LCG Service Challenge 3
"The Final Countdown"

Jamie Shiers, CERN-IT-GD

GDB, May 2005

Agenda

- Reminder of high-level time-line of SC3
- Deadlines; decision points; detailed time-line
- On-going procedure for monitoring progress:
  - Weekly meetings at CERN (Wed 11:00);
  - Weekly con-calls (Wed 17:30?);
  - Daily (hourly?) informal follow-up…
- ➢ **Need higher level monitoring, particularly July 1st on…**
- This is one of the weak areas that still needs to be addressed in order to deliver a **SERVICE**

The Service is the Challenge

Agenda For April Tier1 Visits

- Goals and Timelines of the LCG Service Challenges
- Review of SC1 and SC2
- Summary of LHC Experiments' Computing Models
- Outline of SC3 and SC4
- ➢ After that it's the **FULL PRODUCTION SERVICE!**
- Plans for involving Tier2 sites in the Service Challenges
- Detailed SC3 planning

LHC Computing Grid

Distrubuted Production Environment for Physics Data Processing

The LCG Service Challenges: Ramping up the LCG Service

Jamie Shiers, CERN-IT-GD-SC

http://cern.ch/jamie/SC.ppt

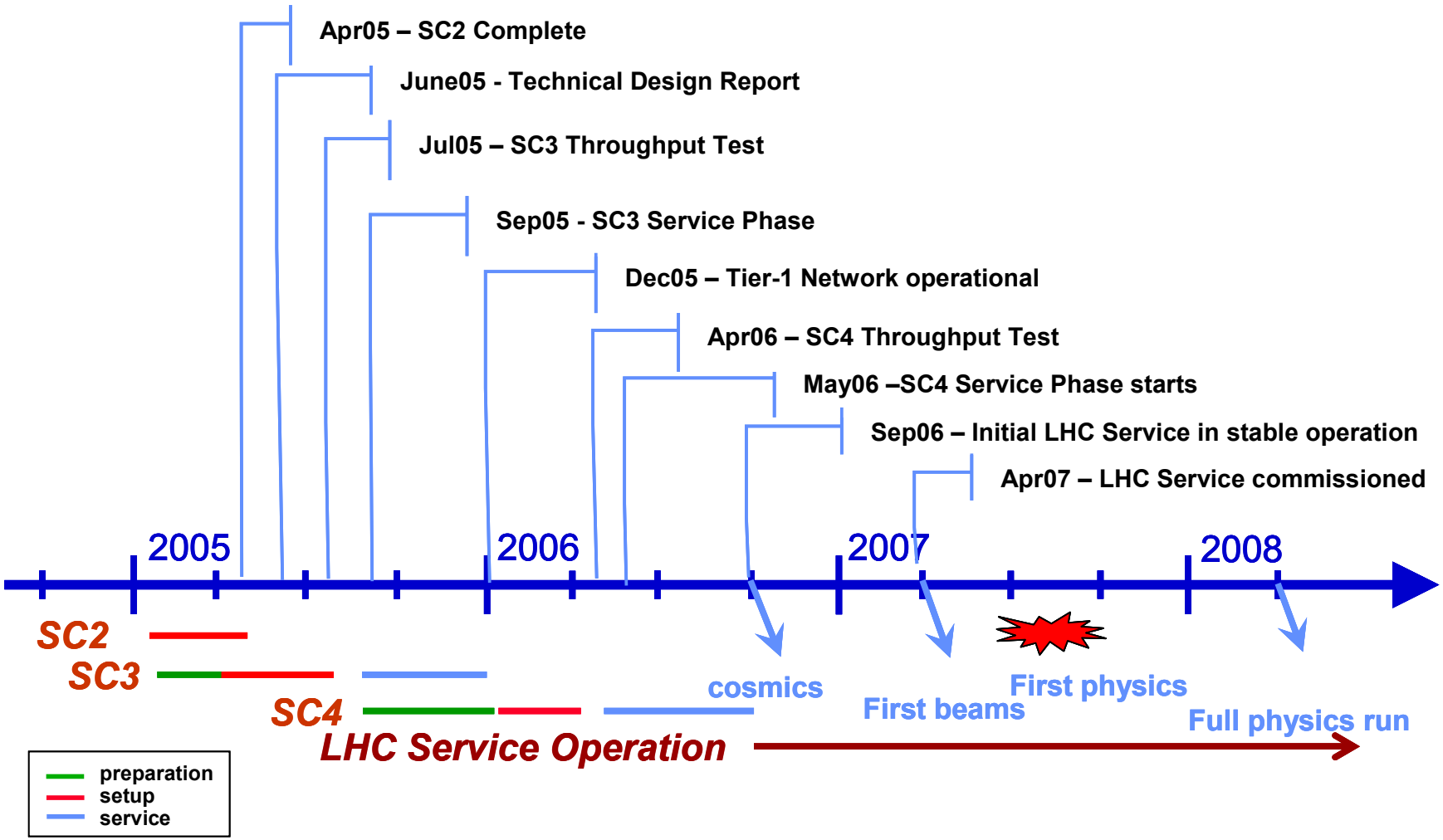
April 2005

pp / AA data rates (equal split)

Centre	*ALICE*	*ATLAS*	*CMS*	*LHCb*	*Rate into T1*	*Rate into T1 (AA)*
ASCC, 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	72.2	11.3
FNAL, USA	0	0	1	0	46.5	16.9
TRIUMF, Canada	0	1	0	0	72.2	11.3
NIKHEF/SARA, Netherlands	1	1	0	1	158.5	80.3
Nordic Centre	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…

LCG Deployment Schedule



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!

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 – 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
  - **My understanding: all currently named T1s will participate in managed transfers, even if not throughput phase**
- 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, HEPiX, BNL 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 – Deadlines and Deliverables

- May 31st 2005: basic components delivered and in place
- June 2005: integration testing
- June 13 – 15: SC3 planning workshop at CERN – experiment issues
- June 30th 2005: integration testing successfully completed
- July 1 – 10: start disk – disk throughput tests
  - Assume a number of false starts / difficulties
- July 11 – 20: disk tests
- July 21 – 27: tape tests
- July 28 – 31: T2 tests

Basic Components For Setup Phase

- **Each T1 to provide 10Gb network link to CERN**
- **Each T1 + T0 to provide SRM 1.1 interface to managed storage**
  - This goes for the named T2s for the T2-T1 transfer tests too
- T0 to provide File Transfer Service; also at named T1s for T2-T1 transfer tests
- Baseline Services Working Group, Storage Management Workshop and SC3 Preparation Discussions have identified one additional data management service **for SC3**, namely the LFC
  - Not all experiments (ALICE) intend to use this
  - Nor will it be deployed for all experiments at each site
- However, as many sites support multiple experiments, and will (presumably) prefer to offer common services, this can be considered a basic component
  - Table by site below

Dedicated connections for SCs

Tier1	Location	NRENs	Status dedicated link
ASCC	Taipei, Taiwan	ASnet, SURFnet	1 Gb via SURFnet, testing
BNL	Upton, NY, USA	ESnet, LHCnet	622 Mbit shared
CNAF	Bologna, Italy	Geant2, GARR	1 Gb now, 10 Gb in Sept
FNAL	Batavia, ILL, USA	ESnet, LHCnet	10 Gb, tested
IN2P3	Lyon, France	Renater	1 Gb now, 10 Gb in Sept
GridKa	Karlsruhe, Germany	Geant2, DFN	10 Gb, tested
SARA	Amsterdam, NL	Geant2, SURFnet	10 Gb, testing
NorduGrid	Scandinavia	Geant2, Nordunet	Not participating yet
PIC	Barcelona, Spain	RedIris, Geant2	Not participating yet
RAL	Didcot, UK	Geant2, Ukerna	2 x 1 Gb via SURFnet soon
Triumf	Vancouver, Canada	Canet, LHCnet	1 Gb via SURFnet, testing

Status of SRM

Ian Bird
IT-GD

LHCC Referees meeting
10th May 2005

Agreement on functionality

- Basic understanding
  - SRM V1.1 is not sufficient
  - Full functional set of SRM V2.1 and later is not required
  - "LCG-required" functionality agreed – Baseline services group and Storage Management workshop
- For SC3
  - V1.1 is sufficient
- For SC4
  - LCG-set is required
  - Workshop put in place a group (developers and experiments) to plan and monitor progress
  - More on this later…

LCG-required SRM functions

- SRM v1.1 insufficient – mainly lack of pinning
- SRM v3 not required – and timescale too late
- Require Volatile, Permanent space; Durable not practical
- Global space reservation: reserve, release, update (mandatory LHCb, useful ATLAS,ALICE). Compactspace NN
- Permissions on directories mandatory
  - Prefer based on roles and not DN (SRM integrated with VOMS desirable but timescale?)
- Directory functions (except mv) should be implemented asap
- Pin/unpin high priority
- srmGetProtocols useful but not mandatory
- Abort, suspend, resume request : all low priority
- Relative paths in SURL important for ATLAS, LHCb, not for ALICE

SRM status at Tier 1 sites

- CERN
  - Castor – in production; update to new Castor for SC3
- FNAL
  - In production; dCache
- BNL
  - dCache – in production, used in SC2. ??? status ???
- CC-IN2P3
  - dCache – under test; planned for SC3
- PIC
  - Castor – in production; update to new Castor as CERN
- CNAF
  - Castor – in production; update to new Castor as CERN
- FZK
  - dCache – testing; not yet in production – planned for SC3
- ASCC
  - Castor – not yet in production – install this month on new hardware.
- RAL
  - dCache used in production (disk only). Tape backend used successfully in SC2. This should be production for SC3.
- NIKHEF/SARA
  - dCache/SRM on test cluster. Expect production in SC3.
- NDGF
  - No information

CASTOR SRM

- New implementation of SRM v1.1. needed for the new stager
- Two parallel implementations ongoing:
  - Re-use of RAL SE
    - RAL efforts working on the porting
  - Re-use of LCG SRM framework written for the LDPM
    - CASTOR dev team working the porting
- Progress checkpoints
  - Mid-May: Demonstrate some minimal functionality (beyond 'ping'), e.g. 'get'
  - End-May: demonstrate fully functional SRM
- Interoperability tests with other clients should be coordinated by LCG and carried out in June

dCache SRM

- Pinning is available
- Provides durable (resilient) space – but needed to allow other VO's access to space
- No space reservation
- Roles – ACL's
  - Integration with VOMS and implementation of permissions planned
- Directory operations
  - Some implemented, others (rm, mv) need clarification
- Getprotocols is implemented
- Abort is implemented, but not suspend/resume
- Relative paths

SRM con-calls

- Call held during SMWS worked well – good attendance from developers and from experiments
- Since then it has been hard to find a suitable slot for further calls
- ➢ How important is this? Unless sufficient effort is provided, likely to decay back to previous state where little or no progress was made…
- Need to agree on SRM 1.1++ specification and give sufficient time for implementation and deployment
- SRM annual meeting (September in TJL) deadline(?) for agreeing on specification
- SC4 starts April 2006 – component delivery end February 2006

Network & SRM Status

- Propose that each T1 verbally states (re-iterates?) status and provides written text to Jeremy Coles for the minutes
- Some sites keep a status page on the web
- Propose that this is adopted by all sites for reference
- Comments? Suggestions?

Additional Components

- Whilst the above presents a 'simple' model, there are clearly additional applications / 'services' that need to be provided
- These include "Agents and Daemons" (next);
- Applications Area-related services;
  - E.g. COOL for ATLAS;
- Applications Area-software;
  - E.g. GEANT4, ROOT POOL, SEAL, …
- Experiment-specific software and services…
  - E.g. ATLAS HVS, book-keeping systems
- I believe that 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…

VOMS

- Plan was to move in steps to a VOMS service running on an Oracle database
  - Oracle port of VOMS core, both client & server, delivered April 29
- Cannot achieve a production setup in time for component delivery for SC3 (end May 2005)
- A new date of June / (July?) compatible with Service Phase
  - Fallback is to run VOMS on MySQL
- This clearly needs more work…
- More on VOMS & SC3 at http://dimou.home.cern.ch/dimou/lcg/voms/SC3-FAQ.html

Agents and Daemons

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
- 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…)

Status of Core Components

- **T1s to give update on network links and SRM implementation**
- CERN SC3 network infrastructure is in place (Vlado's talk)
- Final decision on file transfer h/w setup this week…
- SRM 1.1 for new CASTOR: checkpoint meeting Friday
- CASTOR SC3 production deployment proposal: June 1-3 (as early as poss.)
- FTS: s/w has been released, h/w resources for T0 SC3 services identified and being prepared; some service-issues to be resolved in coming days in parallel with service setup.
  - We are on track … just …
- LFC: s/w (existing code base) has been in test for more than 6 months (CERN) as well as at DESY (for ZEUS) and a few other sites. h/w resources as above
- June 1st: IT 'physics groups' meeting to emphasis goals / timelines / priority of LCG roll-out incl. SCs
- Once again, these issues are being monitored daily…

LCG Service Challenges: Planning for Tier2 Sites

Update for HEPiX meeting

Jamie Shiers
IT-GD, CERN

T2 Executive Summary

- Tier2 issues have been discussed extensively since early this year
- The role of Tier2s, the services they offer – and require – has been clarified
- The data rates for MC data are expected to be rather low (limited by available CPU resources)
- The data rates for analysis data depend heavily on analysis model (and feasibility of producing new analysis datasets IMHO)
- LCG needs to provide:
  - Installation guide / tutorials for DPM, FTS, LFC
- **Tier1s need to assist Tier2s in establishing services**

	Number of T1s	Number of T2s	Total T2 CPU	Total T2 Disk	Average T2 CPU	Average T2 Disk	Network In	Network Out
			KSI2K	TB	KSI2K	TB	Gb/s	Gb/s
ALICE	6	21	13700	2600	652	124	0.010	0.600
ATLAS	10	30	16200	6900	540	230	0.140	0.034
CMS	6 to 10	25	20725	5450	829	218	1.000	0.100
LHCb	6	14	7600	23	543	2	0.008	0.008

Tier2 and Base S/W Components

1) Disk Pool Manager (of some flavour…)
   - e.g. dCache, DPM, …
2) gLite FTS client (and T1 services)
3) Possibly also local catalog, e.g. LFC, FiReMan, …
4) **Experiment-specific s/w and services ('agents')**

1 – 3 will be bundled with LCG release.
Experiment-specific s/w will not…

[N.B. we are talking **interfaces** and not implementation]

Tier2s and SC3

- Initial goal is for a small number of Tier2-Tier1 partnerships to setup agreed services and gain experience
- This will be input to a wider deployment model
- Need to test transfers in both directions:
  - MC upload
  - Analysis data download
- Focus is on **service** rather than "throughput tests"
- As initial goal, would propose running transfers over at least several days
  - e.g. using 1GB files, show sustained rates of ~3 files / hour T2->T1
- More concrete goals for the Service Phase will be defined together with experiments in the coming weeks
  - Definitely no later than June 13-15 workshop

Initial Tier-2 sites

- For SC3 we aim for (updated from input at May 17 GDB):

Site	Tier1	Experiment
Legnaro, Italy	CNAF, Italy	CMS
Milan, Italy	CNAF, Italy	ATLAS
Turin, Italy	CNAF, Italy	Alice
DESY, Germany	FZK, Germany	ATLAS, CMS
Lancaster, UK	RAL, UK	ATLAS
Imperial, UK	RAL, UK	CMS
Edinburgh, UK	RAL, UK	LHCb
US Tier2s	BNL / FNAL	ATLAS / CMS

- Training in UK last Friday (13th) and in Italy end May (26-27)
- Other interested parties: Prague, Warsaw, Moscow, ..
- **Addressing larger scale problem via national / regional bodies**
  - GridPP, INFN, HEPiX, US-ATLAS, US-CMS, Triumf (Canada)
- Cannot handle more for July tests, but please let us know if you are interested! (T1+T2 partnerships)

T2s – Concrete Target

- We need a **small number** of **well identified** T2/T1 partners for SC3 as listed above
- Initial target of end-May is not realistic, but not strictly necessary either…
- Need prototype service in at least two countries by end-June
- Do not plan to strongly couple T2-T1 transfers to T0-T1 throughput goals of SC3 setup phase
- **Nevertheless, target one week of reliable transfers T2->T1 involving at least two T1 sites each with at least two T2s by end July 2005**

LFC and FTS by Tier1

Centre	*ALICE*	*ATLAS*	*CMS*	*LHCb*	*LFC*	*FTS*
ASCC, Taipei	0	1	1	0	Y	N
CNAF, Italy	1	1	1	1	Y	Y
PIC, Spain	0	1	1	1	?	N
IN2P3, Lyon	1	1	1	1	Y	N
GridKA, Germany	1	1	1	1	Y	Y
RAL, UK	1	1	1	1	Y	Y
BNL, USA	0	1	0	0	?	Y
FNAL, USA	0	0	1	0	Globus RLS?	Phedex
TRIUMF, Canada	0	1	0	0	?	N
NIKHEF/SARA, Netherlands	1	1	0	1	Y	N
Nordic Centre	1	1	0	0	?	n
Totals	6	10	7	6		

SC3 – Setup Phase Recap

- Discussions over the past months have led to clarification on the services required for the 'infrastructure'
- All of these need to be in place for the Setup Phase – July
- The basic requirements for most sites – 10Gb network connectivity to CERN + production-quality SRM 1.1 service – are not new
- LFC and FTS for those sites concerned require Oracle or MySQL backend service
- Deployment model at CERN:
  - Dedicated disk server for the Oracle database service for each
  - Farm node for middle tier per VO plus one spare
  - How this deployment model works can be considered part of the 'Service' challenge…
- Don't be disappointed – there is still an awful lot to do just for these services…
- And this is the basic infrastructure, on top of which we need to consider the experiments' needs…

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 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.
- 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.

➢ **Dates: June 13 – 15: B160 1-009 for first 2 days then 513 1-024**

  ➢ **4 x ½ days on Experiment-specific issues (ALICE, ATLAS, CMS, LHCb)**

http://agenda.cern.ch/fullAgenda.php?ida=a051784

SC3 Experiment Status

- Experiment goals for the Service Phase are becoming clearer
- Need to finalise plan in coming weeks, including overall schedule and confirmation of resources at all sites
  - Assumed to come from existing pledges
- Need to be very clear on responsibilities
- Setup for service phase overlaps the July SC3 setup phase
  - And has in some cases already started…
- Once again, the goal of the SCs is to deliver THE SERVICE

Summary

- We are just about on target with delivering the service components required for SC3
  - Possible slippage is in days.. but we will try to get there on time…
- Further clarification between now and June workshop on precise goals of experiments, additional experiment software and services, resource allocation and schedule
- Expect detailed presentations at June workshop, including experience with current services
  - e.g. ATLAS experience from ROME production
- We need to catalog a list of issues to address, rather than just 'a feeling of frustration…'

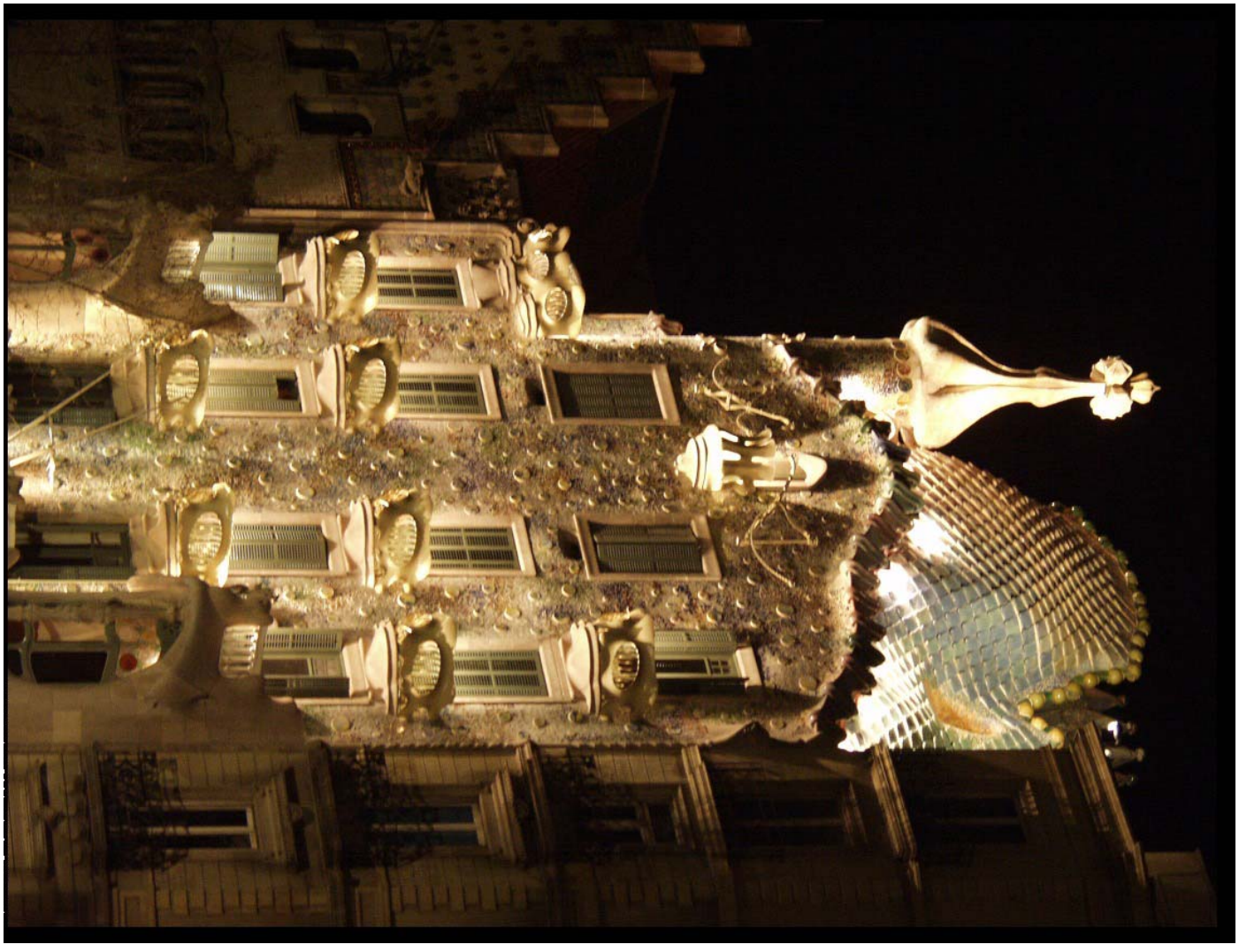
Conclusions

- 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

For SuperComputing
2005

Deploying the LHC Computing Environment

Deploying the LHC Computing Environment



Deploying the LHC Computing Environment

Deploying the LHC Computing Environment

Deploying the LHC Computing Environment

Deploying the LHC Computing Environment

Deploying the LHC Computing Environment

You get the picture…