# Database Meeting 2022-08-24

### Date

24 Aug 2022

### Attendees

• Igor Gaponenko Fritz Mueller Fabrice Jammes Andy Salnikov Joanne Bogart Andy Hanushevsky John Gates

## Notes from the previous meeting:

• Database Meeting 2022-08-17

### Discussion items

Discussed	Item	Who	Notes
•	Project news	Fritz Mueller	<ul> <li>the last DMLT was focused on the upcoming reviews</li> <li>another "hot" topic is the DM all-hands meeting in Chile (preliminarily March 13-17, 2023):         <ul> <li>Yusra sent around the questionary for those who may be interested in participating (online versus in-person, arriving the weekend before or staying the weekend after)</li> <li>It's formally for DM "construction" (not for "operation") which might be an issue for some of us.</li> </ul> </li> </ul>
•	NCSA to SDF (S3DF) migrati on	Igor Gapon enko Fr itz Mueller	Igor Gaponenko on the status of the test Qserv instance:  • got 6 nodes for Qserv, finalized the configuration • unpacked and deployed a snapshot of Qserv instance large6 that was taken at NCSA around May 17th • the snapshot does not include DP02 (the catalog would need to be ingested) • Qserv hasn't been started as I need to do some work on the tooling
			Fritz Mueller there is an interest to set the TAP service at RSP



Status of the Qserv integrat ion tests

team

(techni discuss ion)

Igor Gaponenko on the current status:

- a collection of tables (5 databases) within Qserv source tree within itest\_src
- a purpose of many queries is not well understood (or documented) with documentation links pointing to TRAC
  - Fritz Mueller we may still have the original TRAC pages migrated to Confluence
  - o Action item: need to document each query
- · some queries are meant to test the non-existing functionality of Qserv (some sort of the "wish list" for future improvements?)
  - Fritz Mueller those might be based on the initial survey of what functionality was expected from Qserv
- some may exist for testing the home-grown SQL parser (before migrating to ANTLR4
- o Fritz Mueller some might be added as bugs were discovered in the parser, or for bugs in the query rewriter tests
- about 50% of the test queries are presently disabled (marked as FIXME, etc.)
- some of those (disabled tests) are needed to cover the current functionality of Qserv
- some might be disabled when migrating the tests to the new lite container or because the required functionality wasn't present in the Replication/Ingest system at the time of the migration
- there is quite a bit of duplication between the tests (and catalogs)
- some data ( CSV ) files are compressed, while others aren't. It's not clear why and what it's meant to test.
- only 3 (out f 5) catalogs are presently tested

#### Conclusions:

- it's a bit of a mess in there
- · Qserv coverage is not complete (or excessive) in some areas
- the BIGGEST problem (for myself) was with using very specific table names that imply certain semantics in the context of LSST ("Object", "Source", etc.). Although the initial motivation behind that decision is clear, this naming convention is presenting a big obstacle in understanding what is actually being tested in Qserv. The semantics of some LSST tables has been changed since the original Data Model.

#### The proposal to be discussed:

- · revisit the test cases
- eliminate duplicates and obsolete tests
- add in the missing tests
- come up with the Qserv-specific naming convention for the tables to reflect their role within Qserv ("director", "child", "refmatch", "fully-replicated", etc.)
- refine the table schemas (and data) to leave only the essential columns (required for Qserv and for the referential integrity of the schemas), and add a few of the "payload" columns as needed where tests require row selection based on those values (shared scans, testing the WHERE clause etc.)

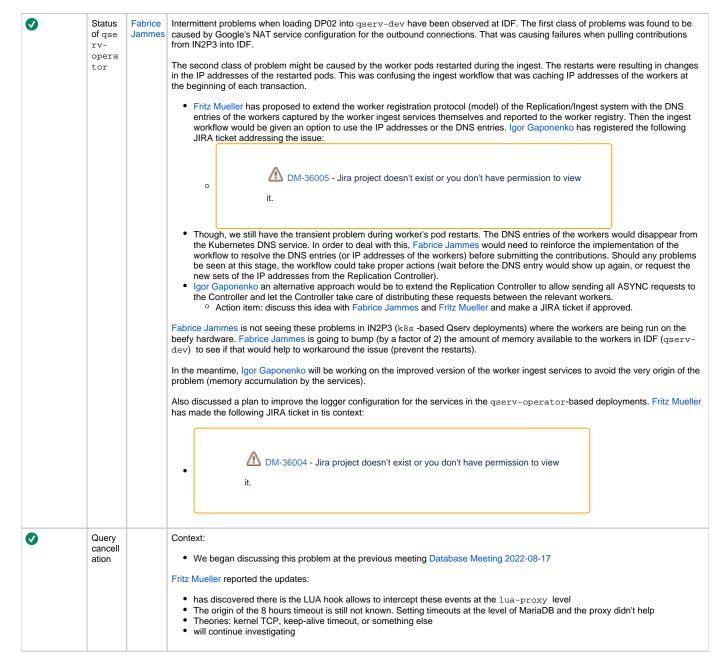
#### Fritz Mueller:

- · some tables should carry the semantics (time series queries, etc.). So we do need a way to keep the semantics (at least) for
- we could document those within the source tree using RST
- · we need to do a systematic revisit of the tests to see what's missing

Fritz Mueller Add the micro dataset to be automatically deployed with Qserv before running any integration tests. This dataset could be used for basic (interactive?) testing of Qserv after it gets deployed.

Igor Gaponenko proposed to implement the synthetic dataset generator (driven by YAML) as an alternative for the present collection of static test catalogs. The dataset would be generated by the Python script at the run-time of the integration test. Or, it could be pregenerated if needed. This option allows generating catalogs of any scale (number of databases, tables, columns in the tables, the number of rows) Also:

- the test queries could be generated by the script accordingly
- this technique could be also used for the small-to-mid term scalability & performance testings of Qserv



## Action items