# WLCG Services in 2009

**Jamie.Shiers@cern.ch**

~ ~ ~

**dCache WLCG T1 Data Management Workshop,
15th January 2009**

# Overview

- I have given myself a somewhat more general title than that proposed by Jos

- I intend to cover the main operational issues as witnessed by the WLCG service in recent months...

- Plus the expected "challenges" in the up-coming period

- I will focus on the story as we tell it to the LHCC referees and other high-level review bodies

- Plus where and how we get this information...

# Expectations

- I am not a believer in "emotive words" when describing services

- Concrete, achievable, measurable objectives are much more useful and relevant (often referred to as "S.M.A.R.T.")

- We have some *de-facto* agreed "Key Performance Indicators" (which I'll explain...)

- Plus some Service Targets derived – over many years of practical experience and dialogue with the experiments – that we use to measure the performance of the service

# Critical Service Follow-up

This activity was triggered by the experiments at the WLCG Overview Board in 2007. The proposal below has been formally accepted by the CMS Computing Management and presented to the MB and other bodies on several occasions.

It continues to be the baseline against we regularly check our response, particularly in the case of alarm tickets.

| Time | | Time |
|------|--|------|
| En | | |
| 30 | | |
| 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% |

# How Are We Doing?

- Better and better, every day, every week and every month...

💣 **But there are still avoidable holes and some areas of significant concern**

- The experiments are doing a lot of useful work, day in, day out, work-day, weekend and holiday!

☺ **And they appreciate it and acknowledge it!**

- I still believe that we can – easily – do better in some areas, with **LESS** effort and stress...

# How?

- We still see "emergency" interventions that could be "avoided" – or at least foreseen and scheduled at a more convenient time

- We still see scheduled interventions that are not sufficiently well planned – that run well into "overtime" or have to be "redone"

- Or those that are not well planned or discussed and have a big negative impact on ongoing production

- The concrete examples that I have in mind refer to CERN and cover more or less uniformly the 25 year period that I have been working there...

- But they are not just at CERN – take some time to plan your interventions: the experiments do understand; it will be less painful for you and typically less effort than alternatives...

- ☺ And that way I won't talk about you at the up-coming MB...

# Getting More Concrete…

☺ **Most of the time the WLCG services run well**

☺ **Response to problems is usually (well) within the agreed – somewhat aggressive IMHO – targets**

💣 **But the problem is "what happens when things go (badly) wrong?"**

💣 **In the best case there might be an unscheduled intervention for some hours – possibly following up to a weekend of effective downtime**

# GGUS Summary – Last Week

| VO | USER | TEAM | ALARM | TOTAL |
|---|---|---|---|---|
| ALICE | **1** | 0 | 0 | 1 |
| ATLAS | 26 | 5 | 0 | 31 |
| CMS | 3 | 0 | 0 | 3 |
| LHCb | 2 | 0 | **1** | 3 |

- Tickets for ALICE are very rare → detail

- We always look at alarm tickets → LHCb

# When This Happens…

- The most important thing is to keep people informed

- You might not know what the problem is and hence not be able to estimate when things will be solved so say it!

  - "We are currently investigating the problem. More news at xx:xx"

- An update at the daily WLCG operations call – either by dialing in or by e-mail – is valuable and appreciated

- **Think also of providing a Service Incident Report – this (IMHO) is useful for you as service providers, plus your colleagues at other labs and your users**

- It doesn't have to be heavy weight – just a brief analysis of the problem, resolution, lessons learned, actions, …

# When It Gets Worse…

- "It won't happen to us"

- It is even more important to follow the above

# In Other Words…

✓ **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.

# 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

- 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)

# Back to dCache…

- It is important for everybody that services supporting the main production Use Cases (import/export of data, reprocessing, MC where applicable) run smoothly and predictably – which does not mean without "hiccoughs" – with reasonable operational effort as soon as possible

- In reality, whilst data exchange continues to be well tested, **reprocessing** (also multi-VO where applicable), not to mention **analysis**, is still somewhat uncharted territory

- Which means that there will be problems. And more problems. If you don't believe me just wait…

13

# So What Do We Do?

- The most obvious strategy is to push for as early and as thorough testing of multi-VO re-processing activities as soon as possible

- I don't (and didn't at the time of LEP either) believe in arbitrary numbers like # tape mounts, throughput etc. – what counts is "Can you support production" under realistic and agreed conditions

- It may be necessary to implement some policies to limit or avoid inter- or intra- VO "interference"

- For the 2009 run – and for ATLAS at least that starts with cosmics from April – there is no time for "buying your way" out of the problem…

- And that's before we get to analysis…

# We Ain't Seen Nothing Yet...

- If we can get the services into a reasonable state by $<(<)$ mid-year we have some chance of being ready for much larger numbers of users, much more chaotic access patterns and much more "disorganized" attempts to access and process data

- Which is actually what we need to support – the users' ability to extract science from the data and to maximize the potential of the detectors and the machine

- Thus I believe it is urgent that we clarify what "analysis services" mean for all sites

# So What Does This Mean?

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

2. Clarity – where not (well always…)

➢ **Please use 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**

    - **So far BNL and PIC (amongst dCache WLCG T1 sites) join very regularly, NIKHEF often – others rarely or never…**

- **They really do help not only with information flow but also in maintaining and calm(-er) and (more) relaxed atmosphere…**

# **Summary**

- We started off being a Data Grid

- We continue to be a Data Grid

- We will always be a Data Grid

- Data is what sets us apart

- Data is the challenge