WLCG Service Interventions

https://twiki.cern.ch/twiki/bin/view/LCG/WLCGServiceLog

Intervention Log

- There are clear correlations between the service intervention log gleaned from EGEE broadcasts (and other sources ☹) and SAM tests
- This is hardly surprising – if e.g. the DB used by FTS/LFC is down, the SAM tests will show failures
- We need to understand *why* things go down – *what* are the key causes – and fix them
- This requires clear and systematic reporting – in particular to the weekly operations meeting

Building Robust Services

- In 2005, did an *"a priori"* analysis of main services and their criticality
- This led to a deployment plan, including h/w requirements, operations issues etc. ([WLCG service dashboard](#))
  - Some m/w issues were identified but never resolved…
- Typically load-balanced servers + Oracle clusters if needed
  - 💣 Limited amount of H/A Linux – which has some limitations!
  - ☹ e.g. nodes must be on same network switch
- We have now quite extensive experience running production services, and hence an *"a posteriori"* analysis is called for
- ➢ **But this must be a coordinated, end-to-end analysis, and not just individual (component) services (if what the user sees is not working, its not working.)**
- See also EGEE II recommendation 42 section 1.3 (see notes)

Key Grid / Physics Services

- Storage management services have been a cause of concern for some time and have had a corresponding impact on e.g. effective inter-site transfer throughput
- It is hard to look 'inside' a site, but deployment complexity, single-points-of-failure and shortage of resources are believed to be key causes of problems
- Data management and other services typically include a database backend – these too require particular attention (appropriate resources, configuration, expertise, …)
- A not insignificant number of problems are related to **'database housekeeping'** issues – this is independent of the type of DB used (Oracle, PostgreSQL, MySQL, …)
  - Table defragmentation, pruning, clean-out of old records etc.
  - Oracle Certified Professional training for Tier1 DBAs?

Proposal

- We plan to present the results of the review of CERN services – together with techniques for handling the main types of intervention – at the WLCG workshop (operations BOF) in BC ([FTS 2.0 wiki](#))
- We propose that the key areas of data management and storage management be addressed systematically across at least Tier0 and Tier1 sites
- The w/s in BC seems a good place to kick this off…
- (Possibly a follow-up (small) w/s in November?)
- Production deployment early 2008 (first round)

Conclusions

- With the agreement of a roll-out schedule for SRM v2.2 services, the deployment phase of WLCG services for the initial running period of the LHC is close to completion.

➤ **The next main challenge is improving service quality to an acceptable and affordable level – essential if we are to optimally exploit the LHC.**

- Aside from infrastructure (power, cooling, network) failures, the primary causes of unscheduled interventions and service loss are:

  – Storage management services;

  – Data management and other database-related services.

- Techniques for building robust services are mature and well understood.

➤ **We just need to "make it so".**

One more slide to go…

Final Remark

- Unfortunately, we tend to focus a lot on when things go wrong and not *vice-versa*
- I'd like to highlight a recent event at SARA which follows from a discussion at the Stockholm operations w/s (maybe):
- ***"The problems with pnfs is fixed now and the FTS channel CERN-SARA is put to Active again."***
- This is exactly the right way – set channel inactive, broadcast, fix, etc. (Use features of the service!)
- ✓ **Thanks a lot, Ron!**

CERN Service Review

- For each service need current status of:

    - Power supply (redundant including power feed? Critical? Why?)
    - Servers (single or multiple? DNS load-balanced? HA Linux? RAC? Other?)
    - Network (are servers connected to separate network switches?)
    - Middleware? (can middleware transparently handle loss of one of more servers?)
    - Impact (what is the impact on other services and / or users of a loss / degradation of service?)
    - Quiesce / recovery (can the service be cleanly paused? Is there built-in recovery? (e.g. buffers) What length of interruption?)
    - Tested (have interventions been made transparently using the above features?)
    - Documented (operations procedures, service information)