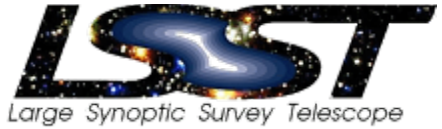
DC3b Planning and Scoping

Jeff Kantor
DM System Manager

Tim Axelrod
DM System Scientist

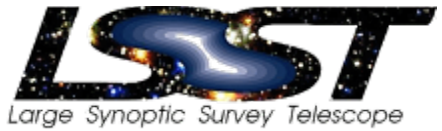
The DM Team

DC3a Results Highlights



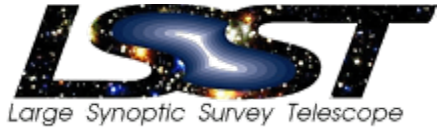
- **Alert Production (formerly Nightly Pipelines)**
  - **Added Instrument Signature Removal stages**
    - Now doing all stages except final alert generation
  - **Added SDQA base classes, initial metrics**
  - **All stages improved performance over DC2 (10 – 50%)**
  - **Latency is approximately 4 minutes/visit (latency gap 3 minutes)**
  - **Image/pixel processing is approximately 1 minute (throughput gap ~40s)**
  - **Algorithm performance improvements are still possible and necessary**
  - **Cannot rely on algorithm and multi-core improvements alone to meet spec**
    - Multi-core performance at best a factor of 2 by 2016
    - Contingency is more than enough to purchase additional CPUs to meet demand
    - Alternative processing architectures must be examined (GPUs, cell) for critical parts of image processing
    - Slice (data) parallel is not enough, must go Stage (algorithm) parallel in some areas e.g. template warping
    - Some database operations/Association Pipeline will need to be parallelized for scaling
- **Application Framework/Middleware**
  - **Now mature enough to get non-DM scientists/developers using them to develop pipelines**
  - **Still need to enhance support "stand-alone" development/debugging with minimal middleware on single node/laptop**

DC3a Results Highlights continued



- **Infrastructure**
  - **Used dedicated 10-node (48 CPU) cluster for all pipelines, data, control**
  - **Will need to increase for Data Release processing to handle self-calibration, database scaling tests, and other load increases**
- **Development Process, Build Environment, and Tools**
  - **Need to reconfigure trac to support new workflow and additional projects (e.g. image simulation)**
  - **Need to integrate standards checker tool with test/build process**
  - **Need to do code reviews more formally**
- **Management/Organization**
  - **Very late to develop resource leveled plan, nail down scope, significant schedule overrun (50%)**
  - **80 use cases / 53 activities in scope (vs DC2 70 use cases)**
  - **7.5 staff-years expended by 28 individuals (~30% avg. participation)**
  - **107k lines of code produced (c++, python, SQL, java, scripts)**
  - **Need to get broader input/participation from Science Collaborations/ Partners on algorithms**

DC3b Overall Goals and Objectives



- Identify and prototype key image processing operations on GPUs
- Develop first release of Data Release Production (coaddition, detection, photometric calibration, astrometric calibration pipelines)
- Expand Pipelines functionality with day MOPS integrated into LSST software framework
- Improve quality of data outputs per Science Collaboration input/analysis
- Expand Science Data Quality Analysis System (SDQAS) to play significant role in enabling above (at least an average of 1 metric/pipeline stage)
- Continue scaled tests of data transfer, data processing, database ingest to 15% of final LSST requirements
- Integrate new team members (U Penn, others)
- Answer key questions for PDR as documented in DM R&D Plan (e.g. scalable database architecture)
- Extend DC2 Application Framework and Middleware to support Data Release Production
- Conduct stand-alone scaled tests of data query (with map reduce/bigtable and DBMS SQL) and file life cycle management (with iRODS)

Items in red have PDR deadlines, others are due May, 2010