WLCG Database Technical Evolution Group Future Work Priorities

GDB meeting

Dave Dykstra, Fermilab

dwd@fnal.gov

Database TEG topics

- Conditions (CORAL, COOL, Frontier, Squid)
- DB operations (Oracle)
- Online DBs (PVSS mostly)
- Distributed apps (NoSQL)

Plans related to Conditions DB, COOL, and CORAL

- None of the experiments are planning to substantially change their Conditions DB model or support tools
- ATLAS, CMS, and LHCb expect that COOL and CORAL will continue to be supported by CERN
  - ATLAS expects the same for CoralServer
- ATLAS reviewing volume & complexity of their Conditions which are significantly above CMS

Plans related to Frontier service

- Used to distribute conditions database worldwide for ATLAS & CMS with a RESTful http protocol
- LHCb planning to use it in the future
- CMS planning to add second server site (@BNL)
- ATLAS has 6 server sites, reevaluate in 2013
- CMS planning on adding failover proxy Squids at CERN and at new server site
  - Prevents site Squid proxy failures from affecting sites that are working normally

Plans related to Squid

- Site Squid proxy caches are currently closely monitored by CMS Frontier team
  - TEG recommends moving this monitoring to WLCG
- Also recommend that each site use one set of Squids for all production, approved applications (known behavior, Frontier & CVMFS for now)
  - Sites that allow opportunistic use of their job slots by other applications should have a separate set of Squids for those applications
- Recommend WLCG standard for jobs to discover Squids when they arrive on a site

Oracle DB operations plans

- No major changes planned
- Gradually scale up
- Replace use of Streams with Active Data Guard where possible, although 11g Streams are improved over 10g Streams

Online Databases plans

- All experiments use PVSS commercial product for detector monitor data, plan to keep using it
  - Major upgrade 4.0 coming, CERN taking active role
  - May conflict with timing of long LHC shutdown
- CMS & ATLAS use different methods of simultaneous loading of Conditions to HLT nodes
  - CMS: Frontier+Squid; ATLAS: CoralServer+Proxy
  - Consolidation possible but no plans to do so

NoSQL Databases plans

- CMS and especially ATLAS have done much good work evaluating them
  - ATLAS using Cassandra in production
  - CMS using CouchDB and MongoDB in production
  - Both have some production use of Hadoop's HDFS
- Hadoop's MapReduce identified as potentially most disruptive if used to analyze event data
  - CERN-IT plans to set up a cluster
- Build LHC-wide communities around the tools

For more info

- https://twiki.cern.ch/twiki/bin/view/LCG/WLCGTEGDatabase