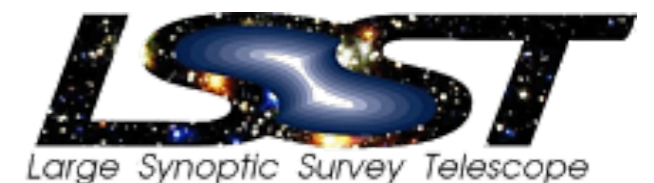# LSST and the SLAC Scalable Data Systems Group

Kian-Tat Lim

Scientific Computing Workshop

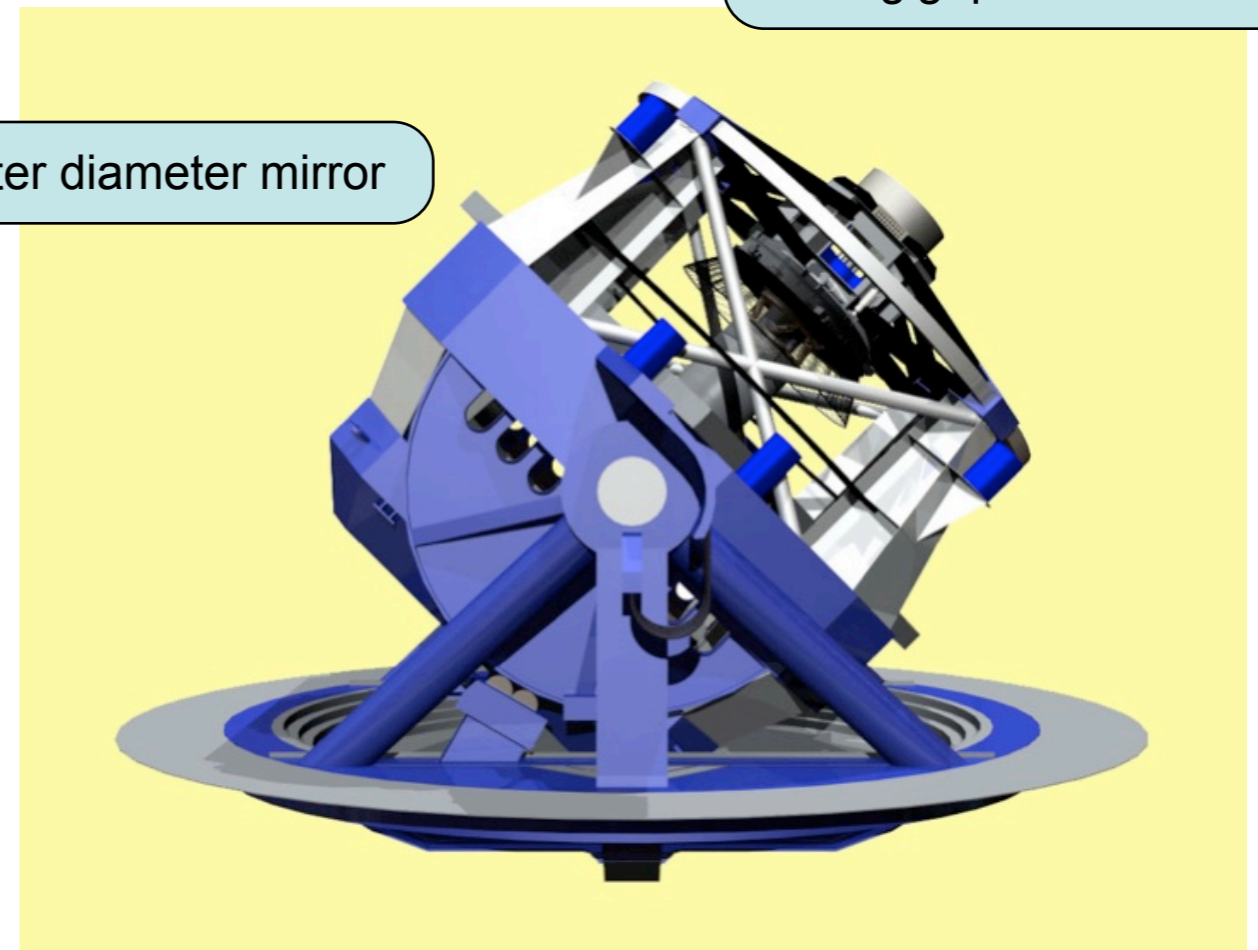**June 20, 2011**

SLAC

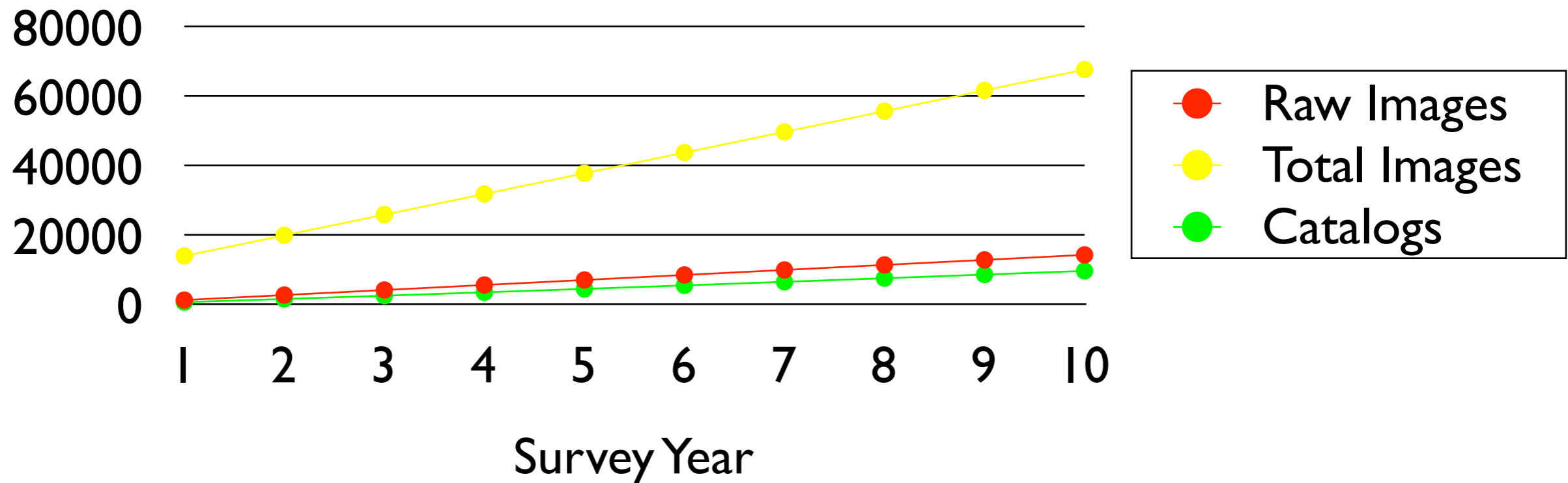**Large Synoptic Survey Telescope**

Movie of (half) the sky

3.2 gigapixel camera

8.4 meter diameter mirror

SLAC

Scientific Computing Workshop
June 20, 2011

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

SLAC

Scientific Computing Workshop
June 20, 2011

Data Access Framework

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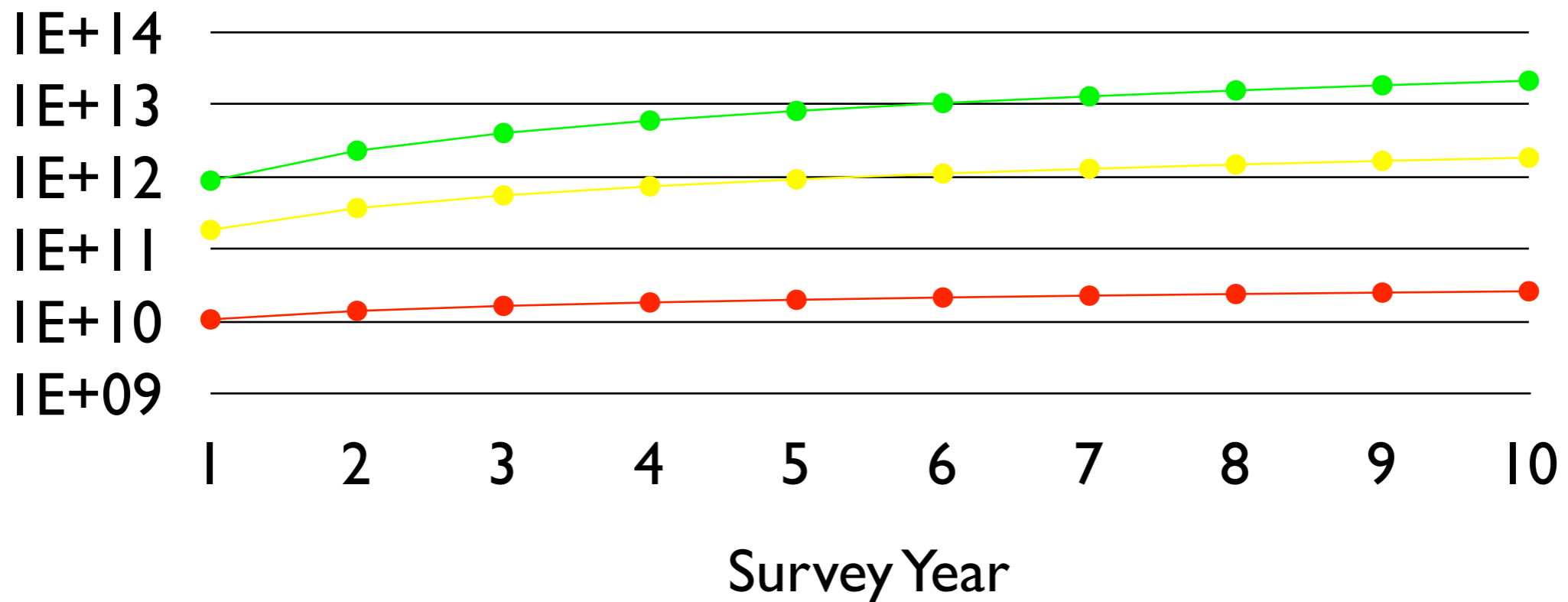- **Image access**
  - **iRODS**
  - **REDDnet**

SLAC

Scientific Computing Workshop
June 20, 2011

Scalable Database: qserv
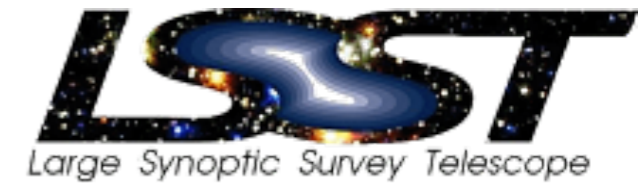
●Catalogs
- –Object: 20+ billion rows
- –Source: 2+ trillion rows
- –ForcedSource: 70+ trillion rows

Legend:
- Objects (red)
- Sources (yellow)
- ForcedSources (green)

Y-axis: 1E+14, 1E+13, 1E+12, 1E+11, 1E+10, 1E+09

X-axis (Survey Year): 1 2 3 4 5 6 7 8 9 10
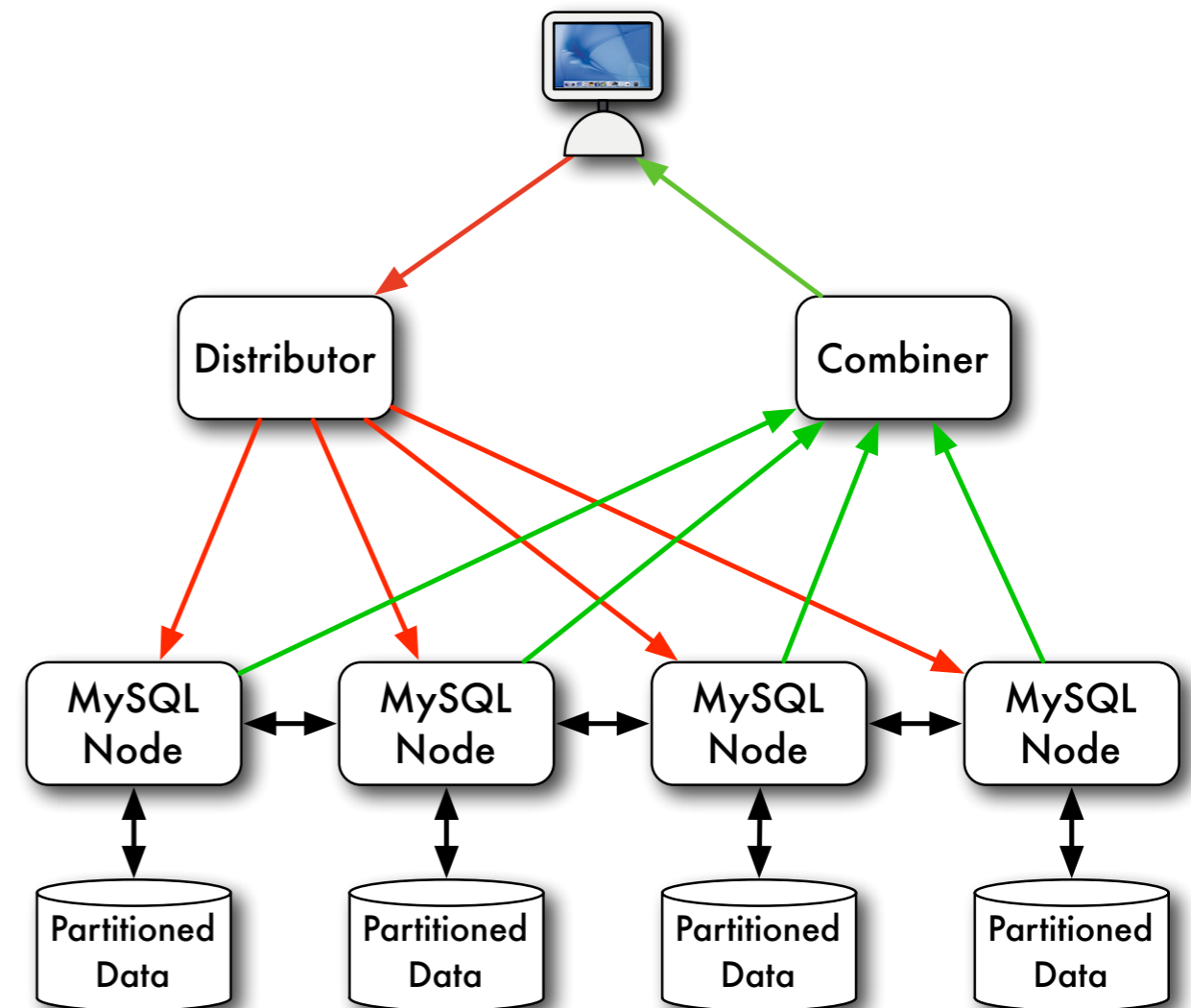
SLAC

qserv Goals

Scalability

(includes elasticity, reliability)

Affordability

Queryability

Manageability

Map/Reduce on RDBMS

Overlapping partitions

Scientific Computing Workshop
June 20, 2011

SLAC
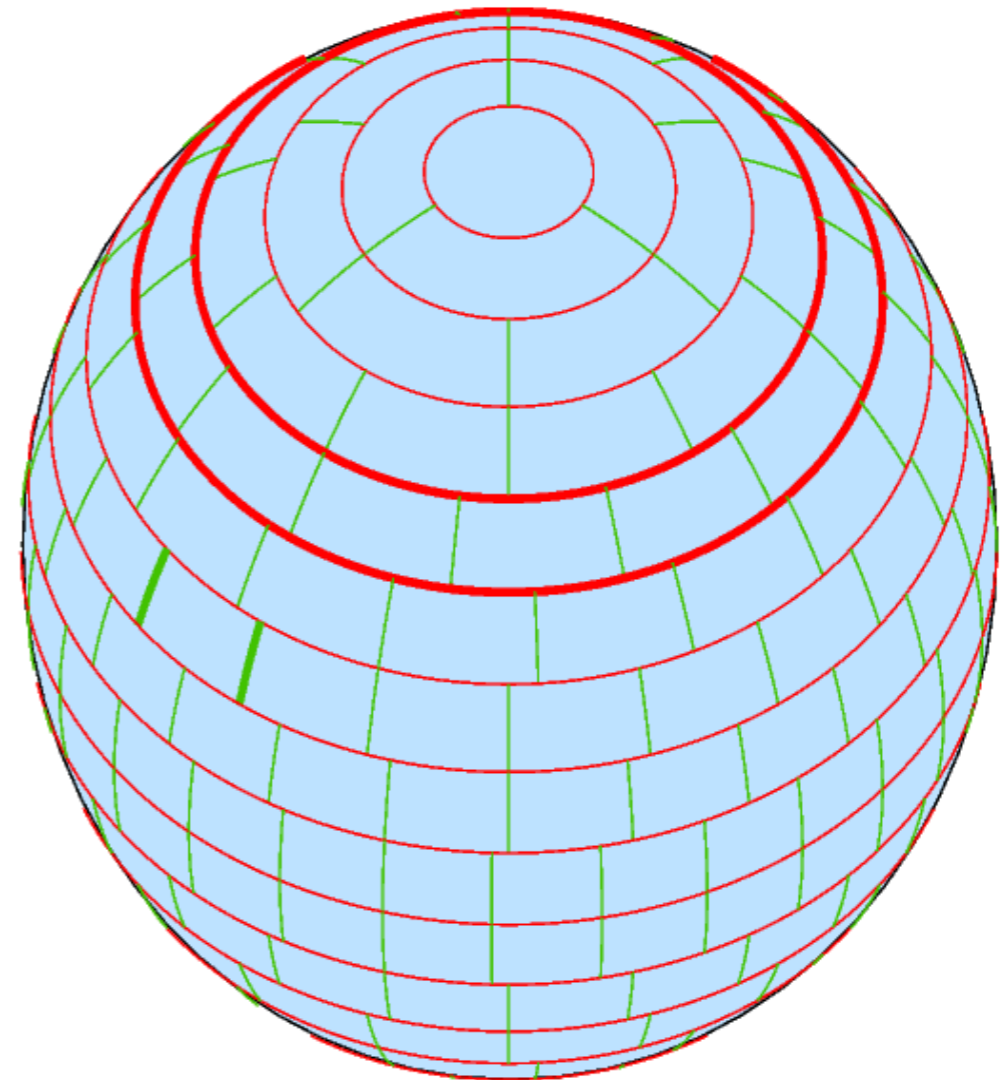
Shared scans
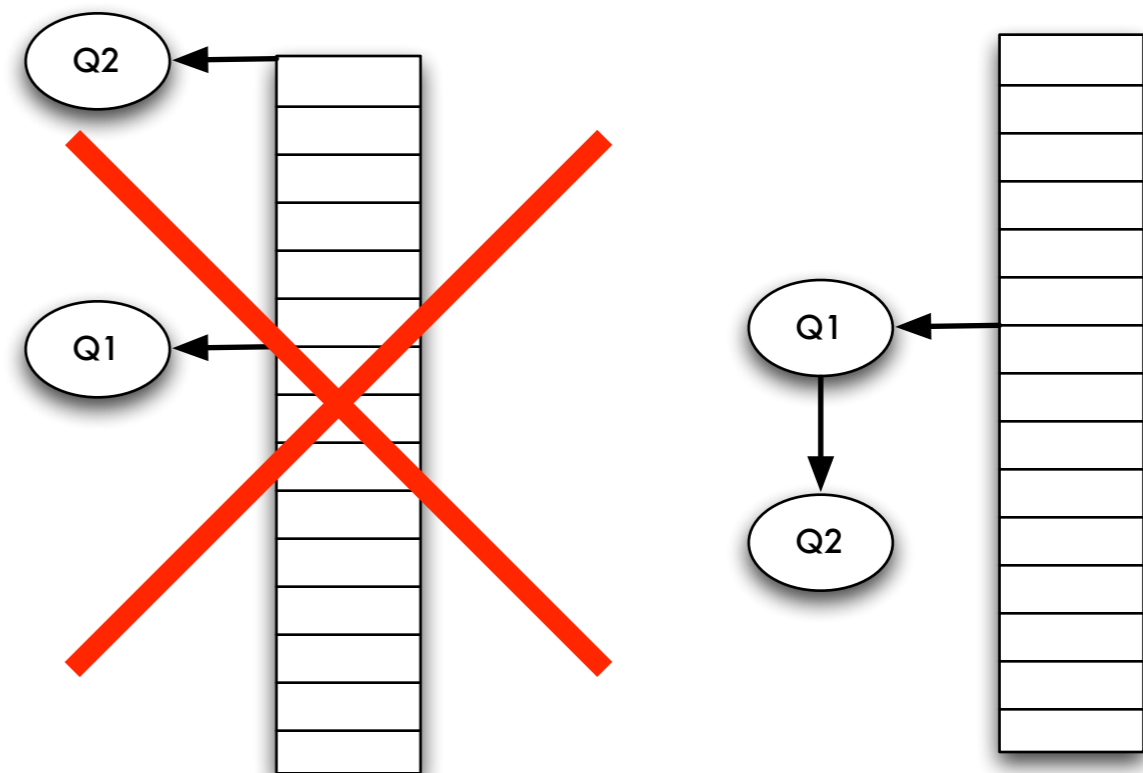
qserv Implementation
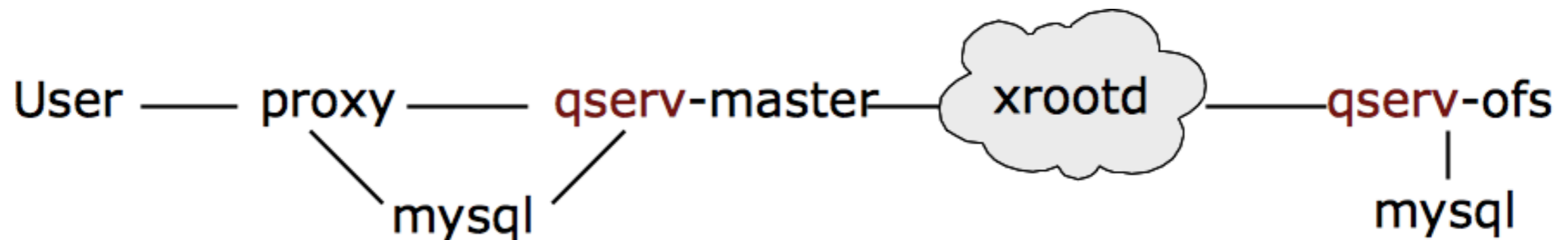
- **MySQL Proxy UI**
- **Master: C++ and MySQL**
- **Communication: xrootd**
- **Worker: C++ and MySQL**

User —— proxy —— qserv-master —— xrootd —— qserv-ofs

mysql (under proxy)   mysql (under qserv-ofs)

Scientific Computing Workshop
June 20, 2011

150 node cluster

Showed query functionality

Showed scalability

SLAC

Scientific Computing Workshop
June 20, 2011

XLDB

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

SciDB

- **Array-oriented database**
- **First production release last week**
- **Loader for HDF5 files**

SLAC

Scientific Computing Workshop
June 20, 2011

Summary

Preparing for petabytes

SLAC

Scientific Computing Workshop
June 20, 2011