Database Meeting 2016-02-17

Date

17 Feb 2016

Attendees

- John Gates, Andy Salnikov, Fritz Mueller, Brian Van Klaveren, Vaikunth Thukral, Serge Monkewitz, Fabrice Jammes, Unknown User (kelsey), Jacek Becla, Kian-Tat Lim

Discussion items

DM Leadership changes announcement and discussion

- Jacek will be taking on broader project management responsibilities for entire DM
- Fritz to back-fill management for qserv team
- K-T returning from interim Project Engineer to system architect role

Project planning

- Looking like we will insert a 3-month cycle to break alignment between end-of-cycle and LSST project meetings
- Review of Jacek's preliminary resource loading matrix for next (3 month) cycle – no objections offhand
- We should elaborate epics involved in this 3-month cycle (e.g. large-results) with as much detail as is currently known
- Jacek would like to institute cross-team end-of-sprint demos/reports
- Jacek concerned about generic, under-specified epic and story definitions – please focus and drive toward more specific definitions

Shared-scan scheduler

- Design and implementation changes are in progress subsequent to clarification of multiple-DR requirements last week. John has emailed the group with some design ideas seeking feedback. Scan balancing and prioritization with multiple DRs requires a somewhat more complicated design.
- Memory manager "flexi-lock" design needs a revisit to more exactly meet John's needs (design worked and subsequently agreed-upon during meeting, Andy will implement.)
- Next steps
  - slot in "MemManReal" memory manager implementation which has been provided by Andy
  - update css in qserv_testdata with new shared scan metadata to get integration test coverage
  - spin up the new implementation on in2p3 cluster and begin to test

Secondary Index prototyping

- Bulk update (sorted interleaved insert/update) experiments running now. Mike was getting slower results than expected (turned out to be mistaken use of LOAD LOCAL, server-side loading subsequently during meeting looked much better). Jacek recommends looking into inno-db "reserve" options to pre-flight the load to help performance.

Thread ID

- LWP-based utility provided by Andy
- Would be nice to roll into Log package so we don't have to elaborate every log point in qserv code.
- Ideas developed later in meeting to use TLS ctor, enable via call from C++ code so not enabled all the time for all Log clients. Andy S will handle.

PanStarrs

- NCSA resources as currently specified will only support non-replicated, single-copy data
- Should have sufficient resources at IN2P3 to experiment with replicated data

IN2P3

- Dell willing to provide them with new nvm-express storage for experimentation (like ssd, but directly on pic-express without intervening controller). Might be interesting to try this out with secondary index. Fabrice to ask for one machine with 2TB of the new tech to experiment.
- 25 nodes of cluster have been updated now to Centos 7; we can do the rest after we test them out.
- Docker updates have been requested, but may take a week or two to get this into puppet scripts and deployed.
- Fabio inquired about cross-datacenter copying via network. Right now plans are not to copy data over network.

Multi-node integration tests on single machine (via docker)

- Vaikunth and Josh have now each run it successfully
- Need to talk to Josh at JTM about integrating this into CI. Tricky with current scripts because they use eups build strategy with "qserv-dev" tag explicitly – maybe use lsstw instead for CI to support multiple branches

Vertical Partitioning

- Test nearly complete. More or less linear result for 10-20 joins. Will run tests for order 50 joins, also test on a different storage engine, then write up tech-note with results.

XSwap

- Decisions made on which info needed from logs
- Vaikunth working with Fabrice to run test queries
- Query-id needed very soon in order to make any use of syslog-gathered info

Sizing model update

- Probably falls to Fritz' plate from Jacek because Jacek won't have time to drive it
- Fritz okay to drive issue administratively, but will need technical help because outside areas of expertise
- Andy H. recommends we include requirement for 40G link per node
- K-T informs next official refresh is scheduled for Aug., but warns that software and data center designs continue to be made based on current known-obsolete models