TABLE 1: Size summary based on LDM-141

Table	Bytes/row	Rows (DR1 -> DR11)	DR1 (TB)	× Growth	DR10 (PB)
Object_Lite	1840	$2.26^{10} - > 4.74^{10}$	42	2.1	0.08
Object_Extra	20393	$2.26^{10} - > 4.74^{10}$	461	2.1	0.9
Source	453	$4.51^{11} - > 9.01^{12}$	204	20.0	4.0
ForcedSrc	41	$1.20^{12} - > 5.01^{13}$	49	42	2.0
DiaObject	1405	$7.94^{08} - > 1.54^{10}$	1.1	19.4	0.002
DiaSource	417	$2.26^{09} - > 4.52^{10}$	0.9	20	0.002
DiaForcedSource	49	$1.50^{10} - > 3.01^{11}$	0.7	20	0.001
Year 1 raw images:$3PB$, tables:$\sim 1PB$, half for Object_Extra,$0.2PB$ Sources					
Year 10 raw images:$30PB$, tables:$\sim 7PB$,$4PB$ Sources,$2.0PB$ Forced ,$1PB$ Object_Extra					

Discussion of Object vs. Source table queries and data distribution

William O'Mullane, Fritz Mueller

2019-07-02

1 Introduction

Writing concise and testable requirements is very difficult. Writing requirements in 2005 for a system to run in 2022 is extremely difficult but is the case for DM. Assumptions are made about requirements and how to implement them, but the perspective of the requirement writer and implementer are usually not identical. Over a long period this could diverge significantly and choices made years ago may not be so valid anymore. So we should continually challenge requirements and ensure they are still valid and that we are interpreting them correctly.

2 The catalog access question

The DataBase requirements are in LSE-61 and LDM-555. A series of use cases has been collected in DMTN-086. It is interesting to consider the size of tables summarized in Table 1.

It is also worth considering the usage for those tables as in Table 2. **PST should closely examine this table, which came from the AMCL and consider if the numbers are correct/-plausible.**

TABLE 2: Potential/estimated usage of proucts in Table 1 and images, this table came from the AMCL originally.

Data Product	Cardinality	Volumei [PB]	Usuage Frequency	Discovery Potential	Replicas
Object_Lite	40M	0.1	95%	20%	0.08
Object_Extra	40M	0.9	4%	24%	0.9
Source	9T	4.0	0.9%	50%	4.0
ForcedSrc	50T	2.0	0.1%	3%	2.0
Image coadds	55K	0.3	0.01%	2%	0.002
Image raw	5.5M	30.0	0.001%	1%	0.002

2.1 Baseline approach to catalog interaction SQL/ Qserv

Qserv is a custom massively parallel database built by LSST(SLAC) for LSST. This has been built on the assumption (requirement) that astronomy on the catalog will be done as queries. Qserv provides SQL access (with some query limitations) to all catalogs, including visits e.g. force photometry/light curves as depicted in Figure 1. Some implementation remains for Qserv, e.g. MyDB-like functionality, authentication.

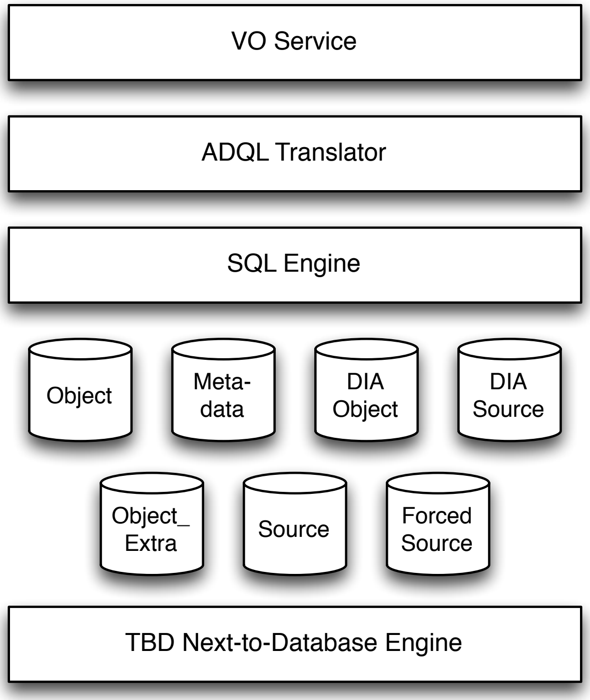

FIGURE 1: Qserv architecture as baselined in LDM-135 and currently being implemented
.

2.2 Heterogeneous Data Access

An alternative heterogeneous approach has been proposed which could be followed if we assume that most SQL like queries would be on the Object catalog, or perhaps even the Object-Lite catalog. Then the Source and other large tables would be stored in something like Parquet files, and accessed with one of the map reduce type systems such as Spark or Dask. This is depicted in Figure 2.

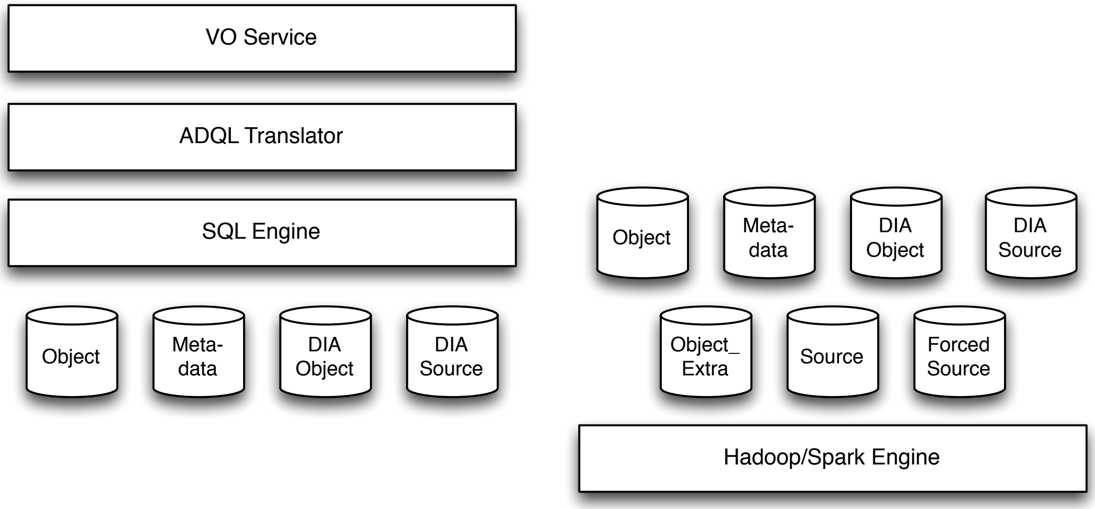

FIGURE 2: Heterogeneous catalog access with SQL for the Object catalog but files only for Sources

We have to be clear SQL would then be restricted to what could fit in a DB (petabyte DBs are possible today.) There are a few astronomy oriented database implementations which could handle the 40TB Object Lite catalog e.g. SqlServer or Postgres. These also have implementations for MyDB and VO protocols like TAP.

MyDB would then be MyDB .. not in Spark[1] or Dask[2] - it would be query-able of course.

There may be a higher cost on hardware for this approach, as it will be less efficient for supporting concurrent access to the source table by more than a few users at a time.

[1] https://spark.apache.org/
[2] https://dask.org/

3 Recent Developments

Since the baseline plan has not to date been altered, Qserv development has continued per plan. A couple of recent developments are worthy of note, however:

- The DRP pipelines group has recently adopted Parquet as the file format on record for intermediate data products.
- Preliminary response from reviewers and early adopters of the LSP has been overwhelmingly positive with respect to the possibility of using the Dask framework for notebook-based analysis of catalog data products. This, in combination with the above item, make it seem both desirable and likely that DRP products be made available in Parquet format in addition to any other plans.
- Investigations into implementation strategies for the baselined (but very underspecified) ”next-to-db” functionality have yielded the conclusion that users want and expect a level of sophistication in non-SQL / programmatic access to data products that already exists in popular off-the-shelf solutions but would be difficult to achieve in a project-built, tightly integrated solution.
- A proof-of-concept collaborative exploration with Google was conducted [DMTN-078], in which Qserv was demonstrated to be both practically field-able and reasonably performant on Google cloud infrastructure. During this exploration, ”shoot-outs” were also conducted between Qserv and Google BigQuery. The results were a mixed bag; each technology out-performing the other on certain types of queries in terms of cost and performance [Document-31100]. There is apparently yet still some life to be had from an MPP database system.

4 Conclusion

The above developments taken together would seem to indicate that DM will need to field something architecturally similar to the heterogeneous design presented above, *regardless* of the choice of particular SQL engine on the left (Qserv or non) and/or the extent of data products to be hosted within (Object/Object-Lite vs. the whole ball of wax). The right-hand side of the heterogeneous design is, in fact, currently being pursued in the context of LSP/LDF development.

With construction nearing completion, and considering the staffing for Qserv development (3.5 FTE), the *management-only* recommendation would be not to change db horses at this time and to cross the start line into operations according to current plans. If real usage patterns then indicate SQL-oriented access is in fact effectively limited to just the Object table, we should be prepared to swap the database on the left-hand side for something more ideal for operating at that reduced scale; any database hardware resources thus freed could then be absorbed into the computation platform on the right.

A final reason to continue with Qserv is the unique ability to query the Source table - the community have never had such a facility and it may be a route to discovery. if we agree with Table 2 up to 50% of our science potential lies in that table so providing this as an experiment in DR1 would seem a reasonable investment in potential science return.

If we wished to reduce risk, and in the era of in-kind contributions, having a Postgres or other implementation of ObjectLite available with a TAP interface would be a super contribution - this may bring its own interop problems with MyDB etc. of course.

Should we agree with Table 2 the project would benefit greatly from having multiple sources of Object table available from partner institutions. This would allow us to distribute 95% of our query load to other locations and allow scientists the easiest access to the most frequently consulted data. An open data policy would have made this easy to achieve, a data rights driven policy should still consider opening the object catalog up - consider it a sort of advertisement at the cost of perhaps allowing non data rights holders a shot at 20% of the science discoveries.

A References

References

[LDM-555], Becla, J., 2017, *Data Management Database Requirements*, LDM-555, URL https://ls.st/LDM-555

[LDM-141], Becla, J., Lim, K.T., 2013, *Data Management Storage Sizing and I/O Model*, LDM-141, URL https://ls.st/LDM-141

[LDM-135], Becla, J., Wang, D., Monkewitz, S., et al., 2017, *Data Management Database Design*, LDM-135, URL https://ls.st/LDM-135

[LSE-61], Dubois-Felsmann, G., Jenness, T., 2018, *LSST Data Management Subsystem Requirements*, LSE-61, URL https://ls.st/LSE-61

[DMTN-078], O'Mullane, W., Swinbank, J., Gelemann, M., Wu, X., Mueller, F., 2018, *Potential proofs of concept for Google Cloud*, DMTN-078, URL https://dmtn-078.lsst.io,
LSST Data Management Technical Note

[DMTN-086], Slater, C., 2018, *Next-to-the-Database Processing Use Cases*, DMTN-086, URL https://dmtn-086.lsst.io,
LSST Data Management Technical Note

[Document-31100], Thomson, J.R., 2019, *LSST Benchmarkin of Qserv and BigQuery*, Document-31100, URL https://ls.st/Document-31100

B Acronyms used in this document

Acronym	Description
AMCL	AURA Management Council for LSST
Baseline	The point at which project designs or requirements are considered to be 'frozen' and after which all changes must be traced and approved
Center	An entity managed by AURA that is responsible for execution of a federally funded project
DB	DataBase
DM	Data Management
DMTN	DM Technical Note
DRP	Data Release Production
Data Management	The LSST Subsystem responsible for the Data Management System (DMS), which will capture, store, catalog, and serve the LSST dataset to the scientific community and public. The DM team is responsible for the DMS architecture, applications, middleware, infrastructure, algorithms, and Observatory Network Design. DM is a distributed team working at LSST and partner institutions, with the DM Subsystem Manager located at LSST headquarters in Tucson.
Document	Any object (in any application supported by DocuShare or design archives such as PDMWorks or GIT) that supports project management or records milestones and deliverables of the LSST Project

FTE	Full Time Equivalent
Handle	The unique identifier assigned to a document uploaded to DocuShare
LDF	LSST Data Facility
LDM	LSST Data Management (Document Handle)
LSE	LSST Systems Engineering (Document Handle)
LSP	LSST Science Platform
LSST	Large Synoptic Survey Telescope
MPP	Massively Parallel Process
Object	In LSST nomenclature this refers to an astronomical object, such as a star, galaxy, or other physical entity. E.g., comets, asteroids are also Objects but typically called a Moving Object or a Solar System Object (SSObject). One of the DRP data products is a table of Objects detected by LSST which can be static, or change brightness or position with time.
PST	Project Science Team
PSTN	Project Science Technical Note
Project Science Team	an operational unit within LSST that carries out specific scientific performance investigations as prioritized by the Director, the Project Manager, and the Project Scientist. Its membership includes key scientists on the Project who provide specific necessary expertise. The Project Science Team provides required scientific input on critical technical decisions as the project construction proceeds
Qserv	Proprietary Database built by SLAC for LSST
SLAC	No longer an acronym; formerly Stanford Linear Accelerator Center
SQL	Structured Query Language
Source	A single detection of an astrophysical object in an image, the characteristics for which are stored in the Source Catalog of the DRP database. The association of Sources that are non-moving lead to Objects; the association of moving Sources leads to Solar System Objects. (Note that in non-LSST usage ”source” is often used for what LSST calls an Object.)
TAP	Table Access Protocol
VO	Virtual Observatory