## **OSG Connect**

Thursday 9:00 AM

Bala Desinghu (bala.desinghu@rutgers.edu)

Rutgers University, Office of Advanced Research Computing University of Chicago, OSG User Support Team (former)

### Overview



What is OSG Connect

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

## **OSG Integrates Computing Resources**

The OSG facilitates access to distributed high throughput computing for research in the US.

Integrates computing and storage resources from about 100 sites in the U.S. These resources are owned by virtual organizations (VOs)

Arzona Mexico Olima Artanasa Teoria Cardonal Mexico Missouri Remucky Variation of Cardonal Mexico Missouri Remucky

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.

### OSG Provides Opportunistic Resources

- OSG supports a default virtual organization (VO) called "OSG"
- List of VO's
- 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!

### **OSG Connect:**

### Entry point to access the OSG opportunistic resources





- Enables US researchers to access the OSG compute resources
- Jobs are submitted via HTCondor
- Provides online guides, remote human support, software, and data support
- Submit locally, run globally

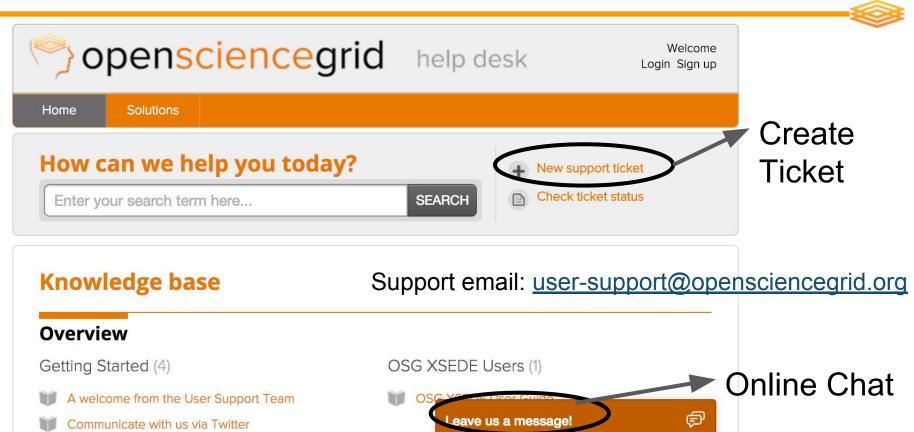
### Overview



What is OSG Connect (Questions?)

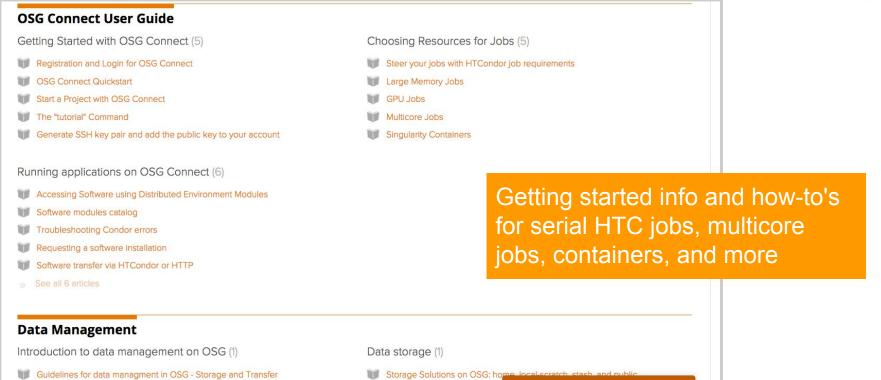
- OSG Connect Services
  - Help Desk and User Support
  - Software Support
  - Data Management: Storage and Transfer

## Help Desk: <a href="https://support.opensciencegrid.org">https://support.opensciencegrid.org</a>



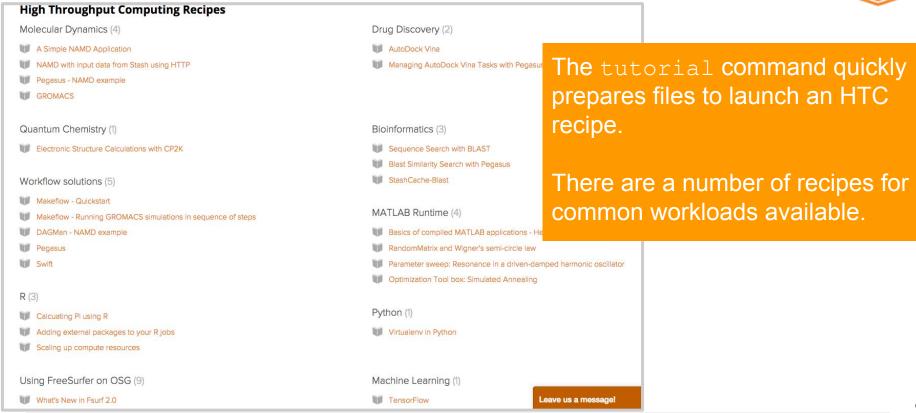
## Help Desk Articles - Basic Topics





## Help Desk Articles - HTC Recipes





### **OSG Connect Submit Hosts**



Submit Host for this workshop: training.osgconnect.net

For permanent accounts: login.osgconnect.net (login01, login02, login03)



- Today you need to be able to ssh to: username@training.osgconnect.net
- Please let an instructor know if you are not able to log in.
- The workshop account is valid for a month. If you are interested in a long-term account, please signup: <a href="http://osgconnect.net/signup">http://osgconnect.net/signup</a>

### **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
flock.opensciencegrid.org
```

### Tutorials on OSG Connect: tutorial Command



- Tutorials are maintained in <u>Github</u> and downloaded on demand
- Each tutorial's README is in the OSG Support site
  - http://osg.link/connect/userguide
  - 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



```
sh$ tutorial quickstart
Installing quickstart (master)...
Tutorial files installed in ./tutorial-quickstart.
Running setup in ./tutorial-guickstart...
sh$ cd tutorial-quickstart/
sh$ ls
        osg-template-job.submit short.sh tutoria102.submit
Images
log README.md tutorial01.submit tutorial03.submit
```

### Overview



What is OSG Connect

- OSG Connect Services
  - Help Desk and User Support (Questions?)
  - Software Support
  - Data Management: Storage and Transfer

## Software Support: OASIS



built

Laptop or Local resources

Remote Worker Machine

(run environment is different from the built environment)

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

- OASIS (OSG Application Software Installation Service)
- Singularity containers (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!



# Accessing OASIS on the Submit Host (training.osgconnect.net/login.osgconnect.net)



### See the available packages in OASIS

module avail

```
sh$ module avail
       ANTS/1.9.4
                                               lapack/3.5.0
                                                                           python/2.7
  ANTS/2.1.0
                         eemt/0.1
                                               lapack/3.6.1
                                                                           python/3.4
                         elastix/2015
  MUMmer/3.23
                                               libXpm/3.5.10
                                                                           python/3.5.2
  OpenBUGS/3.2.3
                         entropy/2017.03.16
                                               libgfortran/4.4.7
                                                                           qhull/2012.1
                                               libtiff/4.0.4
                         espresso/5.1
                                                                           root/5.34-32-py34
  R/3.2.0
                          espresso/5.2 (D)
                                               11vm/3.6
                                                                           root/5.34-32
  R/3.2.1
                                               11 \text{vm} / 3.7
                         ete2/2.3.8
                                                                           root/6.06-02-pv34 (D)
[\ldots]
```

# Accessing OASIS on the Submit Host (training.osgconnect.net/login.osgconnect.net)



```
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
```

## Accessing OASIS for your job

requirements = (HAS MODULES =?= true)



### In your execution script file:



module load package-name

Submit Host Remote Worker Machine (OASIS available)

Remote Worker Machine (no OASIS)

### **Basic OASIS Commands**

Load a software module:

module load package-name

List loaded modules:

module list

Unload a module (to prepare for another)

module unload package-name





## Questions?

### Overview



What is OSG Connect

- OSG Connect Services
  - Help Desk and User Support
  - Software Support (Questions?)
  - Data Management: Storage and Transfer
     (we cover some basics, more details in Derek's talk)

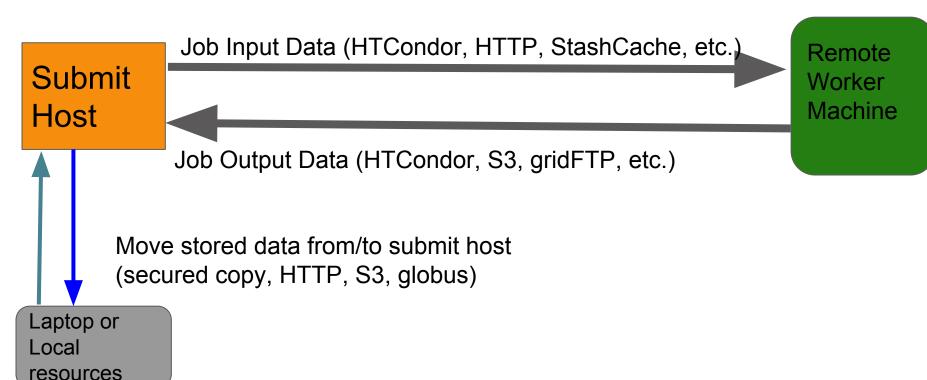
## Data Storage on OSG Connect

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

- 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!

### Data Transfer in OSG Connect





Method	Recommended File Size	Command	Purpose
rrans	iemng inpi	ut Data for y	our Job

**HTCondor** 

tools

HTTP and UNIX

StashCache

Submit

Host

GridFTP

< 100 MB

< 1 GB

> 1 GB

< 50 GB

> 1 GB, < 50 GB

stashcp

Input Data (HTCondor, HTTP, StashCache, etc.)

qfal-copy

ransi	ferring	Input	Data for	r your Job

ransf	err	ng	Input	t Data	a for	your	Job

rans	terring	Input	Data to	r your c	lob

transfer input files

wget, curl, **or** rsync

Input data from home, local-scratch,

Input data from ~/public for HTTP

local-scratch, public or stash (rsync)

Experts with large workflows. Contact

Machine

Remote Worker

26

tools (wget, curl), or home,

Input data from ~/public

us if you want to use it.

public or stash

## Transferring Output Data from your Job

Method	Recommended File Size	Command	Purpose
HTCondor	< 100 MB	transfer_output_files	Transfer data to submit directory
UNIX Tools	< 1 GB	rsync, scp, <b>etc</b> .	Transfer data to home, local-scratch, stash, etc.
GridFTP	> 1 GB, < 50 GB	gfal_copy	Experts with large workflows. Contact us if you want to use it.

Submit Host

Output Data (HTCondor, HTTP, GridFTP)

Remote Worker Machine

### Data Transfer from OSG Connect



Method	Data Size	Tools
Secure Copy Protocol	< 1GB	scp, putty, WinSCP, gFTP, etc.
Globus	> 1GB	Globus web service or globus CLI

NOTE: Globus transfer is available through the OSG Connect Globus "endpoint". You will need to a Globus personal endpoint to transfer to your laptop.

Submit Host

Move stored data from/to submit host

Laptop or Local resources

## OSG Connect Exercises (SchoolPage)



ssh username@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
- BONUS: Try yesterday's examples using OASIS (matlab, python, etc)
- 1.4 Submit the tensorflow example "tf-matmul.py" on the OSG and see how it works.

## Software Support - Containers



built

Laptop or Local resources

Remote Worker Machine

(run environment is different from the built environment)

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

- OASIS (OSG Application Software Installation Service)
- Singularity containers

## Software portability



- Build with compiler tools (make, cmake, etc.)
- OASIS portable modules via CVMFS
- Containers (Dockers, Singularity, Rockers, etc.)
- Virtual Machines

### How much to pack?

	Containers
Size	Small (about 10 - 20 times) compared to VM
Speed	Starts in milliseconds
Overhead	negligible

## Singularity



### Singularity:

can run either a Docker or a Singularity image.

does not need a daemon process to run an image

can run workloads as MPI or OpenMP jobs

### Singularity in OSG:

available in 75% of OSG machines

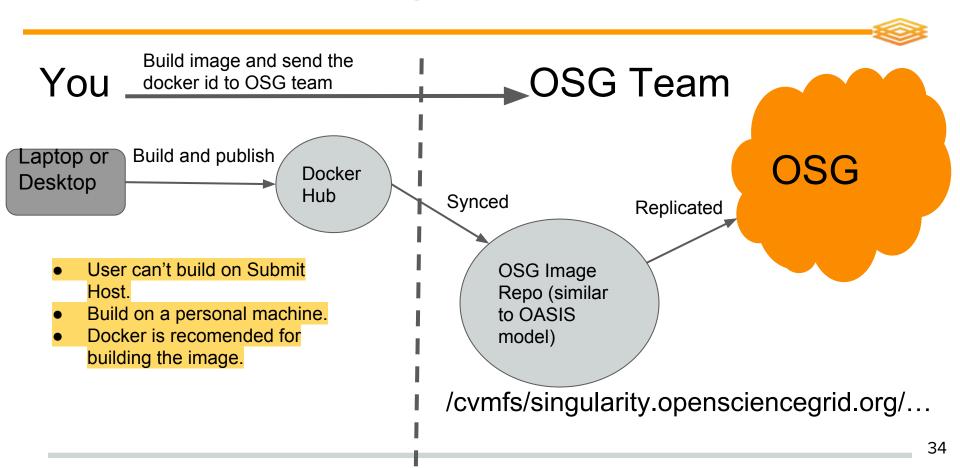
user defined images are supported

## Software support on the OSG



	User	OSG support team
Compilation	Builds the package on submit host or remote workers.	May help to resolve software dependency questions on local or remote workers.
OASIS	Sends a software installation request to OSG team.	Installs the requested package in OASIS which are available on the OSG (~ 90% machines)
Container	Builds the required image on personal machine (laptop/desktop), publish on docker hub, and give the dockerID to the OSG team.	Distributes the images on the OSG (~75% machines). Also, helps to build the image.
VM	(I have not seen any production runs)	May have some support for special cases on some sites

## User created images



## Build image with docker: a simple example



### What you need?

- Docker installed on your personal machine
- Create an account on docker hub (DOCKER\_USER\_ID)
- Docker file (FROM, RUN, etc.)
- Learn basic docker commands (build, push, pull, search, etc.)

```
sh$ $ cat Dockerfile

# Dockerfile - Simple example

FROM ubuntu:latest

MAINTAINER Bala "bala.desinghu@rutgers.edu"

RUN apt-get update

RUN apt-get install -y python python-pip wget

RUN pip install numpy

sh$ $ docker build -t my-python-image .

sh$ docker push $DOCKER_USER_ID/my-python-image
```

## Singularity on OSG: Documentation

- Further details:
  - Help Desk articles:
    - https://goo.gl/FmVkKN
    - https://bit.ly/2IUzWAZ
  - Derek's Blog: <a href="https://goo.gl/LBtBbw">https://goo.gl/LBtBbw</a>



A brief discussion about an example use case (TensorFlow)

## Running TensorFlow Jobs using Singularity





An open-source software library for Machine Intelligence

- 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
  - o 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$ 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+001
                       1.00000024e+00]]
 [ -4.76837158e-07
```

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)

### Overview



What is OSG Connect

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



## OSG Connect Exercises (SchoolPage)



ssh username@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
- BONUS: Try yesterday's examples using OASIS (matlab, python, etc)
- 1.4 Submit the tensorflow example "tf-matmul.py" on the OSG and see how it works.



## Thank You

## Review: OSG Submit Locations

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