Campuses and OSG Services

Tim Cartwright

University of Wisconsin–Madison

OSG Campus Coordinator

OSG Deputy Executive Director

OSG School Director

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

Introduction

- **Yesterday:** talked about services we provide to researchers, e.g.,
  - Open Science Compute Federation (OSCF), esp. Open Science Pool
  - Open Science Data Federation (OSDF)
- **Today:** *how* we help campus IT organizations (writ large) share their compute and storage capacity
- Why might campuses share?
  - Join a larger ecosystem and contribute beyond their borders
  - Connect with and learn from other campuses
  - Maximize usage of capacity

Sharing Compute

Sharing Compute — Overview

- **Campuses** provide raw compute capacity
  - Any batch system
  - Any “shape” of capacity
- **OSG services** create a consistent, usable runtime environment
  - We call this an Execution Point (EP)
  - Gathered into pools
  - Available via Access Points

Sharing Compute — Options

- **OSG Glide-in jobs via a Hosted CE (Compute Entrypoint)**
  - Nearly all new campus integrations use this method
  - Will detail next
- **OSG Glide-in jobs via a CE that you run**
  - But don't do this!
  - Extra work on your end, only necessary in uncommon situations
- **OSG Backfill Container**
  - May be useful if you manage computing with, e.g., Kubernetes
  - Few campuses use this approach but contribute a lot

Sharing Compute via Hosted CE

- **Campus IT/cluster requirements**
  - An OSG account that can submit into your cluster
  - SSH access to that account (using public/private keys)
  - Shared user home directories between submit and worker nodes
  - Scratch space, either on each node or shared using, e.g., BeeGFS
  - Unrestricted outbound network connectivity from worker nodes
- We ***host and operate*** the remaining required services
  - OSG GlideinWMS Factory sends *glide-in* jobs to Hosted CE
  - Hosted CE uses SSH connection to submit those jobs to your cluster
  - When run, the glide-in job sets up the EP environment and joins a pool

Sharing Compute — *Optional* Extras

- Once integrated, extra software and services may increase usage
- **Apptainer** (née Singularity) – extra software on each worker node
  - Container runtime without root privileges
  - If glide-in jobs detect a working install, will run payload jobs in containers
  - Some payload jobs *require* container support – thus you can run more
  - https://osg-htc.org/docs/worker-node/install-apptainer/
- **Frontier Squid** caching proxy – extra service next to cluster
  - Caches HTTP and HTTPS fetches from worker nodes
  - May reduce WAN bandwidth and improve throughput
  - https://osg-htc.org/docs/data/run-frontier-squid-container/

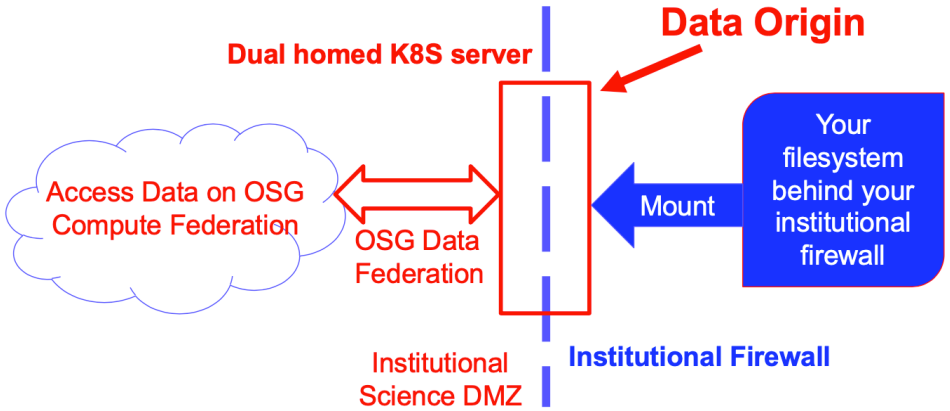
Sharing Storage

Sharing Storage — Overview

- Open Science Data Federation (OSDF) – our storage services
  - OSDF Origins – where data are placed
  - OSDF Caches – where data are accessed
  - Appears as one global namespace
- **Note:** Newer set of services than OS**C**F
  - Integration of campus capacity is particularly new
  - Don’t let that be a deterrent – we will find ways to make things work!

Sharing Storage — Export Your Data

- Export your data, read-only, into the OSDF
  - You provide access to one or more parts of your filesystem
  - You choose whether data are public or protected
  - We manage an OSDF Origin for you – makes data accessible via OSDF
  - Can support writing back to Origin for your users (you provide auth)



Sharing Storage — Other Options

- Donate storage capacity to us, and we manage user access
  - Still exploring and experimenting with this option, but it is likely the future
  - Interested? Contact us (support@osg-htc.org) and let's figure it out!
- Provide storage capacity as OSDF cache
  - You provide storage capacity
  - We manage the OSDF Cache service that uses it

Collaborating

Collaborating — Overview

- Typical phases of a Campus/OSG collaboration:
  - **Planning** – Before proposal, order, or integration
  - **Integration** – Technical and human aspects
  - **Operations** – Monitoring, accounting, and support
- Recently added a Campus Coordinator role (me!)
  - Single point-of-contact throughout the lifecycle
  - Can connect to all PATh teams

Collaborating — Planning

We provide guidance on how to integrate campus capacity with a national CI ecosystem (e.g., OSPool, OSDF)

- **Consult** on topics like the ones covered here, with more details
  - Options for integration with PATh services
  - CI architecture and system requirements
- We provide ***letters of collaboration*** for proposals, incl. NSF CC*

Collaborating — Integration

- Meet once to understand goals and form technical plans
- Carry out the integration plan
  - Recent improvement (OSCF): Aim for first jobs in just 1 working meeting
- We provide detailed technical documentation
  - https://osg-htc.org/docs/
  - Reviewed and updated periodically, plus ad hoc as needed
- Support you throughout process
  - Contact us at any time: support@osg-htc.org
  - Always have Campus Coordinator to turn to

Collaborating — Operations

- We proactively monitor our services and accounting for issues
  - When we find issues in our services, we resolve and let you know
  - When we find issues that may be yours, we let you know and follow up
- Check in with active campuses
  - Listen for new opportunities, including pain points we should address
  - Identify potential connections with other campuses, etc.
- Continue to provide support
  - Respond to questions, concerns, requests
  - Provide software and service updates
  - Improve documentation

Collaborating — Your Responsibilities

- Notify us of changes to your infrastructure, especially:
  - Downtime, scheduled or otherwise (when you can)
  - Major software changes (e.g., to OS, batch system, etc.)
  - Networking changes, esp. firewall updates that could affect us
  - Significant changes to hardware
- Respond to our operational requests in a timely manner
  - In the rare case of campus-specific security concerns, we will provide desired response time
- For big changes, we can start with planning again, etc.

Summary

- We strive to make integration of compute and storage capacity as painless and effective as possible
- Tell us where it hurts, so at least we have the opportunity to fix it
- **To get started: support@osg-htc.org**
  - Or to me, OSG Campus Coordinator: Tim Cartwright <cat@cs.wisc.edu>

Questions?

This material is based upon work supported by the National Science Foundation under Grant Nos. 1836650 and 2030508. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.