



Large Synoptic Survey Telescope

www.lsst.org

LSST Data Products: Enabling LSST Science

Mario Juric¹, J. Kantor¹, T. S. Axelrod², G. P. Dubois-Felsmann³, J. Becla³, K-T Lim³, R. Lupton⁴, the LSST Project and Science Collaborations

¹ Large Synoptic Survey Telescope ²University of Arizona ³SLAC National Accelerator Laboratory ⁴Princeton University

The Large Synoptic Survey Telescope (LSST; <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, from studies of the Solar System, to examining the nature of dark energy. 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 'PI mode' in the classical sense. Instead, the ultimate, science-enabling, deliverable of LSST will be the fully reduced data. This poster presents the baseline design and contents of LSST data products. There will be three main categories. "Level 1" 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" 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. Finally, approximately 10% of LSST's computing capability will be made available to the community for generation of "Level 3" data products. These will be used to perform custom analyses not fully enabled by Level 1/2, while taking advantage of co-location of computation with the entire LSST data set.

Open Data, Open Source – Community Resource: LSST data products, including images and catalogs, will be available with **no proprietary period** to the astronomical communities of the **United States, Chile, and international partners**. Transient alerts will be **made available for world-wide distribution within 60 seconds**, using standard VO protocols. **LSST data processing software will be open source (GPL v3).**

Scientific research based on LSST data products will be done by the community.

LSST Data Products: LSST is an integrated survey system. The ultimate science-enabling deliverable of the LSST is the fully reduced data. LSST will make available three categories of data products:

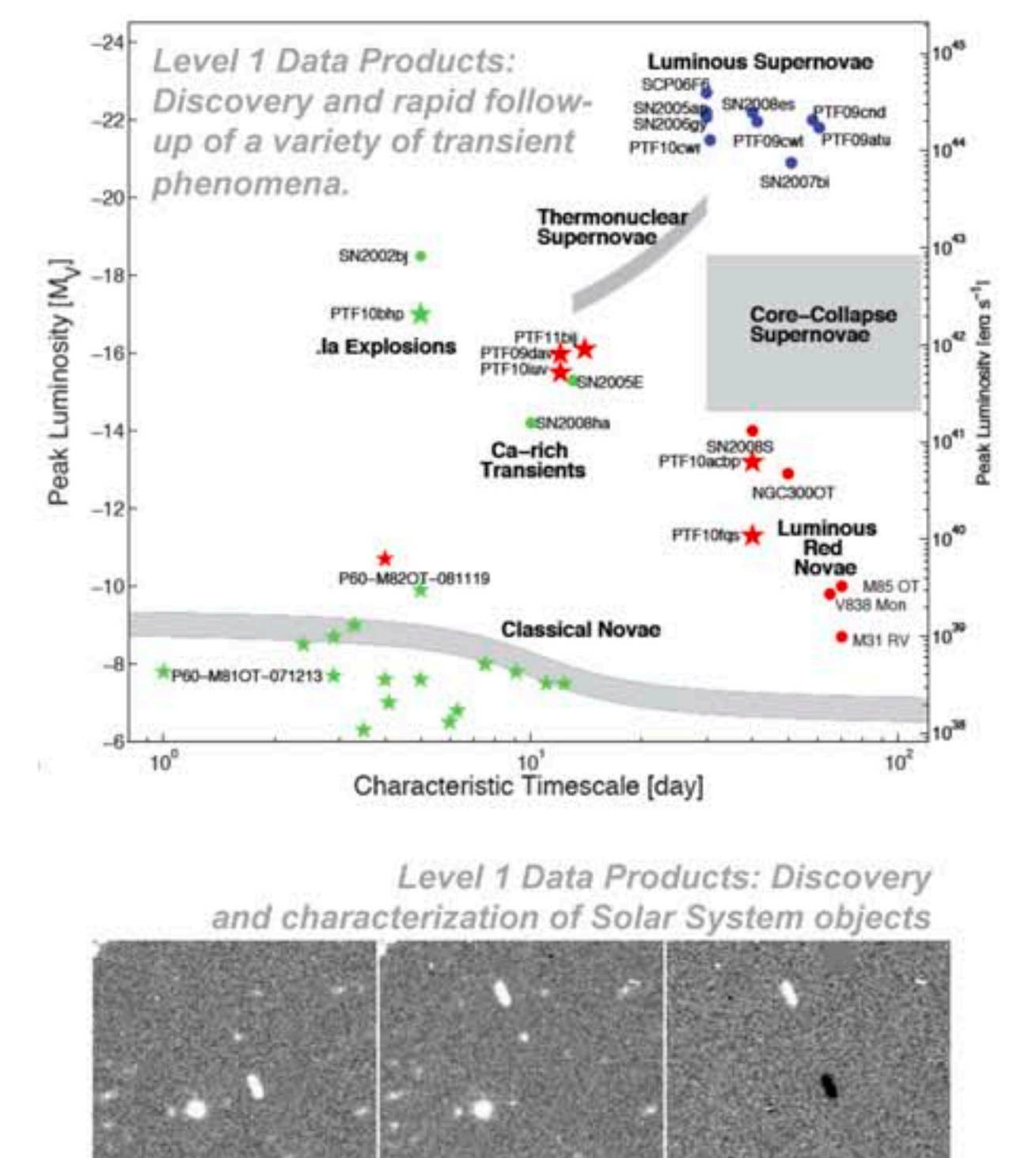
- Level 1 (“nightly”)** data products are generated continuously every observing night, including alerts to objects that have changed brightness or position (excluding the known asteroids).
- Level 2 (“annual”)** data products will be made available as annual **Data Releases** and will include images and measurements of positions, fluxes, and shapes, as well as variability information such as orbital parameters for moving objects and an appropriate compact description of light curves.
- Level 3 (“user-created”)** data products will be created by the community, including project teams, using suitable Applications Programming Interfaces (APIs) that will be provided by the LSST Data Management System. The Data Management System will also provide at least 10% of its total capability for user-dedicated processing and user-dedicated storage. The key aspect of these capabilities is that they will reside “next to” the LSST data, avoiding the latency associated with downloads. They will also allow the science teams to use the database infrastructure to store their results.

Level 1 (nightly) Products: LSST Data Management system will be capable of detecting and issuing 2 million or more alerts per night, for objects detected in images differenced against a deep template. The alerts will be issued in 60 seconds, and made available for broadcast world-wide as VOEvent messages.

For each alert, the following information will be made available:

- Position, flux, size, and shape.
- Prior variability information and variability *characterization* (e.g., low-order light-curve moments), with a “stretch goal” of providing full light curves in all bands. There are no current plans to perform object *classification*.
- Image cut-outs centered on the object.

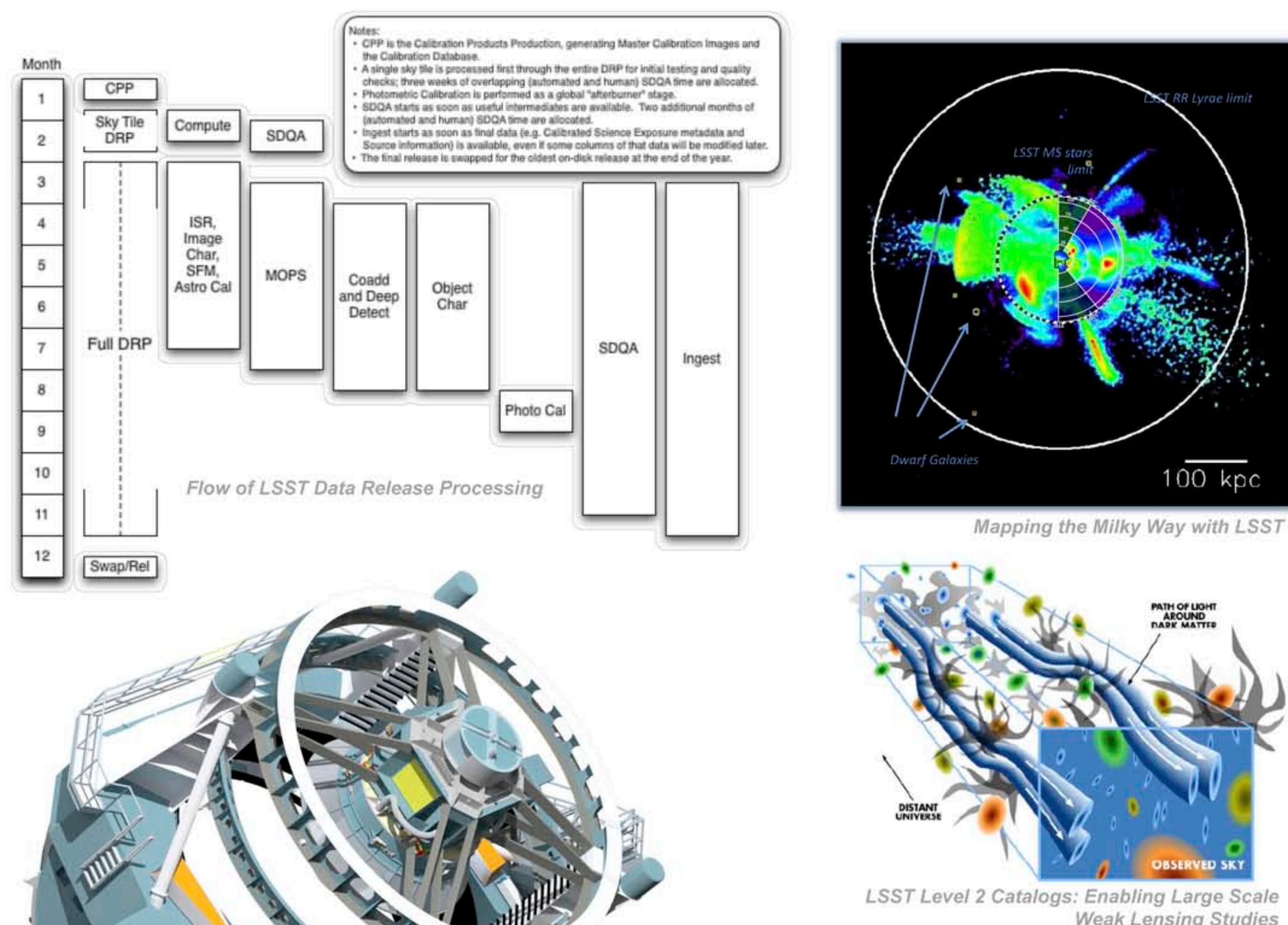
Solar system objects will be detected in difference images as sources with positive flux, and linked together into orbits by the Moving Objects Processing System (MOPS). The orbits will be made available within 24 hours.



Level 2 (annual) Products: These data products, published as **Data Releases (DRs)**, will include **well calibrated and characterized images, associated calibration products, and catalogs**, delivered in an easy-to-use remotely accessible database. They are intended to **facilitate “static sky” science and time domain science for which immediate follow-up is not required**. Two examples – mapping the Milky Way and weak lensing science – are shown in figures below. The quality of measurements delivered in Level 2 will be generally higher than their Level 1 counterparts.

Bottom left diagram shows a typical flow of processing during data release production. Every data release will reprocess all data taken since the start of the survey. Except for the first year when there will be two, **the data releases will be made on a yearly cadence**. There will be no proprietary period; a release will be made available to the community as soon as it's ready.

We expect the catalogs and processed images will be the most requested and used LSST data release products. They will be made available through a web based Science User Interface (SUI), as well as appropriate remote access APIs and protocols (e.g., SIAP, TAP, etc.).



Level 2 Catalogs: Data provided in Level 2 catalogs are designed to facilitate studies requiring a well-calibrated, consistent, wide, and deep data set.

The most important Level 2 database tables are shown on the bottom right panel. The **Object** table keeps the derived information about astronomical objects (e.g., the proper motion of a star), while the **Source** table keeps the individual measurements (e.g., the measured position). The **Difference Image Analysis Source (DIASource)** table contains measurements performed on difference images. The **Forced Source** table contains measurements at predetermined positions (e.g., where objects were detected in other catalogs or exposures). Finally, the **Moving Object** table contains the orbits determined for minor bodies of the Solar System. Detailed description of the schema can be found using the **LSST Schema Browser** (shown on the right).

The decisions on what quantities to measure and include in Level 2 catalogs are based on two guiding principles:

- “Minimize Pixels Needed for Science”:** The large majority of LSST science cases should be enabled by just the catalog, as reprocessing images is costly and difficult for the end-user and should be avoided.
- “Provide simple but useful, commonly used, external or derived, quantities”:** Examples include $E(B-V)$ values for each object, or an estimation of photometric redshift using widely accepted, published, algorithms.

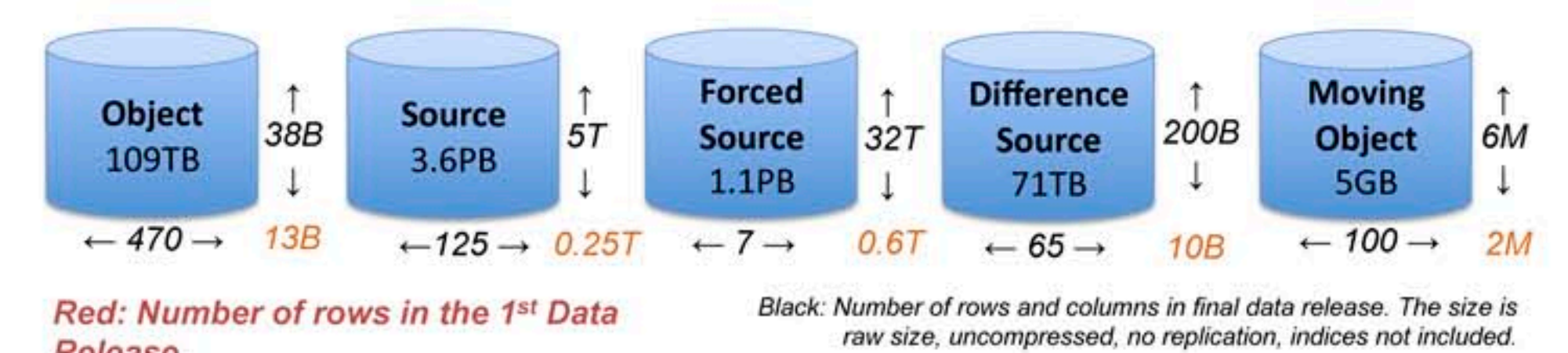
Level 3 (“user-created”) Products: Level 3 products will be created by the community using LSST's hardware and user-contributed software. They are intended to enable processing that would greatly benefit from being co-located with the LSST data. Potential examples include searching for extremely faint Kuiper belt objects using “shift-and-stack”, or detailed characterization of the morphology of interacting galaxies. Note that simply querying Level 2 database does not rise to the level of a “Level 3” effort.

LSST will provide at least 10% of its total storage and computing resources for Level 3 use. An open application and time allocation process will be used to manage access to Level 3 resources. As an example, at the time of DR2 (~year 2022), 20 TFLOPS of compute capability, 300 TB of database, and 1 PB of general-purpose storage space will be made available for Level 3 uses. Furthermore, the LSST Archive will be located in the National Petascale Computing Facility at NCSA and significant non-LSST resources are expected to be available on site.

Level 3 data products may migrate to Level 2, if found to be universally useful and their authors agree to contribute them to the public. LSST will provide support with the necessary software engineering to facilitate such contributions.

<http://lsst1.ncsa.uiuc.edu/schema/index.php?sVer=baseline> (<http://goo.gl/vXPfZ>)

Table List		Details for table Source	
AAA_Version_3_2_4		Table to store high signal-to-noise “sources”. A source is a measurement of Object's properties from a single image that contains its footprint on the sky.	
AngMap		name	type
CalibSource		not null	unit
CcdMap		ucd	description
Ccd_Detector			
DiaSource			
Durations			
Filter			
ForcedSource			
FootprintArea			
LeapSeconds			
Logs			
mops_Event_OrbitDerivation			
mops_Event_OrbitIdentification			
mops_Event_TrackletAttribution			
mops_Event_TrackletPrecovery			
mops_Event_TrackletRemoval			
mops_MovingObjectToTracklet			
mops_SSM			



Final Image Archive	345 PB	All Data Releases Includes Virtual Data (315 PB)
Final Image Collection	75 PB	Data Release 11 (Year 10) Includes Virtual Data (57 PB)
Final Catalog Archive	46 PB	All Data Releases
Final Database	9 PB 32 trillion rows	Data Release 11 (Year 10) Includes Data, Indexes, and DB Swap
Final Disk Storage	228 PB 3700 drives	Archive Site Only
Final Tape Storage	83 PB 3800 tapes	Single Copy Only
Number of Nodes	1800	Archive Site Compute and Database Nodes
Number of Alerts Generated	6 billion	Life of 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.