LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-CONF-217858

Designing a multi-petabyte database for LSST

J. Becla, A. Hanushevsky, S.Nikolaiv, G. Abdulla, A. Szalay, M. Nieto-Santisteban, A. Thakar, J. Gray

December 21, 2005

Designing a multi-petabyte database for LSST
Orlando, FL, United States
May 24, 2005 through May 31, 2005

This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

Designing a multi-petabyte database for LSST

**Jacek Becla, Andrew Hanushevsky (SLAC),
Sergei Nikolaev, Ghaleb Abdulla (LLNL),
Alex Szalay, Maria Nieto-Santisteban, Ani Thakar (Johns Hopkins),
Jim Gray (Microsoft Research)**

The 3.2 giga-pixel LSST camera will produce over half a petabyte of raw images every month. This data needs to be reduced in under a minute to produce real-time transient alerts, and then cataloged and indexed to allow efficient access and simplify further analysis. The indexed catalogs alone are expected to grow at a speed of about 600 terabytes per year.

The sheer volume of data, the real-time transient alerting requirements of the LSST, and its spatio-temporal aspects require cutting-edge techniques to build an efficient data access system at reasonable cost. As currently envisioned, the system will rely on a database for catalogs and metadata. Several database systems are being evaluated to understand how they will scale and perform at these data volumes in anticipated LSST access patterns.

This paper describes the LSST requirements, the challenges they impose, the data access philosophy, and the database architecture that is expected to be adopted in order to meet the data challenges.