Production Analyst Report

Bodhitha Jayatilaka

Fermilab
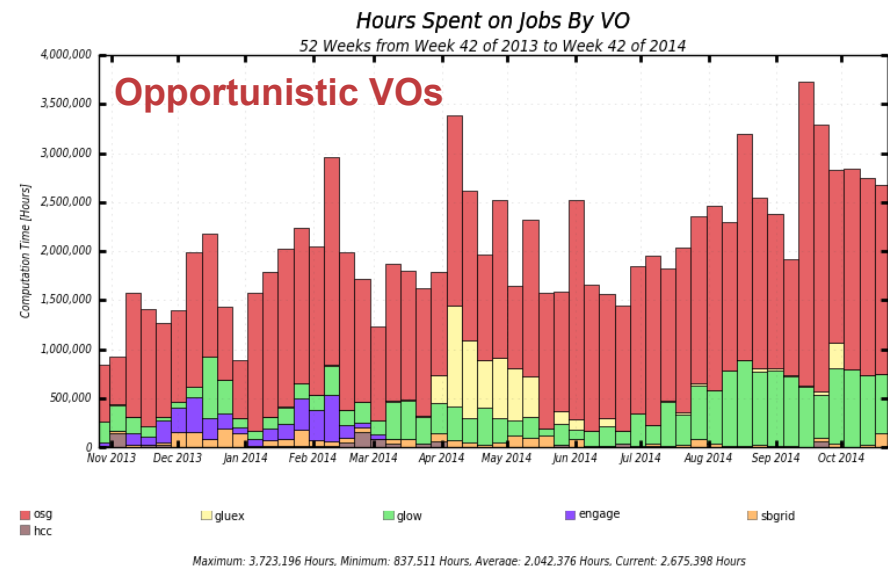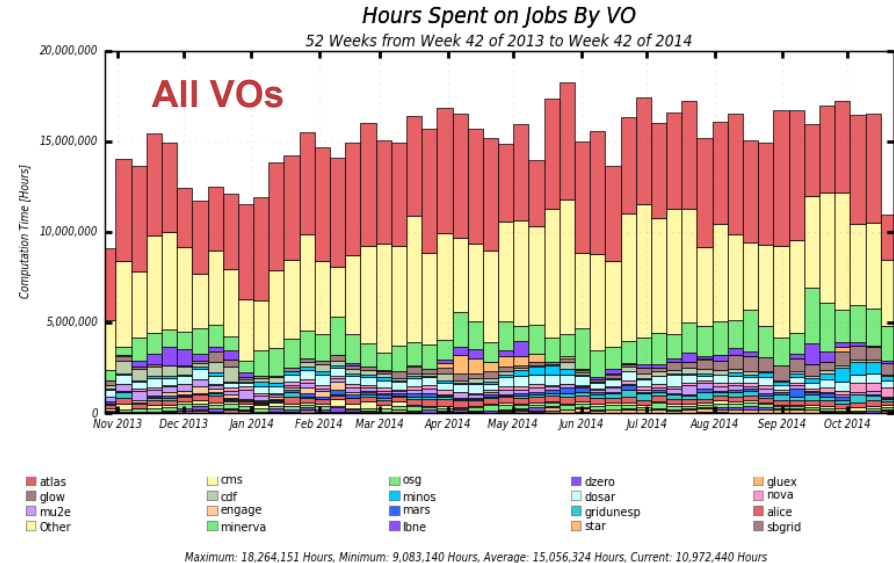
OSG Council Meeting

Norman, OK

October 22, 2014

What are we trying to do?

- **Goal**: *provide access to DHTC for US researchers who are not part of an existing virtual organization*

- Challenge #1: *Knowing how much is available*
  - Only real metric for capacity comes from historical availability
  - Understanding/predicting how availability can change

- Challenge #2: *Expanding capacity*
  - No single "opportunistic ecosphere" at the OSG
  - Have to address at a site-level

- Challenge #3: *Finding users*
  - Large groups are already part of VOs or using HPC
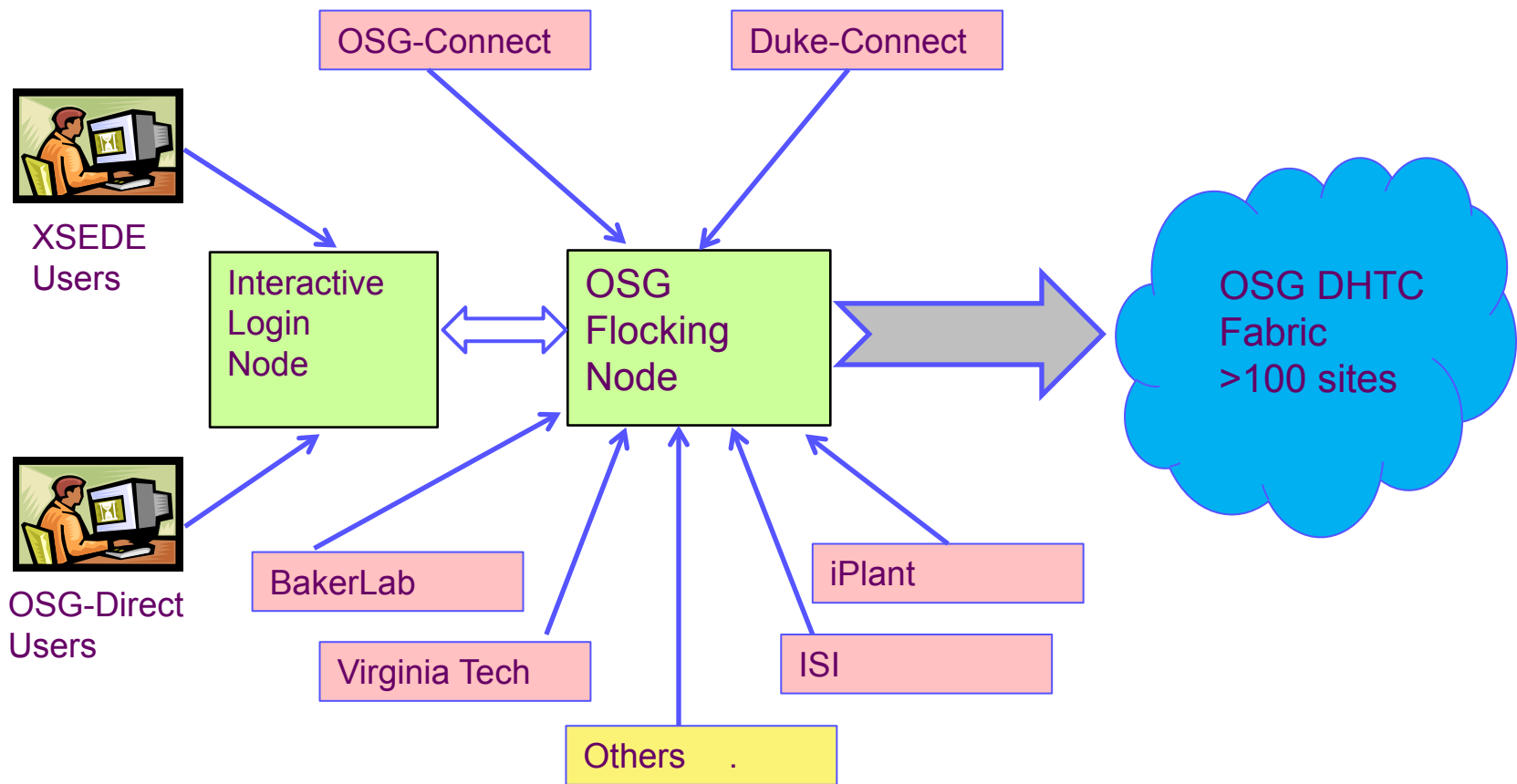 - Cost/benefit of setting up not advantageous to small users

A reminder

- Opportunistic VOs* constitute over 100M hours per year
  - OSG VO ~72M hours in 12 months ending 10/1/14

- Aim to grow the Open Facility
  - 20% of all wall hours at the OSG
  - 100M hours for OSG VO
  - We believe these goals can be achieved **without** hindering owner-VO cycles.

*Opportunistic = osg, hcc, glow, engage, sbgrid, gluex

Hours Spent on Jobs By VO
52 Weeks from Week 42 of 2013 to Week 42 of 2014

All VOs

atlas	cms	osg	dzero	gluex
glow	cdf	minos	dosar	nova
mu2e	engage	mars	gridunesp	alice
Other	minerva	lbne	star	sbgrid

Maximum: 18,264,151 Hours, Minimum: 9,083,140 Hours, Average: 15,056,324 Hours, Current: 10,972,440 Hours

Hours Spent on Jobs By VO
52 Weeks from Week 42 of 2013 to Week 42 of 2014

Opportunistic VOs

| osg | gluex | glow | engage | sbgrid |
| hcc | | | | |

Maximum: 3,723,196 Hours, Minimum: 837,511 Hours, Average: 2,042,376 Hours, Current: 2,675,398 Hours

On-ramp to the OSG

Since last time...

- Added several large sites in past three months (e.g. MIT, UFlorida, BU)

- xd submit host upgrade in early September
  - Eliminated several I/O bottlenecks

- Averaging 320k hours/day since submit host upgrade
  - Well on-track to 100M/year

- Several recent lulls in LHC processing
  - OSG VO effectively demand-limited

Daily Hours By Project
93 Days from Week 29 of 2014 to Week 42 of 2014
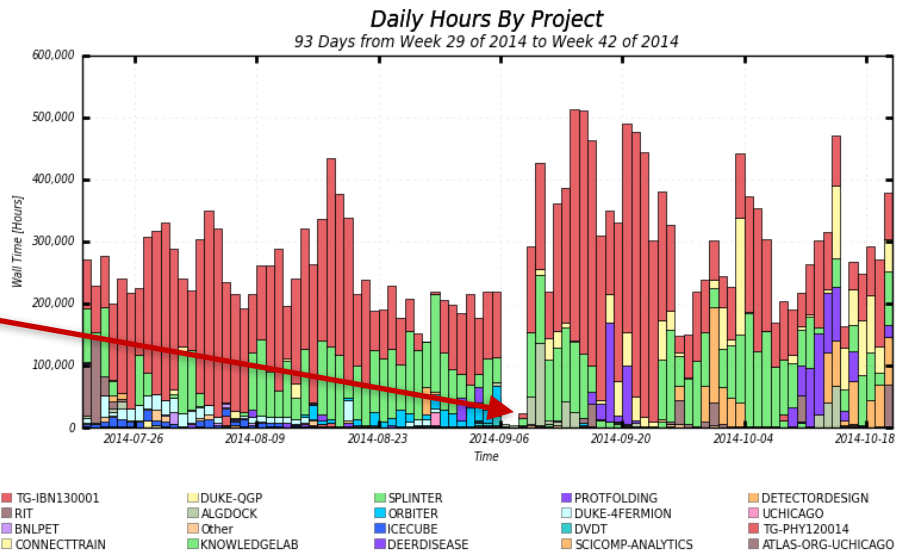
Maximum: 513,858 Hours, Minimum: 2,963 Hours, Average: 271,754 Hours, Current: 379,649 Hours

Wall Hours by Facility (Sum: 26,186,226 Hours)
14 Weeks from Week 29 of 2014 to Week 42 of 2014
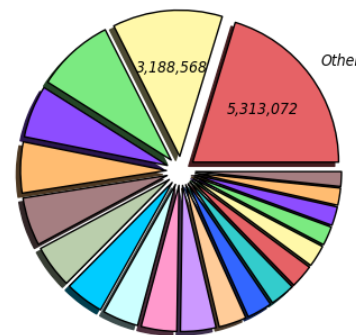
Adding "incentive"

Open Science Grid

Opportunistic Resources provided by the top 10 OSG Sites for the OSG Open Facility (2014-09-01 - 2014-9-30).

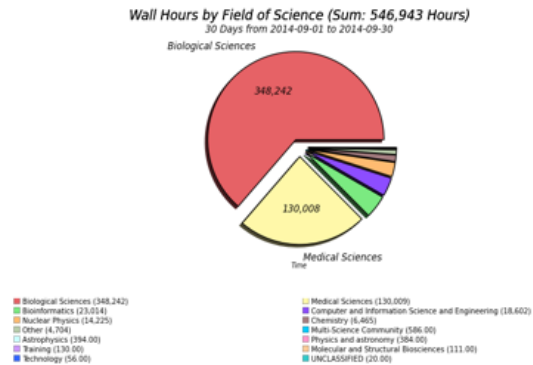| SITE	INSTITUTION (ORG)	RECENT PERIOD RANK	RECENT PERIOD HOURS	PRIOR PERIOD RANK	PRIOR PERIOD HOURS
USCMS-FNAL-WC1	Fermi National Accelerator Laboratory (USCMS Tier1)	1	1,295,720	7	395,587
MWT2	University of Chicago, Indiana University, University of Illinois, Urbana-Campaign (USATLAS Midwest Tier2)	2	1,291,273	4	725,769
UCSDT2	University of California San Diego (USCMS Tier2)	3	1,050,015	1	1,091,183
FNAL_FERMIGRID	Fermi National Accelerator Laboratory (Campus Grid gateway)	4	832,573	3	853,682
MIT_CMS	Massachusetts Institute of Technology (USCMS Tier2)	5	546,943	8	351,981
CIT_CMS_T2	California Institute of Technology (USCMS Tier2)	6	532,943	2	920,799
UFlorida-HPC	University of Florida (USCMS Tier2)	7	509,245	63	0
Tusker	The Holland Computing Center at the University of Nebraska-Omaha (HCC Campus Grid)	8	488,855	6	520,391
Nebraska	The Holland Computing Center at the University of Nebraska-Licoln (USCMS Tier2)	9	479,262	5	697,329
UConn-OSG	University of Connecticut at Storrs (GLUEX)	10	191,622	11	158,809

Top 10 sites (included in OSG newsletter starting this month)

Example **monthly site report** showing individual projects run at site

Open Science Grid

SITE MIT_CMS (2014-09-01 - 2014-9-30):

MIT_CMS has provided 546,917 wall hours to the OSG Open Facility during 2014-09-01 - 2014-9-30.
Last period this site provided 351,981 wall hours.

Wall Hours by Field of Science (Sum: 546,943 Hours)
30 Days from 2014-09-01 to 2014-09-30

- Biological Sciences (348,242)
- Bioinformatics (23,014)
- Nuclear Physics (14,225)
- Other (4,704)
- Astrophysics (394.00)
- Training (130.00)
- Technology (56.00)
- Medical Sciences (130,009)
- Computer and Information Science and Engineering (18,602)
- Chemistry (6,465)
- Multi-Science Community (586.00)
- Physics and astronomy (384.00)
- Molecular and Structural Biosciences (111.00)
- UNCLASSIFIED (20.00)

PROJECT NAME	FIELD OF SCIENCE	PI	VO	SUBMISSION HOST	WALL HOURS
TG-IBN130001	Biological Sciences	Donald Krieger	osg	XD-LOGIN	348,242
SPLINTER	Medicine	Robert Quick	osg	XD-LOGIN	109,152
ProtFolding	Bioinformatics	Jinbo Xu	OSG-Connect	OSGCONNECT	22,599
DetectorDesign	Medical Imaging	John Strologas	osg	XD-LOGIN	20,857
Orbiter	Mathematics	Anton Betten	OSG-Connect	OSGCONNECT	17,632
Duke-QGP	Nuclear Physics	Steffen A. Bass	osg	XD-LOGIN	14,199
AIGDock	Chemistry	David Minh	OSG-Connect	OSGCONNECT	6,465
KnowledgeLab	Other	James Evans	OSG-Connect	OSGCONNECT	4,704

Grid at-a-glance

Site	Total	Percent Opportunistic	Opportunistic VOs				Opp Total	alice	argoneut	atlas
			osg	glow	hcc	sbgrid				
Total	66,866,854	18%	9,063,256	2,872,441	58,130	84,007	12,077,834	1,207,664	4,902	22,653,381
							0			
BNL-ATLAS	7,270,804	0%	12,314	12,606	428		25,348			7,243,035
USCMS-FNAL-WC1	7,262,273	19%	1,309,640	99,406	21	5,604	1,414,671			
FNAL_FERMIGRID	6,979,072	11%	785,470		100		785,570		4,902	
MWT2	6,276,315	25%	1,243,052	345,871	8,345		1,597,268			4,618,192
AGLT2	2,894,536	10%	125,289	156,981	1,604	18,628	302,502			2,592,034
Nebraska	2,749,221	35%	603,337	336,608	11,488		951,433			4
WT2	2,640,235	0%					0			2,639,158
CIT_CMS_T2	2,637,477	42%	560,717	519,225	7,955	12,329	1,100,226			
MIT_CMS	2,590,132	29%	707,365		11,499	30,074	748,938			
UCSDT2	2,501,575	53%	1,015,564	305,790	447	10,707	1,332,508			
OU_OSCER_ATLAS	2,196,143	5%	110,326				110,326			379,050
GLOW	1,844,302	9%	163,142		809		163,951			
BU_ATLAS_Tier2	1,838,380	7%	129,132				129,132			1,709,248
Tusker	1,551,295	83%	782,248	504,438	6,131		1,292,817			
UFlorida-HPC	1,521,795	36%	553,738				553,738			
Purdue-Rossmann	1,445,416	0%					0			
SWT2_CPB	1,236,987	1%	2,841	4,548			7,389			1,229,598
GridUNESP_CENTRAL	1,200,735	12%	91,249	52,469	55	2,219	145,992			
HU_ATLAS_Tier2	1,052,671	0%					0			1,045,873
FZU	929,242	0%					0			
Purdue-Carter	821,428	0%					0			
Purdue-Conte	717,363	0%					0			
NERSC-PDSF	670,150	0%					0	669,585		
Purdue-Hansen	630,460	0%					0			
Purdue-Hadoop	614,952	21%	62,136	67,344		344	129,824			
LC-glcc	538,079	0%					0	538,079		
SPRACE	485,025	32%	76,061	80,094		1,129	157,284			
UTA_SWT2	456,913	12%	29,428	25,111			54,539			402,374
OU_OCHEP_SWT2	421,000	0%					0			421,000
LUCILLE	373,815	0%					0			373,815
UColorado_HEP	366,150	0%					0			
Vanderbilt	326,853	2%		5,313			5,313			
UConn-OSG	279,509	91%	136,196	115,762	3,667		255,625			
Crane	222,293	100%	64,813	156,068	838		221,719			
FNAL_GPGRID_1	202,488	0%					0			

Link to full spreadsheet

Usage trends: projects
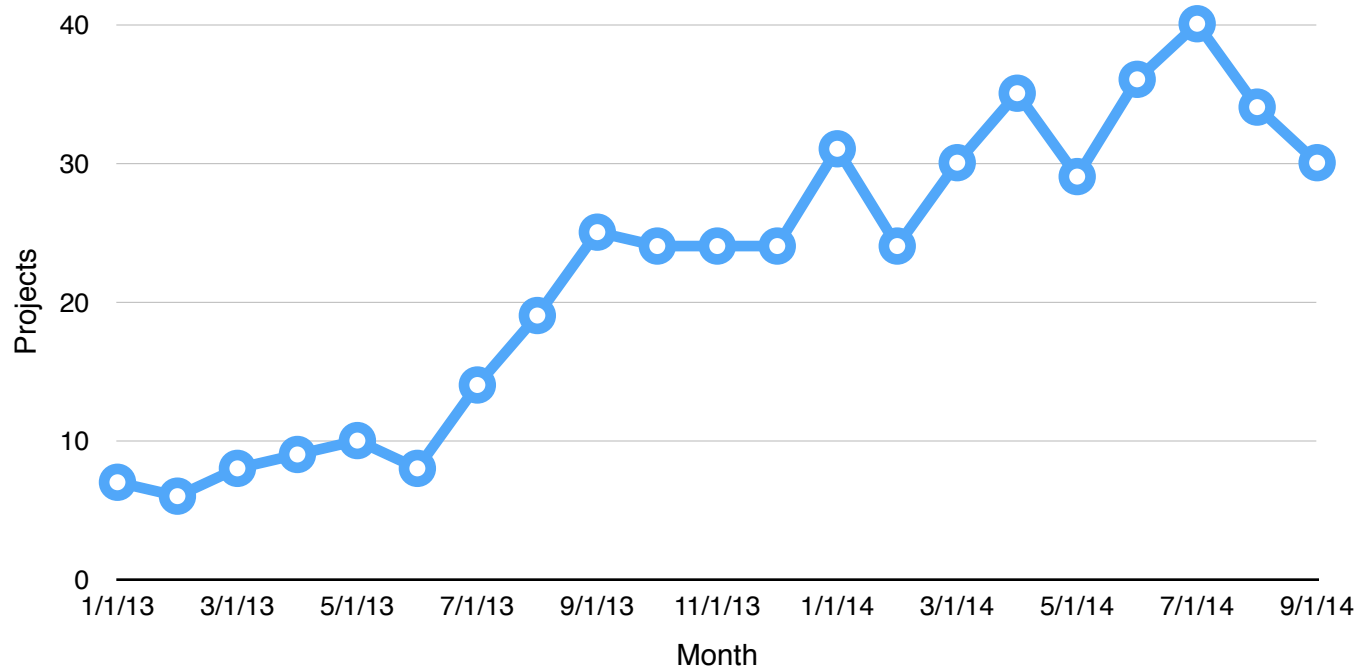
- ~30-40 unique projects running on OSG VO every month
- Number of users not growing as fast as hours delivered
  - Several periods during which OSG VO is demand-limited
  - Need to take more active approach in finding users?

ACI-REFs: introduction

- **A**dvanced **C**yber**i**nfrastructure **R**esearch and **E**ducation **F**acilitators

- Distributed across 6 campuses:
  - Clemson University, University of Hawaii, University of Southern California, University of Utah, University of Wisconsin and Harvard University

- NSF funded program designed to:
  - *"[forge] a nationwide alliance of educators to empower local campus researchers to be more effective users of advanced cyberinfrastructure (ACI). . . [and] work with the 'long tail' of ACI users—those scholars and faculty members who traditionally have not benefitted from the power of massively scaled cluster computing."*

Partnering with ACI-REFs

*Effort lead by **E. Harstad** (UNL)*

- By partnering with ACI-REFs, identify researchers who may benefit from OSG DHTC
  - Equip ACI-REFs with the tools and knowledge they need to feel comfortable recommending OSG to researchers at their institutions
  - Provide 2nd layer of support for ACI-REFs as they include OSG in their solutions for local campus researchers
- Progress
  - Step 1: Meet with ACI-REFs to establish partnership (done Oct. 8)
    - Present introduction to OSG; gauge feedback and interest
    - ACI-REFs enthusiastic and interested in an in-depth OSG tutorial
  - Step 2: Provide OSG tutorial covering common aspects of OSG usage (starting in early November)
    - One per ACI-REF campus (beginning with Clemson)

Conclusion

- Opportunistic usage at the OSG now exceeds 10M hours/month

- During periods of low LHC experiment computing activity, OSG VO is increasingly demand-limited

- Efforts moving to finding new potential users

  – ACI-REFs effort key in this process

- Plan to soon circulate policy proposal on opportunistic usage

- Efforts made in expanding OSG VO can be applied to other opportunistic VOs as well

  – *e.g.* Fermilab intensity frontier experiments exploring opportunistic sphere for peak usage

Backup

Increasing Opportunistic Cycles

1. Grow overall size of OSG

    – Rate of new sites approaching OSG is increasing

    – Existing ATLAS/CMS sites growing anticipating Run 2

2. Increase opportunistic share at existing sites

    • Are existing sites not providing all opportunistic hours they can

    • Are opportunistic jobs not configured to take advantage of existing resource

3. Eliminate inefficiencies in existing infrastructure

    • e.g., at Factory-level

Categorizing sites

- Broadly break sites down into three categories

  1. Sites accepting OSG VO jobs "fully opportunistically"

  2. Sites accepting OSG VO jobs with a cap/quota

  3. Sites not accepting OSG VO jobs

- Metrics used to determine

  – Glidein monitoring per site

  – Correlations between opportunistic VOs and dominant VO for a site

"Type 1"

"Type 2"
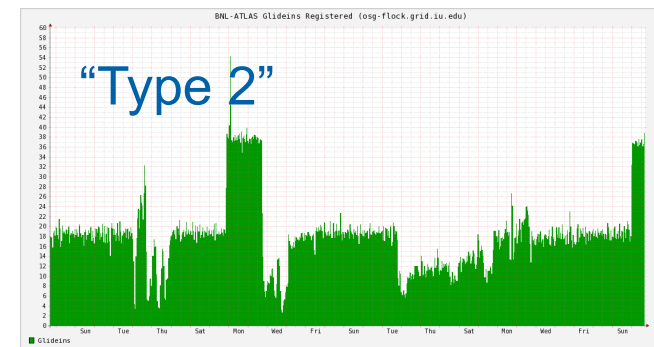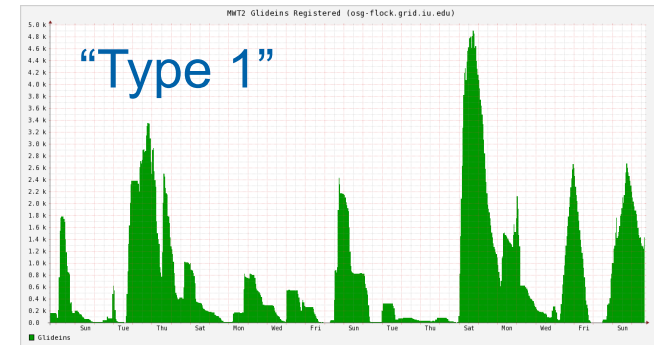
	cms	osg	engage
cms	1.00000	-0.83145	-0.68202
osg	-0.83145	1.00000	0.60380
engage	-0.68202	0.60380	1.00000

Projects using OSG VO

https://oim.grid.iu.edu/oim/project