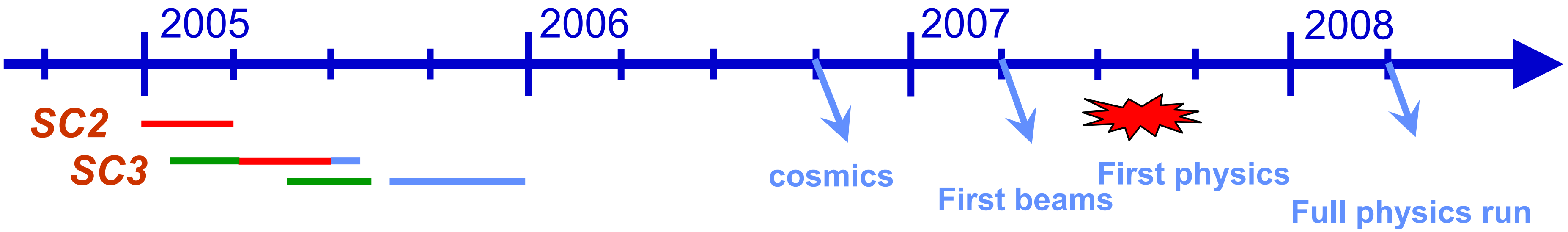
50% Computing Model Validation Period

The service exercised in SC3 is made available to experiments as a stable, permanent service for computing model tests

Additional sites are added as they come up to speed

End-to-end *sustained* data rates –

- 500 Mbytes/s at CERN (aggregate)
- 60 Mbytes/s at Tier-1s
- Modest Tier-2 traffic



SC3 – Milestone Decomposition

- File transfer goals:
  - Build up disk – disk transfer speeds to 150MB/s
    - SC2 was 100MB/s – agreed by site
  - Include tape – transfer speeds of 60MB/s
- Tier1 goals:
  - Bring in additional Tier1 sites wrt SC2
    - PIC and Nordic most likely added later: SC4?
- Tier2 goals:
  - Start to bring Tier2 sites into challenge
    - Agree services T2s offer / require
    - On-going plan (more later) to address this via GridPP, INFN etc.
- Experiment goals:
  - Address main offline use cases ***except*** those related to analysis
    - i.e. real data flow out of T0-T1-T2; simulation in from T2-T1
- Service goals:
  - Include CPU (to generate files) and storage
  - Start to add additional components
    - Catalogs, VOs, experiment-specific solutions etc, 3D involvement, …
    - Choice of software components, validation, fallback, …

SC3 – Experiment Goals

- Meetings on-going to discuss goals of SC3 and experiment involvement
- Focus on:
  - First demonstrate robust infrastructure;
  - Add 'simulated' experiment-specific usage patterns;
  - Add experiment-specific components;
  - Run experiments offline frameworks but don't preserve data;
    - Exercise primary Use Cases ***except*** analysis (SC4)
  - Service phase: data is preserved…
- **Has significant implications on resources beyond file transfer services**
  - Storage; CPU; Network… Both at CERN and participating sites (T1/T2)
  - May have different partners for experiment-specific tests (e.g. not all T1s)
- **In effect, experiments' usage of SC during service phase = data challenge**
- Must be **exceedingly clear** on goals / responsibilities during each phase!

SC3 Preparation Workshop

- This (proposed) workshop will focus on very detailed technical planning for the whole SC3 exercise.

- **It is intended to be as interactive as possible, i.e. not presentations to an audience largely in a different (wireless) world.**

- There will be sessions devoted to specific experiment issues, Tier1 issues, Tier2 issues as well as the general service infrastructure.

- Planning for SC3 has already started and will continue prior to the workshop.

- This is an opportunity to get together to iron out concerns and issues that cannot easily be solved by e-mail, phone conferences and/or other meetings prior to the workshop.

➢ **Is there a better way to do it? Better time?**

SC3 – Experiment Involvement Cont.

- Regular discussions with experiments have started
  - ATLAS: at DM meetings
  - ALICE+CMS: every ~2 weeks
  - LHCb: no regular slot yet, but discussions started…
- Anticipate to start first with ALICE and CMS (exactly when TDB) ATLAS and LHCb around October
  - T2 sites being identified in common with these experiments
    - More later…
  - List of experiment-specific components and the sites where they need to be deployed being drawn up
    - Need this on April timeframe for adequate preparation & testing

Experiment plans - Summary

- SC3 phases
  - Setup and config - July + August
  - Experiment software with throwaway data - September
  - Service phase
    - ATLAS – Mid October
    - ALICE – July would be best…
    - LHCb – post-October
    - CMS – July (or sooner)
- Tier-0 exercise
- Distribution to Tier-1
- …

A Simple T2 Model

N.B. this may vary from region to region

- Each T2 is configured to upload MC data *to* and download data *via* a given T1
- In case the T1 is logical unavailable, wait and retry
  - MC production might eventually stall
- For data download, retrieve via alternate route / T1
  - Which may well be at lower speed, but hopefully rare
- Data residing at a T1 other than 'preferred' T1 is transparently delivered through appropriate network route
  - T1s are expected to have at least as good interconnectivity as to T0
- Each Tier-2 is associated with a Tier-1 who is responsible for getting them set up
- Services at T2 are managed storage and reliable file transfer
  - DB component at T1; user agent also at T2
- 1GBit network connectivity – shared (less will suffice to start with, more maybe needed!)

Prime Tier-2 sites

- For SC3 we aim for

Site	Tier1	Experiment
Bari, Italy	CNAF, Italy	CMS
Turin, Italy	CNAF, Italy	Alice
DESY, Germany	FZK, Germany	ATLAS, CMS
Lancaster, UK	RAL, UK	ATLAS
London, UK	RAL, UK	CMS
ScotGrid, UK	RAL, UK	LHCb
US Tier2s	BNL / FNAL	ATLAS / CMS

- Responsibility between T1 and T2 (+ experiments)
- CERN's role limited
  - Develop a manual "how to connect as a T2"
  - Provide relevant s/w + installation guides
  - Assist in workshops, training etc.
- Other interested parties: Prague, Warsaw, Moscow, ..
- ➢ Also attacking larger scale problem through national / regional bodies
  - GridPP, INFN, HEPiX, US-ATLAS, US-CMS

Tier2 Region	Coordinating Body	Comments
Italy	INFN	A workshop is foreseen for May during which hands-on training on the Disk Pool Manager and File Transfer components will be held.
UK	GridPP	A coordinated effort to setup managed storage and File Transfer services is being managed through GridPP and monitored via the GridPP T2 deployment board.
Asia-Pacific	ASCC Taipei	The services offered by and to Tier2 sites will be exposed, together with a basic model for Tier2 sites at the Service Challenge meeting held at ASCC in April 2005.
Europe	HEPiX	A similar activity will take place at HEPiX at FZK in May 2005, together with detailed technical presentations on the relevant software components.
US	US-ATLAS and US-CMS	Tier2 activities in the US are being coordinated through the corresponding experiment bodies.
Canada	Triumf	A Tier2 workshop will be held around the time of the Service Challenge meeting to be held in Triumf in November 2005.
Other sites	CERN	One or more workshops will be held to cover those Tier2 sites with no obvious regional or other coordinating body, most likely end 2005 / early 2006.

LCG Project, Grid Deployment Group, CERN

Conclusions

- To be ready to fully exploit LHC, significant resources need to be allocated to a series of **Service Challenges** by all concerned parties

- These challenges should be seen as an **essential** on-going and **long-term** commitment to achieving production LCG

- The countdown has started – we are already in (pre-)production mode

- Next stop: 2020