Open Science Grid

OSG Status & Accomplishments

Kent Blackburn
California Institute of Technology

Open Science Grid Joint Oversight Team Meeting
February 20th 2007

Open Science Grid

Outline

- OSG Funding
- Status of the OSG project & organization.
- Status from the Resources Managers (of which I am one).
- Milestones to date.
- Usage of the infrastructure.
- More specifics on usage, plans, technologies etc. in later talks.

Open Science Grid

OSG Funding

- OSG is funded by 6 program offices - DOE ASCR, DOE HEP, DOE NP, NSF MPS, NSF OCI, NSF OISE.
- Broadly speaking the funding supports (Ruth will cover the details):
  - OCI : Engagement of new communities
  - ASCR: CS aspects of Software (VDT, Security, Storage), Troubleshooting and Integration, Training and outreach.
  - NP: Security and Integration
  - OISE: Outreach, Interoperability - with Scandinavia.
  - HEP/MPS: Operations, Software (VDT), Extensions, Security, Integration, Administration and Communications.

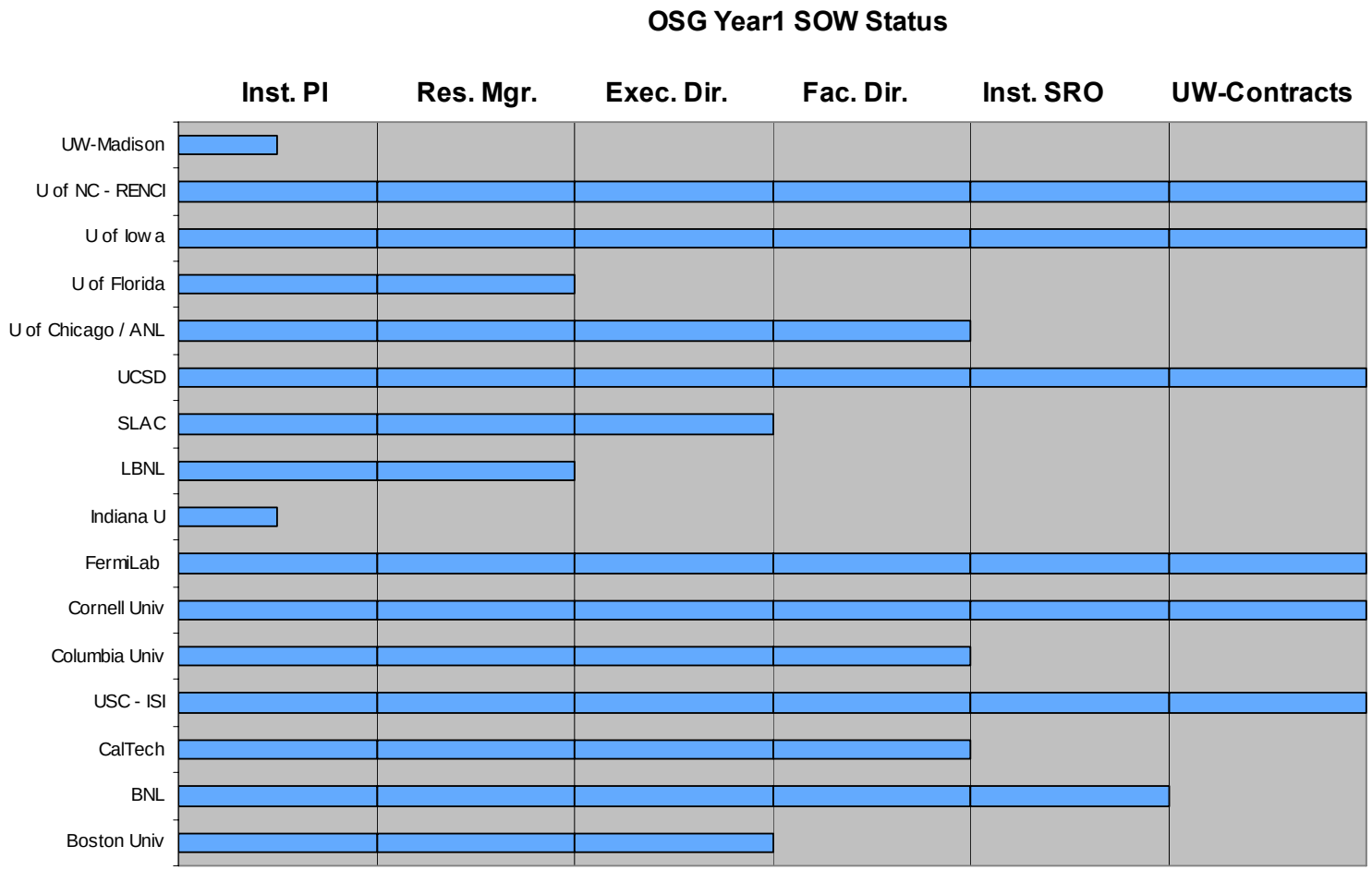
OSG Project Organization

- Major focus of first six months of funded project on transitioning from an “ad-hoc” collaboration to a managed project.
  - Better understood roles and responsibilities in the management chart
  - Filled in missing roles e.g. User Support Coordinators; Identified other holes e.g. Site Coordinator.
  - Agreement on Year 1 Project Plan, WBS, Change control process.
  - Adjusted effort and funding profile based on need - e.g. ISI for LIGO deliverables.
  - Wrote Statements of Work (SOWs) for 16 institutions

Statements of Work

- Major accomplishment to get details of commitments and deliverables agreed to and “tied” to the WBS.
- Process through the signatories much slower than we would have guessed.
- Major changes by Institution SROs slows things down at the end:
  - Caltech
  - Reaction to “open source software” statement.
  - Staff turnover.

Status of Statements Of Work

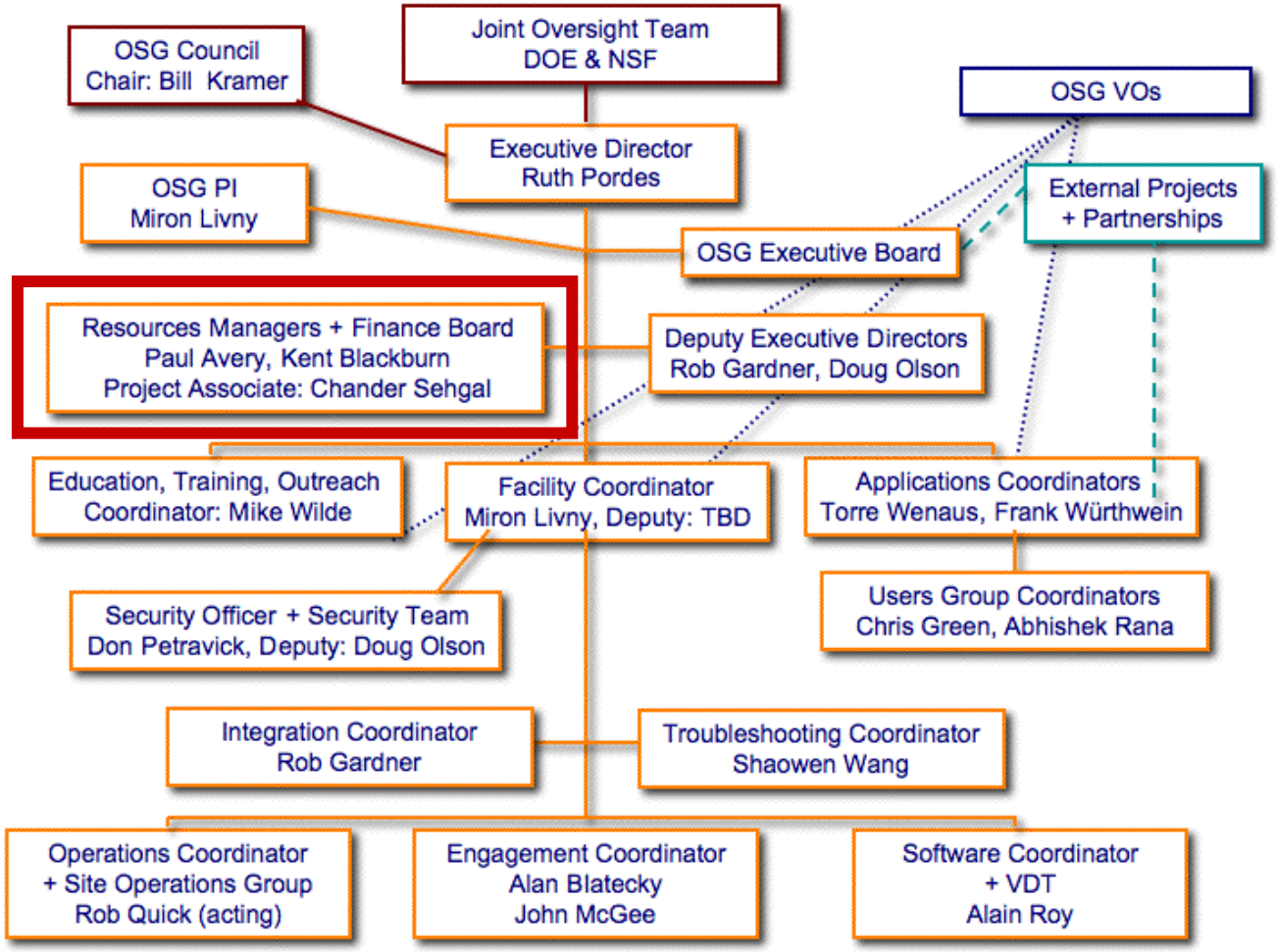


Progress on the Signature Trail

Resources Managers

- “Resources Manager and Finance Board – manages all matters related to the costs and resources of the OSG program of work.”
- Advises the Executive Director
- Handles all change requests.
- Responsibilities cover Consortium and External Project contributions also.
- Responsibilities include agreements and OSG (VO) allocations.
- Responsibilities helped by 1 FTE project staff- Project Associate.

ORGANIZATION CHART



Finance Board

- "The Finance Board evaluates contributions, including contributions pledged from the stakeholder organizations, for the relations with the funding agencies, for contract policies, and for all administrative matters. The Resources Manager is chair of the Finance Board."
- All important decisions are reported to the Executive Board and the OSG Council.
- Requests for a change in the financial allocations can be made to the Finance Board by the Executive Board, stakeholder organizations, or project resource managers…, a Resource Change Proposal is submitted to the Executive Board for approval. Once agreed to by the Executive Board it is confirmed by the Council.

  Initial change requests are for funding for EOT presentation at Taiwan Grid Conference (agreed to) and for $5K hardware for EOT training (in discussion).

Change Control Process

1 The OSG Change Control Process

The purpose of this document is to describe the process and thresholds for changes in the financial, schedule or technical scope and schedule for the OSG project.

This process is governed by the Project Executive Plan [OSG-408] which says: "Once the project plan is baselined through internal and external reviews, changes to the technical program, cost and/or schedule must be submitted in writing. The Resources Manager gathers input from the project management, participants and the user community and makes recommendations on the feasibility, cost, and impact of proposed changes. Changes are agreed by the Executive Director and endorsed by the Executive Board and Council. The documentation generated in considering any accepted request will be filed with the Council. "

That is the Resources Managers are responsible for working with the requestor of the change and the rest of the OSG to determine the impact and dependencies of a change in financial allocations, schedule and/or technical deliverables.

OSG implements a light-weight change control process where an agreement can be made rapidly, as needed, and there is a record of the decisions made and the impact and dependencies gathered. Change Control documents are kept in the OSG document repository, and may be accessed by a named list of OSG staff.

We define three levels of change control and explain the authorities assigned for each level:

Level 0	The change control process is not invoked.
Level 1	Area Coordinators have authority to decide and inform the Resources Managers after the decision.
Level 2	Resources Managers give input to the Executive Director who decides and then obtains the agreement of the Executive Board and endorsement of the Council. The Joint Oversight Team is informed of the change.

Area Coordinators are: Facility Coordinator, , Education Coordinator, Applications Coordinators, Resource Managers, Executive Director.

Significant Change to the Technical Baseline: This is defined by consensus and discussion within the weekly meetings of the Executive Team (ET) . Any member of the ET or Council can request that the issue be addressed by the Resources Managers as a Level 1 or Level 2 change. All changes in $ allocations require the signature of the Resource Managers, Executive Director and OSG PI.

Changes to the Change Control Procedures are recommended to the Executive Director, agreed to by the Executive Board and endorsed by the Council.

1.1 Level 0 Threshold

The Change Control Process is not invoked. This applies for changes which do not trigger the Level 1 or Level 2 threshold. In these cases the change is recorded by the area coordinator (Facility, EOT, or Applications Coordinators or Executive Director (through effort, area or institutional reports) but no signatures are required. Use cases include reallocating funds between activities within an institution, reassigning effort within an institution that are below the thresholds below.

1.2 Level 1 Change Control Threshold

The following changes will trigger a Level 1 Change Control Process:

- Schedule slip of more than 3 months.
- Movement of >$15K within an institution or any change in $ allocation to an institution .
- Significant changes to the technical baseline.

1.3 Level 2 Change Control Threshold

The following changes will trigger a Level 2 Change Control Process:

- Schedule slip of more than 6 months.
- Change of 20% of the currently allocated funding to an institution or activity.
- Significant changes to the technical baseline or changes in the OSG performance and scope metrics.

2 Change Control Request and Agreement Template

TYPE OF CHANGE: [ ] SCHEDULE [ ] BUDGET [ ] SCOPE [ ] SOW

If Funding Change:

Current Annual Funding Amount:

Requested Change to Annual Funding Amount:

DESCRIPTION OF CHANGE

Take as much space as required here.

REQUIRED SIGNATURES:

ORIGINATOR *(name)* _____ DATE: _____

WBS Level 3 Coordinator *(name)* _____ DATE: _____

SIGNATURES REQUIRED IF CHANGES EXCEED THRESHOLD LEVELS

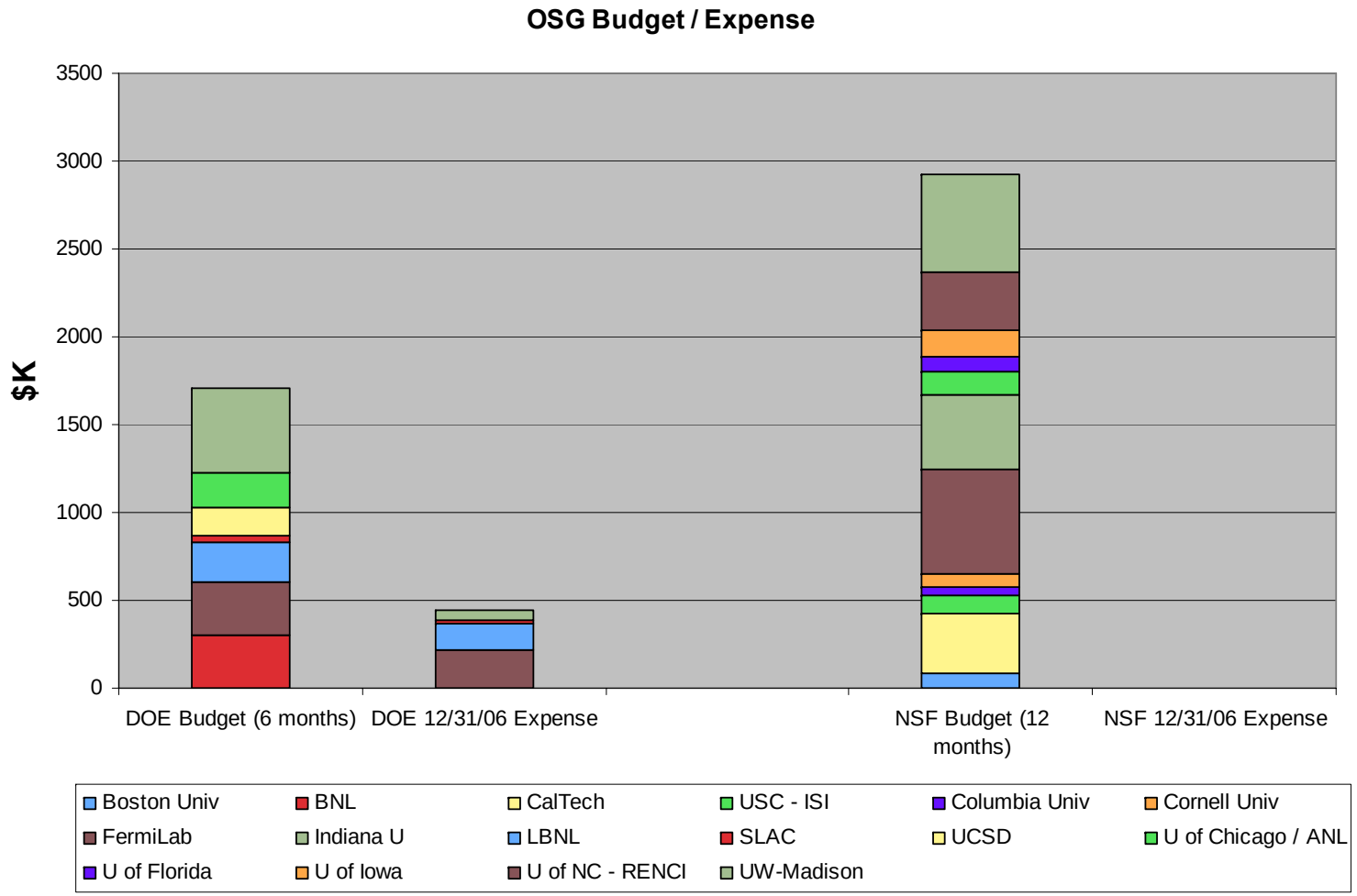
- Resources Managers *(Paul Avery, Kent Blackburn)* _____
- Executive Director *(Ruth Pordes)* _____
- OSG PI (Miron Livny) _____

3 List of Year 1 Milestones Under Change Control

Finance Board, Finances

- First Finance Board meeting Dec 14th Caltech; Minutes are open
- Agreed on ~Quarterly meetings + as-needed. Next meeting in April + some time at the all hands meeting.
- Acknowledge very useful guidance and input from Jim Yeck to work of the Resources Managers and Finance Board.
- Tracking expenditures quarterly.

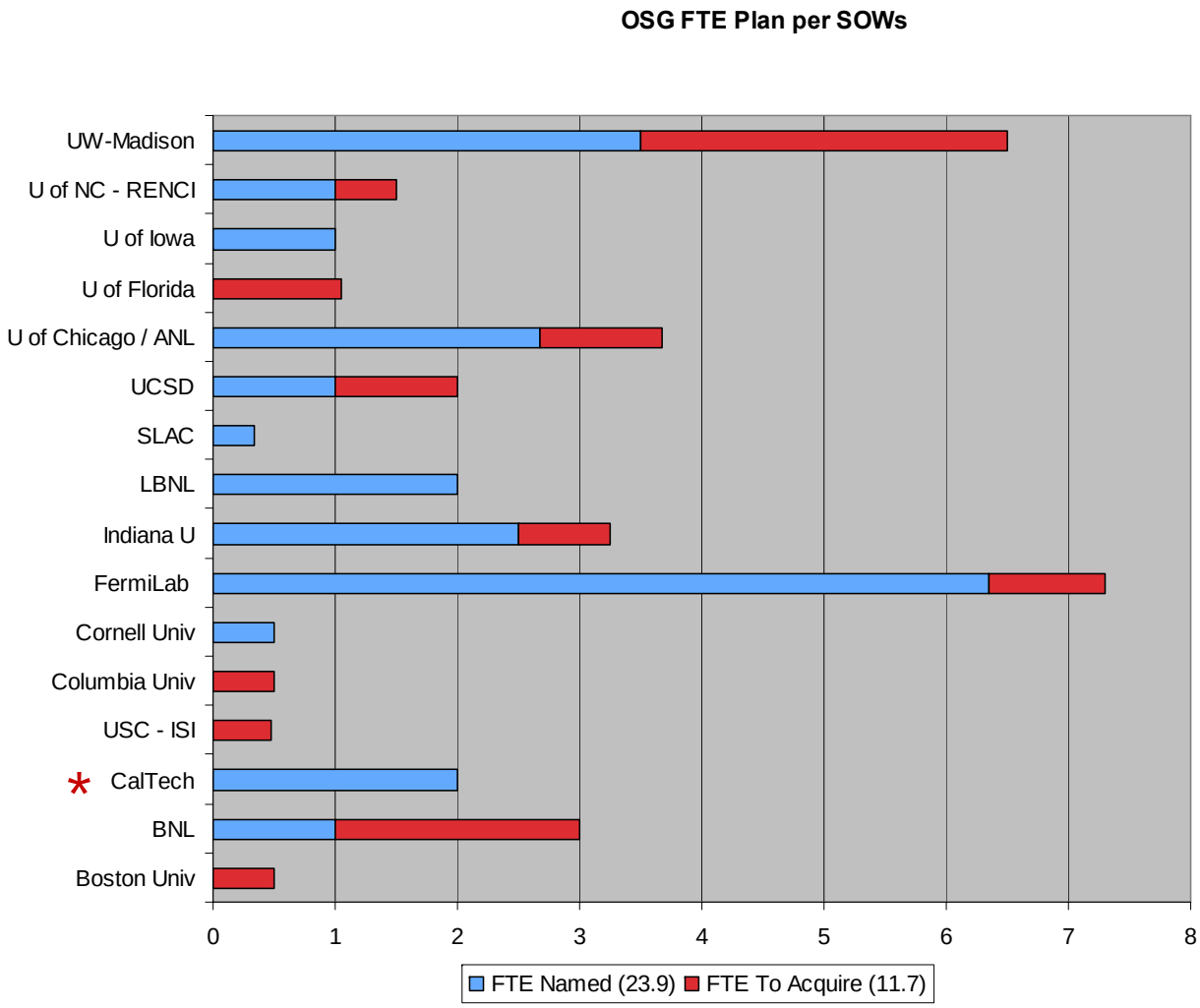
EXPENSES AS OF 12/31/06



Reporting Effort & Accomplishments

- SOWs include agreement to monthly individual effort reports and quarterly Institutional PI reports (as well as area coordinators).
- Reports posted.
- Effort being tracked.
- Six month project report submitted to Agencies.
- Action item to rationalize reporting into useful and ongoing structure.

Status of FTE Availability



*** Caltech will be replacing its existing two staff (*not shown in count above*)**

Status of High Level Milestones to Date - From the Project Plan & WBS

- Milestones exist and are worked to.
- Reasons for missing them get discussed and feedback into re-planning.
- More detail in Ruth's talk.

WBS	Name	Date
1.1.1.2	Define Operational Metrics for Year 1 **Draft under review**	1/1/07
1.1.3.1.1	Release Security Plan √	1/1/07
1.1.5.2.3	Release OSG 0.6.0 **Provisioning and final testing in progress**	2/27/07

Overall Status of the OSG Infrastructure

- Sustained use of the infrastructure continues with the throughput in number of jobs, data, and CPU×hours/day increasing (with hiatus over new year).
- We are working on improving validation tests to ensure VOs are indeed supported at sites that advertise such support.
- In general the production infrastructure Runs Stably

VIRTUAL ORGANIZATIONS (2/1/07)

- Collider Detector at Fermilab (CDF)
- Compact Muon Solenoid (CMS)
- *CompBioGrid (CompBioGrid)*
- D0 Experiment at Fermilab (DZero)
- Dark Energy Survey (DES)
- *Distributed Organization for Scientific and Academic Research (DOSAR)*
- *Engagement (Engage)*
- Fermi National Accelerator Center (Fermilab)
- *Functional Magnetic Resonance Imaging (fMRI)*
- Geant4 Software Toolkit (geant4)
- *Genome Analysis and Database Update (GADU)*
- *Georgetown University Grid (GUGrid)*
- *Great Plains Network (GPN)*
- Grid Exerciser (GEx) (GridEx)
- Grid Laboratory of Wisconsin (GLOW)
- *Grid Research and Education Group at Iowa (GROW)*
- *Group Researching Advances in Software Engineering at Buffalo (NYSGrid)*
- *Interactions in Understanding the Universe Initiative (i2u2)*
- International Virtual Data Grid Laboratory (iVDGL)
- Laser Interferometer Gravitational-Wave Observatory (LIGO)
- *nanoHUB Network for Computational Nanotechnology (NCN) (nanoHUB)*
- *Northwest Indiana Computational Grid (NWICG)*
- *Open Science Grid (OSG)*
- OSG Monitoring Information System (MIS)
- OSG Operations Group (Ops)
- Sloan Digital Sky Survey (SDSS)
- Solenoidal Tracker at RHIC (STAR)
- United States ATLAS Collaboration (USATLAS)

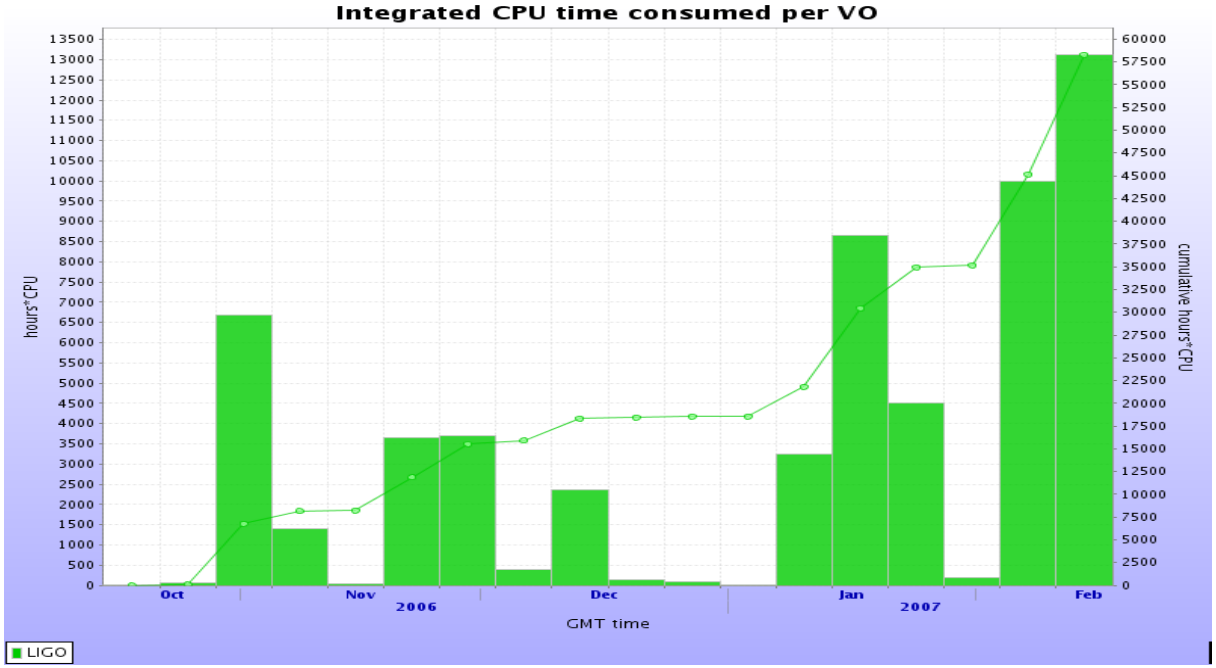
•*Non-physics*
•Partner Grids
•VO crosses OSG & TeraGrid

US ATLAS, US CMS, WLCG

- Torre's talk gives
  - the ATLAS and CMS accomplishments.
  - Network and data movement use (LHC & LIGO)
- The US LHC management stated in their Jan review that OSG has met the throughput and performance needs for the 2006 data challenges, and has performed above the level expected by the fractional US contributions to their whole distributed systems.

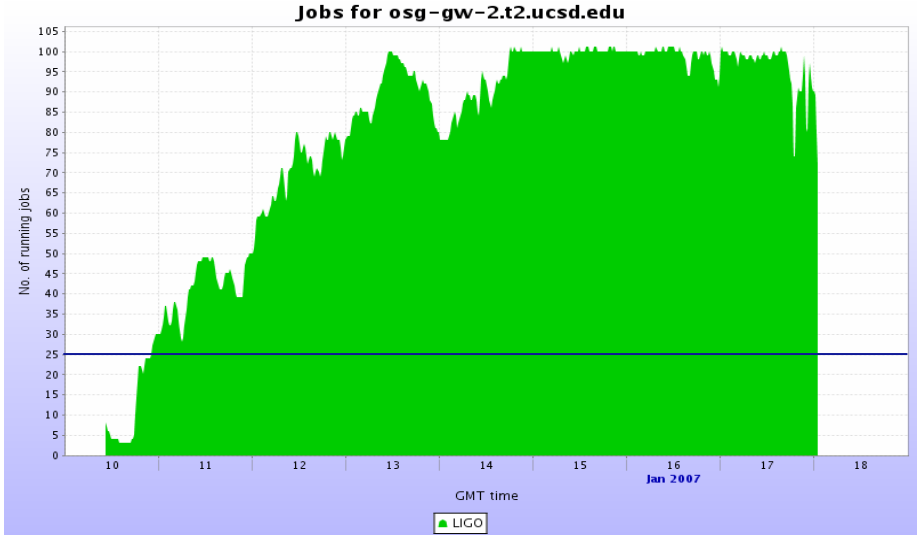
LIGO

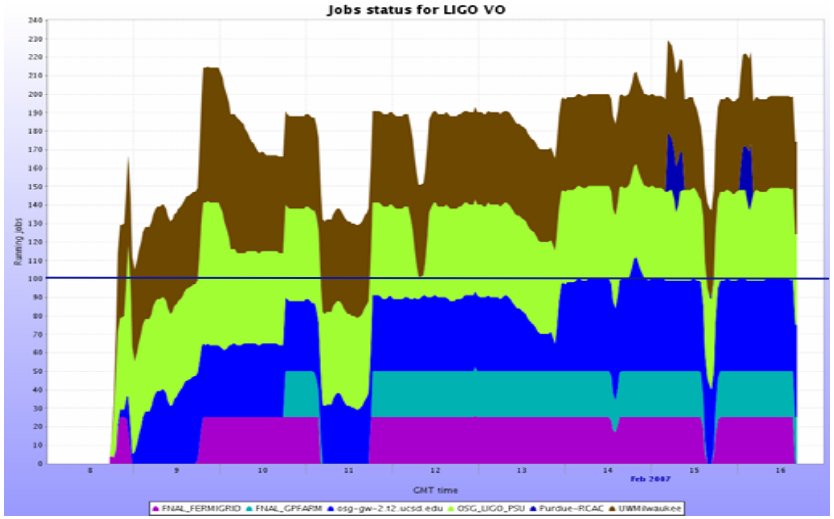
- Making steady progress migrating in-house applications onto the OSG Grid
- Working closely with Pegasus developers at ISI to enhance workflow planning in support of LIGO workflows on the OSG
  - Pegasus workflow planner enhancements acknowledging OSG accepted for publication
- Seeing steady increases in utilization of opportunistic cycles on the OSG



LIGO Milestones

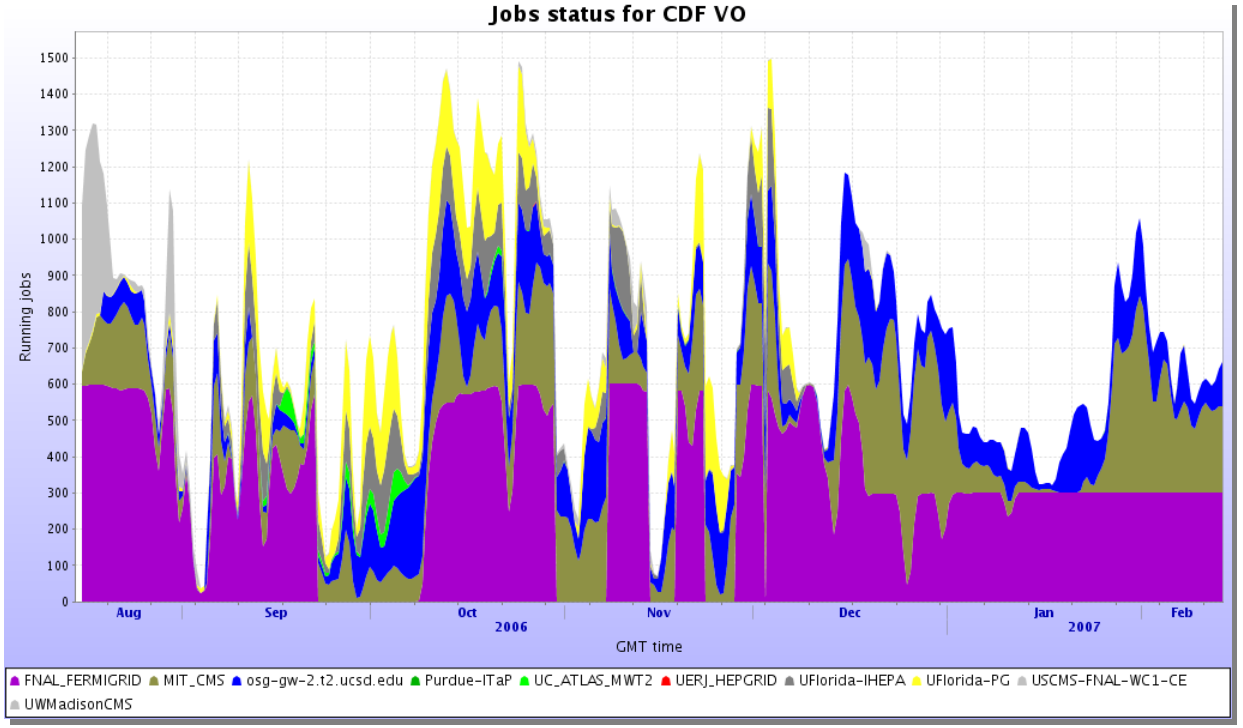
- Opportunistic use of OSG adapting LIGO workflows.
- Completed First LIGO Milestone in January.
  - Sustained workflow at >25 job slots for one week at UCSD Tier-2 site.
  - Joint OSG collaborative project with LIGO, DISUN, ISI.
  - Exercised newly provided Pegasus functionality provided by USC-ISI.
- Just completed the second milestone well ahead of June timeline.
  - Sustained multi-site workflows at >100 job slots for one week .





CDF

- CDF continuous using of OSG for monte carlo production and acknowledge the benefit to the scientific output.
- ~38 CPU x years a month average across 10 OSG sites, 5 of which are “opportunistic use”. (some/many dips due to monitoring problems?)



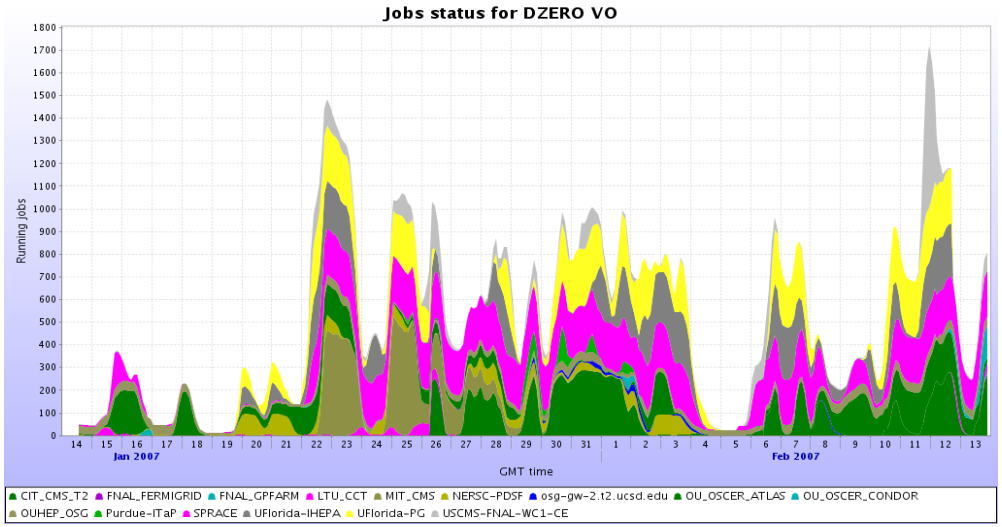
CDF first to agree on process for acknowledging OSG contributions

- Begin forwarded message:
- **From:** Ashutosh Kotwal kotwal@phy.duke.edu
- **Date:** February 9, 2007 9:52:12 AM.
- **Subject: Re: Citations of OSG use and benefit**
- Dear Ruth,
- This is a good idea.CDF has been using OSG resources heavily for some time for all MC production. Therefore I would think that most, if not all, recent CDF papers can be listed. Regards
- ,Ashutosh
- On Fri, 9 Feb 2007, Ruth Pordes wrote:
- Dear all,we have started a web page to list papers that rely on OSG science use and benefit. We will be reporting on this page at the upcoming Joint Oversight Team meeting in 10 days. http://www.opensciencegrid.org/?pid=1000077.
- Please could you send any such publications to the osg-webmasters to include on the page?Also, if you look the page includes the citation that we are asking groups to include on such papers. the CDF spokespeople have agreed to this. We appreciate your taking this up internally with your collaborations and experiments. Anne will also be talking to people about this at the all hands meeting.
- Citing OSG in your PublicationsWhenever you make use of Open Science Grid resources, services or tools, we would be grateful to have you acknowledge this use in any resulting publications.
- We suggest the following or similar text:"This research was done using resources provided by the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science."thanksRuth--

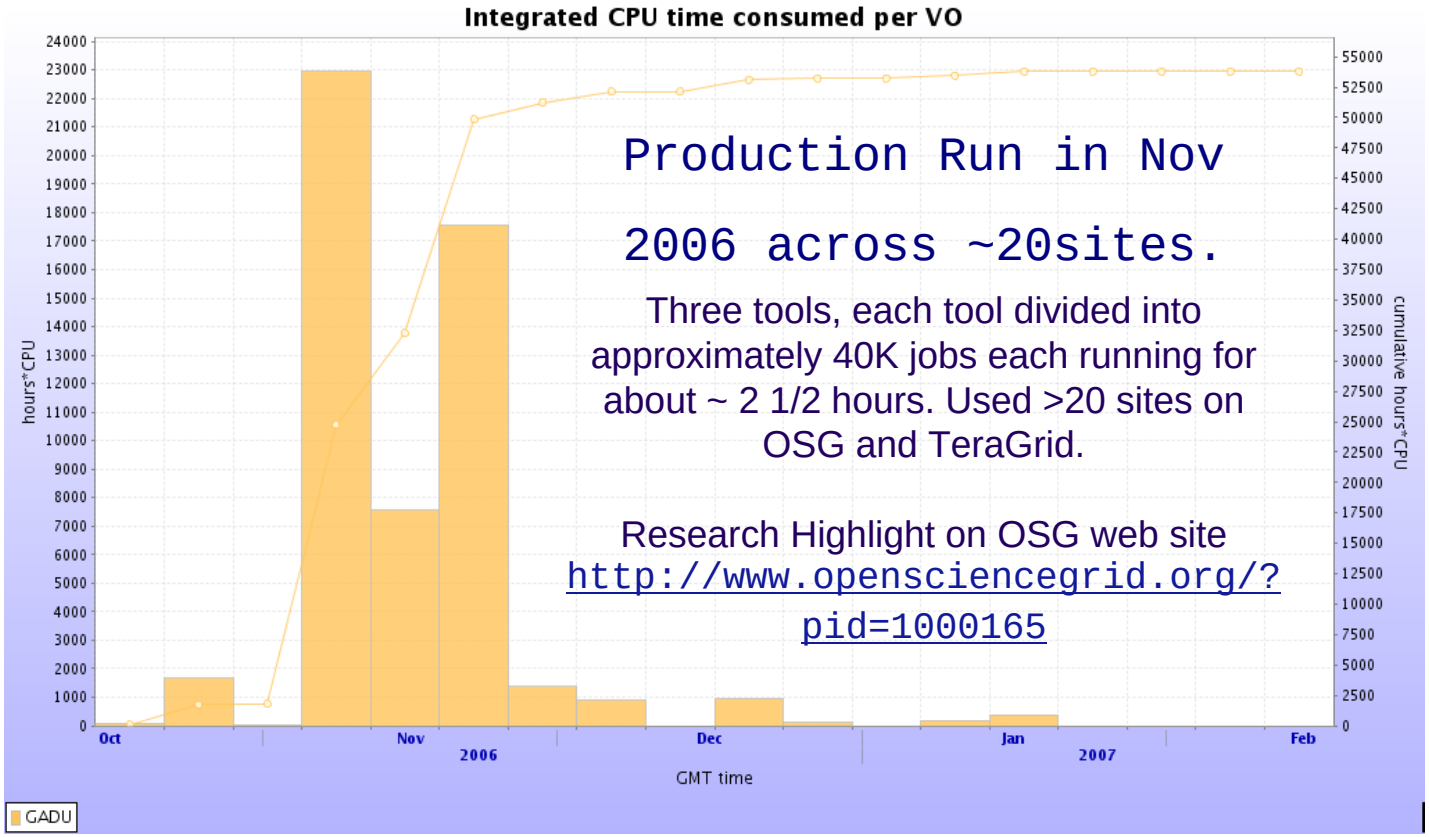
D0 Production Reprocessing - Looking for Maximum Throughput on EGEE and OSG

- *Message to Council:* D0 is starting a reprocessing run that will effectively be able to use between 1500 and 3000 CPUs over the Open Science Grid over three months. Requires:
  - Support for two types of jobs will be running production and merging, production jobs last 6-12 hours and merging jobs are slightly longer.
  - Worker nodes to have outgoing network access.
  - Worker nodes should have about 6GB of scratch storage space per job.
  - 4 TB of total disk space over the OSG to store input and output files. This space can be distributed among participating sites.
  - For prestaging of our raw data it would be desirable to have at least 1 TB of disk cache at each site connected with 1 Gbit link to all CPUs.

- Testing on 15 sites.
- Now running at ~60% efficiency,
- They will be conducting some error analysis soon and report findings to us.

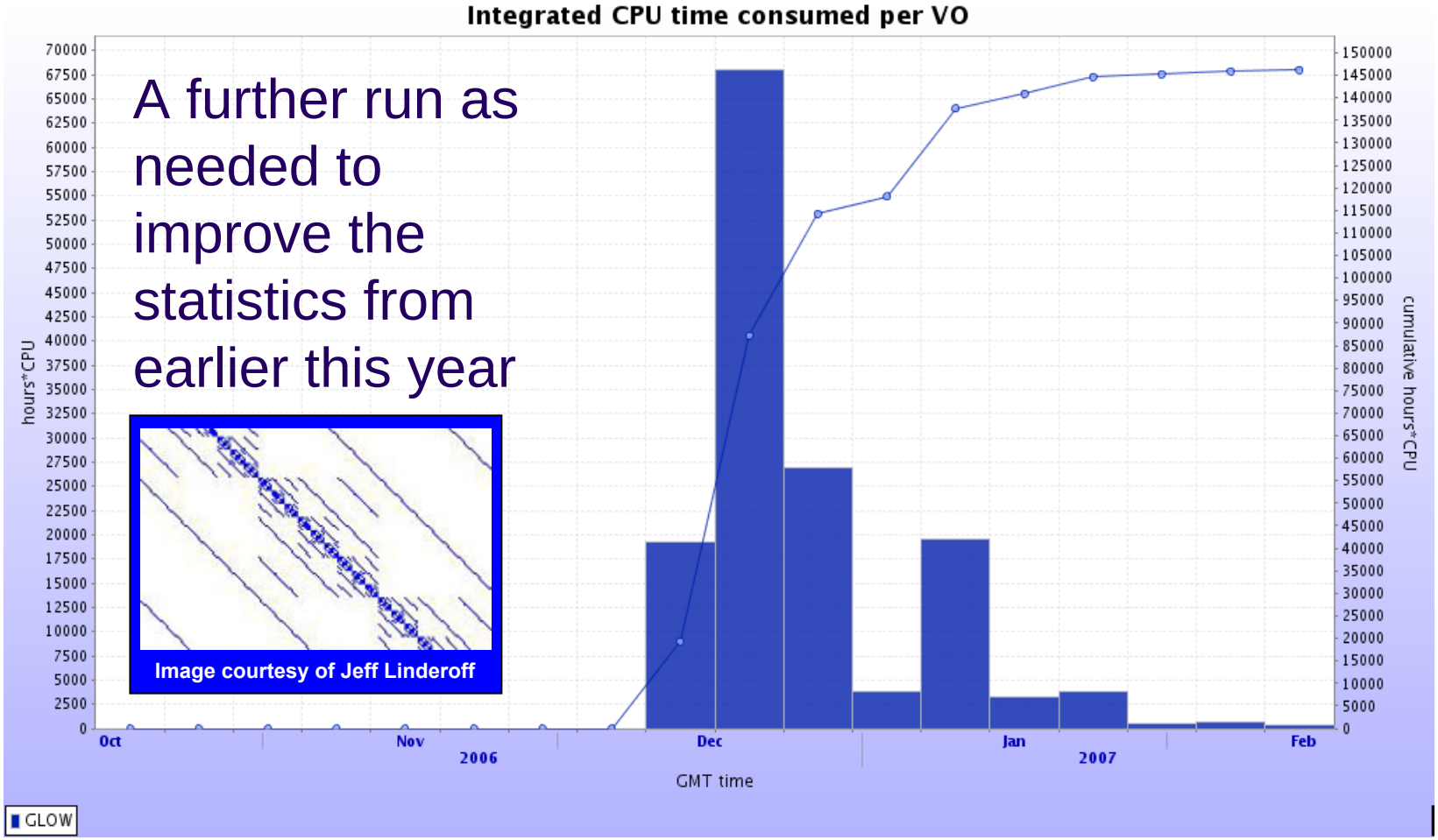


"small" "opportunistic" benefit - reaping behind the big guys: 1) GADU



"Margie Romine, a microbiologist at the Pacific Northwest National Laboratory in Richland, Washington, has used GADU to help her study the genetic code of the bacterium Shewanella oneidensis MR-1, whose metal- and radionuclide-reducing capability can impact the movement of such materials in the environment".

2) FOOTBALL POOL PROBLEM



3) Molecular Chemistry

- CHARMM -
  - single investigator support under OSG VO, using generalized PANDA workload management; currently under test.
- Monitoring

Jobs: 46 total jobs, last at 2007-02-09 13:55
Sites used: CHARMM (46)
Job types run: test (46)
Groups: all atlas usatlas

Usage	1 day (quota)	7 day (quota)	30 day (quot
Analysis	None (300)	None (2100)	None (9000)
User production	None (30)	None (210)	None (900)
Express	None (150)	None (1050)	None (4500)

Summary of all jobs for the last 20 days in any state at any site Go

46 jobs. Click job number to see details.
States: defined:0 assigned:0 waiting:0 activated:0 running:0 transferring:0 holding:0 finished:15 failed:31
Users: Benjamin Timothy Allen Miller:46
Releases: NULL:46
Sites: CHARMM:46
Datasets used by selected jobs

pathena job sets with multiple jobs in this selection:

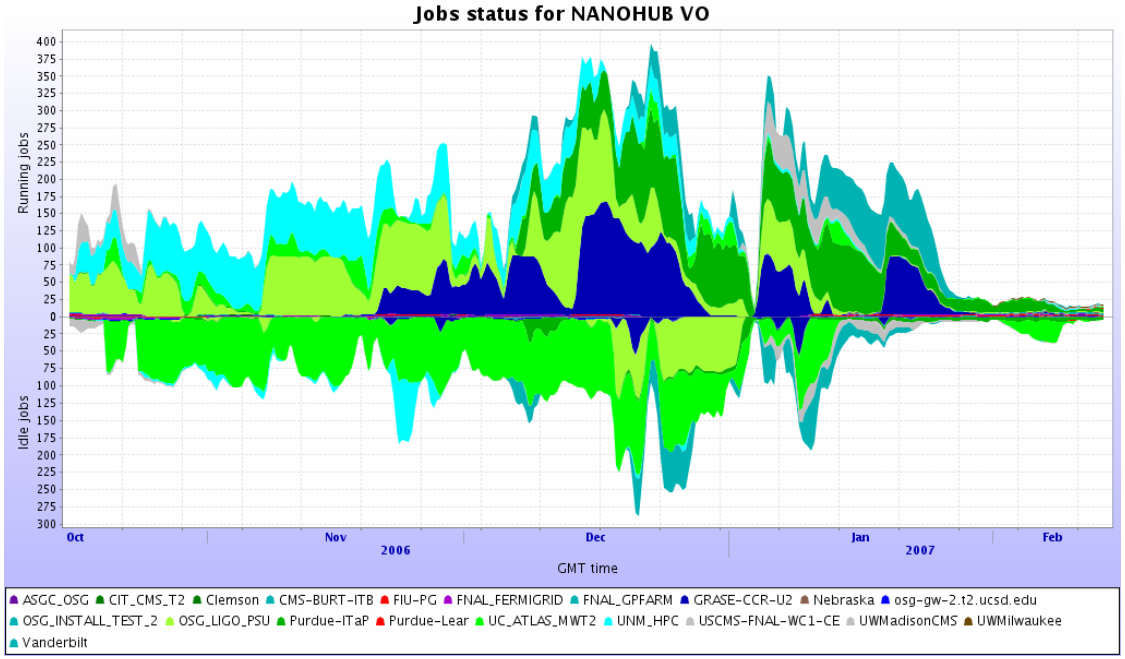
User:jobID	Panda jobs	libDS buildJob
Benjamin Timothy Allen Miller:0	936795,936794,919379,919378,913092,...	

Jobs:

PandaID, Owner	Job	Status	Created	Time to start	Duration	Ended/ Modified	Site, Type
936795 Benjamin Timothy Allen Miller	pathena jobID=0 trans-charmm-1.sh	failed	02-09 13:03	0:00:58	0:31:25	02-09 13:35	CHARMM test
	Error details: trans: Unspecified error, consult log file						
	Out: NULL						
936794 Benjamin Timothy Allen Miller	pathena jobID=0 trans-charmm-1.sh	failed	02-09 13:03	0:00:54	0:31:27	02-09 13:35	CHARMM test
	Error details: trans: Unspecified error, consult log file						
	Out: NULL						
919379 Benjamin Timothy Allen Miller	pathena jobID=0 trans-charmm-1.sh	failed	02-08 13:57	1:38:02	0:17:53	02-08 15:53	CHARMM test
	Error details: trans: Unspecified error, consult log file						
	Out: NULL						
919378 Benjamin Timothy Allen Miller	pathena jobID=0 trans-charmm-1.sh	failed	02-08 13:57	1:38:02	0:17:42	02-08 15:52	CHARMM test
	Error details: trans: Unspecified error, consult log file						
	Out: NULL						
913092	pathena jobID=0 trans-charmm-1.sh	finished	02-08 00:09	0:02:58	0:15:03	02-08 00:27	CHARMM test

4) Nanohub

- Challenge in the application: runs for up to 40 days; no application level checkpointing; Input/output sandboxes not robust.
- Monitoring of completed runs



Open Science Grid

Other Status:

- Software, etc. - Miron's talk
- EOT, etc. - Mike's talk
- Partnerships - Ruth's talk
- Engagement - John's talk

The End