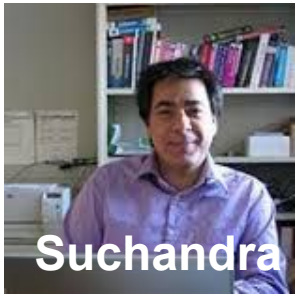
User Support, XD & Campus Infrastructures

Rob Gardner • University of Chicago

OSG All Hands Meeting, Clemson University, March 14-17, 2016

User Support and Campus Team



CI Connect
ATLAS Midwest
Tier2 Center

38M XSEDE cpu-hrs	52M Direct cpu-hrs	24M Connect cpu-hurs	200 Active users
96 PI/Projects	29 Publications	4 Training events	190 app modules
7 Submit hosts	6 Campus grids	81 Institutions	3.0 FTE

OSG shared infrastructure 115M CPU hours

2/1/2015-2/1/2016

OSG-Connect Projects 2015/02/01 - 2016/02/01

Project Name	PI	Institution	Feild of Science	Wall Hours
AlGDock	David Minh	Illinois Institute of Technolog	Chemistry	5240846
numfpi	Jerry Tessendorf	Clemson University	Computer and Information Scien	3679098
FFValidate	Vijay Pande	Stanford University	Chemistry	3556135
z2dqmc	Snir Gazit	University of California Berke	Physics	2241066
BioGraph	Alex Feltus	Clemson University	Biological Sciences	1322596
EvolSims	Oana Carja	University of Pennsylvania	Biological Sciences	1286555
CentaurSim	Nathan Kaib	Northwestern University	Astrophysics	1050911
IceCube	Francis Halzen	University of Wisconsin	Astrophysics	901601
PainDrugs	Pei Tang	University of Pittsburgh	Medical Sciences	785932
SourceCoding	Ahmad Golmohammadi	New Mexico State University	Engineering	753884
Errorstudy	Christopher Richards	USDA Agricultural Research	Molecular and Structural Biosci	573448
microphases	Patrick Charbonneau	Duke University	Chemistry	516830
ConnectTrain	Robert William Gardner Jr	University of Chicago	Training	377906
OSG-Staff	Chan			329686
mab	Vivek			277125
ProtEvol	Prem			163930
uchicago	Robe			156912
RicePhenomics	Hark			122319
SouthPoleTelescope	John			118223
cms-org-nd	Robe			116187
MS-EinDRC	Jacol			114335
FutureColliders	Serg			84903
SDEalgorithms	Haris			84376
duke-swcstaff	Robe			73112
NSNM	Vadir			57537
EvolvingAI	Jeff			47928
Swift	Mich			32650
scicomp-analytics	Robe			30288
DemandSC	Ferna			26637
atlas-org-uchicago	Robe			19472
KnowledgeLab	Jame			16335
Paniceae-trans	Jacol			15687
SNOplus	Josh			14614
ERVmodels	Fabri			13151
z2qmc	Snir			9598
ContinuousIntegration	Robe			8537
PathSpaceHMC	Frank			8434
UserSchool2015	Robe			8163
ExhaustiveSearch	Sam			4889
PTMC	Dere			3263
Phylo	Siava			2315
hABCNWHI	Yvon			1822
NSLS2ID	Dean			1486
MiniWorkshopUC15	Robert William Gardner Jr	University of Chicago	Computer and Information Scien	643
freesurfer	Donald Krieger	University of Chicago	Medical Sciences	638
SWC-OSG-IU15	Robert William Gardner Jr	IUPUI	Community Grid	286
ProbTracx	Dr. Bruce P. Hermann	University of Wisconsin	Neuroscience	277
pipediffusion	Panthea Sepehrband	Santa Clara University	Materials Science	270
DelhiWorkshop2015	Robert William Gardner Jr	University of Chicago	Physics	244
EHEC	Chuck Kaspar	University of Wisconsin-Mad	Microbiology	86
ASPU	ilyoup kwak	University of Minnesota	Bioinformatics	47
RADICAL	Shantenu Jha	Rutgers University	Computer and Information Scien	44
HTCC	Rob Quick	Indiana University	Community Grid	15
UserSchool2014	Tim Cartwright	OSG	Multi-Science Community	9
NeoflAnnot	Petra Lenz	University of Hawaii at Mano	Biological Sciences	5
atlas-wg-Exotics	Robert William Gardner Jr	ATLAS	High Energy Physics	4
OSGOpsTrain	Rob Quick	Open Science Grid	Community Grid	3
cms-org-fnal	Lothar Bauerdick	Fermi National Accelerator L	High Energy Physics	3
atlas-org-illinois	Mark Neubauer	University of Illinois	High Energy Physics	1
gem5	Dean Tullsen	University of California San I	Multi-Science Community	1
Total				24253300

OSG-Direct Projects 2015/02/01 - 2016/02/01

Project Name	PI	Institution	Feild of Science	Wall Hours
SPLINTER	Rob Quick	Indiana University	Medical Sciences	32746961
Duke-QGP	Steffen A. Bass	Duke University	Nuclear Physics	9591924
sPHENIX	Martin Purschke	Brookhaven Nation	Nuclear Physics	5589611
UPRRP-MR	Steven Massey	Universidad de Pue	Bioinformatics	1593405
IU-GALAXY	Rob Quick	Indiana University	Bioinformatics	1266000
BNL-PHENIX	Matthew Snowball	Brookhaven Nation	Nuclear Physics	519229
DetectorDesign	John Strologas	University of New M	Medical Sciences	376994
Pheno	Stefan Hoeche	SLAC	High Energy Physic	260755
DeerDisease	Lene Jung Kjaer	Southern Illinois Ur	Biological Sciences	157624
OSG-Staff	Chander Sehgal	Fermilab	Computer and Infor	138872
UNC-RESOLVE-pr	David Stark	UNC Chapel Hill	Physics and astron	34840
IBN130001-Plus	Donald Krieger	University of Pittsbi	Neuroscience	2002
Total				52278216

OSG-XD Projects 2015/02/01 - 2016/02/01

Project Name	PI	Institution	Feild of Science	Wall Hours
TG-IBN130001	Donald Krieger	University of Pittsbi	Biological Sciences	34454812
TG-DMR130036	Emanuel Gull	University of Michig	Materials Science	1269854
TG-CHE140110	John Stubbs	University of New E	Chemistry	1048149
TG-AST140088	Francis Halzen	University of Wisco	High Energy Physic	552688
TG-AST150012	Gregory Snyder	Space Telescope S	Mathematical Scier	420545
TG-CHE140098	Paul Siders	University of Minne	Chemistry	318432
TG-AST150044	Jennifer Lotz	Space Telescope S	Astrophysics	128107
TG-PHY150040	Francis Halzen	University of Wisco	Physics and astron	83901
TG-MCB150090	Emiliano Brini	SUNY at Stony Brc	Molecular Bioscien	77481
TG-AST150033	Juliette Becker	University of Michig	Astrophysics	68490
TG-MCB140160	David Rhee	Albert Einstein Coll	Molecular and Stru	39368
TG-GEO150003	Jon Pelletier	University of Arizor	Geographic Informa	36641
TG-TRA130011	John Chrispell	Indiana University (	Other	21074
TG-DMR140072	Adrian Del Maestro	University of Vermc	Materials Science	18098
TG-OCE140013	Yvonne Chan	University of Hawa	Ocean Sciences	17884
TG-PHY120014	Qaisar Shafi	University of Delaw	Physics and astron	17594
TG-TRA130030	Neranjan Edirisingl	Georgia State Univ	Mathematical Scier	3933
TG-AST150046	Suzanne Hawley	University of Washi	Mathematical and F	3020
TG-MCB140232	Alan Chen	SUNY at Albany	Molecular and Stru	598
TG-CCR140028	Shantenu Jha	Rutgers; the State	Computer and Infor	455
TG-TRA100004	Andrew Ruether	Swarthmore Colleg	Training	204
TG-MCB060061N	Jeffry D. Madura	Duquesne Universi	Molecular and Stru	49
TG-MCB140268	Graziano Vernizzi	Siena College	Molecular and Stru	19
TG-STA110011S	Stephen McNally	University of Tenne	Other	1
Total				38581397

A few projects selected from OSG Connect
Training and Educational Projects

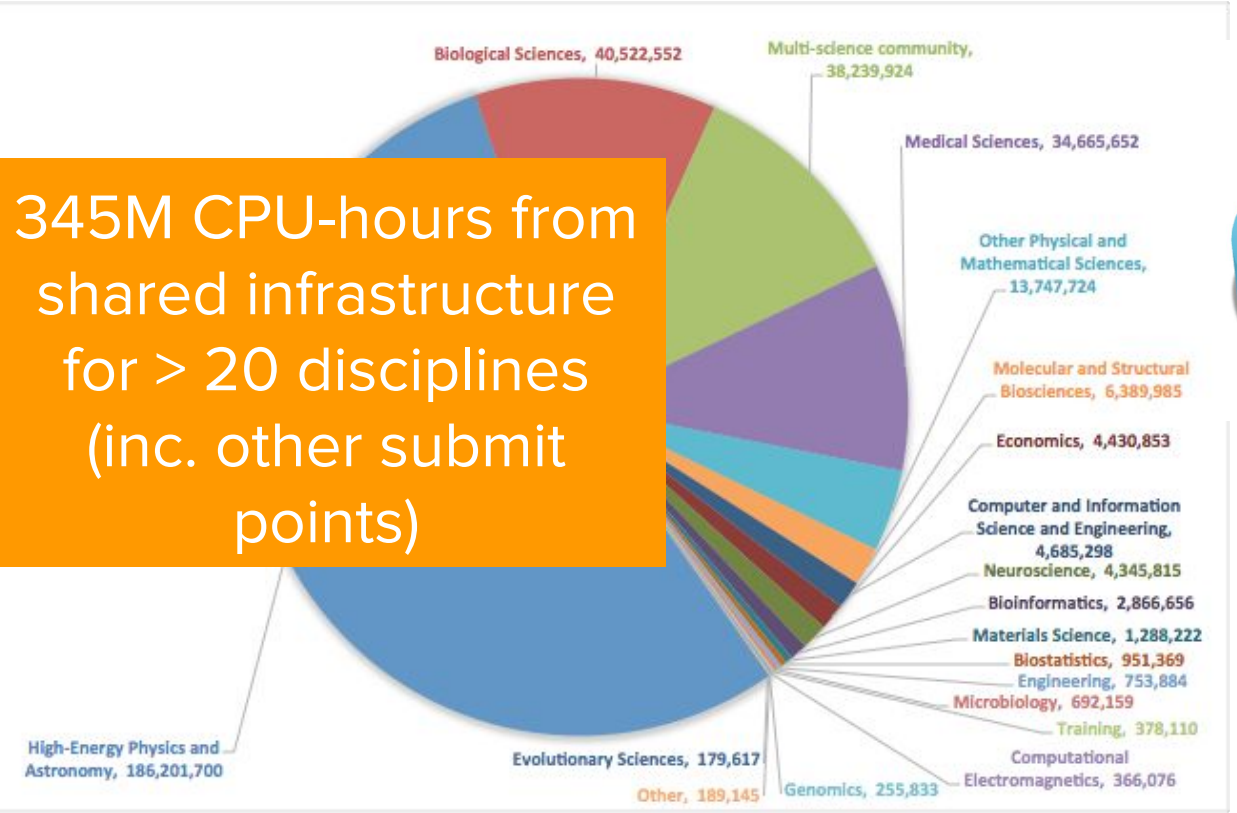
Project Name	PI	Institution	Feild of Science	Wall Hours
ConnectTrain	Robert William Gar	University of Chica	Training	377906
Duke-SWC-OSG1	Robert William Gar	Duke University	Multi-Science Com	234514
duke-swcstaff	Robert William Gar	Duke University	Multi-Science Com	73112
UserSchool2015	Robert William Gar	University of Wisco	Education	8163
MiniWorkshopUC1	Robert William Gar	University of Chica	Computer and Infor	643
SWC-OSG-IU15	Robert William Gar	IUPUI	Community Grid	286
DelhiWorkshop201	Robert William Gar	University of Chica	Physics	244
UserSchool2014	Tim Cartwright	OSG	Multi-Science Com	9
Total				694877

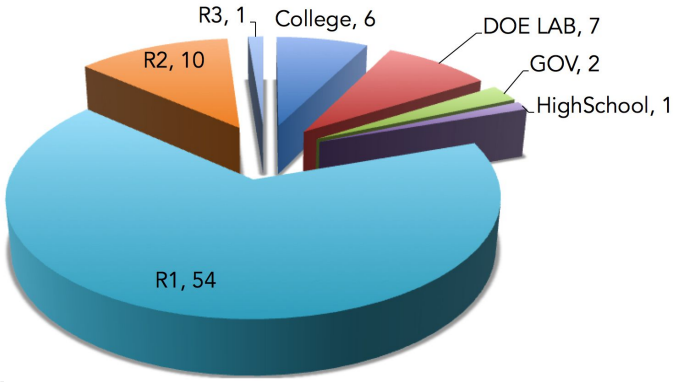
Campus Projects

Project Name	PI	Institution	Feild of Science	Wall Hours
Duke-SWC-OSG1	Mark DeLong	Duke University	Multi-Science Com	234514
uchicago	Robert William Gar	University of Chica	Multi-Science Com	156912
cms-org-nd	Kevin Lannon	University of Notre	High Energy Physic	116187
duke-swcstaff	Mark DeLong	Duke University	Multi-Science Com	73112
Total				580725

Impact across disciplines and institutions

345M CPU-hours from shared infrastructure for > 20 disciplines (inc. other submit points)

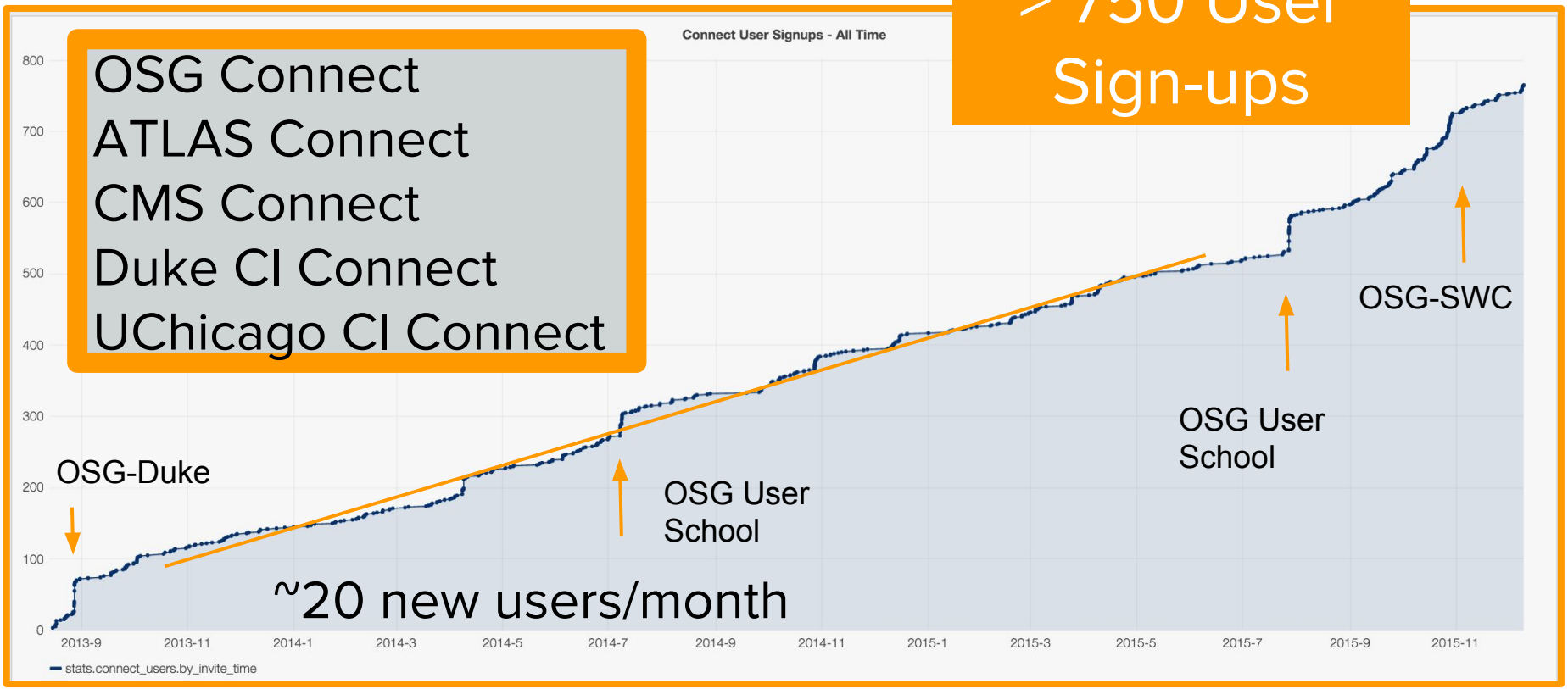


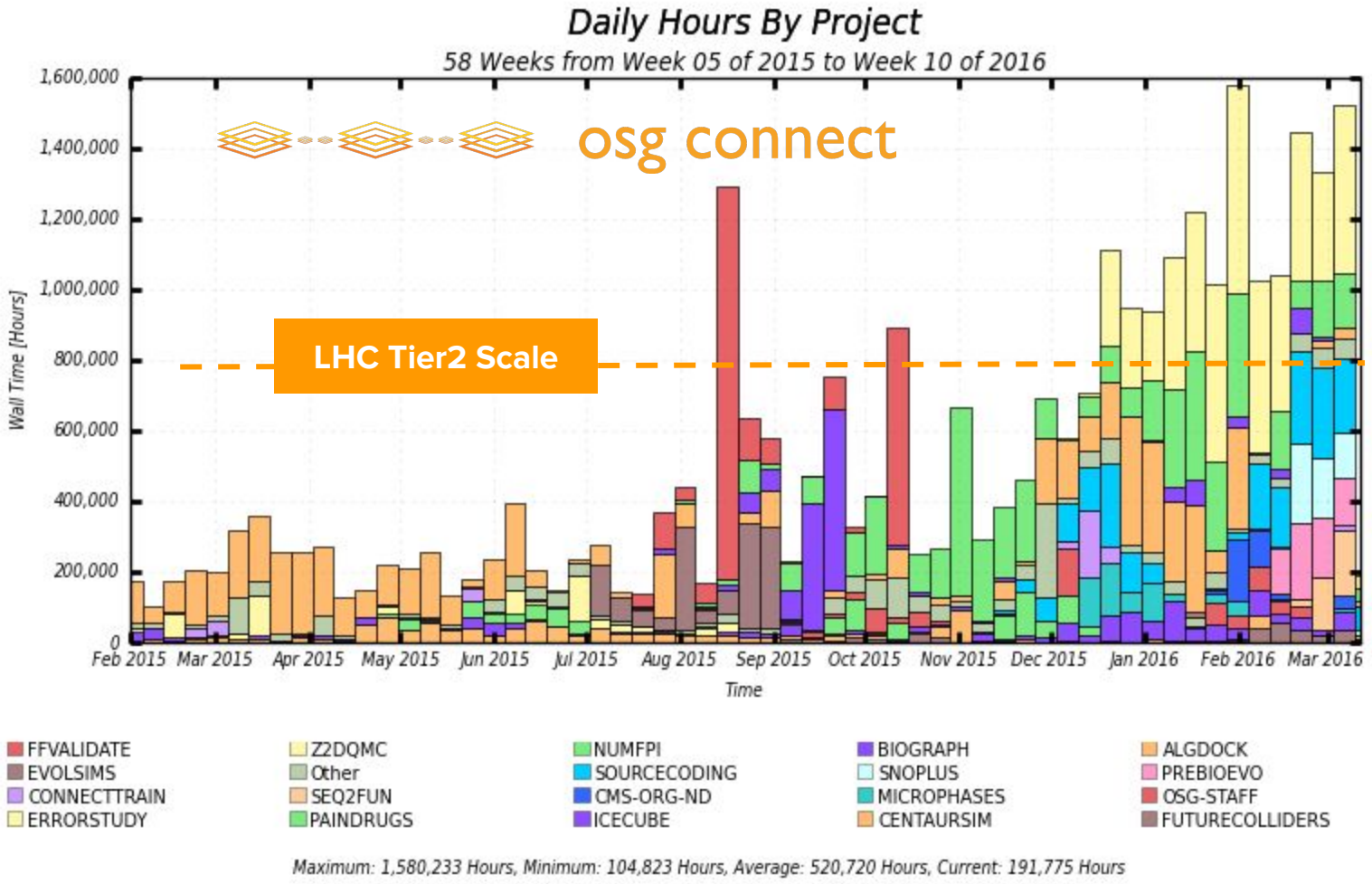


PIs by Institution Carnegie classification (opportunistic users)

> 750 User Sign-ups

OSG Connect
ATLAS Connect
CMS Connect
Duke CI Connect
UChicago CI Connect




Daily Hours By Project
58 Weeks from Week 05 of 2015 to Week 10 of 2016
osg connect
LHC Tier2 Scale
Wall Time [Hours]
1,600,000
1,400,000
1,200,000
1,000,000
800,000
600,000
400,000
200,000
0
Feb 2015
Mar 2015
Apr 2015
May 2015
Jun 2015
Jul 2015
Aug 2015
Sep 2015
Oct 2015
Nov 2015
Dec 2015
Jan 2016
Feb 2016
Mar 2016
Time
FFVALIDATE
EVOLSIMS
CONNECTTRAIN
ERRORSTUDY
Z2DQMC
Other
SEQ2FUN
PAINDRUGS
NUMFPI
SOURCECODING
CMS-ORG-ND
ICECUBE
BIOGRAPH
SNOPLUS
MICROPHASES
CENTAURSIM
ALGDOCK
PREBIOEVO
OSG-STAFF
FUTURECOLLIDERS
Maximum: 1,580,233 Hours, Minimum: 104,823 Hours, Average: 520,720 Hours, Current: 191,775 Hours

Science

Science publications (partial list)

Project: ProtEvol

1. Shah P, McCandlish DM, and Plotkin JB. “Contingency and entrenchment in protein evolution under purifying selection”. PNAS 112: E3226–E3235
2. Otwinoski J, McCandlish DM, Shah P, and Plotkin JB. “Identifying site-specific amino acid preferences for modeling protein evolution” (In Prep)

Project: AIGDock

1. Trung Hai Nguyen and David D. L. Minh, "Intermediate Thermodynamic States Contribute Equally to Free Energy Convergence: Demonstration with Replica Exchange.” (In preparation)
2. David D. L. Minh, “Protein-Ligand Binding Potential of Mean Force Calculations with Hamiltonian Replica Exchange on Alchemical Interaction Grids” (http://arxiv.org/pdf/1507.03703.pdf

Project: ErrorStudy

1. Patrick A. Reeves, Cheryl L. Bowker, Christa E. Fettig, Luke R. Tembrock, and Christopher M. Richards, “Effect of error and missing data on population structure inference using microsatellite data” (In preparation)

Project: numpfi

1. Paul Kilgo and Jerry Tessendorf. “Accelerated path generation and visualization for numerical integration of Feynman path integrals for radiative transfer”. Joint International Conference on Mathematics and Computation, Supercomputing in Nuclear Applications and the Monte Carlo Method, April 2015.
2. Paul Kilgo and Jerry Tessendorf. Toward validation of a Monte Carlo rendering technique. Special Interest Group on Graphics and Interactive Techniques, August 2015. (Poster).

Project: FutureColliders

1) S.V. Chekanov, I. Pogrebnyak, D. Wilbern, Cross-platform validation and analysis environment for particle physics, arXiv:1510.06638

2) S.V.Chekanov, J.Dull. Energy range of hadronic calorimeter towers a and cells for high-pT jets at a 100 TeV collider, arXiv:1510.06638

Project: CentaurSim

Kaib, Nathan A.; Chambers, John E. “The fragility of the terrestrial planets during a giant-planet instability”, Monthly Notices of the Royal Astronomical Society, Volume 455, Issue 4, p.3561-3569 (2016)

Project: SPT

N. Whitehorn, T. Natoli et al, "Millimeter Transient Point Sources in the SPTpol 100 Square Degree Survey," (to be submitted to The Astrophysical Journal).

Project: RicePhenomics

Knecht, AC, Campbell, MT, Caprez, A, Swanson, DR: Image Harvest: An open source platform for high-throughput plant image processing and analysis. Journal of Experimental Botany. [Accepted, pending minor revision].

XD-Project: TG-CHE130091

1. Paul D. Siders, "Conformational free energy of alkylsilanes by nonequilibrium-pulling Monte Carlo simulation," Molecular Simulation, accepted September 2015. doi: 10.1080/08927022.2015.1083101

2. Presented "Conformational free energy in a chromatographic stationary phase by nonequilibrium pulling within Gibbs-Ensemble Monte Carlo simulation" at the Virtual Conference on Computational Chemistry, University of Mauritius, August 2015.

Optical data communication & compression

- David Mitchell, New Mexico State
  - Ahmad Golmohammadi (EE graduate student)
  - Important for digital space and satellite communication & wireless data transmission
  - Whole system simulations - transmitter, decoder, receiver & stochastic noise, data compression
  - Computations are well suited to HTC. Ahmad:

> I am so happy that could do my simulation that much fast. Instead of waiting for several weeks, right now I am getting them in hours. Thank you very much, Best wisehs

Piyasat Nilkaew → catalyzed by RMACC meeting in Boulder Aug '15

Randomness in Evolution

- Understanding evolution at molecular scale in DNA with combination of mathematical modeling and simulation
- How quickly does a genome fix a mutation?
- Role of randomness versus natural selection?

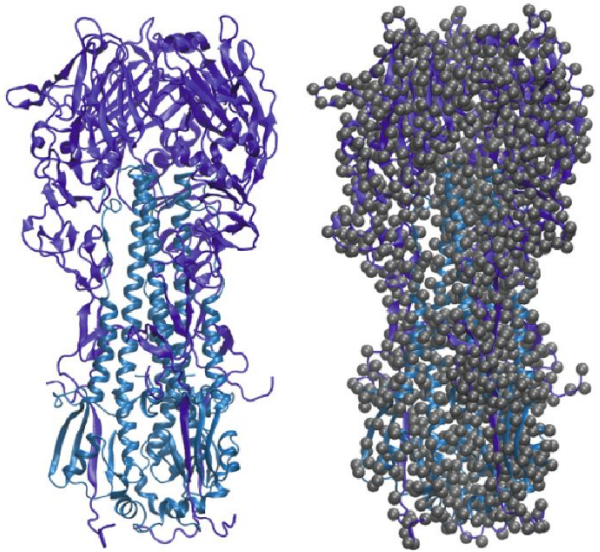


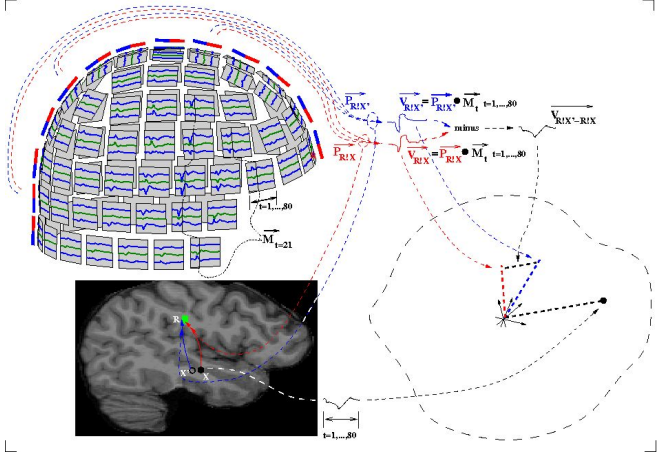
Image courtesy Joshua Plotkin. A computationally predicted structure of an influenza virus protein. Plotkin's research group uses the Open Science Grid to study questions in evolutionary biology and ecology.

Joshua Plotkin, Penn

"We use the OSG to run computer simulations for complex processes. By simulating evolution in populations, we can study hypothetical situations that you can't study in the wet lab or field."

Human Neuroimaging on the Open Science Grid

(c.f. Don Krieger's talk this week)

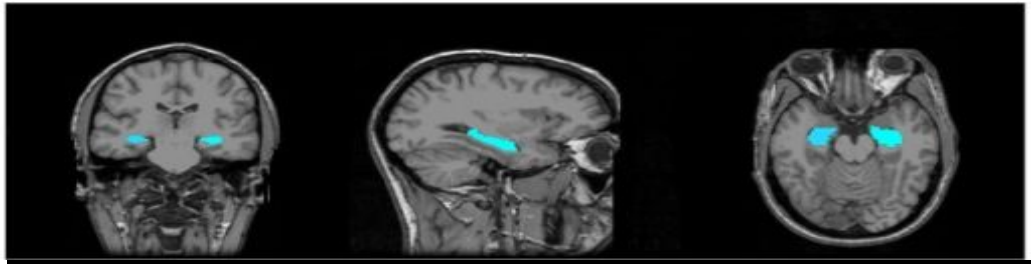


Referee consensus is a new approach to solving large systems of simultaneous mixed linear/nonlinear equations (ACM, Intl J Adv Comp Sci, Explanation of the figure on YouTube).

- Solution of the equations requires search of a high dimensional space.
- The discovery and use of a cost function computed in the "source" space rather than in the "measurement" space enables:
  - solution for a single source at a time,
  - efficient implementation on the grid.

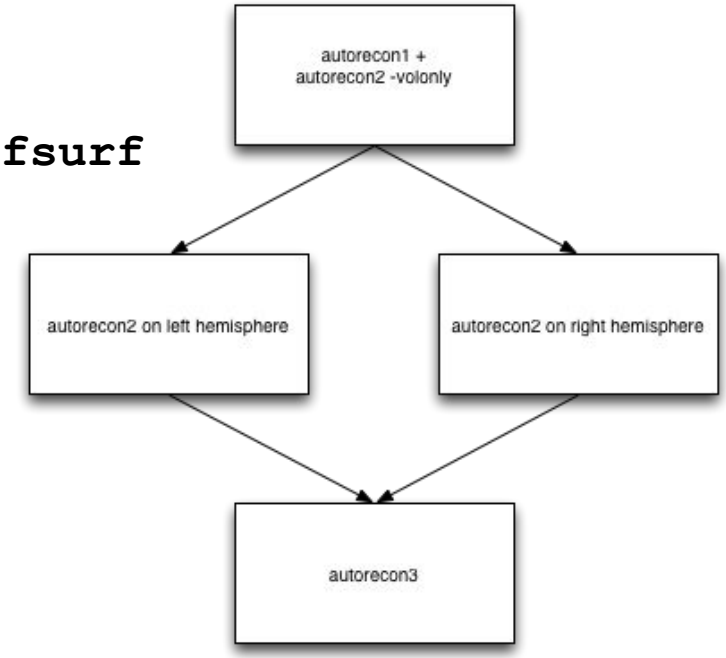
Computational **Neuroimaging Service**

- OSG partnership with University of Pittsburgh Medical School to provide open computational brain image analysis platform **FreeSurfer** analysts



```
fSurf --runFile SubNo_01_defaced.mgz --singleCore
fSurf --runFile SubNo_02_defaced.mgz --multiCore
fSurf --showJobsStatus
fSurf --fetchResults SubNo_02_defaced.mgz
```

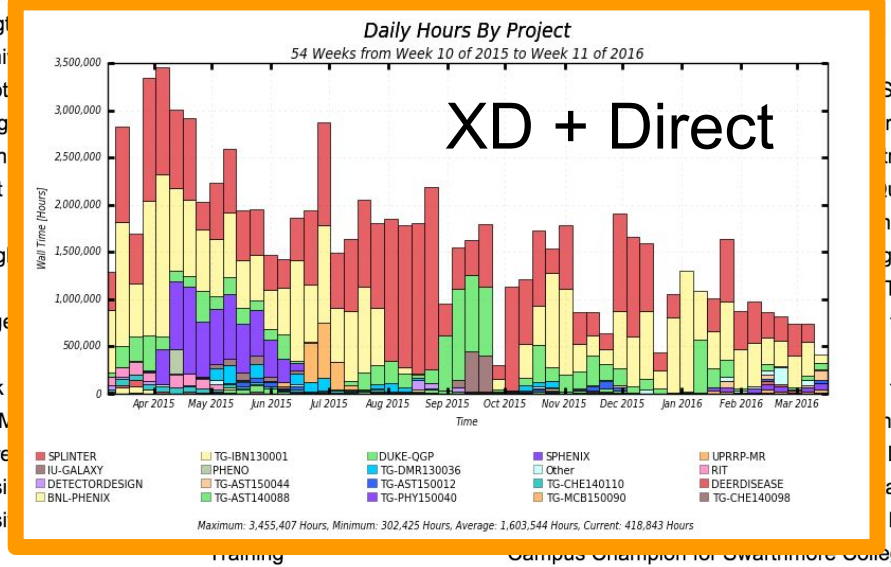
`fsurf`



Command line and web API - S. Thapa

OSG XSEDE - 24 active projects

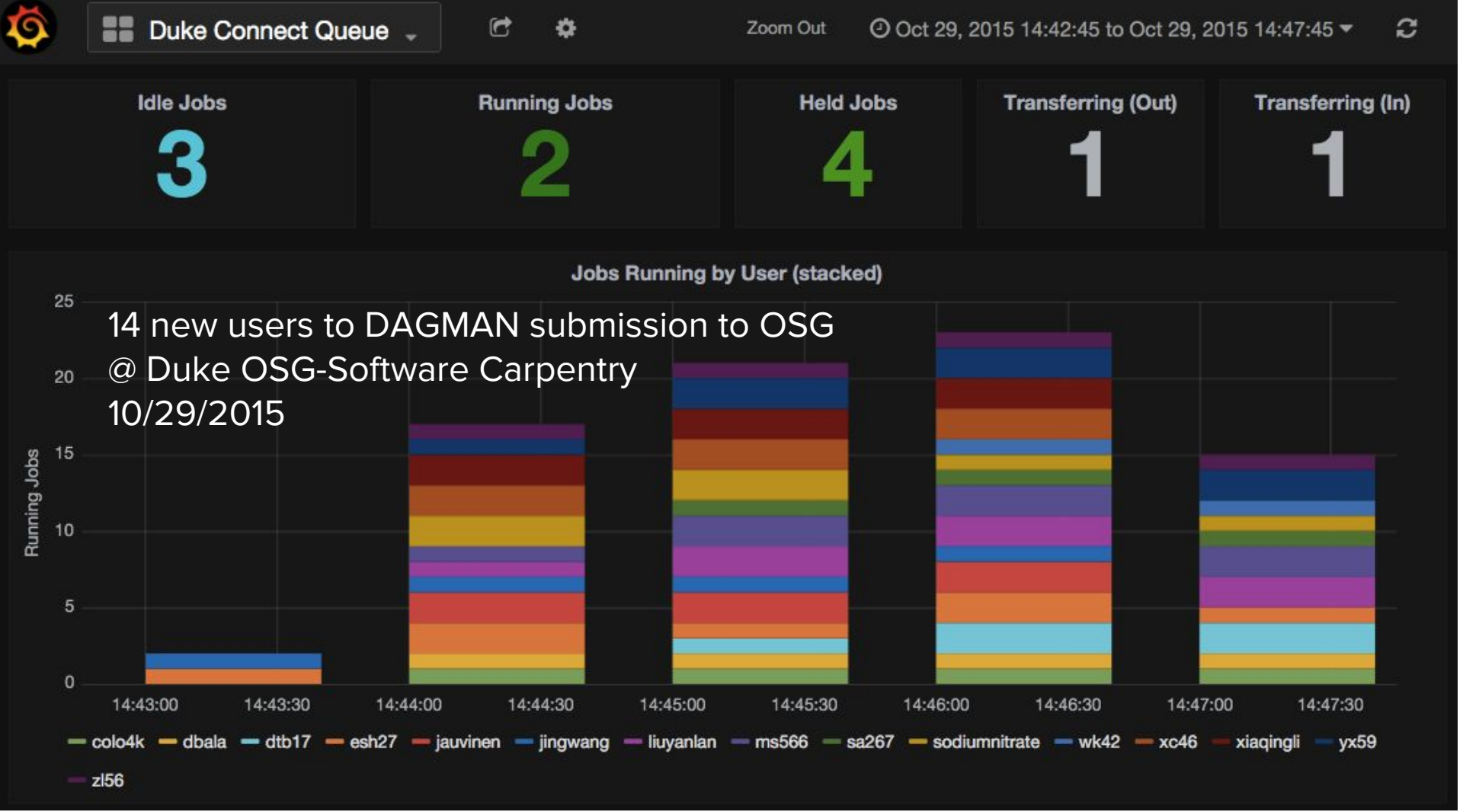
Project ID	CPU-hours	PI	Institution	Field of Science	Title
TG-AST140088	110,621	Francis Halzen	University of Wisconsin-Madison	Astronomical Sciences	IceCube Monte Carlo production
TG-AST150012	312,370	Gregory Snyder	Space Telescope Science Institute	Astronomical Sciences	Testing galaxy physics by modeling observations with simulations
TG-AST150033	28,167	Juliette Becker	University of Michigan	Astronomical Sciences	Oscillations of Relative Inclination Angles in Compact Extrasolar P
TG-AST150044	101,162	Jennifer Lotz	Space Telescope Science Institute	Astronomical Sciences	Constraining the Initial Conditions of Major Galaxy Mergers Using
TG-AST150046	2,340	Suzanne Hawley	University of Washingt		
TG-CCR140028	3	Shantenu Jha	Rutgers, the State Uni		
TG-CHE140098	5,371	Paul Siders	University of Minnesot		Supercritical Fluid Chroma
TG-CHE140110	871,898	John Stubbs	University of New Eng		rogeneous DNA Hybridiza
TG-DMR130036	606,439	Emanuel Gull	University of Michigan		tron Systems with Two-Pa
TG-DMR140072	12,423	Adrian Del Maestro	University of Vermont		uantum Fluids
TG-GEO150003	7,109	Jon Pelletier	University of Arizona		ny
TG-IBN130001	27,133,268	Donald Krieger	University of Pittsburg		ging
TG-MCB060061N	1	Jeffry D. Madura	Duquesne University		Transporters, and Extend
TG-MCB140160	33,475	David Rhee	Albert Einstein College		to search for aptamers
TG-MCB140232	358	Alan Chen	SUNY at Albany		
TG-MCB150090	199,137	Emiliano Brini	SUNY at Stony Brook		free energies of biologica
TG-OCE140013	339	Yvonne Chan	University of Hawaii, M		n Computation to Detect C
TG-PHY120014	10,592	Qaisar Shafi	University of Delaware		Dark Matter and Grand U
TG-PHY150040	23,337	Francis Halzen	University of Wisconsi		a analysis and detector u
TG-PHY160001	0	Terrance Figy	Wichita State Universi		Production in Association
TG-TRA100004	118	Andrew Ruether	Swarthmore College	Training	Campus Champion for Swarthmore College
TG-TRA130011	5,855	John Chrispell	Indiana University of Pennsylvania	Training	Campus Champion Indiana University of Pennsylvania
TG-TRA130030	2,891	Neranjan Edirisingh	Georgia State University	Computational Mathematics	Campus champion allocation
TG-TRA150018	0	Stephen Wolbers	Oregon State University	Training	Campus Champion for Oregon State



User Support

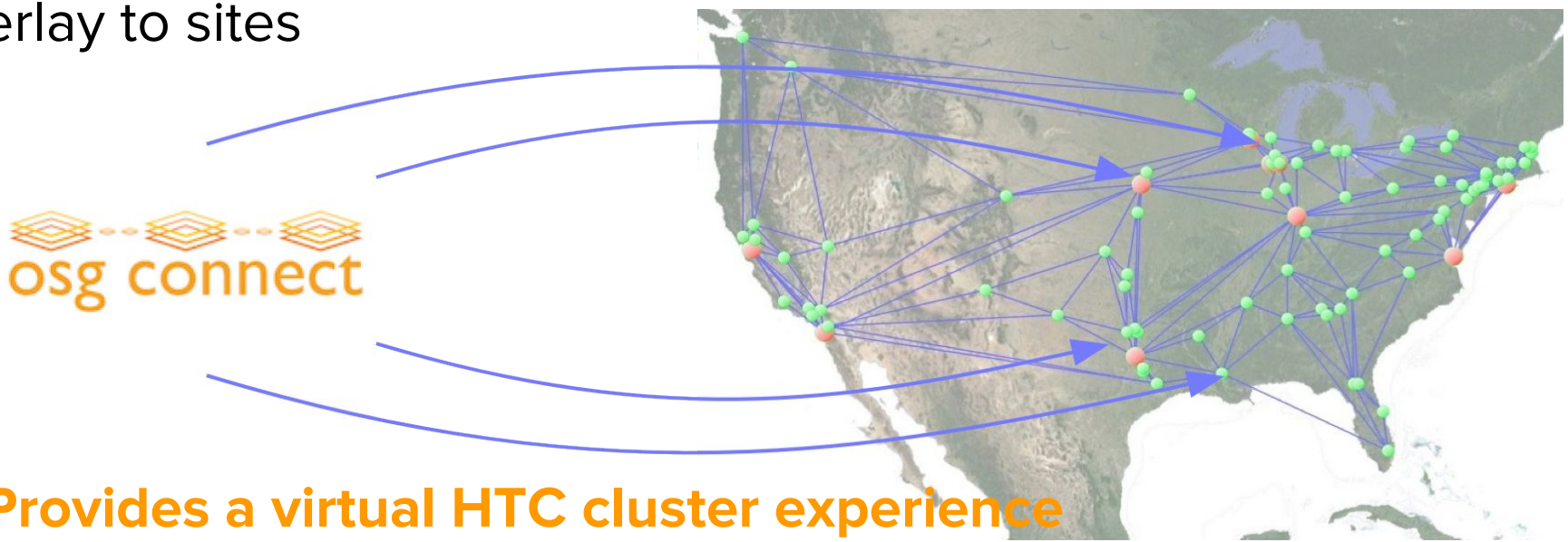
Joint OSG Software Carpentry Workshops

- Two members of OSG User Support team are Software Carpentry instructors
- Extended standard core curriculum with a day of HTC computing best practices


Duke Connect Queue
Zoom Out
Oct 29, 2015 14:42:45 to Oct 29, 2015 14:47:45
Idle Jobs
3
Running Jobs
2
Held Jobs
4
Transferring (Out)
1
Transferring (In)
1
Jobs Running by User (stacked)
14 new users to DAGMAN submission to OSG
@ Duke OSG-Software Carpentry
10/29/2015
Running Jobs
25
20
15
10
5
0
14:43:00
14:43:30
14:44:00
14:44:30
14:45:00
14:45:30
14:46:00
14:46:30
14:47:00
14:47:30
colo4k
dbala
dtb17
esh27
jauvinen
jingwang
liuyanlan
ms566
sa267
sodiumnitrate
wk42
xc46
xiaqingli
yx59
zl56

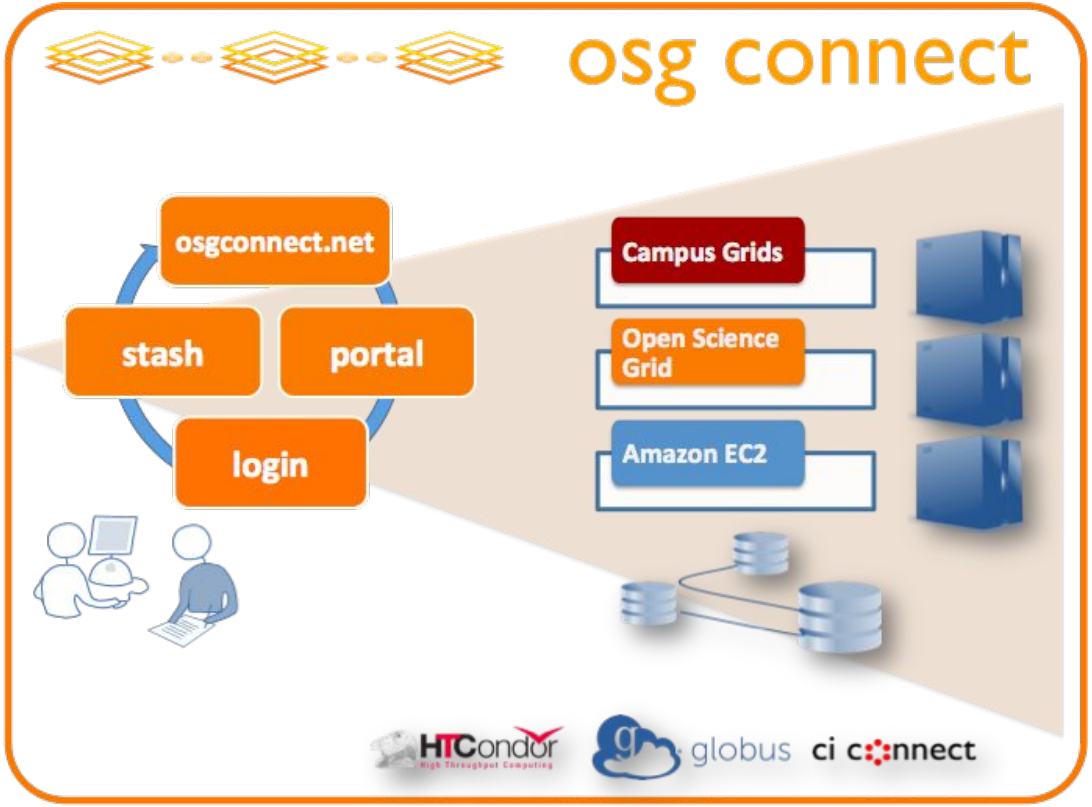
OSG Connect Service

Local campus identity (CILogon) ‣ OSG Connect identity (Globus) ‣ virtual organization (OSG) ‣ HTCondor Glidein Overlay to sites



⇒ **Provides a virtual HTC cluster experience**

OSG Connect is an easy way to get started



OSG as a **campus cluster**

- ★ Login host
- ★ Job scheduler
- ★ Software
- ★ Storage

OSG Connect Service

- For users without an institutional submission point
- login node for job management,
  `login.osgconnect.net`
- Stash is a temporary storage service
  - Globus Online, HTTP, Xrootd
  - Posix accessible from login nodes
  - Origin server for StashCache
- Uses OASIS software repository for user-installed software

User software Installed on the OSG

- Repository for common, pre-installed user installed software
- Accessed with a **`module`** command
  - identical software on all clusters
  - apps/libraries installed

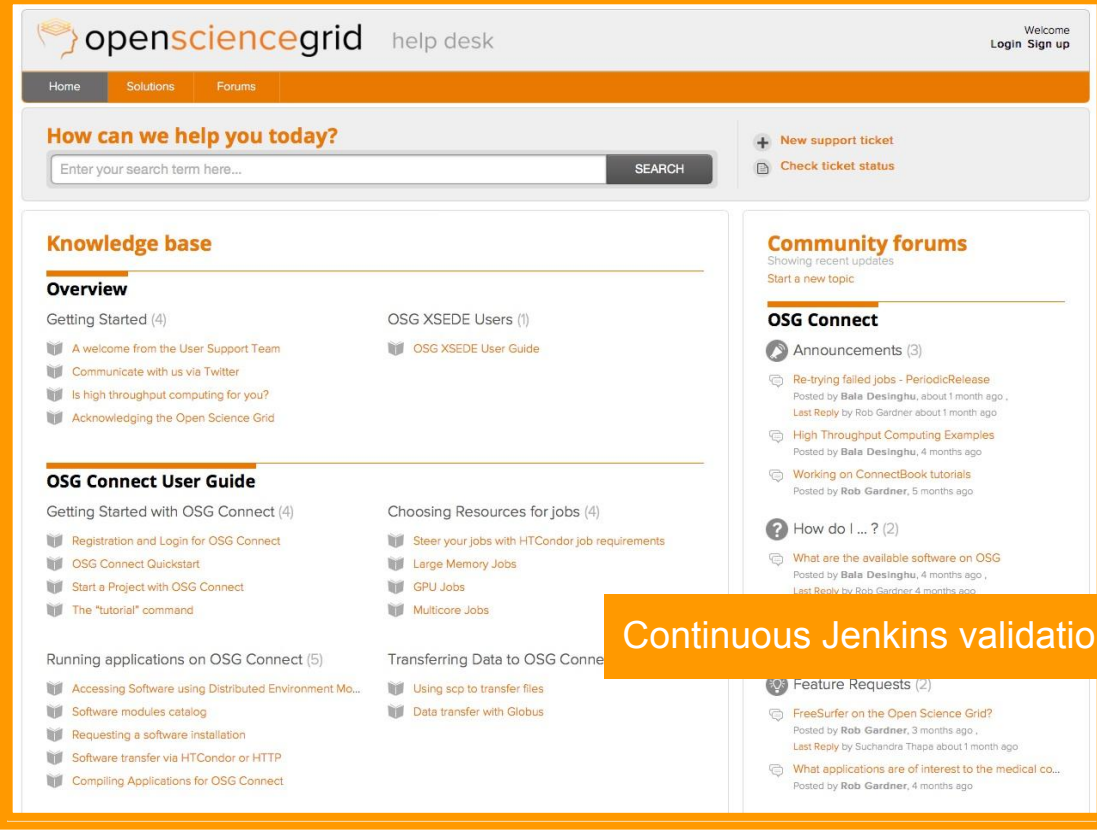
```
#!/bin/bash
switchmodules oasis
module load R
module load matlab
...
```

Integrated KB, tutorials, github-managed

Helpdesk:
https://support.opensciencegrid.org

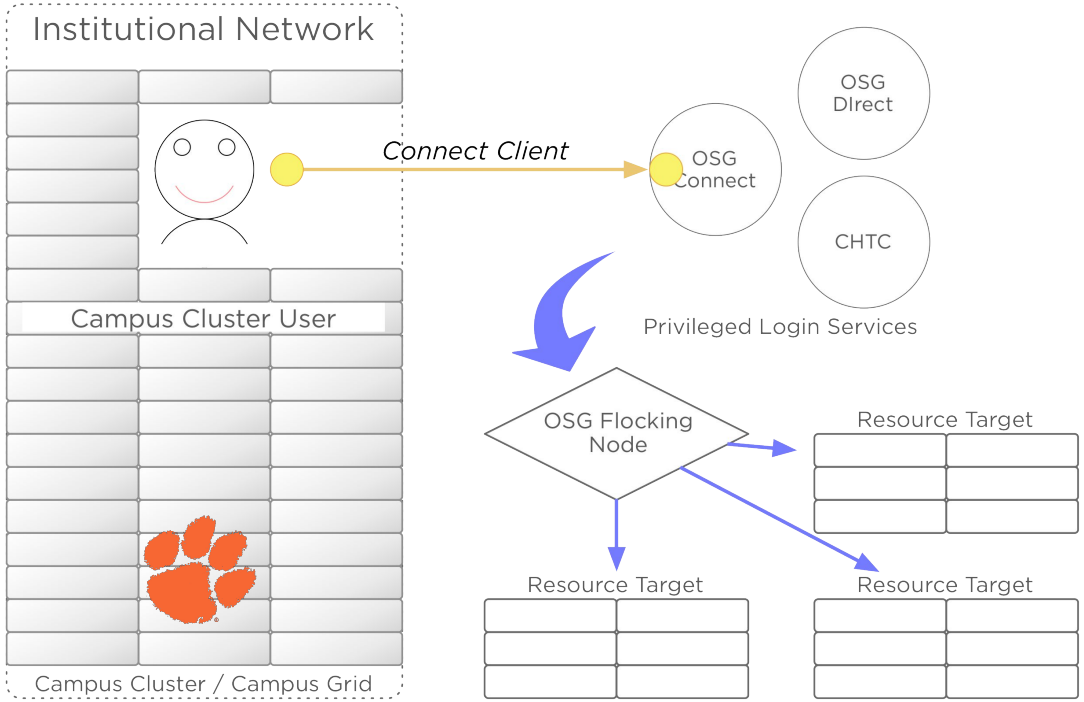
- Knowledge Base
  - tutorials
  - HTC recipes
  - “How do I…?” articles
  - searchable
- Support channels
  - Web chat
  - Email

user-support@opensciencegrid.org



Continuous Jenkins validation

Connect Client brings OSG pools to campus



Work from “home”

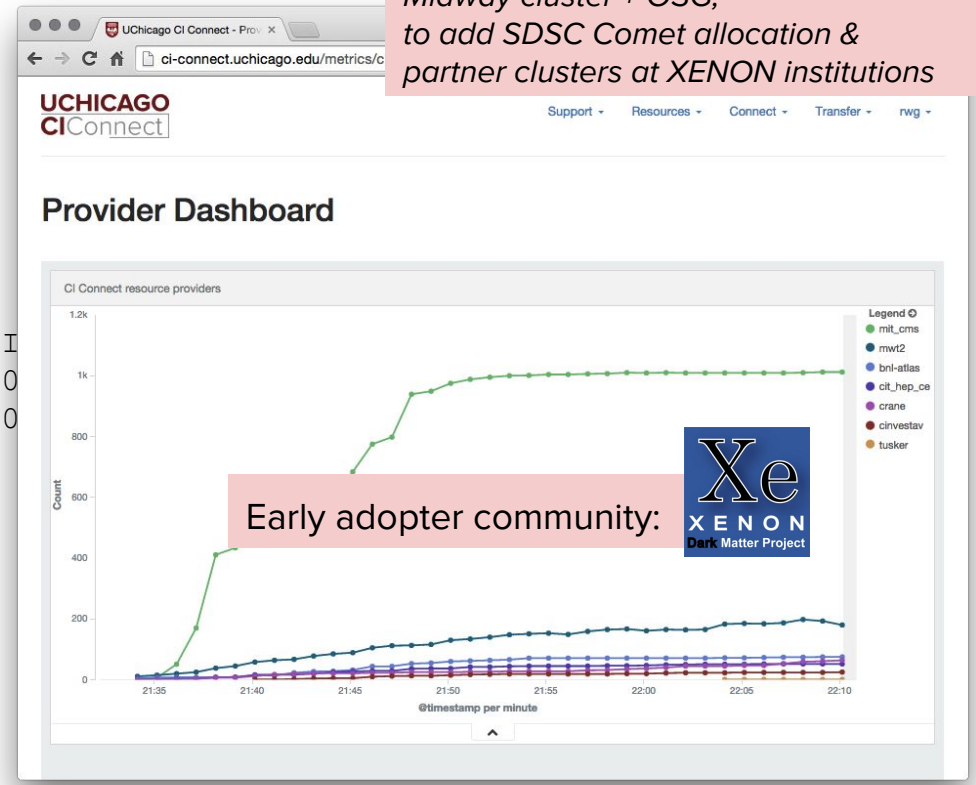
Submit locally, run globally

Workshop Thursday
OSG + ACI-Ref

Organized into a ‘local’ queue

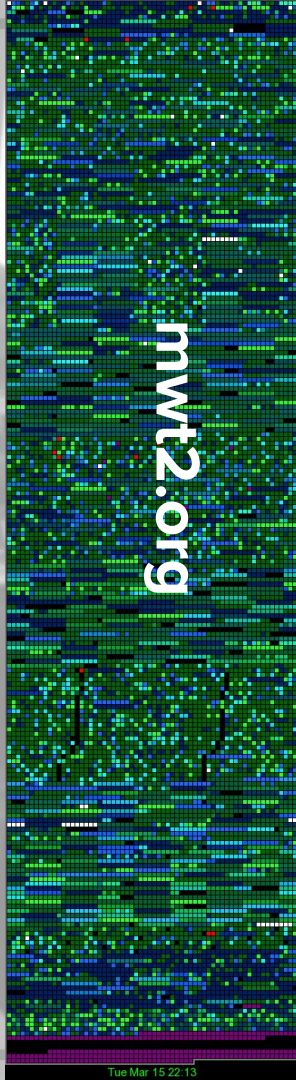
```
$ module load connect-client
$ connect setup
$ connect test
$ connect submit myjob.sub
$ connect q rwg
-- Submitter: login.ci-connect.uchicago.edu :
uchicago.edu
 ID      OWNER            SUBMITTED     RUN_TI
252624.0   rwg             9/2  14:21    0+00:0
252624.1   rwg             9/2  14:21    0+00:0
...
$ connect status
$ connect pull (results)
```

UChicago CI Connect Service:
Midway cluster + OSG;
to add SDSC Comet allocation & partner clusters at XENON institutions



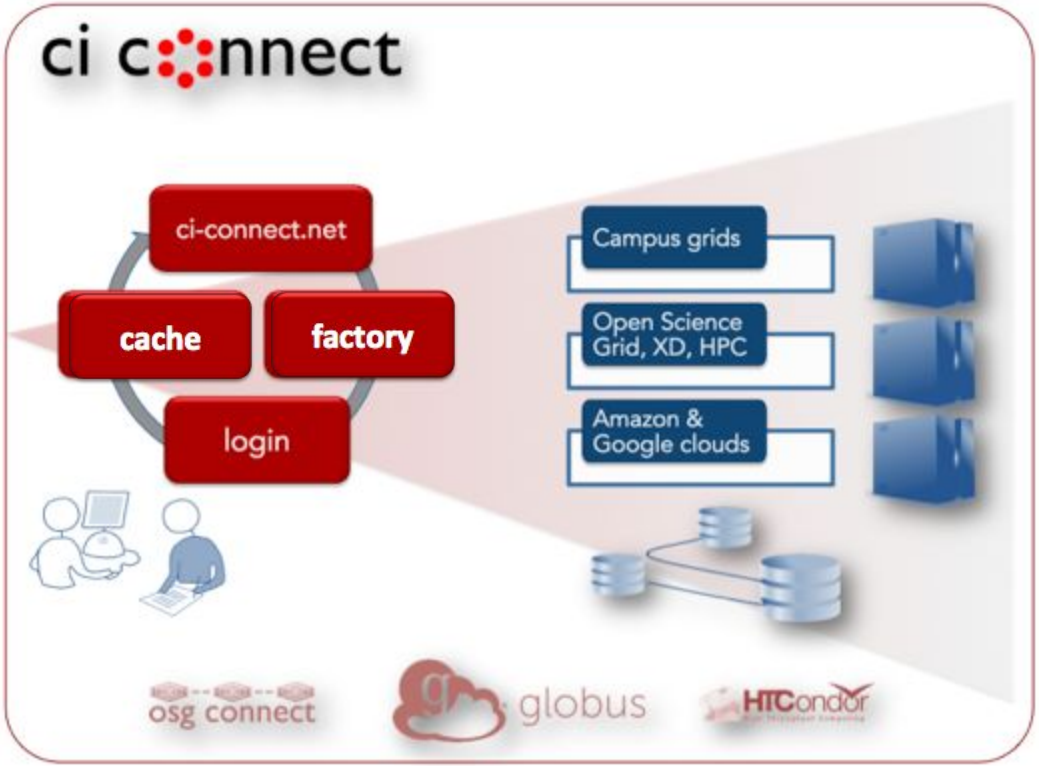
Submitted from UChicago Research Computing Center cluster “Midway”

Campus Cyberinfrastructure



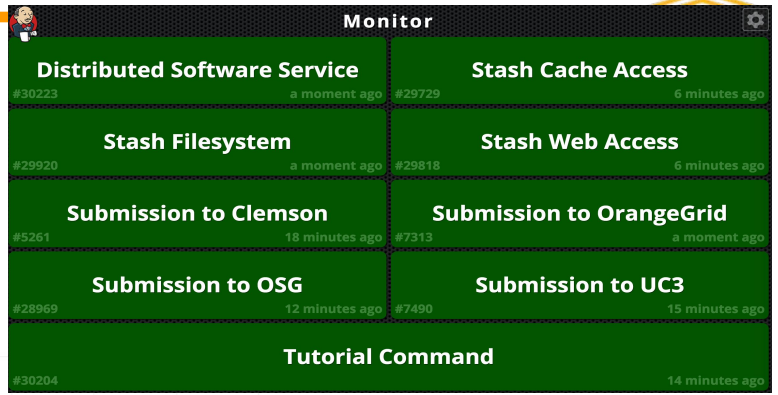
CI Connect

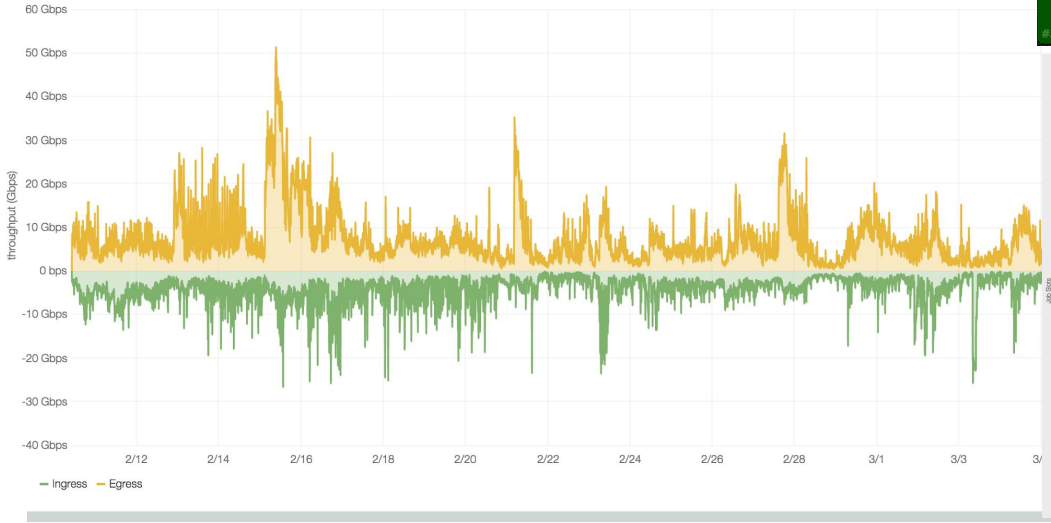
- **Integration** of mature software
- Concetrated expertise
- Minimize footprint on resources
- Deliver as service

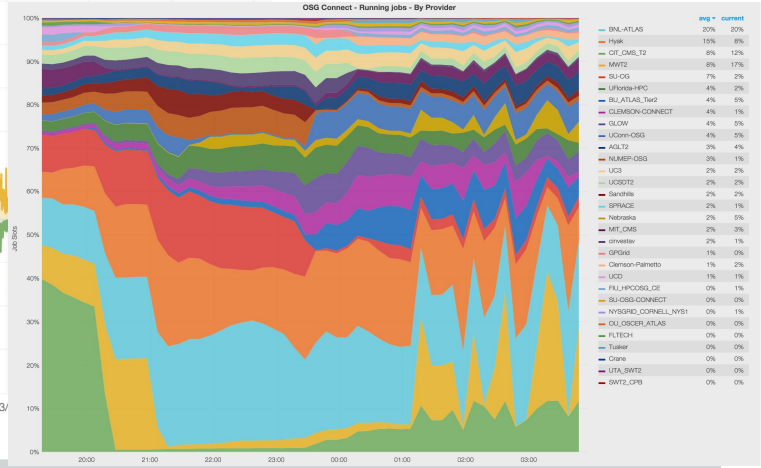


*-Connect Support Infrastructure

- 6 login servers (schedd's)
- 25 Connect (bosco) factories
- 1000 TB of CephFS storage
- 80 Gbps SciDMZ (I2 & ESnet)



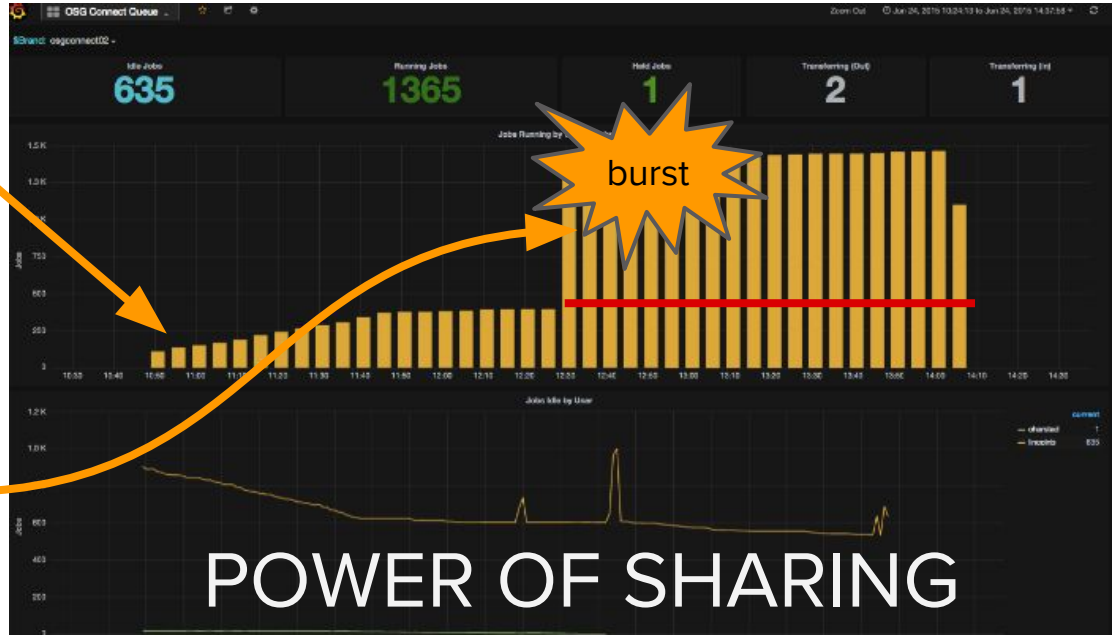




Campus Users (@ Clemson) + OSG

Submission to Palmetto cluster (local)

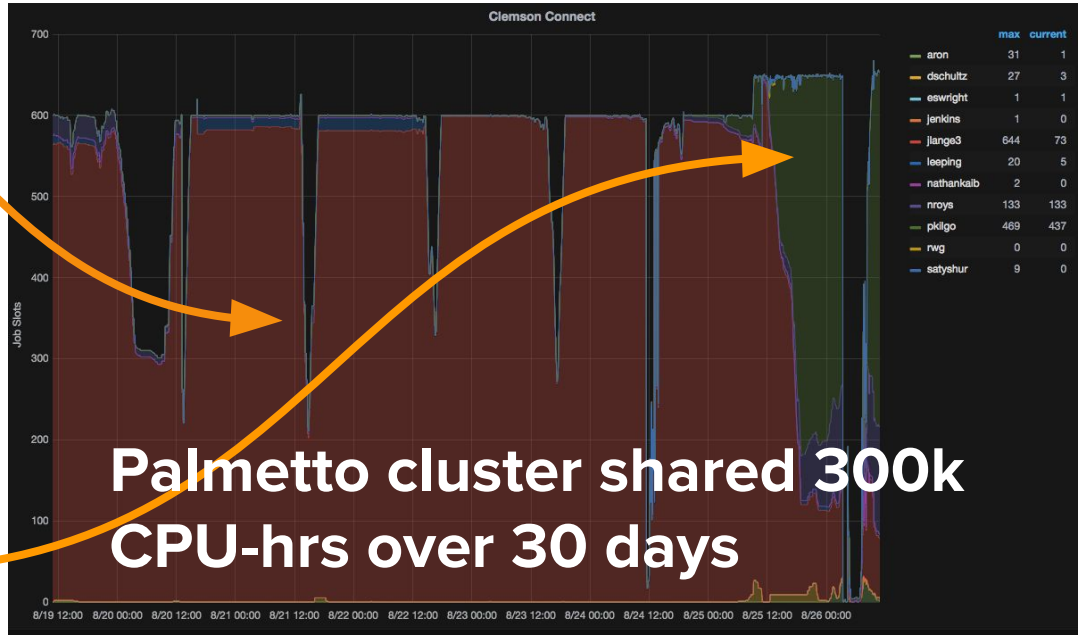
add OSG nodes

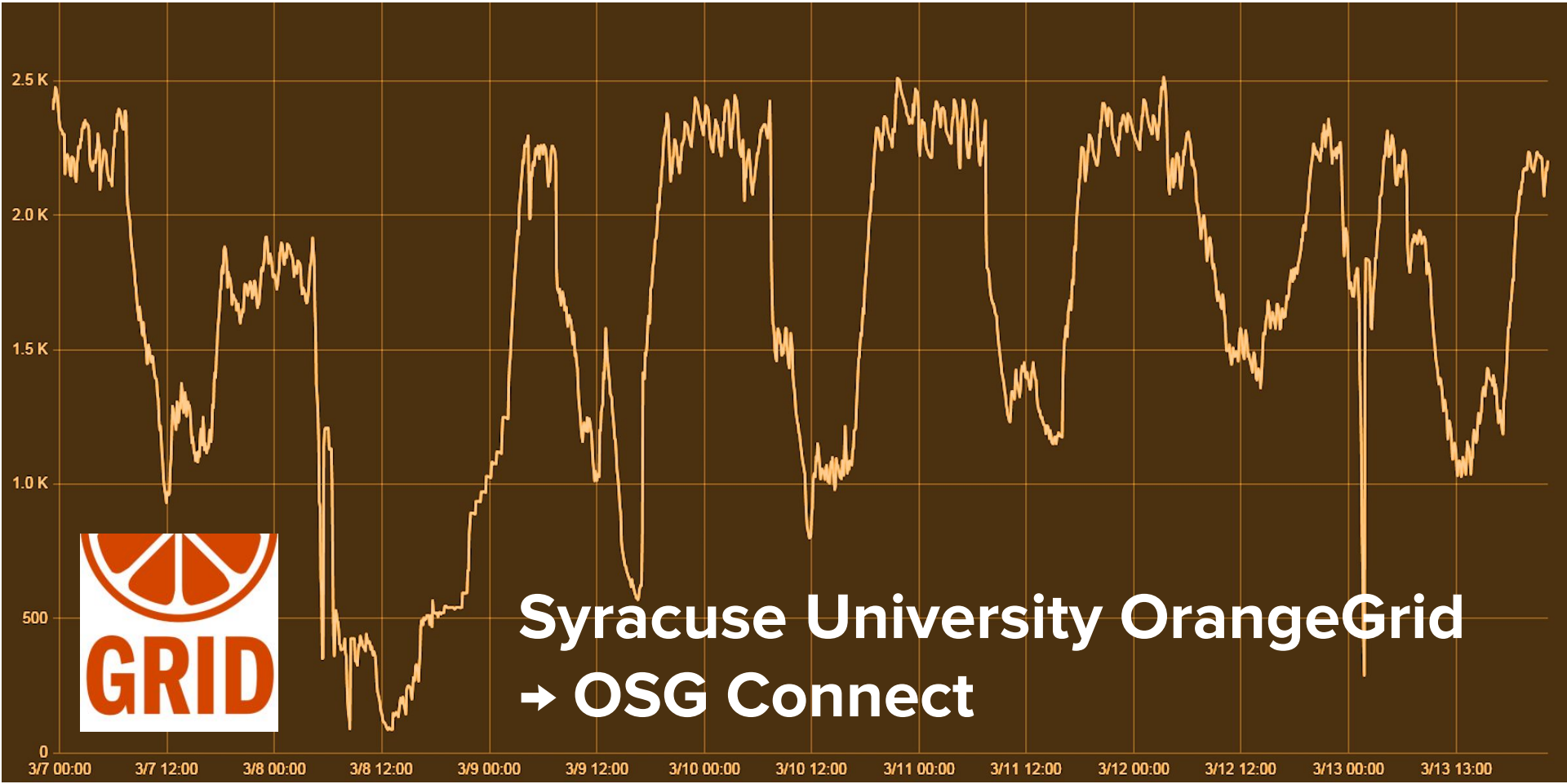


Sharing local resources with communities

GLOW (UWisconsin-based campus researchers)

OSG Connect



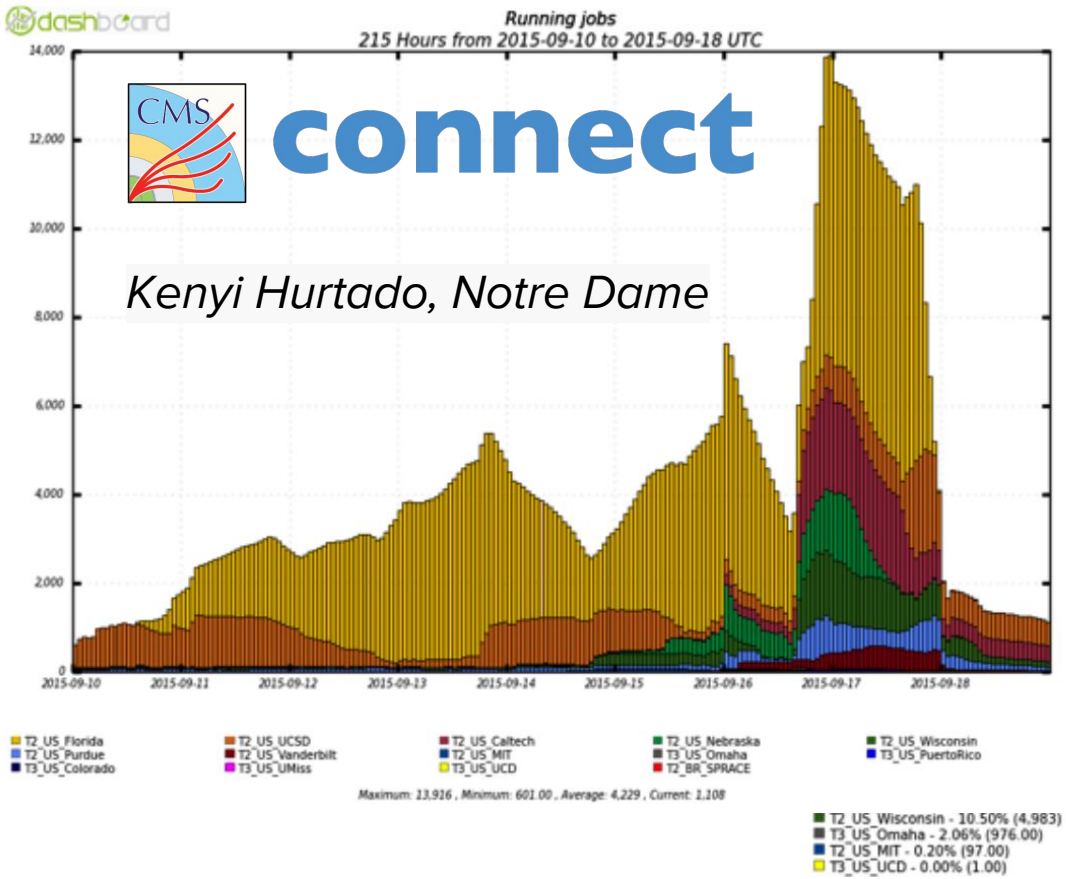

2.5 K
2.0 K
1.5 K
1.0 K
500
0
3/7 00:00
3/7 12:00
3/8 00:00
3/8 12:00
3/9 00:00
3/9 12:00
3/10 00:00
3/10 12:00
3/11 00:00
3/11 12:00
3/12 00:00
3/12 12:00
3/13 00:00
3/13 13:00
GRID
Syracuse University OrangeGrid
→ OSG Connect

CMS Connect

Kenyi Hurtado, Notre Dame

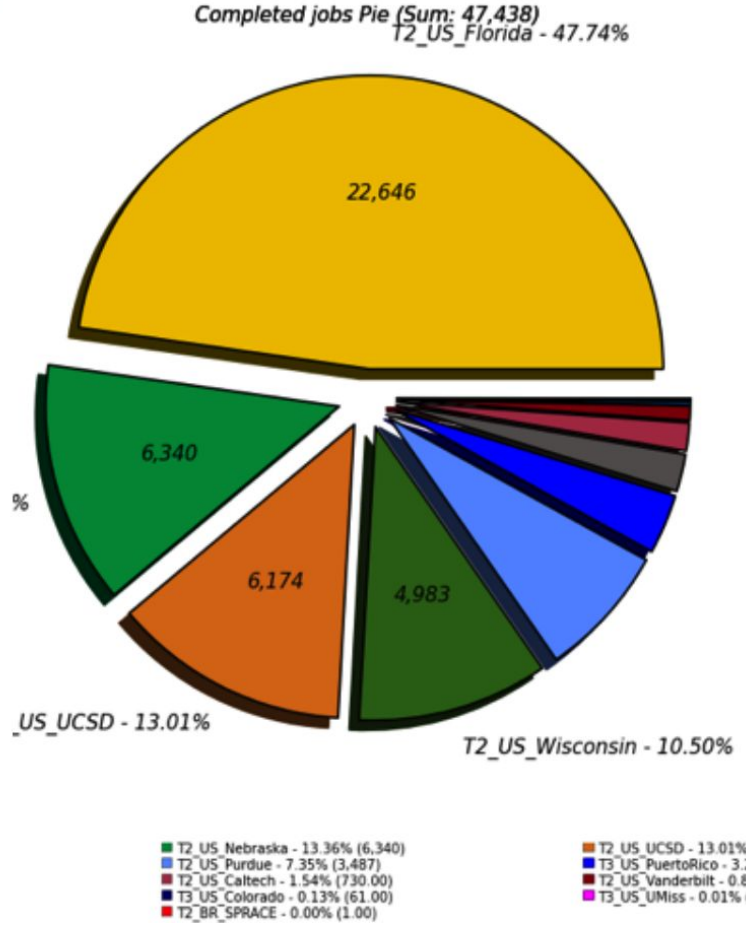
This is a complement to the existing tools in CMS with access to the same CMS Global resources for condor-like workflows that are frequently submitted to local clusters like e.g the LPC CAF at Fermilab or local University CMS analysis clusters.

T2 US Sites historical view example

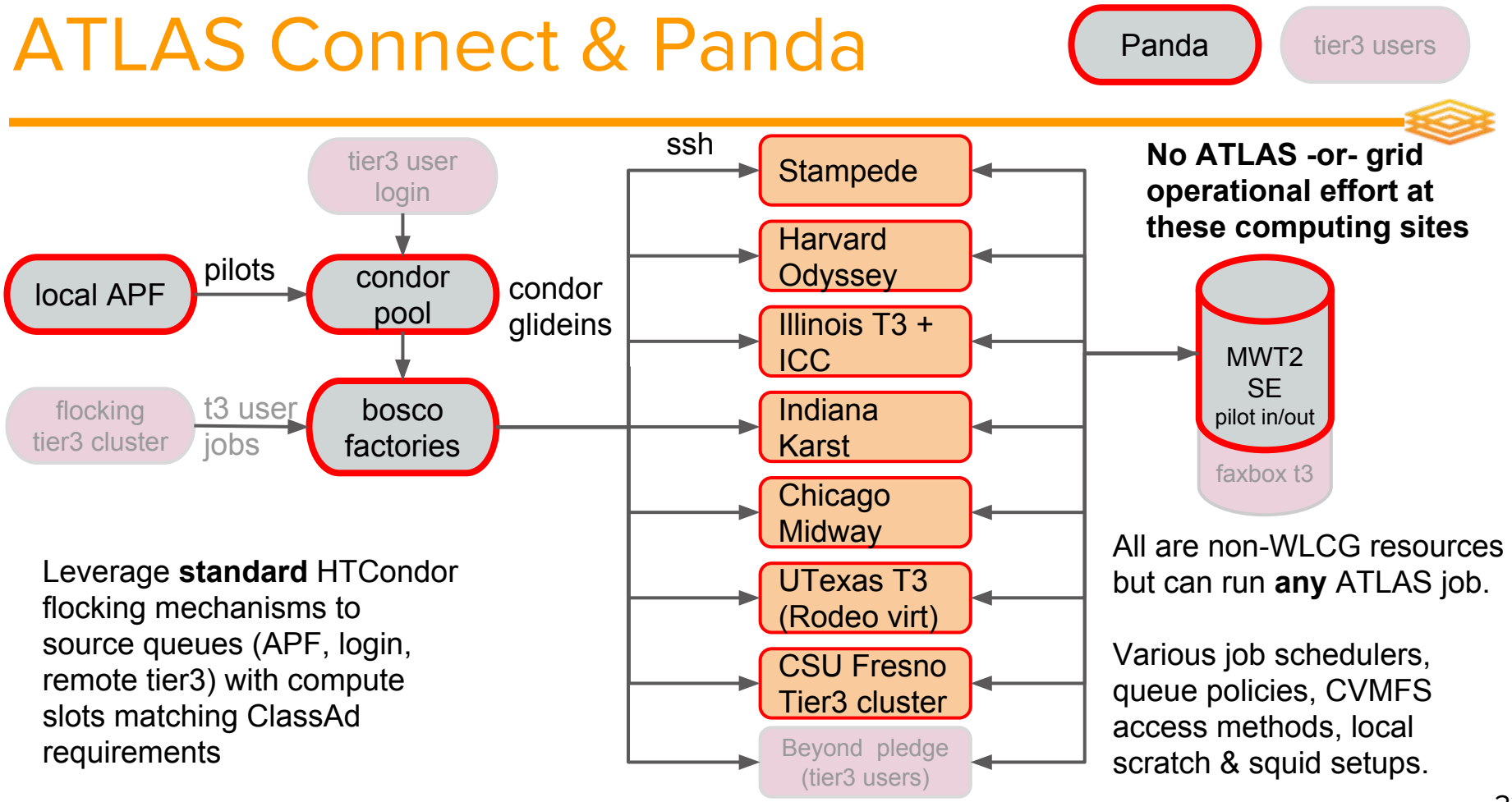


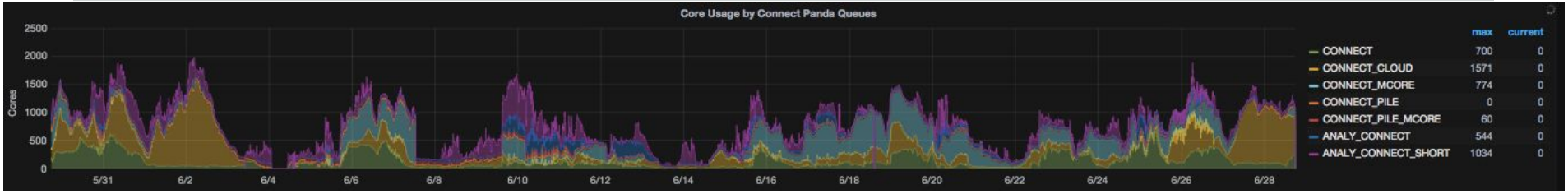
CMS connect

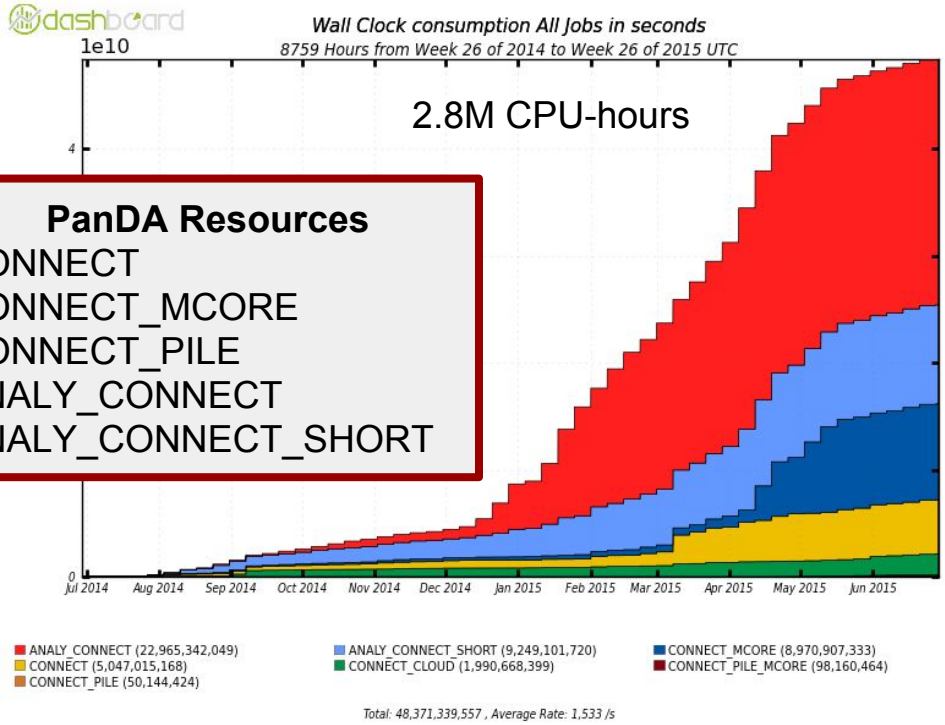
Kenyi Hurtado, Notre Dame



ATLAS Connect & Panda







PanDA Resources

CONNECT
CONNECT_MCORE
CONNECT_PILE
ANALY_CONNECT
ANALY_CONNECT_SHORT

Current scale:
~ ¼ of typical Tier2

~2000 running jobs at peak

But at low operating cost

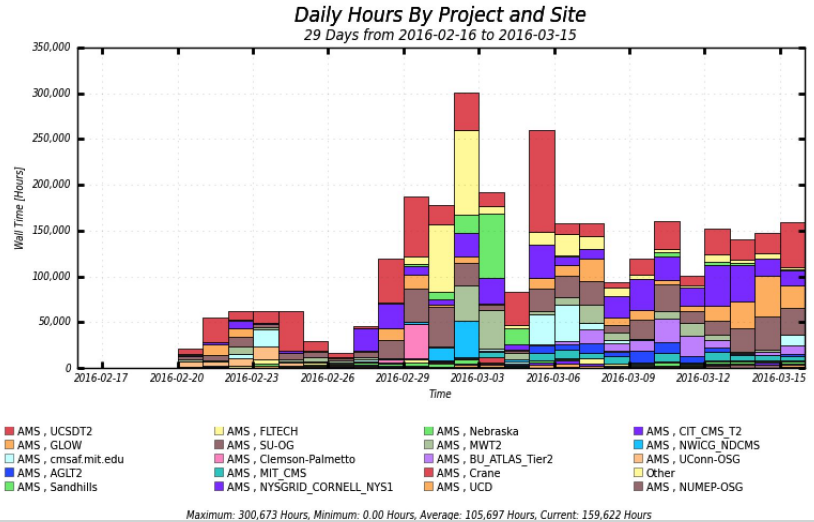
Only limited by available allocations

Easy to plug in additional resources or grow with new allocations

Present @ next US ATLAS Institution Board meeting to solicit interest from university research computing centers

MIT OSG Campus Infrastructure

- Supporting AMS space station experiment
- Up and running in hours - 18k two days later
- Plans are to run Bosco to other campus resources
- Eventually deploy HTCondor-CE



Campus Cyberinfrastructure futures

Leverage!

- Leverage investments from innovations and other NSF investment, e.g. from ACI DIBBS, CC-DNI, CC-*
- Push deployment of StashCache as primary delivery mechanism for campus PIs
- Hosted HTCondor-CE-Bosco
- Future projects - e.g. “Ubiquitous CI substrate”

San Diego State University
Stanford University
Texas A&M University
The Florida State University
The University of California, Davis
The University of Florida
The University of North Carolina at Chapel Hill
The University of Oregon
The University of Southern Californi's Information Sciences Institute (ISI)
University of Wisconsin–Madison
Wayne State University
Auburn University
Case Western Reserve University
Dartmouth College
Lehigh University
North Carolina State University
Oregon State University
Pittsburgh Supercomputing Center
Stephen F. Austin State University
Texas A&M University–Corpus Christi
The Johns Hopkins University
The University of California, Berkeley
The University of California, Santa Cruz
The University of Chicago
The University of Connecticut
The University of Connecticut Health Center
The University of Dayton
The University of Houston
The University of Idaho
The University of Illinois at Urbana–Champaign
The University of Michigan
The University of Montana
The University of New Hampshire
The University of Notre Dame
The University of Pennsylvania
The University of Puerto Rico, Río Piedras
The University of Southern California
The University of Utah
The University of Virginia
Wittenberg University
Arizona State University
Carleton College
Earlham College
Juniata College
Keystone Initiative for Network Based Education and Research (KINBER)
Northern Illinois University
Northwest Indian College
Princeton University
Saint Anselm College
South Dakota State Universit
St. Olaf College
Texas State University
The Old Dominion University Research Foundation
The University of Arkansa
The University of California, Riverside
The University of Cincinnati
The University of Pittsburgh
The University of Puerto Rico at Mayagüez
The University of South Alabama
The University of South Dakota
The University of Texas at Arlington
The US National Center for Atmospheric Research
University of Louisiana at Lafayette
University of Maryland, Baltimore
Wabash College
Washington State University
West Virginia University
Multiple Award Winners (2012-2014)
Clemson University
Colorado State University
Duke University
Florida International University
Georgetown University
Indiana University
Kansas State University
Louisiana State University and Agricultural and Mechanical College
Renaissance Computing Institute (RENCI)
The California Institute of Technology
The Ohio State University
The Pennsylvania State University
The University of Central Florida
The University of Colorado Boulder
The University of Missouri
The University of Nebraska–Lincoln
The University of Oklahoma
The University of Tennessee

120
NSF ACI CC-* DNI
SciDMZ Campuses

Tulane University
University of Chicago
University of California, San Diego
University of Maryland
University of Washington
Vassar College
Virginia Polytechnic Institute and State University
Yale University
Vanderbilt University
The University of Hawai'i at Mānoa
The University of New Mexico
The University of Kentucky
The University of Massachusetts Lowell
Northwestern University
Syracuse University
The University of Wyoming
The University of California, Irvine
The University of Massachusetts Amherst
The University of Alabama in Huntsville
Cornell University
The University of Michigan
Washington University
Boise State University
Colorado State University - Pueblo
Franklin and Marshall College
Portland State University
Univ of Minnesota Twin Cities
Univ of Alabama Birmingham
University of Missouri Kansas City
University of Texas Dallas
University of Alabama Tuscaloosa
UCLA
Marshall University
Colorado School of Mines
Malone University
Northern New Mexico College
University of Arkansas Pine Bluff
Harrisburg Area Community College
Bucks County Community College
Utah Valley University
Baylor College of Medicine
University of Arizona

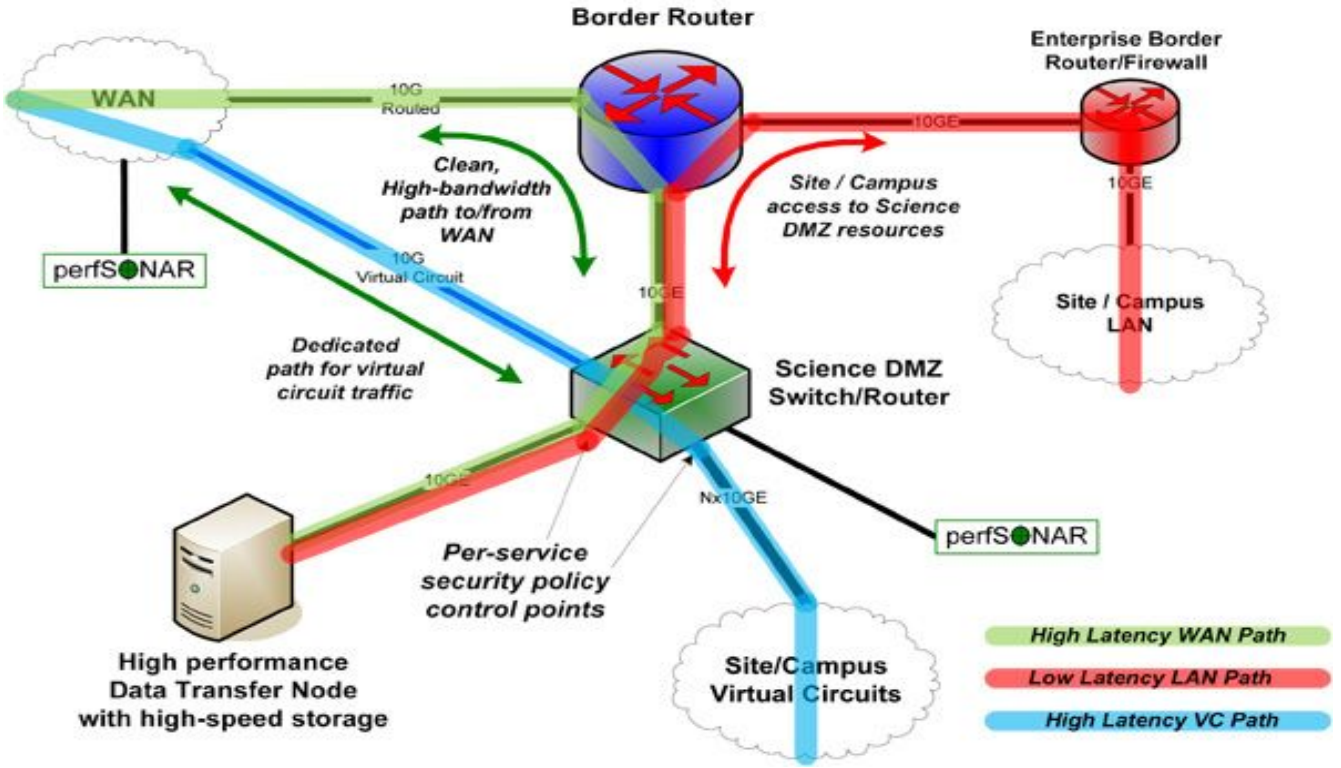
Ubiquitous & Easy “CI Substrate”

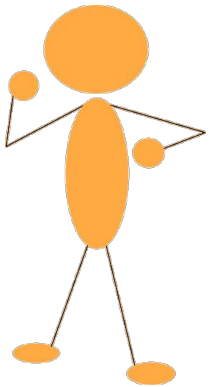
- Pioneer a new phase of advanced cyberinfrastructure deployment, allowing sites to flexibly evolve and sustain both on-premise and commercial cloud-based infrastructure
- Hosted services, such as CEs, data caches, squid, etc., could be centrally deployed onto “CI substrates” within a trusted CI zones and remotely operated, upgraded, and optimized for performance
- Extend to shared, opportunistic university clusters and cloud resources

Distributed Virtualized Data Centers

- Reduce IT footprint and ops burden
  - Centralize deployment & ops; reduce local admin cost
- Explore virtualized data center frameworks
  - E.g. container management over bare metal or VMs
- “Blue sky” goal
  - Establish a “trusted pattern” for a “CI substrate” on sites
  - Create distributed virtualized data center(s) overlaying the fabric substrate

SciDMZ

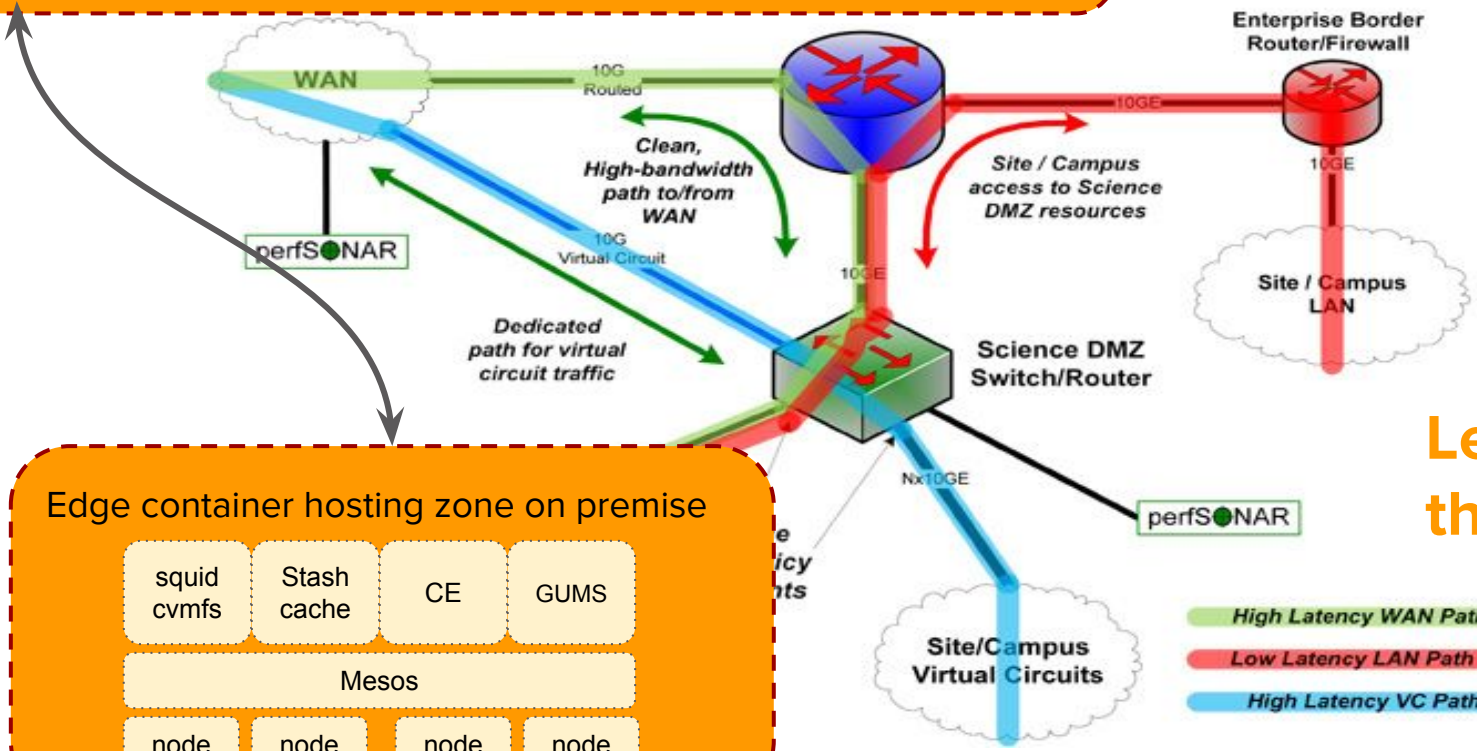

Border Router
Enterprise Border Router/Firewall
WAN
10G Routed
10GE
10GE
Clean, High-bandwidth path to/from WAN
Site / Campus access to Science DMZ resources
perfSONAR
10G Virtual Circuit
10GE
Site / Campus LAN
Dedicated path for virtual circuit traffic
Science DMZ Switch/Router
10GE
NxT0GE
perfSONAR
Per-service security policy control points
High performance Data Transfer Node with high-speed storage
Site/Campus Virtual Circuits
High Latency WAN Path
Low Latency LAN Path
High Latency VC Path



SuperDMZ

CI Connect central ops console:

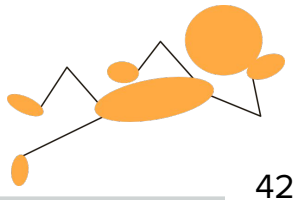
$ slate install osg-squid.3.3 --sites UAB Wellesley Wyo



Edge container hosting zone on premise

squid cvmfs	Stash cache	CE	GUMS
Mesos			
node	node	node	node

Let OSG do the work!



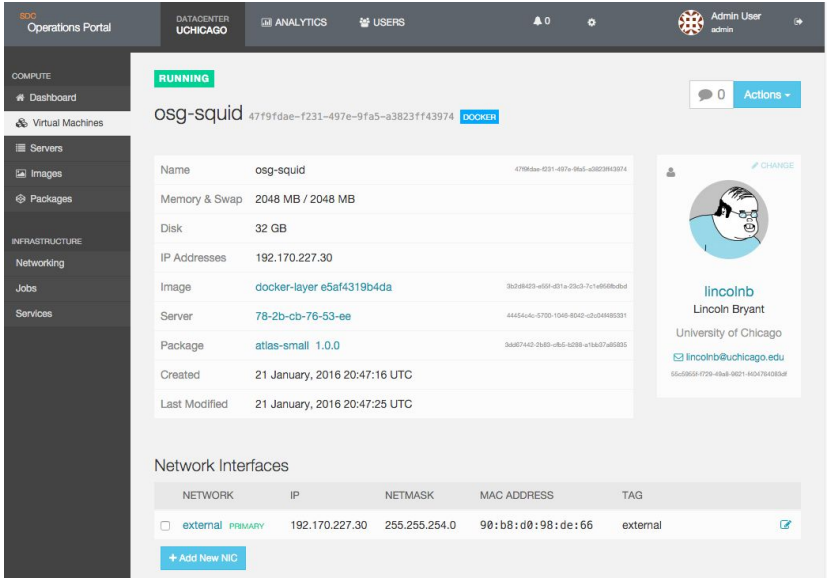
Example: Frontier Squid - Dockerfile

```
FROM lincolnbryant/osg-base-3.3-el6
MAINTAINER Lincoln Bryant <lincolnb@uchicago.edu>
# See https://twiki.grid.iu.edu/bin/view/Documentation/Release3/InstallFrontierSquid

RUN yum install -y frontier-squid initscripts

VOLUME ["/var/cache/squid"]

COPY customize.sh /etc/squid/customize.sh
RUN chown squid: /etc/squid/customize.sh && chmod +x /etc/squid/customize.sh

EXPOSE 3128 3401

CMD /sbin/runuser -s /bin/bash squid /usr/sbin/fn-local-squid.sh start && tail -f
/var/log/squid/*.log
```

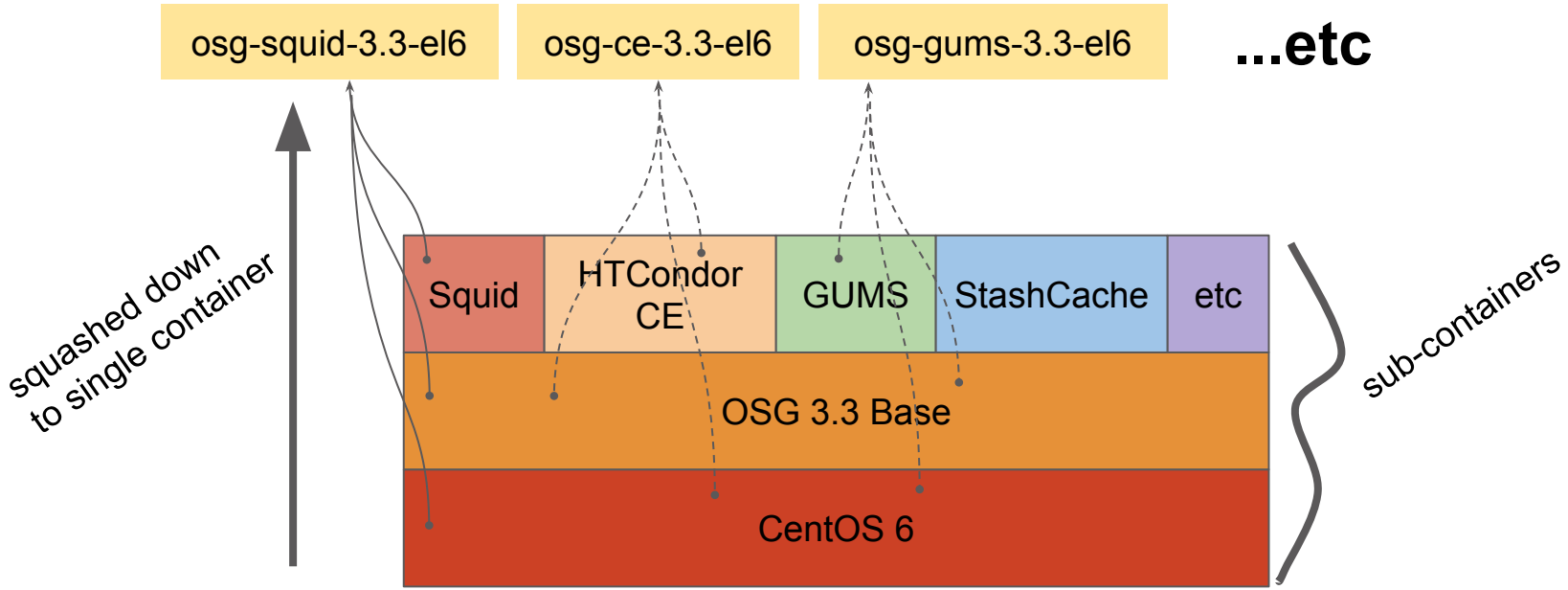
Example: Frontier Squid - Launching

- The container can be launched on another machine, or via Docker's remote API to a cloud resource

```
$ docker run -p 3128:3128/tcp -p 3401:
3401/udp -ti -e IP_BLOCKS="10.0.0.0/8
192.170.226.0/23" -e MEMORY_MB=2048 -e
CACHE_MB=32768 lincolnbryant/osg-squid-3.3-
el6

Generating /etc/squid/squid.conf

Initializing Cache...

2016/01/21 20:45:07| Creating Swap
Directories

Starting 1 Frontier Squid...

done


...
```



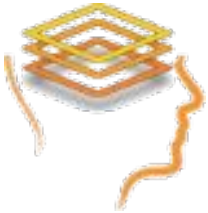
Goal: Containerize full OSG Stack



Ease delivery of OSG services where they need to be

Summary of User Support, XD, Campus

- Lowering barries to entry
  - OSG as a Service - for individual PI’s
  - Training fundamentals of distributed high throughput
  - CI Connect pattern for multi-campus collaborations
- Future: innovating ubiquitous CI substrate for small campuses



opensciencegrid

user-support@opensciencegrid.org

support.opensciencegrid.org

www.opensciencegrid.org/links