LSST Database Challenges

The Large Synoptic Survey Telescope (LSST) is a proposed ground-based 8.4-meter telescope. It is the largest astronomical experiment planned for the this decade. Once in operation (year 2018), its 3.2 giga-pixel camera will produce approximately half a petabyte of archive images every month. These data need to be reduced in under a minute to produce real-time transient alerts, and then added to the cumulative catalog for further analysis. During the expected 10-year LSST lifetime, the survey is expected to catalog over 50 billion stars and galaxies based on tens of trillions individual astronomical source detections. All the sources, stars and galaxies will be analyzed by a wide range of users, ranging from professional astronomers, through amateur astronomers to the general public.

The raw images will be kept in files. All data derived from the images as well as the metadata including file-related metadata will be kept in a database. It is expected that the vast majority of all analysis will be done using the database.

Data Flow

The LSST telescope will be located on Cerro Pachón in Chile. The only database-related activity there will involve capturing the hardware configuration, calibration, and weather conditions into a small on-line database. This data together with raw images will be sent to the Base Camp – a local processing center in Chile. The main activity at the Base Camp will be real-time alert generation. The alert generation will involve continuously updating a 100+ terabyte database at a rate of ~3,000 rows/sec and creating some 3,000 new rows/sec. The raw data together with the on-line database will also be sent to the Main Archive in the USA. The Main Archive will re-run alert generation and update the Unreleased Catalog – a set of tables containing all unreleased data, sized in the hundreds of terabytes. Several times a year, all images for each observed patch of sky will be co-added in a process called Deep Detection. Co-adding images will allow the detection of O(100x) more sources than the difference image approach used at the Base Camp. The main products of each Deep Detection run will include the three biggest LSST Catalogs: the Object Catalog the Source Catalog, and the ForcedSource Catalog. The former will contain all detected and classified astronomical objects, and the latter two will contain information about all detected sources. Sources with high signal-to-noise ratio will be stored in "wide" (information-reach) Source table, while low SNR Sources will be stored in "narrow" ForcedSource table. Once a year catalogs produced by Deep Detection will be released, and the Unreleased Catalog will be purged at that time. The size of the Object Catalog for each release will be measured in the tens of billions of entries, the corresponding Source Catalog will be measured in the hundreds of billions; and the ForcedSource in tens of trillions. Catalog in the total data size will be well into the petabytes.

It is expected that a full replica of all data generated by the Main Archive will be also kept at a different location. In addition, data will be sent to multiple Analysis Centers. The number of analysis centers, their locations and their exact roles are under discussion. One of the possible scenarios is that each such center will serve a fraction of the LSST data and will specialize in one type of analysis.

Key Database Challenges

Key LSST database challenges include:

- data correlations
- querying multi-petabyte database
- integrating database queries with pixel data
- data uncertainties
- real-time transient alerting requirements
- spatio-temporal aspects
- all data public with easy access
- provenance and metadata
- incremental large scale updates

Data Correlations

Unlike most systems, astronomical data is highly correlated: objects on the sky affect properties of other objects (imagine a bright star next to a faint one, or a fast moving asteroid passing in front of a star). For that reason, partitioning data and parallelizing queries becomes a real challenge.

Querying Multi-petabyte Database

Based on current estimates, at the end of the LSST survey in 2028, the LSST will manage close to 400 trillion (3.7E14) database rows scattered across many tables, databases and servers. This number includes multiple versions of the same data processed with different algorithms and kept as different data releases, but it does not account for any data replication. The size of the LSST database is expected to reach multi-tens of petabytes. With all persistent overheads, including indexes, it is expected one full replica will reach ~40 petabytes.

We expect to see many complex, large volume, ad-hoc, analytical queries. The most challenging types of queries include near neighbor (*find certin type of objects near other objects, across entire* sky), and time series (*for example, finding pairs of objects with "similar" time* series). The query load is very difficult to predict, based on past survey; the unprecedented LSST scale will very likely enable new types of data analyses.

Integrating Database Queries with Pixel Data

Scientists will very likely need to process pixel data in ways that are not included in the standard data release. The aim is to offer non-standard lookup, database query, and resulting pixel arithmetic all in one interactive package.

Data Uncertainties

Only a small percentage of all classified astronomical objects will be classified with 100% probability. Vast majority will be uncertain: for example an object might be classified as “we are 70% certain it is a star, 27% it is a galaxy, and 3% -unknown”. Different scientists will then want to process data using different cut outs, for example some might ask for stars classified with probability > 90%, and some will want all stars classified with probability > 75%”, and the system has to be flexible enough to handle that efficiently. Of course the results based on uncertain data will be approximate, a measurement based on 20 galaxies classified as galaxy with 100% probability is not equal to a measurement on the same number of galaxies classified with lower probability.

Real-time Transient Alerting

The LSST will generate transient alerts within 60 sec after observation of each pair of images (called a visit). In order to generate alerts, images have to be analyzed, and detected sources have to be matched against all known astronomical objects that LSST ever recorded in the observed patch of sky. In practice, from the database perspective, it will involve per visit (every 37 sec):

- persisting newly detected sources, up to 100,000
- selecting up to 10 million rows from the Object Catalog and cross-matching
- the new sources against known objects inside the database
- performing any additional filtering using corresponding past
- measurements of just detected sources (potentially including
- measurements earlier the same night)
- updating certain fields in the matching objects (expected max number of
- matches for every visit is ~100,000, average ~40,000)
- inserting O(1000) new objects and deleting O(100) objects.

In order to meet the 60 sec alerting requirement, all of the above need to be completed in under 10 sec to allow sufficient time for non-database related activities that need to run as part of the alert generation.

Spatio-Temporal Aspects

Most astronomical surveys deal only with spatial data. LSST will explore both spatial and temporal aspects of the data: each piece of the visible sky will be visited 1,000 times during the expected 10 year lifetime of LSST, opening a movie-like window on objects that change or move. For the database, that implies a need for efficient lookups across both dimensions.

Public Data

All LSST data will be made public with minimal delay, and should be easily accessible. That implies a wide range of users and consequently a wide range of queries to be supported, ranging from many short queries (public access) to complex queries scanning the entire data set (professional astronomers).

Provenance and Metadata

We need to store metadata (information about the data) for all data elements, including every image, source, and object. The metadata consists of two parts:

- provenance (execution trace, or in other words information "how we came up with the data")
- information about data not captured by the provenance, including summary statistics.

This type of metadata has to be flexible (perhaps as general as key/value pairs), and it has to scale to billions of entries. Tracking data provenance is extremely important, because the intermediate data products used to generate final results and even some final data products will not be saved, and we must be able to reconstruct them on demand as needed. Keeping the intermediate products would increase the size of the LSST data to a truly unmanageable extent. As an example, difference images generated at the Base Camp and used for alert generation will not be saved, but we must be able to reconstruct them. To reconstruct them, we must preserve the exact software and hardware configuration used at the time when these images are produced. The case of difference images alone will save us from persisting 50 more petabytes, at the expense of extra complexity in the software supporting provenance.

Incremental Large-Scale Updates

All released data will be immutable, primarily to guarantee reproducibility of published results. Data immutability is good from the database perspective, as it opens many possibilities for optimization. However, one of the catalogs, the Unreleased Catalog, will have to be constantly updated. The Unreleased Catalog will contain up to one year's worth of data, which translates to hundreds of billions of rows. Performing incremental updates to such a catalog and, in particular, to any indexes on the catalog, poses a unique, unprecedented set of challenges.