LARGE SYNOPTIC SURVEY TELESCOPE

Large Synoptic Survey Telescope (LSST) Data Management

Data Backbone Design

Michelle Gower and Kian-Tat Lim

DMTN-122

Latest Revision: 2019-07-14

Abstract

The Data Backbone (DBB) is a key component that provides for data storage, transport, and replication, allowing data products to move between enclaves. This service provides policy-based replication of files (including images and flat files to be loaded into databases as well as other raw and intermediate files) and databases (including metadata about files as well as other miscellaneous databases but not including the large Data Release catalogs) across multiple physical locations, including the Base, Commissioning Cluster, NCSA, and Data Access Centers. It manages caches of files at each endpoint as well as persistence to long-term archival storage (e.g. tape). It provides a registration mechanism for new datasets and database entries and a retrieval mechanism compatible with the Data Butler.

Change Record

Version	Date	Description	Owner name
1	2019-07-08	Initial version extracted from LDM-152.	Kian-Tat Lim

Document curator: Michelle Gower

Document source location: `https://github.com/lsst-dm/dmtn-122`

Contents

Data Backbone Design

1 Introduction

The Data Backbone (DBB) provides data management (storage, tracking, and replication) for LSST data products that reside in non-computational storage tiers. It provides policy-defined operations for intra-site and inter-site data distribution, access latency requirements dependent upon the lifecycle of the data, data protection and recovery, data retention and eviction given cadences and predicted and observed on-demand usage, and efficient bulk recall of data which are spatially, temporally, or otherwise related in support of processing or export.

This document describes the baseline design of the LSST Data Backbone (DBB), including the following components:

- Data Backbone Ingest/Metadata Management

- Data Backbone Lifetime Management

- Data Backbone Transport/Replication/Backup

- Data Backbone Storage

The relationships between the Data Backbone components are illustrated in Figure 1.

Requirements for the Data Backbone are defined in LDM-635.

The Data Backbone serves as an abstraction against storage components whose similar requirements may otherwise have been met through duplication of effort during construction and operations. From the perspective of data producers and consumers, the Data Backbone provides a common, well-defined concept of how all participating storage components are expected to function, which greatly simplifies understanding of the system as opposed to understanding differing personalities of a large number of components.

The Data Backbone links all of the computational enclaves and the Data Access Centers, acting as the spine that supports them all.

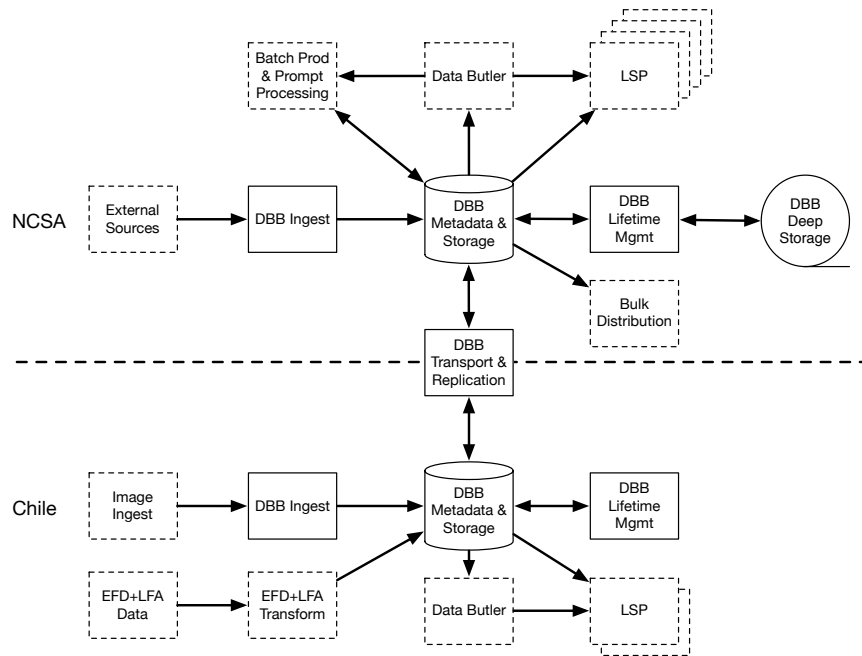Rucio (Rucio) is being considered as an overall DBB system for files.

FIGURE 1: Data Backbone

2   Replication and Transport

The Data Backbone spans the Base Site, Archive Site, all Data Access Centers and all sites participating in annual data release processing. Data products can enter the Data Backbone at any location as permitted by policy and are subject to timely distribution, access-latency guarantees, and eviction as defined by the policy. The Data Backbone provides data protection of data products while resident with the backbone.

Movement operations supported by the Data Backbone include:

- Staging: data is copied in and out of the Data Backbone in coordination with an external management system; primarily used for workflow orchestration.

- Caching: data location and lifetime is managed by policy; caches are populated by use.

- Mirroring: 1:1 replication of data is dictated by policy.

- Export: data is copied out of the Data Backbone to end users;

The Data Backbone does not include public access restrictions, responsibilities regarding proprietary data periods or users with data access rights, or authorization or authentication of

external users or services. This functionality is provided by layers on top of the Data Backbone, in the LSST Science Platform (LSP), Identity Management, and Bulk Distribution components.

Tiers within the Data Backbone include a custodial store with assurance of data preservation (e.g. tape but possibly other technologies) and an access tier that may have lower latency.

Replication between sites and transfer to the custodial store is currently envisioned to be handled by layered utilities, so the DBB does not necessarily present a single-filesystem view. File transport technologies such as Globus Transfer (Globus) with GridFTP and RESTful interfaces are being considered.

3   Location and Metadata

The Data Backbone tracks the locations of all replicas of data ingested into it, along with their metadata and provenance. This information is stored in global, replicated database tables within the Consolidated Database instances.

4   Files

The Data Backbone holds all files that are part of the Science Image Archive, including raw data and processed data products, as well as additional files such as the Engineering and Facilities Database Large File Annex, files associated with the Calibration Database, etc. It is also currently envisioned to contain files representing the archival contents of the catalog data products, either canonical files that are ingested into database servers or backup files dumped from canonical databases.

These files will be kept on a high-performance, scalable file store and archived in a reliable long-term file store. The baseline design uses GPFS (GPFS) and HPSS (HPSS), but drawbacks to these have been identified. Investigations of alternate storage technologies such as object stores (including Amazon Glacier (Amazon)), Campaign Storage (Campaign Storage), and Quobyte (Quobyte) have been performed, but future work remains in this area.

5   Databases

The Data Backbone holds the content for all databases that are part of the Science Catalog Archive that is visible to data rights holders. It is also responsible for maintaining many of the database instances that serve this data. Exceptions are currently envisioned to include the large-scale Data Release catalogs that are loaded only into the Qserv software, described separately in LDM-135), which are considered as outside the Data Backbone. The associated metadata will be part of the Data Backbone, even if it is also replicated into the Qserv software. The Calibration Database, the reformatted Engineering and Facility Database, and the (external-facing) Prompt Products Database are all currently envisioned tto be part of the Data Backbone.

Just like files, these databases need to be managed in terms of replication, disaster recovery, and lifetime. The underlying mechanisms for data storage and transport and the interfaces to the data are significantly different, however. Accordingly, all databases are stored in appropriate database management systems that provide their own native mechanisms for replication and backup. These include the Consolidated Database implemented in Oracle (Oracle), which is a large instance of such a database server containing science data, metadata, provenance, and production tracking information, particularly for Data Release Production. Additional databases such as the internal, Alert Production-only database containing DIAObjects, DIASources, and DIAForcedSources; tracking information for the Image Ingest and Processing system; or measurements for the Quality Control systems may reside in their own specialized storage outside the Data Backbone..

5.1   Interfaces

The primary access to the Data Backbone files is currently envisioned to be via mounted POSIX filesystems, but additional web service methods for retrieving files are contemplated.

The primary access to the Data Backbone databases is via direct queries to the relevant database servers. For portability and future-proofing, such queries should typically be intermediated with a package such as SQLAlchemy so that the underlying database software can be changed.

An ingest utility will transfer files into the Data Backbone and load their metadata and provenance into the appropriate DBB database tables. No file will be made available to be retrieved before its contents and metadata are complete. (It may be possible for file metadata to exist without file content if the fact that content does not yet exist or will never exist is recorded

and made available along with the metadata.)

Since the DBB file metadata tables are large, complex, and hosted on a consolidated database server instance, their service level is not currently envisioned to be guaranteed to be high enough to directly support observatory operational processes. Accordingly, access to raw data from the LSSTCam, ComCam, and Auxiliary Telescope Spectrograph for use by observatory systems is provided by the Observatory Operations Data Service, which can maintain a higher level of availability.

LSP instances, including Portal, JupyterLab, and Web API nodes, will have access to DBB filesystems (either through mounts or through web services) and direct client access to DBB databases.

Since Science Pipelines codes will need to run in multiple environments including both batch worker nodes and LSP instances but also including development environments, they use the Data Butler [LDM-135] to isolate themselves from the details of access interfaces.

A   References

References

Amazon, Amazon Glacier – Cloud Archive, URL `https://aws.amazon.com/glacier/`

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

Campaign Storage, Campaign Storage, URL `http://campaignstorage.com/`

Globus, Globus Transfer API Documentation, URL `https://docs.globus.org/api/transfer/`

[LDM-635], Gower, M., Butler, M., Lim, K.T., 2018, *Data Management Data Backbone Services Requirements*, LDM-635, URL `https://ls.st/LDM-635`

GPFS, IBM Spectrum Scale, URL `https://www.ibm.com/us-en/marketplace/scale-out-file-and-object-storage`

HPSS, HPSS – High Performance Storage Systems, URL `http://hpss-collaboration.org/`

Oracle, Oracle – Database 12c, URL `https://www.oracle.com/database/index.html`

Quobyte, Quobyte – Data Center File System, URL `https://www.quobyte.com/`

Rucio, Rucio Distributed Data Management Documentation, URL `http://rucio.cern.ch/`

B Acronyms