Database Technical Evolution Group (extract for GDB)

Dario Barberis & Dave Dykstra

Report section editors: Rainer Bartoldus, Luca Canali, Gancho Dimitrov, Giacomo Govi, Mario Lassnig, Simon Metson, Andrea Valassi

Databases for Conditions Data

- CORAL (ATLAS, CMS, LHCb) and COOL (ATLAS, LHCb) are examples of very successful common projects
- The CMS conditions software relies on CORAL for accessing Oracle (both directly and via Frontier) and SQLite.
  - CMS expects that CERN IT will provide the users with an adequate support for the CORAL framework, which should be devoted mainly in bug fixing and in improving performance bottlenecks when identified.
- ATLAS relies on COOL for the conditions database infrastructure, on CORAL to access the database layer, and on CoralServer for database access in the online environment.
  - ATLAS expects that these products will be supported by CERN as long as they are used by ATLAS and other experiments.
- LHCb needs CORAL and COOL to continue to be supported, as they are essential components of the LHCb software.

Access to Conditions Data

- CMS, ATLAS and soon LHCb use Frontier/Squid for Conditions DB access.
  - Central monitoring of worldwide Squids is very important to keep the Squids operating properly. The monitoring is now done by computers operated by CMS Frontier, but we recommend that a plan be made to transition the Squid central monitoring to WLCG
  - Locating the Squids is currently done separately per experiment and application, but we recommend that there be a WLCG standard way for jobs to locate Squids
  - We recommend that sites share Squids for all production services (currently Frontier and CVMFS)
  - Frontier/Squid should be recognized as a WLCG service and treated accordingly (GOCDB, GGUS, central rpm repository, monitoring)

NoSQL Recommendations

- It seems to us evident that there are valid use cases for providing and supporting at least one of the NoSQL technologies at CERN
- In order to be able to properly advise developers within the experiments groups, the CERN-IT-DB group should test the most common NoSQL products for the already known use cases and acquire expertise with them
  - Technology tracking and market survey should also be part of this task.

➡ Discussion on this point: CERN-IT-DB vs CERN-IT vs WLCG vs community support

- CERN IT should deploy a suitably sized Hadoop cluster
  - Focus on Hadoop rather than fragment effort over a variety of NoSQL tools
    - Other tools can, and will, be run ad-hoc by experiments as necessary
  - Hadoop clients, including pig/hive available on user interfaces (lxplus?)
  - Reasonably sized HBase installation
  - We make no operational requirements on the cluster, and appreciate that it will require training etc. for ops staff, so may run at low service level initially.
  - In the end it may need development, integration and production clusters
- Experiments would like to be involved in deployment discussions
- Build a community around the tools
  - Best practice doesn't really exist at CERN; have a forum to communicate what is learnt
  - Other groups may be interested in using these tools (Dashboard seems like a good candidate for example) but a central service is needed before expanding the user base