

# LCG Service Challenge 3: Goals and Milestones



Jamie Shiers, CERN-IT-GD  
15 March 2005



# Agenda

- Reminder of motivation for Service Challenges
- High-level milestones for SC3
- Drill-down to T1/T2 view and Experiment view
- Possible detailed milestones
- Summary and Conclusions

# 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

# 2005 Q1 - SC2

## SC2 - Robust Data Transfer Challenge

Set up infrastructure for 6 sites

- Fermi, NIKHEF/SARA, GridKa, RAL, CNAF, CCIN2P3

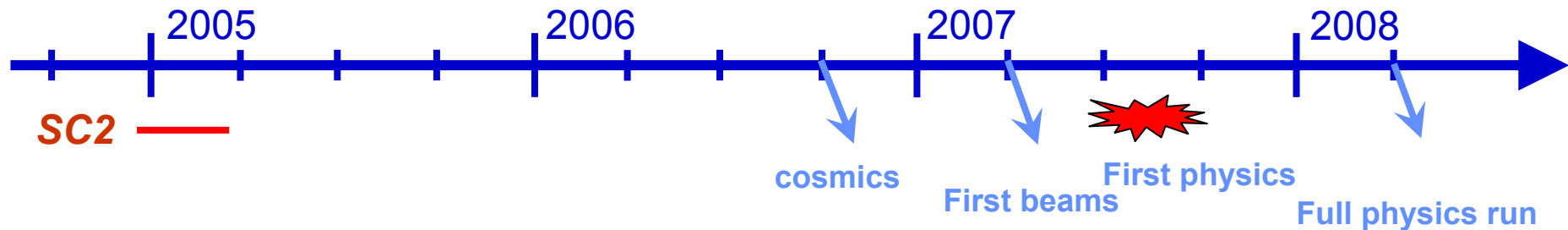
Test sites individually at 100MB/s (agreed by site)

- at least two at 500 MByte/s with CERN

Agree on sustained data rates for each participating centre

Goal - by end March sustained 500 Mbytes/s aggregate at CERN

**These are the old slides unedited by the Ministry of Truth**



# 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

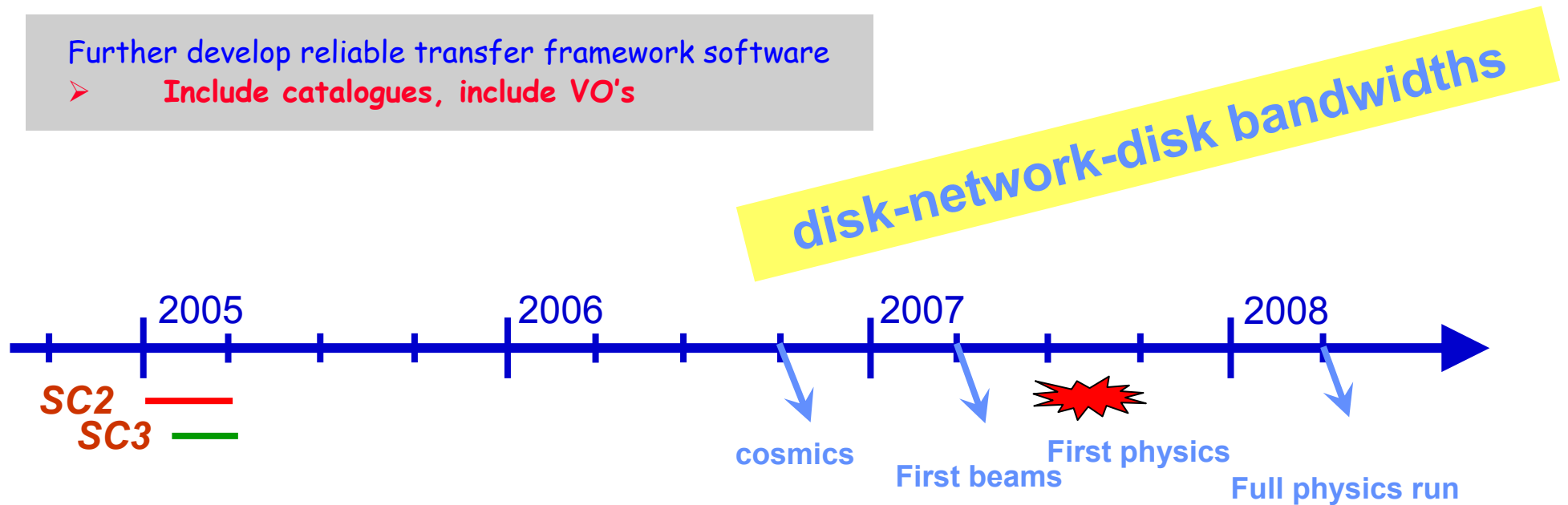
- The current 500 MByte/s facility used for SC2 will become the *testbed* from April onwards (10 ftp servers, 10 disk servers, network equipment)

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



# 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 -

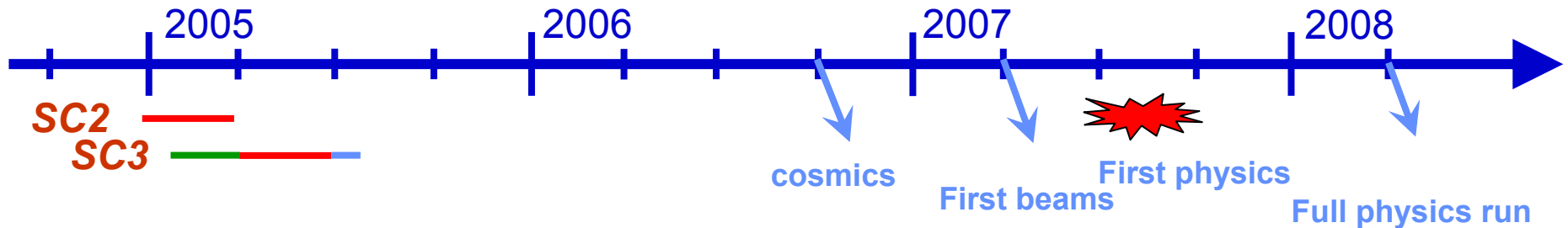
- Disk at T0 to tape at all T1 sites at 60 Mbyte/s
- Data recording at T0 from same disk buffers
- Moderate traffic disk-disk between T1s and T2s

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 ??

tape-network-disk  
bandwidths



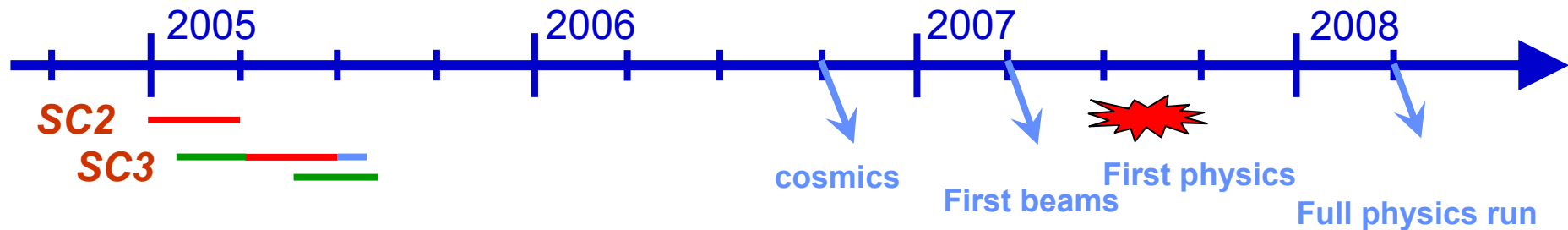
## 2005 Q2-3 - SC3 additional centres

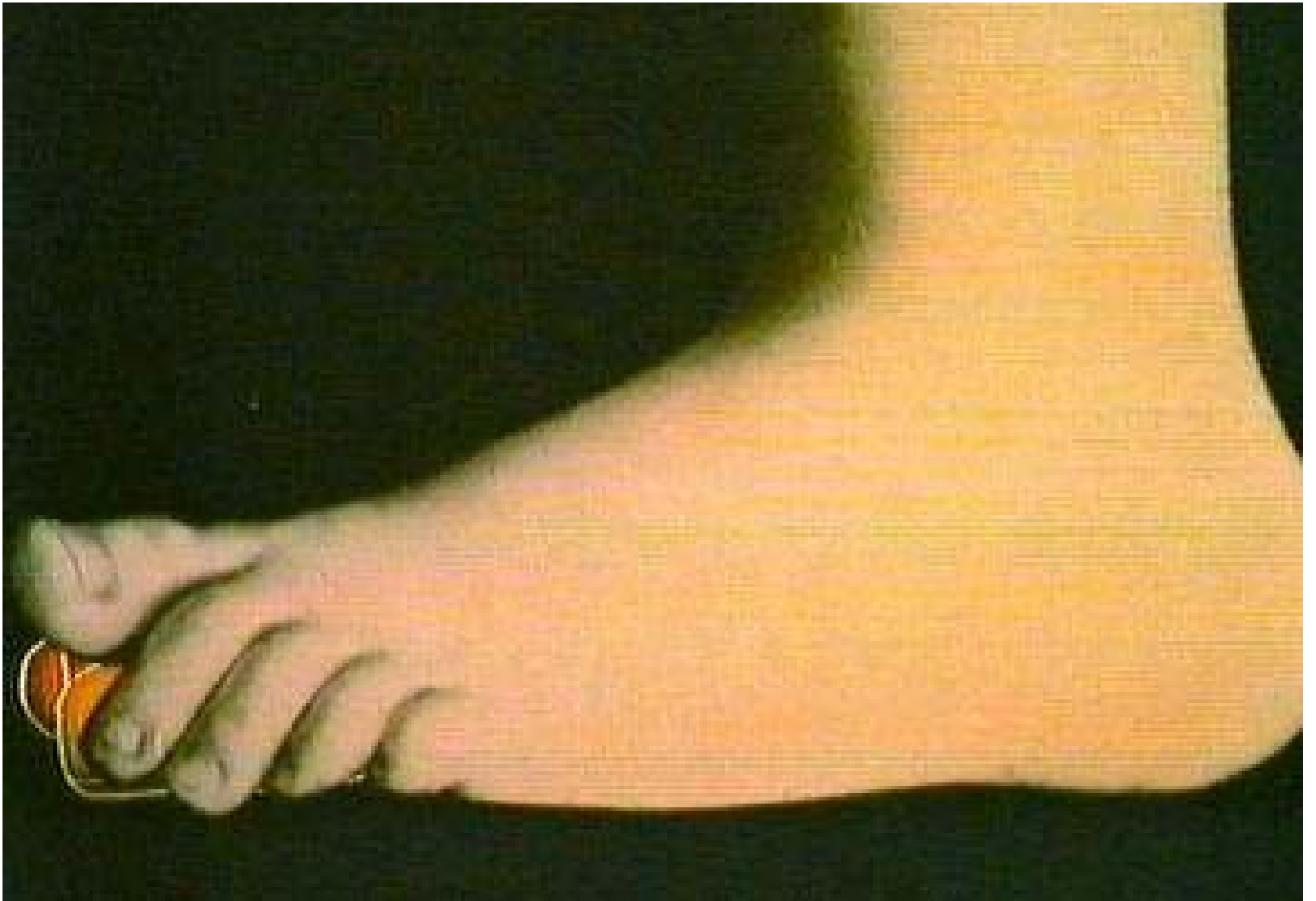
In parallel with SC3 prepare additional centres using the 500 MByte/s test facility

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

Further develop framework software

- Catalogues, VO's, use experiment specific solutions







## Progress Since Feb SC Meeting

- SC3 meetings have started with the experiments
- T2 discussions at GridPP T2 Deployment Board
- T2 discussions during CNAF visit

### Primary goal of Service Challenges is:

- Ramp up LCG service to meet full production needs of experiments
- Service Challenges are part of the process - not the goal itself!

# 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 - 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, ...

## T2 Sites - Introduction

- Roles of Tier2 sites
- Services they require (and offer)
- Timescale for involving T2s in LCG SCs
- Simplest (useful) T2 Model
- Possible procedure

## The Problem (or at least part of it...)

- SC1 - December 2004
  - SC2 - March 2005
- } Neither of these involve T2s or even the experiments – just basic infrastructure
- SC3 - from July 2005 involves 2 Tier2s
    - + experiments' software + catalogs + other additional stuff
  - SCn - completes at least 6 months prior to LHC data taking. Must involve all Tier1s and ~all Tier2s
  - Not clear how many T2s there will be
  -  Maybe **50**; maybe **100** - still a *huge* number to add!
    - ALICE: 15? ,ATLAS: 30, CMS: 25, LHCb: 15; overlap?

## Tier2 Roles

- Tier2 roles vary by experiment, but include:
  - Production of simulated data;
  - Production of calibration constants;
  - Active role in [end-user] analysis
- Must also consider services offered to T2s by T1s
  - e.g. safe-guarding of simulation output;
  - Delivery of analysis input.
- No fixed dependency between a given T2 and T1
  - But 'infinite flexibility' has a cost...

## Basic T2 Services

- T2s will need to provide services for data up- and down-load
- Assume that this uses the same components as between T0 and T1s
- Assume that this also includes an SRM interface to local disk pool manager
  - This / these should / could also be provided through LCG
- Networking requirements need to be further studied but current estimates suggest 1Gbit connections will satisfy needs of a given experiment
  - 💣 Can be dramatically affected by analysis model, heavy ion data processing etc

# 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



## Possible T2 Procedure(?)

- Work through existing structures where possible
  - INFN in Italy, GridPP in the UK, HEPiX, etc.
  - USATLAS, USCMS etc.
    - Probably need also some 'regional' events
    - Complemented with workshops at CERN and elsewhere
  
- Work has already started with GridPP and INFN
  - ☺ Choosing candidate sites for SC3 involvement (together with expts)
  - ☺ Longer term goal of adding all T2s over coming year

## Which Candidate T2 Sites?

- Would be useful to have:
  - Good local support from relevant experiment(s)
  - Some experience with disk pool mgr and file transfer s/w
  - 'Sufficient' local CPU and storage resources
  - Manpower available to participate in SC3+
    - And also define relevant objectives?
  - 1Gbit/s network connection to T1?
    - Probably a luxury at this stage...
- First T2 site(s) will no doubt be a learning process
- Hope to (semi-)automate this so that adding new sites can be achieved with low overhead

## From GridPP Summary...

- Coordinator for T2 work
- Network expert
  - Understand how sites will be linked and ensure appropriate network is provisioned
- Storage expert
  - Someone to understand how the SRM and storage should be installed and configured
- Experiment experts
  - To drive the experiment needs
  - How are we going to do the transfers?
  - What do the experiments want to achieve out of the exercise?

## T2 Summary

- Activities launched in the UK and Italy
  - Other regions will follow - building on experience from these
  - Will continue to discuss these issues as we visit sites / regions
- ☺ Very encouraged by how this activity is taking off

# 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 started to discuss goals of SC3 and experiment involvement
- Presentations on these later today...
- 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; ...

## SC3 - Experiment Involvement Cont.

- Need to agree work plan and timeline with each experiment over coming weeks
- Unrealistic to schedule all simultaneously
- Resources (as always) and avoiding unrealistic expectations likely to be main problem(s)...
- But.. as always.. end point is fixed!
- Cannot afford to defer too many items to future challenges!

# 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, ...



## Service Goals

- Expect relatively modest increase in 'service' components
  - File catalog based on agreement from Baseline Services WG
  - Other services agreed by BSWG
- Experiment-specific components and impact on other services, e.g. Distributed Database Services, need to be clarified as soon as possible
- Similarly, requirements for processing power and storage at all sites involved (T0, T1, T2)

# 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, ...

## Ramp-up of FTS and T1s

- As expected, this is proving to be a significant amount of work!
- Need to keep up momentum:
  - Increasing file transfer performance;
  - Adding tape;
  - Building up network links;
  - Bringing additional T1s into the challenges...

## File Transfer Milestones

- Need to understand how infrastructure at CERN will build up to address file transfer goals of SC3
  - Internal / External network infrastructure
  - File transfer Server nodes
  - Access to Experiment files
- For each Tier1 need to understand how local infrastructure and network connectivity will be build up to meet requirements
- For Tier2s, assume "Minimal Usable Model" (next)...



- Additional threads started to address:
  - Experiment involvement;
  - Bringing T2s in SC3;
  - Longer-term goals of bringing all T2s into the LCG Service (Challenges)
- The enthusiasm and support provided to these new activities is much appreciated
- We have a lot of work ahead...
- ...but the problem is beginning to become tractable(?)

**Backup Slides**

## 2005 Sep-Dec - SC4 preparation

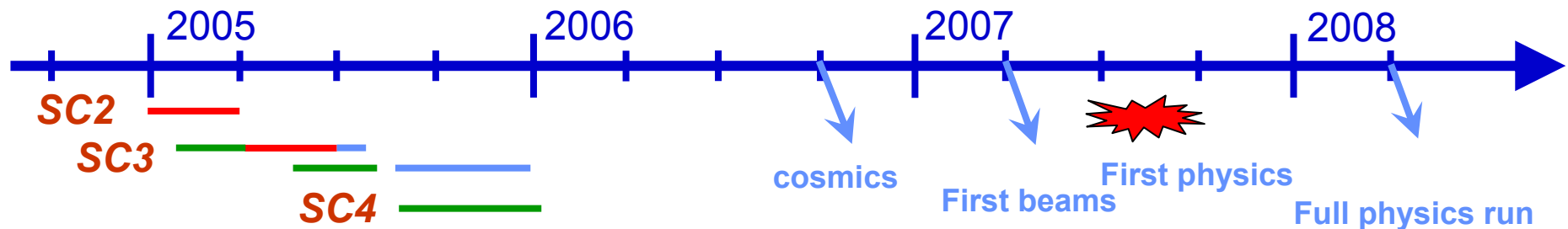
In parallel with the SC3 model validation period,  
in preparation for the first 2006 service challenge (SC4) -

Using 500 MByte/s test facility

- test PIC and Nordic T1s
- and T2's that are ready (Prague, LAL, UK, INFN, ..)

Build up the production facility at CERN to 3.6 GBytes/s

Expand the capability at all Tier-1s to full nominal data rate



# 2006 Jan-Aug - SC4

## SC4 - full computing model services

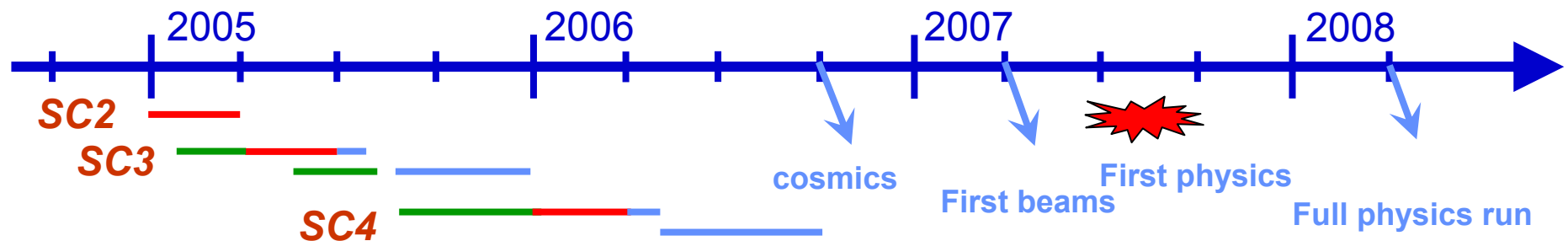
- Tier-0, ALL Tier-1s, all major Tier-2s operational at full target data rates (~2 GB/sec at Tier-0)
- acquisition - reconstruction - recording - distribution, PLUS ESD skimming, servicing Tier-2s

Goal - stable test service for one month - April 2006

## 100% Computing Model Validation Period (May-August 2006)

Tier-0/1/2 full model test - All experiments

- 100% nominal data rate, with processing load scaled to 2006 cpus



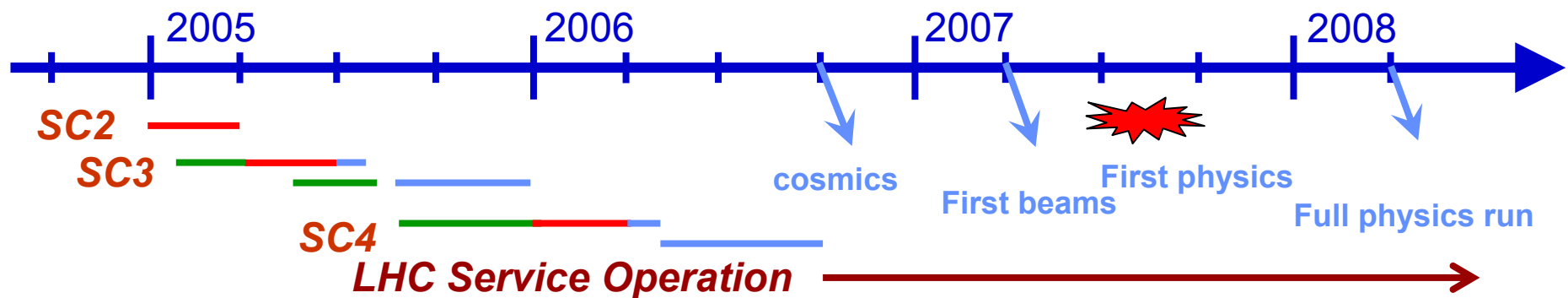


## 2006 Sep - LHC service available

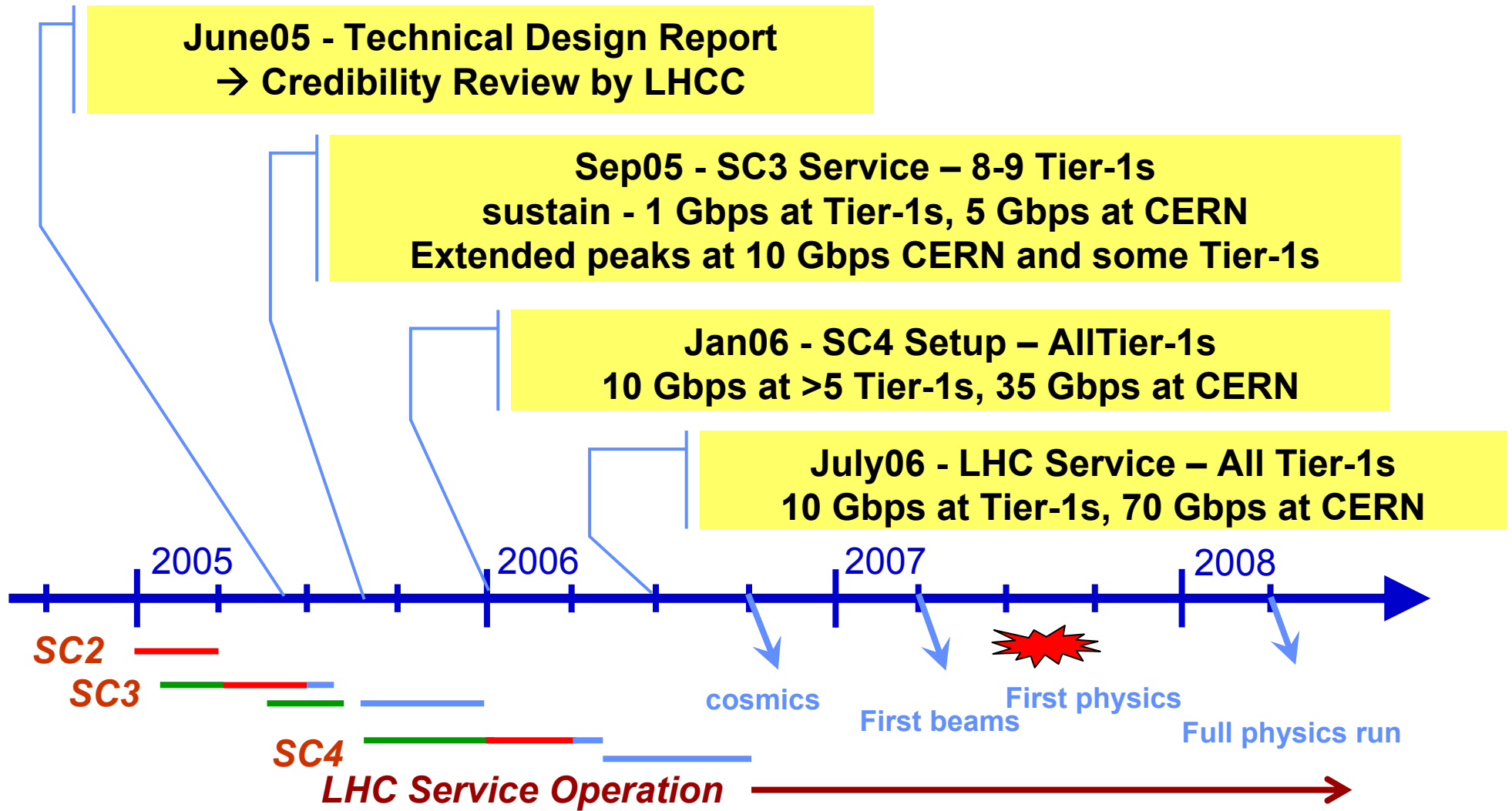
The SC4 service becomes the permanent LHC service - available for experiments' testing, commissioning, processing of cosmic data, etc.

All centres ramp-up to capacity needed at LHC startup

- TWICE nominal performance
- Milestone to demonstrate this 3 months before first physics data  
→ April 2007



# Key dates for Connectivity



# Key dates for Services

