Extending the Reach and Scope of Hosted CEs

OSG All Hands Meeting
March 20, 2018

Suchandra Thapa
Robert Gardner
University of Chicago

Derek Weitzel
University of Nebraska

Open Science Grid

1

Introduction

- Hosted Compute Elements (CEs) were introduced about a year and a half ago to give sites an easier way to contribute cycles to OSG
- Sites also get a deeper view on their contributions to OSG
- Since then Hosted CEs have extended the sites and resources that can be integrated into OSG
  - **Greater geographical reach**
  - **Sites that differ from the "typical" OSG site**
  - **HPC resources on XSEDE**

The Hosted CE Approach

- Using HTCondor Bosco CE (i.e. 'Hosted CE'), the CE administration can be cleanly separated from the cluster administration
- Cluster admins only need to provide ssh access to cluster
- OSG staff can maintain the hosted CE and associated software and deal with OSG user/site support

Providing local view of science

- After OSG jobs started running, admins can track their cluster's contributions to OSG users using GRACC
- Multiple views on contributions to OSG users:
  - By field of science
  - By project or VO
  - By researcher's institution

CE hosting infrastructure

- Minimal requirements
- Hosted CE can be run on a fairly small VM (1 core / 1GB)
  - Memory usage for typical hosted CE is less than 512MB
  - Hosted CE VMs cpu have been more than 80% idle
 - Max network traffic is fairly low (<200Kb/s)

Greater Geographical Reach

- Low cost of entry has allowed sites to contribute despite time zone and logistical difficulties
- Example: IUCAA Sarathi (LIGO India)
  - Located in Pune, India
  - 12:30 hour difference (1 day lag in email responses)
  - Didn't want to require admins to need to learn about the internal details about OSG glidein infrastructure
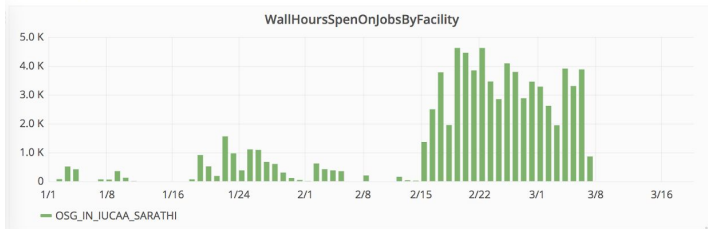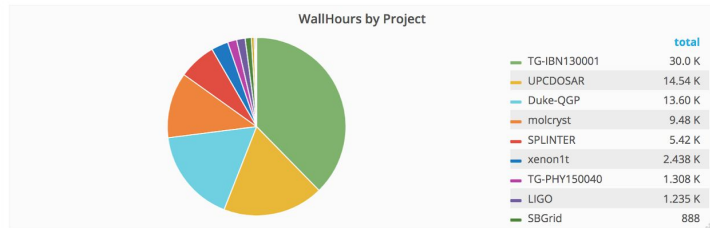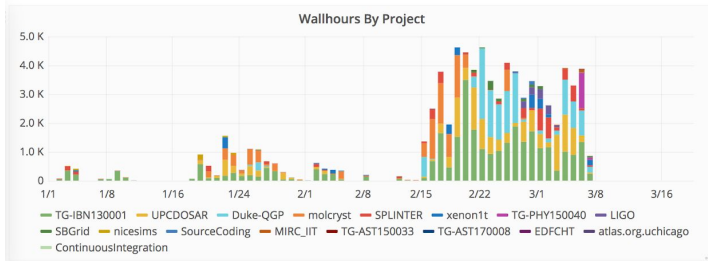
IUCAA Sarathi Cluster (LIGO - India)
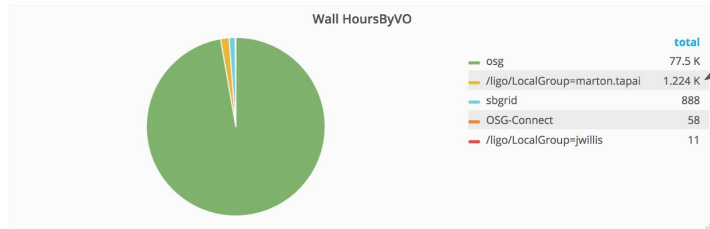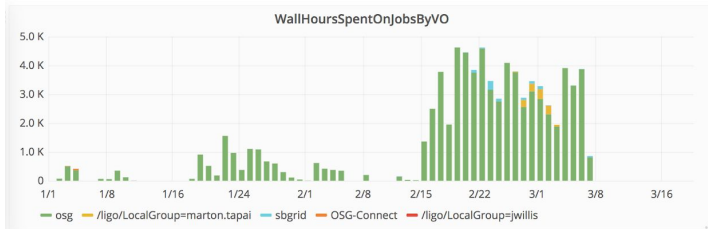
LIGO users running under a LIGO specific account through OSG

~80k wall hours provided from India this year!

Expanding the variety of sites

- Bulk of sites contributing to OSG tend to be national labs or large research institutions
  - A lot are brought in by ATLAS or CMS
- Due to the lower cost of entry when using hosted CEs, other types of sites can now contribute:
  - University of Utah
  - North Dakota State University
  - Georgia State University
  - Wayne State

Example: University of Utah
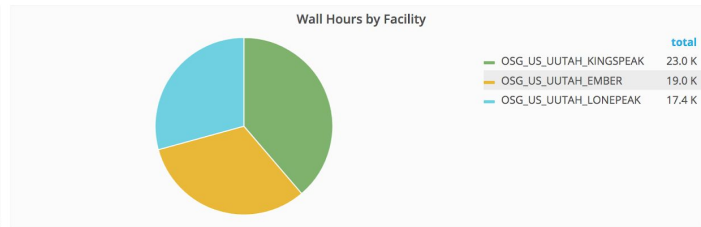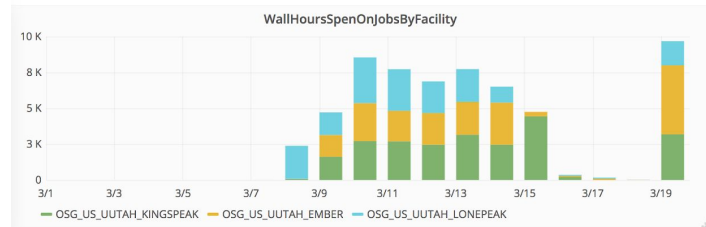
- Several clusters on campus

+

- Time needed to become familiar with OSG CE operations / Glidein troubleshooting

=

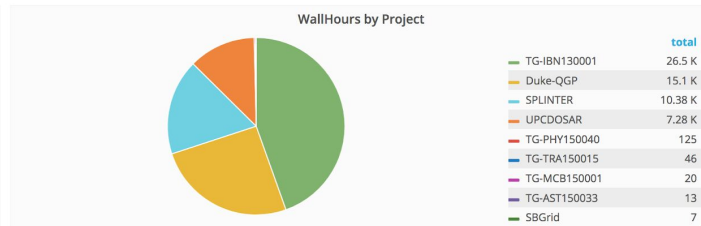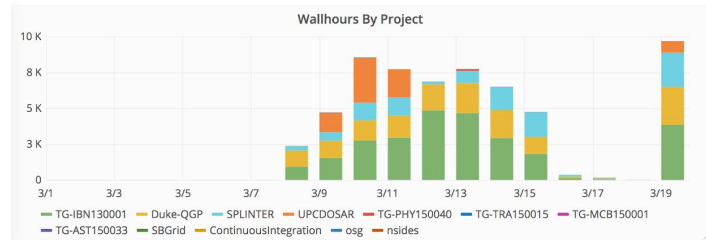- Significant barrier to entry in order to contribute to OSG
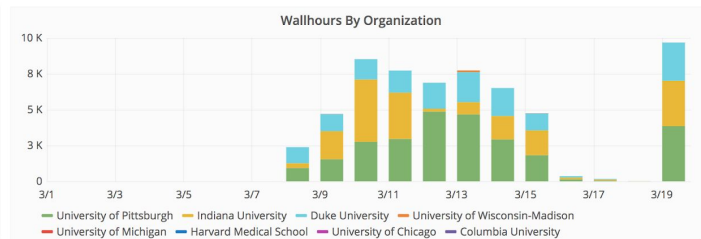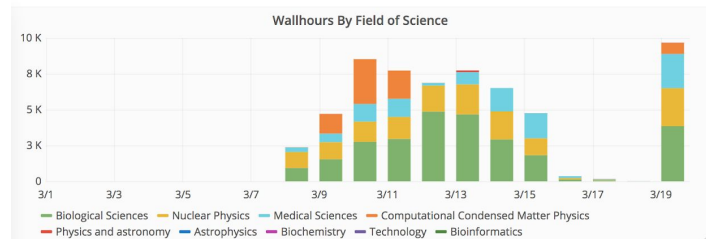
Utah contributions

By Facility

WallHoursSpenOnJobsByFacility

Wall Hours by Facility

	total
OSG_US_UUTAH_KINGSPEAK	23.0 K
OSG_US_UUTAH_EMBER	19.0 K
OSG_US_UUTAH_LONEPEAK	17.4 K

By Project

Wallhours By Project

WallHours by Project

	total
TG-IBN130001	26.5 K
Duke-QGP	15.1 K
SPLINTER	10.38 K
UPCDOSAR	7.28 K
TG-PHY150040	125
TG-TRA150015	46
TG-MCB150001	20
TG-AST150033	13
SBGrid	7

By Field Of Science

Wallhours By Field of Science

Wallhours By Organization

All three clusters brought into production over the last 2 weeks

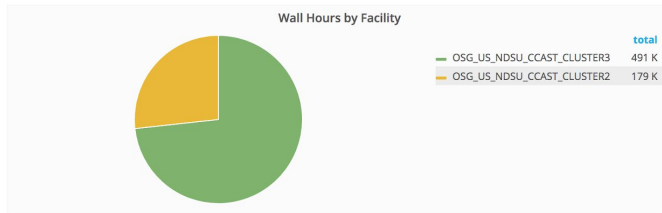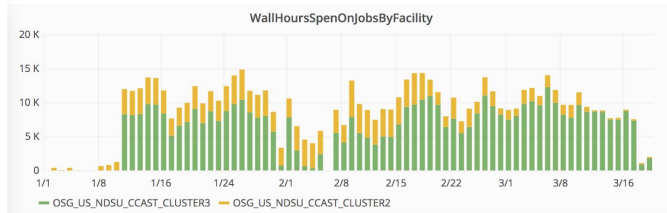Still tweaking jobs, looking at using multicore jobs to more effectively backfill and get more cores

Already contributed ~60k cpu hours, in top 2 institutions contributing through Hosted CEs

10

North Dakota State University
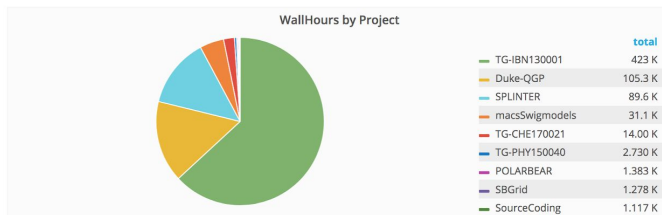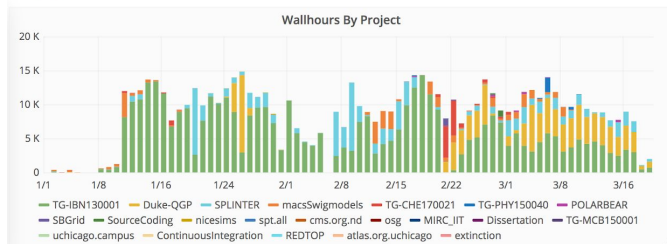
Two clusters, CCAST3 was brought online beginning of year

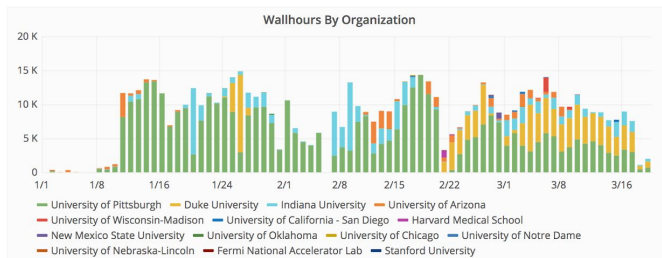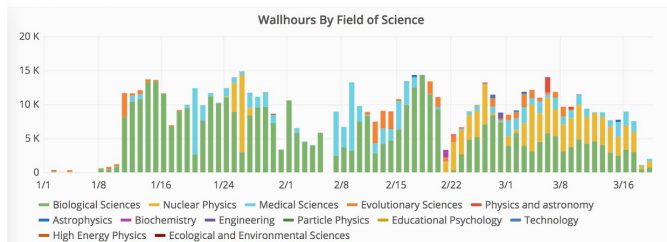Single core jobs on CCAST2, 8 core jobs on CCAST3

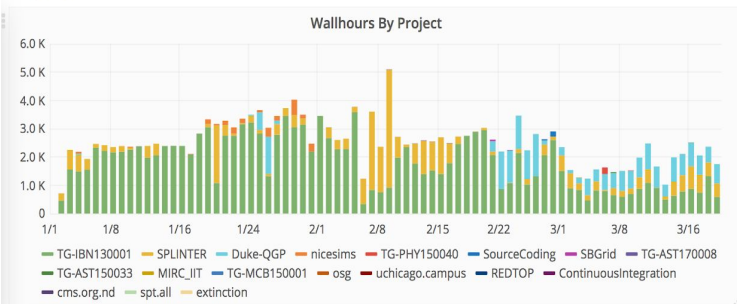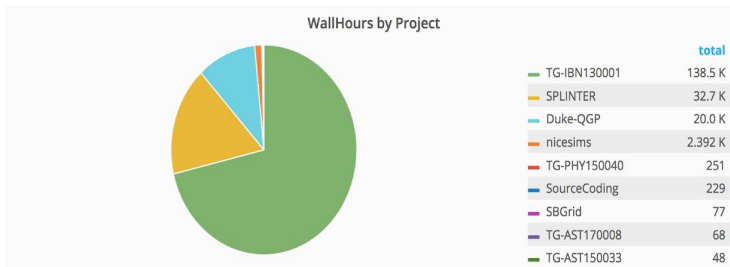670K wall hours delivered, one of top hosted CE sites

Georgia State University
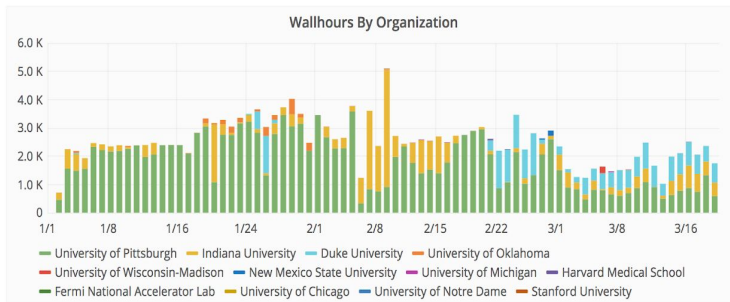

194K wall hours delivered since Jan 1

18 Projects helped

Provided cpu to 11 fields of science

12 Institutions ran jobs on resource

Wayne State University
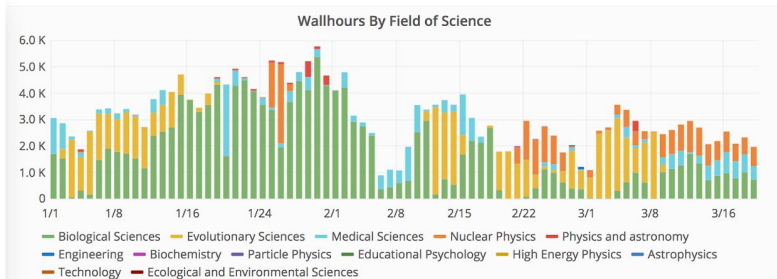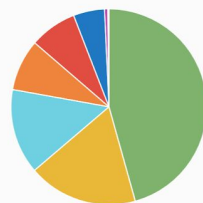
By Project

Wallhours By Project

6.0 K / 5.0 K / 4.0 K / 3.0 K / 2.0 K / 1.0 K / 0

1/1 1/8 1/16 1/24 2/1 2/8 2/15 2/22 3/1 3/8 3/16

TG-IBN130001 — macsSwigmodels — GLOW — SPLINTER — Duke-QGP — TG-CHE170021 — nicesims
TG-PHY150040 — SourceCoding — SBGrid — TG-MCB150001 — TG-TRA150015 — cms.org.nd — uchicago.campus
MIRC_IIT — Dissertation — osg — REDTOP — IceCube — spt.all — ContinuousIntegration — atlas.org.uchicago
extinction
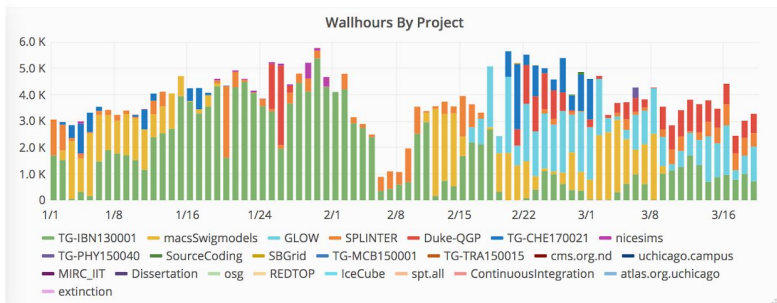
WallHours by Project

	total
TG-IBN130001	137.2 K
macsSwigmodels	54.4 K
GLOW	42.3 K
SPLINTER	25.5 K
Duke-QGP	23.4 K
TG-CHE170021	15.4 K
nicesims	1.604 K
TG-PHY150040	395
SourceCoding	133

By Field Of Science

Wallhours By Field of Science

6.0 K / 5.0 K / 4.0 K / 3.0 K / 2.0 K / 1.0 K / 0

1/1 1/8 1/16 1/24 2/1 2/8 2/15 2/22 3/1 3/8 3/16

Biological Sciences — Evolutionary Sciences — Medical Sciences — Nuclear Physics — Physics and astronomy
Engineering — Biochemistry — Particle Physics — Educational Psychology — High Energy Physics — Astrophysics
Technology — Ecological and Environmental Sciences

Wallhours By Organization

6.0 K / 5.0 K / 4.0 K / 3.0 K / 2.0 K / 1.0 K / 0

1/1 1/8 1/16 1/24 2/1 2/8 2/15 2/22 3/1 3/8 3/16

University of Pittsburgh — University of Arizona — Indiana University — Duke University — University of Oklahoma
University of Wisconsin-Madison — New Mexico State University — Harvard Medical School — University of Notre Dame
University of Nebraska-Lincoln — Fermi National Accelerator Lab — University of Wisconsin — University of Chicago
Stanford University

300k cpu hours delivered since Jan 1
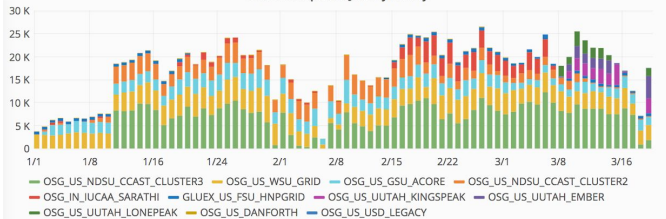
Ran jobs from 24 projects, 13 fields of science, and 14 institutions

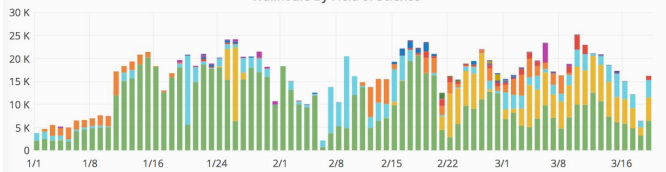13
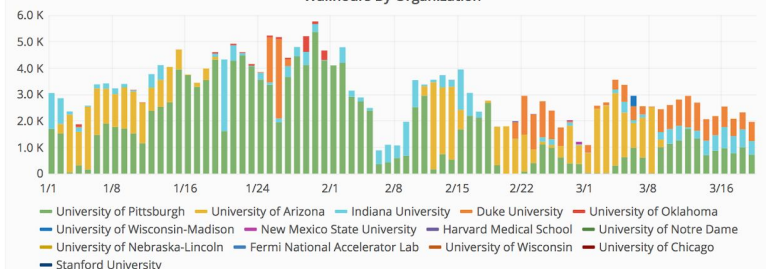
Total Contributions

WallHoursSpenOnJobsByFacility

Legend
OSG_US_NDSU_CCAST_CLUSTER3 — OSG_US_WSU_GRID — OSG_US_GSU_ACORE — OSG_US_NDSU_CCAST_CLUSTER2
OSG_IN_IUCAA_SARATHI — GLUEX_US_FSU_HNPGRID — OSG_US_UUTAH_KINGSPEAK — OSG_US_UUTAH_EMBER
OSG_US_UUTAH_LONEPEAK — OSG_US_DANFORTH — OSG_US_USD_LEGACY
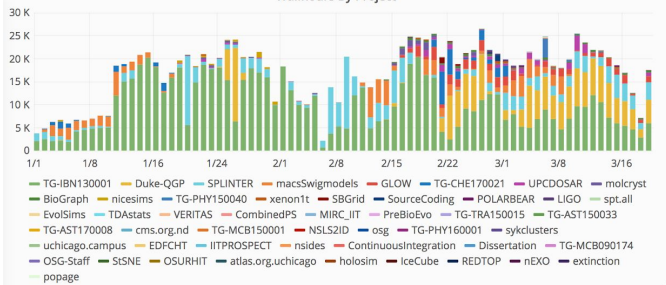
Wall Hours by Facility
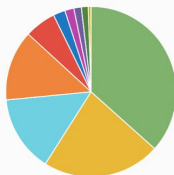
Facility	total
OSG_US_NDSU_CCAST_CLUSTER3	491 K
OSG_US_WSU_GRID	300 K
OSG_US_GSU_ACORE	194 K
OSG_US_NDSU_CCAST_CLUSTER2	179 K
OSG_IN_IUCAA_SARATHI	79.7 K
GLUEX_US_FSU_HNPGRID	30.5 K
OSG_US_UUTAH_KINGSPEAK	23.1 K
OSG_US_UUTAH_EMBER	19.1 K
OSG_US_UUTAH_LONEPEAK	17.5 K

By Project

Wallhours By Project

WallHours by Project

Project	total
TG-IBN130001	767 K
Duke-QGP	180 K
SPLINTER	166 K
macsSwigmodels	89.0 K
GLOW	42.3 K
TG-CHE170021	32.6 K
UPCDOSAR	22.5 K
molcryst	10.22 K
BioGraph	9.69 K

By Field Of Science

Wallhours By Field of Science

Wallhours By Organization

>1.3M wall hours delivered since Jan 1

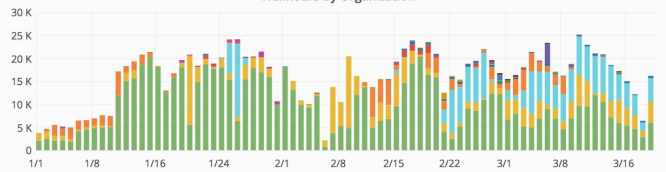Averaging about 111K wall hours a week. About 10-15% of weekly opportunistic usage by OSG Connect users

Ran jobs from 25 fields of science and 35 institutions

14

Integrating HPC resources into OSG

- Major cultural differences between HPC resources and OSG resources
  - MultiFactor Authentication (MFA) using tokens
  - Software access and distribution
  - Allocations

Bridging the Gap

- Solutions:
  - Authentication -> get MFA exceptions or use IP address as a factor
  - Software access -> Stratum-R
  - Job Routing -> Multi-user BOSCO
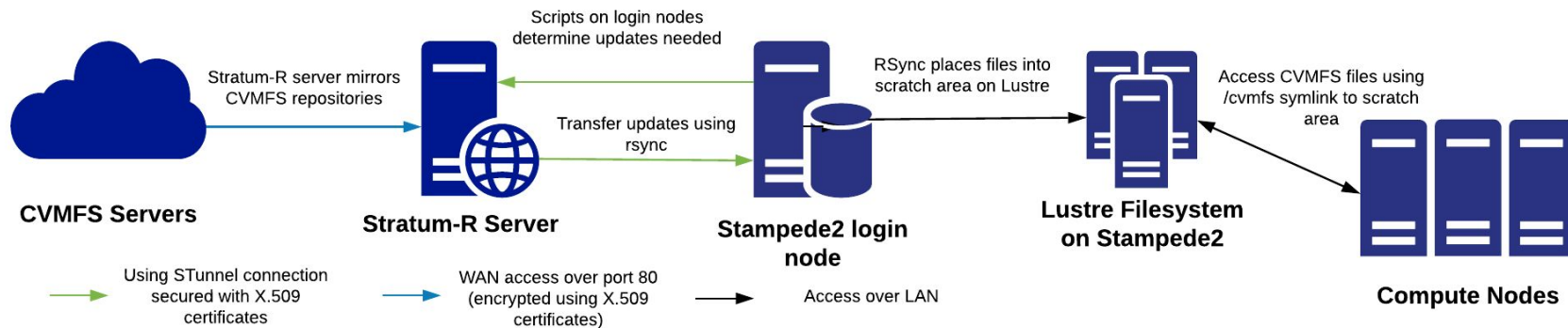
User Authentication

- HPC resources are increasingly moving to using MFA
  - OSG software doesn't have any way to incorporate token requirements into job authentication
- Solutions:
  - Use submit site's IP as one factor.
    - All job submissions come from a fixed IP.
    - Can use a ssh public key or proxy as another factor
  - Get a MFA exception for accounts
    - Sites often have procedures requesting this for science gateways or similar facilities to use.

Software distribution and access

- VOs and users are increasingly using CVMFS to distribute software and data
  - HPC resources usually aren't willing to install and maintain CVMFS on their compute nodes
- Stratum-R allows for replication of selected CVMFS repositories
  - Requires some effort from admins but not much
 - Successfully used on Bluewaters, Stampede, Stampede2

Stratum-R

CVMFS Servers

Stratum-R server mirrors CVMFS repositories

Scripts on login nodes determine updates needed

Stratum-R Server

Transfer updates using rsync

Stampede2 login node

RSync places files into scratch area on Lustre

Lustre Filesystem on Stampede2

Access CVMFS files using /cvmfs symlink to scratch area

Compute Nodes

Using STunnel connection secured with X.509 certificates

WAN access over port 80 (encrypted using X.509 certificates)

Access over LAN

Routing jobs to allocations

- Due to allocations, must route jobs to proper user accounts on HPC resources
- BOSCO's default configuration is to use a single user on remote resource for **all** job submissions
- With some modifications to config files, JobRouter entries, and other bits, can have jobs going to different users on remote resources
  - This allows for jobs to use different allocations, partitions, configurations
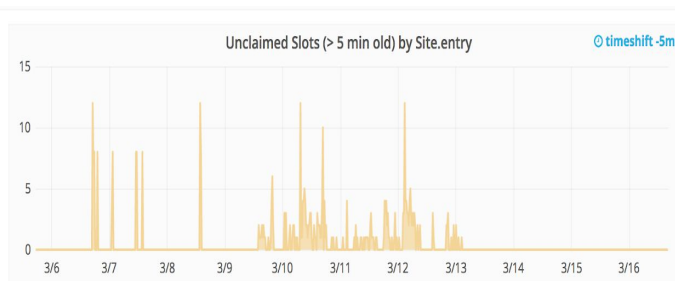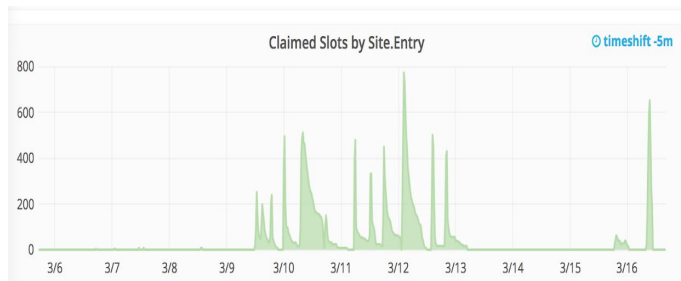
HTCondorCE BOSCO Job Routing
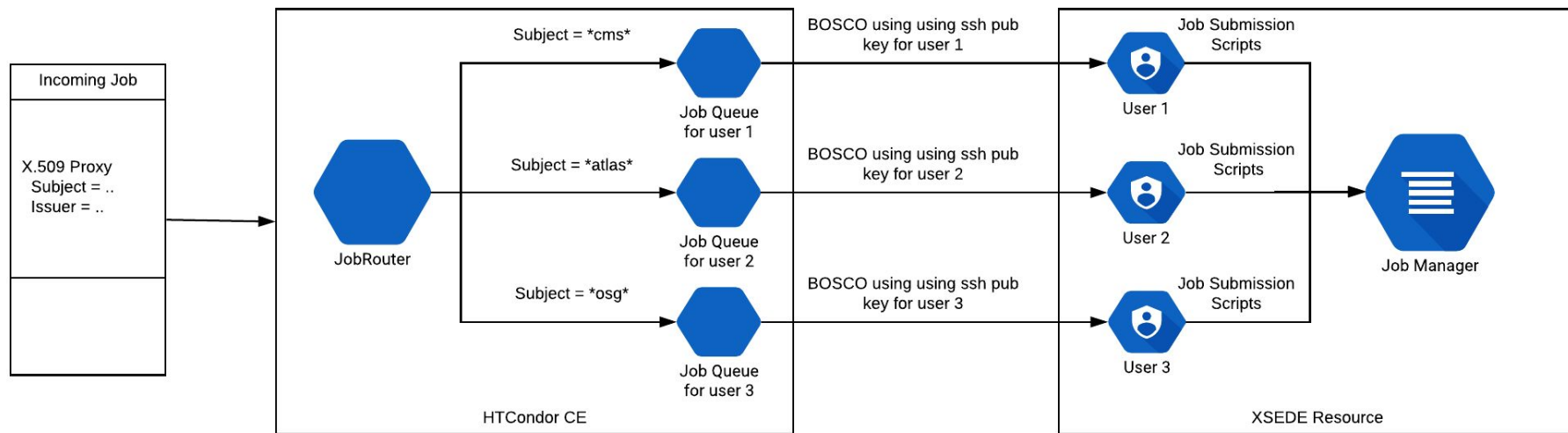
Running CMS jobs on XSEDE

Stampede2
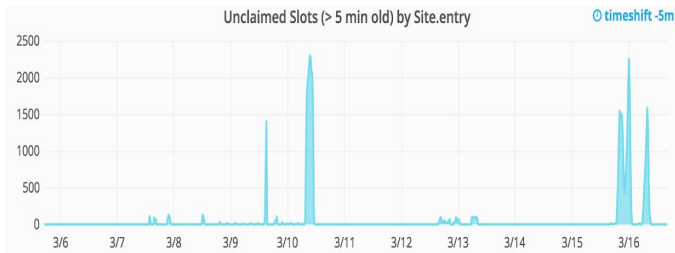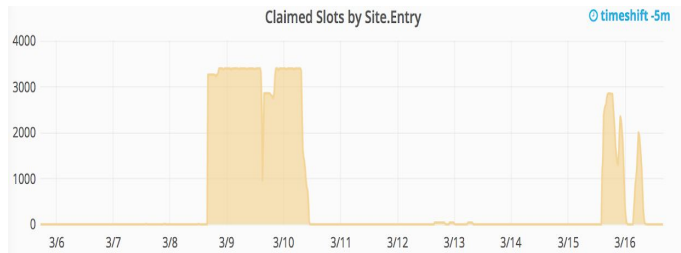
Bridges

Still validating and testing CMS workflows on Bridges and Stampede2

Conclusions

- Hosted CEs offer OSG the opportunity to obtain cycles and engage with new types of sites and resources increasing the diversity and reach of OSG.
  - Smaller universities and institutions
  - XSEDE resources (direct allocations for users)

More information

- [Support document for cluster admins](#)
- [BOSCO](#) [CE](#)

Acknowledgements

- Derek Weitzel
- Factory Ops (Jeff Dost, Marian Zvada)
- David Lesny
- Mats Rynge
- CMS HepCloud Team (Dirk, Ajit, Burt, Steve, Farrukh)
- Lincoln Bryant - Infrastructure support
- Rob Gardner