LARGE SYNOPTIC SURVEY TELESCOPE

Large Synoptic Survey Telescope (LSST)

LDF File Systems Baseline Overview

Don Petravick

DMTN-051

Latest Revision: 2017-08-03

Abstract

This is a descriptive and explanatory document, not a normative document. This document explains the proposed baseline as presented in the DM replan in July, 2017, referred to just "baseline" in the prose that follows.

Change Record

Version	Date	Description	Owner name
	2017-08-03	Initial draft.	Don Petravick
1.0	2017-08-09	Convert to Latex as DMTN-051.	Tim Jenness

Document source location: `https://github.com/lsst-dm/dmtn-051`

Contents

LDF File Systems Baseline Overview

1 Overview

The baseline for the LSST provides for two types of file systems – "site" file systems, which are local to LDF systems at the Base or at LDF systems at NCSA, and "data backbone file systems" that hold the permanent record of data facility operations.

In the baseline, all file systems described below are POSIX file systems. A subset of these file systems are backed up, as indicated.

2 Site File Systems

File System	Where	Backed up?
Login	NCSA	Yes
Project	NCSA	No
Scratch	NCSA	No
PDAC-user	NCSA	Yes
USDAC-User	NCSA	Yes
Backup	NCSA	n/a
Datasets	NCSA	Yes
Kubernetes containers	NCSA	Yes
Production login	NCSA	Yes
L1 Alert Log File system	NCSA	No
L1 Input/Output Cache	NCSA	No
Template Cache	NCSA	No
Login	Base	Yes
Project	Base	No
Scratch	Base	No
Chilean DAC-USER	Base	Yes
L1 application Support	Base	Yes
Observatory Operations File System	Base	No
Datasets	Base	Yes
Kubernetes containers	Base	Yes
Backup	Base	n/a

Login file systems serve users at each site, and are backed up.

Project file systems provide a place for output of informal, development-oriented outputs of running codes. These are not backed up, on the assumption that data here can be regenerated.

Scratch file systems hold temporary program output files, and are subject to advertised purge policies. This distinguishes scratch file systems from Project file systems, which are not purged by system administrator applying a policy.

Dataset file systems hold designated datasets that supports development. Designated datasets are backed up, whilst some datasets are easily obtain-able from the internet, and do not require a backup. However, files in the dataset file systems are not considered to be part of the formal record of the survey. Any files that would become part of the formal system would be replicated and kept in the data backbone, discussed below.

The DAC file systems (PDAC, USDAC, CHILEANDAC) hold user-created files for the corresponding data access centers. The file systems need to be backed up.

Backup file systems provide for copies of recent data. Backup file systems have independent implementations from the systems they backup. Backup file systems have a role as secondary disaster recovery data recovery resources, but are not primary sources of storage for disaster recovery.

The template Cache, L1 Input/Output file system, and L1 Alert Log play a special role in supporting prompt processing at NCSA.

The L1 application support file system supports LDF L1 services at the base.

The Observatory Operations file server is an intermediate file system at the Base. The spectrograph, Comcam and LSSTCam archivers write to this file system, which is exported to designated observatory computers, including the commissioning cluster. Data from this file system is further ingested into the raw file partition of the data backbone.

3 Data Backbone File Systems

Data backbone file systems [LDM-230] have these main distinguishing characteristics.

- These file systems hold the permanent file-based record of the survey.

- Data on these file systems are synchronized between the base and NCSA sites according to policy.

- These file systems are subject to change-controlled disaster recovery plans.

- All files are supported by location, meta-data, and provenance database entries.

- The integrity of these data residing on these file systems are routinely monitored using cataloged meta-data.

- Data are only entered into these file systems by processes run in change-controlled production processes.

- File systems are presented as POSIX file systems.

- Designated data may only be accessed by access methods which stage data from tape.

- Only data from data backbone file systems are used to create formal production products.

- Only data from data backbone file systems are used to produce file based products specified in the DPDD.

The baseline provides for the following file systems:

- The **Raw** file system holds pixel data from the spectrograph, COMcam, and LSSTcam, and files ingested from the Engineering and Facility Database large file annex [LTS-210].

- The **Calibrations** file system holds computed calibration products. Raw data leading to calibrations, such as raw flat and bias frames are held in the raw file system.

- The **Production Products** file system contains files other than calibration file products produced with the formal production system.

- The **Computed Engineering Products** file system holds files related to the miscellaneous production processes, not covered above.

References

[1] **[LSE-163]**, Jurić, M., et al., 2017, *LSST Data Products Definition Document*, LSE-163, URL https://ls.st/LSE-163

[2] **[LTS-210]**, Mills, D., 2015, *Engineering and Facility Database Design Document*, LTS-210, URL https://ls.st/LTS-210

[3] **[LDM-230]**, Petravick, D., Gelman, M., 2017, *Concept of Operations for the LSST Data Facility Services*, LDM-230, URL https://ls.st/LDM-230