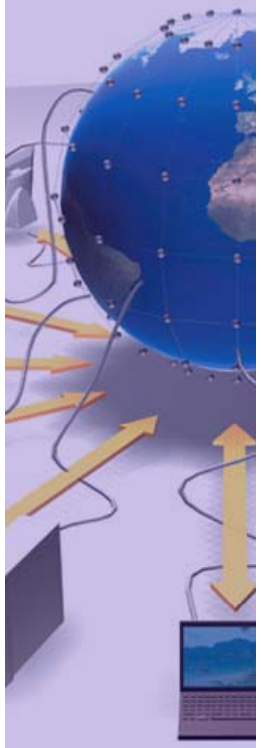
GS Grid Support

CERN IT Department



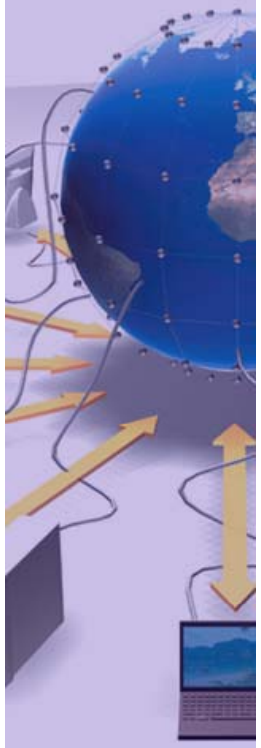
Service Reliability & Critical Services

January 15th 2008

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

CERN

Experiment Requests



- For the most critical services, maximum downtime of **30'** has been requested
- As has been stated on several occasions, including at the WLCG Service Reliability workshop and at the OB, maximum downtime of **30'** is impossible to guarantee at affordable cost
- **30'** – even for maximum time for a human to begin to intervene – cannot be guaranteed
  - e.g. IT department meeting of yesterday!

➢ **But much can be done in terms of reliability – by design! (See next slides…)**

- A realistic time for intervention (out of hours – when they are likely to occur!) is **4 hours**
- Christmas shutdown text typically says ½ a day

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Reliable Services – The Techniques

- ☺ **DNS load balancing**
- ☺ **Oracle "Real Application Clusters" & DataGuard**
- H/A Linux (less recommended… because its not really H/A…)
- 💣 **Murphy's law of Grid Computing!**
- Standard operations procedures:
  - Contact name(s); basic monitoring & alarms; procedures; hardware matching requirements;
- ➢ **No free lunch! Work must be done right from the start (design) through to operations (much harder to retrofit…)**
- Reliable services take **less** effort(!) to run than unreliable ones!
- 💣 **At least one WLCG service (VOMS) middleware does not currently meet stated service availability requirements**
- 💣 **Also, 'flexibility' not needed by this community has sometimes led to excessive complexity (complexity is the enemy of reliability) (WMS)**
- ➢ **Need also to work through experiment services using a 'service dashboard' as was done for WLCG services (see draft service map)**

Service Reliability: Follow-up Actions

1. Check m/w (prioritized) against techniques – which can / do use them and which cannot? → Priorities for development (service)
2. Experiments' Lists of critical services: service map (FIO+GD criteria)
3. Measured improvement – how do we do it?
4. VO Boxes → VO services
5. Tests – do they exist for all the requested 'services'? → SAM tests for experiments
6. ATLAS & CMS: warrant a dedicated coordinator on both sides
7. Database services: IT & experiment specific
8. Storage – does this warrant a dedicated coordinator? Follow-up by implementation
9. Revisit for Tier1s (and larger Tier2s)
10. **Overall coordination? → LCG SCM → GDB → MB/OB**
11. Day 1 of WLCG Collaboration workshop in April (21st)
12. Long-term follow-up? → solved problem by CHEP 2009
13. "Cook-book" – the current "knowledge" is scattered over a number of papers – should we put it all together in one place? (Probably a paper of at least 20 pages, but this should not be an issue.)

VOBOX Hardware:

- Resource requirements and planning
  - it is not always easy to have an additional disk on demand because “/data” becomes full
- Hardware warranty
  - Plan for hardware renewal
  - Check warranty duration before moving to production
- Hardware naming and labeling
  - Make use of aliases to facilitate hardware replacement
  - Have a “good” name on the sticker
    - **e.g. All lxbiiii machines may be switched off by hand in case of a cooling problem**

Some “critical services” run over Xmas were just that – and nodename hard-coded in application!

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

On-going Follow-up

- We have one day (Monday 21st April) dedicated to follow-up on **measured** improvement in service reliability
- Using the Grid Service Map, we will focus on the most critical services and work through the lists of all experiments
- It requires **work** on all sides (service provider, developers, experiments) to make concrete progress
- For the services where the guidelines are already implemented, production experience is **consistent** with the requests from the experiments
- We should use the features of the Grid to ensure that the overall service reliability is consistent with the requirements (i.e. no SPOFs)

☝ Individual components may fail, but the overall service can and should continue!