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)

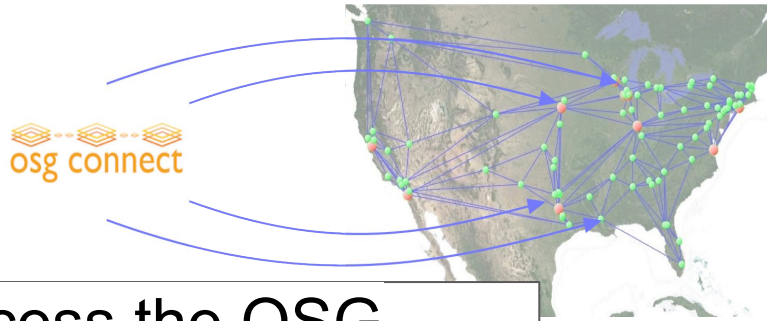
A Virtual Organization (VO) is a set of groups or individuals defined by some common cyber-infrastructure need. This can be a scientific experiment, a university campus or a distributed research effort.

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



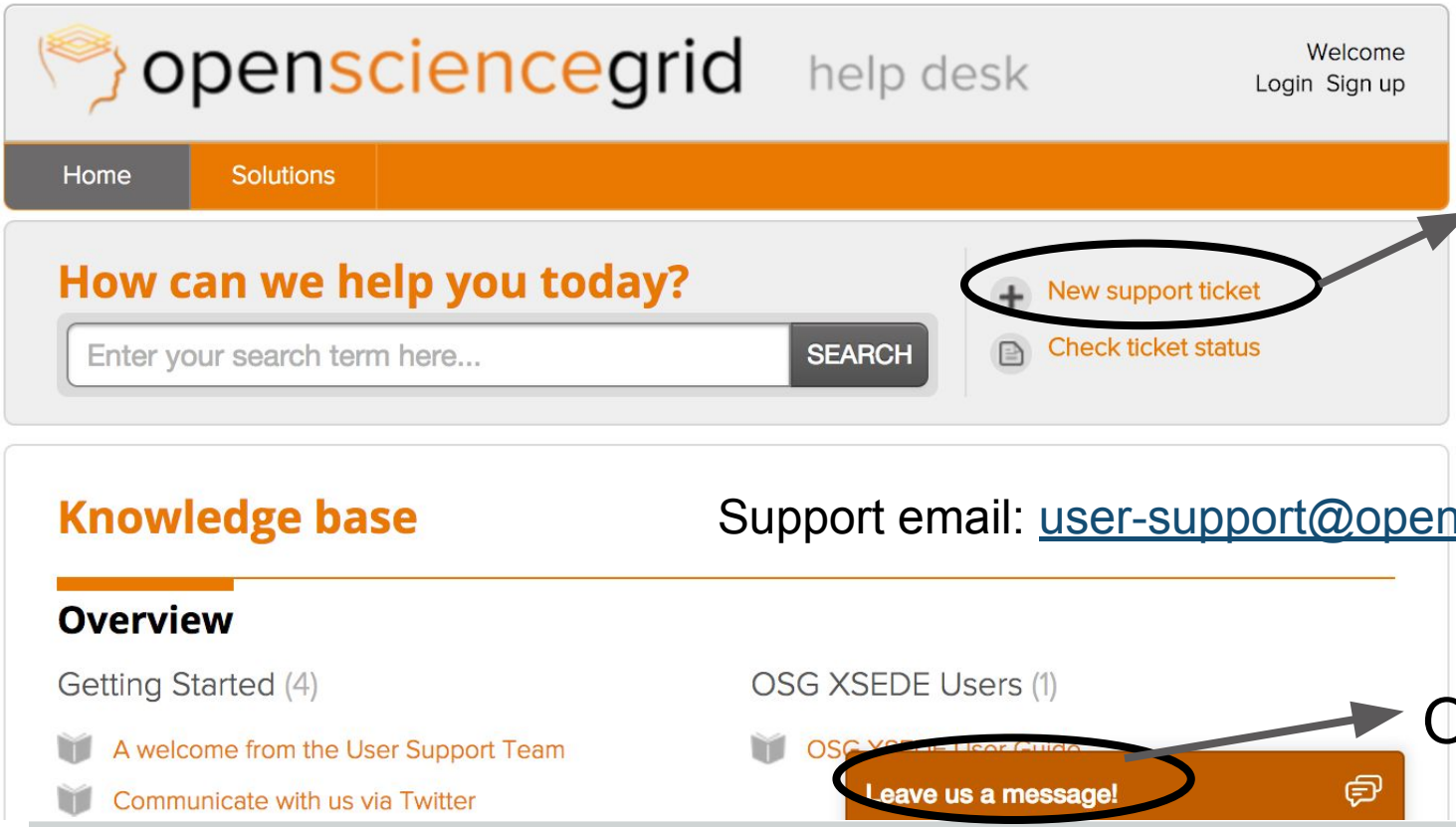
- Enables US researchers to access the OSG compute resources
- Jobs are submitted via HTCondor
- Provides online guides, remote human support, software, and data support
- Submit locally, run globally

Overview

- What is OSG Connect (Questions?)

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

Help Desk: https://support.opensciencegrid.org

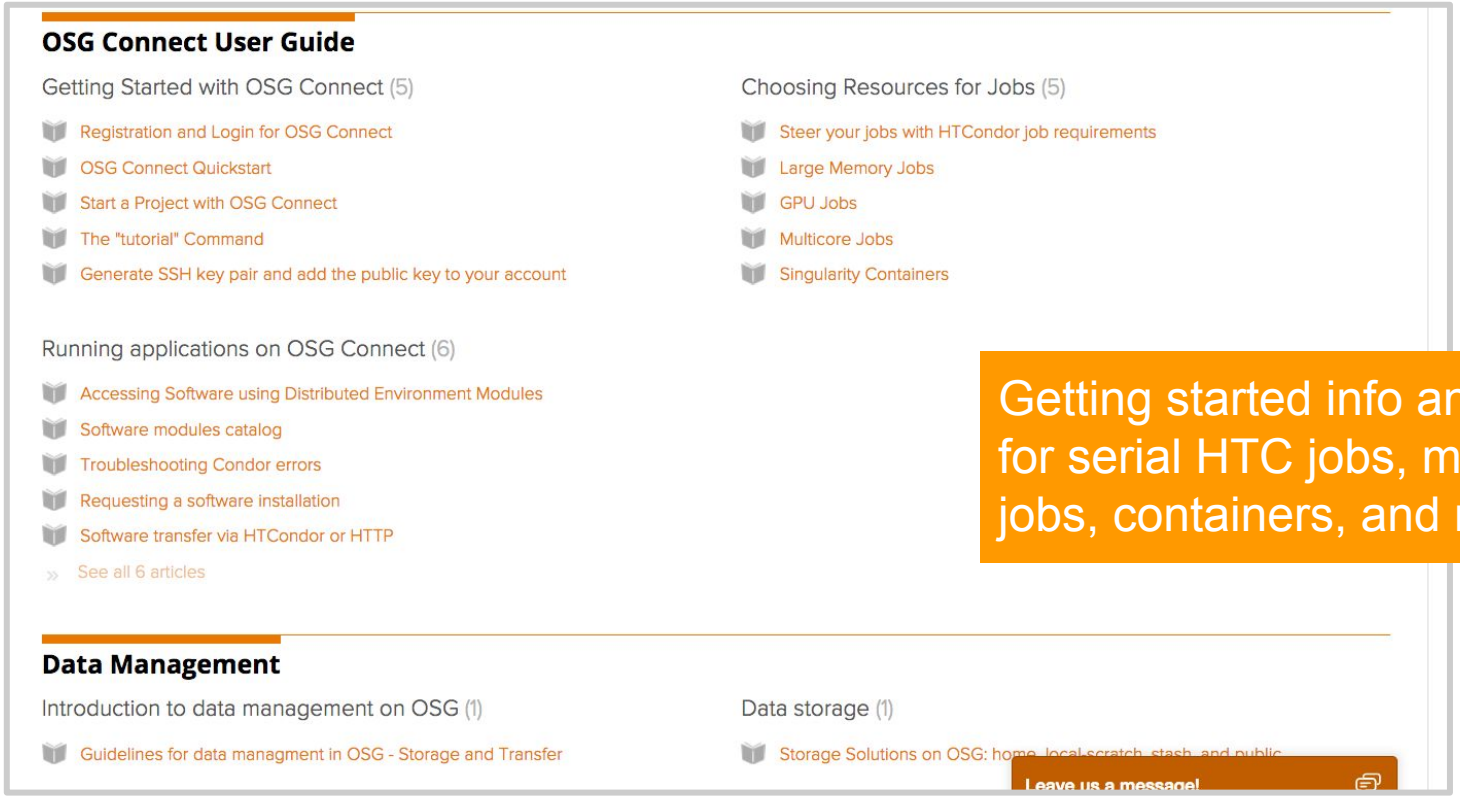


Create Ticket

Support email: user-support@opensciencegrid.org

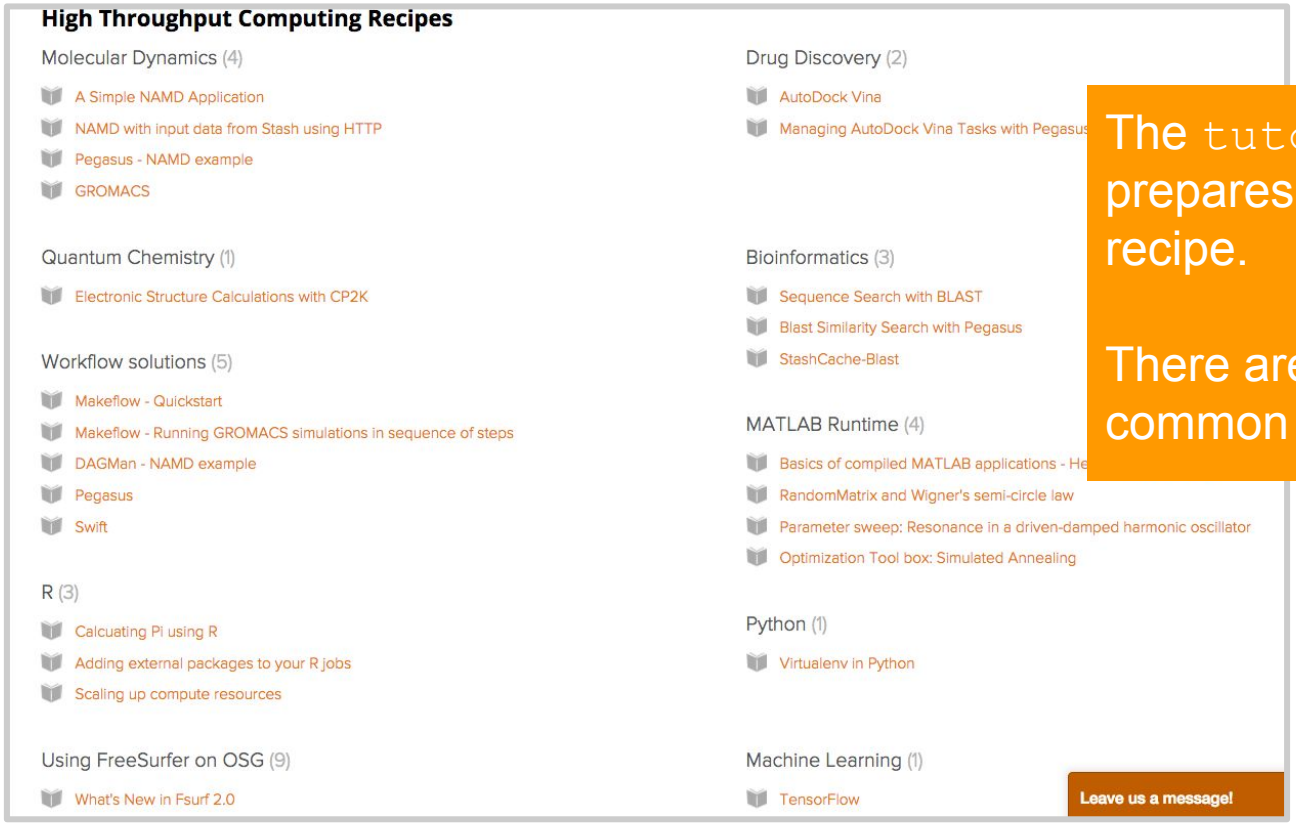
Online Chat

Help Desk Articles - Basic Topics



Getting started info and how-to's for serial HTC jobs, multicore jobs, containers, and more

Help Desk Articles - HTC Recipes



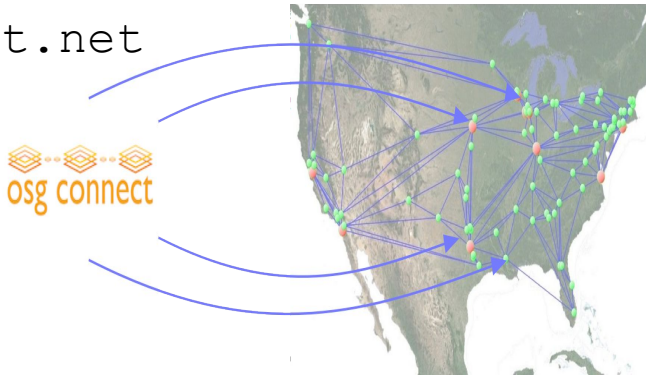
The `tutorial` command quickly prepares files to launch an HTC recipe.

There are a number of recipes for common workloads available.

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: 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
flock.opensciencegrid.org
```

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/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

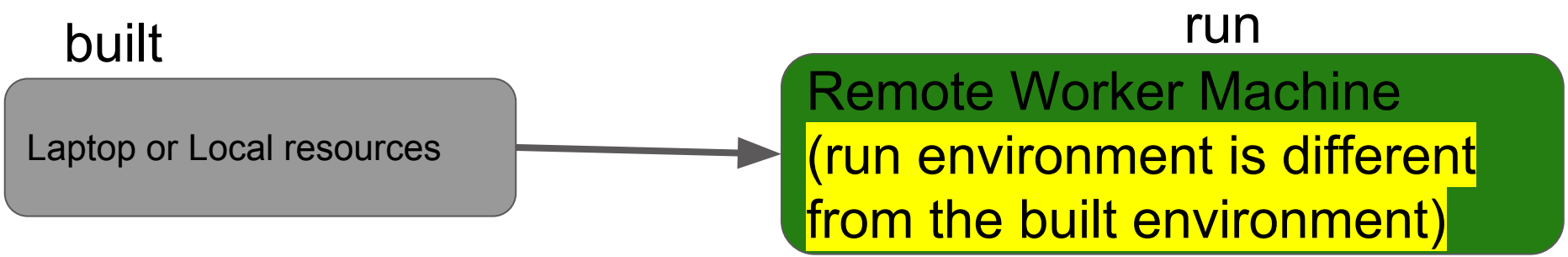
```
sh$ tutorial quickstart
Installing quickstart (master)...
Tutorial files installed in ./tutorial-quickstart.
Running setup in ./tutorial-quickstart…

sh$ cd tutorial-quickstart/

sh$ 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 (Questions?)
  - Software Support
  - Data Management: Storage and Transfer

Software Support: OASIS



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

- OASIS (**O**SG **A**pplication **S**oftware **I**nstallation **S**ervice)
- 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
------------------ /cvmfs/oasis.opensciencegrid.org/osg/modules/modulefiles/Core--------------------
   ANTS/1.9.4                 ectools                  lapack/3.5.0                python/2.7        (D)
   ANTS/2.1.0          (D)    eemt/0.1                 lapack/3.6.1          (D)   python/3.4
   MUMmer/3.23                elastix/2015             libXpm/3.5.10               python/3.5.2
   OpenBUGS/3.2.3             entropy/2017.03.16       libgfortran/4.4.7           qhull/2012.1
   R/3.1.1             (D)    espresso/5.1             libtiff/4.0.4               root/5.34-32-py34
   R/3.2.0                    espresso/5.2       (D)   llvm/3.6                    root/5.34-32
   R/3.2.1                    ete2/2.3.8               llvm/3.7                    root/6.06-02-py34 (D)
[...]
```

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
```

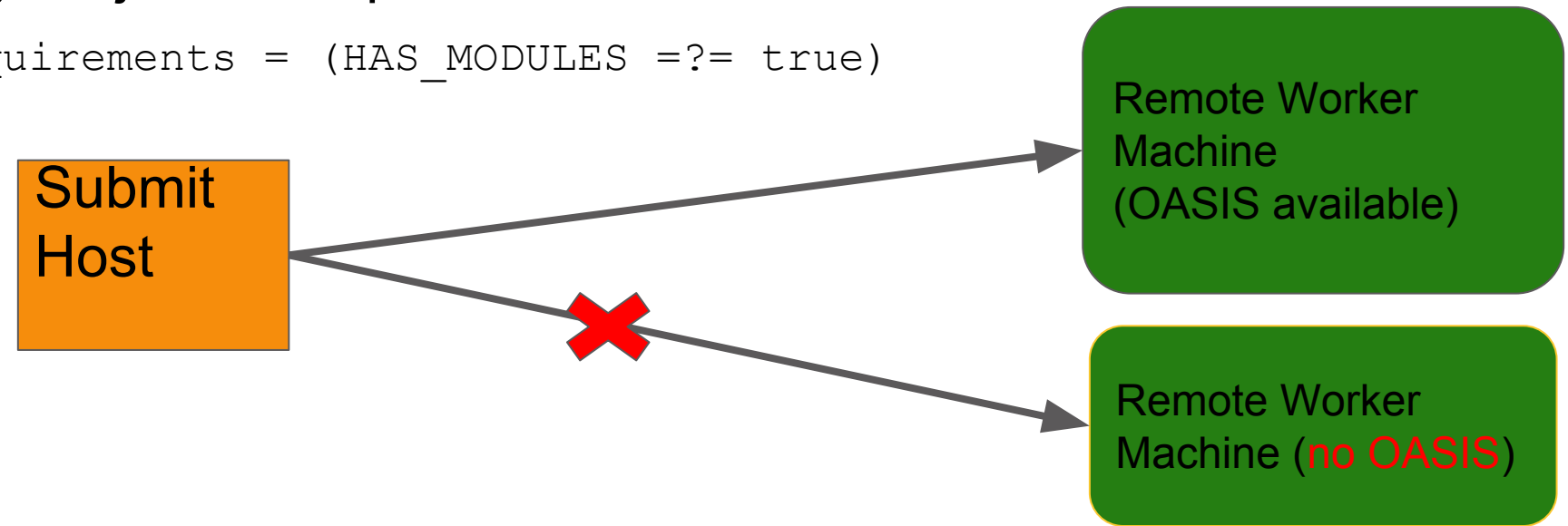
Accessing OASIS for your job

In your execution script file:

```
module load package-name
```

In your job description:

```
requirements = (HAS_MODULES =?= true)
```



Basic OASIS Commands

- Load a software module:

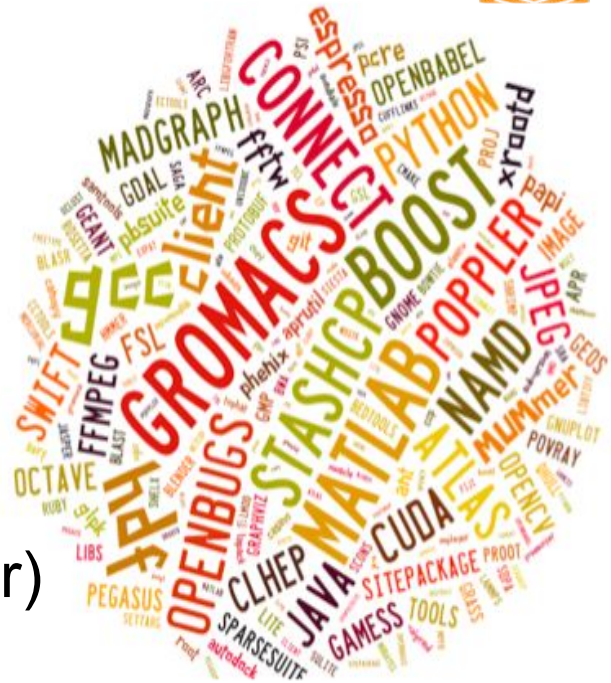
  `module load package-name`

- List loaded modules:

  `module list`

- Unload a module (to prepare for another)

  `module unload package-name`



Overview

- What is OSG Connect

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

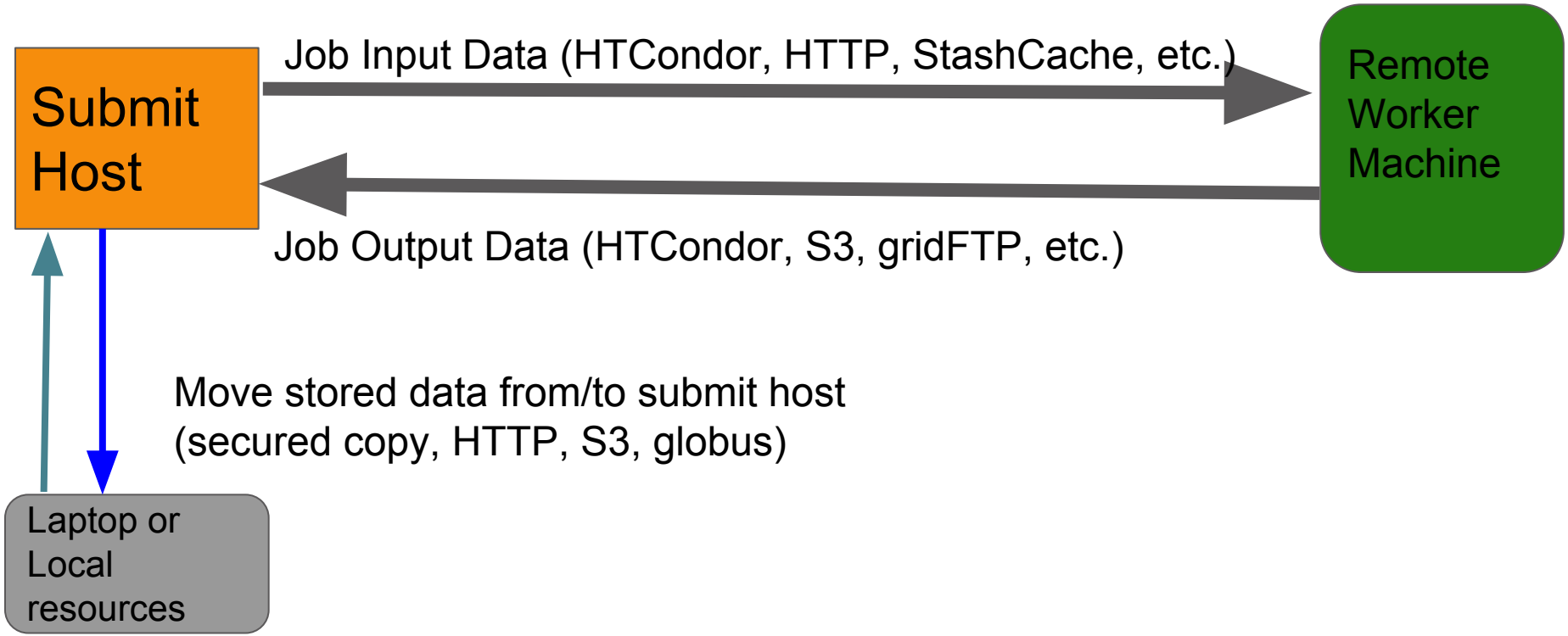
  (we cover some basics, more details in Derek's talk)

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

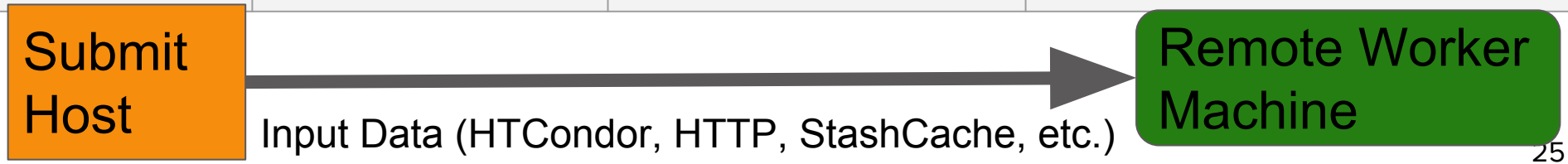
- Both **stash** and **public** are on the same filesystem.
- Public directory has an http interface and the files are world readable.
- Let us know if you need more!

Data Transfer in OSG Connect



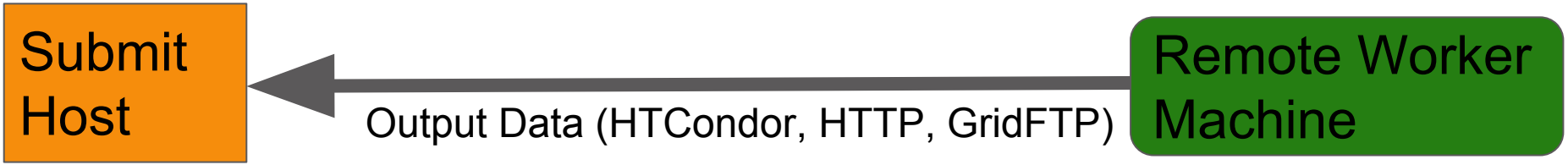
Transferring Input Data for your Job

Method	Recommended File Size	Command	Purpose
HTCondor	< 100 MB	`transfer_input_files`	Input data from home, local-scratch, public or stash
HTTP and UNIX tools	< 1 GB	`wget`, `curl`, or `rsync`	Input data from ~/public for HTTP tools (`wget`, `curl`), or home, local-scratch, public or stash (`rsync`)
StashCache	> 1 GB, < 50 GB	`stashcp`	Input data from ~/public
GridFTP	> 1 GB < 50 GB	`gfal-copy`	Experts with large workflows. Contact us if you want to use it.



Transferring Output Data from your Job

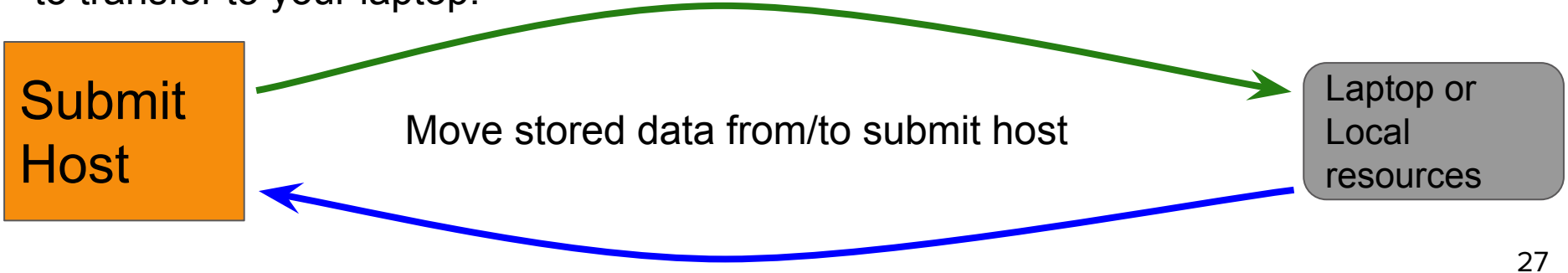
Method	Recommended File Size	Command	Purpose
HTCondor	< 100 MB	`transfer_output_files`	Transfer data to submit directory
UNIX Tools	< 1 GB	`rsync`, `scp`, etc.	Transfer data to home, local-scratch, stash, etc.
GridFTP	> 1 GB, < 50 GB	`gfal_copy`	Experts with large workflows. Contact us if you want to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



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.



OSG Connect Exercises (SchoolPage)

```
ssh username@training.osgconnect.net
```

1.1 Get acquainted with OSG Connect

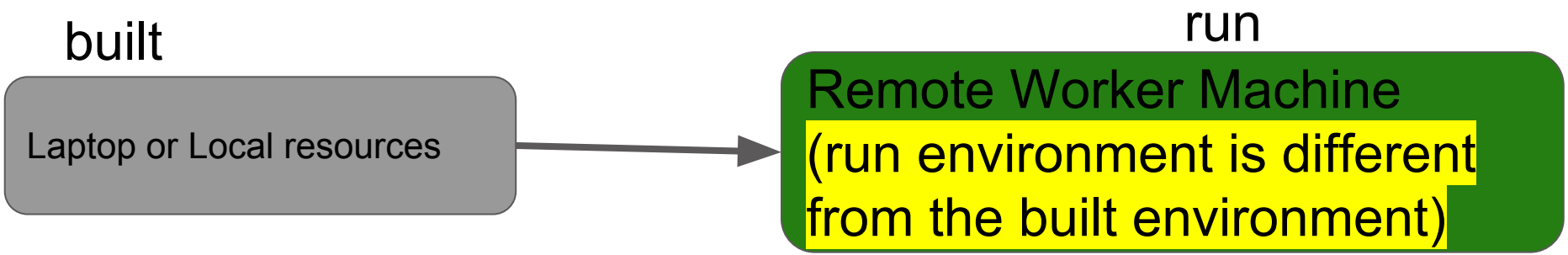
1.2 Do the “OSG Connect Quickstart”

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



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

- OASIS (**O**SG **A**pplication **S**oftware **I**nstallation **S**ervice)
- 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

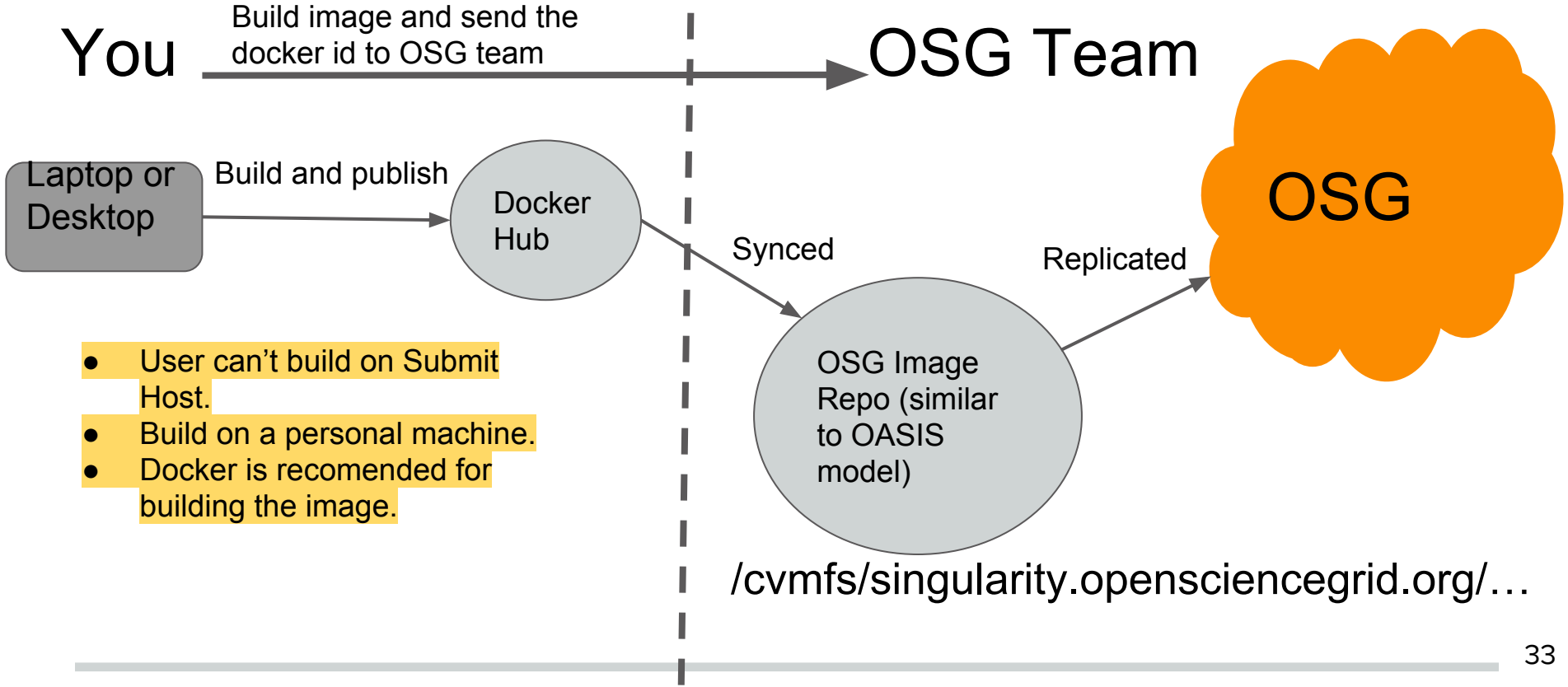
Singularity in OSG:

- 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



- User can't build on Submit Host.
- Build on a personal machine.
- Docker is recomended for building the image.

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: https://goo.gl/LBtBbw

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
  - GPU version: based on NVIDIA's CUDA image, with TensorFlow added

Getting TensorFlow Tutorial

```
sh$ $ tutorial tensorflow-matmul


Installing tensorflow-matmul (master)...

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

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

sh$ cd ./tutorial-tensorflow-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/opensceincegrid/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/opensceincegrid/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

Questions?

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