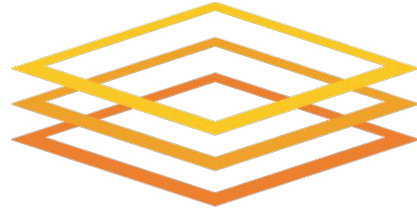# Changes to OSG

and how they affect US WLCG sites

Jeff Dost (OSG Operations / UCSD) HEPIX Spring 2019

Introduction

- In spring of 2018, central OSG services were transitioned out of Indiana University and into other participating OSG institutions
- This past year there has been some reorganization, so some services have a new home and / or name and interface
- The goal of this talk is to introduce the site facing services offered by the Open Science Grid today

Retired Services

- ## The OSG CA has been retired
  - ### Sites should use IGTF InCommon or Let's Encrypt for hostcerts
    - Old style service certs with CN=servicename/hostname are not supported in InCommon, sites should use hostcerts for services where IGTF is required
  - ### User Certs are generally not needed for OSG, but WLCG users should obtain certs from the CERN CA
    - Non-WLCG users can use CILogon Basic
- ## Other retired services:
  - ### Central RSV service
  - ### VOMS Admin Servers (non-WLCG VOs)
- ## More information about the transition:
  - ### https://opensciencegrid.org/technology/policy/service-migrations-spring-2018/

This slide heavily borrows from Brian Lin's HOW 2019 talk: OSG Software: The year in review

Resources for new Sites

- Documentation:

  https://opensciencegrid.org/docs/

- Support:
  - via ticket:
    - WLCG related: ggus.eu (support unit: OSG Software Support)
    - Non-WLCG freshdesk: https://support.opensciencegrid.org/support/tickets/new
  - via email: help@opensciencegrid.org

- 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 (stored in private repo, **not** in GitHub)
  - 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

Open Science Grid

- 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#subcluster-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
- 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
- OSG also provides a cvmfs repo for singularity images:

  /cvmfs/singularity.opensciencegrid.org

* 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 CMS and 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?