OSG Connect

Thursday 9:00 AM
Bala Desinghu (bm@uchicago.edu)
University of Chicago, OSG User Support Team
Overview

● What is OSG Connect

● OSG Connect Services
  ○ Help Desk and User Support
  ○ Software Support
  ○ Data Management: Storage and Transfer
Open Science Grid (OSG)

A framework for large scale distributed resource sharing addressing the technology, policy, and social requirements of sharing computing resources.

Integrates computing and storage resources from over 120 sites in the U.S. These resources are owned by virtual organizations (VOs).

A Virtual Organization (VO) is a set of groups or individuals defined by some common cyber-infrastructure need. This can be a scientific experiment, a university campus or a distributed research effort.
Open Science Grid (OSG)

- OSG supports a default virtual organization (VO) called "OSG"
- If you are not already part of a VO, you can join OSG via OSG Connect
- OSG welcomes any researcher affiliated with an U.S. institution!

Tell colleagues who don’t have a local VO!
Open Science Grid (OSG)

- Local VO: Users with Campus/Org
- OSG Connect: Affiliated with U.S. research university/institute
- XD Connect: Users with XSEDE allocations

> 120 sites
## Review: OSG Submit Locations

<table>
<thead>
<tr>
<th>Available to:</th>
<th>Local</th>
<th>OSG Connect</th>
<th>XD Connect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Researchers affiliated with institution</td>
<td>Affiliates of U.S. research orgs</td>
<td>Users with XSEDE allocation</td>
</tr>
<tr>
<td>Compute resource</td>
<td>Opportunistic + Allocations</td>
<td>Opportunistic</td>
<td>Allocations</td>
</tr>
<tr>
<td>Limit on CPU hrs</td>
<td>Unlikely</td>
<td>No</td>
<td>Yes (per allocation)</td>
</tr>
<tr>
<td>User support</td>
<td>Local staff</td>
<td>Help Desk, Documentation, Support chat, email, and tickets</td>
<td>OSG Connect Support and, if available, campus champion</td>
</tr>
<tr>
<td>Submit location</td>
<td>Local submit server</td>
<td>login.osgconnect.net</td>
<td>xd-login.opensciencegrid.org</td>
</tr>
</tbody>
</table>
OSG Connect

- OSG VO
- Jobs are submitted via HTCondor
- Provides online guides, remote human support, software, and data support
- Submit locally, run globally
Overview

● What is OSG Connect

● OSG Connect Services
  ○ Help Desk and User Support
  ○ Software Support
  ○ Data Management: Storage and Transfer
Help Desk and User Support

- Knowledge Base
  - User guides/tutorials
  - HTC Recipes
- Forums
- “How do I…?” articles
- Interactive online chat
- Workshops and outreach
- Help Desk: https://support.opensciencegrid.org
- Support email: user-support@opensciencegrid.org
Help Desk Tickets

Create Ticket

Online Chat
Help Desk Articles - Basic Topics

**OSG Connect User Guide**

**Getting Started with OSG Connect (5)**
- Registration and Login for OSG Connect
- OSG Connect Quickstart
- Start a Project with OSG Connect
- The "tutorial" Command
- Generate SSH key pair and add the public key to your account

**Choosing Resources for Jobs (5)**
- Steer your jobs with HTCondor job requirements
- Large Memory Jobs
- GPU Jobs
- Multicore Jobs
- Singularity Containers

**Running applications on OSG Connect (6)**
- Accessing Software using Distributed Environment Modules
- Software modules catalog
- Troubleshooting Condor errors
- Requesting a software installation
- Software transfer via HTCondor or HTTP
- See all 6 articles

**Data Management**

**Introduction to data management on OSG (1)**
- Guidelines for data management in OSG - Storage and Transfer

**Data storage (1)**
- Storage Solutions on OSG: home, local, scratch, stash, and public

---

Getting started info and how-to's for serial HTC jobs, multicore jobs, containers, and more
The tutorial command quickly prepares files to launch an HTC recipe.

There are a number of recipes for common workloads available.
Submit Host for this workshop: user-training.osgconnect.net

For permanent accounts: login.osgconnect.net

- Today you need to be able to ssh to:
  username@user-training.osgconnect.net

- Please let an instructor know if you are not able to log in.

- The workshop account is valid for two weeks. If you are interested in a long-term account, please signup after the workshop: http://osgconnect.net/signup
OSG Connect Submit Hosts

- Job Manager: HTCondor
- Workflow Managers: DAGMan, Pegasus, and, Makeflow
- Use `condor` commands and submit files as usual!
- One exception: `condor_status`

```
condor_status -pool osg-flock.grid.iu.edu
```
Tutorials on OSG Connect: tutorial Command

- Tutorials are maintained in Github and downloaded on demand
- Each tutorial’s README is in the OSG Support site
  - [http://osg.link/connect/userguide](http://osg.link/connect/userguide)
  - [http://osg.link/connect/recipes](http://osg.link/connect/recipes)
- These are recommended for learning new techniques on OSG Connect
tutorial Command

```
sh$ tutorial

tutorial

usage: tutorial list - show available tutorials
      tutorial info <tutorial-name> - show details of a tutorial
      tutorial <tutorial-name> - set up a tutorial

Currently available tutorials:
AutoDockVina ............... Ligand-Receptor docking with AutoDock Vina
R .......................... Estimate Pi using the R programming language
R-addlibSNA ................ Shows how to add R external libraries for the R jobs
ScalingUp-Python ............ Python example to optimize a function on grid points
```
tutorial Command

```
ssh$ tutorial quickstart
Installing quickstart (master)...
Tutorial files installed in ./tutorial-quickstart.
Running setup in ./tutorial-quickstart...
ssh$ cd tutorial-quickstart/
ssh$ ls
Images osg-template-job.submit short.sh tutorial02.submit
log README.md tutorial01.submit tutorial03.submit
```
Overview

● What is OSG Connect

● OSG Connect Services
  ○ Help Desk and User Support
  ○ Software Support
  ○ Data Management: Storage and Transfer
Software Support

There are several ways to run software on OSG. (See Christina’s talk yesterday). We will focus on two approaches.

- OASIS (OSG Application Software Installation Service)
- Singularity containers (Advanced material, time permitting)
What is OASIS?

- Repository for common user software
- Compiled and maintained by the user support team
- The repo contains about 180 software packages and libraries, including most commonly used open source science and engineering tools
- Available across ~90% of OSG sites
- Let us know if you need a package installed!
OASIS is easy to use

- Accessed with the `module` command
- Make sure the execution site has OASIS installed by including this in your HTCondor job description file:
  ```
  requirements = (HAS_MODULES =?= true)
  ```
- In your execution script file:
  ```
  module load package-name
  ```
### Accessing OASIS on the Submit Host

```bash
sh$ module avail
```

```
------------------ /cvmfs/oasis.opensciencegrid.org/osg/modules/modulefiles/Core-------------------

<table>
<thead>
<tr>
<th>Module</th>
<th>Version</th>
<th>Dependent Version</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANTS/1.9.4</td>
<td></td>
<td>ectools</td>
<td>lapack/3.5.0</td>
</tr>
<tr>
<td>ANTS/2.1.0</td>
<td>(D)</td>
<td>eemt/0.1</td>
<td>lapack/3.6.1</td>
</tr>
<tr>
<td>MUMmer/3.23</td>
<td></td>
<td>elastix/2015</td>
<td>libXpm/3.5.10</td>
</tr>
<tr>
<td>OpenBUGS/3.2.3</td>
<td></td>
<td>entropy/2017.03.16</td>
<td>libgfortran/4.4.7</td>
</tr>
<tr>
<td>R/3.1.1</td>
<td>(D)</td>
<td>espresso/5.1</td>
<td>libtiff/4.0.4</td>
</tr>
<tr>
<td>R/3.2.0</td>
<td></td>
<td>espresso/5.2</td>
<td>(D) llvm/3.6</td>
</tr>
<tr>
<td>R/3.2.1</td>
<td></td>
<td>ete2/2.3.8</td>
<td>llvm/3.7</td>
</tr>
</tbody>
</table>

[...]
```
Basic OASIS Commands

- Load a software module:
  
  \texttt{module load package-name}

- List loaded modules:

  \texttt{module list}

- Unload a module (to prepare for another)

  \texttt{module unload package-name}
Using OASIS on the Submit Host

```
sh$ module load R
sh$ which R
/cvmfs/oasis.opensciencegrid.org/osg/modules/R/3.1.1/bin/R
sh$ Rscript --version
R scripting front-end version 3.1.1 (2014-07-10)
sh$ module list
Currently Loaded Modules:
  1) R/3.1.1
```
Overview

● What is OSG Connect

● OSG Connect Services
  ○ Help Desk and User Support
  ○ Software Support
  ○ Data Management: Storage and Transfer
    (we cover some basics, more details in Derek’s talk)
## Data Storage on OSG Connect

<table>
<thead>
<tr>
<th>System</th>
<th>Default Limit</th>
<th>Purpose</th>
<th>Network mounted</th>
<th>Backed Up</th>
</tr>
</thead>
<tbody>
<tr>
<td>home</td>
<td>20 GB</td>
<td>Quick data access and not for submitting jobs</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>local-scratch</td>
<td>25 GB</td>
<td>Large temporary storage and I/O for your jobs. Files older than 30 days are automatically removed.</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>stash</td>
<td>200 GB</td>
<td>Large storage and accessible via Globus to/from your campus or laptop</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>public</td>
<td>10 GB</td>
<td>Sharing data and transfer input data via HTTP or stashcp</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

- Both **stash** and **public** are on the same filesystem.
- Public directory has an http interface and the files are world readable.
- Let us know if you need more!
## Transferring Input Data for your Job

<table>
<thead>
<tr>
<th>Method</th>
<th>Recommended File Size</th>
<th>Command</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTCondor</td>
<td>&lt; 100 MB</td>
<td>transfer_input_files</td>
<td>Input data from home, local-scratch, public or stash</td>
</tr>
<tr>
<td>HTTP and UNIX tools</td>
<td>&lt; 1 GB</td>
<td>wget, curl, or rsync</td>
<td>Input data from ~/public for HTTP tools (^wget, curl^), or home, local-scratch, public or stash (^rsync^)</td>
</tr>
<tr>
<td>StashCache</td>
<td>&gt; 1 GB, &lt; 50 GB</td>
<td>stashcp</td>
<td>Input data from ~/public</td>
</tr>
<tr>
<td>GridFTP</td>
<td>&gt; 1 GB &lt; 50 GB</td>
<td>gfal-copy</td>
<td>Experts with large workflows. Contact us if you want to use it.</td>
</tr>
</tbody>
</table>
## Transferring Output Data from your Job

<table>
<thead>
<tr>
<th>Method</th>
<th>Recommended File Size</th>
<th>Command</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTCondor</td>
<td>&lt; 100 MB</td>
<td><code>transfer_output_files</code></td>
<td>Transfer data to submit directory</td>
</tr>
<tr>
<td>UNIX Tools</td>
<td>&lt; 1 GB</td>
<td><code>rsync, scp, etc.</code></td>
<td>Transfer data to home, local-scratch, stash, etc.</td>
</tr>
<tr>
<td>GridFTP</td>
<td>&gt; 1 GB, &lt; 50 GB</td>
<td><code>gfal_copy</code></td>
<td>Experts with large workflows. Contact us if you want to use it.</td>
</tr>
</tbody>
</table>

**Submit Host**

Output Data (HTCondor, HTTP, GridFTP)

**Remote Worker Machine**
Data Transfer from OSG Connect

<table>
<thead>
<tr>
<th>Method</th>
<th>Data Size</th>
<th>Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secure Copy Protocol</td>
<td>&lt; 1GB</td>
<td>scp, putty, WinSCP, gFTP, etc.</td>
</tr>
<tr>
<td>Globus</td>
<td>&gt; 1GB</td>
<td>Globus web service or globus CLI</td>
</tr>
</tbody>
</table>

NOTE: Globus transfer is available through the OSG Connect Globus “endpoint”. You will need to a Globus personal endpoint to transfer to your laptop.
OSG Connect Exercises (twiki)

ssh username@user-training.osgconnect.net

1.1 Get acquainted with OSG Connect
1.2 Do the “OSG Connect Quickstart”
1.3 Run ‘Gromacs’ via the OASIS module
1.4 Submit the tensorflow example “tf-matmul.py” on the OSG and see how it works.

BONUS: submit jobs from osg-learn using other software in OASIS (matlab, python, etc)
Advanced Material follows - Containers & TensorFlow
Singularity Containers

- Workload runs inside a container
- In fact, ~75% of OSG VO jobs run inside a container without the user even knowing about it
  - Provides a consistent environment and job separation
- User defined images are supported!
  - Complex software deployments
  - Non-Red-Hat-based Linux operating systems, e.g. Debian, Ubuntu, etc.
User created Singularity Containers

● You:
  ○ Define image in Docker
    ■ We have a set of images you can base your image on
  ○ Push image to Docker Hub

● Our system will:
  ○ Regularly check Docker Hub for updates on registered images
  ○ Automatically synchronize the latest version to CVMFS
  ○ Make the image available in exploded form under /cvmfs/singularity.opensciencegrid.org/…
    ■ This is a very efficient distribution mechanism for images
Singularity on OSG: Documentation

- Further details:
  - Help Desk article: https://goo.gl/FmVkKN
  - Derek’s Blog: https://goo.gl/LBtBbw

A brief discussion about an example use case (TensorFlow)
TensorFlow Installation
- TensorFlow is a very active project which requires up-to-date Python modules and system libraries - Makes it a difficult installation on long-term supported Red Hat Enterprise Linux distributions

TensorFlow Singularity Solution - OSG provides vetted TensorFlow images
- CPU version: directly imported from Docker image release by TensorFlow project
- GPU version: based on NVIDIA’s CUDA image, with TensorFlow added
Getting TensorFlow Tutorial

```
sh$ $ tutorial tf-matmul

Installing tf-matmul (master)...

Tutorial files installed in ./tutorial-tf-matmul.

Running setup in ./tutorial-tf-matmul...

sh$ cd ./tutorial-tf-matmul

sh$ ls

README.md  tf_matmul.py  tf_matmul.submit  tf_matmul_wrapper.sh
```
Using Singularity to run TensorFlow

Running the singularity container on the submit host

```sh
sh$ python tf_matmul.py
Traceback (most recent call last):
  File "tf_matmul.py", line 3, in <module>
    import tensorflow as tf
ImportError: No module named tensorflow

sh$ singularity shell /cvmfs/singularity.opensciencegrid.org/tensorflow/tensorflow:latest
sh$ python tf_matmul.py
result of matrix multiplication
===============================
[[  1.00000000e+00   0.00000000e+00]
 [ -4.76837158e-07   1.00000024e+00]]
===============================
```

Throws error because tensorflow is only available in the container environment and not as a regular package.

Start a container and a shell inside the container

Inside the container, the job execution is successful.
Requesting TensorFlow Containers on OSG

Running on the remote worker machine

- Take a look at the job description file
- Requirements = HAS_SINGULARITY == True (Find a machine that has singularity installed)
- +SingularityImage = 
  
"/cvmfs/singularity.opensciencegrid.org/tensorflow/tensorflow:latest" (use the container image on cvmfs)
OSG Connect Exercises (twiki)

ssh username@user-training.osgconnect.net

1.1 Get acquainted with OSG Connect
1.2 Do the “OSG Connect Quickstart”
1.3 Run ‘Gromacs’ via the OASIS module
1.4 Submit the tensorflow example “tf-matmul.py” on the OSG and see how it works.

BONUS: submit jobs from osg-learn using other software in OASIS (matlab, python, etc)