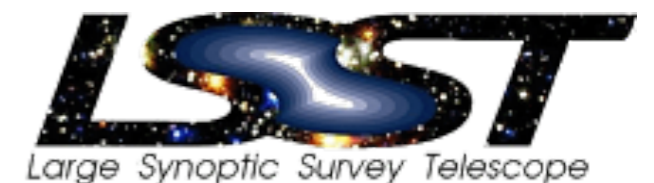

LSST and the SLAC Scalable Data Systems Group

Kian-Tat Lim

Scientific Computing Workshop

June 20, 2011

SLAC

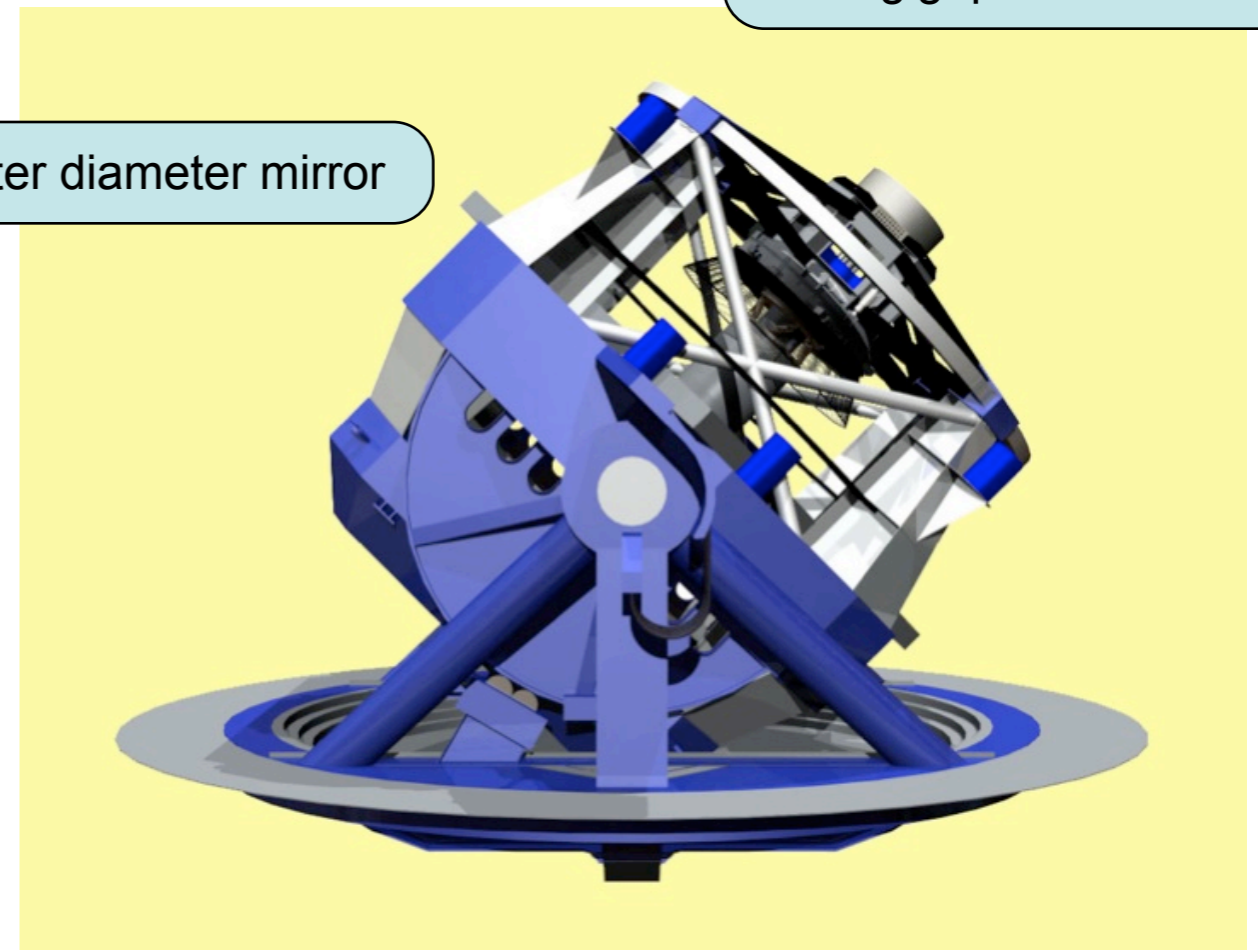


Large Synoptic Survey Telescope

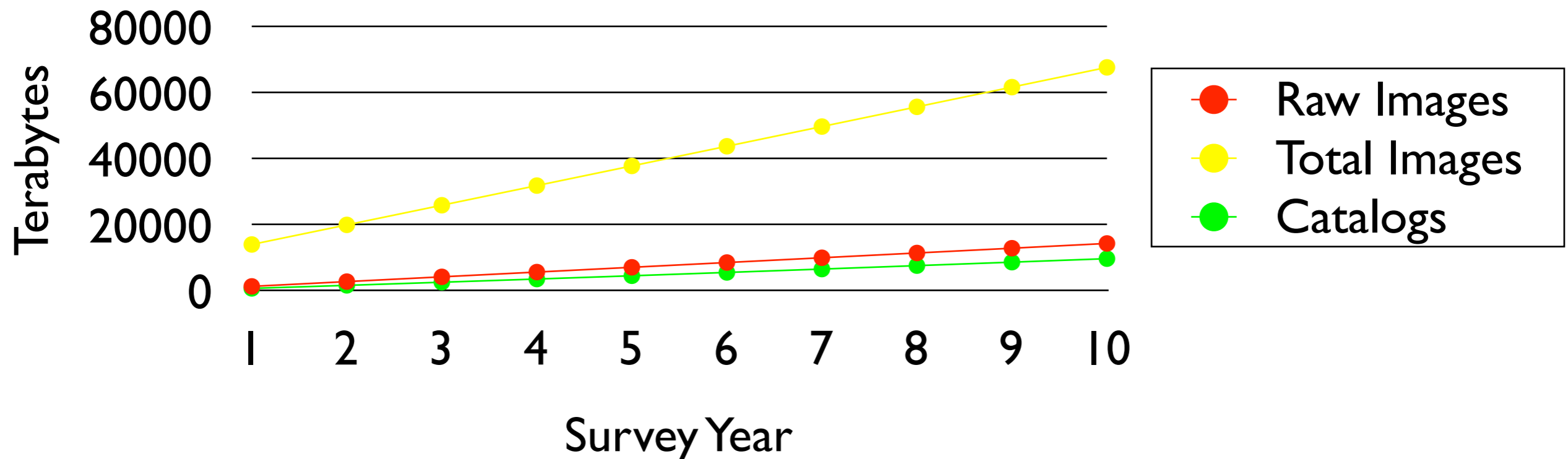
Movie of (half) the sky

8.4 meter diameter mirror

3.2 gigapixel camera



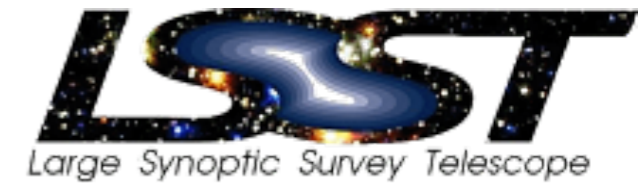
- **3 gigapixels every 15 sec**
 - 15 PB compressed raw images
 - 68 PB compressed total images
 - 9.8 PB compressed catalogs



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



Scalable Database: qserv

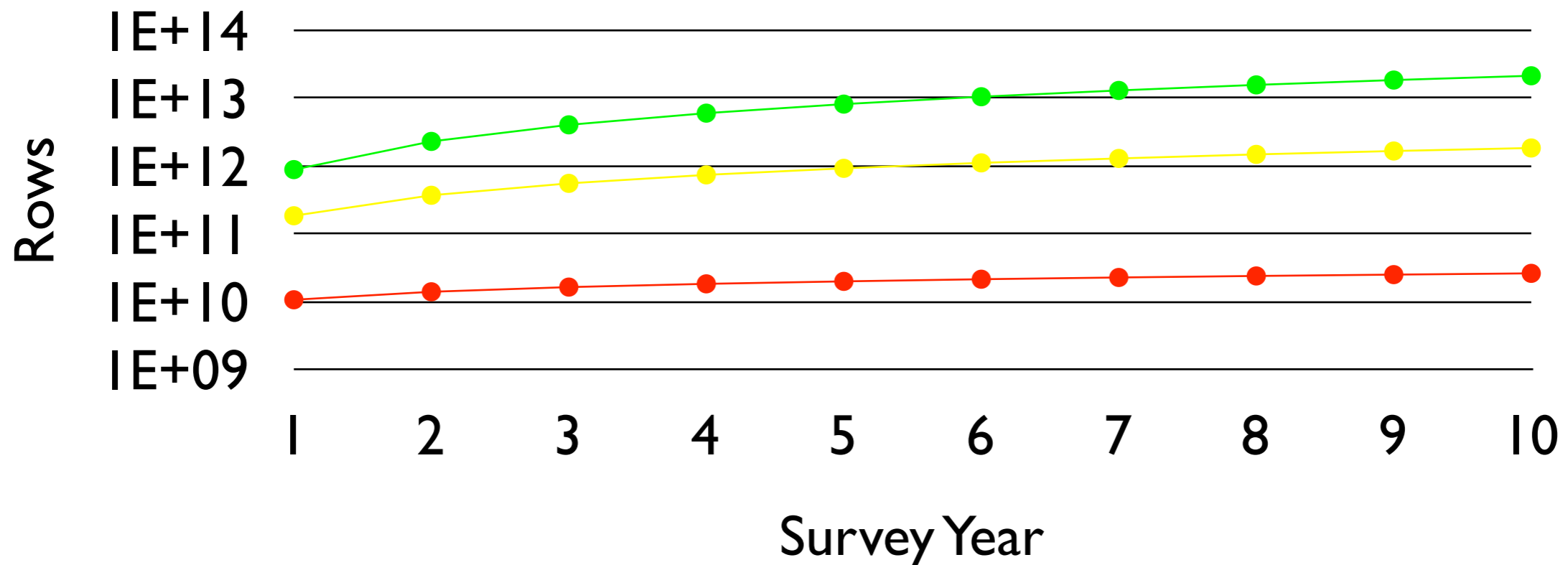


● Catalogs

–Object: 20+ billion rows

–Source: 2 trillion rows

–ForcedSource: 20+ trillion rows



Scalability

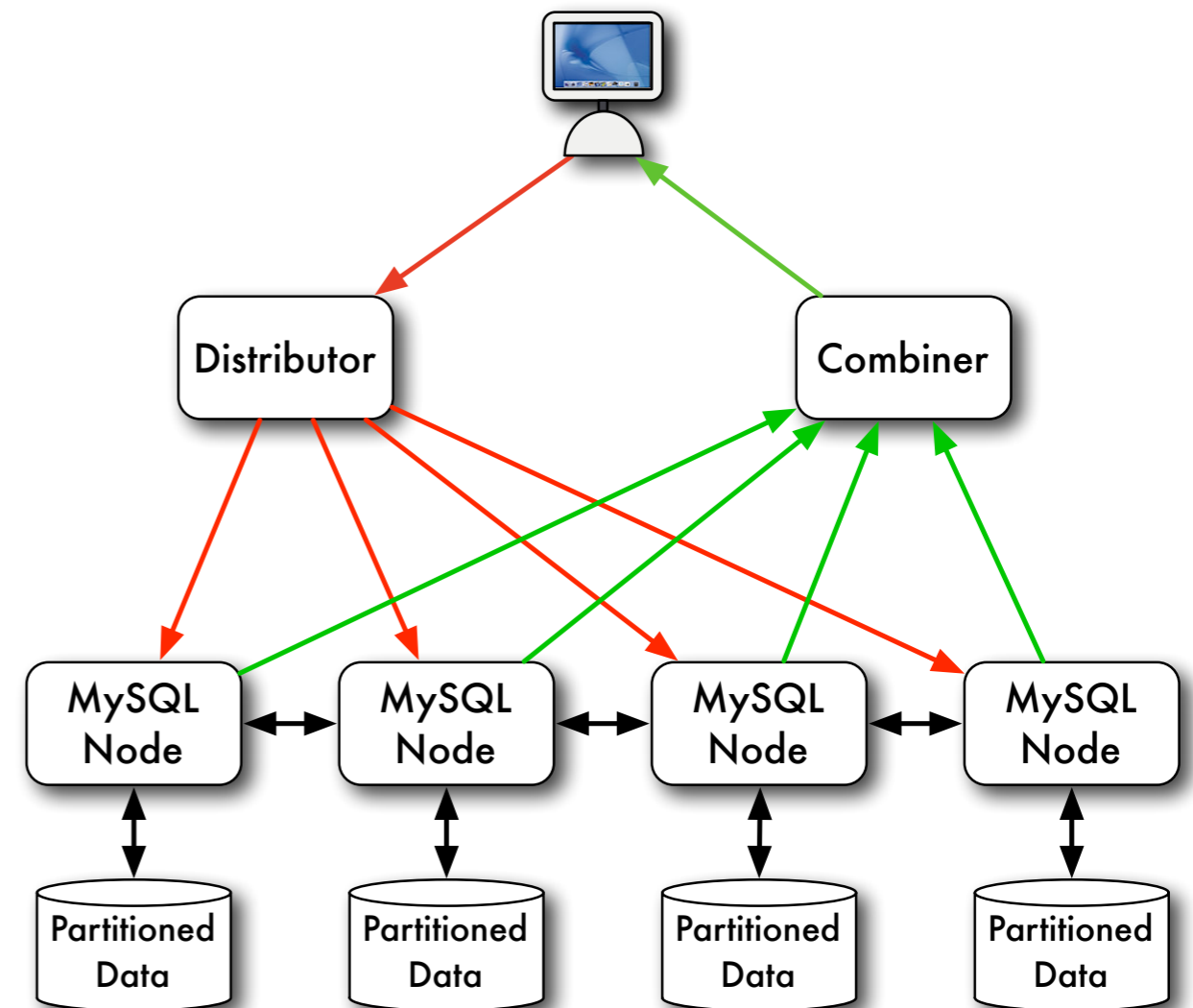
(includes elasticity, reliability)

Affordability

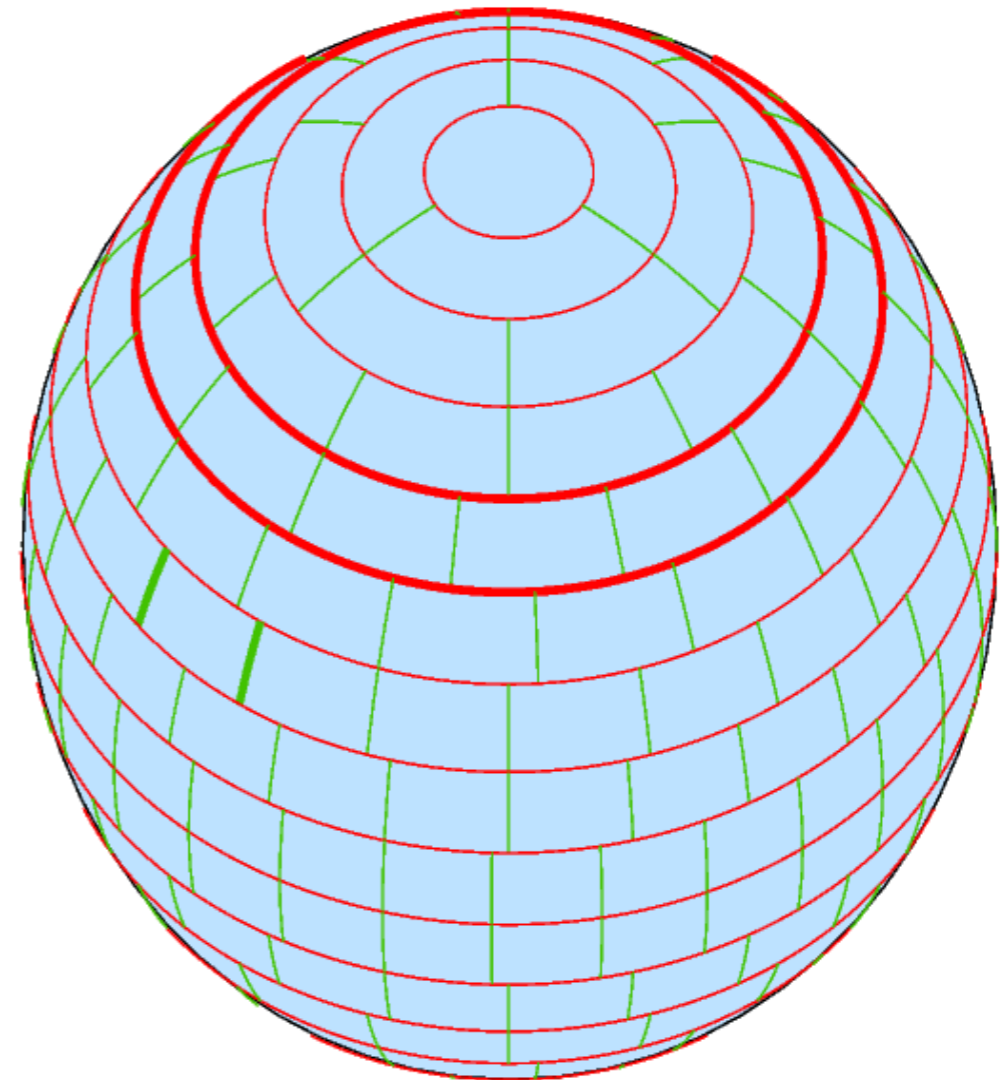
Queryability

Manageability

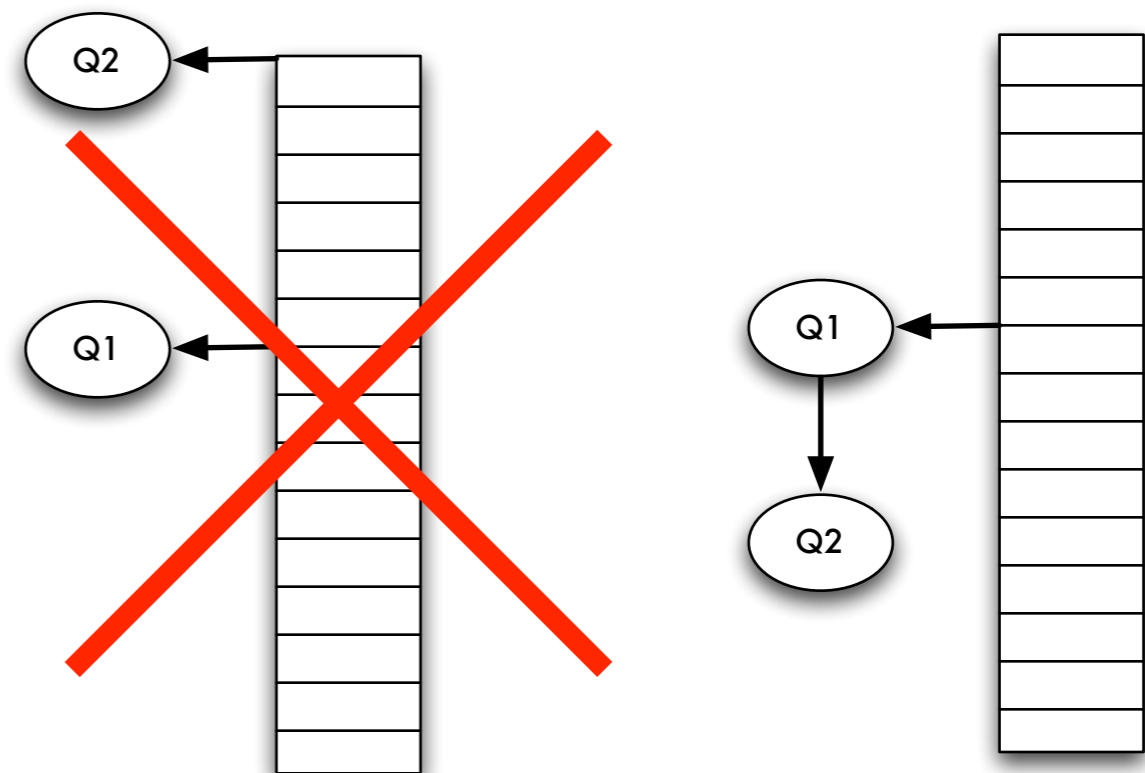
Map/Reduce on RDBMS



Overlapping partitions

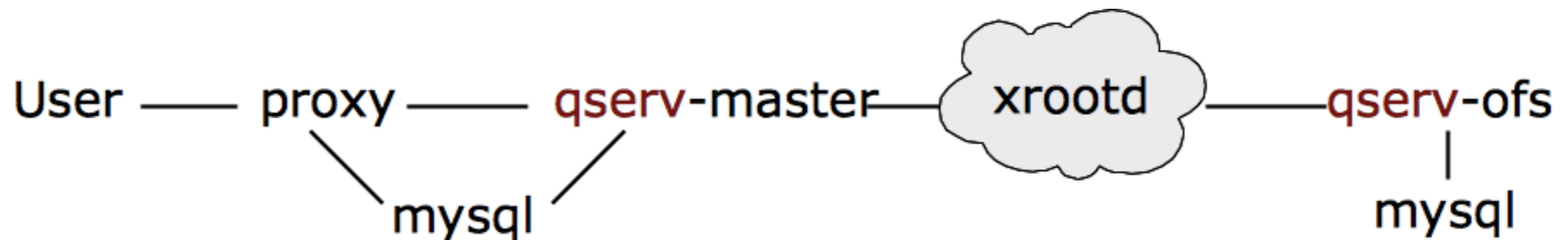


Shared scans



qserv Implementation

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



150 node cluster

Showed query functionality

Showed scalability

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



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



Preparing for petabytes