



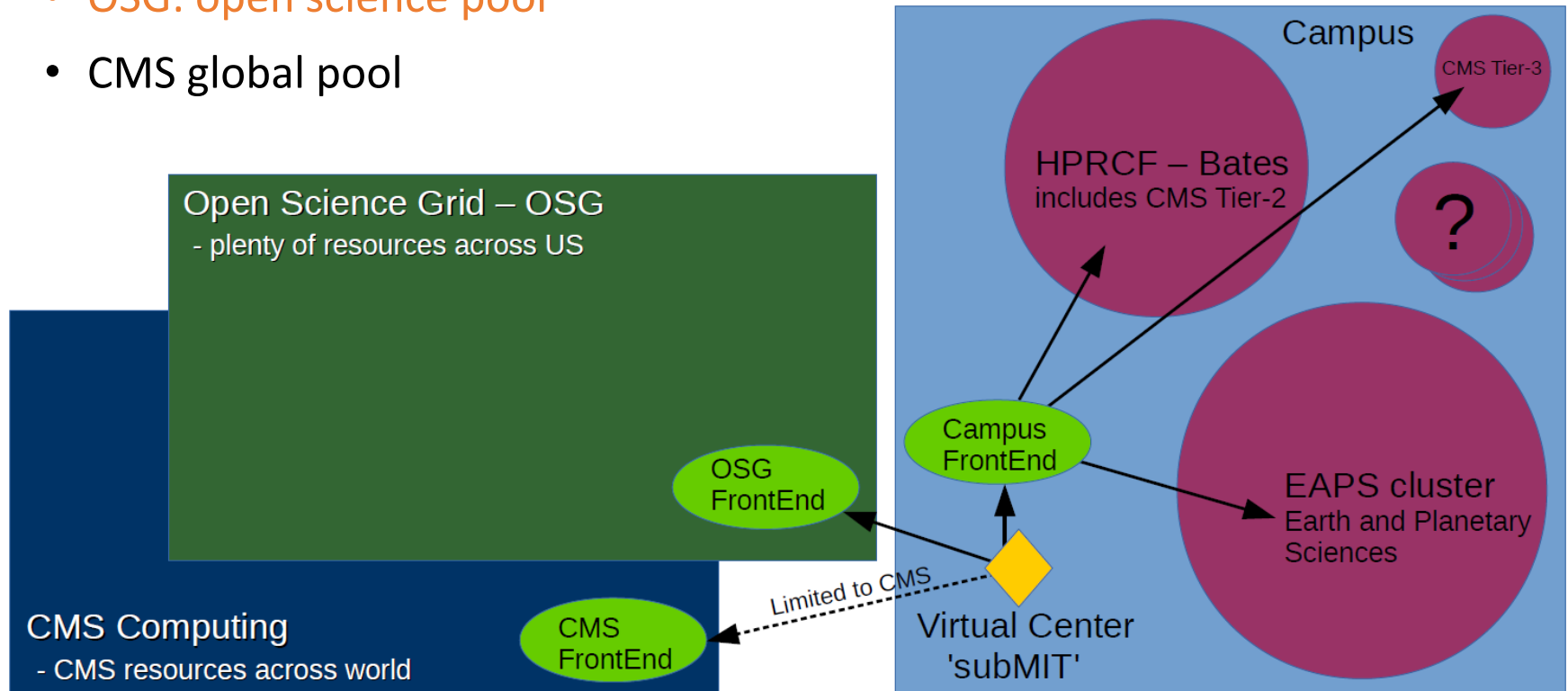
The Open Science Grid and EAPS

Zhangqier Wang
subMIT Workshop
Jan. 26th 2022



Condor Resources

- The submit04.mit.edu is the condor central machine, which connects to different external resources
- Campus FrontEnd: BOSCO
 - Tier2, Tier3, **EAPS Clusters**
- **OSG: open science pool**
- CMS global pool





OSG: Open Science Pool

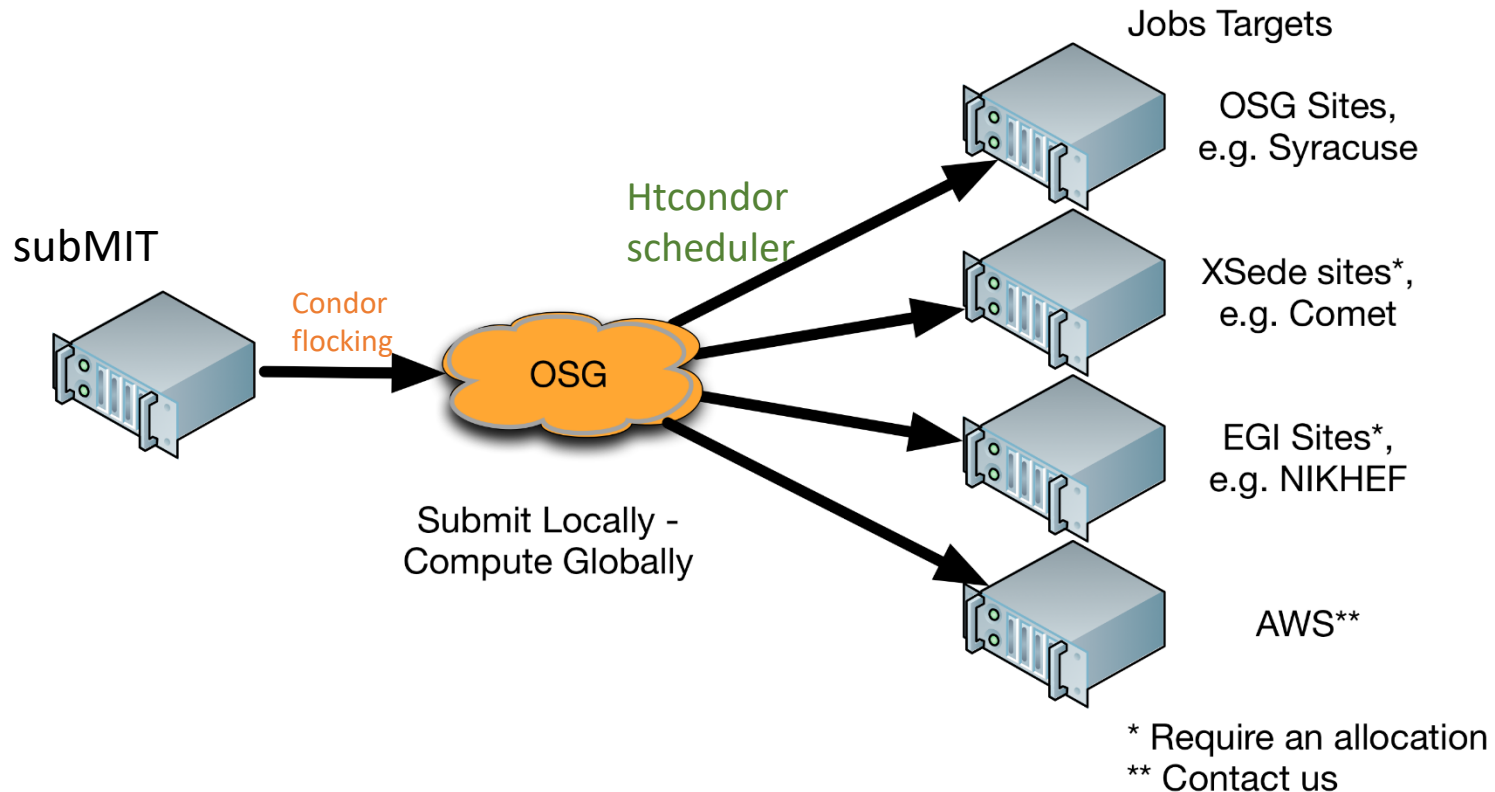


- The Open Science Pool (OSPool) is a virtual cluster operated by the OSG, with shared computing and data resources using distributed high-throughput computing (dHTC).
- The Open Science Pool includes capacity contributed by dozens of campus, national labs, and non-profit organizations
- Researchers can submit computational work to the OSPool via access points operated locally to their campuses, or via access points operated as part of the [OSG Connect](#) service.
 - SubMIT is one of the access points
 - Access is free and open to all subMIT users
 - OSPool opportunistic access
 - i.e., with lower priority than the jobs which use the clusters as dedicated resources.



Open Science Grid

- SubMIT condor central machine submit04.mit.edu connects to OSPool through condor flocking with IDTOKEN authentication.





OSPool Resources



- The computing resources for the OSPool come from more than 120 institutions
- The OSPool delivered more than 280 million compute hours in 2021, individual users achieving more than 30 million hours without fee or allocation requirement
- On average, there are more than 40,000 total cores available across the pool, with peaks of more than 70,000 cores, and with hundreds of [GPUs available](#).
- Nodes from each contributing cluster may differ in CPU and/or GPU models, number of cores, RAM, etc.
 - Each job requests the computing resources it needs (cores, memory, disk, GPU features, etc.)



Open Science Grid



Jobs Go to OSPool



- Ideal size of computations for the OSPool
 - individual jobs that complete **within 20 hours** on one or few cores
 - software that can be distributed across each job in the form of **static binaries, self-contained installations, or containers**
 - **less than 20GB** of input and output, per-job

	Ideal Jobs!	Still very advantageous	Maybe not, but get in touch!
Expected Throughput, per user	1000s concurrent cores	100s concurrent cores	Let's discuss!
CPU	1	< 8	> 8 (or MPI)
Walltime	< 10 hrs*	< 20 hrs*	> 20 hrs
RAM	< few GB	< 40 GB	> 40 GB
Input	< 500 MB	< 10 GB	> 10 GB**
Output	< 1 GB	< 10 GB	> 10 GB**
Software	pre-compiled binaries, containers	Most other than ->	Licensed Software, non-Linux



Submit Jobs to OSPool



- Submit condor jobs to OSPool:
 - Special requirement in the condor script example:
 - `Requirements = (OSGVO_OS_STRING == "RHEL 7"`
`&& BOSCOCluster != "t3serv008.mit.edu"`
`&& BOSCOCluster != "ce03.cmsaf.mit.edu"`
`&& BOSCOCluster != "eofe8.mit.edu")`
 - `+ProjectName = "MIT_submit"`
- The additional requirement can be found in [OSG support web](#)
- A few common requirements:
 - **CVMFS:** `CVMFS_oasis_opensciencegrid_org_REVISION >= 3983`
 - **Singularity:** `HAS_SINGULARITY`
 - **GPU jobs:** `CUDACapability`
- Using Software on the OSG
 - Containers, software compilation, ...
 - Details: <https://support.opensciencegrid.org/support/solutions/folders/5000260921>

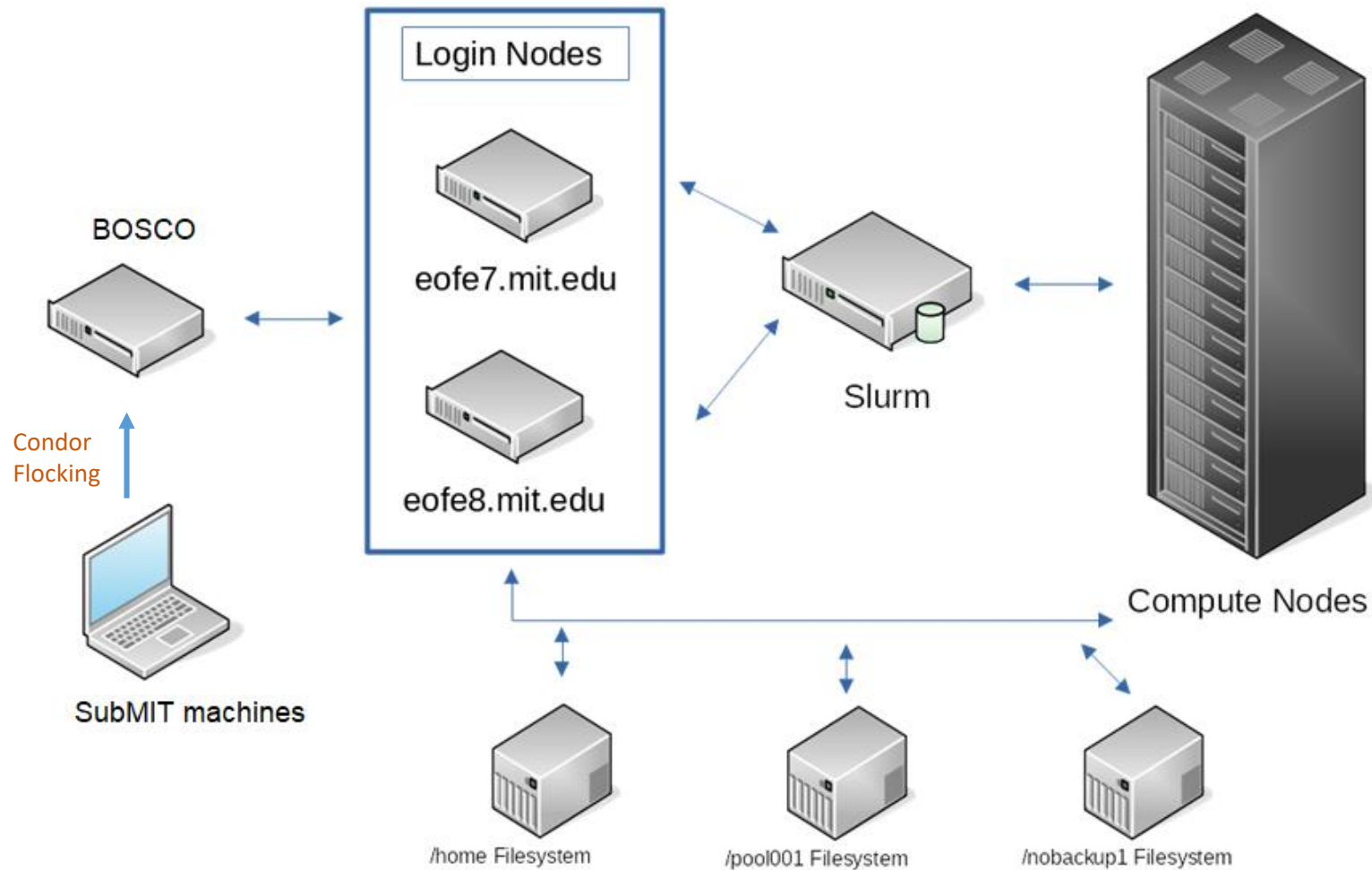


EAPS Clusters



- From Earth, Atmospheric and Planetary Sciences
- The [Engaging Cluster](#) is a shared high performance compute cluster used by MIT and its affiliates for research and computing purposes
 - EAPS “central machine”: eofe8.mit.edu
 - Scheduler: Slurm
- Submit condor jobs to EAPS:
 - Special requirement in the condor script:
 - **Requirements = (BOSCOCluster == "eofe8.mit.edu") &&.....**
- **Work in Progress:**
 - jobs at EAPS are not having priority, which may be rescheduled when higher priority jobs come in. It might take much longer than expected







EAPS Software



- You can run ordinary jobs on EAPS.

Also, you can do:

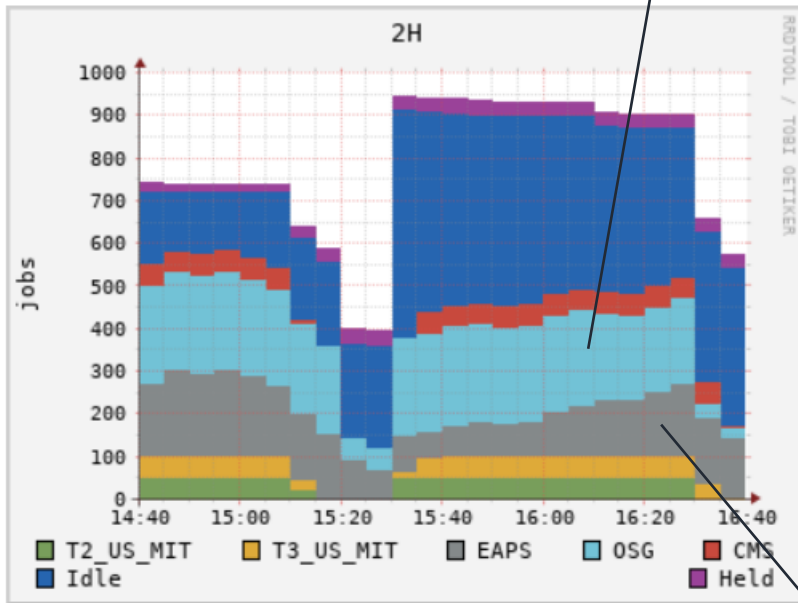
- Loading software modules
 - There is a wide selection of software installed on the engaging cluster that are available to be used by users. Software installed across the cluster are available as [environmental modules](#).
- These modules enable the cluster to have multiple versions of the same software available to users.
 - The available environmental modules with: **module avail**
- Example
 - **Installing Python Packages Locally with Pip**
 - loaded a Python module with the command:
 - **module load python/[version]**
 - install python packages with the command:
 - **pip install [package_name] --user**



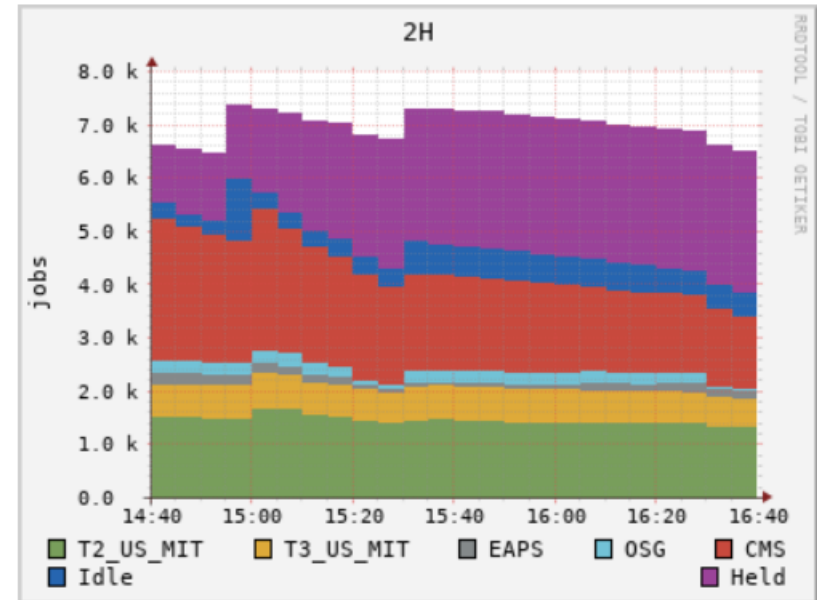
Jobs in Monitoring

Monitoring: <http://submit04.mit.edu/condormon/>

wangzqe



Total



EAPS

The End