Introduction to the Open Science Grid

HCC Kickstart

October 22nd, 2020

Emelie Fuchs <efuchs@unl.edu>

OSG Research Facilitator

UNL Holland Computing Center – Applications Specialist

Outline

- What is the OSG?
- Who uses OSG?
- Owned vs. Opportunistic Use
- Characteristics of OSG-Friendly Jobs
- Is OSG Right for Me?

The Open Science Grid

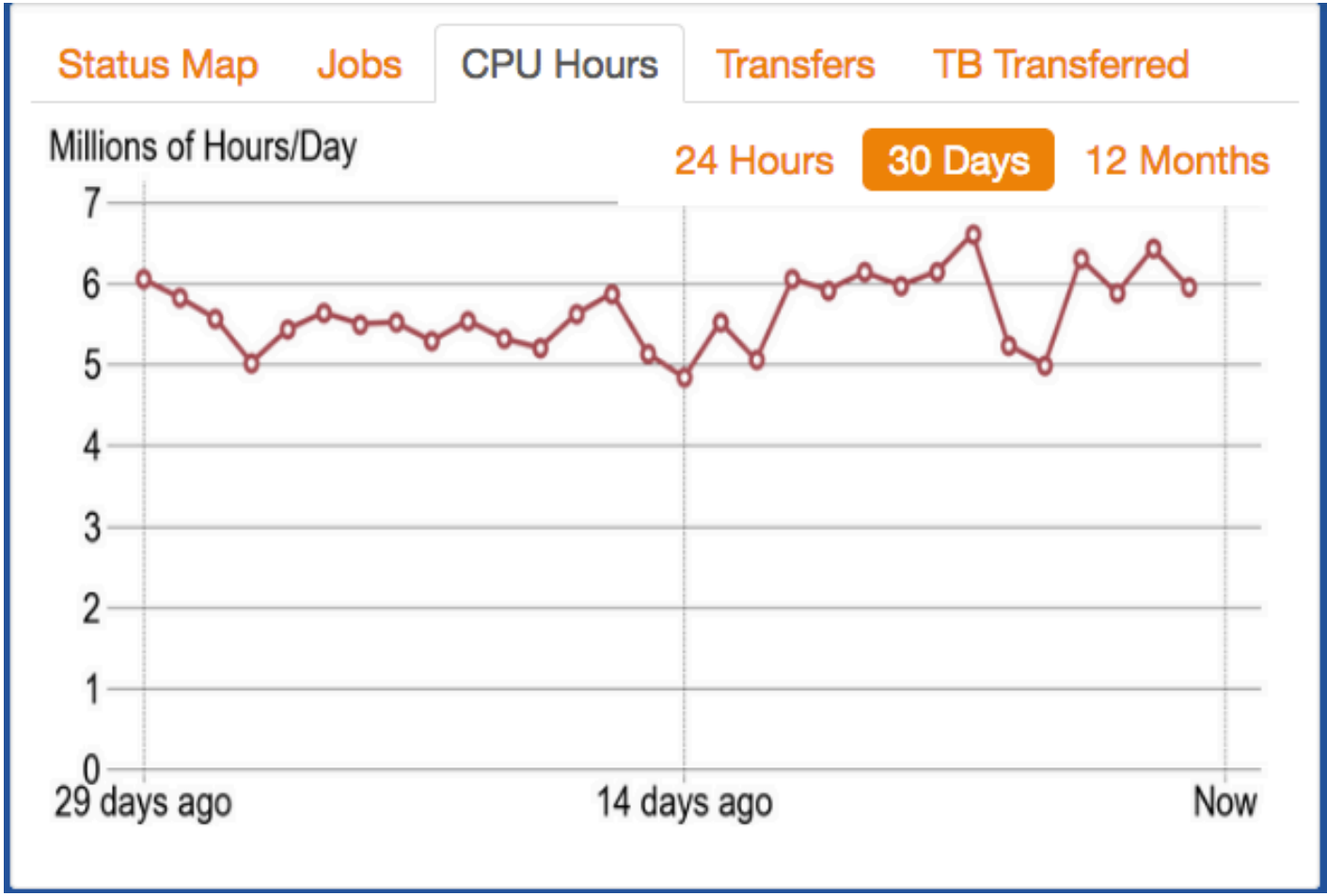
A framework for large scale distributed resource sharing addressing the technology, policy, and social requirements of sharing computing resources.

- The OSG is a consortium of software, service and resource providers and researchers, from universities, national laboratories and computing centers across the U.S., who together build and operate the OSG project.
- Funded by the NSF and DOE.

> 50 research communities
> 130 sites
> 100,000 cores accessible

The Open Science Grid

Over 1.8 billion CPU hours per year!!



In the last 24 Hours	
369,000	Jobs
5,956,000	CPU Hours
4,677,000	Transfers
1,085	TB Transfers
In the last 30 Days	
9,301,000	Jobs
169,553,000	CPU Hours
181,004,000	Transfers
26,581	TB Transfers
In the last 12 Months	
107,759,000	Jobs
1,855,960,000	CPU Hours
2,473,279,000	Transfers
421,000	TB Transfers

Who is Using the OSG?

- Astrophysics
- Biochemistry
- Bioinformatics
- Earthquake Engineering
- Genetics
- Gravitational-wave physics
- Mathematics
- Nanotechnology
- Nuclear and particle physics
- Text mining
- Covid-19

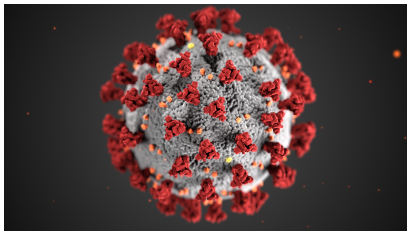


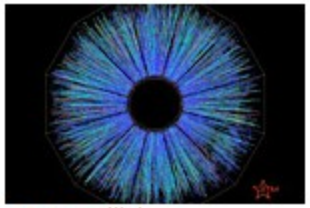
Image Credit
https://www.cdc.gov/

- And more…

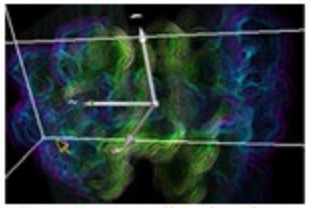
ATLAS Detector
Copyright CERN
Permission Information

SDSS Telescope
Image Credit Fermilab
Permission Information

CDMS photo
Image Credit Fermilab
Permission Information



STAR Collision
Image Credit Brookhaven National Laboratory/STAR Collaboration
Permission Information



BioMOCA Application in nanoHUB
Image Credit Shawn Rice, Purdue University
Permission Information

CMS Detector
Copyright CERN
Permission Information

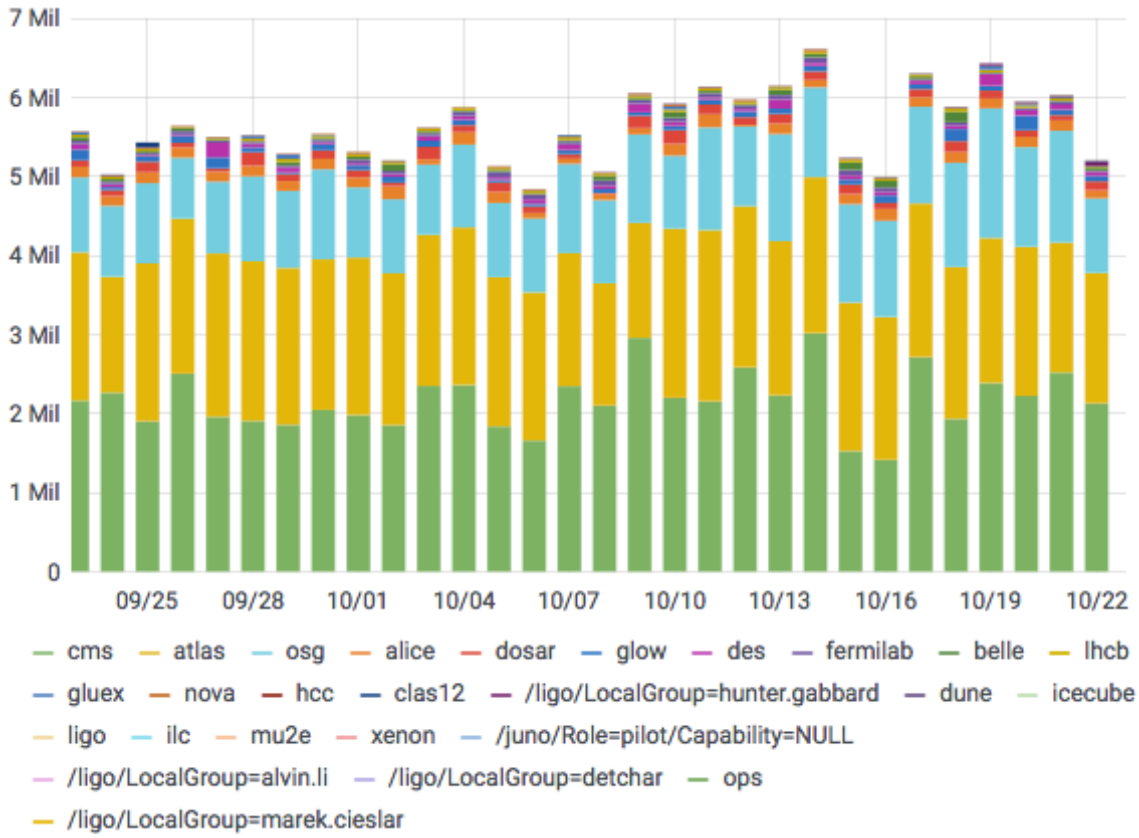
Auger photo
Image Credit Pierre Auger Observatory
Permission Information

MiniBooNE photo
Image Credit Fermilab
Permission Information

DZero Detector
Image Credit Fermilab
Permission Information

OSG Usage

CPU Hours by VO in past 30 days



- VO = Virtual Organization
- Most OSG use is on ***dedicated resources*** (used by resource owners) – 'atlas', 'cms'
- About ***15% is opportunistic*** use – 'osg', 'hcc', 'glow'

OSG Jobs

- **High Throughput Computing**
  - Sustained computing over long periods of time. Usually serial codes, or low number of cores threaded/MPI.

  vs. High Performance Computing
  - Great performance over relative short periods of time. Large scale MPI.

- ***Distributed* HTC**
  - No shared file system
  - Users ship input files and (some) software packages with their jobs.

- **Opportunistic Use**
  - Applications (esp. with long run times) can be *preempted* (or killed) by resource owner's jobs.
  - Applications should be relatively short or support being restarted.

OSG Jobs

- **High Throughput Computing**
  - Sustained computing over long periods of time. Usually serial codes, or low number of cores treaded/MPI.

  vs. High Performance Computing
  - Great performance over relative short periods of time. Large scale MPI.

- ***Distributed* HTC**
  - No shared file system
  - Users ship input files and (some) software packages with their jobs.

- **Opportunistic Use**
  - Applications (esp. with long run times) can be *preempted* (or killed) by resource owner's jobs.
  - Applications should be relatively short or support being restarted.

OSG Jobs

- **High Throughput Computing**
  - Sustained computing over long periods of time. Usually serial codes, or low number of cores treaded/MPI.

  vs. High Performance Computing
  - Great performance over relative short periods of time. Large scale MPI.

- ***Distributed* HTC**
  - No shared file system
  - Users ship input files and (some) software packages with their jobs.

- **Opportunistic Use**
  - Applications (esp. with long run times) can be *preempted* (or killed) by resource owner's jobs.
  - Applications should be relatively short or support being restarted.

OSG Jobs

- Run-time: 1-12 hours
- Single-core
- Require <2 GB Ram
- Statically compiled executables (transferred with jobs)
- Non-proprietary software

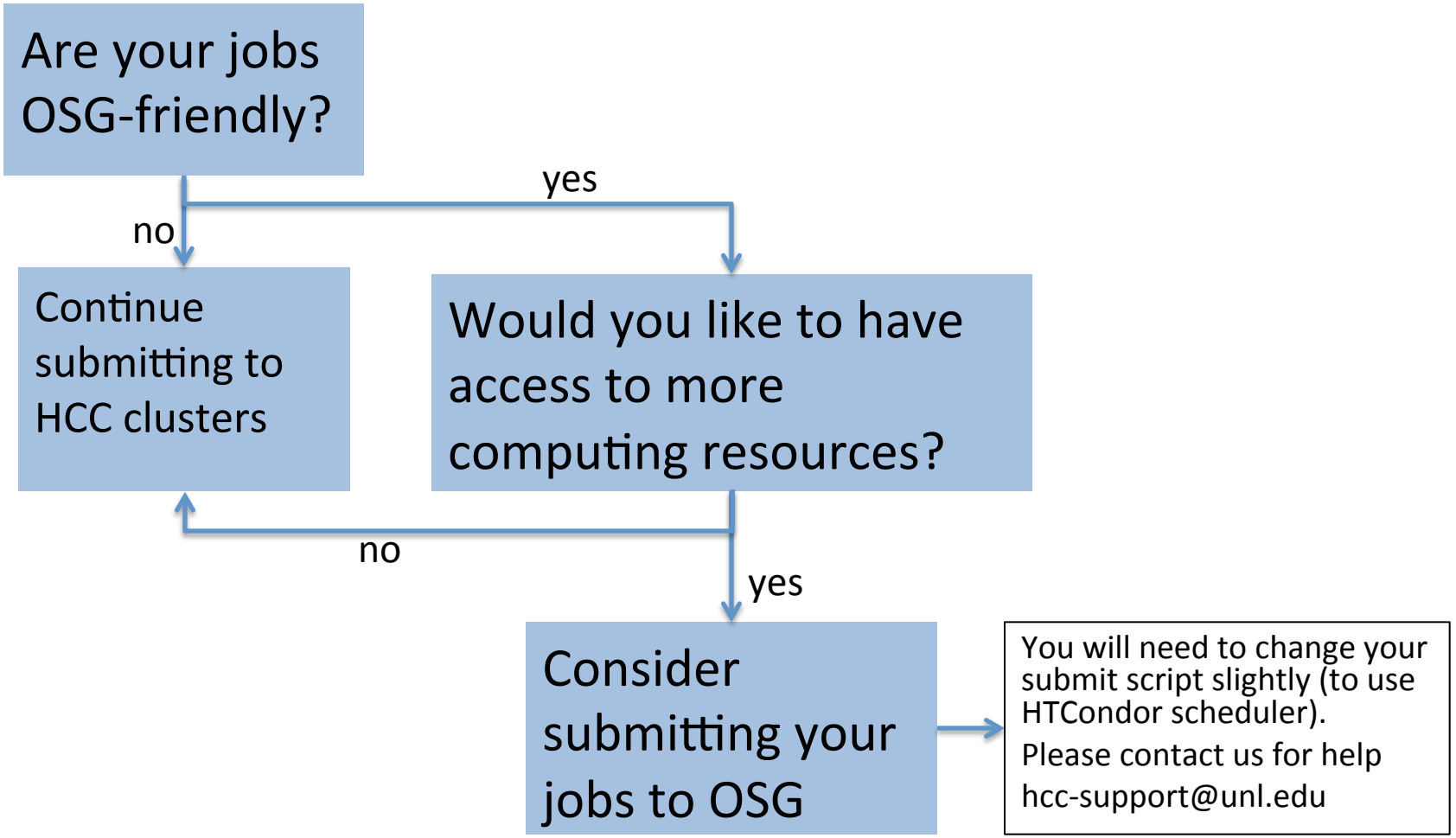
OSG Jobs

- Run-time: 1-12 hours
- Single-core
- Require <2 GB Ram
- Statically compiled executables (transferred with jobs)
- Non-proprietary software

These are not hard limits!

- Checkpointing – for long jobs that are preempted
  - Many applications support built-in checkpointing
  - Job image is saved periodically so that it can be restarted on a new host after it is killed (without losing the progress that was made on the first host)
- Limited support for larger memory jobs
- "Partitionable" slots – for parallel applications using up to 8 cores
- Modules available – a collection of pre-installed software packages
- Can run compiled Matlab executables

Is OSG right for me?



For more information on the Open Science Grid:

https://www.opensciencegrid.org/

For instructions on submitting jobs to OSG:

https://go.unl.edu/hcc-osgsubmit