Highlights of Astronomy, Volume 16
XXVIIIth IAU General Assembly, August 2012
T. Montmerle, ed.

doi:10.1017/S174392131401285X

LSST Data Management: Entering the Era of Petascale Optical Astronomy

Mario Juric[1] and Tony Tyson[2]

[1]Large Synoptic Survey Telescope,
950 North Cherry Ave, Tucson, AZ 85719,
email: mjuric@lsst.org

[2]University of California, Davis,
Physics Department, University of California, One Shields Avenue, Davis, CA 95616 USA
email: tyson@physics.ucdavis.edu

Abstract. The Large Synoptic Survey Telescope (LSST; Ivezic *et al.* 2008, http://lsst.org) is a planned, large-aperture, wide-field, ground-based telescope that will survey half the sky every few nights in six optical bands from 320 to 1050 nm. It will explore a wide range of astrophysical questions, ranging from discovering killer asteroids, to examining the nature of dark energy. LSST will produce on average 15 terabytes of data per night, yielding an (uncompressed) data set of 200 petabytes at the end of its 10-year mission. Dedicated HPC facilities (with a total of 320 TFLOPS at start, scaling up to 1.7 PFLOPS by the end) will process the image data in near real time, with full-dataset reprocessing on annual scale. The nature, quality, and volume of LSST data will be unprecedented, so the data system design requires petascale storage, terascale computing, and gigascale communications.

1. Introduction

By visiting each patch of 18,000 square degrees of sky over 800 times in pairs of 15 sec exposures, LSST will explore a wide range of astrophysical questions, from studies of the Solar System, to examining the nature of dark energy, and will revolutionize time domain astrophysics. LSST is an integrated survey system. The observatory, telescope, camera and data management systems will be built to conduct the LSST survey and will not support a 'PI mode' in the classical sense. Instead, the ultimate, science-enabling, deliverable of LSST will be the fully reduced data – catalogs and images. The machine learning algorithms developed for automated discovery will find wide application.

2. LSST Data Products

The data management challenge for the LSST Observatory is to provide fully-calibrated public databases to the user community to support the frontier science expected of LSST (LSST Science Book, 2009), while simultaneously enabling new lines of research not anticipated today. The nature, quality, and volume of LSST data will be unprecedented, so the data management system (DMS) design features petascale storage, terascale computing, and gigascale communications. The computational facility and data archives of the LSST DMS will rapidly make it one of the largest and most important facilities of its kind in the world (see Table 1). New algorithms will have to be developed and existing approaches refined in order to take full advantage of this resource, so "plug-in" features in the DMS design and an open data/open source software approach enable both science and technology evolution over the decade-long LSST survey.

Table 1. LSST Data Management Computing System Size

Quantity	Size	Comment
Cumulative Image Archive	345 PB	Total over all Data Releases, including virtual data
Cumulative Catalog Archive	46 PB	Total over all Data Releases, incl. database indices
Final Image Collection	75 PB	In final data release (DR 11), including virtual data
Final Database	32 trillion rows (9 PB)	In final data release (DR 11)
Final Disk Storage	228 PB (3700 drives)	Archive Site only (at NCSA)
Final Tape Storage	83 PB (3800 tapes)	Single copy only
Number of Nodes	1800	Archive Site, includes both compute and database nodes
Alerts Generated	6 billion	Alerts generated over the life of the survey

LSST Data and Computing at a Glance: The sizes of various components of LSST data management systems and data products. "Virtual data" is data that is dynamically recreated on-demand from provenance information.

LSST will deliver or enable three different classes of data products (LSST Science Requirements Document, 2011):

- *Level 1* ("nightly") data products will be generated continuously every observing night and include measurements such as alerts to objects that have changed brightness or position. They will be broadcast world-wide using VO protocols.
- *Level 2* ("yearly") data products will be made available as annual Data Releases and will include images and measurements of quantities such as positions, fluxes, and shapes, as well as variability information such as orbital parameters for moving objects and an appropriate compact description of light curves. The exact contents of Level 2 products will be set by the desire to minimize the necessity to independently reprocess the image data.
- The LSST will enable the creation of *Level 3* ("user-created") data products, by making available approximately 10% of its computing capability to the community. These "community cycles" will be used to perform custom analyses not fully enabled by Level 1/2, taking advantage of co-location of computation with the entire LSST data set.

3. Open Data, Open Source – Community Resource

All LSST research will be done by the community using LSST data products, and not by the LSST Project. In particular, LSST data products, including images and catalogs, will be made available with no proprietary period to the astronomical communities of the United States, Chile, and to LSST's international partners. Transient alerts will be made available for world-wide distribution within 60 seconds, using standard VO protocols.

LSST data processing stack, including the image processing stack, pipeline middleware, and a highly distributed database, will be free software. Highly advanced prototypes are already available at http://dev.lsstcorp.org/cgit. A novel database architecture is being developed, enabling fast efficient search in space and time.

Acknowledgements

The authors would like to acknowledge the support of the National Science Foundation under Grant No. AST-1227061, *"LSST - Concept & Development"*."

References

Ivezic, Z., Tyson, J. A., Acosta, E., *et al.* 2008, arXiv:0805.2366
LSST Science Collaboration, Abell, P. A., Allison, J., *et al.* 2009, arXiv:0912.0201
LSST Science Requirements Document, http://www.lsst.org/files/docs/SRD.pdf