LSST and the SLAC Scalable Data Systems Group

Kian-Tat Lim

Scientific Computing Workshop

June 20, 2011

SLAC

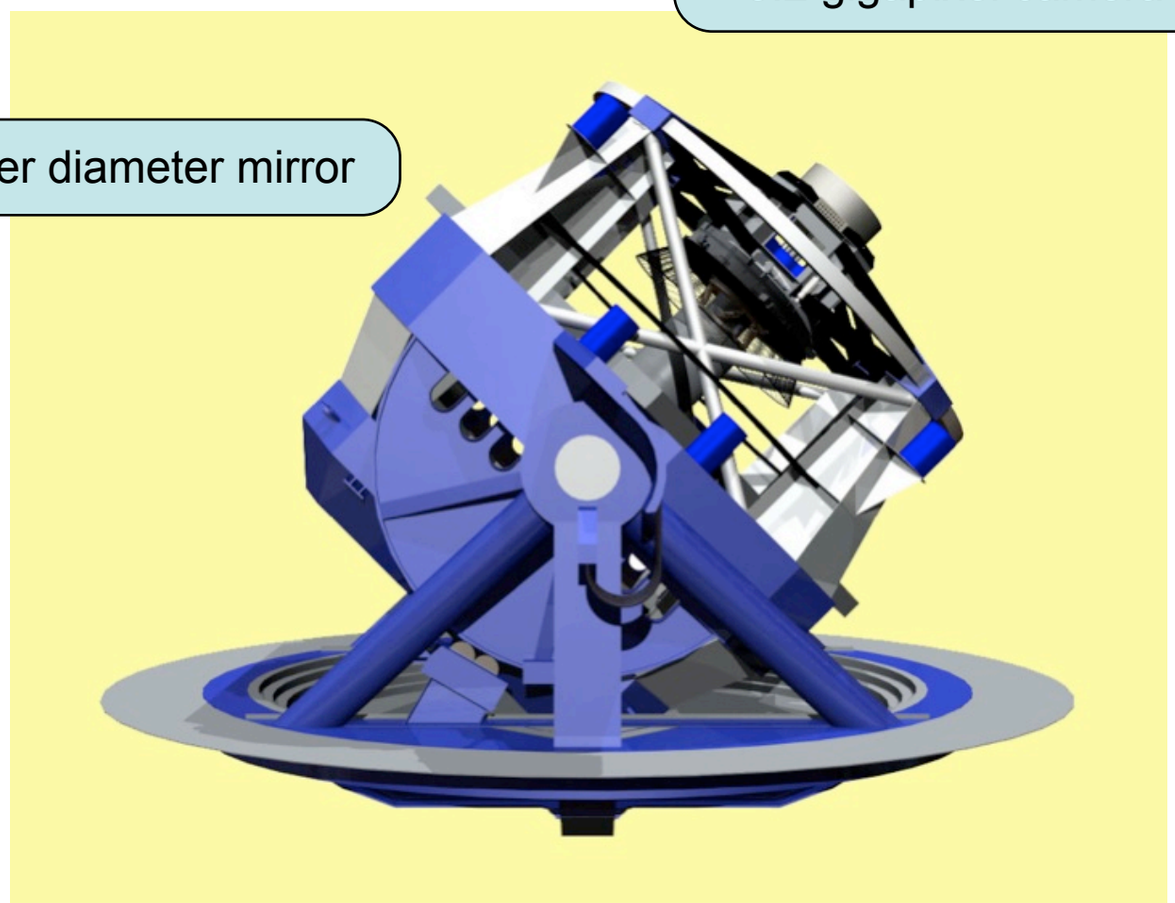
Large Synoptic Survey Telescope

LSST: What Is It

Large Synoptic Survey Telescope

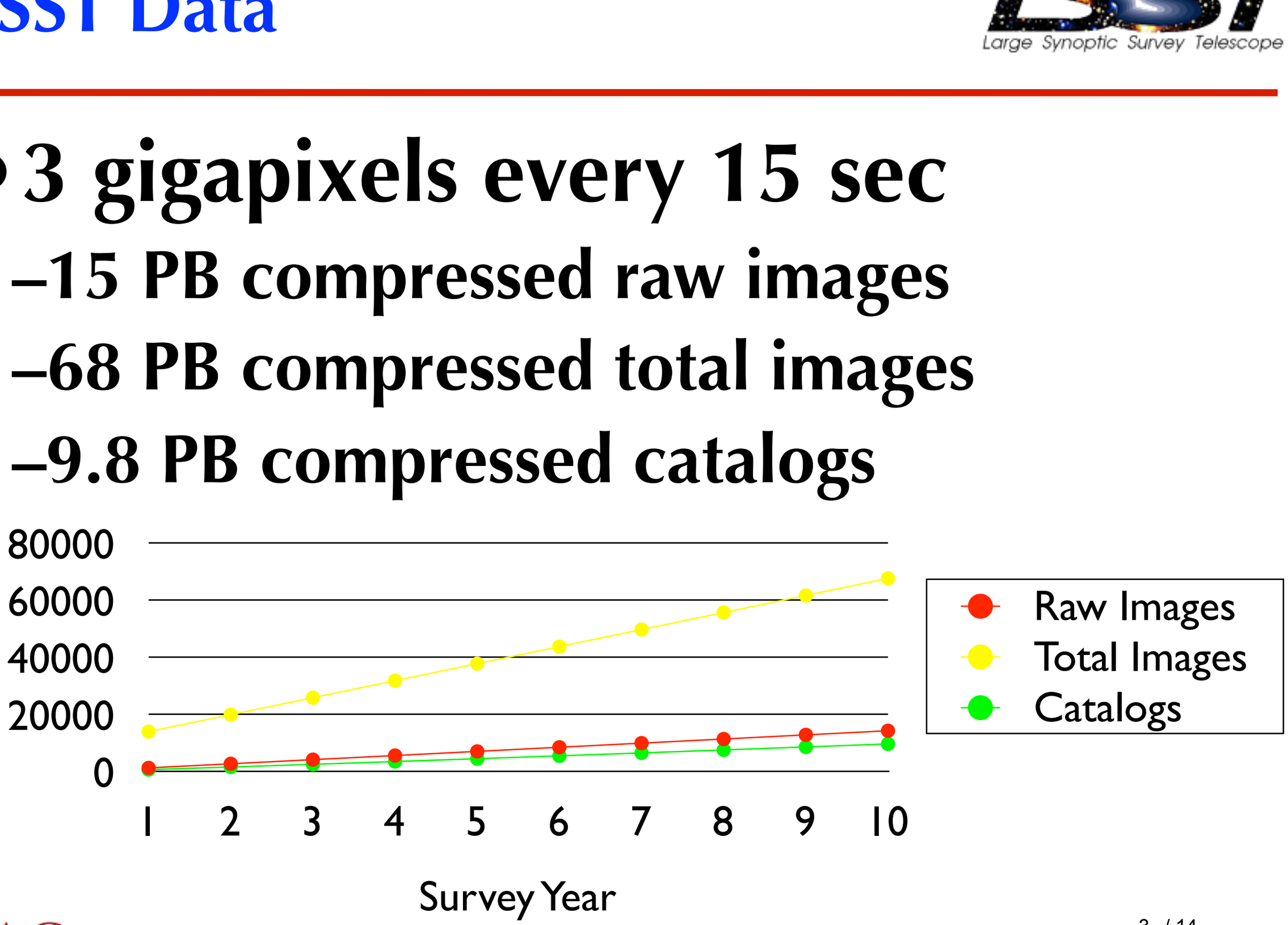
Large Synoptic Survey Telescope

Movie of (half) the sky



LSST Data

- 3 gigapixels every 15 sec
  - –15 PB compressed raw images
  - –68 PB compressed total images
  - –9.8 PB compressed catalogs

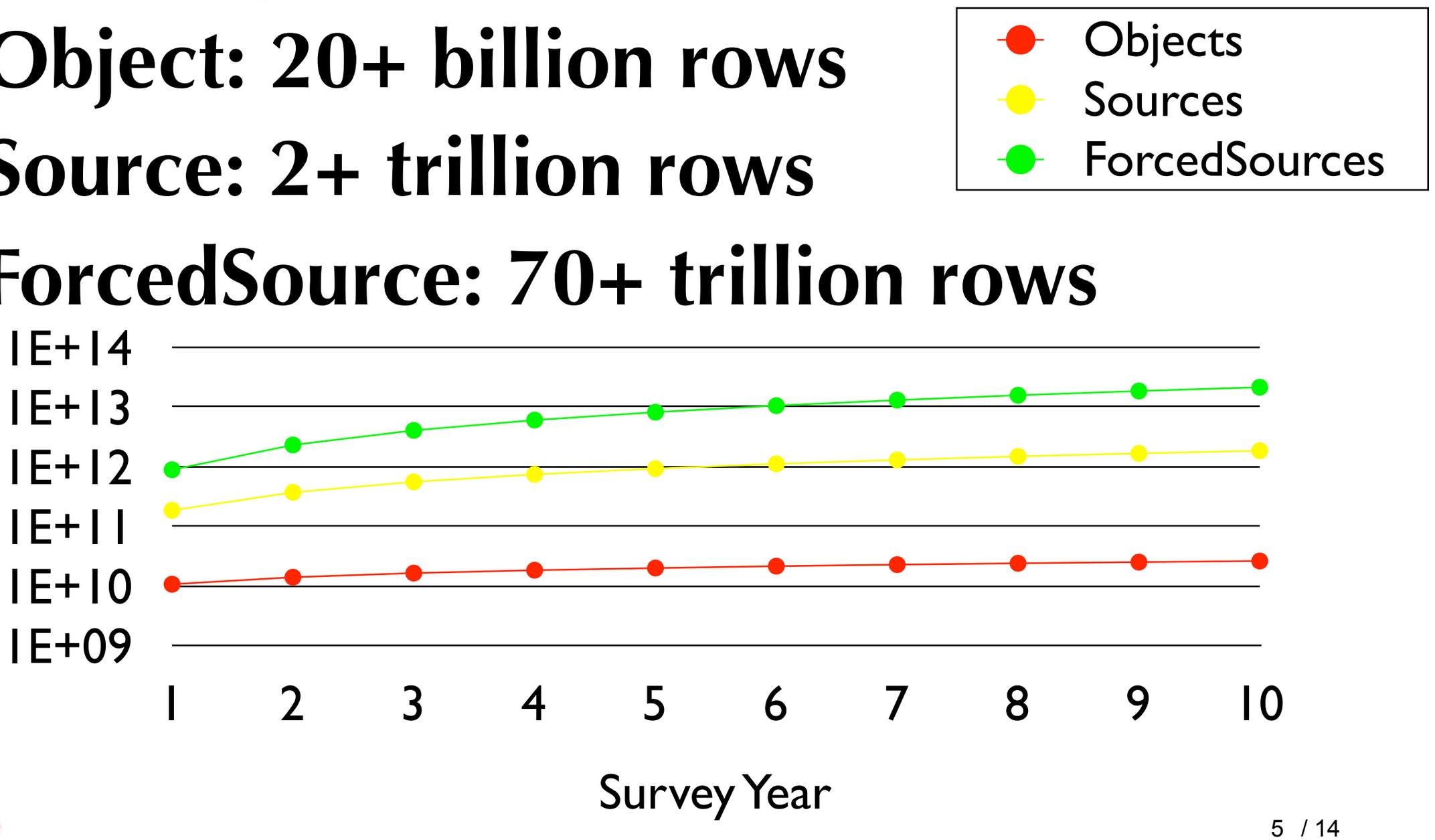


Data Access Framework

- **Butler**
  - –Retrieves datasets by key/value pairs
- **Persistence Framework**
  - –Boost serialization, FITS files, database
- **Image access**
  - –iRODS
  - –REDDnet

Scalable Database: qserv

- **Catalogs**
  - **Object: 20+ billion rows**
  - **Source: 2+ trillion rows**
  - **ForcedSource: 70+ trillion rows**



qserv Goals

Large Synoptic Survey Telescope

Scalability

(includes elasticity, reliability)

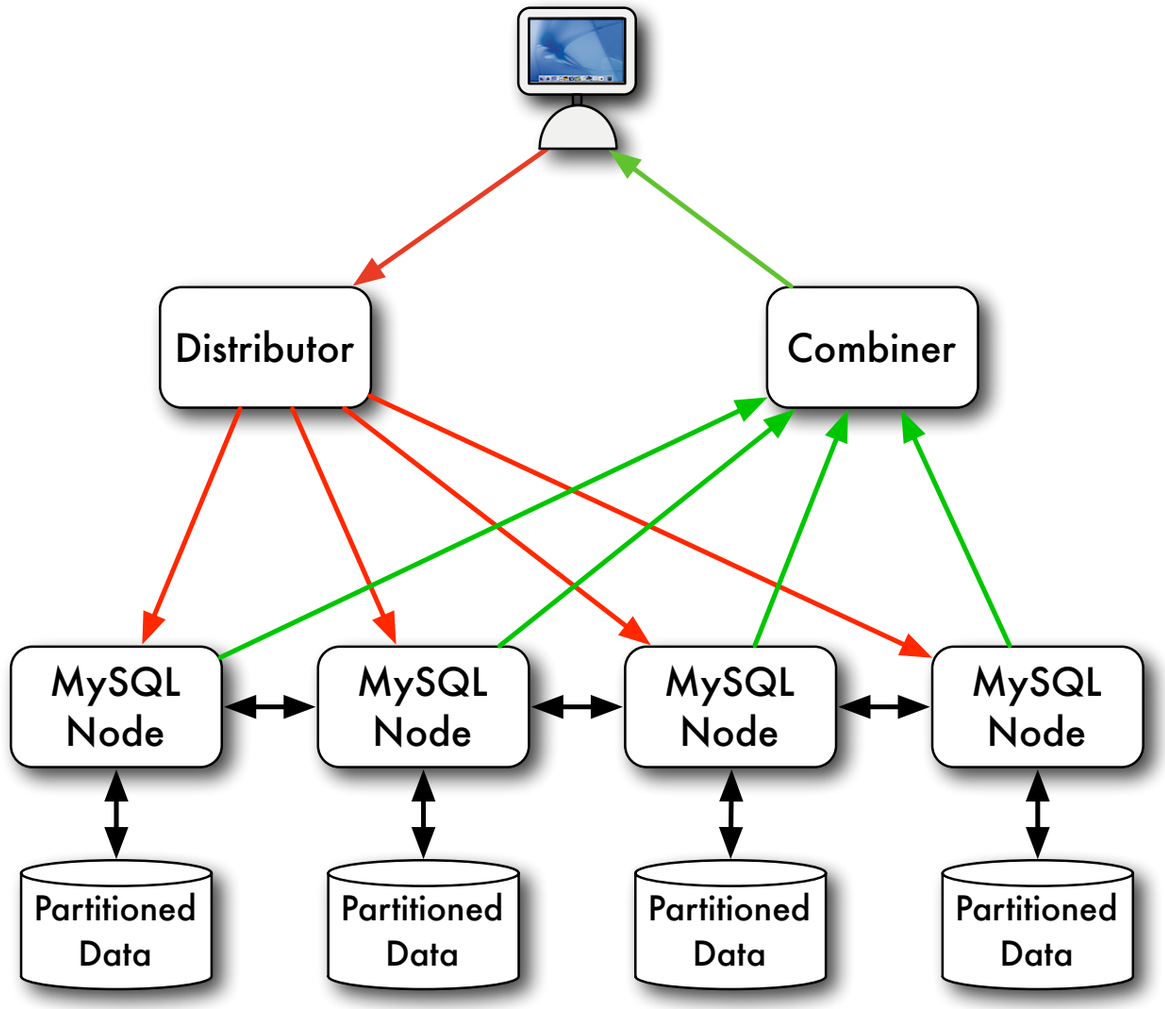
Affordability

Queryability

Manageability

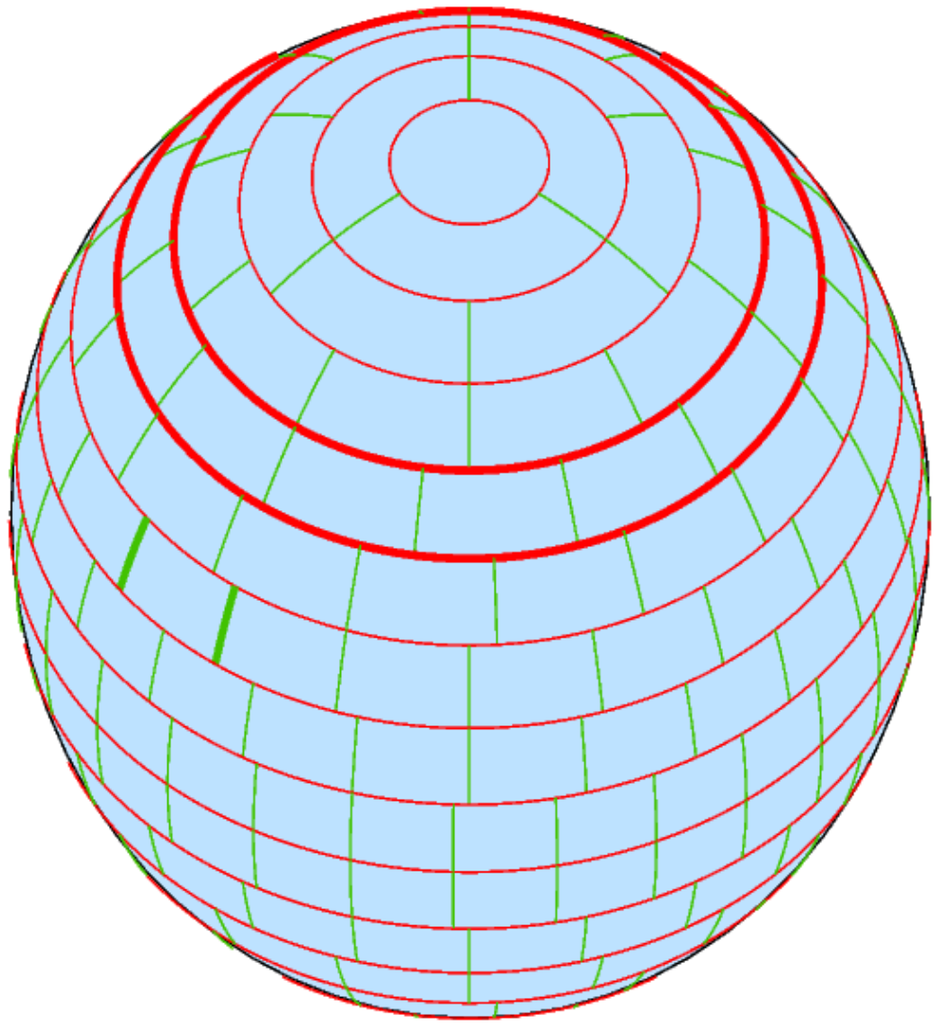
qserv Architecture

Map/Reduce on RDBMS



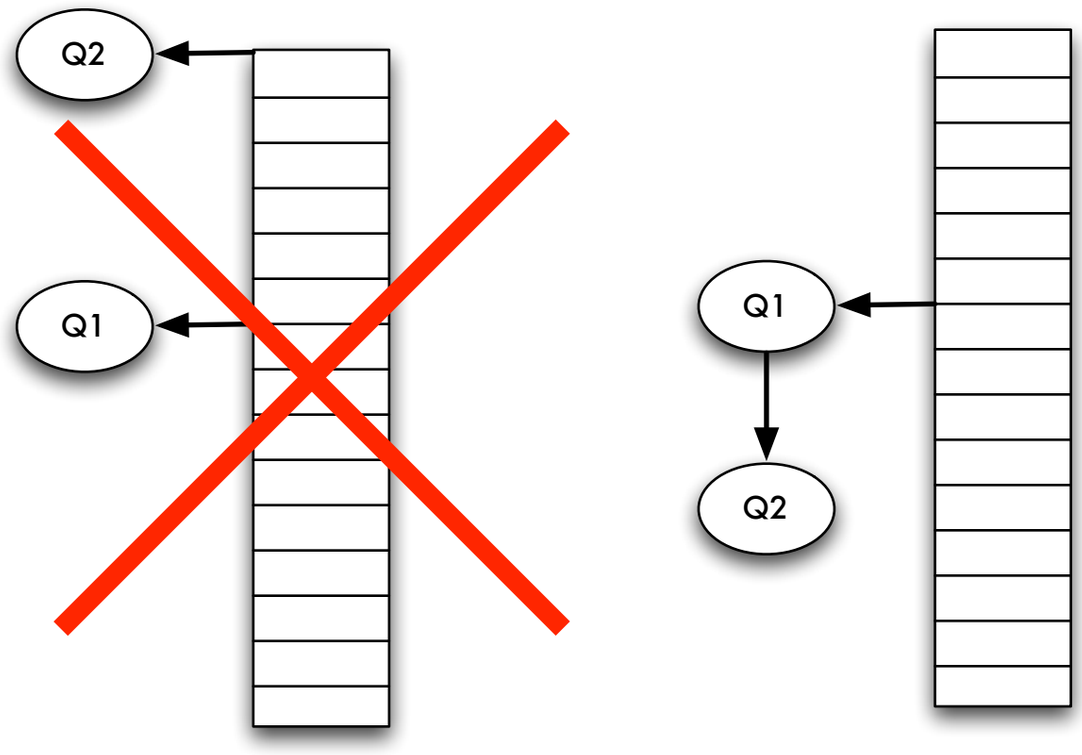
qserv Architecture

Overlapping partitions



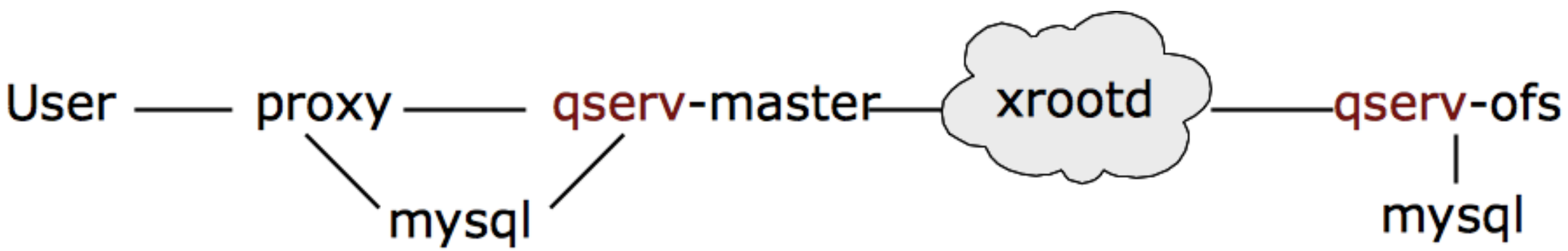
qserv Architecture

Shared scans



qserv Implementation

- MySQL Proxy UI
- Master: C++ and MySQL
- Communication: xrootd
- Worker: C++ and MySQL



150 node cluster

Showed query functionality

Showed scalability

XLDB

- Workshop and conference series
- Academics, scientists, but also industry and vendors
- Practical experience
- Spinoff workshop in Europe

SciDB

- Array-oriented database
- Initial release last week
- Loader for HDF5 files

Summary

Large Synoptic Survey Telescope

Preparing for petabytes