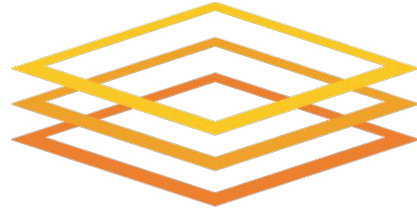

Site-Facing Services

For OSG Resource Providers



Jeff Dost (OSG Operations) HOW 2019



Introduction

- The goal of this talk is to introduce the site facing services offered by the Open Science Grid today
- This past year there has been some reorganization, so some services have a new home and / or name and interface
- See Brian Lin's Tuesday talk for details on the transition:
 - [OSG Software: The year in review](#)



Resources for new Sites

- Documentation:
<https://opensciencegrid.org/docs/>
- Support:
 - via email: help@opensciencegrid.org
 - via Freshdesk ticket:
<https://support.opensciencegrid.org/support/tickets/new>
- Yum repository:
 - <https://opensciencegrid.org/docs/common/yum>
 - This repo contains the software needed to install on the CE and the worker nodes



OSG Topology (formerly OIM)

- Topology is the catalog of all OSG sites
 - collection of YAML files stored in the OSG GitHub
- New sites should register to Topology for:
 - OSG accounting data
 - Site contact information
 - WLCG accounting (APEL CPU usage, downtime info)
- <https://opensciencegrid.org/docs/common/registration/>
- Topology is also where sites declare service downtimes



GRACC

- GRACC (GRid ACcounting Collector) is the database of all site usage data (Compute hrs, Data transfer)
- In order to collect data, sites must:
 - Register with topology
 - ensure services are correctly configured with osg-configure
<https://opensciencegrid.org/docs/other/configuration-with-osg-configure/#site-information>
- Resource group and resource name in osg-configure site info must match fields as registered in topology
- <https://gracc.opensciencegrid.org/>



CE Collector

- Sites should advertise attributes that describe their compute clusters, such as:
 - number of cores / node
 - memory / node
 - max walltimes
 - allowed VOs
- This information propagates to a HTCondor collector at `collector.opensciencegrid.org`
- Pilot submission systems use this information to properly configure pilots to run at sites
- <https://opensciencegrid.org/docs/other/configuration-with-osg-configure#submitter-resource-entry>



Hosted CE

- Intended for smaller sites that do not have the effort to install and maintain their own CE
- CE is run on OSG provided host, maintained by OSG Operations staff
- CE requires ssh login to Site cluster submit host
- <https://opensciencegrid.org/docs/compute-element/hosted-ce/>



OASIS

- cvmfs repository at [/cvmfs/oasis.opensciencegrid.org](https://cvmfs/oasis.opensciencegrid.org)
- Uses:
 - VOs can stage application software / data there
 - sites can use it to obtain WN software
- Sites should make cvmfs available the workers
- For standard* cvmfs installs, assumed site has
 - local squid
 - FUSE on the WNs

* Other methods of exporting CVMFS data on the WNs exist if the standard squid + fuse solution doesn't fit site needs



StashCache

- XRootD federation for VOs to make application data available across OSG
- VOs can provide data origins, e.g. OSG Connect origin is based at UChicago
- Sites wanting to support StashCache should provide cvmfs, user jobs can access data via `/cvmfs/stash.osgstorage.org`
- Optionally Sites can install XRootD caches, to reduce network overhead and decrease access latency

<https://opensciencegrid.org/docs/data/stashcache/install-cache/>



OSG Glidein Factory

- Standard pilot submission infrastructure that most OSG VOs use to submit to the grid
- OSG submission model is pilot based, which means user jobs aren't submitted directly to CEs
- Pilot jobs are instead submitted and claim site resources for a finite amount of time, and pull in jobs that match those resources
- OSG Factory Operations team works closely with sites to ensure pilots run correctly
- Support contact info: osg-gfactory-support@physics.ucsd.edu



Questions?