Fermilab U.S. DEPARTMENT OF ENERGY Office of Science

# Access to Offsite Resources

Bo Jayatilaka, Tanya Levshina

FIFE Workshop 2016

June 20, 2016

FABRIC FOR FRONTIER EXPERIMENTS

# What is OSG?

Frank Wuerthwein:

*"We create a uniform environment across a heterogeneous set of resources that is distributed globally. Submit locally – Run Globally" - at OSG AHM 2016*

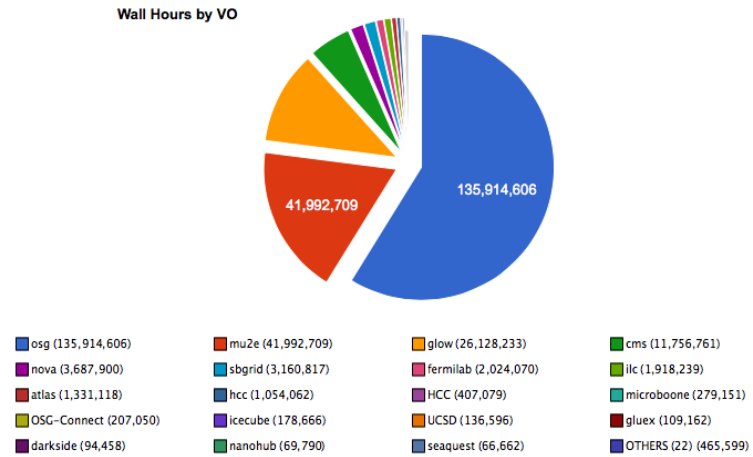
- The vision of the Open Science Grid Consortium (OSG) is one of a persistent national infrastructure for US science.
- The OSG includes:
  - scientific collaborations
  - scientific computing centers
  - distributed high throughput computing research and deployment projects

# Why OSG?

- New technologies
- Testing software on new OS
- OSG Software Stack
- Centrally running services
- OSG Security Team
  - CILogon OSG CA
  - OSG CA Distribution
  - Operational security, interoperability with other grids
- GOC provides a single point of operational support for the OSG

# OSG:What's out there?

- Many OSG sites allow opportunistic use of spare cycles:
  - 231M hours across all VOs and all sites in the past year
  - 42M by mu2e alone
  - Most permit the Fermilab VO (any offsite jobs submitted via jobsub)
- Individual experiments with collaborating institutions that have grid sites:
  - Mostly ATLAS/CMS sites either on OSG or EGI
  - In most cases accessible via the same OSG interface
  - Considered dedicated resources for that experiment
  - Most cases not by the entire Fermilab VO
- Allocation-based HPC (supercomputers)
  - Also can be accessible via OSG
  - Currently limited to two (NSF) sites

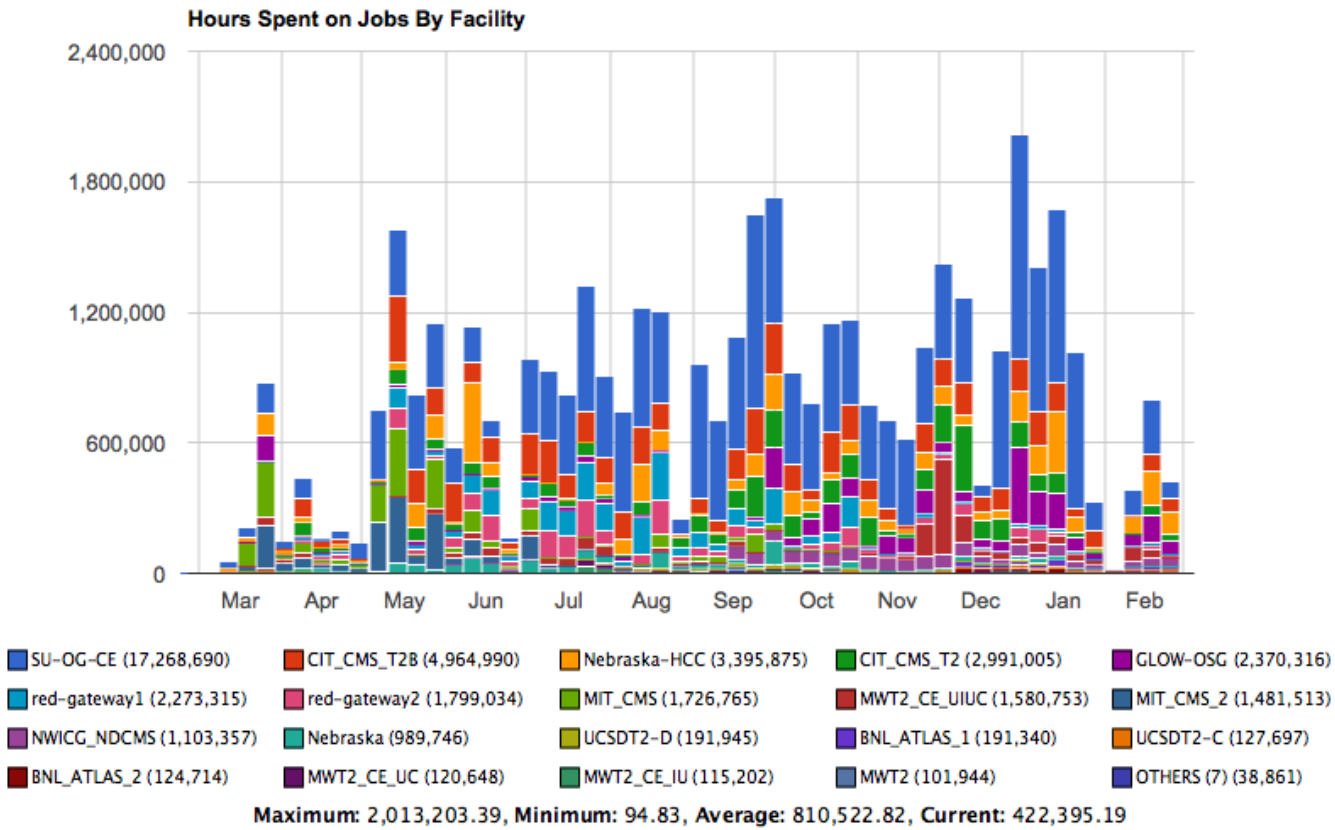



**June 2015-June 2016**
**Opportunistic use by all VOs**
**230,982,728 hours**

# OSG: General opportunistic

- Currently **18 OSG sites** support the generic Fermilab VO
  - ATLAS (4)
  - CMS (9)
  - university HPC clusters (5)
- What to expect on these sites
  - All will have CVMFS
  - Multiple sites are now offering access to CVMFS/StashCache (useful for flux files access)
  - Most sites offer ~2GB RAM/core and ~35GB disk space but expect some variation
  - Most sites offer **multicore** slots: a way to get large memory jobs in. Most sites allow to request up to 8 cores.
  - Some of these sites are very large: Syracuse (SU-OG) will rival all of GPGrid at times just in opportunistic OSG running

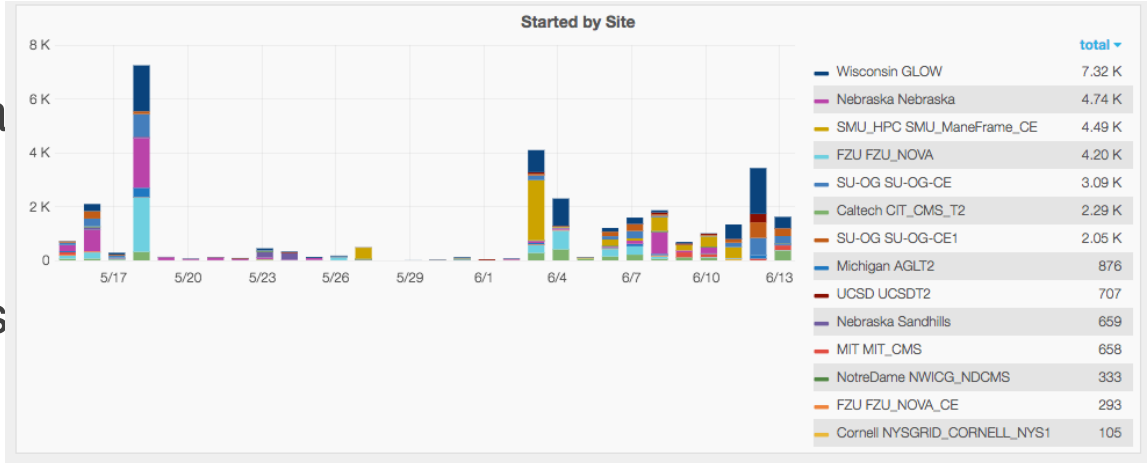
# OSG: General opportunistic (example)

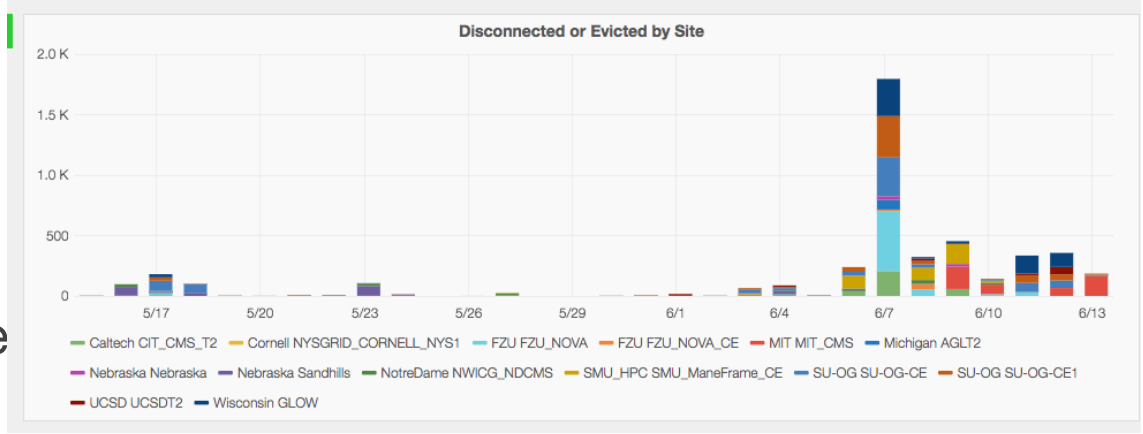



42M hours, **none** of these sites offered mu2e any dedicated resources

# OSG: Preemption

- Most sites will **preempt**
  - Owner VOs can claim a slot back at any time
  - Your job will restart elsewhere if this occurs
- Most sites do not offer any specific quota for Fermilab VO jobs
  - All opportunistic VOs are treated on the same footing

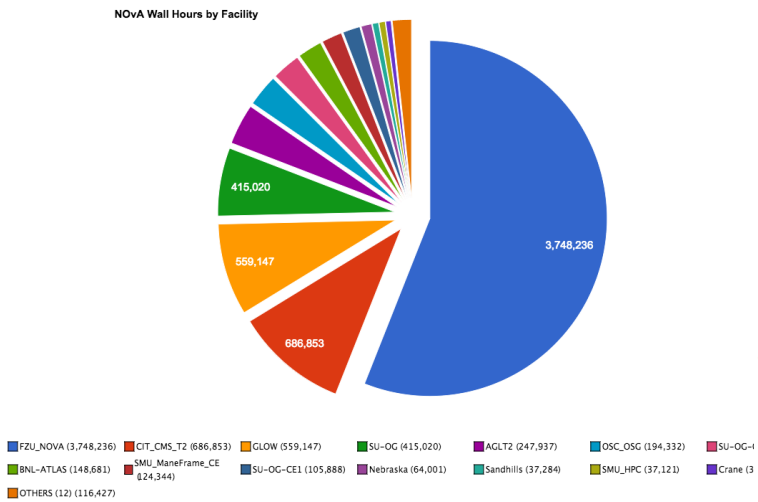


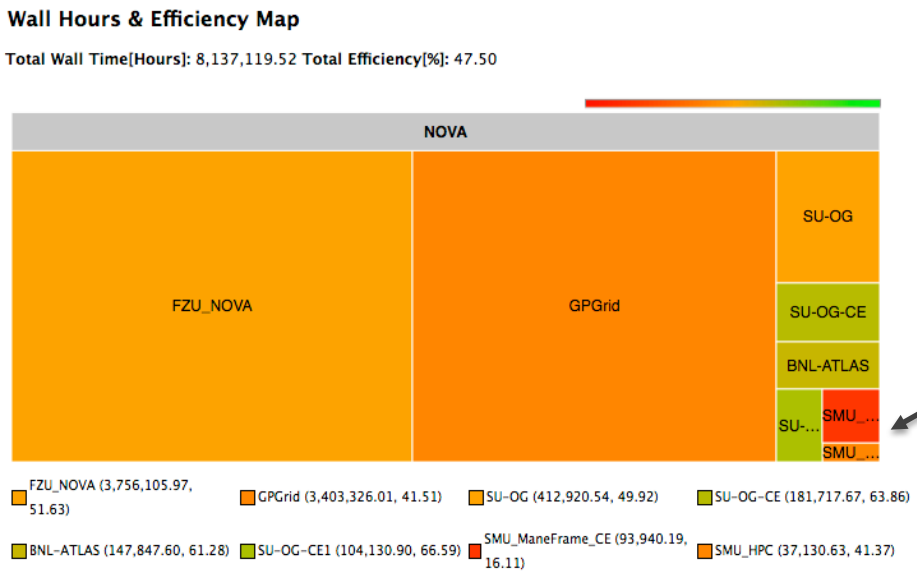



# Dedicated Sites

- Sites that treat **one particular experiment** apart from the generic Fermilab VO. Could mean that experiment:
  - gets **dedicated resources** at the site
  - gets **priority** over other outside Vos
  - **does not get preempted**
- Examples
  - NOvA: SMU, FZU
  - MicroBooNE: Bern, Manchester, Lancaster (in progress)
  - CDF: CNAF
- Several ways to do it:
  - Site becomes an OSG site (installs OSG CE and other services from OSG Software Stack)
  - "Non-OSG" (European) sites could be used if enabled in the OSG Factory
    - e.g., Manchester an existing EGI/ATLAS site->Enable in the OSG factory->uboone job submissions to the site then work

# Dedicated Sites: Is it worth the trouble?




16%

Dedicated Sites:

- FZU
- SMU
- Multiple GOC tickets
  - Transfer rate issues
  - GUMS configuration
  - Job is not starting

# Dedicated Site: Acceptance Criteria

- Before accepting “dedicated” site:
  - How big is a cluster?
  - What OS is installed?
  - How efficient are transfers from remote site to/from dCache?
  - How much experience does sys admin have?
  - Do you have a local PI that is really interested to deliver cycles for your experiment.
  - Does site allow outbound connection from worker node?
- How can experiments help here?
  - If you have a collaborating institution with grid resources willing to become a dedicated site **contact us**:
    - Email: fife-support@fnal.gov
    - SNOW ticket: FIFE-SUPPORT

# OSG/XSEDE: Allocation-based

- These are "national cyberinfrastructure" supercomputers
- HPC resources with interconnects, O(10k-100k) CPU cores
- Access via **allocations**
  - CPU time allotment to a PI based on proposal process
  - Some will also have local/institutional quotas as well
- NSF: XSEDE (Extreme Science and Engineering Discovery Environment)
  - Allocation process is described at https://portal.xsede.org/allocations-overview
    - PI must be a researcher or educator at a U.S.-based institution, including federal research labs
    - PIs with support from any funding source, not just NSF, are encouraged to apply
  - Allocation requests are reviewed quarterly and awarded for 1 year increments
  - Review process is central for **all resources**
  - Proposals are considered only on **scientific merit**
    - End-goal must be peer-reviewed scientific result
    - Resource availability is **not considered** in review process
  - Startup allocations granted for development on a lighter review basis (abstract only)
    - You will need to get a startup allocation first before you will be be able to request a "real" one

# Allocation Based Resources

- OSG currently enables access to two XSEDE resources: Comet (San Diego), Stampede (Austin)
- Comet:
  - 47k cores Intel(Haswell)
  - 247TB memory
  - 634TB compute node SSD
  - Opportunistic (backfill) access via UCSDT2 site on OSG
  - Any CVMFS repository available on UCSD is available on Comet
  - If your jobs already run on opportunistic OSG resources they already land here
  - Allocation-based access coming soon
- Stampede:
  - 100k cores Intel (Sandy Bridge)
  - 270TB memory
  - Allocation-based access via OSG already used by LIGO (about 2M hours)
  - CVMFS installation requires manual addition of new repositories
  - MINOS soon testing institutional allocation via UT Austin
- FIFE frontend version which supports allocation ID currently in testing

# Summary

- Many opportunistic cycles are available on the OSG.
- 18 sites are accepting jobs from Fermilab.
- Experiment may have get access to additional HPC resources via XSEDE allocation.
- FIFE Support will be glad to help to include experiments dedicated sites into OSG resource pool.
  - Site should satisfy min requirements.
- You could improve job success rate by tailoring your job memory, cpu, duration and transfer requests.
- FIFE Support is ready to help to run on the OSG not only production teams but individual users running analysis jobs.