LHC: First Contact

Global Grid Operations - What it means for the LHC Grid and the HEP Community

Jamie.Shiers@cern.ch

Grid Support Group, IT Department, CERN

Introduction

- On September 12th 2008, the **eyes of the world** were on CERN as probably never before
- After some 15 years of construction, the Large Hadron Collider (LHC) was finally ready for first **circulating beams** and for the **associated detectors** to record their first ‘real’ events
- This talk is not about the **heroic achievements** that made this possible, but rather about the computing and storage infrastructure that was put in place to manage and process the **vast quantities of data** that would soon be produced
- This is the story of the Worldwide LHC Computing Grid (WLCG) – **were we, or are we, ready?**

Ready for What?

- To unc
  need t
  require
  judge
- This is
  and cc
- We are
  of the
  autom
  interve
  anoma

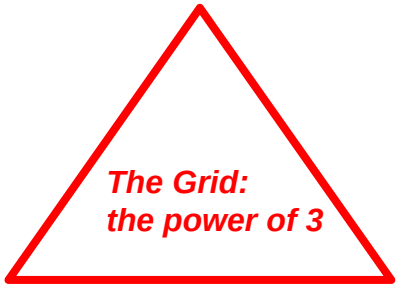
The Requirements

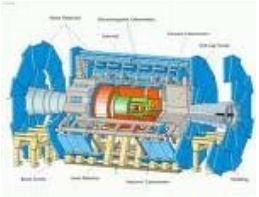
- **Resource requirements**, e.g. ramp-up in Tier*N* CPU, disk, tape and network
  - Look at the Computing TDRs;
  - Look at the resources pledged by the sites (MoU etc.);
  - Look at the plans submitted by the sites regarding acquisition, installation and commissioning;
  - ➢ **Measure what is currently (and historically) available; signal anomalies.**
- **Functional requirements**, in terms of services and service levels, including operations, problem resolution and support
  - Implicit / explicit requirements in Computing Models;
  - Agreements from Baseline Services Working Group and Task Forces;
  - Service Level definitions in MoU;
  - ➢ **Measure what is currently (and historically) delivered; signal anomalies.**
- **Data transfer rates** – the Tier*X* ←→ Tier*Y* matrix
  - Understand Use Cases;
  - ➢ **Measure …**

And test extensively, both 'dteam' and other VOs

WLCG Viewpoint

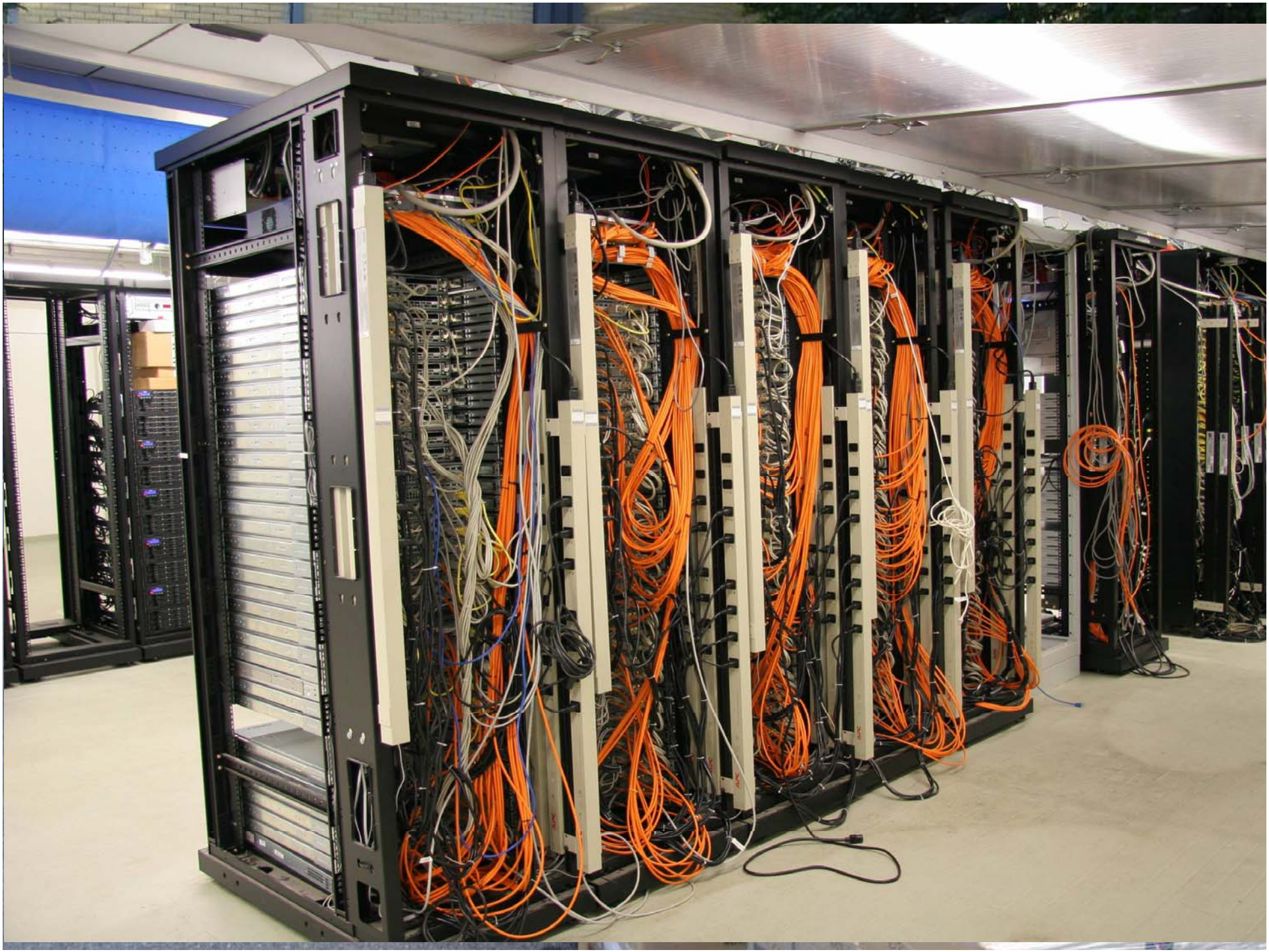
- From the point of view of the WLCG project – which includes the experiment spokesmen and computing coordinators (as well the sites) in the project management structure – the **LHC machine**, the **experiments** and the **WLCG project** (itself a collaboration) work together to achieve a common goal:
- To allow the science to be extracted swiftly & efficiently





Experiment Viewpoint

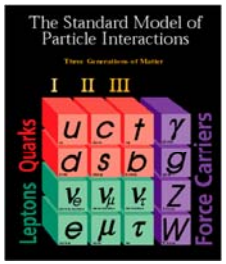
- From the point of view of the experiments, the LHC **machine** and the **experiments** work together to achieve a common goal
- Oh, yes, and
  - [The WLCG service] ***"should not limit ability of physicist to exploit performance of detectors nor LHC's physics potential"***
  - ***"…whilst being stable, reliable and easy to use"***
- Whilst this is clearly simplified and/or exaggerated, this makes an important point that we must not forget: we (WLCG) **must provide** a **service** – that is the only reason we exist!



LHC & WLCG

What and Why?

Physics Motivation



- We currently have a good and very accurate model that has been extensively validated by experiment
- ☹ But it is – at best – incomplete (or possibly **wrong**), leaving some important open questions:
  - Mass;
  - Matter vs anti-matter;
  - Dark Matter;
  - Dark Energy

- The LHC has been built as a ***Discovery Machine*** to hopefully answer these questions – and perhaps raise some more!

CERN

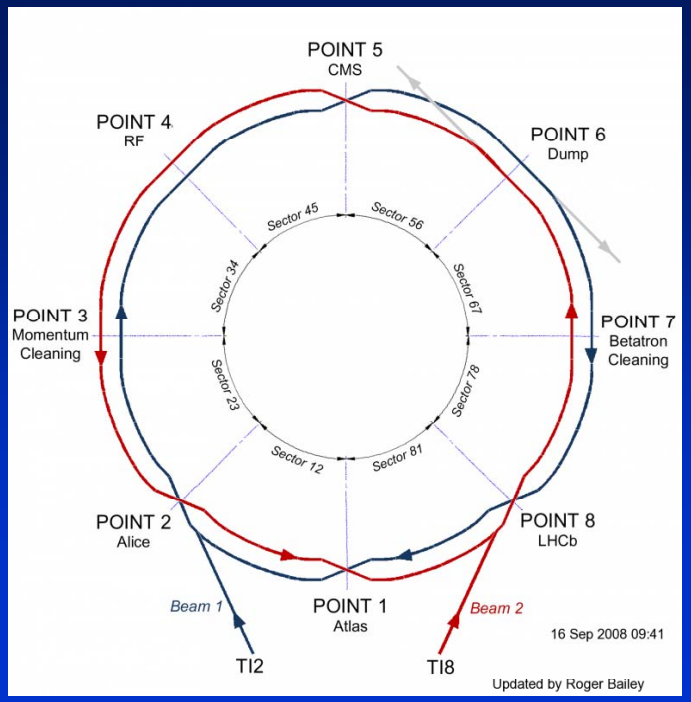
- To some people, CERN is simply a geographic location
  - Latitude: 46°13'59" N
  - Longitude: 6°3'20" E
- For me this description is more than incomplete – it is simply ***wrong!***
- I do not believe that you can really understand what CERN is (about) unless you also consider:
  - The scientific research programme;
  - The close collaboration with a large number of institutes worldwide – that CERN serves and for whom it exists;
  - Its outreach programme and technology transfer; ...
- We need to recognise (explain, evangelise) the role of science in society – it is not just for science!

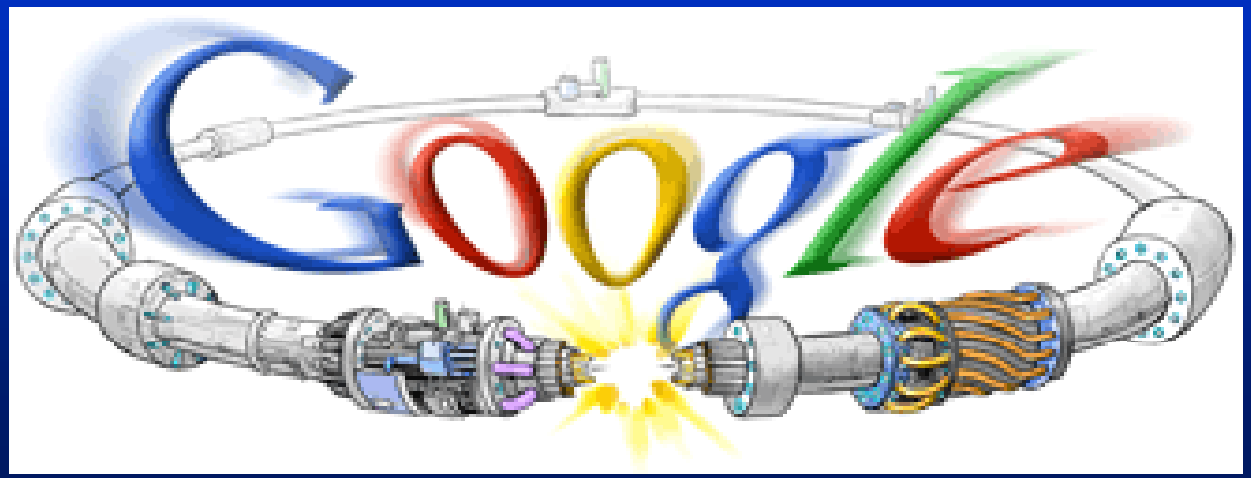
The LHC Machine

- The LHC machine – actually ***two*** concentric accelerators – is an excellent example of large scale collaboration and a tribute to human ingenuity
- It is made possible by a variety of technologies and phenomena – including **superconductivity** and **superfluidity** – whose properties are inherent to the design of the machine
- First proposed around the late 1970s, it has been some 15 years in construction
- It builds on experience with previous colliders at CERN – notably the proton/anti-proton collider some 25 years ago – but also LEP, whose tunnel it borrows…
- I personally have been working on LHC computing since 1992 – roughly 2/3 of my career at CERN!

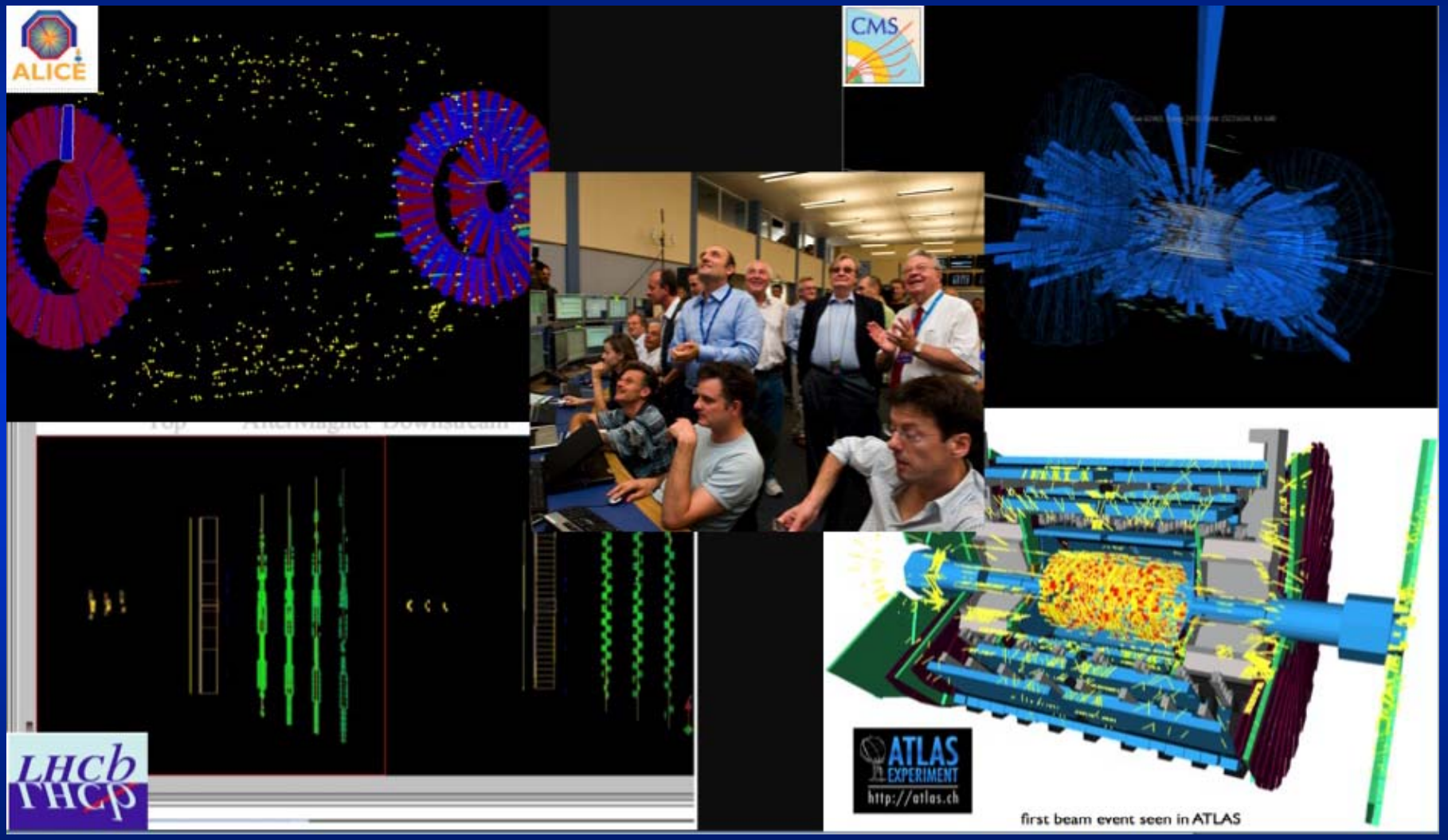
September 10th

- Achieved
  - Beam 1 injected IP2
  - Threaded around the machine in 1h
  - Trajectory steering gave 2 or 3 turns
  - Beam 2 injected IP8
  - Threaded around the machine in 1h30
  - Trajectory steering gave 2 or 3 turns
  - Q and Q' trims gave a few hundred turns





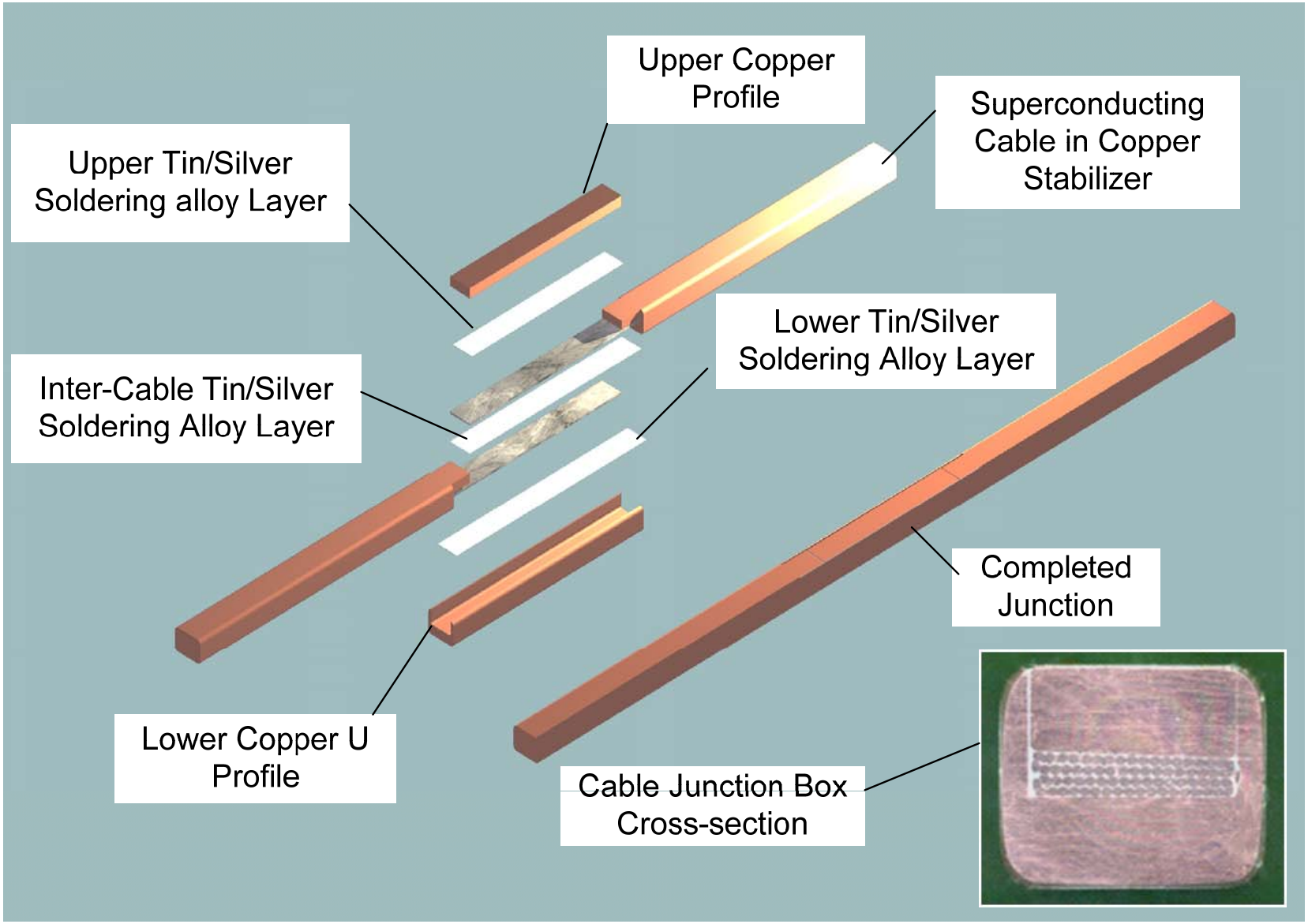
First Beam Events



Electrical joints on 12 kA bus bars

History of interconnections in sector 3-4 (Oct 2006-July 2007) shows no particular cause of defect, but worst working conditions of all machine:
- low temperature and humidity in tunnel
- low productivity of industrial staff (Jan 2007) following contract policy of company

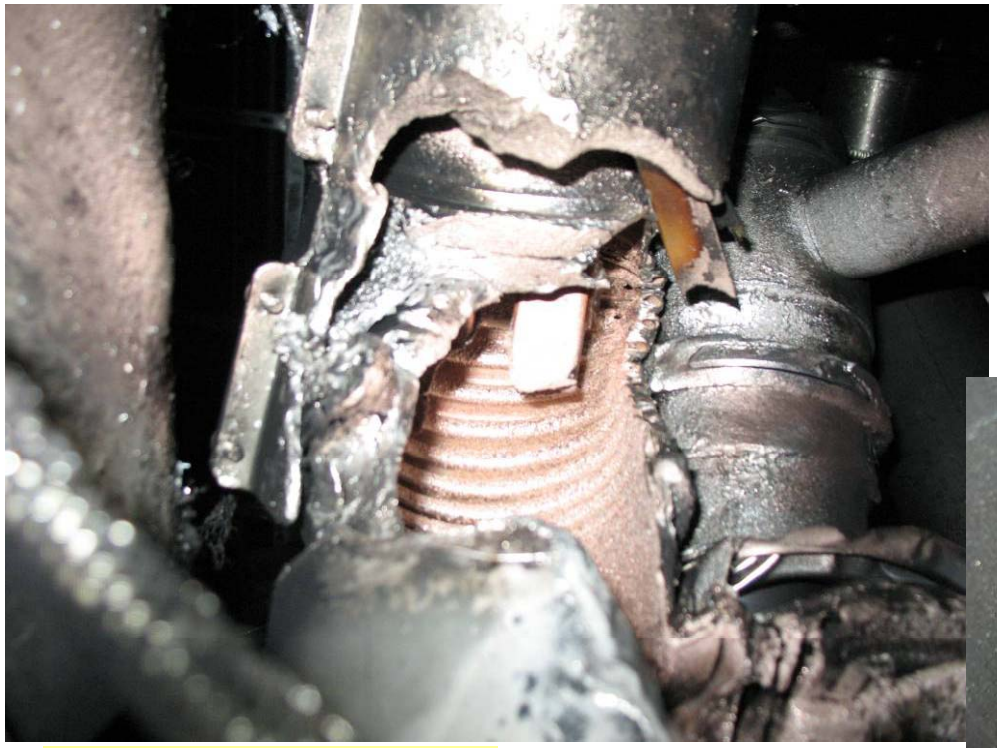
Electrical joint in 12 kA bus bar



Collateral damage: magnet displacements

QQBI.27R3

Collateral damage: secondary arcs



QQBI.27R3 M3 line

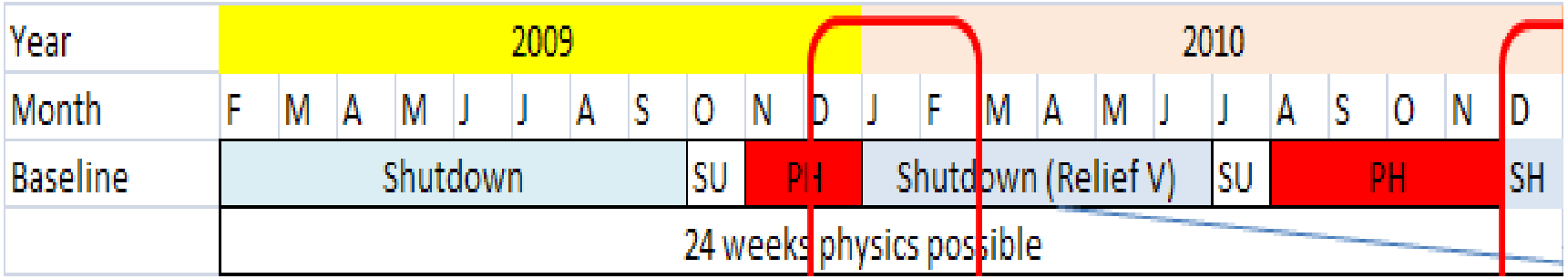
QBBI.B31R3 M3 line

Collateral damage: ground supports

Physics Running Time

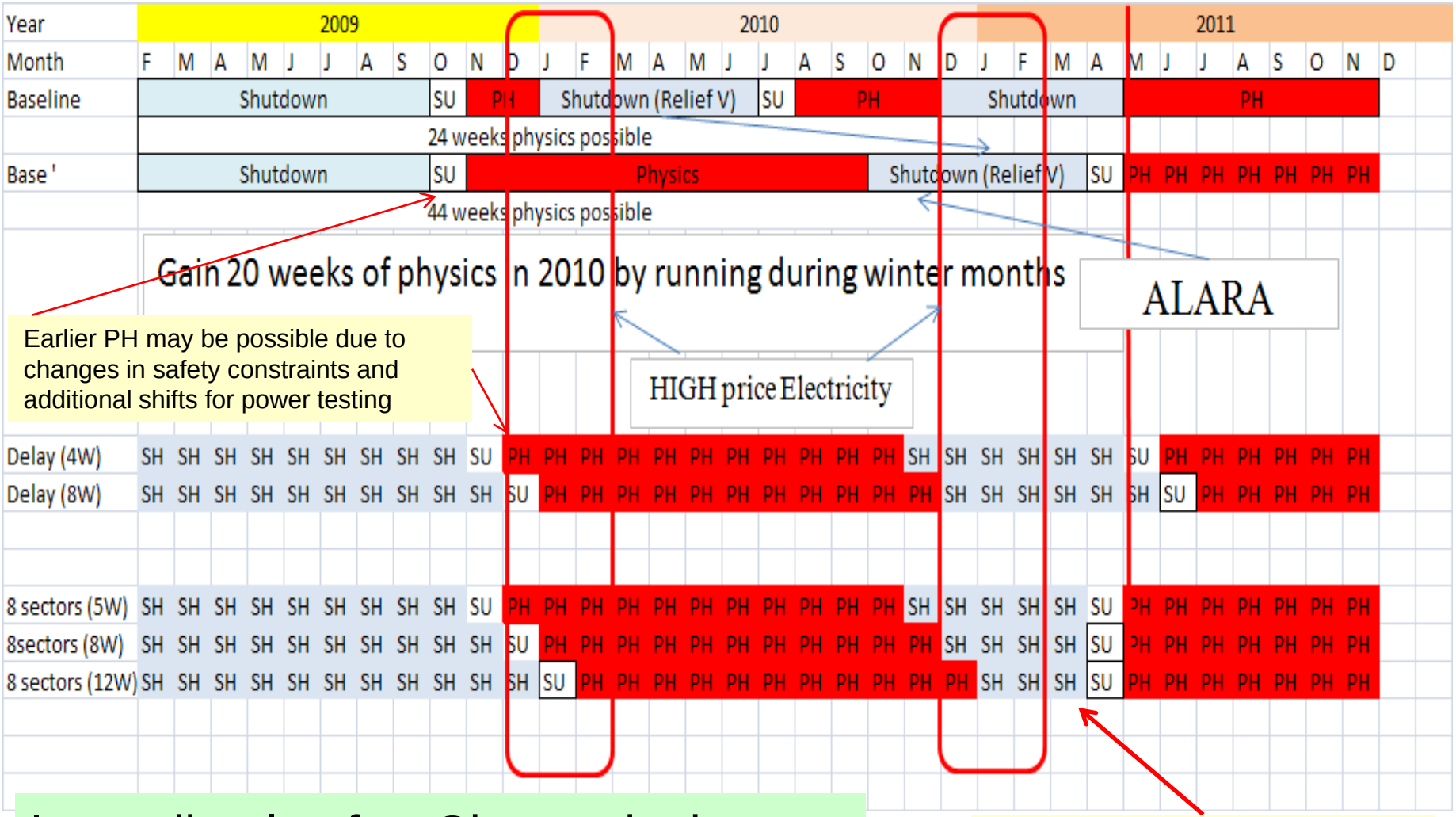
With Strictly No running of the machines in the winter months

- Present baseline schedule
  - schedule allows very limited physics in 2009/2010 (24 weeks)
  - Any slip of >1 month in the S34 repair will delay first LHC physics till August/September 2010!!
  - Repair schedule has no contingency



- **Must** have the possibility of running during winter months

Summary on Schedule



Immediately after Chamonix the management decided on scenario A

Here it is assumed that these shutdowns will be long enough in case of problems seen during the preceding PH running

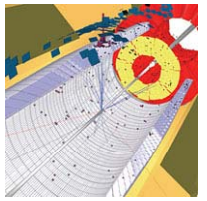
What does that mean in English?

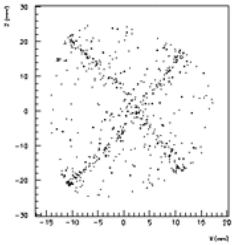
- Effectively the 2009 data taking run has “slipped up against” the 2010 run
- This has significant implications for the WLCG Service / Operations
- **In a nutshell, what we have in place now – with a small number of already scheduled enhancements – will have to take us through the combined 2009 + 2010 data taking runs of the LHC**
- **This includes the transition period from EGEE III to EGI which must be non-disruptive to data taking, processing and analysis!**

WLCG Overview

A Quick Summary / Reminder of the
WLCG Computing Model

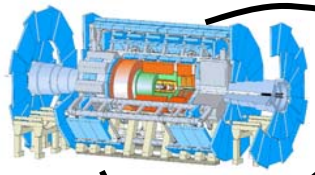
WLCG





- For the purpose of this talk I will use a similarly loose definition of the **W**orldwide **L**HC **C**omputing **G**rid
- This formally consists of a collaboration between the 4 main LHC "experiments" and a set of institutes that provide computing resources and services to these communities
  - Defined in a "Memorandum of Understanding" signed by all parties;
  - Includes services and service levels offered, resource pledges for coming years
- Also a set of management infrastructures and operational boards to plan, deploy, operate and evolve the services (**close collaboration with EGEE etc**.)
- IMHO, essential to also include "friends" in this informal definition – other (VOs, sites, services) with various couplings to the "core business"
  - e.g. GEANT4 – main simulation tool in HEP and (way) beyond…
  - SIXT – simulation tool for LHC accelerator (sixtrack)
  - Lattice QCD simulation – a number of serious scientific publications (see slide notes)
- Many physicists / institutes also involved in other experiments and / or disciplines
- The "boundaries" – if they even exist – are tenuous & flexible…
  - e.g. CERN Grid Support group works with a wide range of disciplines;
  - Grid Data & Storage Management products (dCache, DPM, …) – even if developed at HEP labs for HEP users are also used by – and extended for – many disciplines…

Data Handling and Computation for Physics Analysis



detector

event filter (selection & reconstruction)

reconstruction

raw data

event summary data

processed data

event reprocessing

batch physics analysis

analysis

analysis objects (extracted by physics topic)

event simulation

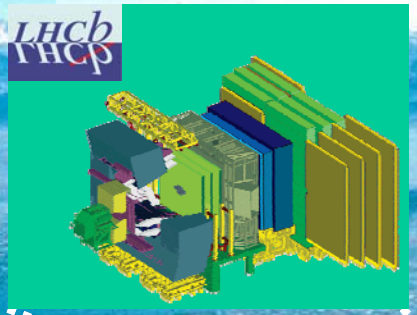
simulation

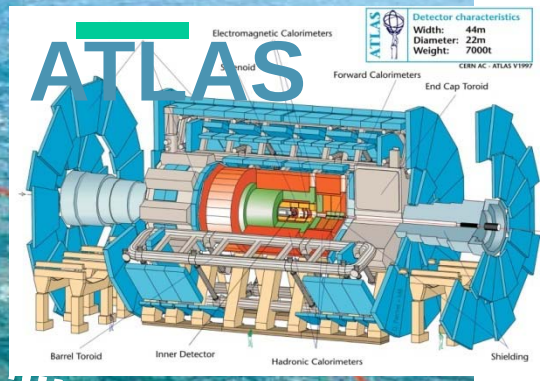
interactive physics analysis

LHC: One Ring to Bind them…

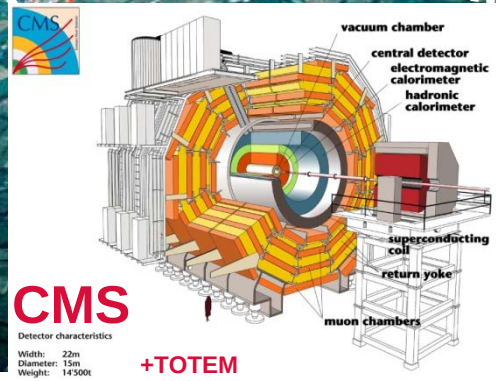
LHC : 27 km long
100m underground

pp, B-Physics,
CP Violation

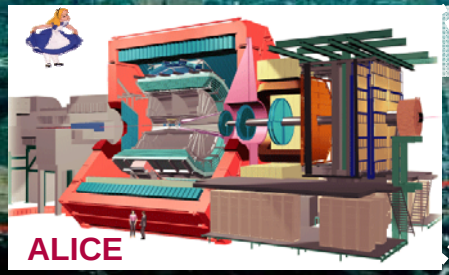




General Purpose,
pp, heavy ions

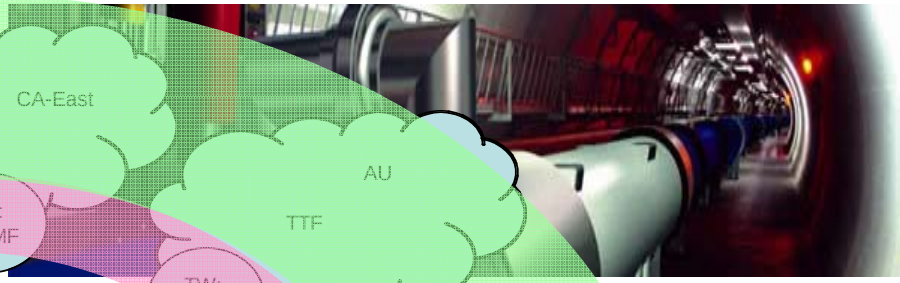


+TOTEM



Heavy ions, pp

ATLAS Cloud Model



Tier-0 – the accelerator centre

- Data acquisition & initial processi
- Long-term dat
- ➢ **Data Distrib** entres

Tier-1 – "o isition
proces lity

- Managed Mass Stora
- → grid-enab
- **All re-**
- Data-heavy
- Nat regional support

Tier-2 – centres

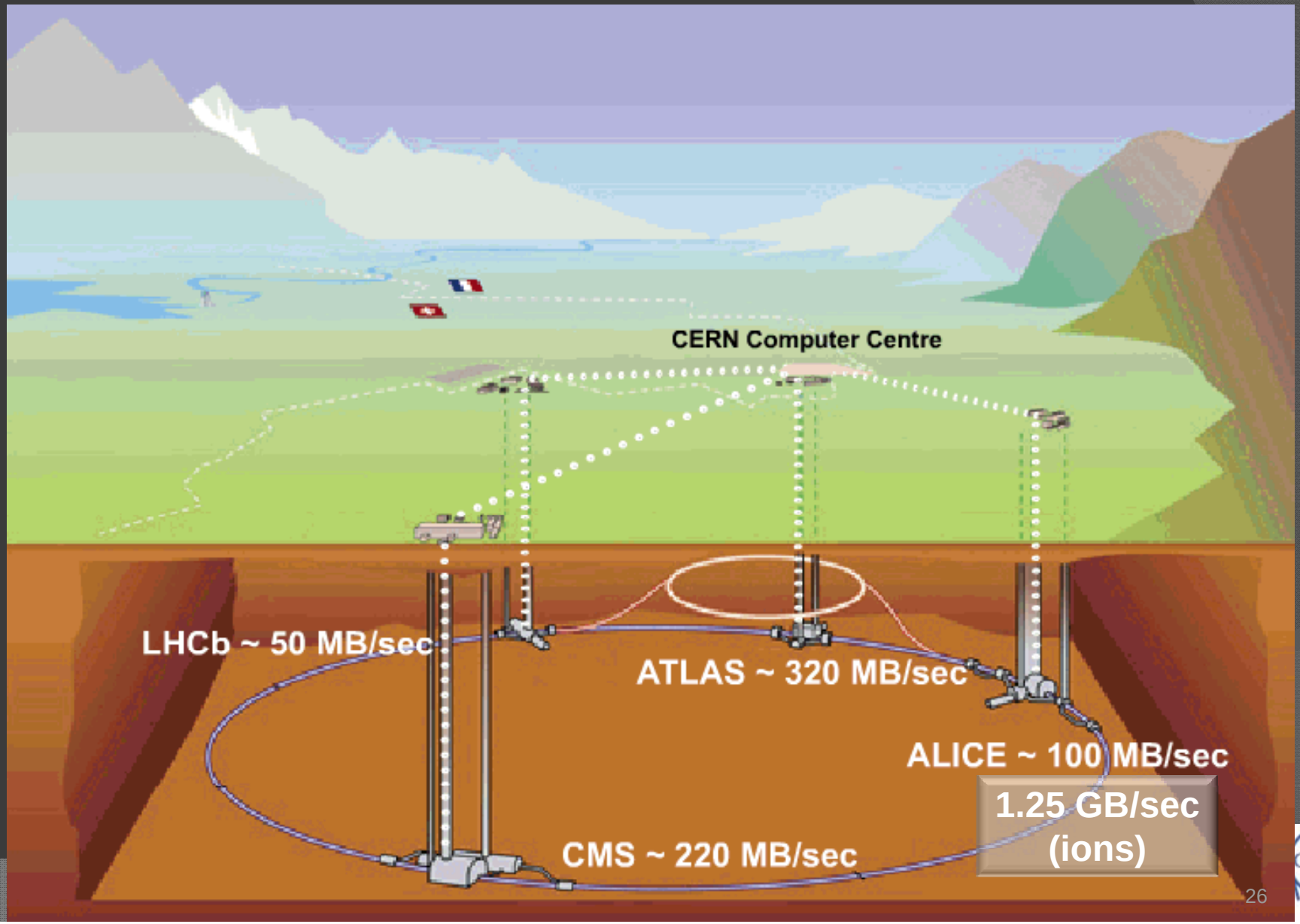
- Simulation
- End-user analysis – batch and interactive
- ➢ **Services, including Data Archive and Delivery, from Tier-1s**

Canada – Triu
France – IN2P3 (L
Germany –Karlsruhe
Italy – CNAF (Bologna)
Netherlands – NIKHEF/SA
Nordic countries – distribut

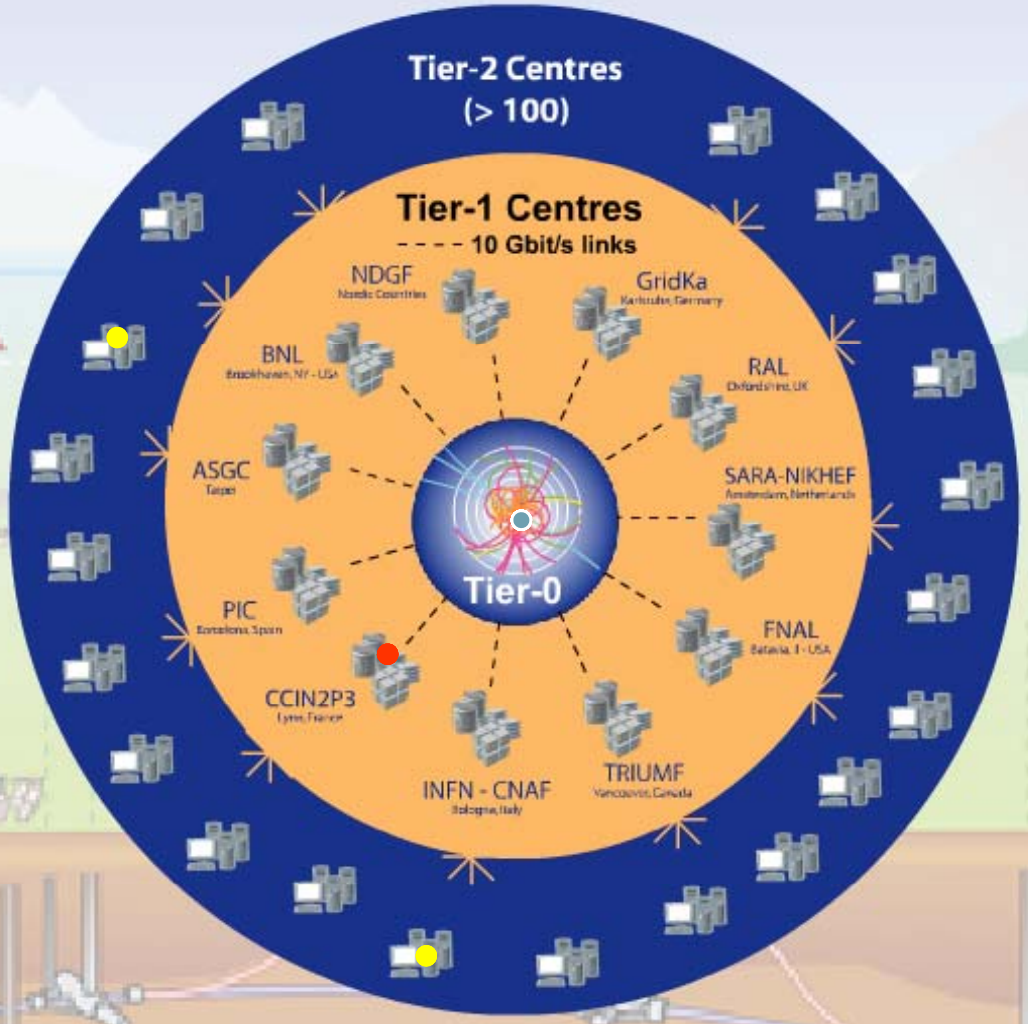
PIC (Barcelona)
Taiwan – Academia Sinica (Taipei)
(NY)

les.robertson@cern.ch

Tier 0 at CERN: Acquisition, First pass reconstruction, Storage & Distribution



Tier 0 – Tier 1 – Tier 2



Tier-0 (CERN):
- Data recording
- Initial data reconstruction
- Data distribution

Tier-1 (11 centres):
- Permanent storage
- Re-processing
- Analysis

Tier-2 (>200 centres):
- Simulation
- End-user analysis


BNL
ASGC/Taipei
CCIN2P3/Lyon
NIKHEF/
SARA
FNAL
TRIUMF/BC
RAL
PIC
NDGF
CNAF
TIER2s
CERN
FZK

WLCG Service

How has the service stood up to real production usage?

Service Status – The Story So Far…

- **One year ago** we had still not demonstrated that we could sustain all production workflows from all 4 LHC experiments simultaneously
- This prompted the "legendary question":
- 💣 **What happens when the LHC is operating?**
- This led to the "Common Computing Readiness Challenge(s)" that were exercised during the first half of 2008
  - **Agreed metrics, targets & reporting mechanisms…**
- The conclusion from the challenge (February and May) was that:
  - ✓ **We met the goals** (even if overlap from all experiments less than optimal) ***but***
  - **Real data taking will be different!**
- The real – and very frightening – prospect had CCRC'08 been less successful would have been de-scoping!
- ☺ **This option was ruled out already by the February run**

- **IMHO – the success of CCRC'08 is a landmark in the fable of grid computing**
  - **and obviously to the many people who contributed to this**

How We Measured Our Success

- Agreed up-front on specific targets and metrics – these were 3-fold and helped integrate different aspects of the ser

Experience shows this is not enough! Computing models of experiments must also be considered…

nts for each
tes to ensure
re in place;

2. ed by the experiments – al p their production, with an analysis of the impact of service egradation or interruption **(WLCG Design, Implementation & Deployment standards)**
3. WLCG "**Memorandum of Understanding**" (MoU) targets – services to be provided by sites, target availability, time to intervene / resolve problems ...

- **Clearly some rationalization of these would be useful – significant but not complete overlap**

Problem Response Time and Availability targets Tier-1 Centres

Service	Maximum delay in responding to operational problems (hours)			Availability
	Service interruption	Degradation of the service		
		> 50%	> 20%	
Acceptance of data from the Tier-0 Centre during accelerator operation	12	12	24	99%
Other essential services – prime service hours	2	2	4	98%
Other essential services – outside prime service hours	24	48	48	97%

les.robertson@cern

Critical Service Follow-up

- Targets (not commitments) proposed for Tier0 services
  - Similar targets requested for Tier1s/Tier2s
  - Experience from first week of CCRC'08 suggests targets for **problem resolution** should not be too high (if ~achievable)
    - The MoU lists targets for responding to problems (12 hours for T1s)
  - ¿ **Tier1s: 95% of problems resolved <1 working day ?**
  - ¿ **Tier2s: 90% of problems resolved < 1 working day ?**

➢ **Post-mortem triggered when targets not met!**

Time Interval	Issue (Tier0 Services)	Target
End 2008	Consistent use of all WLCG Service Standards	100%
30'	Operator response to alarm / call to x5011 **/ alarm e-mail**	99%
1 hour	Operator response to alarm / call to x5011 **/ alarm e-mail**	100%
4 hours	Expert intervention in response to above	95%
8 hours	Problem resolved	90%
24 hours	Problem resolved	99%

CCRC'08 Post-Mortem High-lights

- **The bottom line** – we believe that the experience in 2008 so far confirms that we have a working service model and that we are ready to ***face*** the challenges of data taking from pp collisions in the LHC
  - ✓ Most aspects of the service work well most of the time
  - ✓ We have a proven track record in resolving even the most daunting of problem in an acceptably short time
- **What is really interesting is what happens when things go wrong – and how we can improve on this in the future**

Strengths

- CCRC'08 and accompanying experiment "dress rehearsals" have in most cases demonstrated that the services / sites / experiments are ready for higher loads than are expected from 2008 pp data taking

☺ **The middleware process is working well!**

☺ **The database services are working well!**

➢ **We have a well tested service model and have demonstrated steady improvement over a long time**

Weaknesses

- Some of the services – including but not limited to storage / data management – are still not sufficiently robust. (Process? Deployment?) We have (so far) failed to define and regularly update a clear table of versions + release + patch level. This is nevertheless the target, with a weekly update at the joint EGEE-OSG-WLCG operations meeting
- Communication is still an issue / concern. This requires work / attention from everybody – it is not a one-way flow.
- Not all activities (e.g. reprocessing, chaotic end-user analysis) were fully demonstrated even in May, nor was there sufficient overlap between all experiments (and all activities). Work continues (July and beyond)...
- There were a large number (IHMO too many) Tier0 service upgrades in June – not always well scheduled and / or motivated. We must balance stability with needed fixes

Opportunities

- There is no technical reason why we cannot solve the non-technical problems in the storage area (i.e. define recommended versions that have been released and tested – not “dreams”!)
- Communication – certainly no silver bullet expected. Need solutions that scale to the number of sites / players involved, that can adapt to constraints of time zones and affordable technology (audio & video conferencing, for example...)
- Improvements in monitoring and automation to reduce human expert involvement to a sustainable level (medium – long-term?)
- We still need to maintain a high(-er) level view of the overall WLCG service – a purely component view is not compatible with a highly complex service with many inter-dependencies

Threats

- The biggest threat that I see is to fall back from reliable service mode into “fire-fighting” at the first sign of (major?) problems.
- This in the past has been accompanied by memos to the highest level, triggering time and effort consuming response / post-mortems, but is not sustainable and is much less efficient than the proven service mode.
- This requires close collaboration and concerted effort – as has been the case through many years of data and service challenges, and as we have seen at previous machines.
- Daily operations meeting as a focus / dispatching point plus constant interactions with experiments / sites.

S.W.O.T. Summary

- CCRC'08 has proven to be a very valuable exercise for demonstrating readiness for 2008 data taking, including identifying (and fixing) holes in the service
- With justification, we can be confident of our readiness – from steady operation through to unexpected "crises" (which we will quickly defuse & resolve...)
- Communication & coordination have been key
- It has been – at least at times – very hard work, but also extremely rewarding!
- May collisions commence...

WLCG Key Performance Indicators

- Since the beginning of last year we have held week-daily **conference calls** open to all experiments and sites to follow-up on short-term operations issues
- These have been well attended by the **experiments**, with somewhat more patchy attendance from sites but minutes are widely and rapidly read by members of the WLCG Management Board and beyond
- A weekly summary is given to the Management Board where we have tried to evolve towards a small set of **Key Performance Indicators**
- These currently include a summary of the GGUS tickets opened in the previous week by the LHC VOs, as well as more important service incidents requiring follow-up: **Service Incident Reports (aka post-mortems)**

GGUS Summary

VO concerned	USER	TEAM	ALARM	TOTAL
ALICE	3	0	0	3
ATLAS	16	16	0	32
CMS	13	0	0	13
LHCb	9	2	0	11
Totals	41	18	0	59

☺ No alarm tickets – this may also reflect activity

- Increasing use of TEAM TICKETS

☝ Regular test of ALARM TICKETS coming soon!

- See https://twiki.cern.ch/twiki/bin/view/LCG/WLCGDailyMeetingsWeek090223#Tuesday under AOB

Intervention Summary (fake)

Site	# scheduled	#overran	#unscheduled	Hours sched.	Hours unsched.
Bilbo	5	0	**1**	10	4
Frodo	1	1	0	2	**22**
Drogo	**27**	0	0	**165**	0

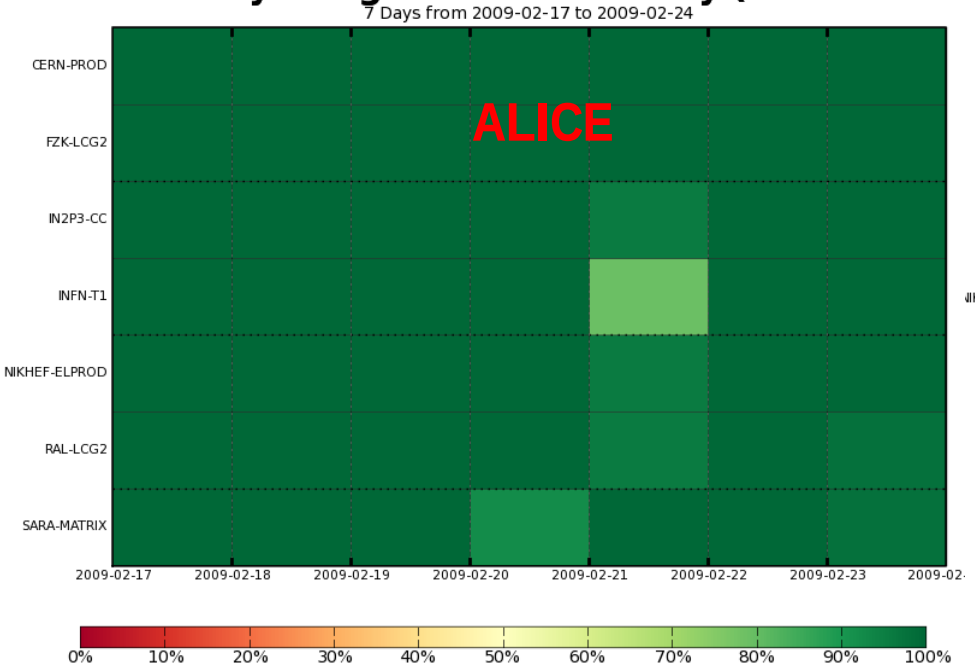
- As with GGUS summary we will drill-down in case of exceptions (examples high-lighted above)
- Q: what are reasonable thresholds?
- **Proposal:** look briefly at **ALL** unscheduled interventions, **ALL** overruns and “high” (TBD) # of scheduled

(Some) Unscheduled Interventions

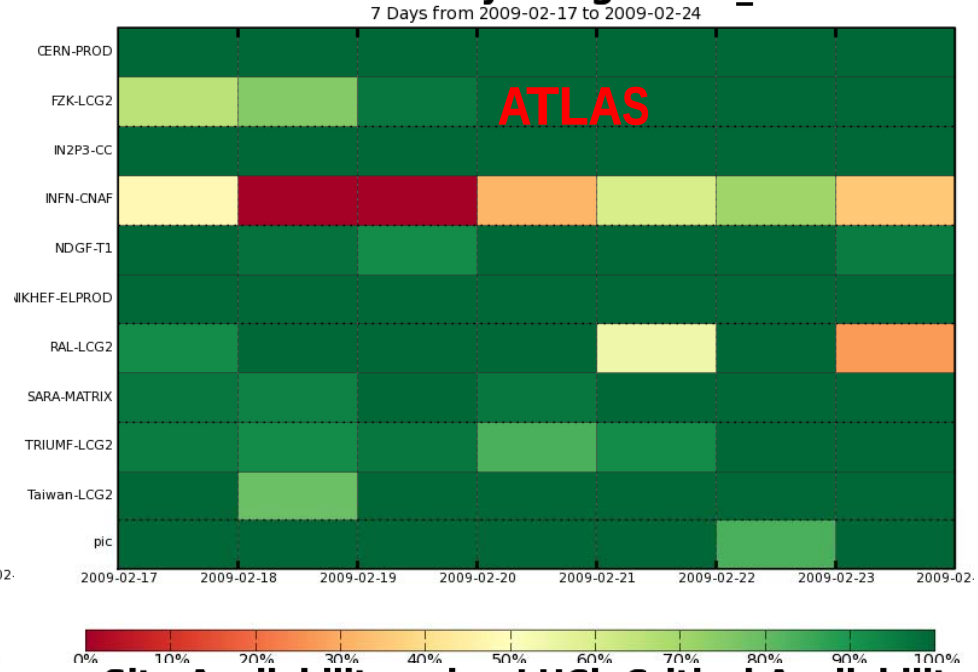
Site	Reason
NL-T1 (SARA-MATRIX)	A DDN storage device partially crashed and needs a cold reboot and some additional actions. We are uncertain how long it will take. The SARA CE's may be affected. Period announced 23-02-2009 09:30 – 11:15 Intervention terminated 23-02-2009 12:20
NDGF	Some dCache pools offline from time to time due to bad hardware causing spontaneous reboots. Period announced 20-02-2009 15:22 – 23-02-2009 15:22 Terminated 23-02-2009 16:25

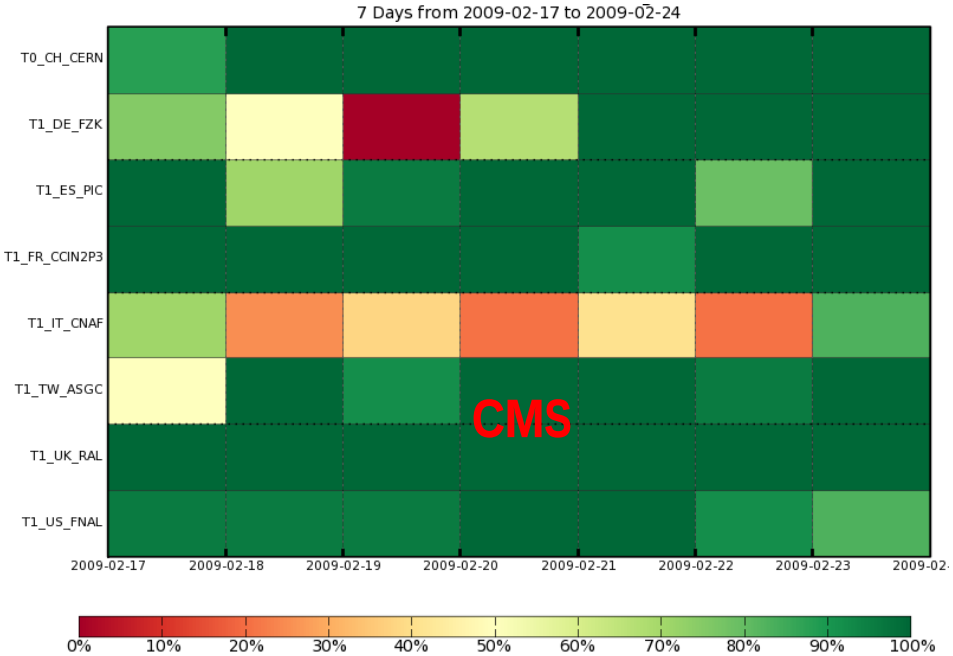
- We need to automatically harvest this information and improve follow-up reporting
- **A convenient place to provide such a report is at the daily WLCG operations call!**

ite Availability using WLCG Availability (FCR critical

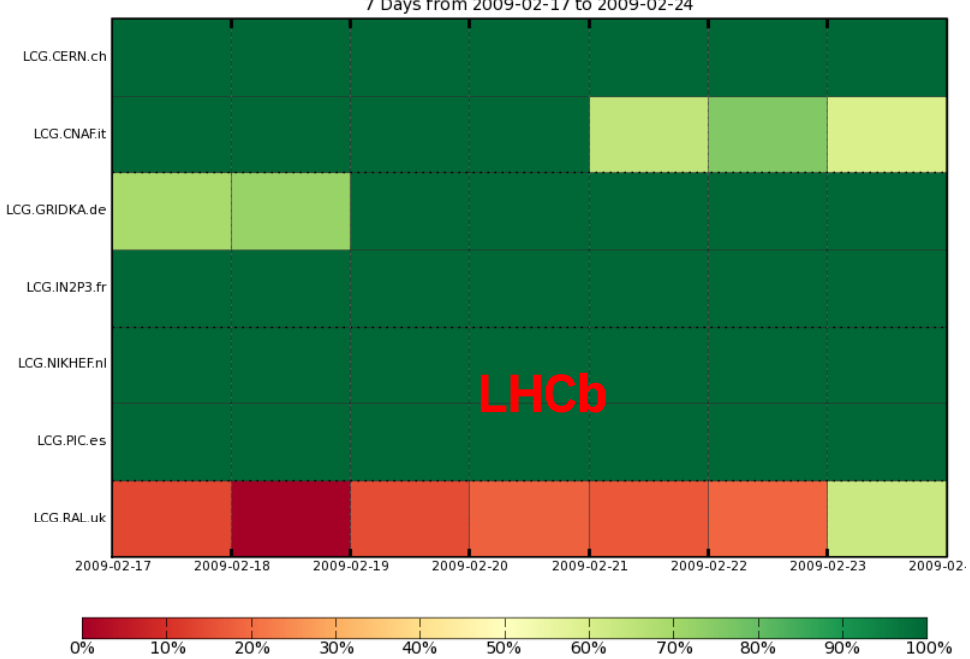


Site Availability using WLCG_SRM2





Site Availability using LHCb Critical Availability



WLCG Service

Immediate Concerns and Challenges

My top three concerns…

1. Dealing with the **realities** of data taking and production which are **bound** to be different to planned exercises including **many more users** and the **pressures** of getting some results;
2. Handling **larger** changes such as new **architectures** that are bound to come (or are already here);
3. **Manpower** and **funding** issues in the post-EGEE III era – **we face significant reductions in the 'middle' of the first data taking run of the LHC**

2009 Data Taking – The Prognosis

- Production activities will work **sufficiently well** – many Use Cases have been tested extensively and for prolonged periods at a level equal to (or even greater than) the peak loads that can be expected from 2009 LHC operation
  - **There will be problems but we must focus on restoring the service as rapidly and as systematically as possible**

💣 **Analysis activities are an area of larger concern – by definition the load is much less predictable**

- Flexible Analysis Services and Analysis User Support will be key

☝ **In parallel, we must transition to a post-EGEE III environment – whilst still not knowing exactly what this entails…**

☺ **But we do know what we need to run stable Grid Services!**

WLCG Service Summary

- **Great strides** have been made in the past year, witnessed by key achievements such as wide-scale production deployment of SRM v2.2 services, successful completion of CCRC'08 and support for experiment production and data taking

- Daily **operations** con-calls – together with the weekly summary – are key to follow-up of service problems

- Some **straightforward** steps for improving service delivery have been identified and are being carried out

➢ **Full 2009-scale testing of the remaining production + analysis Use Cases is urgently required – without a successful and repeatable demonstration we cannot assume that this will work!**

How Can We Improve?

✓ **Change Management**
- Plan and communicate changes carefully;
- Do not make untested changes on production systems – these can be extremely costly to recover from.

✓ **Incident Management**
- The point is to learn from the experience and hopefully avoid similar problems in the future;
- Documenting clearly what happened together with possible action items is essential.

➢ **All teams must buy into this: it does not work simply by high-level management decision (which might not even filter down to the technical teams involved).**

- **CERN IT plans to address this systematically (ITIL) as part of its 2009+ Programme of Work**

Concrete Actions

1. Review on a regular (3-6 monthly?) basis open Oracle "Service Requests" that are significant risk factors for the WLCG service (Tier0+Tier1s+Oracle)
   - The first such meeting is being setup, will hopefully take place prior to CHEP 2009
2. Perform "technology-oriented" reviews of the main storage solutions (CASTOR, dCache) focussing on service and operational issues
   - Follow-on to Jan/Feb workshops in these areas; again report at pre-CHEP WLCG Collaboration Workshop
3. Perform Site Reviews – initially Tier0 and Tier1 sites – focussing again and service and operational issues.
   - Will take some time to cover all sites; proposal is for review panel to include members of the site to be reviewed who will participate also in the review before and after their site

The Goal

- The goal is that – by end 2009 – the weekly WLCG operations / service report is quasi-automatically generated 3 weeks out of 4 with no major service incidents – just a (tabular?) summary of the KPIs

➢ **We are currently very far from this target with (typically) multiple service incidents that are either:**
  - New in a given week;
  - Still being investigating or resolved several to many weeks later

- By definition, such incidents are characterized by severe (or total) loss of service or even a complete site (or even Cloud in the case of ATLAS)

Service Priorities

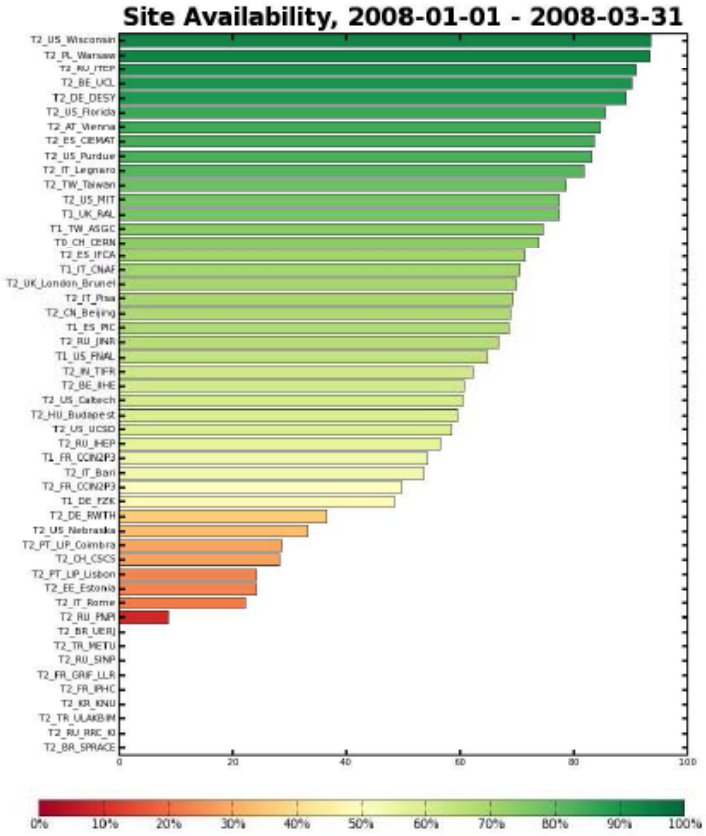
1. **Stability – up to the limits that are currently possible**

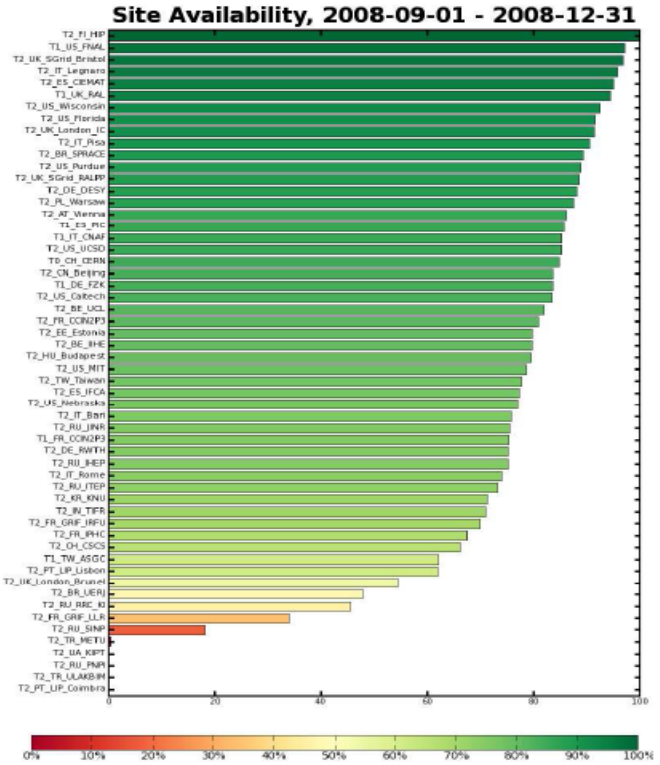
2. **Clarity – where not (well always…)**

- **All sites and experiments should use consistently the existing meetings and infrastructure where applicable**
  - **Join the daily WLCG operations con-call regularly – particularly when there have been problems at your site and / or when upcoming interventions are foreseen**
  - **Always submit GGUS tickets – private emails / phone calls do not leave a trace**
  - **Use the established mailing lists – quite a few mails still "get lost" or do not reach all of the intended people / sites**
- **The LCG Home Page is your entry point!**

Constant improvement of the quality of the infrastructure

Comparison of the CMS site availability based on the results of SAM tests specific for CMS VO
First and last quarter of 2008.





Remaining Challenges

- For me, the key remaining challenge is to handle large numbers of analysis users – who cannot, by definition, be “scheduled” or coordinated in the same way as production activities
- This brings new problems, particularly in the area of Support
- Tools – such as Ganga (covered elsewhere in this conference) – can surely hide much of the complexity and lower support costs, but IMHO will not be sufficient…

Remaining Questions

- Are Grids too complex?
- Are Clouds too Simple?
- ***Do Grids have to be too complex?***
- ***Do Clouds have to be too simple?***

IMHO we can learn much from the strengths and weaknesses of these approaches, particularly in the key (for us) areas of data(base) management & service provision. This must be a priority for the immediate future….

Current Data Management vs Database Strategies

Data Management

- Specify only interface (e.g. SRM) and allow sites to chose implementation (both of SRM and backend s/w & h/w mass storage system)

Databases

- Agree on a single technology (for specific purposes) and agree on detailed implementation and deployment details

WLCG experience from both areas shows that you need to have very detailed control down to a low level to get the required performance and scalability.
How can this be achieved through today's (or tomorrow's) Cloud interfaces?
Are we just dumb???

Some related sessions…

- Ganga/Diane:
  - Demo session (Tuesday afternoon from 16:00 to 20:00). Slot [100]
  - The toolkits will be shown for the 1st time in collaboration with the Fusion cluster.
- Dashboards:
  - Tutorial Friday morning at 9:00, Room Leopardi.
  - “Dashboard tutorial - Site Monitoring for sites serving LHC VOs”
  - URL: http://indico.cern.ch/sessionDisplay.py?sessionId=119&slotId=0&confId=40435#2009-03-06

Conclusions

- Compared to the criteria in Ian Foster's "What is a Grid? A 3-point checklist" WLCG gets **full marks**!
- The concepts of **collaboration** and **community** have been **essential** in achieving this success – over and beyond the technical successes of the underlying infrastructures
- Whilst we have achieved a great deal, the **challenges** that lie ahead are significant, important and by no means VO-specific – **much is applicable to other application communities and probably also different paradigms…**

☺ **Thanks to all who have made this possible…**

Grid Computing in 3 Easy Steps

- Today there are many definitions of *Grid computing*:
- The definitive definition of a Grid is provided by [1] Ian Foster in his article **"What is the Grid? A Three Point Checklist"** [2].
- The three points of this checklist are:

1. Computing resources are not administered centrally;
2. Open standards are used;
3. Non-trivial quality of service is achieved.