

Database Meeting 2017-01-10

Date

10 Jan 2017

Attendees

Local: [Kenny Lo](#), [Brian Van Klaveren](#), [Unknown User \(npease\)](#), [Igor Gaponenko](#), [Andy Salnikov](#), [Vaikunth Thukral](#), [Andy Hanushevsky](#)

Remote: [John Gates](#)

Discussion items

General:

- Team has started attending CS245 (MW 1:50pm to 2:30pm). Although Vaikunth emailed prof Nate spoke with him after class and he was ok with our presence
 - Slides are usually put up after class, prof recommends reading DB book for lectures (Fritz has it currently)
- RFC 107: Team consensus is that since we don't have a lot of python code this change will be ok especially if there is a tool that will wrap the docstrings in older code
 - Some concern over who will actually do the work for older code
- RFC 245: Right now we use debug mostly everywhere, what to do for production? ok to do WARN for now, INFO ok for production?
 - Other comments already on RFC reason well with our requirements
 - AndyS may have some thoughts he will leave on the RFC
 - Need to internally decide which messages need to go on which level
 - The RFC issue may not be sufficiently well defined yet to answer these questions
- LSST cybersecurity course expired over the break for Vaikunth and Igor, AndyS didn't get it at all - email NCSA?

John:

- DM-8226: Still having docker pull issues, Vaikunth will help debug it, need to test this on NCSA to be able to merge
- DM-8426: ORDER BY needs to match output column in the result table, will work on that, major thoughts listed out on JIRA ticket

Kenny:

- DM-8286, DM-8234: making branches for both and starting work on it, still some machine/helpdesk issues
 - Planning meetings with John etc about these Imgserv tickets
- Need Parallels license (and VMWare license for John)
- Thinking about butler, have high level thoughts for future in agreement with current understanding

Brian:

- DM-8665: Can now extract column/tables with queries. Still some bugs in parser, can't parse some types of queries
 - Want to fix this before finishing or probably make "if clause" for "can parse OK else NOT", can probably handle most SUIT stuff
 - Proof of concept is sufficiently complete

Nate:

- DM-8125: Working on it and trying to maintain backward compatibility, not sure if it is possible and probably needs further discussion

Igor:

- Finishing up on LSST20 dataset for KPMs - 82 partitions around the equator have discrepancies but can probably drop this issue since it no longer affects new data
- After loading is finished will test a bit then replicate on John's side of the cluster. Vaikunth is certain he won't restart the container in the middle of loading now
- Secondary index might take a while to populate on this data (~4billion rows and InnoDB)
- Finishing up the documentation for DM-8406
- Gregory wants us to generate even bigger dataset from WISE (DM-8271 in next sprint)

AndyS:

- Supertask is on the backburner for now
- Running L1DB-proto tests on the super machine at in2p3 with different configs, takes time because tests need large amount of data
- Baseline config is likely not enough for large scale operation, considering splitting it into different aspects

AndyH:

- Transitioning to xrdssi v2, documentation is a pain
- AndyH wants to limit size of a request (set to unlimited right now), wants to discuss again with team
- Cannot optimize if request size isn't limited, MIN should be 2GB can probably do MAX up to 16GB
- Currently queries cannot be repeated after 16giga queries due to counter, will wrap beyond that
 - Need it for new czar to work with old worker xrdssi - can have switch for old and new type during transition
- Limited availability in Jan - Gone to CERN next week for 1 week, then San Diego after that, then jury duty last week of Jan

Vaikunth:

- General issues at in2p3 like QTP index not working anymore
- Locked memory also coming from mmfsd(gpfsd, ~2GB) and collectd (~32KB) itself, hence non-zero when not running qserv locked tables
 - Can we take away the gpfsd locked memory?
 - Does it work with the containers on all the machines. is the 2GB recorded across all containers or across the machine
- KPM measurements will be looked at after LSST20 dataset is done loading
- Updates from Jeff Tseng @ Oxford:
 - Making progress on Edinburgh people, collecting info from colleagues about use cases
 - Checking SDSS help pages, old examples of science queries
 - Talking to people to get more manpower for this work
 - Pan-STARRS made data public - Brian says DESC is aware of this so nothing to do about that yet