

LSST Data Management Working Group Report

Flexible Architecture

Flexible Architecture WG Membership

- Rob Pike (Google, chair; r@google.com)
- Ghaleb Abdulla (LLNL abdulla1@llnl.gov)
- Jim Brase (LLNL brase1@llnl.gov)
- Rahul Dave (U Penn rdave@cis.upenn.edu)
- Marcus Miller (LLNL marcusm@llnl.gov)
- Michelle Miller (NOAO michelle@noao.edu)
- Ray Plante (NCSA rplante@ncsa.uiuc.edu)
- Frank Valdes (NOAO fvaldes@noao.edu)

Where do we need to be at the end of R&D?

- What's needed:
 - Definition of data products and the algorithms to produce them
 - Interface specifications, internal and external
 - Specification for self-diagnosis and repair
 - Simulated data available
 - Pipeline simulator built, used for evaluating prototypes
 - Scheduling algorithms designed and simulated
 - Reference platform defined

Issues Addressed

- Control is separate from data
- Same infrastructure for all pipelines
- Resource allocation, advertisement
- Scheduling: interleaving is key. R.A. too.
- Remoteness and scale => self-maintenance
- Handling failure: unified architecture
- Reliability, redundancy, etc.
- Flexibility over lifetime: algorithms change; how does dataset adapt?
- Outsiders injecting pipelines and elements.

Unresolved Issues

- Systems architecture questions (next slide)
- How much redundancy do we build into processing hardware, esp. on mountain?
- What is latency requirement for transient detection?
- Resource management in an architecture of intercommunicating pipelines
- Model for sharing info between simultaneous pipelines.
- How to isolate pipeline model from data model.
- How to handle restart: architecture & policy.

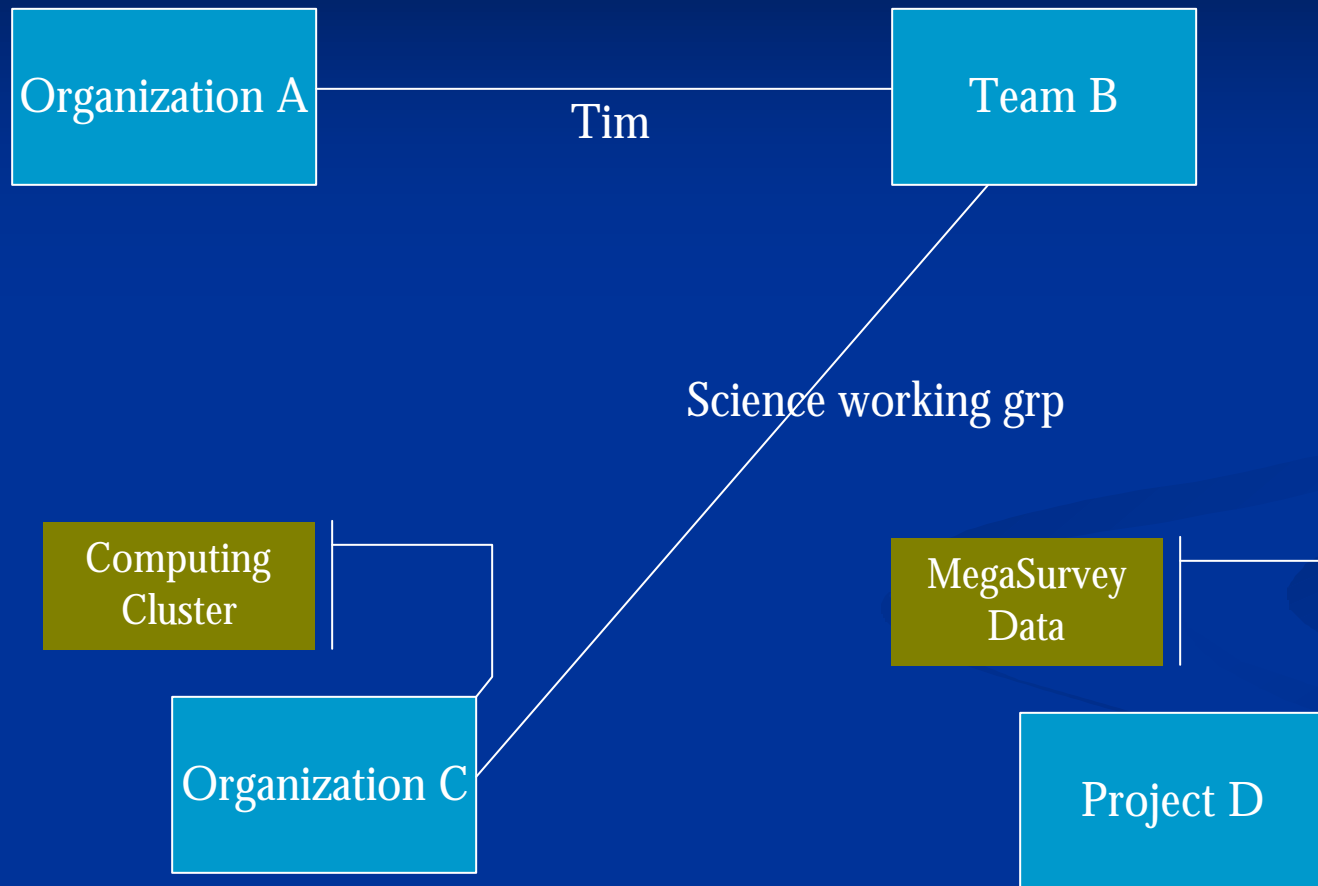
Distributed System Issues

- Overall: master/slave, peer to peer, GRID? Need self-repair.
- Scheduler: design (triggers, dataflow, msg pass)
- Scheduler: data contention
- Scheduler: resource management
- Authentication, security, verification
- Robustness and safety
- Feedback: quality of data, hardware failure, data failure, etc.

Team Makeup & Relationships

- Players: LLNL, NOAO, U Penn, NCSA, Google
- Collaborations: TAOS, MOSAIC, IDES, Super MACHO, ...
- Extant Pipelines: Space Telescope, SDSS, CFH, QUEST, VISTA, ...
- Need more distributed systems researchers.
- (If there are multiple collaborations working in this area, how can they support each other, and the overall goals of this LSST subarea?)

Resource Map Example



Needed Resources

- What's needed to get the R&D job done?
 - Money for salaries
 - Computing facilities
 - Access to survey data streams
 - Experts in distributed systems software
 - Process to coordinate R&D with LSST design
 - Simulator & challenge problem
 - A follow-up workshop to focus on this issue
 - ...
- Any plans already in place to get these resources?
 - Lots of in-kind resources; these should be developed

R&D Tasks / Milestones

- Given all of the above, try to rough out a plan for this area
 - Problem is too unconstrained
 - Create a (formal) standing working group
 - Workshop to create strawman design by 1/05
 - “Prototype approaches 1 and 2 by 6/05”
 - “Downselect to a single approach by 9/05”
 - “Design ready for construction phase by 2/06”