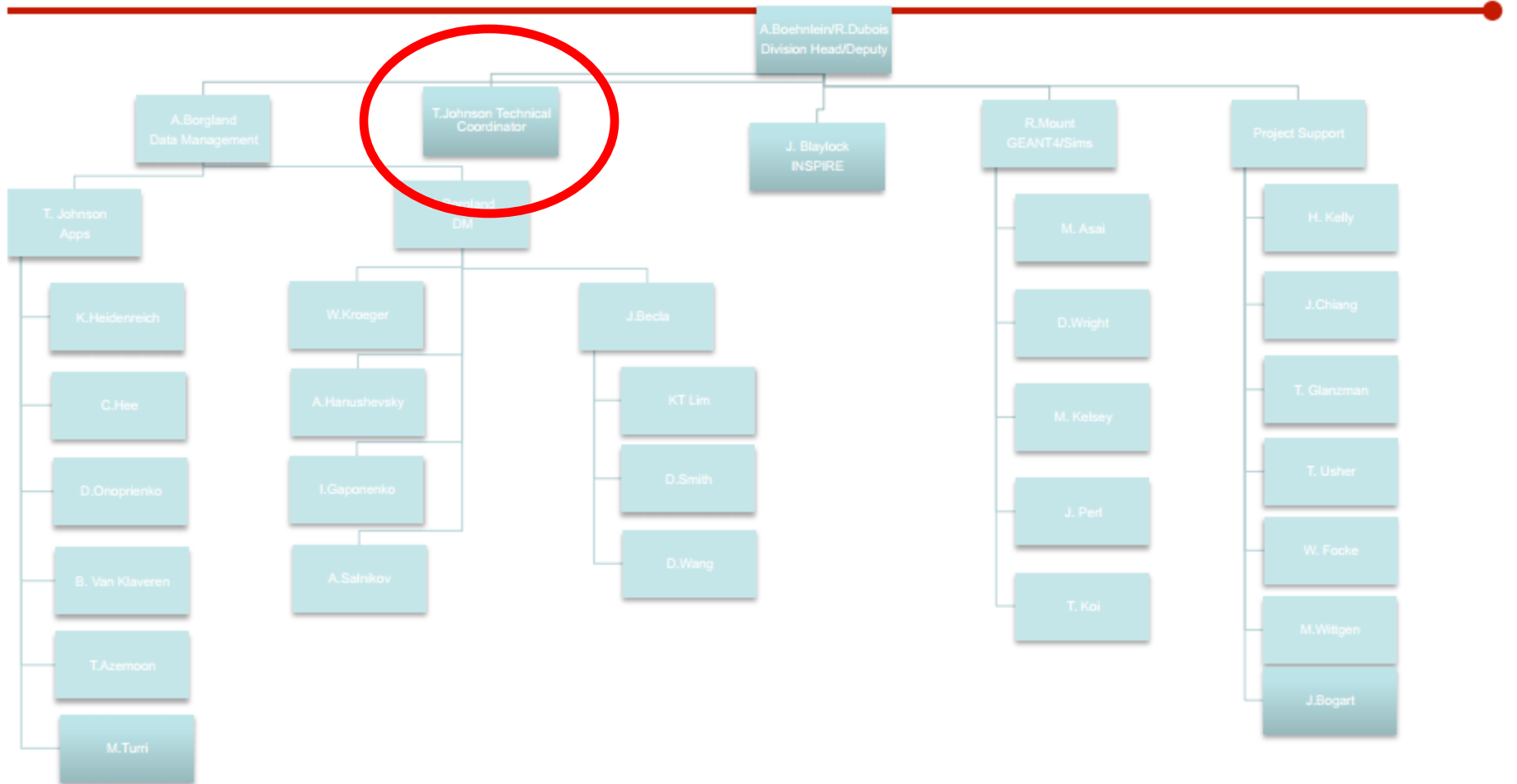


# SCA “Technical Coordination” Goals and Plans

Tony Johnson

May 18 2012

# SCA Org Chart



# Goals

- Encourage development of SCA as a lab-wide resource for scientific computing
  - After starting as a group of people doing scientific computing for many different projects at the lab
    - Identify existing projects which can be used beyond their initial scope
    - Identify developing technologies which we can harness for use by multiple groups
    - Develop collaborations with other groups at SLAC and broader science/industry community
- SCA should be seen as a leader and “partner of choice” for scientific computing at the lab
  - A natural place to go for expertise and support

# Organization

- SCA “Tech Coord” is just me
  - Idea is to look at work being done in entire SCA division (and computing in the lab as a whole)
    - Identify places where cross-department projects make sense
  - To get anything done will need to pull expertise and effort **and ideas** from entire SCA division

# Looking at Existing Projects

- Developing criteria/metrics for project support levels within SCA
  - Take into account need to support existing experiments/projects vs need to develop new initiatives
  - Support single experiment/projects vs set of SLAC based projects vs extended community
  - Internal funding vs external funding
- Understanding existing and planned projects within SCA
  - Including projects which SCA could/should be involved in
  - Try to identify common requirements which can be "factored out"
    - Maximize reuse, eliminate unnecessary/unintentional duplication
- Defining technical direction for new projects
  - Attempt to balance leveraging of existing skill-sets vs. understanding and exploiting new industry trends
  - Need people's input on what we should be doing
  - Form "task-force" to evaluate some new technologies
- Developing targets of opportunity for new projects and collaborations
  - Set of projects ready when new funding opportunities present themselves
  - Developing priority list for existing/new projects
- Creation of (new) project teams
  - Possibly bring people together from different parts of SCA, and people from outside of SCA

# Example Project Summary

- Who/What/When
- Technologies Used
- Status
- Future plans and opportunities
- Needs to be kept up-to-date
  - by a project leader?
- Links to publications, web sites, related projects

<b>Name:</b>	Automated processing Pipeline (aka Fermi Pipeline)
<b>What:</b>	<p>A system for automatically managing arbitrary graphs of processing jobs. Major features:</p> <ul style="list-style-type: none"> <li>• Automated submission and monitoring of batch jobs with high reliability</li> <li>• Maintains full history of all data processing</li> <li>• Ability to parallelize processing of subgraphs of jobs</li> <li>• Ability to embed python scripts to perform simple computations between job steps</li> <li>• Ability to rerun jobs (whether successful or not)</li> </ul> <p>The pipeline consists of the core pipeline server, line mode client, web interface, and job control daemons. Currently supported batch systems: LSF, BQS, GridEngine, Condor.</p>
<b>Related Projects:</b>	Data Catalog
<b>Who:</b>	<p>Current SCA Developers: Tony Johnson, Brian Van Klaveren                  Past SCA Developers: Dan Flath, Charlotte Hee, Karen Heidenreich                  Key SCA Users: Tom Glanzman, Warren Focke</p>
<b>Used by:</b>	<ul style="list-style-type: none"> <li>• Fermi Gamma Ray Space Telescope for Prompt (L1) Processing, Monte Carlo, Science Processing, reprocessing</li> <li>• EXO for MC simulation, data processing, reprocessing</li> <li>• CDMS for MC Simulation at SLAC and SMU</li> <li>• CTA for MC Simulation at SLAC</li> </ul>
<b>Status:</b>	Stable/Supported. Pipeline core designed for use by experiments and projects at SLAC, use of pipeline core beyond SLAC could be considered if a strong use case is discovered. Supported for use by current and new SLAC projects. Job Control daemons can be installed at remote sites to allow jobs to be submitted from SLAC to other sites. No major new features are currently planned but incremental improvements, bug fixes and minor feature requests will be supported.
<b>Planned Work:</b>	<ul style="list-style-type: none"> <li>• Extension to support submission of jobs to Grid via Dirac. This work is being performed mainly by Fermi and CTA collaborators in Europe.</li> <li>• Performance improvements</li> <li>• Completion of job cancellation features.</li> <li>• Support for "split mode" (jobs within a task being submitted to different sites)</li> </ul>
<b>Possible Future Work:</b>	<ul style="list-style-type: none"> <li>• Remove dependency on oracle.</li> <li>• More performance improvements.</li> <li>• More interactive web interface.</li> <li>• JSON interface for improved integration with languages like Python.</li> <li>• Better job throttling (limit max number of running jobs, based on resource usage).</li> <li>• Handle more batch job types, including perhaps Open Science Grid.</li> </ul>
<b>Last Updated:</b>	May 2012 by tonyj

# Sample New Project Ideas (not necessarily good ideas)

- General Computing (in collaboration with computing division)
  - Improved support for developer tools (Subversion, GIT, Fisheye, Hudson/Jenkins, ...)
    - Developer “journal-club” to bring together software developers from across the lab
  - Interactive web based scientific computing status page
  - Better integration between Crowd, Group Manager, Single sign-on (webauth?)
  - Account management
  - mathjax plugin for confluence
  - LSF with virtualization
- Data Management
  - Data Portal for Photon Science (and Astro physics)
    - Best features from Fermi data catalog, JCSG, CXIDB, ICAT, Dirac, ...
  - Why stop at presenting data location via the web?
    - Should be able to drill into and operate on data
    - In a collaborative way
  - Web based scientific data visualization tools
    - Extension/compatible with Google Visualization API
  - Virtual Observatory integration
- <your-ideas-belong-here>

# Possible Technologies

- Technology selection criteria
  - Lifespan/stability
  - Adoption/Support
  - Popularity/Mindset
  - Language neutrality/Interoperability
- Web/HTML5/GWT/Dart/JSON
- Google visualization API
- C++/Java/Python/JavaScript/Scala/Lua/...
- HDF5/Fits
- IPython/IPython Notebook
- Virtual Observatory
- MatLab
- LSF virtualized cluster support
- <your-favorite-technology-belongs-here>



# Scientific Computing Status Display

- **Features/Requirements**

- Status display should be an interactive web application, that can be run on a desktop or on a captive display (e.g. in the computer center lobby)
- Emphasis should be on scientific computing -- but should include some infrastructure status
- Web application will for the most part fetch data from web services/databases provided by other groups
  - E.g. Networking, Nagios, LSF, planned outages, message of the day, ...
  - Not all features need to be ready to get started
- Status display should be extensible, easy to add new modules
- Since web based can include links for more detailed information
- Could potentially include modules for display of experiment/project info (e.g. LCLS, Fermi, EXO, ....)
- Could run offsite -- but is there any point?
- Security concerns -- do we need to limit access (would be much better if this could be avoided)?

- **Bonus Features**

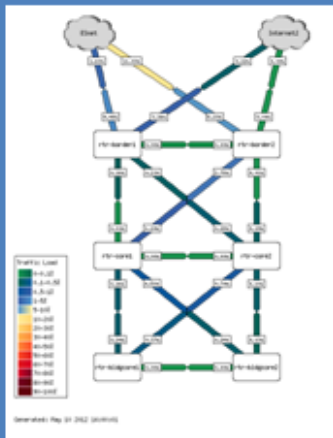
- Will involve bringing together information from many groups
- One off project – can experiment with new web technologies



# Scientific Computing Status Display

## Scientific Computing Status

### Network



### Batch System



Host Metrics: batch 2025, server 860, desktop 278,  
compute 51, fileserver 49, interactive 7

Upcoming Outages

Message of the day

Nagios Summary

LCLS

Fermi  
Last data delivery: hh:mm  
Gamma Rays Delivered: 1,234,456,789

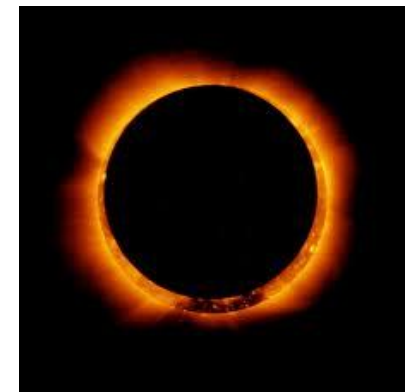
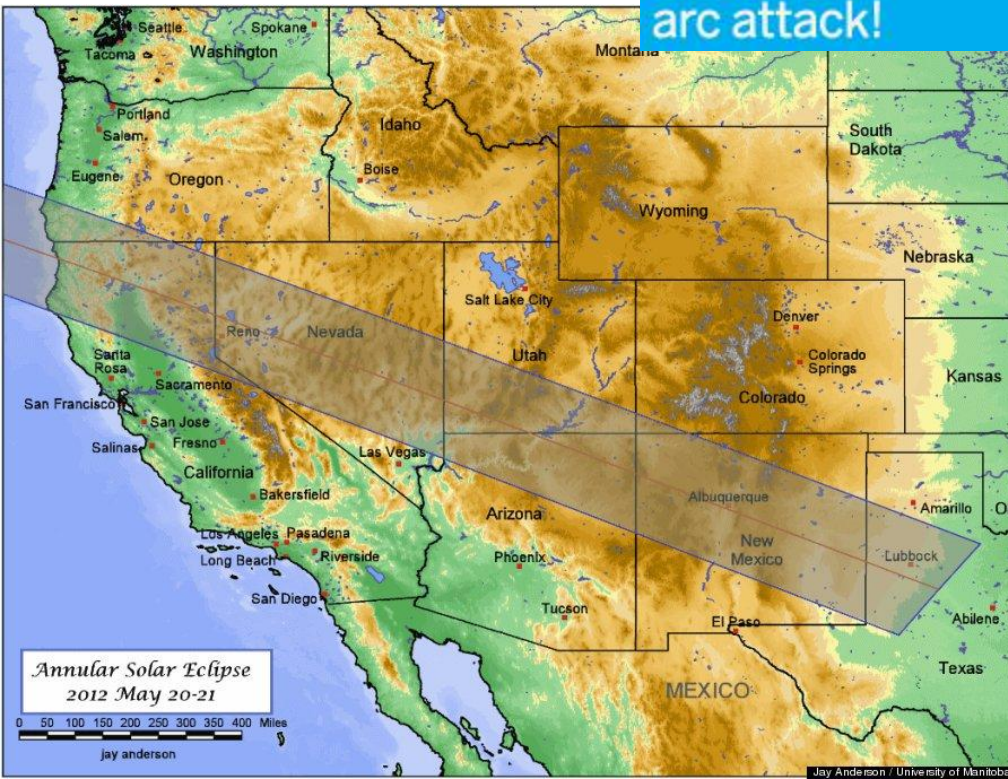
# Next Steps

- Start collecting project templates
  - Combine them onto confluence SCA “portfolio” page
- Short term projects
  - Status Display
  - Developer tools and developer forum
- Medium term projects
  - Look for potential projects and collaborators
    - CHEP next week

Have a good weekend...

**Maker Faire**®

electricity + music =  
arc attack!



**Location:** San Francisco, CA

**Timezone:** Pacific

**Eclipse Begins (Local Time):** 5:16

**Maximum Eclipse:** 6:33

**Eclipse Magnitude:** 0.897