Production Support

Bo Jayatilaka
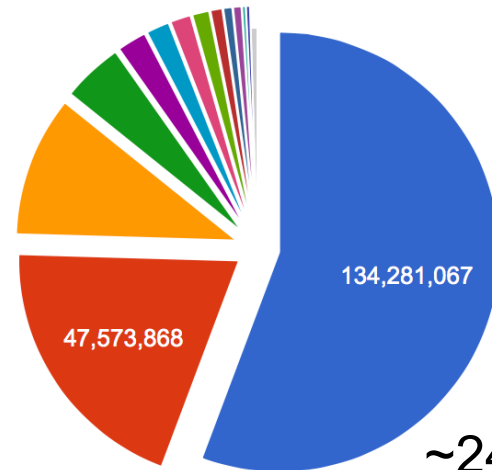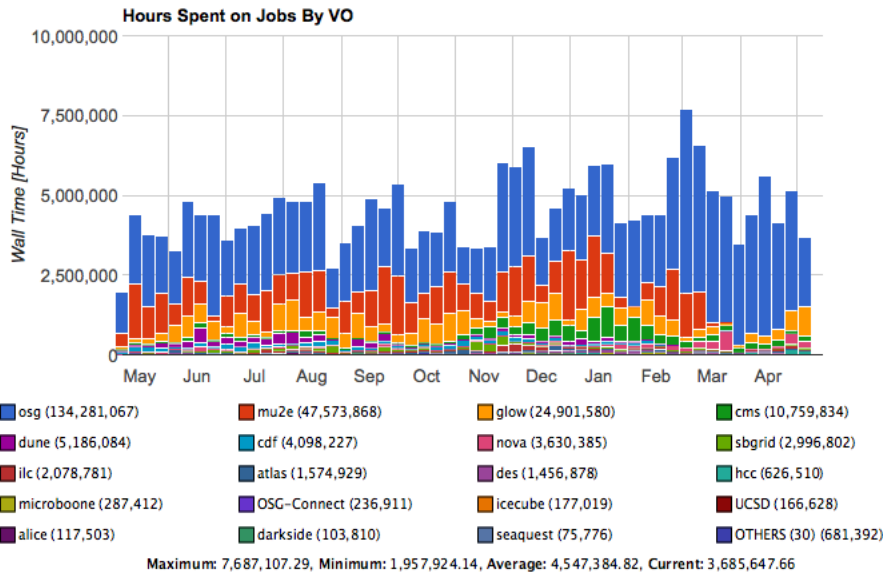*FNAL*

OSG Planning Retreat
San Diego, CA
May 11, 2016

Opportunistic cycles

- **Goal**: delivering 150M opportunistic hours 6/15-6/16

  - Already well past 200M since 6/15

  - Growth of ATLAS+CMS demand has not been substantial overall

    - LHC resources are also growing

  - Large growth in university clusters not associated with ATLAS or CMS is the biggest offset of any real LHC drop in opportunistic

    - Can we continue to find these resources?

Hours Spent on Jobs By VO

osg (134,281,067)	mu2e (47,573,868)	glow (24,901,580)	cms (10,759,834)
dune (5,186,084)	cdf (4,098,227)	nova (3,630,385)	sbgrid (2,996,802)
ilc (2,078,781)	atlas (1,574,929)	des (1,456,878)	hcc (626,510)
microboone (287,412)	OSG-Connect (236,911)	icecube (177,019)	UCSD (166,628)
alice (117,503)	darkside (103,810)	seaquest (75,776)	OTHERS (30) (681,392)

Maximum: 7,687,107.29, Minimum: 1,957,924.14, Average: 4,547,384.82, Current: 3,685,647.66

134,281,067

47,573,868

~240M hours
~187M jobs

Non-LHC growth

Hours Spent on Jobs By VO
53 Weeks from Week 04 of 2015 to Week 04 of 2016
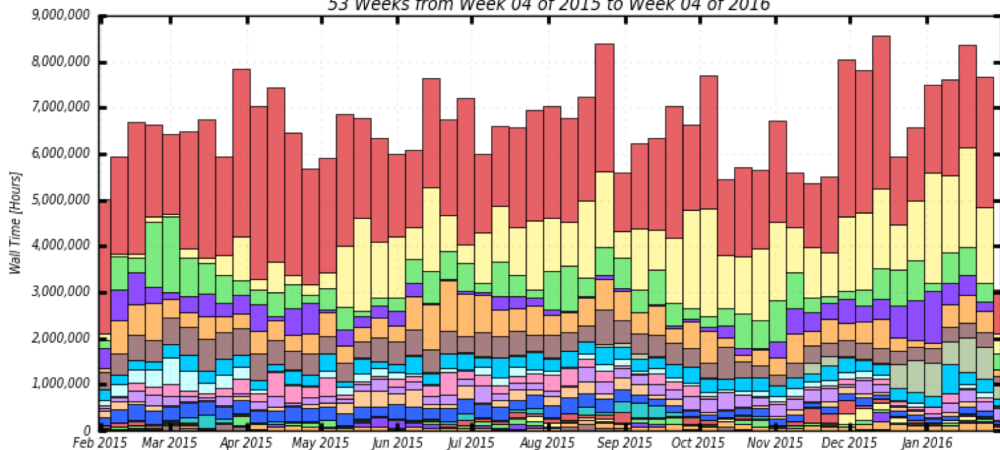
All VOs except ATLAS and CMS
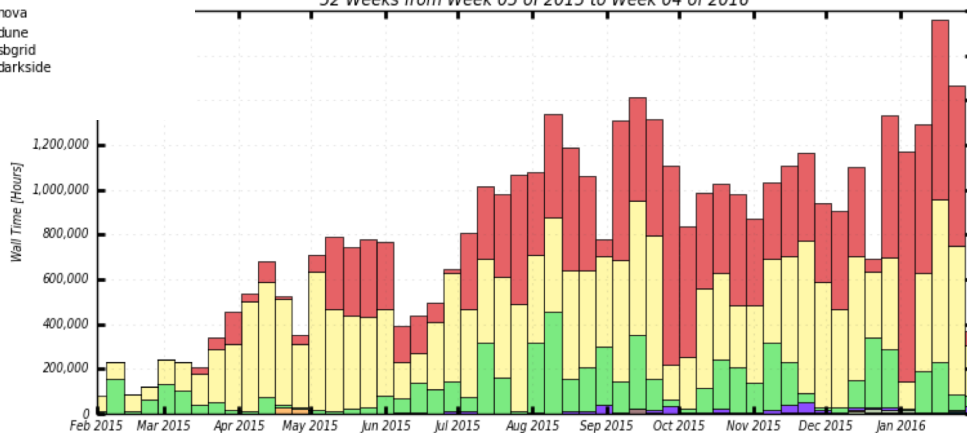
Clemson, Washington, Syracuse

Hours Spent on Jobs By VO
52 Weeks from Week 05 of 2015 to Week 04 of 2016

Scale testing

- **Goal**: demonstrate 50k simultaneous jobs on OSG VO flock
  - Accomplish with aid of sleep pools as CMS did
- Got to this on 12/3/15
  - 51533 peak (~40k real jobs)
- Few peaks of 35k+ jobs since then
 - Would like to push up the demonstrated maximum higher this year
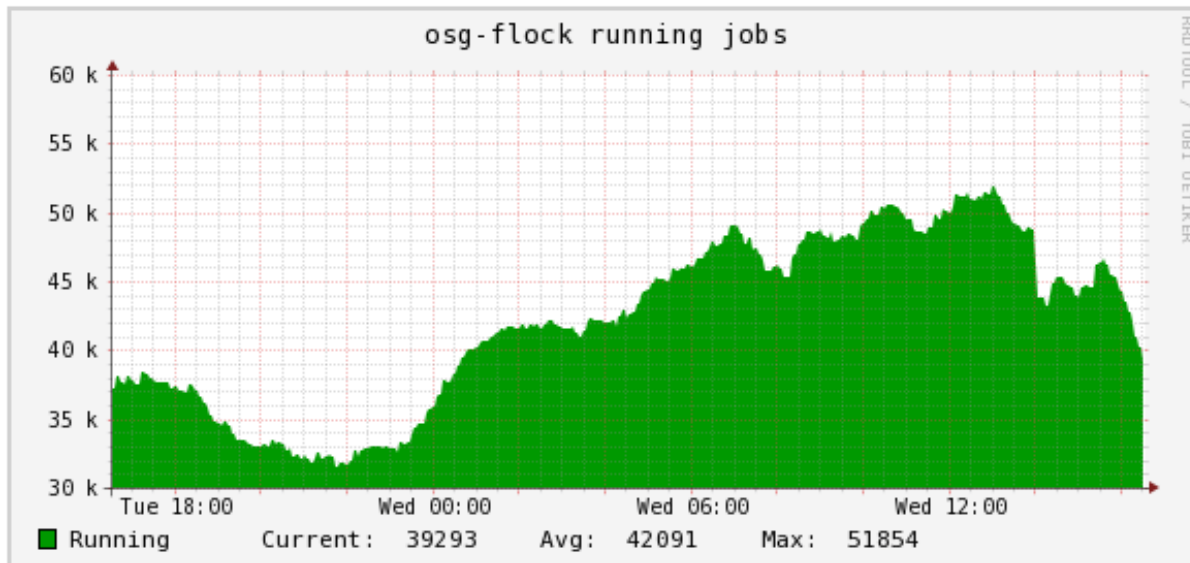
Intensity frontier engagement

- **Goal**: "Expanded use of OSG by the Intensity Frontier Experiments"

- Obvious success here is mu2e: massive campaign leading up to CD-2/3

  - Over 60M hours, **50M opportunistic**

  - **4th largest** VO by compute hours in 2015

- Expanded off-site computing by NOvA, μBooNE, and others as well

  - Upwards of 40% of all FIFE jobs running outside of FNAL

WMS Hours Spent on Jobs By Facility (Glidein)
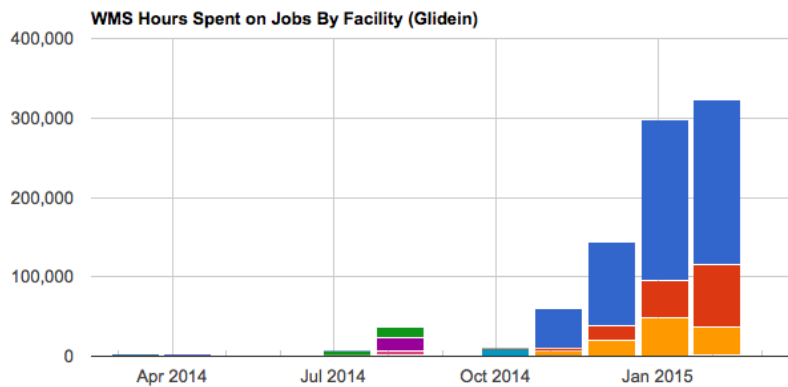
FZU_NOVA (565,824)	HU_ATLAS_Tier2 (147,080)	OSC_OSG (109,126)	Tusker (20,134)
Crane (19,406)	Amazon_AWS (12,425)	WT2 (6,272)	Nebraska (2,131)
BNL-ATLAS (597)	MWT2 (262)	UCSDT2 (257)	SMU_HPC (235)
Sandhills (96)	GLOW (88)	SU-OG (77)	NYSGRID_CORNELL_NYS1 (18)
TTU-ANTAEUS (3)			

Maximum: 322,693.59, Minimum: 161.28, Average: 73,669.10, Current: 322,693.59

WMS Hours Spent on Jobs By Facility (Glidein)
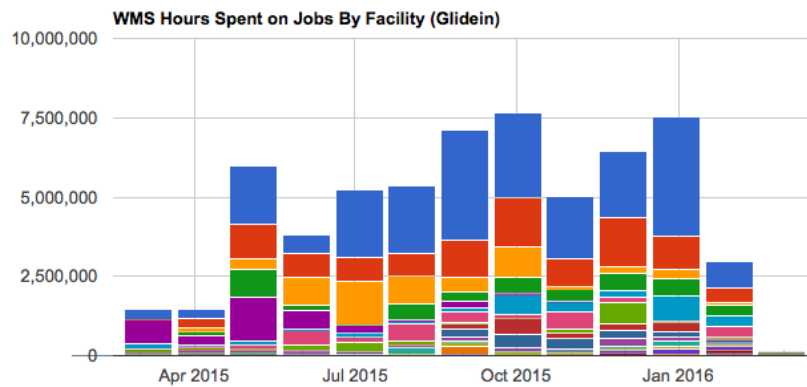
SU-OG (22,111,464)	CIT_CMS_T2 (10,281,859)	Nebraska (5,712,286)	USCMS-FNAL-WC1 (4,248,06
MIT_CMS (3,631,656)	GLOW (2,996,360)	FZU_NOVA (2,920,109)	MWT2 (1,586,303)
Crane (1,546,264)	NWICG_NDCMS (1,531,992)	AGLT2 (1,078,423)	Tusker (785,780)
UCSDT2 (474,746)	BNL-ATLAS (411,675)	Hyak (271,429)	Sandhills (180,231)
OSC_OSG (147,029)	SMU_HPC (94,637)	SMU_ManeFrame_CE (85,519)	OTHERS (9) (49,590)

Maximum: 7,663,842.88, Minimum: 160.98, Average: 4,296,100.87, Current: 114,395.69

Intensity frontier challenges

- Connecting sites with (some) dedicated resources
  - European sites for both NOvA and MicroBooNE registered in OIM and receiving jobs from FIFE
    - Trying stashcache with one of them
  - In many cases sites that are already part of EGI
- Leading a horse to water is not the same as getting it to drink
  - Sociology problem of "but it's harder" is still there
- Memory
  - More and more 3GB+ jobs
  - LAr experiments in particular
  - Using multicore helps

FIFE jobs since 1/2016

>2GB/core
0-2GB/core
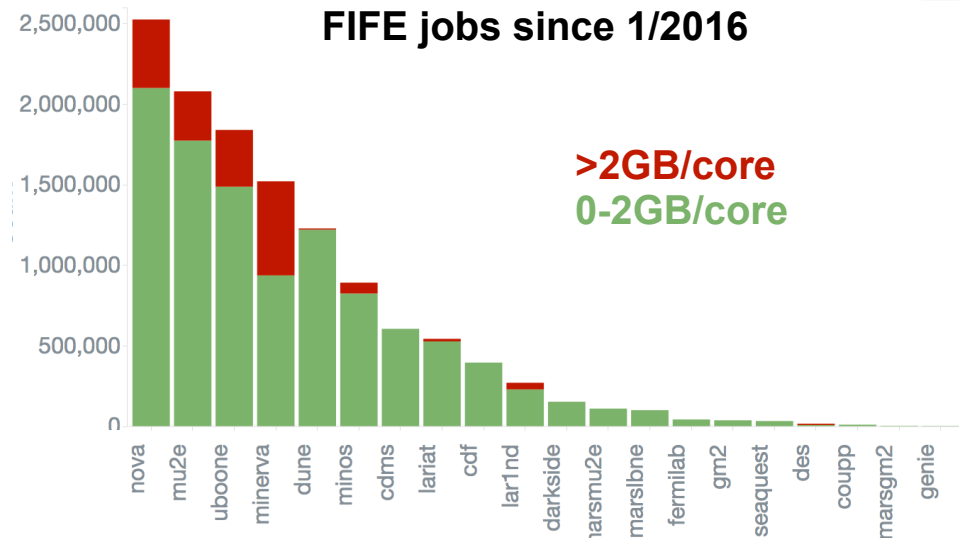
Integrating new communities

- Several large experiments/communities came on board in y4 with relatively rapid startups
  - mu2e, discussed previously
  - LIGO, discussed by Brian yesterday
  - sPHENIX
    - Effectively one user, very rapid startup
  - AMS
- Several large experiments/communities with previous involvement expanded/expanding OSG reach
  - IceCube
  - SBGrid
  - FermiLAT

Engaging institutions

- **Non-LHC university clusters** were a significant growth area
  - Syracuse (partly LIGO), Clemson, Washington were the biggest
- Approach for new institutions starts with a phone call with RWG and BAJ. Suggested steps are usually
  1. Identify potential OSG users and lead them to running jobs (e.g., via OSG-Connect)
  2. Identify potential shareable resources and suggest connecting via Bosco interface (e.g., Connect client)
  3. If 2. is insufficient solution, discuss setting up dedicated OSG CE
- Steps 2-3 are only if there are resources willing to be shared
  - Always make clear that this is **not** required to use OSG
- My personal sense: we should avoid 3) wherever at all possible with university clusters/HPC

Setting up a campus computing site

- Connecting LHC involvement in OSG with campus research IT at MIT (probably should have happened much earlier)

- Used the AMS request as an impetus to connect MIT resources to OSG

- Started with a schedd that submits to the OSG VO flock and direct condor submission to existing CMS Tier-2

- Then connected with an HPC cluster at Earth/Planetary Sciences by way of **Bosco**

  – Moving this to **Bosco-CE** in the coming weeks to finally allow outside jobs to come in

- This infrastructure gave AMS **10 million hours in 2 months**

- Hardest steps were sociology and politics

  – Technology steps were quick and for the most part "just worked"

- p.s., have an API that can move upwards of 15TB/day to dropbox (onto 18PB MIT dropbox account)

HPC

- HPC here refers to allocation-based supercomputers

- LIGO spurred the process here substantially

  - Successfully used allocation at Stampede

- Edgar *et al.* have also interfaced UCSD-T2 CEs with SDSC Comet (one rack) to run opportunistic OSG jobs

- Going forward

  - Streamline support structure (GOC<->XSEDE ticketing)

  - Enable running jobs on Comet on production scales

- Testing CMS on both of these sites as well via startup allocation

- Open question: do we want to enable running actual MPI jobs?

  - e.g., SBGrid would like to run Relion (a cryo-EM application)

Accounting

- Overhauling accounting was another y4 goal

- Have started GRÅCC project to accomplish this

  - To recap: Elasticsearch+RabbitMQ, maintain existing site probes

  - Project management coming out of production support

  - Base goal: full functionality in September. Stretch goal: comfortable retiring Gratia by the end of 2016

- Google charts implementations are there and have been for a while

  - Efficiency tree map has been a hit

  - Others are on ITB

  - **Should we make it a default view?**
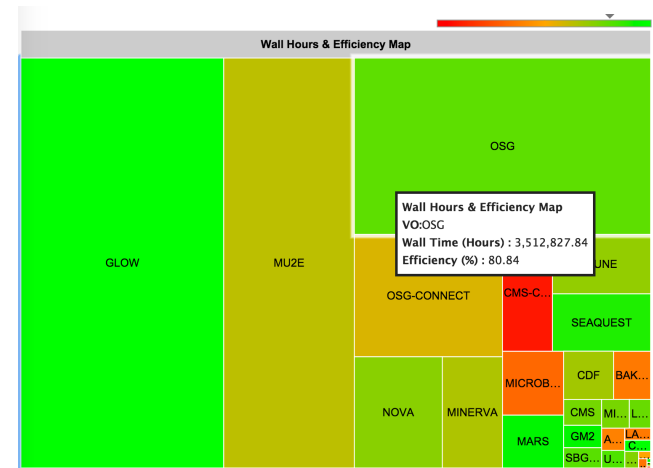
Now for some deep questions

- **What is a VO?**
  - We are moving towards increased use of community VOs even by large collaborations/groups
    - e.g., LIGO and AMS->OSG, IceCube->GLOW, All FIFE experiments->Fermilab
  - This matters for our trust model (banning a VO affects larger-scale operations)
  - This matters for accounting (should project accounting be the default?)
- **What is an OSG job?**
  - CMS connect jobs running on EGI CMS sites? IceCube jobs in Europe?
  - We started this discussion yesterday. My answer is "it depends."

Staffing

Name	Institution	FTE
Alex Zaytsev	BNL	0.10
Ken Herner	FNAL	0.10*
Tanya Levshina	FNAL	0.25
Bo Jayatilaka	FNAL	0.50
Juan Morales	FNAL	1.00**

Formerly
Robert Illingworth (FNAL)
Chander Sehgal (FNAL)

*10% directly on OSG for DES, much more OSG-related via FIFE
**Gratiaweb/Gratia development Departing end of June