HTCondor @ ORNL ALICE T2

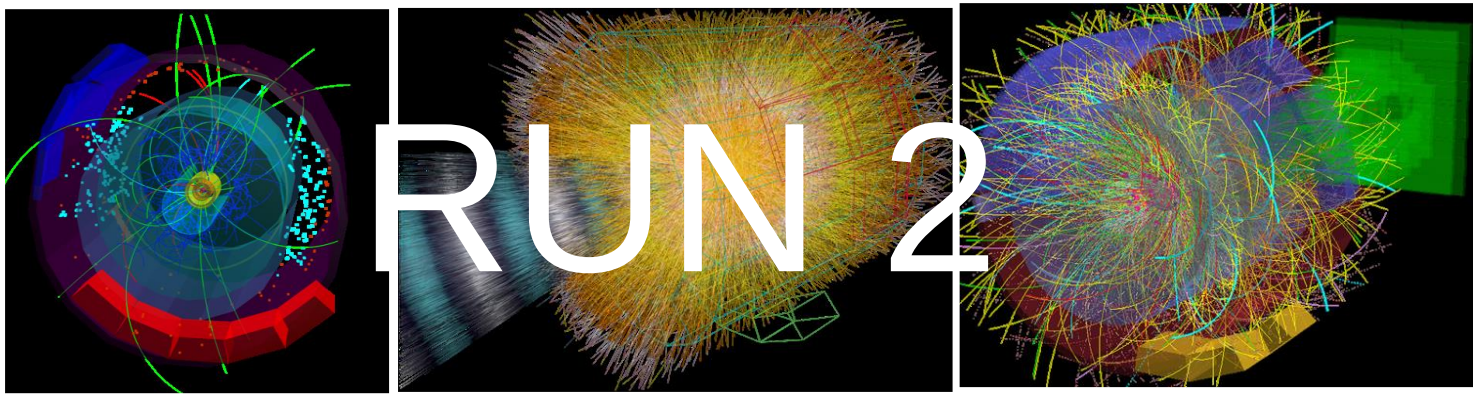
HTCondor Workshop, Autumn 2024, Amsterdam

September 27, 2024

Irakli Chakaberia

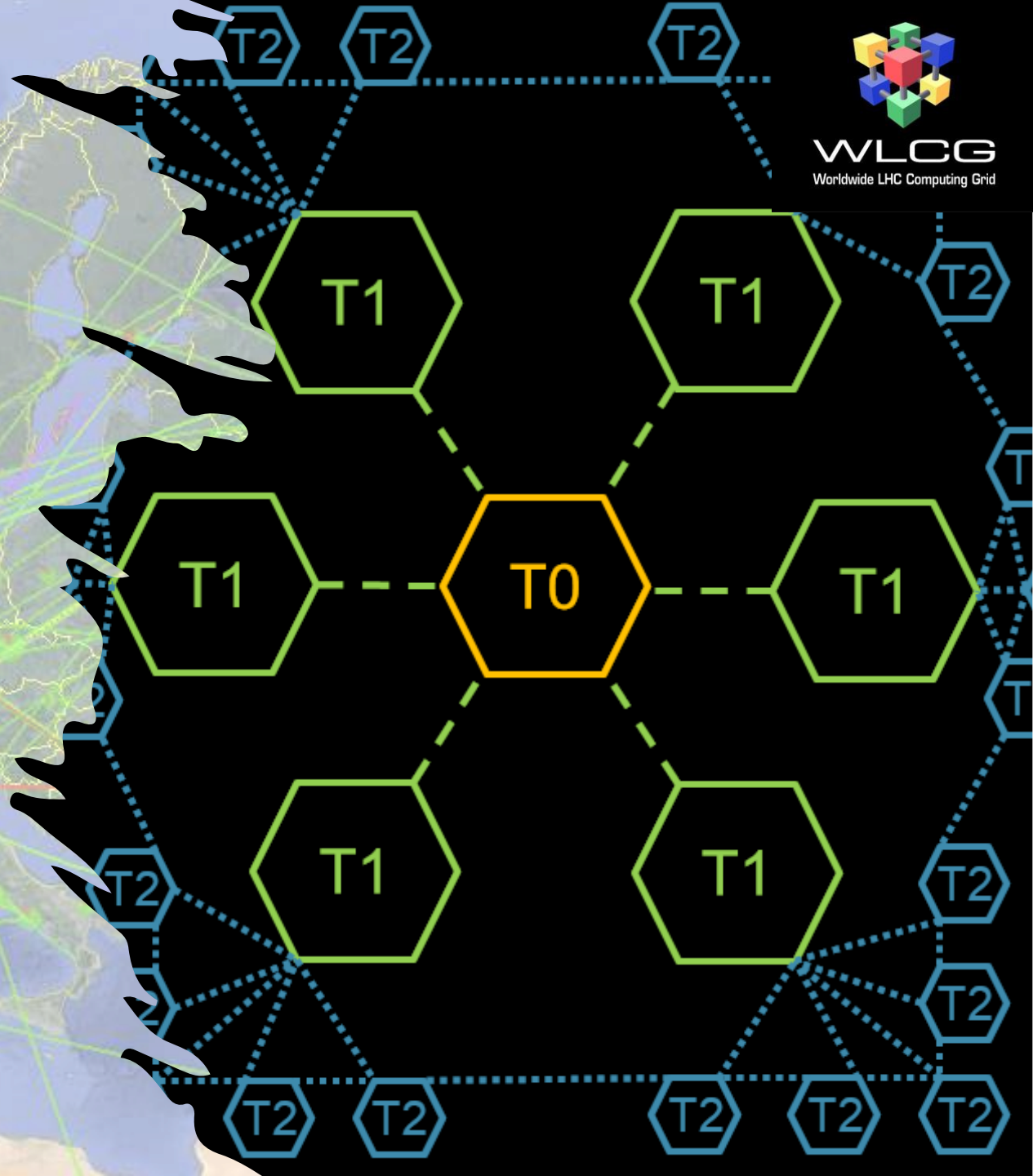
ALICE Experiment @ LHC – Computing POV

- Experiment produces immense amount of data that requires a lot of computing resources
- In LHC Run3 ALICE collects O(100 PB) of data a year
- To sustain computing such computing needs ALICE adopted a grid computing philosophy
- It is part of the Worldwide LHC Computing Grid infrastructure



WLCG Computing infrastructure

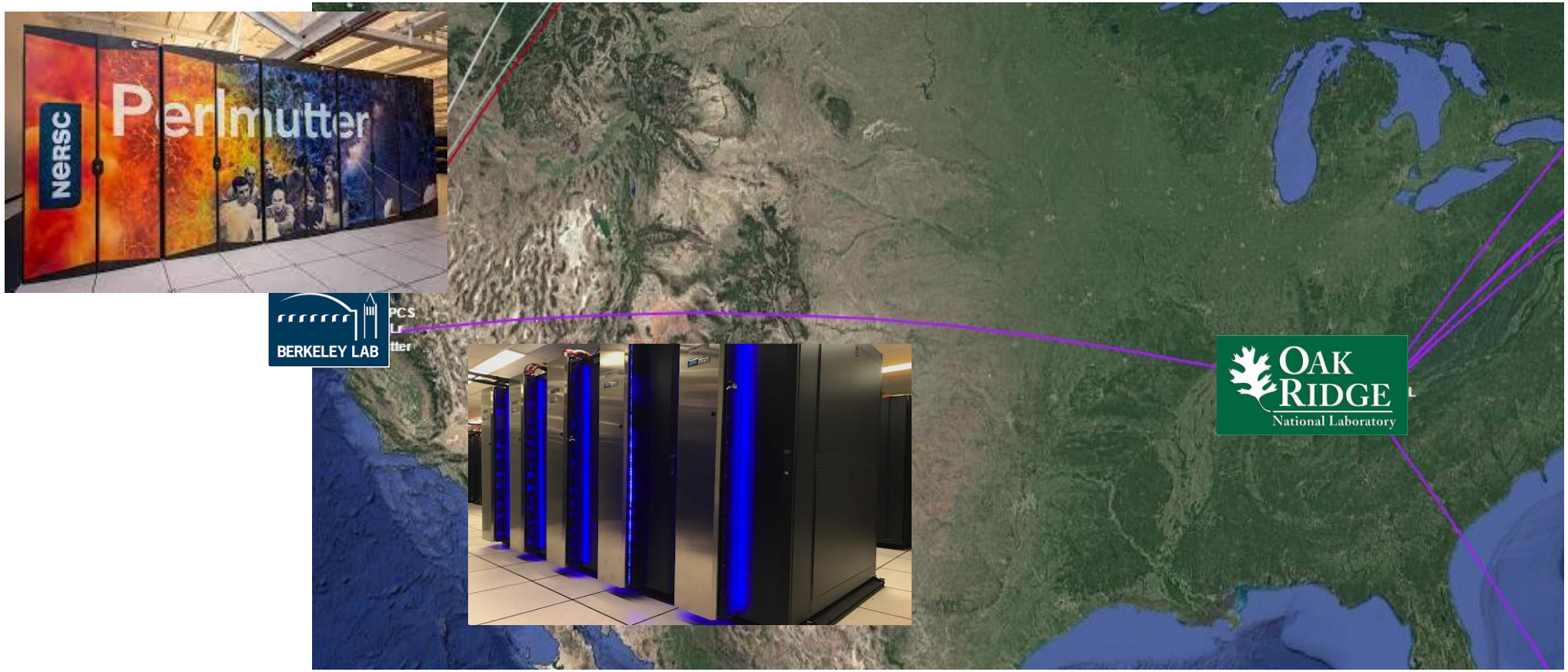
- The Worldwide LHC Computing Grid (WLCG) project is a global collaboration
- The mission of the WLCG project is to provide global computing resources to store, distribute and analyse the ~200 Petabytes of data expected every year of operations from the Large Hadron Collider (LHC) at CERN
- It operates around 170 computing centers in more than 40 countries
- Globally distributed system of computing centers:
  - Configured in a tiered architecture that functions as a single coherent system
  - Tier 0 – Tier 1 – Tier 2 – Tier 3
  - Each center provides Grid-enabled gateways to CPU and storage
  - some of the centers also provide Analysis Facilities
  - Extensive high-quality network allows for communication among all computing centers



ALICE Computing GRID

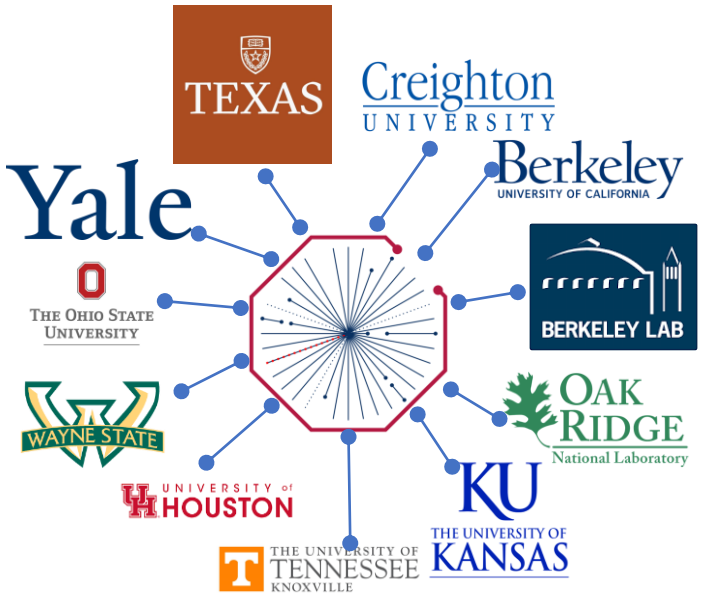
ALICE-USA T2 Sites & Analysis Facility (Prototype)

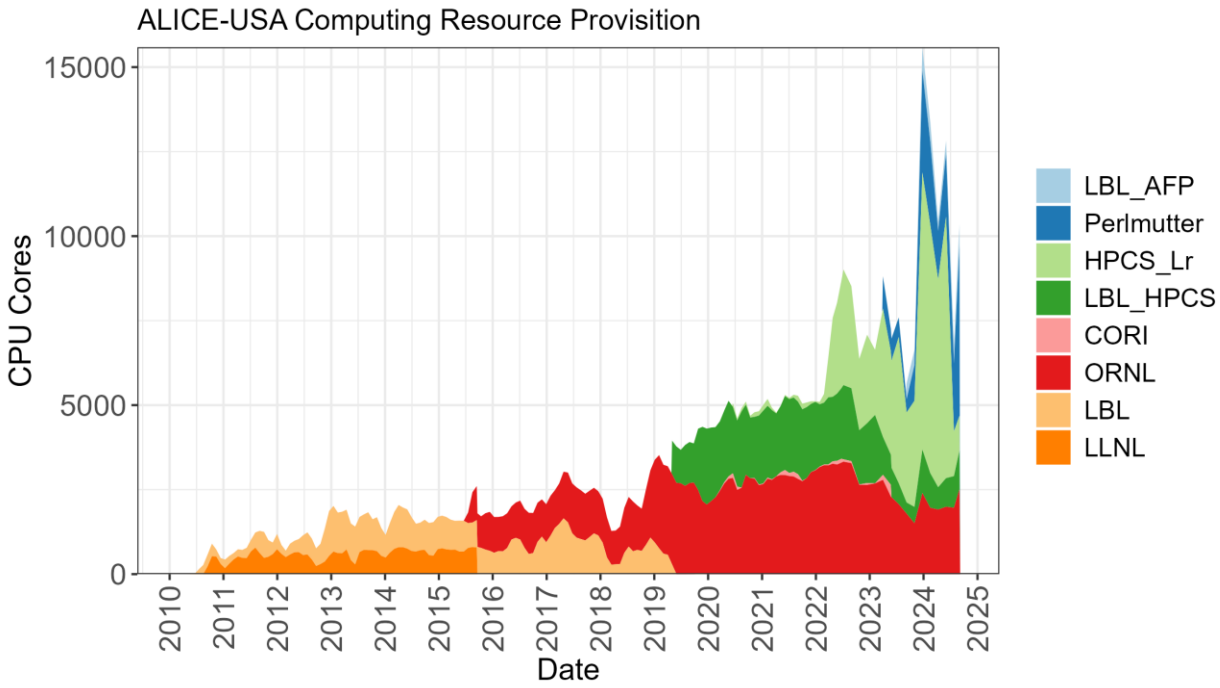
- Project currently operates two T2 sites at ORNL and LBNL and an AF
- In addition, we provide resources on Lawrencium (opportunistic) and Perlmutter HPCs



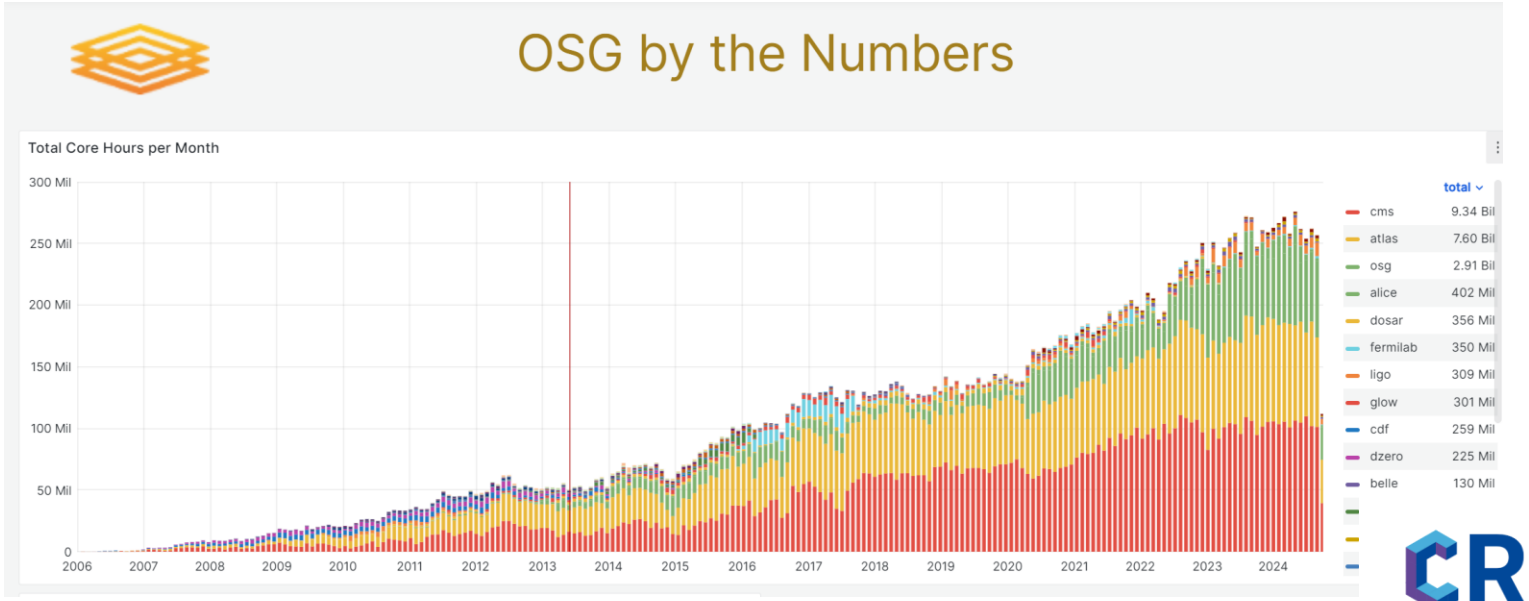
ALICE-USA Computing Project

- ALICE-USA provide about about 9% of total CPU and storage resources on behalf of the US collaborators of the ALICE experiment
- Resources provided by each participating entity (country/research establishment) need to go through a careful accounting process
- In case of ALICE-USA accounting is via OSG to WLCG





Accounting or GRACC and CRIC



CRIC
Computing Resource Information Catalog

WLCG CRIC

Site & Service management

- Create Resource Center Site
- Create Experiment Site
- Create Storage Service
- Create Compute element
- Create queue
- Create protocol
- Validate monthly accounting data
- Site Network topology

Federation management

- Create Federation
- Create Federation pledge
- Create VO requirements

WLCG Operations management

- Generate T1 accounting report
- Generate T2 accounting report
- Generate Country report

How to

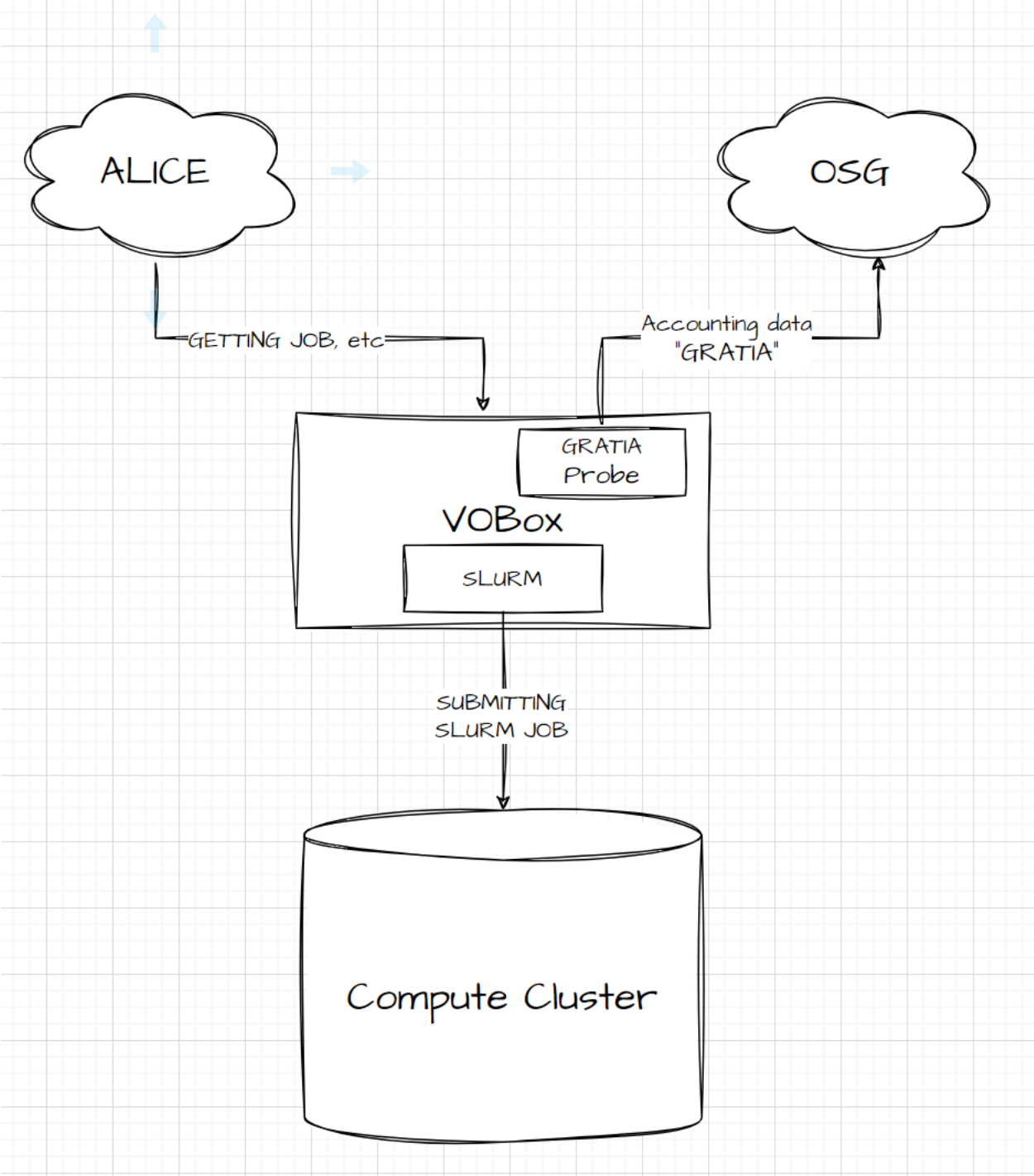
- How to get CRIC access and privileges
- How to manage pledge info in CRIC
- How to manage storage info in CRIC
- CRIC main concepts and data models

Useful links

- CRIC Instances: ATLAS | CMS
- WLCG Home page
- WLCG Dashboards
- WLCG Accounting Utility (WAU)
- EGI Accounting Portal
- GocDB
- GGUS
- VOFeeds: ALICE | ATLAS | CMS | LHCB

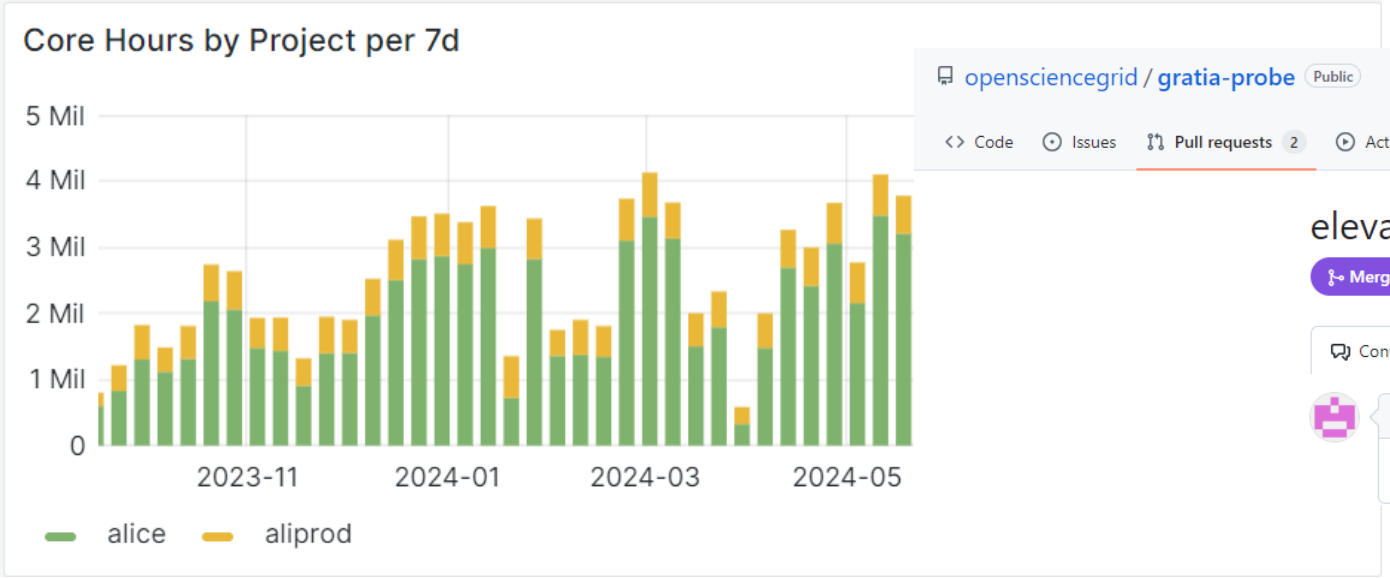
Old setup

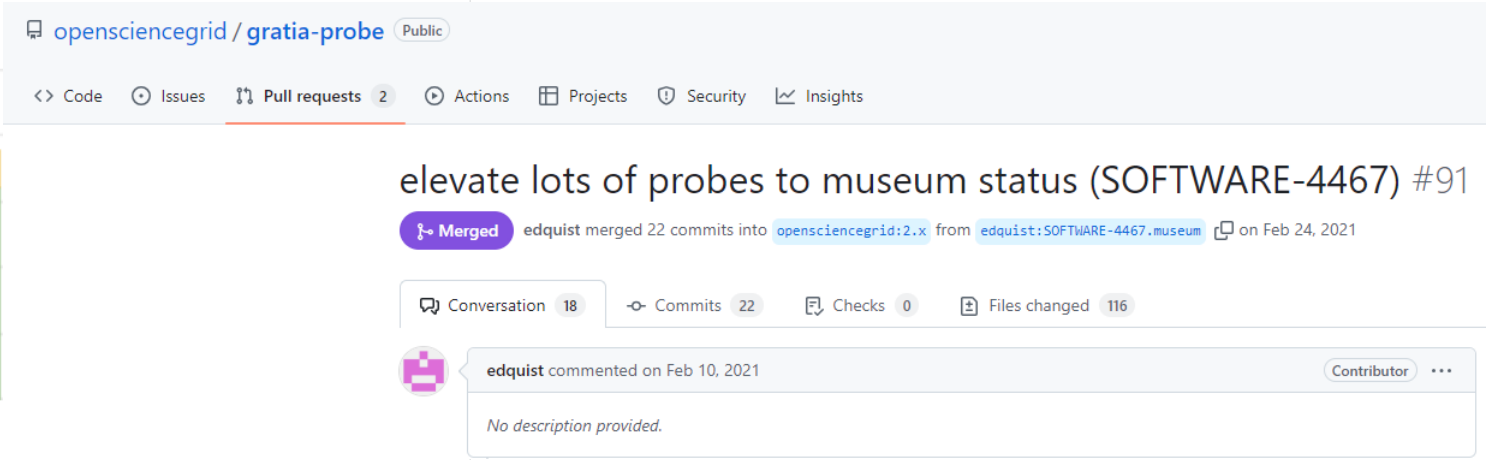
- We ran our T2 using the gratia probe with SLURM for a long time
- Sometimes the probe would get stuck or go down and needed to be restarted and timestamp reset
- But even that was rare and just worked



WLCG/OSG Accounting

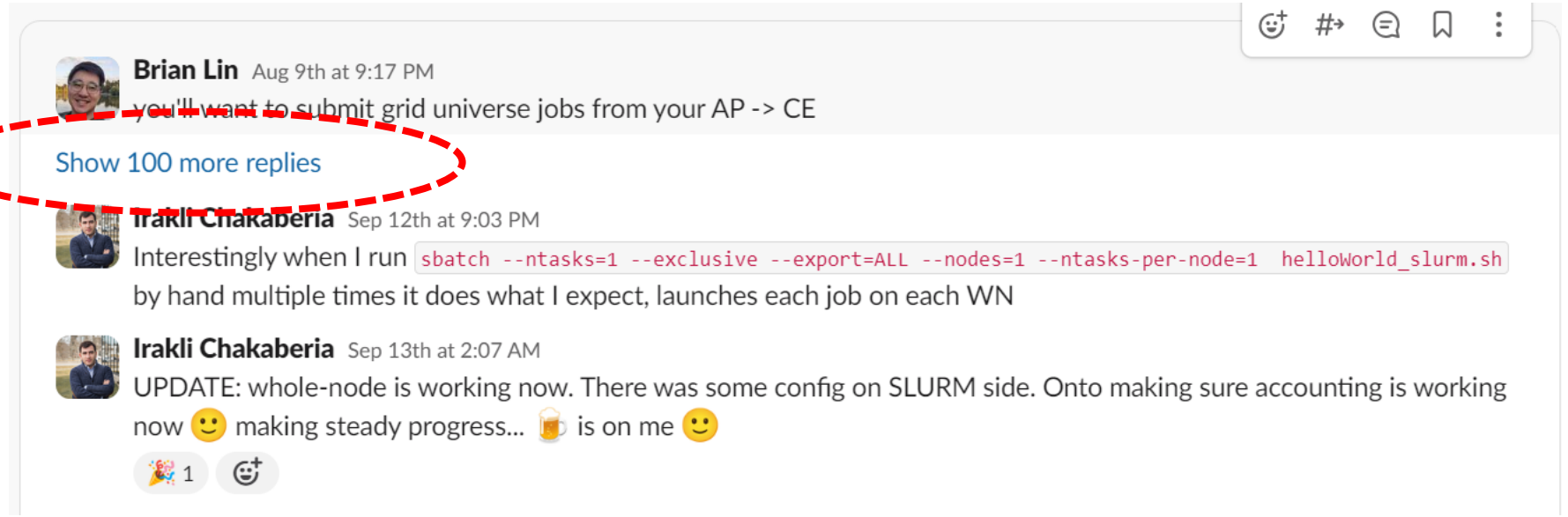
- Reporting/accounting to WLCG is done via OSG
- OSG “gratia” tool collects batch system info and reports to GRACC
- At some point OSG stopped maintaining separate dedicated probes for individual batch systems
- We kept old “gratia” going but OS update necessitated the need for update
- So, we had to move to HTCondor CE





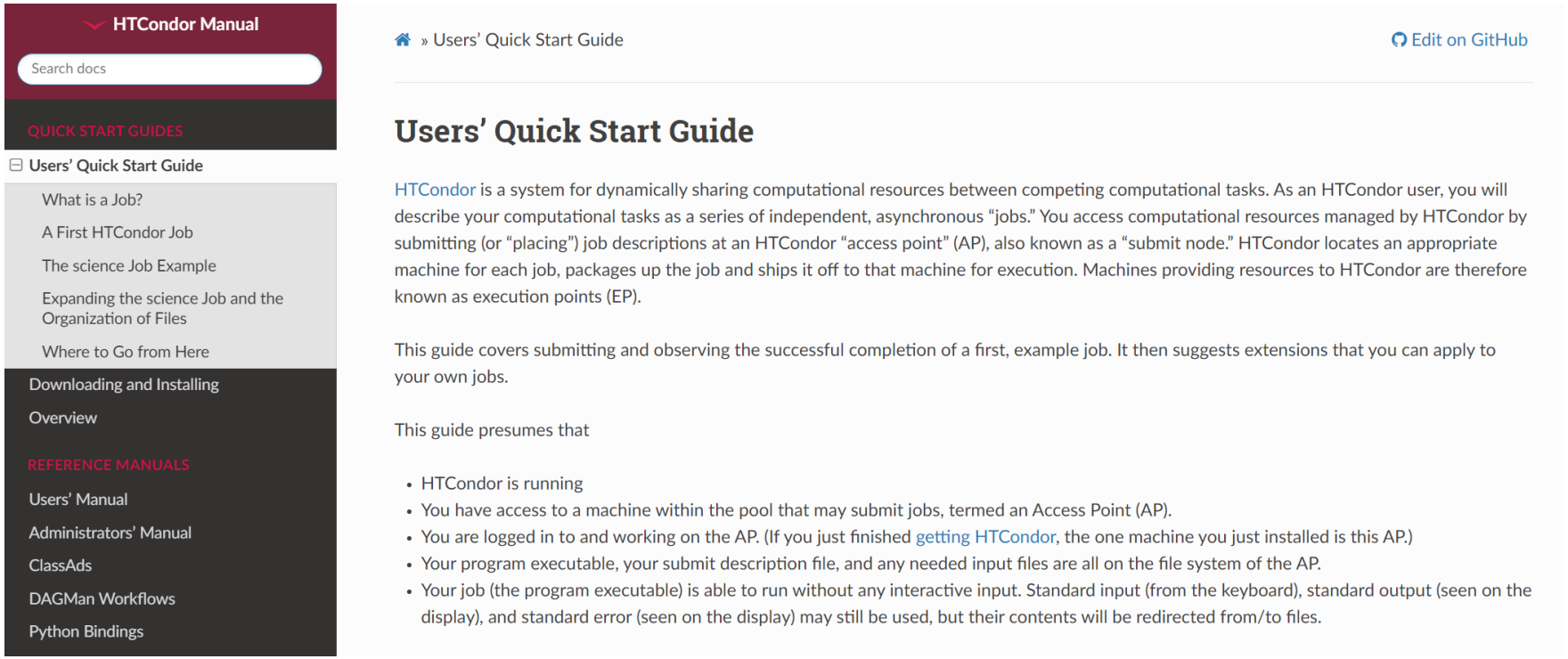
Huge gratitude to…

- OSG/HTCondor Support
- Best support for the community by far…
- In particular
  - Brian Lin and Derek Weitzel



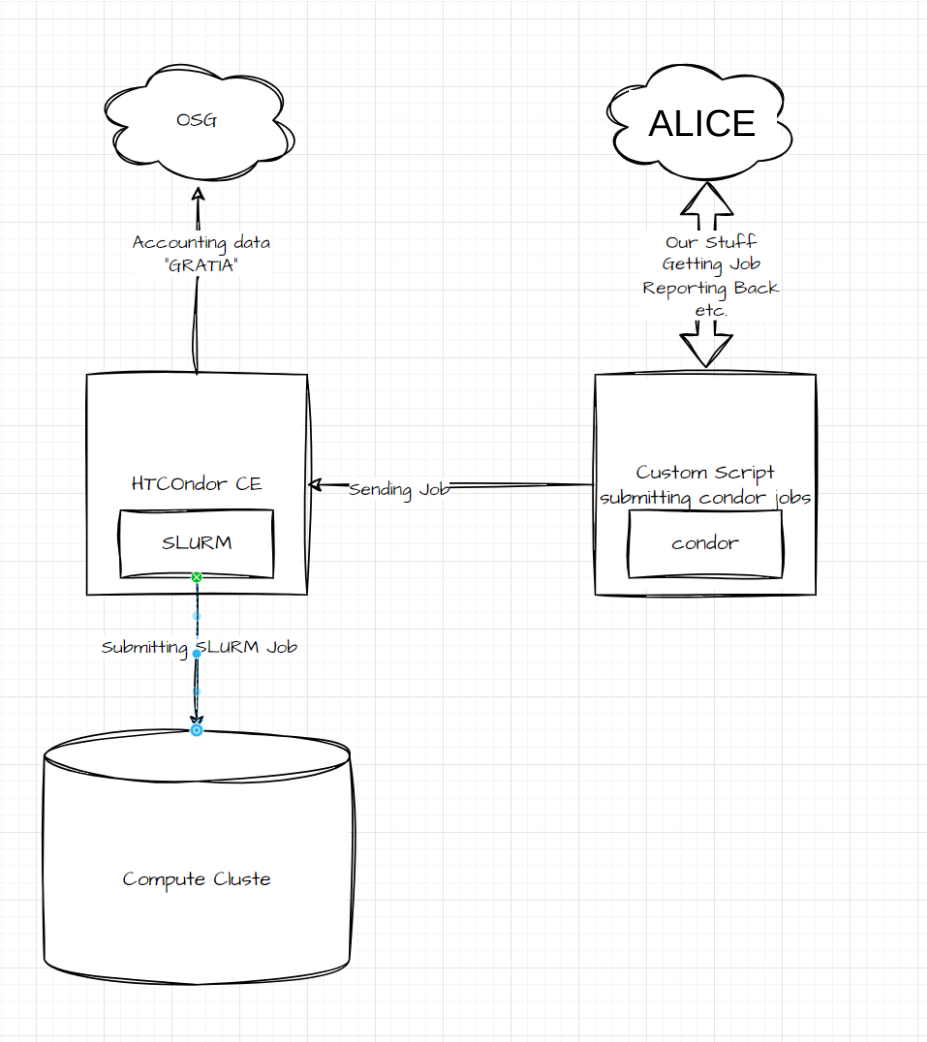
Documentation is good too

- QuickStart Guide is there to…
- …quickly show you that you need to read the friendly manual

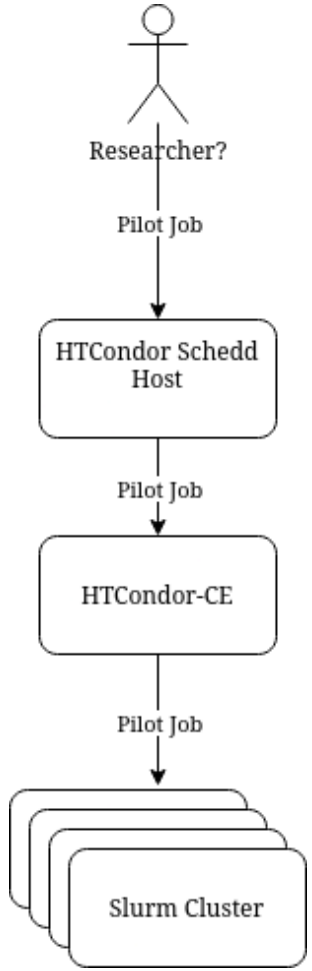


Current Setup

How I see it



How Brian Sees it



Certificates

- I was naive: if I only allow my machine to submit to my cluster via explicit definition of the submitting server I should not need to deal with certificates
- In /etc/condor-ce/config.d/10-osg-attributes-generated.conf - `AllowedVOs = { "vobox.alice.ornl.gov" };`

```
universe = grid
grid_resource = condor slurm9r.alice.ornl.gov
slurm9r.alice.ornl.gov:9619

executable = helloWorld.sh

Log        = helloWorld.$(ClusterId).log
Output     = helloWorld.$(ClusterId).out
Error      = helloWorld.$(ClusterId).error

ShouldTransferFiles = YES
WhenToTransferOutput = ON_EXIT

use_scitokens = true
scitokens_file = .globus/wlcg.dat
queue 1
```

- “grid” mandates otherwise!
- So, I had to figure that part out
- One confusion had to do with the type of certificate
- It seemed that if one deals with the WLCG in some way or form one will need

  InCommon RSA IGTF Server CA 3

- However, we ended up making all work using LE certificate

Configuration

- This slide would have been much busier were showing it a couple of weeks ago… the wounds have healed by now ☺
- From today's point of view, I can see how documentation has the information (looking forward to AI plugin to digest it for me)

On the CE/AP

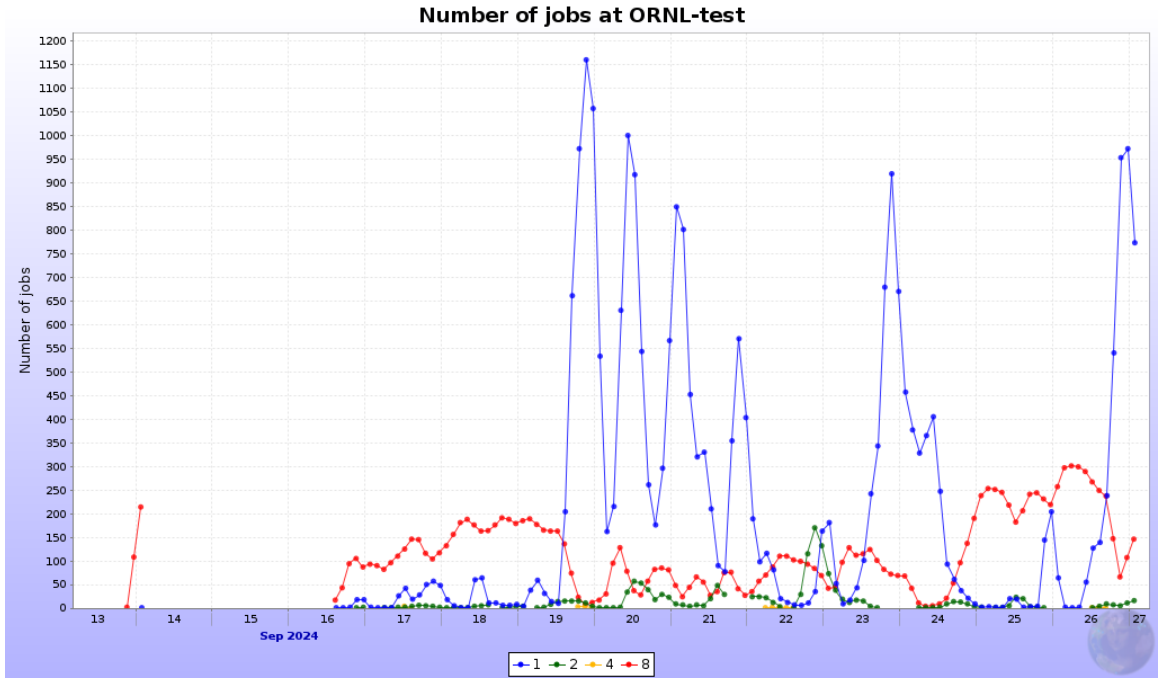
```
/etc/condor-ce/config.d/01-ce-auth.conf
/etc/condor-ce/config.d/99-temporary-debugging.conf
/usr/share/condor-ce/verify_ce_config.py
/etc/osg/config.d/20-slurm.ini
/etc/blah.config
/etc/osg/config.d/40-siteinfo.ini
/etc/osg/config.d/31-cluster.ini
/etc/osg/config.d/31-cluster.ini::q!
/usr/share/condor-ce/config.d/02-ce-slurm-
defaults.conf
/etc/osg/config.d/31-cluster.ini
/etc/condor-ce/config.d/10-osg-attributes-
generated.conf
```

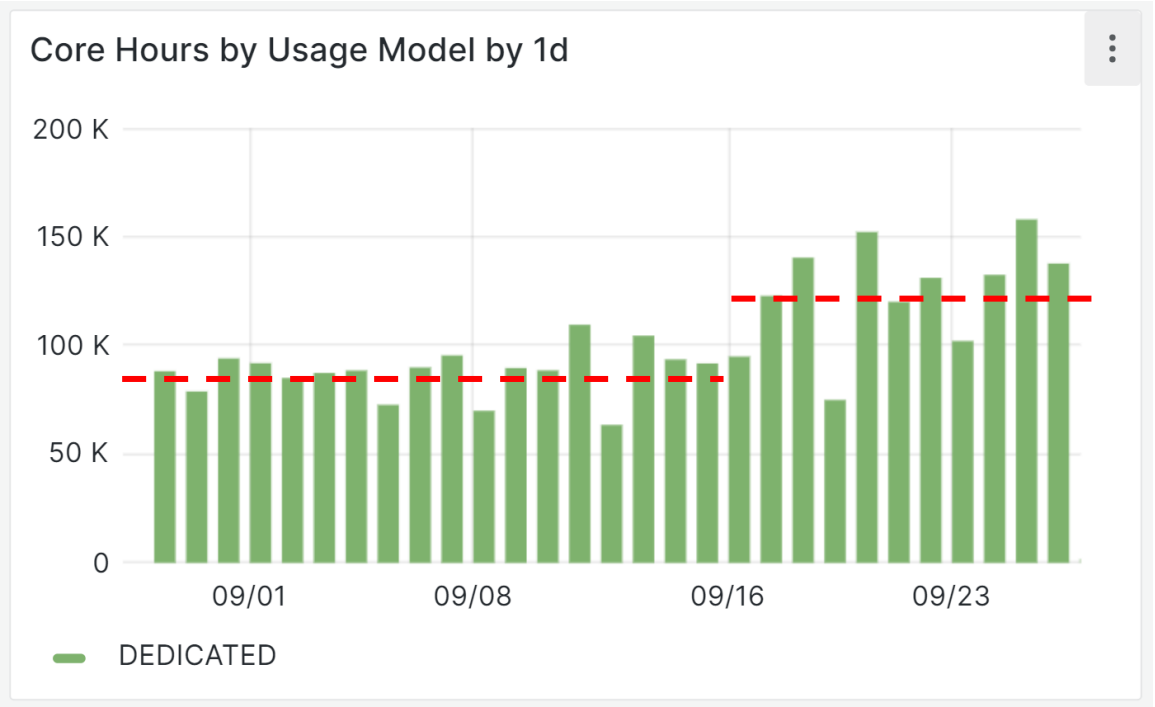
On the VOBox

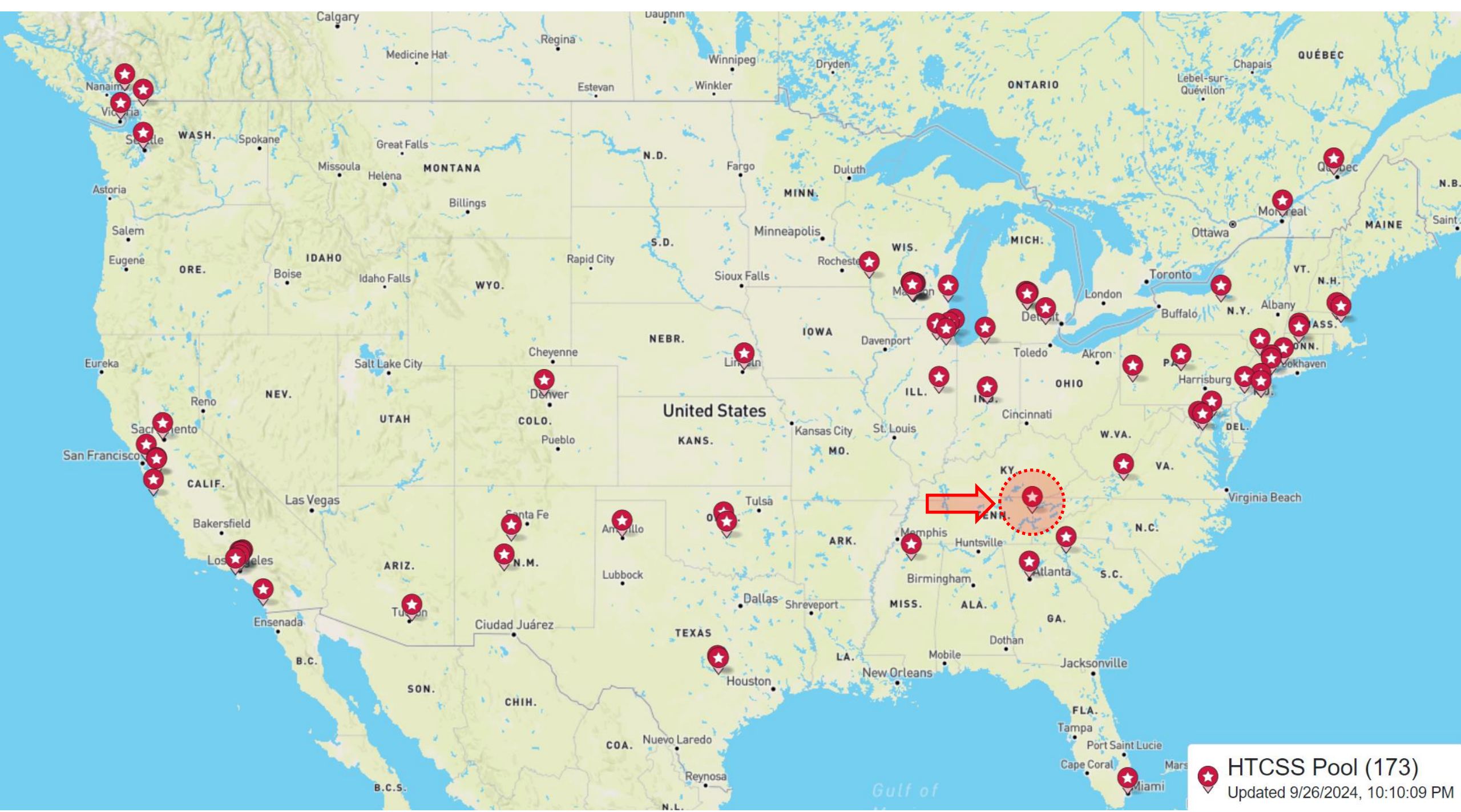
```
/etc/ssh/sshd_config
/etc/condor/config.d/99-ssl.conf
/etc/condor/config.d/99-ip.conf
```

HTCondor @ ORNL

- It took a few weeks of configuring, debugging, reconfiguring, re-debugging, understanding, fixing, coffee, bothering HTCondor people, repeating
- But we got there…






HTCSS Pool (173)
Updated 9/26/2024, 10:10:09 PM
United States
Calgary
Regina
Medicine Hat
Winnipeg
Dauphin
Dryden
Winkler
Estevan
ONTARIO
QUÉBEC
Chapais
Lebel-sur-Quévillon
Nanaimo
Victoria
Seattle
WASH.
Spokane
Great Falls
Missoula
Helena
MONTANA
N.D.
Fargo
Duluth
MINN.
Billings
Astoria
Salem
Eugene
ORE.
IDAHO
Boise
Idaho Falls
WYO.
S.D.
Rapid City
Minneapolis
Sioux Falls
WIS.
MICH.
Ottawa
Montreal
Quebec
MAINE
N.B.
Toronto
London
Buffalo
Detroit
N.Y.
Albany
VT.
N.H.
MASS.
CONN.
Davenport
IOWA
NEBR.
Cheyenne
Lincoln
Denver
Salt Lake City
Eureka
Reno
NEV.
UTAH
COLO.
Pueblo
KANS.
Kansas City
St. Louis
MO.
ILL.
IND.
OHIO
Toledo
Akron
Cincinnati
Harrisburg
PA.
DEL.
W.VA.
VA.
KY.
TENN.
N.C.
S.C.
Virginia Beach
Memphis
Huntsville
Atlanta
Birmingham
ARK.
MISS.
ALA.
GA.
Dothan
Jacksonville
Mobile
New Orleans
LA.
FLA.
Tampa
Port Saint Lucie
Cape Coral
Miami
Gulf of
San Francisco
CALIF.
Las Vegas
Bakersfield
Los Angeles
Ensenada
ARIZ.
Tucson
Santa Fe
N.M.
Lubbock
Tulsa
Amarillo
Dallas
Shreveport
Ciudad Juárez
TEXAS
Houston
Nuevo Laredo
Reynosa
B.C.
SON.
CHIH.
COA.
B.C.S.
N.L.

Near Future Plans

- We will need to replicate the setup at our LBNL T2/AF/HPCs
- Attempt a single VM setup @ LBNL
- From yesterday… perhaps try HTCondor for submission to Perlmutter
  - We currently use NERSC's SuperFacility API that allows us to submit directly to SLURM
  - We had to plug this in into our grid submission system
  - But if HTCondor can make this simple… there are advantages to that…

Thank you