Database Meeting 2016-10-05

Date

05 Oct 2016

Attendees

Local: Igor Gaponenko, Vaikunth Thukral, John Gates, Brian Van Klaveren, Andy Hanushevsky

Remote: Fabrice Jammes, Fritz Mueller

Discussion items

Fritz:

- At DMLT in Tucson - very productive, 5 year re-plan and resource loading needs to be finished 80% by end-of-week
- Management folks present to answer all questions
- Preparation for reviews is underway
- Bringing up DAX services on PDAC should be our point of focus for the upcoming week

Fabrice:

- PDAC - Qserv is installed but still some issues with container management
- Basic installed image/version and integration tests are working
- Loading SDSS data in local instance at Clermont. Handing over that config to Igor just in case
- Stripe82 bulk loading prep is finished, loader is installed on all nodes and data is on GPFS
- Can use integration test scripts to figure out (reverse engineer) loading procedure like Jason Alt did when he visited SLAC (not recommended)
- Group decided to name the database at NCSA "stripe82"
- For backup at PDAC (since no solution exists yet) - can manually copy data dir after loading is finished
- No update on Swarm yet, just minor xrootd config improvements
- Consider Rocket instead of Docker for containers?

John:

- Finished Qserv related issues from last week, on to PDAC
- Reviewing Fabrice's ticket DM-7104
- Having trouble pulling Qserv images from docker hub, Fabrice will look into it
- Need another node for developer specific tasks/monitoring
- Another procurement of machines coming up next month, can ask for one dedicated for that
- Alternate option - make a VM in OpenStack and figure out networking options needed

Brian:

- Working on Docker containers for services at PDAC
- Mostly busy with Slack migration
- SLAC HR told Brian to talk to University HR to create new Big Data positions, for systemwide reclassification definition

Igor:

- Working on QA between data at in2p3 vs PDAC
- coadd, images and forced sources along overlap covers small fraction of the sky, hole in the patch
- Consider having 2 DBs in Qserv - "stripe82" and "stripe82_v1"
- Gregory requested having coadd images accessible on GPFS, which is now done and all coadds appear to be there
- Finishing migration from in2p3
- Trying a loader config in a distributed fashion rather than how it was done earlier
- Chunk to worker mapping - consider preloading mapping before loading data with qserv-admin
- Fritz's DHT prototype could help Igor in figuring out how to parallelize the loader - service model

AndyH:

- Recently back from CERN
- Reading up on sizing models

Vaikunth:

- Continue to debug spurious recorded values. Narrowed down - observed different numbers for same query!
- Diff file shows the different values when query submitted at different times on exact same data (same number of lines in result too)
- Fabien ran his queries directly on the backend REST interface, Vaikunth uses the browser Sense plugin, not sure if that is the cause of the bug /problem
- Vaikunth and Fritz @ LBNL on Nov 1 for XSWAP renewal strategy meeting

Other/Monitoring:

- Nagios @ PDAC is default, we should ask for Ganglia and not sure whether we should use ELK
- Lots of options available, more discussion and follow up between Fritz and Fabrice/Fabien
- Consider moving entire ELK stack config from in2p3 and go all out on monitoring at PDAC? Could also help Vaikunth for XSWAP work