OSG User Support

March 14, 2014

Chander Sehgal

Mission of OSG User Support

The goal of the User Support area is to enable science and research communities from their initial introduction to the OSG to production usage of the Distributed High Throughput Computing (DHTC) services and to provide ongoing support for existing communities' evolving needs.

➢ The recent focus has been to increase the access of US researchers (who are not already a member of an OSG community) to DHTC computing resources to enable their science.

User Support Focus Areas

Open Science Grid

1. Grow access to Distributed High Throughput Computing (DHTC) for US Researchers **

   ➢ OSG as XD level 2 Service Provider

   ➢ Login node for researcher who contact OSG directly

   ➢ Campus "on-ramp" to OSG production fabric via HTCondor flocking

2. Support new communities (e.g. Sites & VOs) who want to join and use OSG

3. Integrate enabling technologies for goals 1 & 2 above

** In addition to the DHTC access already provided by regional OSG teams such as GLOW, HCC, SBGrid, UCSD, UC3 …

Our Method

➢ Inform science communities about OSG and how best to join OSG and leverage DHTC in their own environment

➢ Active support for communities who want to join and use OSG
  ❑ Understand the goals of the research community
  ❑ Possibly embed our staff into their team for a limited period to help them adapt their applications to DHTC
  ❑ Support the new community in resolving technical issues and achieving production goals

➢ Serve as a general entry point to OSG for information and user support

User Support Team

Open Science Grid

Name	Institution	FTE
Mats Rynge	ISI	0.50
Marko Slyz	FNAL	0.60
Tanya Levshina	FNAL	0.25
Alex Zaytsev	BNL	0.10
Chander Sehgal	FNAL	0.40
	Total	**1.85**

❖ Just Added: 0.7 FTE at FNAL - Bo Jayatilaka
❖ Pending: 0.5 FTE at Nebraska
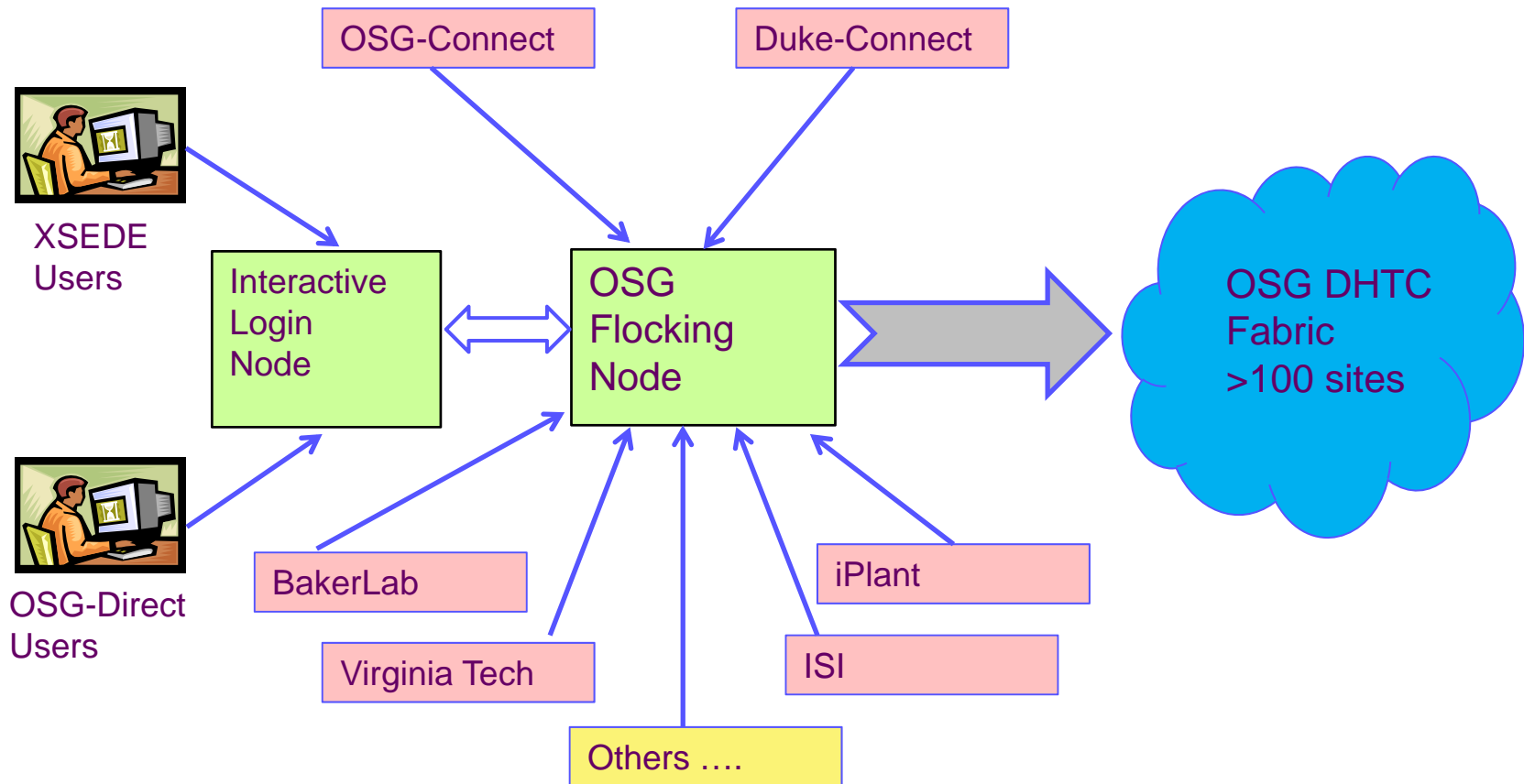
Grow access to DHTC for US Researchers

We endeavor to reduce the complexity of connecting to OSG and make it easier to do scientific computing using DHTC ➔ thus we hope to grow the community of US researchers who are actively leveraging OSG.

The backbone for this initiative is the OSG VO; this VO does not own any computing resources and only exists to harvest unused cycles at OSG sites (aka Opportunistic cycles) and make them available to researchers who are not already affiliated with an OSG community
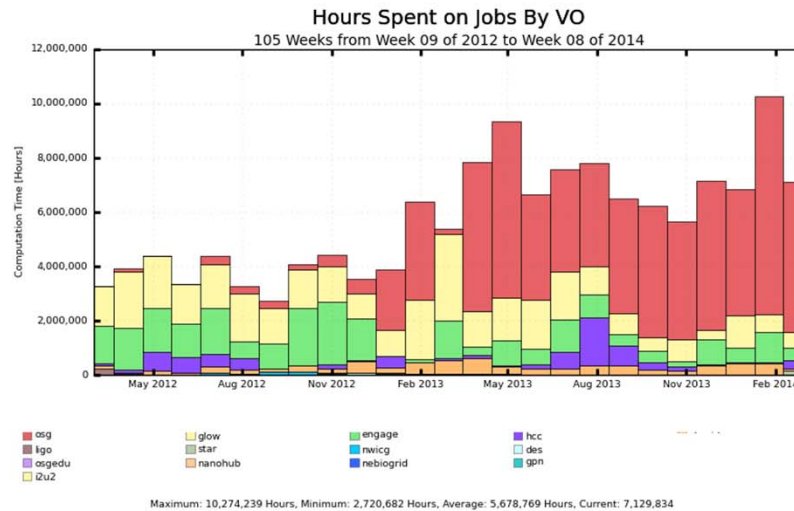
Our Approach
1. Easier to use Access Methods:
   ➢ Login node for XSEDE users and OSG-Direct users
   ➢ "On-ramp" to OSG for "bulk" connections from campus submit hosts and/or campus batch systems via flocking
2. Embedded support for researchers to adapt their applications to run in a distributed DHTC environment; help them -- but not do it for them.

Access to OSG DHTC Fabric via OSG VO

Open Science Grid

OSG-Connect

Duke-Connect

XSEDE Users

Interactive Login Node

OSG Flocking Node

OSG DHTC Fabric >100 sites

OSG-Direct Users

BakerLab

Virginia Tech

iPlant

ISI

Others ….

All access operates under the OSG VO using glideinWMS

Opportunistic Usage Trends

Hours Spent on Jobs By VO
105 Weeks from Week 09 of 2012 to Week 08 of 2014

Maximum: 10,274,239 Hours, Minimum: 2,720,682 Hours, Average: 5,678,769 Hours, Current: 7,129,834
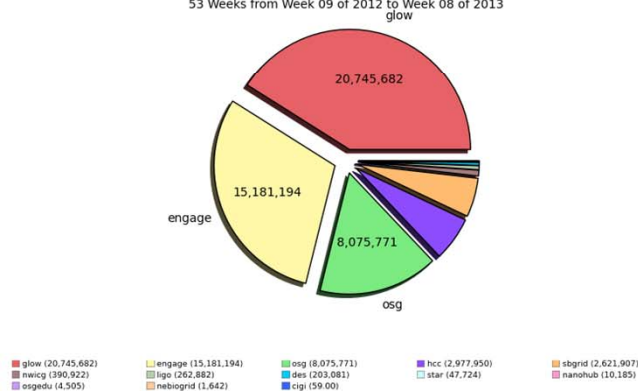
Opportunistic VOs last 2 years

OSG VO delivered 60M hours in the last year (Mar 2013 to Feb 2014) to US Researchers; compared to 8M hours in prior year (Mar 2012 to Feb 2013)
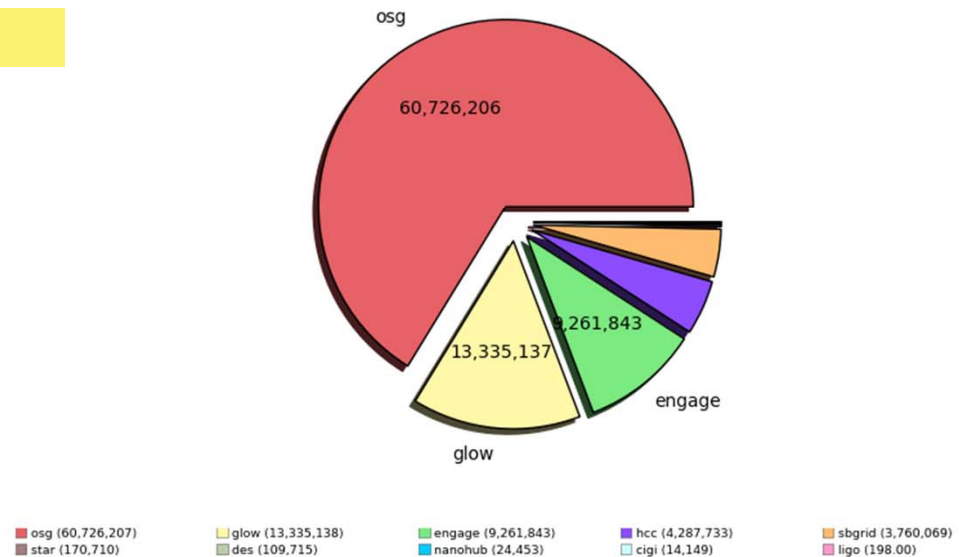

Wall Hours by VO (Sum: 50,523,504 Hours)
53 Weeks from Week 09 of 2012 to Week 08 of 2013

glow
20,745,682
15,181,194
engage
8,075,771
osg

glow (20,745,682)   engage (15,181,194)   osg (8,075,771)   hcc (2,977,950)   sbgrid (2,621,907)
nwicg (390,922)   ligo (262,882)   des (203,081)   star (47,724)   nanohub (10,185)
osgedu (4,505)   nebiogrid (1,642)   cigi (59.00)

March 2012 to Feb 2013

Wall Hours by VO (Sum: 91,690,215 Hours)
53 Weeks from Week 08 of 2013 to Week 08 of 2014

osg
60,726,206
9,261,843
engage
13,335,137
glow

osg (60,726,207)   glow (13,335,138)   engage (9,261,843)   hcc (4,287,733)   sbgrid (3,760,069)
star (170,710)   des (109,715)   nanohub (24,453)   cigi (14,149)   ligo (198.00)

March 2013 to Feb 2014

OSG-XD users March 2013 to Feb 2014

Project Name	PI	Institution	Field of Science	Wall Hours
TG-ATM130009	Phillip Anderson	University of Texas at Dallas	Atmospheric Sciences	41,798
TG-ATM130015	Phillip Anderson	University of Texas at Dallas	Atmospheric Sciences	77,169
TG-CCR120041	Luca Clementi	San Diego Supercomputer Center	Computer and Computation Research	13
TG-CHE130091	Paul Siders	University of Minnesota; Duluth	Chemistry	83,435
TG-CHE130103	Jeremy Moix	Massachusetts Institute of Technology	Chemistry	48,769
TG-DMR130036	Emanuel Gull	University of Michigan	Materials Research	212,059
TG-DMS120024	Benjamin Ong	Michigan State University	Mathematical Sciences	68,908
TG-IBN130001	Donald Krieger	University of Pittsburgh	Biological Sciences	28,842,307
TG-IBN130008	Jorden Schossau	Michigan State University	Biological Sciences	9,042
TG-IRI130016	Joseph Cohen	University of Massachusetts; Boston	Information; Robotics; and Intelligent Systems	70,536
TG-MCB090163	Michael Hagan	Brandeis University	Molecular Biosciences	33,786
TG-MCB090174	Shantenu Jha	Rutgers	Molecular Biosciences	128
TG-MCB100109	Lillian Chong	University of Pittsburgh	Molecular Biosciences	264,362
TG-MCB120070	Joseph Hargitai	Albert Einstein College of Medicine	Molecular Biosciences	84
TG-MCB130072	Robert Quick	Indiana University	Molecular Biosciences	16
TG-OCE130029	Yvonne Chan	University of Hawaii; Manoa	Ocean Sciences	23,878
TG-PHY110015	Pran Nath	Northeastern University	Physics	37
TG-PHY120014	Qaisar Shafi	University of Delaware	Physics	528,308
TG-STA110014S	Nancy Wilkins-Diehr	University of California-San Diego	Other	5
TG-TRA100004	Andrew Ruether	Swarthmore College	Other	444,374
TG-TRA120014	Pol Llovet	Montana State University	Cross-Disciplinary Activities	10,860
TG-TRA120041	Hanning Chen	George Washington University	Computer and Information Science	231
		22 Users	**Total**	**30,760,106**

Mostly in Support of various NSF Programs

OSG-Direct users March 2013 to Feb 2014

Project Name	PI	Institution	Field of Science	Wall Hours
BNLPET	Martin Purschke	Brookhaven National Laboratory	Medical Imaging	22,453
DeerDisease	Lene Jung Kjaer	Southern Illinois University	Biological Sciences	26,422
detectorDesign	John Strologas	University of New Mexico	Medical Imaging	444,405
Duke-QGP	Steffen A. Bass	Duke University	Nuclear Physics	2,453,095
ECFA	Meenakshi Narain	Brown University	High Energy Physics	1,744,646
EIC	Tobias Toll	Brookhaven National Laboratory	Accelerator Physics	410,594
IU-GALAXY	Robert Quick	Indiana University	Bioinformatics	627,204
OSG-Staff	Chander Sehgal	Fermilab	Computer Science	44,023
Pheno	Stefan Hoeche	SLAC	High Energy Physics	712,820
RIT	P. Stanislaw Radziszowski	Rochester Institute of Technology	Computer Science	721,291
SNOplus	Joshua R Klein	University of Pennsylvania	Physics - Neutrino	489
Snowmass	Meenakshi Narain	Brown University	High Energy Physics	9,972,536
SPLINTER	Robert Quick	Indiana University	Medicine	3,190,027
UMich	Paul Wolberg	University of Michigan	Microbiology	1,347,361
UPRRP-MR	Steven Massey	Universidad de Puerto Rico (UPRRP)	Bioinformatics	714,257
		15 Users	**Total**	**22,431,622**

Linked to various DOE programs

Sites & VOs

Assist new communities in joining OSG and enabling them as resource providers and as user communities wanting to use OSG resources.

- ➤ Enabled SURAgrid as new VO
- ➤ Assisting SNOplus & XENON US Collaborators
- ➤ Transitioned users as Engage VO wraps-up

- ➤ PNNL as new site in OSG (and supporting Belle II)
- ➤ OSC as new site in OSG
- ➤ UMD-IGS as new site in OSG

Upcoming
- SLU as a new site
- KSU as a new site

Technology: Public Storage

Smaller VOs have continued to struggle with the problem of data location when their jobs can land on any of >100 sites in a distributed environment like OSG. By using building blocks from iRODS, we provide a "basic" OSG public storage capability

- Enable non-LHC VOs whose computation requires "large" data to use OSG sites more easily
- Ease the task of VO data management:
  - Moving data and software to the sites for caching
  - Retrieving the output data from the sites
  - Providing metadata catalog
- Demonstrated use cases: EIC, Pheno, Snowmass, etc.
- Available as an OSG VO service today; and we plan to move the hosting of this service at the GOC (Indiana) to improve robustness

OSG Public Storage Architecture
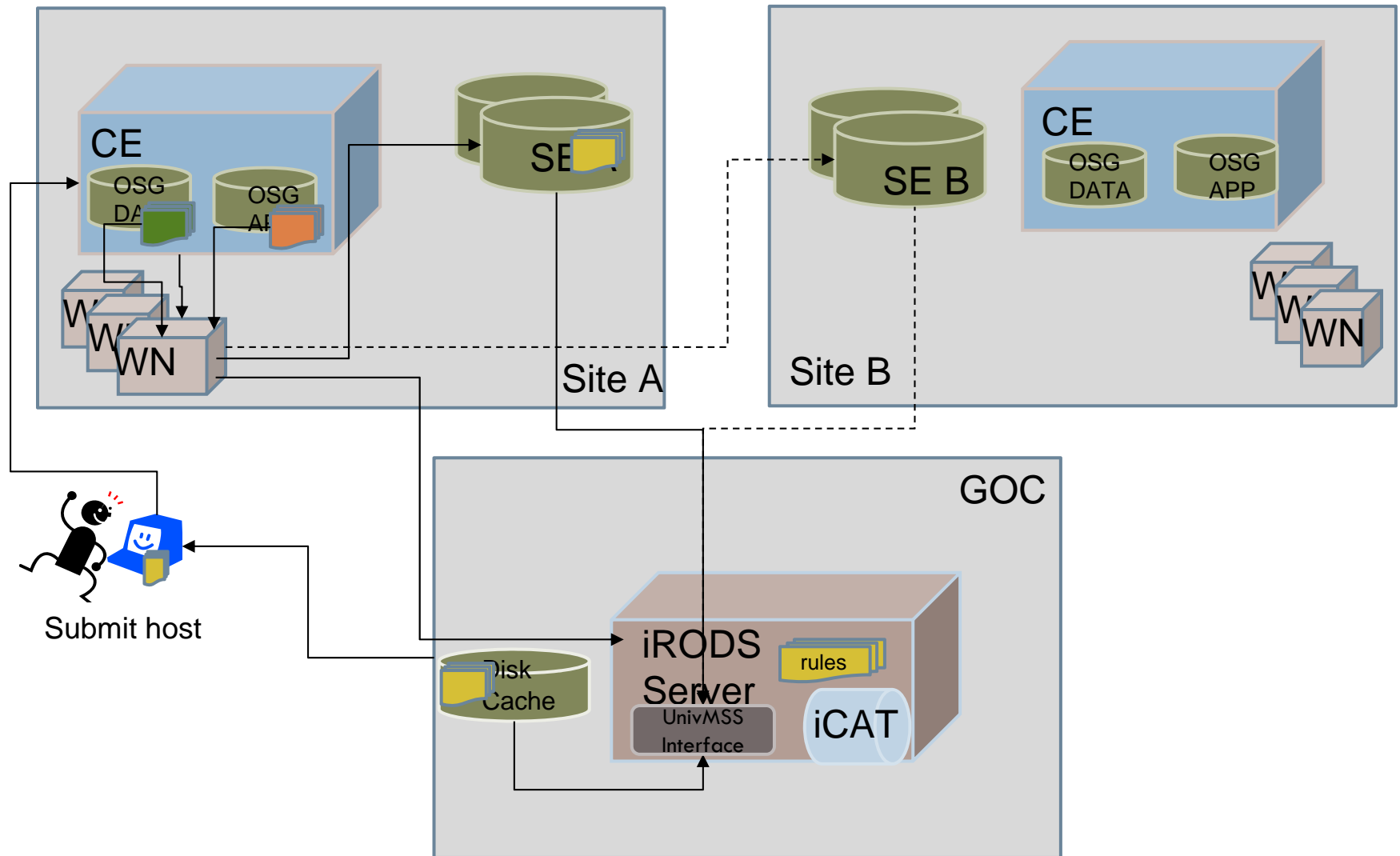
Technology: Field of Science Accounting

➢ Sites have long expressed a need to better understand who is using their sites? And what science is being done?
  – For single science VOs (e.g. atlas, cms, sbgrid...) there was no issue
  – For multi-science VOs (e.g. osg, engage, hcc, glow, uc3….) we have not had a good solution

➢ In partnership with the Gratia project, implemented the Field-of-Science accounting tools
  1. Whenever we enable a new research group, the sponsoring VO or Campus Grid registers the project in OIM
  2. The user includes a line in their job identifying their ProjectName (e.g. +ProjectName = "Snowmass")
  3. The gratia probe at the submit host send records to the Gratia DB including the ProjectName for each job
 4. Appropriate Gratia queries can be used to provide different views of the usage

Example ProjectName in OIM

Name	Snowmass
Description	Simulate hundreds of millions of high-energy proton proton collisions, which mimic the collisions expected at future hadron colliders. This simulated data is used to assess the physics potential of future colliders, allowing US decision makers and funding agencies to prioritize future physics projects.
Organization	Brown University
Department	Physics
Sponsor Campus Grid	OSG-XD
Principal Investigator	Meenakshi Narain
Field Of Science	High Energy Physics

FOS Accounting - Example Report

Usage by all Projects from March 2013 to Feb 2014 – in OSG-VO eco-system

Wall Hours by VO (Sum: 55,549,527 Hours)
366 Days from Week 08 of 2013 to Week 08 of 2014

TG-IBN130001

28,889,270
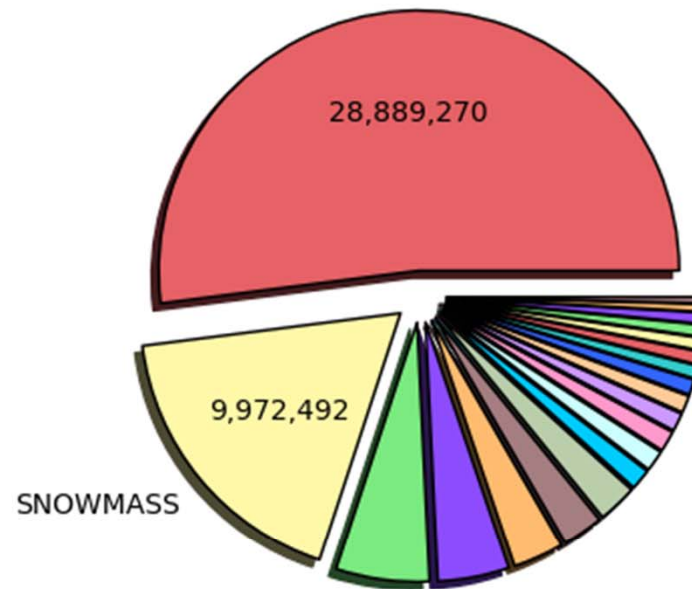
9,972,492

SNOWMASS

🟥 TG-IBN130001 (28,889,270)	🟨 SNOWMASS (9,972,493)	🟩 SPLINTER (3,222,799)	🟪 DUKE-QGP (2,460,614)
🟧 ECFA (1,744,646)	🟫 Other (1,405,658)	⬜ UMICH (1,357,651)	🟦 PHENO (726,135)
⬜ RIT (721,291)	🟪 UPRRP-MR (714,257)	🟪 IU-GALAXY (627,241)	🟧 DUKE-4FERMION (589,469)
🟦 TG-PHY120014 (528,308)	🟦 DETECTORDESIGN (444,405)	🟥 TG-TRA100004 (444,374)	🟨 DUKE (422,230)
🟩 EIC (410,594)	🟪 KNOWLEDGESYS (383,182)	🟧 TG-MCB100109 (264,362)	🟫 NESCENT (220,547)

Challenges and Opportunities

Open Science Grid

Major Challenges

➢ We don't have a good understanding of the OSG opportunistic eco-system and "optimal" methods to quantify and access these opportunistic resources – plan to apply new staff to undertake research in this area

➢ We need better methods to reach potential researchers who can benefit from access to OSG DHTC resources – exploring use of XSEDE campus champions as a channel

Other Opportunities

➢ Integrate a complete submit node software package (RPM) for research communities and campuses who want to administer their own login node and connect to OSG via flocking

➢ Establish a "library" of commonly used software packages and make them available (via CVMFS) to users

➢ Continue to improve methods for user applications that need to access large read-only data sets and produce large output datasets