The LSST Data Processing Pipeline

T. Axelrod (Steward), A. Connolly (U. Pitt.), Z. Ivezić (U. Wash.), J. Kantor (LSSTC), R. Lupton (Princeton), R. Plante (NCSA), C. Stubbs (Harvard), D. Wittman (UCD)

In science observing mode, the LSST telescope and camera system will deliver a 3.2 Gpixel image every 12 sec, a data rate of about 0.5 GByte/sec. The data processing pipeline must process this incoming data to produce the LSST's primary data products: the calibrated image archive; the detection catalog; the object catalog; and real-time alerts. Additionally, the quality of the incoming data must be rapidly assessed and fed back to the observatory control system. The pipeline must be flexible enough to allow addition of new processing stages and replacement of existing algorithms with improved ones, and must be robust in the face of the failure of hardware components such as disk drives. We present a preliminary design of the primary LSST data products, and of the data processing pipeline. The mapping of the pipeline onto computing hardware is discussed, along with estimates of the computational, I/O, and network bandwidths required.

The LSST Data Processing Pipeline

Computational Requirements

Two broad requirements determine the computational capacity required by the LSST data pipeline:

1. The pipeline must not fail behind in processing observations during a night. This requires completely processing during 24 hours a set of images that arrive every 12 sec for at least 10 hours (3000 images/24 hrs).
2. The pipeline must have sufficient excess capacity to support reprocessing of past data. This is difficult to assess in advance of operations, but experience with other surveys suggests that reprocessing will grow to perhaps 2x the nightly throughput.

These determine the required throughput of the pipeline. There is an additional latency requirement, set by the need to generate alerts promptly in response to interesting transient events.

The LSST pipeline can be highly parallel, applying large numbers of CPU's to many tasks.

The Data Products

LSST will produce four classes of publicly available data products:

1. Image Archive
2. Object Catalog
3. Alerts
4. Object Catalog

The relationships between these products, and some of their inner structure, are shown in Fig. 2.

- The Image Archive contains highly processed images, such as image stacks and subtracted images, as well as raw and calibrated images.
- The Detection Catalog contains an entry for every detection generated by the Detection Pipeline. Depending on pipeline parameters, such as the detection threshold, the Detection Catalog may contain large numbers of false detections.
- The Object Catalog is intended to contain information about real astronomical objects, including their time dependent brightnesses in the various filters, motion on the sky, shape parameters, and classification. It is likely to be the most heavily used, and most computationally demanding, of the data products.
- Alerts are notifications sent to subscribers of events that meet predefined criteria. Because the interests and capacity to process alerts vary widely among potential subscribers, we are committed to allowing alerts to be flexibly defined on a per-subscriber basis. All publicly available LSST data products will be made available through a VO-compliant interface. Many aspects of the interface are still to be worked out in conjunction with the VO community.