Database Meeting 2022-05-18

Date

18 May 2022

Attendees

- Igor Gaponenko John Gates Andy Salnikov Unknown User (npease) Fabrice Jammes Fritz Mueller Andy Hanushevsky Colin Slater

Discussion items

Discussed	Item	Who	Notes
✅	Project news	Fritz Mueller Colin Slater	PCW: please, get registered. The registration page: https://project.lsst.org/meetings/rubin2022/register An idea of having a DM meeting in Chile is being discussed by the management. No specific info on the date or the probability of this venue yet.
	Progress on topics discussed at the previous meeting Database Meeting 2022-05-11	team	
✅	DP02	Igor Gaponenko Colin Slater Andy Hanushevsky	Igor Gaponenko: to report on the status and obstacles - it's still not clear to me how to treat correctly the empty (NULL) values in the Parquet files. There was a discussion on this subject last night on the team's Slack channel: https://lsstc.slack.com/archives/G2JPZ3GC8/p1652852985598069 Andy Hanushevsky : commented on pq2csv: - The Parquet files have the schema, including the NULL requirements for select columns - Apparently, there is a bug in the tool that needs to be investigated - Error reporting for incorrect input schema needs to be improved to report the incorrectly configured columns Igor Gaponenko will provide Andy Hanushevsky with a sample Parquet file (DP02 table Source) and the schema file for further investigation Andy Hanushevsky: the NULL option for the columns needs to be documented in the LSST documentation for schemas Fritz Mueller: how are we going to ingest the catalog into qserv-prod? - Igor Gaponenko: it will take a few hours after all intermediate data products (CSV files will be ready and deployed). Fritz Mueller : on geometry optimization (table ObsCore) - this has been tested and working. The tables can be ingested into Qserv, and Qserv can process the optimized queries. - proposed changes for Qserv query analyzer/optimized treating ADQL queries involving these tables
⛔ Moved into the Slack forums.	Development & Qserv infrastructure at SLAC	Fritz Mueller Igor Gaponenko	Igor Gaponenko: to report on the last round of negotiation with Yee - preliminarily agreement on the following configuration of workers: 2U server, single-socket CPU 32C/64HT, 256 GB, 24 disk slots, 12 slots will be populated with NVMe drives of 3.5 GB each. So the total amount of the *raw* storage per worker would be **42 TB**. The *usable* amount after setting up RAID would be 2 drives less. Fritz Mueller: any news on the development machines?
✅	Status of qserv-operator and qserv-ingest	Unknown User (npease) Fabrice Jammes Fritz Mueller	Why two versions of the integration tests? - The qserv-ingest is meant for testing the Kubernetes-based ingest framework. At the end of the ingest workflow the result validation test based on DB Bench is being run. It's used for various purposes, including performance. - The qserv itest is meant for testing Qserv in the Docker compose environment. Are we going to merge both packages into qserv? What implications of this move we should expect?

✅	Improving error reporting in Qserv	Igor Gaponenko John Gates	Context: - ⚠️ DM-34464 - Jira project doesn't exist or you don't have permission to view it. - There are still uncertainties about what needs to be retained in the persistent state (new table `qserv meta.QMessages`. How many details we should retain? - What needs to be done in the long term to improve error reporting to users and to make it usable for Qserv data administrators? Fritz Mueller is going to look for the use of the column `code` by `mysql-proxy`. If it's not used then we should eliminate it.
✅	Partitioning *director* and *child* tables	team	Context: - there is a little bit of confusion on the *overlap* tables of the *child* tables Igor Gaponenko will run extensive tests and report results on the Confluence page (TBC)
⛔ Postponed till next week	Qserv in IDF fails to lock tables in the memory	team	Context: - there was a discussion on the team's Slack channel: https://lsstc.slack.com/archives/G2JPZ3GC8/p1652497266243519 - the problem may affect the *aggregate* performance of the *shared scan* queries

Action items

- ☐