



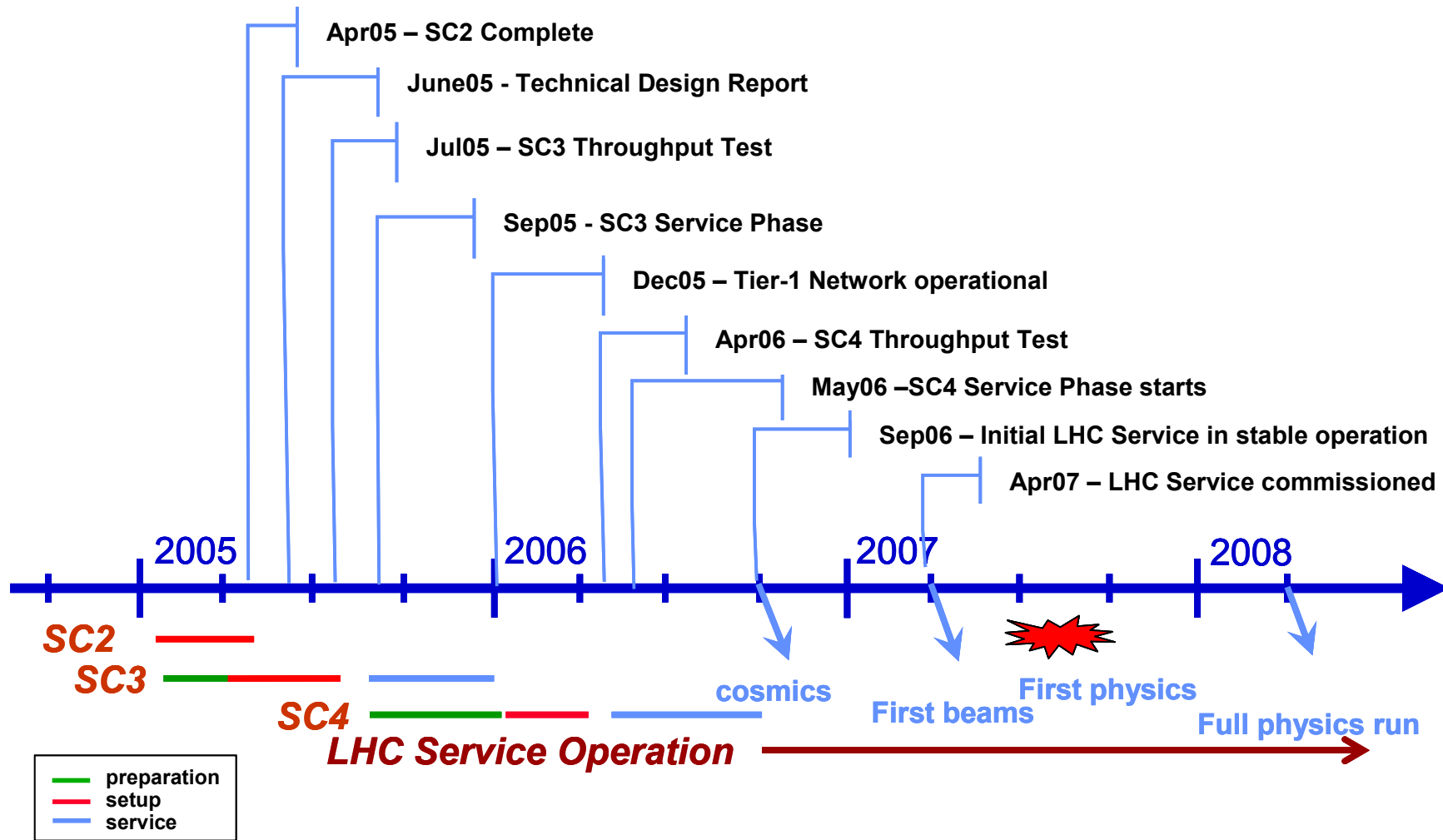
The Service is
the Challenge

Antarctica

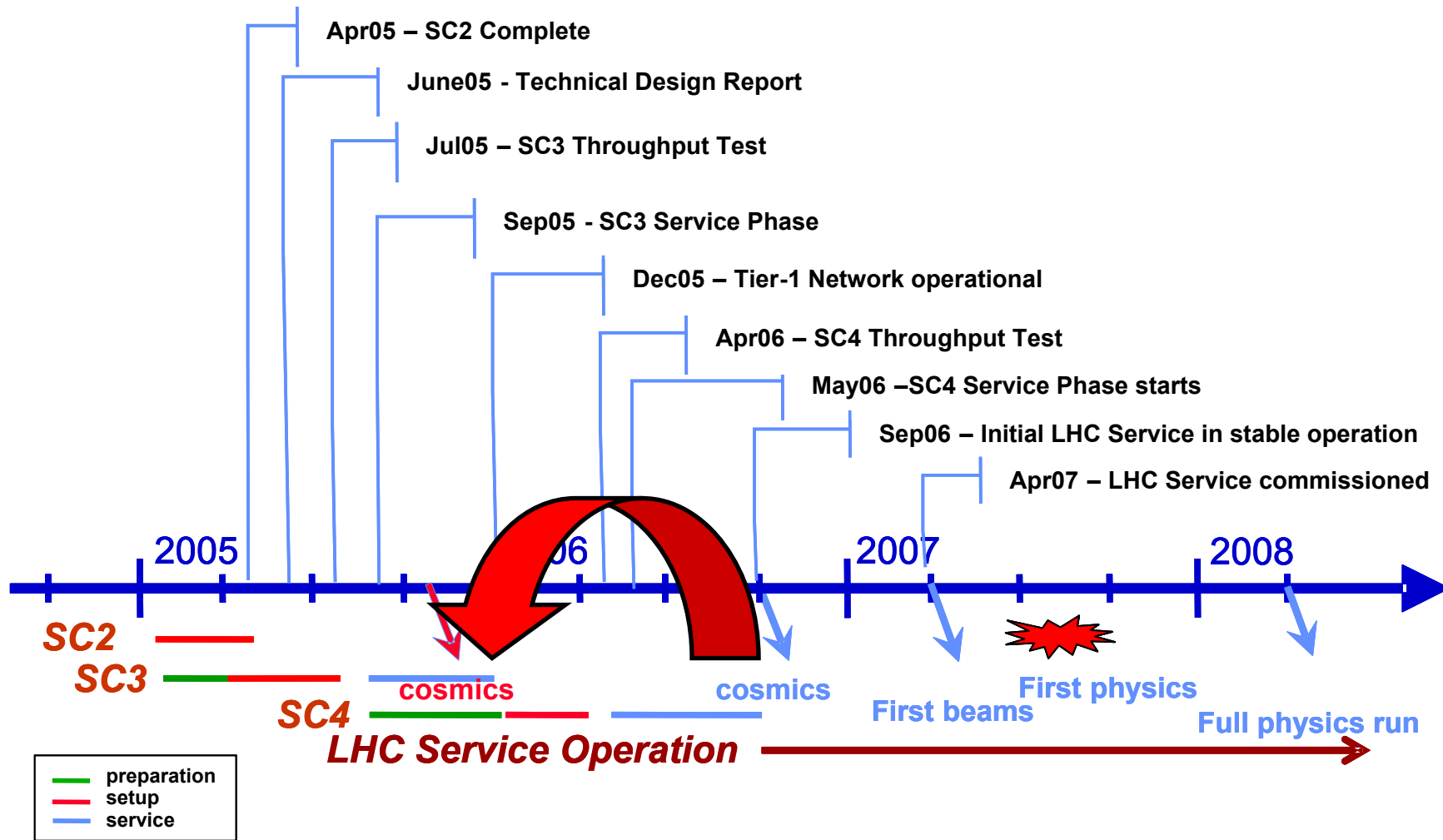
Experiment Requirements - Implications

- What services do sites have to offer?
- What resources do sites have to provide?
- What is the schedule?
- Start with the basic services and work upwards...
- Goal is to specify things in detail...
- Motivation: what is defined is not part of the challenge

LCG Deployment Schedule



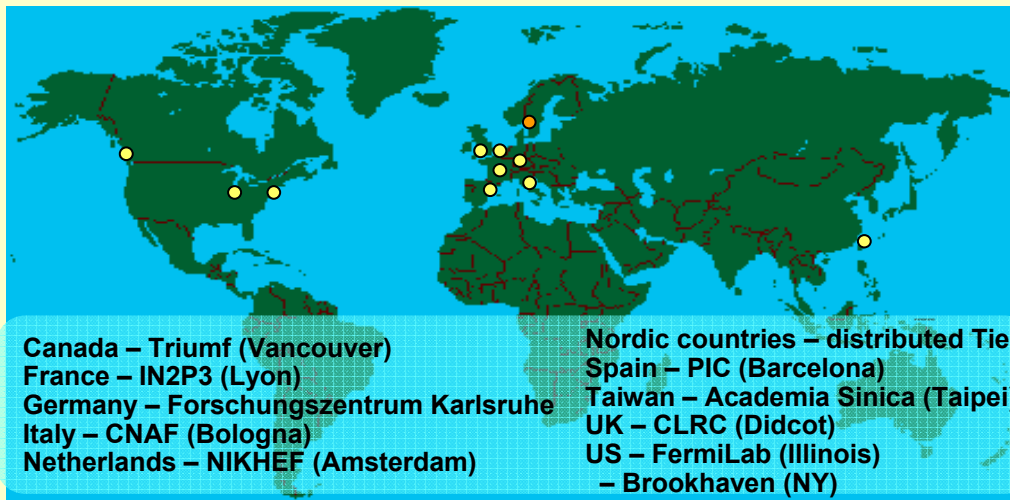
LCG Deployment Schedule



LCG Service Hierarchy

Tier-0 - the accelerator centre

- Data acquisition & initial processing
- Long-term data curation
- Distribution of data → Tier-1 centres



- | | |
|---------------------------------------|---------------------------------------|
| Canada – Triumf (Vancouver) | Nordic countries – distributed Tier-1 |
| France – IN2P3 (Lyon) | Spain – PIC (Barcelona) |
| Germany – Forschungszentrum Karlsruhe | Taiwan – Academia Sinica (Taipei) |
| Italy – CNAF (Bologna) | UK – CLRC (Didcot) |
| Netherlands – NIKHEF (Amsterdam) | US – FermiLab (Illinois) |
| | – Brookhaven (NY) |

Tier-1 - "online" to the data acquisition process → high availability

- Managed Mass Storage -
→ grid-enabled data service
- Data intensive analysis
- National, regional support

Tier-2 - ~100 centres in ~40 countries

- Simulation
- End-user analysis – batch and interactive

Networking

- Latest estimates are that Tier-1s will need connectivity at **~10 Gbps** with **~70 Gbps** at CERN
- There is no real problem for the technology as has been demonstrated by a succession of Land Speed Records
- But LHC will be one of the few applications needing –
 - this level of performance as a service on a global scale
- We have to ensure that there will be an effective international backbone –
 - that reaches through the national research networks to the Tier-1s
- LCG has to be pro-active in working with service providers
 - Pressing our requirements and our timetable
 - Exercising pilot services

Tier0 - Mandatory Services

- Data acquisition & initial processing
- Long-term data curation
- Distribution of data to Tier-1 centres



Mandatory Services

- The (DM) services LCG absolutely must provide (IMHO) are:
 1. Management of the primary copies of the RAW data, ESD etc.
 2. Reliable file transfer service to Tier1s (including networking)...
- Baseline services: as per Baseline Services Working Group
- Additional services include file catalog, conditions database etc.
 - We managed these for LEP and LEP-era experiments
- The focus must be on the mandatory services which simply cannot be called 'rock solid' today

Tier-1 Centres

				ALICE	ATLAS	CMS	LHCb	
1	GridKa	Karlsruhe	Germany	X	X	X	X	4
2	CCIN2P3	Lyon	France	X	X	X	X	4
3	CNAF	Bologna	Italy	X	X	X	X	4
4	NIKHEF/SARA	Amsterdam	Netherlands	X	X		X	3
5	NDGF	Distributed	Dk, No, Fi, Se	X	X			1
6	PIC	Barcelona	Spain		X	X	X	3
7	RAL	Didcot	UK	X	X	X	X	4
8	Triumf	Vancouver	Canada		X			1
9	BNL	Brookhaven	US		X			1
10	FNAL	Batavia, Ill.	US			X		1
11	ASCC	Taipei	Taiwan		X	X		2
				6	10	7	6	

A US Tier1 for ALICE is also expected.

Core Site Services

- **CERN**
 - Storage: Castor/SRM
 - File catalogue: POOL LFC Oracle
- **FNAL**
 - Storage: dCache/SRM
 - File catalogue: POOL Globus RLS
- **CNAF**
 - Storage: Castor/SRM
 - File catalogue: POOL LFC Oracle
- **RAL**
 - Storage: dCache/SRM
 - File catalogue: POOL LFC Oracle?
- **IN2P3**
 - Storage: dCache/SRM
 - File catalogue: POOL LFC Oracle
- **SARA/NIKHEF**
 - Storage: dCache/SRM
 - File catalogue: POOL LFC MySQL(?)
- **PIC**
 - Storage: Castor/SRM
 - File catalogue: POOL LFC MySQL
- **FZK**
 - Storage: dCache/SRM
 - File catalogue: POOL LFC Oracle
- **ASCC**
 - Storage: Castor/SRM
 - File catalogue: POOL LFC Oracle
- **BNL**
 - Storage: dCache/SRM
 - File catalogue: POOL LFC Oracle
- **TRIUMF**
 - Storage: dCache/SRM
 - File catalogue: POOL LFC MySQL(?)
- **NDGF**
 - Storage:
 - File catalogue:

Running FTS service for T2s

Services

- Need to define service level
- Proposal: something 'soft' for SC3 (learning period)
- Something appropriate for post-SC3 service
- Pilot:
 - 8 x 5 best effort response
 - Is backup required?
- SC3 services:
 - 24x7 response
 - Recovery of e.g. DB to yesterday's backup
- Post-SC3 services
 - 24x7 response
 - Recovery to now - 3 hours(?); by VO? By subdirectory (tree)?
- Need to set expectations realistically: time to react; time to resolve etc.
 - Installing e.g. LFC is one thing - offering a service is another..

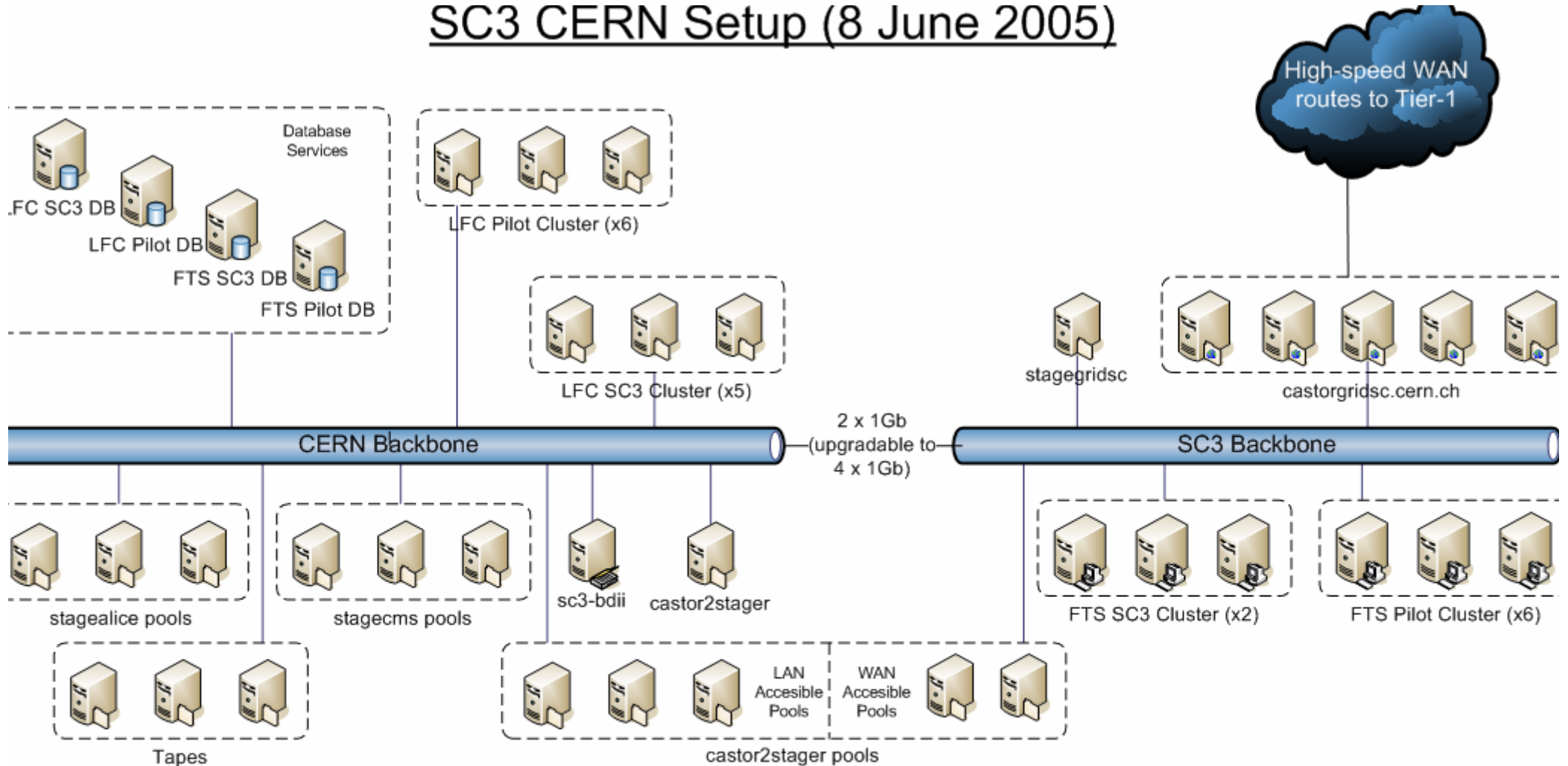
Additional Components

- These include “Agents and Daemons” (next);
- **Applications Area-related services;**
 - E.g. COOL for ATLAS; LHCb, ...
- **Applications Area-software;**
 - E.g. GEANT4, ROOT, POOL, SEAL, ...
- **Experiment-specific software and services... (many using / requiring MySQL...)**
 - E.g. ATLAS HVS, book-keeping systems (and a lot more mentioned during Rome production...)
- **We have to document all of these (and any others on the critical path for SC3)...**
- Cannot afford for production to stop due to a lack of documentation / procedures...
- Many of these also have implications on Tier1s and Tier2s...

Agents and Daemons

- This is something clearly identified during BSWG discussions
- And corresponds to what is running on lxgate machines at CERN today...
- Current assumption: Each experiment requires **ONE** such box at **EACH** site
 - T0, T1, T2 (some expt ppts suggest even more...)
- Must agree on minimal service level:
- Standard system; installation; box-level monitoring; intervention procedures, etc.
- These are - by definition - critical components of the production system and hence must be treated as such
- I believe that we need separate instances of these for SC3
 - And not mix with on-going production / other work
- (I also doubt that lxbatch style machines are the right answer for "service machines" but this can perhaps be deferred...)

SC3 CERN Setup (8 June 2005)



Notes

Configuration Options

1. "Setup/Throughput phase". All data served from local disks on castorgridsc nodes (pools of stagegridsc).
 - 1a. It is also possible to stage from tape (through the same pools), but at a max rate of 2Gb/s
2. "Service Phase". Traffic through experiment stagers inside CERN LAN. Max transfer rate is limited at 2Gb/s

NOTES:

- A. We assume we upgrade the SRM software to the new version capable of talking to either old or new CASTOR stagers. Then this design is independent of using the old or new stager i.e. the pool used for a particular VO could be either an old or new one.
- B. A upgrade to 4x1Gb to the LAN is possible
- C. to change from Configuration 1 to Configuration 2 is done via changes to configuration files on castorgridsc only – no other software configuration or hardware reconfiguration is needed.

SC3 vs evaluation

- Some aspects of what has been presented this week more appropriate for 'pilot services' rather than SC3
 - E.g. LHCb Fireman stress testing
- Need to keep SC3 focused on its core goals
 - Whilst enabling realistic evaluations and understanding that in most cases the same 'resources' are involved

Schedule

- Resource scheduling needs to be confirmed with site responsables
- Assuming split between evaluation / SC3 can be agreed, 'natural' schedule offers itself...
- Need to come back to this in a few days...
- Big caveat (again): resources

SC4

- SC4 starts April 2006
- It ends with the deployment of the FULL PRODUCTION SERVICE
- **Deadline for component (production) delivery: end January 2006**
- Adds further complexity over SC3
 - Additional components and services
 - Analysis Use Cases
 - SRM 2.1 features required by LHC expts
 - All Tier2s (and Tier1s...) at full service level
 - Anything that dropped off list for SC3...
 - **xrootd?? proof??**
- "Analysis" clearly needs significant thought and work
 - Step 1: define and agree common terms for primary Use Cases
 - And never say 'analysis' again... (without qualifying it...)
 - e.g. "AOD production", "interactive ROOT analysis", ...