CCRC – Conclusions from February and update on planning for May

Jamie Shiers

~ ~ ~

WLCG Overview Board, 31 March 2008

Agenda

- **What were the objectives?**
  - **How did we agree to measure our degree of success?**
  - **What did we achieve?**
- **Main lessons learned**
- **Look forward to May and beyond**
  - **April F2F meetings are this week!**
  - **WLCG Collaboration workshop April 21 – 25**
  - **CCRC'08 Post-Mortem workshop June 12 – 13**

Objectives

- Primary objective (**next**) was to demonstrate that we (sites, experiments) could run together at 2008 production scale
  - **This includes testing all "functional blocks":**
    - **Experiment to CERN MSS; CERN to Tier1; Tier1 to Tier2s etc.**

- Two challenge phases were foreseen:
  1. **February :** not all 2008 resources in place – still adapting to new versions of some services (e.g. SRM – **later**) & experiment s/w
  2. **May:** all 2008 resources in place – full 2008 workload, all aspects of experiments' production chains

💣 N.B. failure to meet target(s) would necessarily result in discussions on **de-scoping!**

- Fortunately, the results suggest that this is not needed, although **much** still needs to be done before, and during, May!

CCRC'08 – Motivation and Goals

What if:

- LHC is operating and experiments take data?
- All experiments want to use the computing infrastructure simultaneously?
- The data rates and volumes to be handled at the Tier0, the Tier1 and Tier2 centers are the sum of ALICE, ATLAS, CMS and LHCb as specified in the experiments computing model

- Each experiment has done data challenges, computing challenges, tests, dress rehearsals, …. at a schedule defined by the experiment

- This will stop: we will no longer be the master of our schedule…
  …. Once LHC starts to operate.

- We need to prepare for this … together ….

A combined challenge by all Experiments should be used to demonstrate the readiness of the WLCG Computing infrastructure before start of data taking at a scale comparable to the data taking in 2008.

This should be done well in advance of the start of data taking on order to identify flaws, bottlenecks and allow to fix those.

We must do this challenge as WLCG collaboration: Centers and Experiments

How We Measured Our Success

- Agreed up-front on specific targets and metrics – these were 3-fold and helped integrate different aspects of the service (CCRC'08 wiki):
  1. Explicit "**scaling factors**" set by the experiments for each functional block: discussed in detail together with sites to ensure that the necessary resources and configuration were in place;
  2. Targets for the lists of "**critical services**" defined by the experiments – those essential for their production, with an analysis of the impact of service degradation or interruption **(discussed at previous OBs…)**
  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**

What Did We Achieve? (High Level)

- Even **before** the official start date of the February challenge, the exercise had proven an extremely **useful** focusing exercise, in helping understand missing and / or weak aspects of the service and in identifying **pragmatic** solutions
- Although later than desirable, the **main bugs** in the middleware were fixed (just) in time and many sites upgraded to these versions
- The deployment, configuration and usage of SRM v2.2 went **better** than had predicted, with a noticeable **improvement** during the month
- Despite the high workload, we also **demonstrated** (most importantly) that **we can support** this work with the available manpower, although essentially no remaining effort for longer-term work (**more later…**)
- If we can do the same in May – when the bar is placed much higher – we will be in a **good position** for this year's data taking
- However, there are certainly **significant concerns** around the available manpower at all sites – not only today, but also in the longer term, when funding is unclear (**e.g. post EGEE III**)

ATLAS Critical Services (PDF)

Tier	Service	Criticality	Consequences of service interuption
0	Oracle database RAC (online, ATONR)	Very high	Possible loss of DCS, Run Control, and Luminosity Block data while running. Run start needs configuration data from the online database. Buffering possibilities being investigated.
0	DDM central services	Very high	No access to data catalogues for production or analysis. All activities stops.
0	Data transfer from Point1 to Castor	High	Short (<1 day): events buffered in SFO disks, backlog transferred as connection is resumed. Long (>1 day): loss of data.
...			
0-1	3D streaming	Moderate	No export of database data. Backlog can be transferred as [soon as] connections are resumed.
... more ...			

CMS Critical Services (wiki)

Rank	Definition	Max. Downtime	Comments
11	CMS Stops Operating	0.5 hours	Not covered yet
10	CMS stops transferring data from Cessy		Cessy output buffer time
9	T0 Production stops		min(T0 input buffer/Cessy output buffer) or defined time to catch up
8	T1/T2 Production/analysis stops		
7	Services critical when needed but not needed all the time (currently includes documentation)	0.5	
6	A service monitoring or documenting a critical service	8	
5	CMS development stops if service unavailable	24	
4	CMS development at CERN stops if service unavailable		
... more ...			

ALICE critical services list

- WLCG WMS (hybrid mode OK)
  - LCG RB
  - gLite WMS (gLite VO-box suite a must)
- FTS for T0->T1 data replications
  - SRM v.2.2 @ T0+T1s
- CASTOR2 + xrootd @ T0
- MSS with xrootd (dCache, CASTOR2) @ T1
- PROOF@CAF @ T0

LHCb Critical Services (CCRC08 wiki)

Service	Criticality
CERN VO boxes	10=critical=0.5h max downtime
CERN LFC service	10
VOMS proxy service	10
T0 SE	7=serious=8h max downtime
T1 VO boxes	7
SE access from WN	7
FTS channel	7
WN misconfig	7
CE access	7
Conditions DB access	7
LHCb Bookkeeping service	7
Oracle streaming from CERN	7
... more ...	

WLCG Services – In a Nutshell…

Services	
ALL	WLCG / "Grid" standards
KEY PRODUCTION SERVICES	+ Expert call-out by operator
CASTOR/Physics DBs/Grid Data Management	+ 24 x 7 on-call

- **Summary slide on WLCG Service Reliability shown to OB/MB/GDB during December 2007**
- On-call service established beginning February 2008 for CASTOR/FTS/LFC (not yet backend DBs)
- Grid/operator alarm mailing lists exist – need to be reviewed & procedures documented / broadcast

Pros & Cons – Managed Services

☺ Predictable service level and interventions; fewer interventions, lower stress level and more productivity, good match of expectations with reality, steady and measurable improvements in service quality, more time to work on the physics, more and better science, …

☹ Stress, anger, frustration, burn-out, numerous unpredictable interventions, including additional corrective interventions, unpredictable service level, loss of service, less time to work on physics, less and worse science, loss and / or corruption of data, …

Design, Implementation, Deployment & Operation

CCRC'08 Preparations…

- Monthly Face-to-Face meetings held since time of "kick-off" during WLCG Collaboration workshop in Victoria, Canada
- Fortnightly con-calls with Asia-Pacific sites started in January 2008
- Weekly planning con-calls ← suspended during February: restart?
- **Daily "operations" meetings @ 15:00 started mid-January**

☝ **Quite successful in defining scope of challenge, required services, setup & configuration at sites…**

➢ **Communication – including the tools we have – remains a difficult problem… but…**

- **Feedback from sites regarding the information they require, plus "adoption" of common way of presenting information (modelled on LHCb) all help**

☺ **We were arguably (much) better prepared than for any previous challenge**

- **There are clearly some lessons for the future – both the May CCRC'08 challenge as well as longer term**

LHC Computing is Complicated!

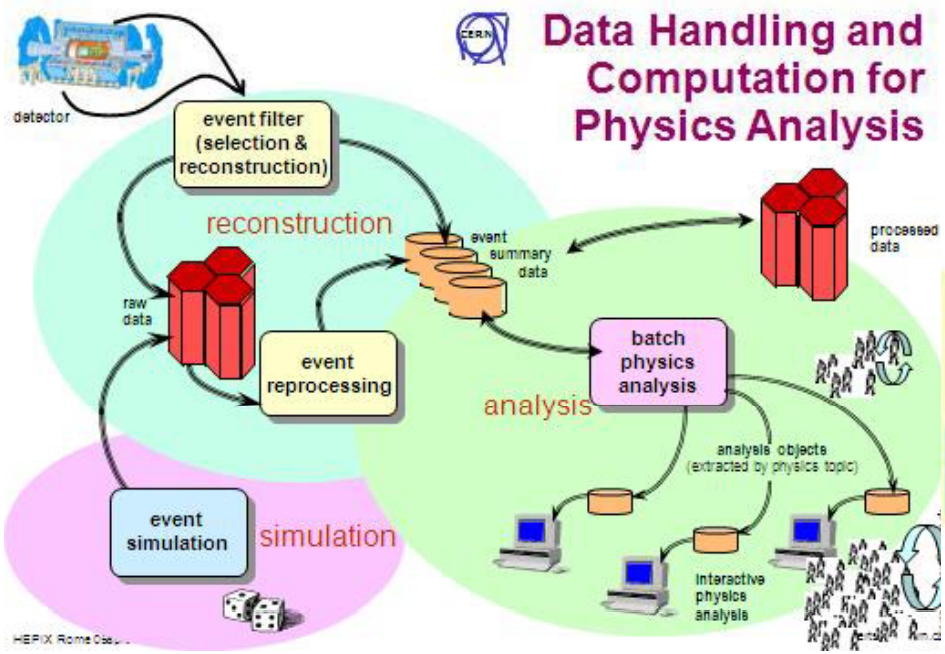
- Despite high-level diagrams (**next**), the Computing TDRs and other very valuable documents, it is very hard to maintain a complete view of all of the processes that form part of even one experiment's production chain
- **Both detailed views of the individual services, together with the high-level "WLCG" view are required…**
- It is ~impossible (for an individual) to focus on both…

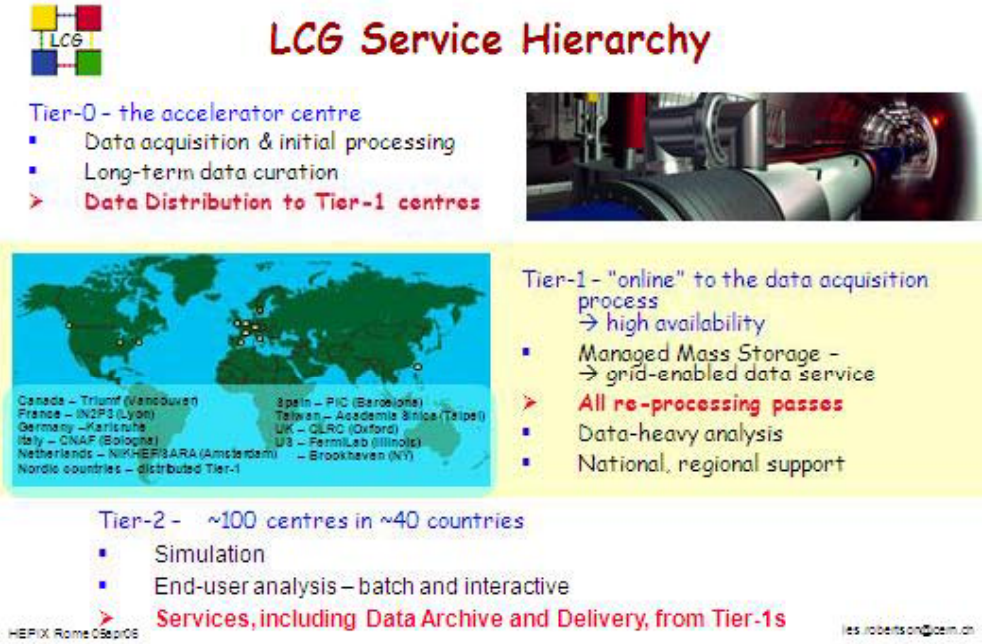
☝ **Need to work together as a team, sharing the necessary information, aggregating as required etc.**

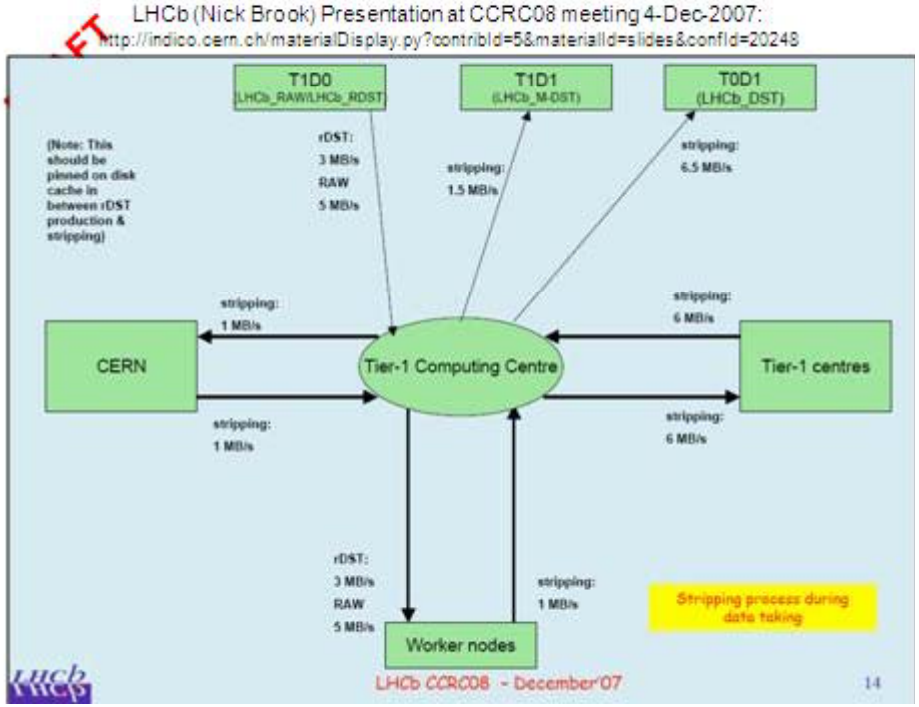
➢ **The needed information must be logged & accessible!**

- (Service interventions, s/w & h/w changes etc.)

☝ **This is critical when offering a smooth service with affordable manpower**





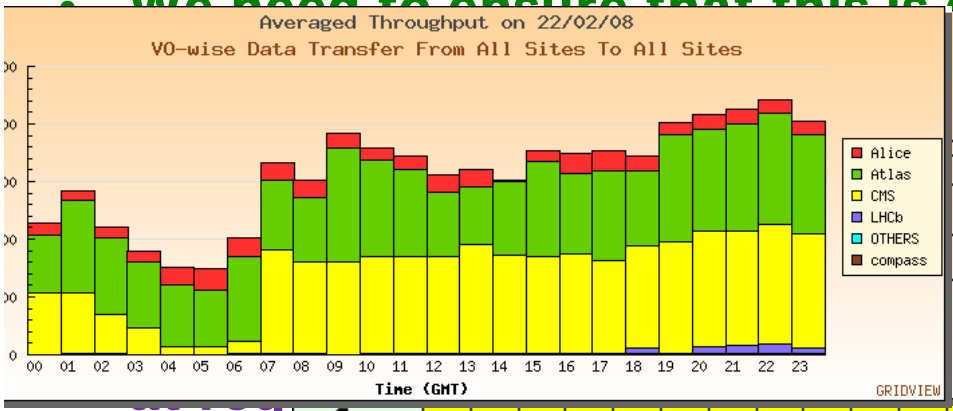


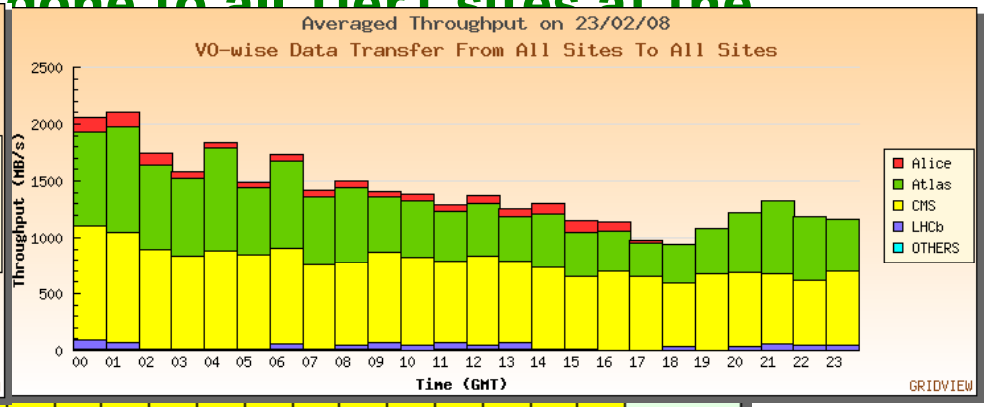
High: Tier-0 Operations

Tier	Service	Criticality	Consequences of service interruption
0	Data transfer from Point1 to Castor	High	Short (<1 day): events buffered in SFO disks, backlog transfered as connection is resumed. Long (>1 day): loss of data.
0	Online-offline database connectivity	High	No export of conditions data, this delays offline calibration and reconstruction. Import of calibration and configuration data proceed via files and does not require this connection.
0	Castor internal data movement	High	Slow down or interruption of Tier-0 processing. Buffers will fill up after 5 days.
0	Tier-0 processing farm	High	Slow down or interruption of Tier-0 processing. Buffers will fill up after 5 days.
0	Oracle database RAC (offline, ATLR)	High	Slow down or interruption of Tier-0 processing. Buffers will fill up after 5 days. No export of database data.

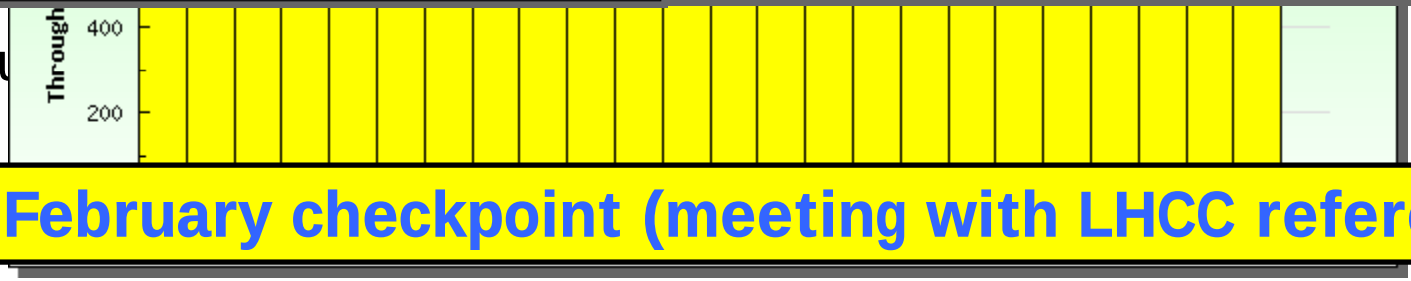
Tier0 – Tier1 Data Export

- We need to sustain 2008-scale exports for at least ATLAS & CMS for at least two weeks
  - The short experience that we have is not enough to conclude that this is a solved problem **[experience to early march suggest this is now OK!]**
  - **The overall system still appears to be too fragile – sensitive to 'rogue users' (what does this mean?) and / or DB de-tuning**
  - **(Further) improvements in reporting, problem tracking & post-mortems needed to streamline this area**
- **We need to ensure that this is done to all Tier1 sites at the**

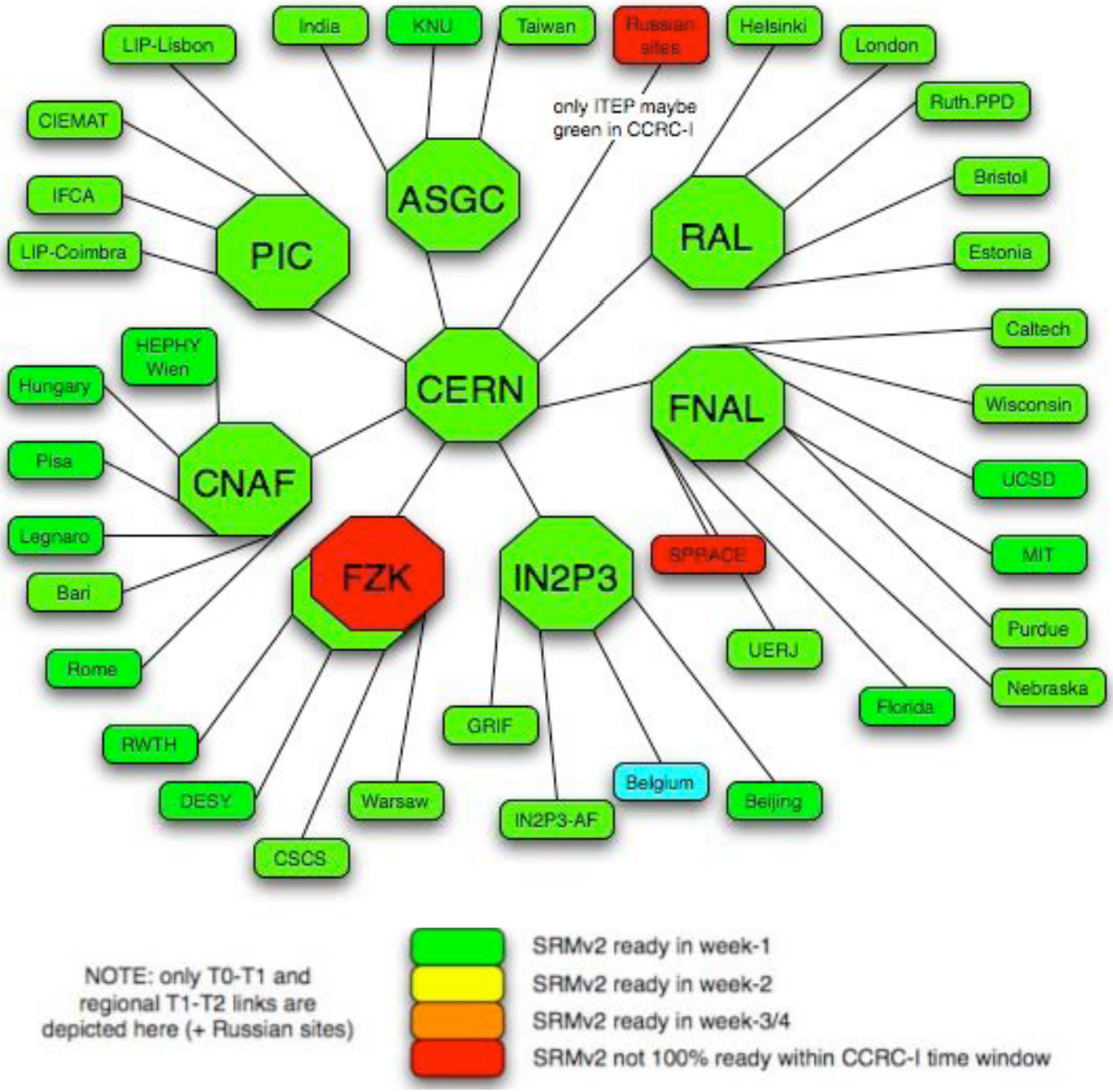




- Re-adju
  - e.g.

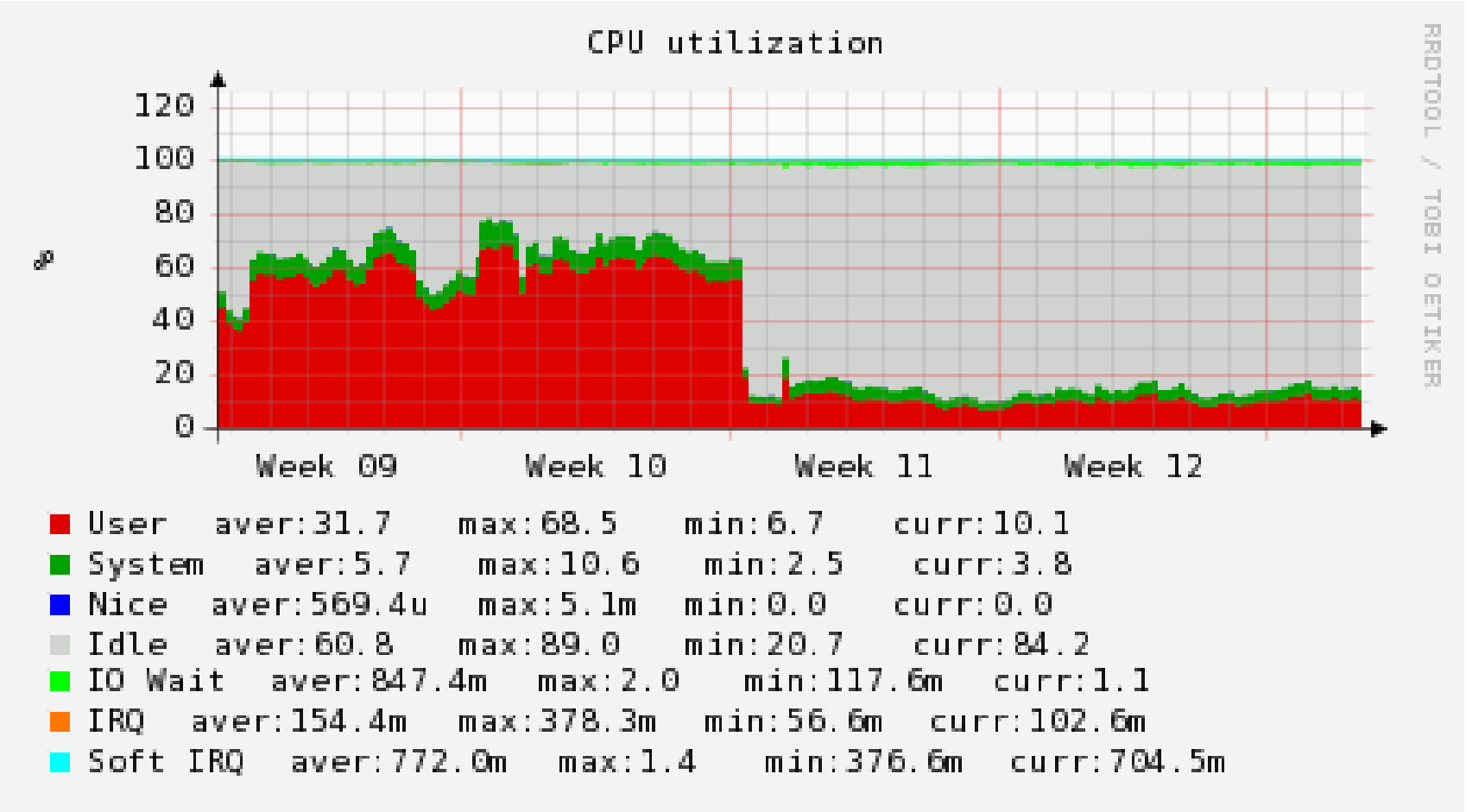


Mid-February checkpoint (meeting with LHCC referees)


India
KNU
Taiwan
Russian sites
Helsinki
London
LIP-Lisbon
only ITEP maybe green in CCRC-I
Ruth.PPD
CIEMAT
Bristol
IFCA
ASGC
RAL
PIC
LIP-Coimbra
Estonia
Caltech
HEPHY Wien
CERN
FNAL
Hungary
Wisconsin
Pisa
CNAF
UCSD
Legnaro
FZK
IN2P3
SPRACE
MIT
Bari
Purdue
UERJ
Rome
Nebraska
Florida
GRIF
RWTH
Belgium
Beijing
DESY
Warsaw
IN2P3-AF
CSCS
NOTE: only T0-T1 and regional T1-T2 links are depicted here (+ Russian sites)
SRMv2 ready in week-1
SRMv2 ready in week-2
SRMv2 ready in week-3/4
SRMv2 not 100% ready within CCRC-I time window

CERN CE Load & Improvements

- Early on in February, CERN LCG CEs regularly had ‘high load’ alarm – scaling issue for the future???



Database Issues

- Databases behind **many** of the most critical services listed by the experiments – "best effort" out of hours at CERN!
- No **change** in Tier0 load seen during February – is the workload representative of 2008 data-taking or will this only be exercised in May? **Later??**
- Oracle Streams **bug** triggered by (non-replicated) table compression for PVSS – **one week** for patch to arrive!
- Interest in moving to Oracle 10.2.0.4 prior to May run of CCRC'08 **if** validated in time by experiments & WLCG – to be discussed at April F2F meetings
  - Motivation: avoid **long delays** (see above) in receiving any needed bug fixes – avoid back-porting to 10.2.0.3 – but is there enough time for realistic testing???
    - **Remember "(shared) cached cursor" syndrome?**

Service Observations (1/2)

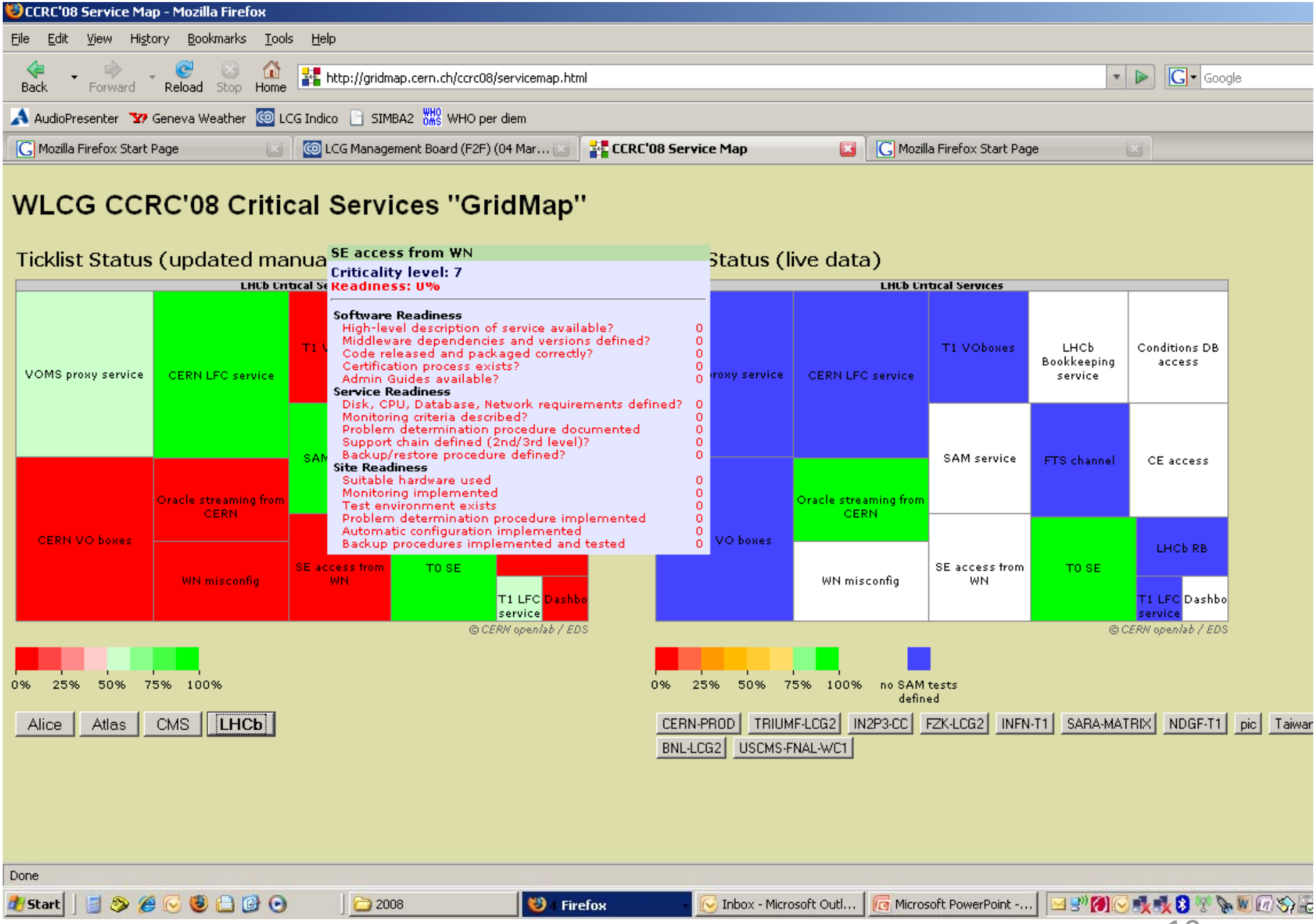
- We must **standardize** and **clarify** the operator/experiment communications lines at Tier0 and Tier1.
- The management board milestones of providing 24x7 support and implementing agreed experiment VO-box Service Level Agreements need to be completed as soon as possible.
- As expected there were many **teething** problems in the first two weeks as SRMv2 endpoints were setup (over 160) and early bugs found after which the SRMv2 deployment worked **generally well**.
- Missing functionalities in the data management layers have been exposed (CCRC'08 "Storage Solutions Working Group" was closely linked to the February activities) and follow-up planning is in place.
- **Few** problems found with middleware – only 1 **serious** (FTS proxy delegation corruption), with work-around in place.
- Tier1s proved fairly reliable: follow-up on their tape operations worked is required.

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 **/ mailing list**	99%
1 hour	Operator response to alarm / call to x5011 **/ mailing list**	100%
4 hours	Expert intervention in response to above	95%
8 hours	Problem resolved	90%
24 hours	Problem resolved	99%


CCRC'08 Service Map - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Back Forward Reload Stop Home
http://gridmap.cern.ch/ccrc08/servicemap.html
Google
AudioPresenter Geneva Weather LCG Indico SIMBA2 WHO per diem
Mozilla Firefox Start Page
LCG Management Board (F2F) (04 Mar...
CCRC'08 Service Map
Mozilla Firefox Start Page
WLCG CCRC'08 Critical Services "GridMap"
Ticklist Status (updated manua
Status (live data)
SE access from WN
Criticality level: 7
Readiness: 0%
Software Readiness
High-level description of service available? 0
Middleware dependencies and versions defined? 0
Code released and packaged correctly? 0
Certification process exists? 0
Admin Guides available? 0
Service Readiness
Disk, CPU, Database, Network requirements defined? 0
Monitoring criteria described? 0
Problem determination procedure documented 0
Support chain defined (2nd/3rd level)? 0
Backup/restore procedure defined? 0
Site Readiness
Suitable hardware used 0
Monitoring implemented 0
Test environment exists 0
Problem determination procedure implemented 0
Automatic configuration implemented 0
Backup procedures implemented and tested 0
LHCb Critical Services
VOMS proxy service
CERN LFC service
CERN VO boxes
Oracle streaming from CERN
WN misconfig
SE access from WN
T0 SE
T1 LFC service
Dashbo
T1 VOboxes
LHCb Bookkeeping service
Conditions DB access
SAM service
FTS channel
CE access
LHCb RB
© CERN openlab / EDS
0% 25% 50% 75% 100%
no SAM tests defined
Alice Atlas CMS LHCb
CERN-PROD TRIUMF-LCG2 IN2P3-CC FZK-LCG2 INFN-T1 SARA-MATRIX NDGF-T1 pic
BNL-LCG2 USCMS-FNAL-WC1
Done
Start
2008
Firefox
Inbox - Microsoft Outl...
Microsoft PowerPoint -...

Experiment View

In Order of Appearance (March F2F…)

- CMS
  - [Very] Detailed presentation of up-front metrics per functional block
  - 100% success not reported, but well understood status
- ATLAS:
  - CCRC was a very useful exercise for ATLAS
  - Achieved most milestones in spite of external dependencies
  - It's difficult to serve the Detector, Physics and IT community!
- ALICE:
  - For ALICE, the CCRC exercise has fulfilled its purpose
    - Focus on data management
    - Brings all experiments together
    - Controlled tests, organization
- LHCb:
  - Initial phase of CCRC'08 was dedicated to development and testing of DIRAC3
  - CCRC'08 now running smoothly
    - Online->T0 and T0-T1 transfers on the whole a success
    - Some issues with reconstruction activity and data upload from the WNs
    - Investigating with Tier-1s recent problem of determining file sizes using gfal
    - Quick turnaround for reported problems

Service Observations (2/2)

- Some particular experiment problems seen at the WLCG level:
  - **ALICE:** Only one Tier1 (FZK) was fully ready, NL-T1 several days later then the last 3 only on the last day (RAL being setup in March)
  - **ATLAS:** Creation of physics mix data sample took much longer than expected and a reduced sample had to be used.
  - **CMS:** Inter-Tier1 performance not as good as expected.
  - **LHCb:** New version of Dirac had teething problems – 1 week delay.

  ➢ Only two inter-experiment interferences were logged: FTS congestion at GRIF caused by competing ATLAS and CMS SEs (solved by implementing sub-site channels) and degradation of CMS exports to PIC by ATLAS filling the FTS request queue with retries.

- We must collect and analyze the various metrics measurements.

☺ **The electronic log and daily operations meetings proved very useful and will continue.**

☹ **Not many Tier1s attend the daily phone conference and we need to find out how to make it more useful** [for them!]

- Overall a good learning experience and positive result. Activities will continue from now on with the May run acting as a focus point.

Well, How Did We Do?

- Remember that prior to CCRC'08 we:

  a) Were not confident that we were / would be able to support all aspects of all experiments simultaneously
  b) Had discussed possible fall-backs if this were not demonstrated
     - **The only conceivable "fall-back" was de-scoping…**

- Now we are reasonably confident of the former
- Do we need to retain the latter as an option?
- Despite being rather late with a number of components (not desirable), things settled down reasonably well
- Given the much higher "bar" for May, need to be well prepared!

Main Lessons Learned

☺ **Generally, things worked reasonably well…**

➢ **Still improvements in communication are needed!**

- Tools still need to be streamlined (e.g. elog-books / GGUS), and reporting automated
- ¿ **Service dashboards – should be in place before May…**
- F2Fs and other meetings working well in this direction!

- **Pre-established metrics extremely valuable!**
  - As well as careful preparation and extensive communication!

☛ **Now continuous production mode –this will continue – as will today's infrastructure & meetings**

Recommendations

- ✓ **To improve communications with Tier2s and the DB community, 2 new mailing lists have been setup, as well as regular con-calls with Asia-Pacific sites (time zones…)**
- Follow-up on the lists of "**Critical Services**" must continue, implementing not only the appropriate monitoring, but also ensuring that the WLCG "standards" are followed for **D**esign, **I**mplementation, **D**eployment and **O**peration
- **Clarify reporting and problem escalation lines (e.g. operator call-out triggered by named experts, …) and introduce (light-weight) post-mortems when MoU targets not met**
- ➢ **We must continue to improve on open & transparent reporting, as well as further automations in monitoring, logging & accounting**
- ☝ **We should foresee "data taking readiness" challenges in future years – probably with a similar schedule to this year – to ensure that full chain (new resources, new versions of experiment + AA s/w, middleware, storage-ware) is ready**

CCRC'08 Summary from February

- The preparations for this challenge have proceeded (largely) smoothly – we have both learnt and advanced a lot simply through these combined efforts
  - As a focusing activity, CCRC'08 has already been very useful
  - We will learn a lot about our overall readiness for 2008 data taking
  - We are also learning a lot about how to run smooth production services in a more sustainable manner than previous challenges
- **It is still very manpower intensive and schedules remain extremely tight: full 2008 readiness still to be shown!**
- **More reliable – as well as automated – reporting needed**
- Maximize the usage of up-coming F2Fs (March, April) as well as WLCG Collaboration workshop to fully profit from these exercises
- **June on: continuous production mode (all experiments, all sites), including tracking / fixing problems as they occur**

Preparations for May and beyond…

- **The May challenge must be as complete as possible – we must continue exercising all aspects of the production chain for all experiments at all sites until first collisions & beyond!**
- **This includes the use of full 2008 resources at all sites!**
- Aim to agree on baseline versions for May during April's F2F meetings
- Based on versions as close to production as possible at that time (and not (pre-)pre-certification!)
- **Aim for stability from April 21st at least!**
  - The start of the collaboration workshop…
- This gives very little time for fixes!

- **Beyond May we need to be working in continuous full production mode!**
- **March & April will also be active preparation & continued testing – preferably at full-scale!**

- **CCRC'08 "post-mortem" workshop: June 12-13**

N.B. – need to consider startup scenarios!

May Preparations

- Monthly F2F meetings continue – April's is tomorrow!
  - Agenda: http://indico.cern.ch/conferenceDisplay.py?confId=30246
- Beyond that, the WLCG Collaboration Workshop (April 21 – 25) has a full day devoted to the May run

☝ **This might seem late, but the experiments have been giving regular updates on their planning for several months in advance since the beginning of CCRC'08 (and before…)**

- In terms of services, there is very little change (some bug fixes, most Tier2s now on SRM v2.2, better monitoring...)

➢ **Largest change is increased scope of experiment testing – plus full 2008 resources at sites!**

Summary

- The February run of CCRC'08 was largely **successful** and introduced the important element of up-front and **measurable** metrics for many aspects of the service

- We still have a lot to do to **demonstrate** full 2008 readiness – May and beyond will be a busy time with no let up in production services prior to then

- We have developed a way of planning and operating the service that works **well** – re-enforce this and build on it incrementally

➢ **This is it – the WLCG Production Service!**

Acknowledgements

- The achievements presented above are the result of a concerted effort over a prolonged period by a large number of people – **some I know, but many I don't…**
  - **Experiments, sites, services, …**

- **It wasn't always easy – it wasn't even always fun!**

- Clearly, in a project of this magnitude, such efforts are needed to **succeed** – and computing is just a tiny part of the whole

- But somehow, I think that this effort should be **acknowledged**, in a way that is real to those who have made such a huge effort, and on whom we rely for the future...