

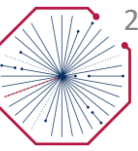
HTCondor @ ORNL ALICE T2

HTCondor Workshop, Autumn 2024, Amsterdam

September 27, 2024

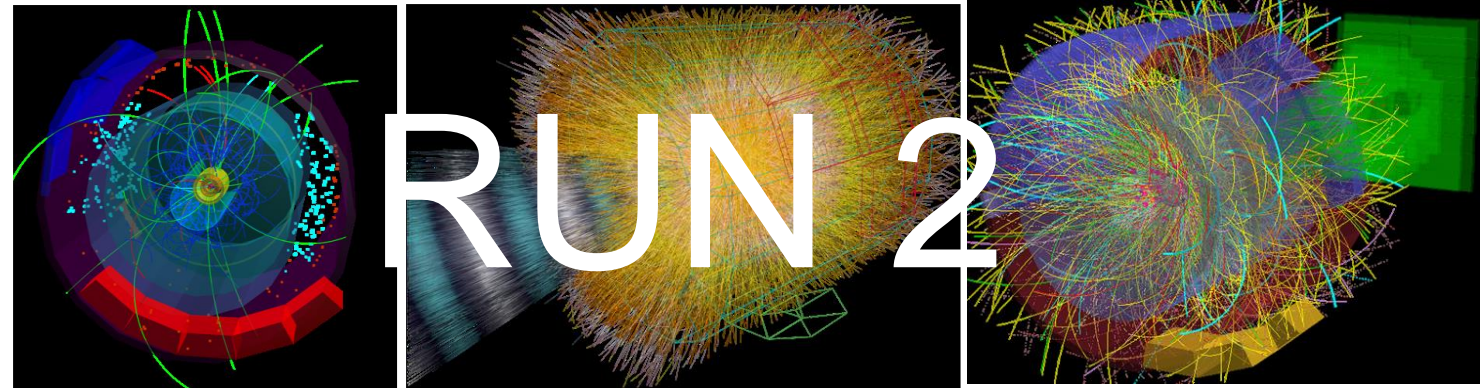
Irakli Chakaberia





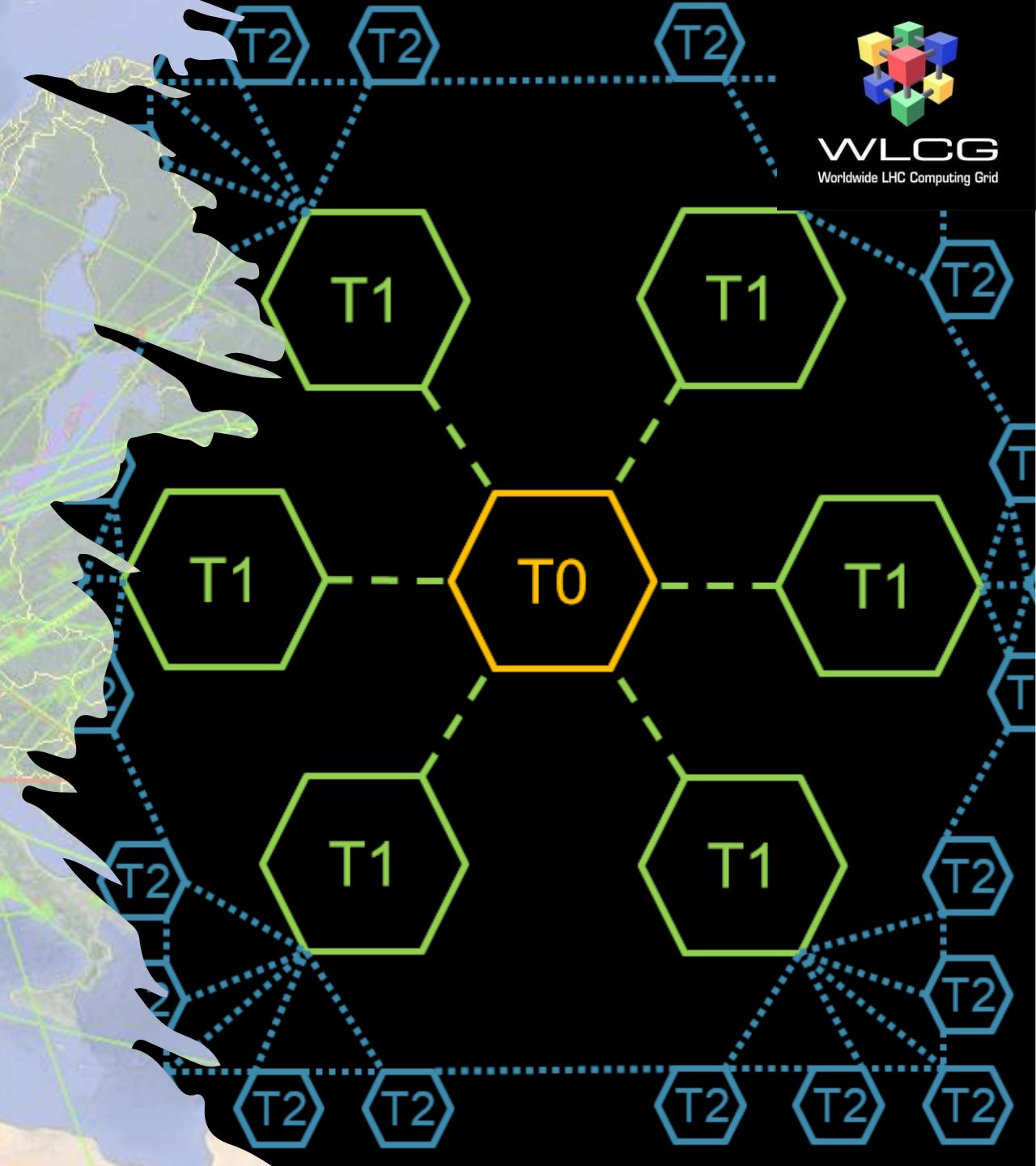
ALICE Experiment @ LHC – Computing POV

- 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



NIKHEF/SARA

US Sites @ LBNL and ORNL

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 Lawrence Livermore (opportunistic) and Perlmutter HPCs

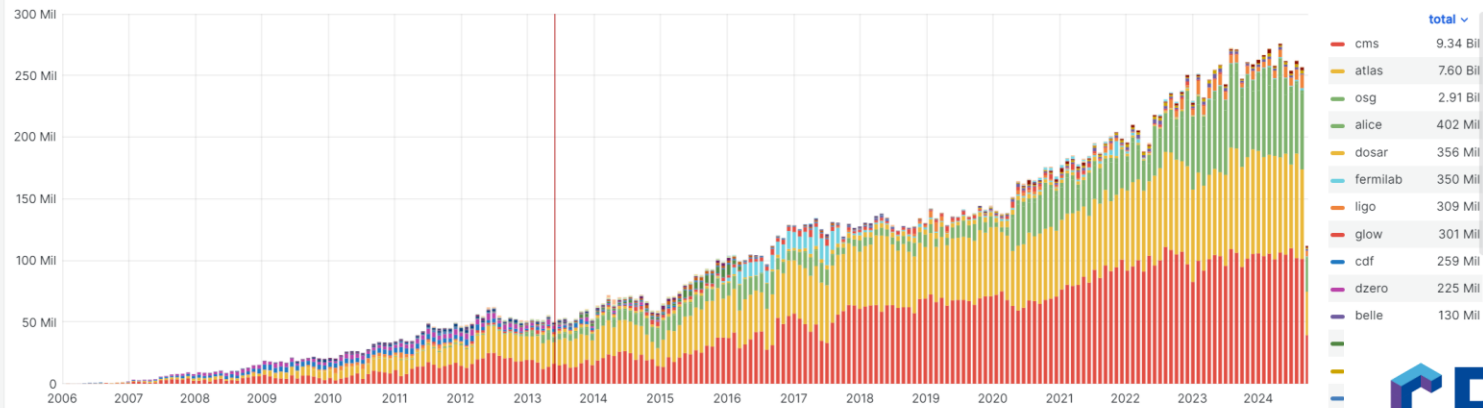


Accounting or GRACC and CRIC



OSG by the Numbers

Total Core Hours per Month



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

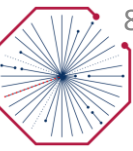
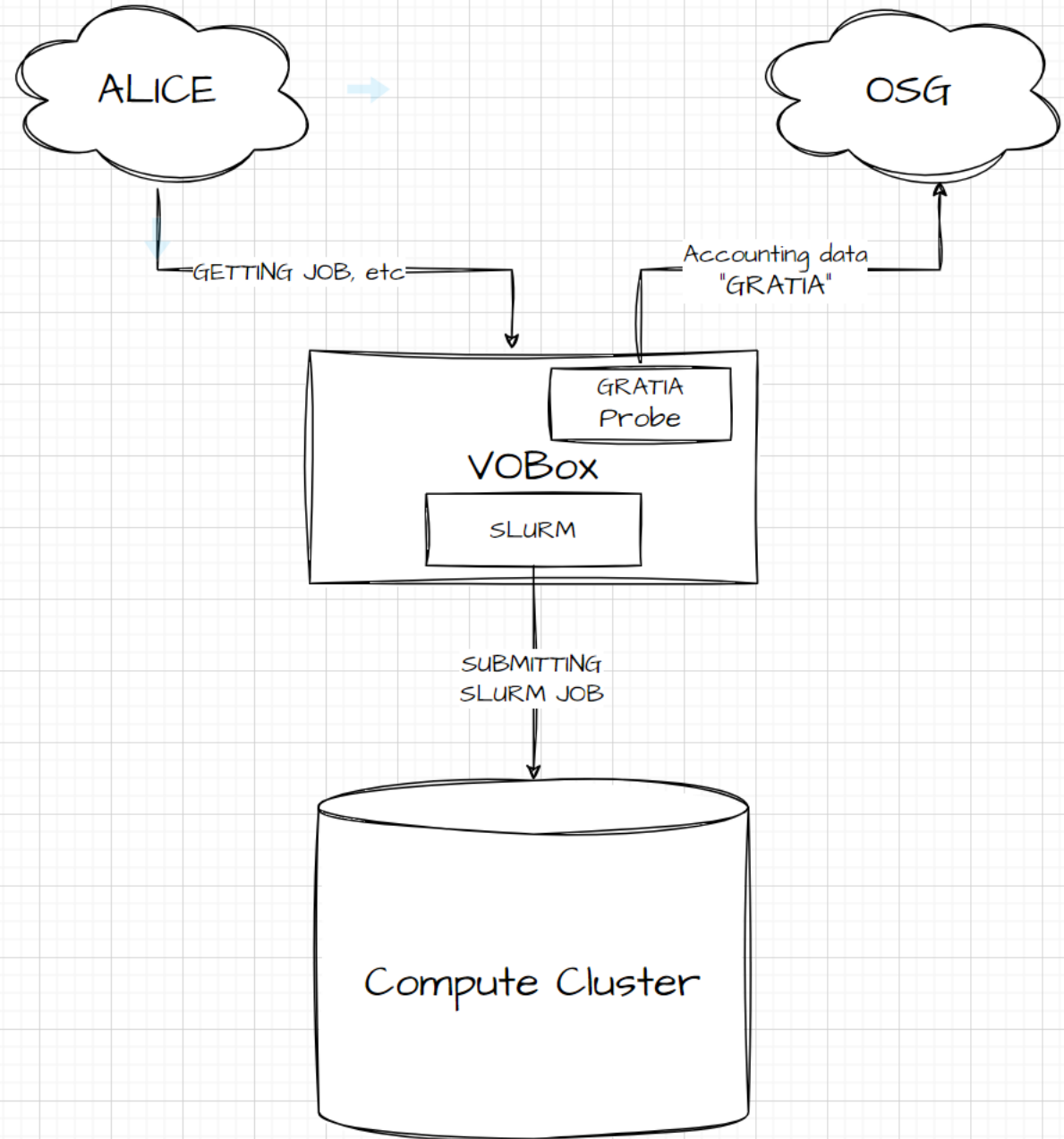
- [How to get CRIC access and privileges](#)
- [How to manage pledge info in CRIC](#)
- [How to manage storage info in CRIC](#)
- [CRIC main concepts and data models](#)

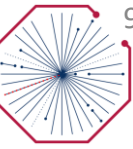
Useful links

- [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

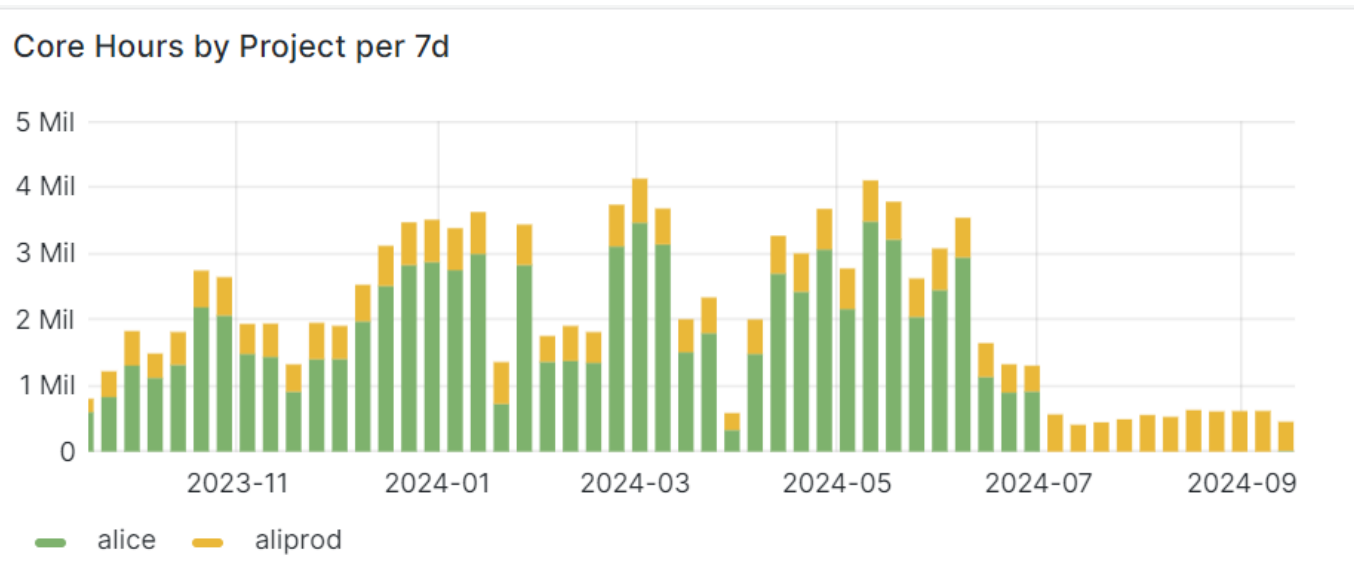
- We ran our T2 using the gratia probe with SLURM for a long time
- Sometimes the probe would get stuck or go down and needed to be restarted and timestamp reset
- But even that was rare and just worked





WLCG/OSG Accounting

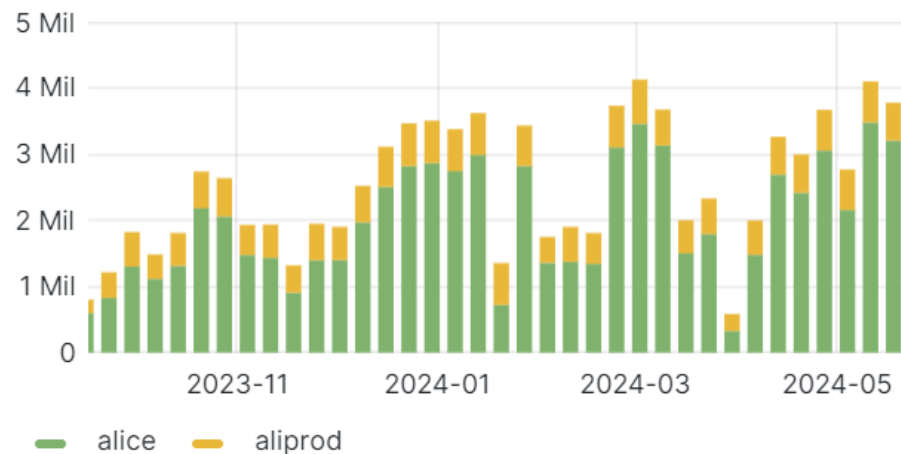
- Reporting/accounting to WLCG is done via OSG
- OSG “gratia” tool collects batch system info and reports to GRACC



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

Core Hours by Project per 7d



openseiencegrid / gratia-probe Public

<> Code Issues Pull requests 2 Actions Projects Security Insights

elevate lots of probes to museum status (SOFTWARE-4467) #91

Merged edquist merged 22 commits into openseiencegrid:2.x from edquist:SOFTWARE-4467.museum on Feb 24, 2021

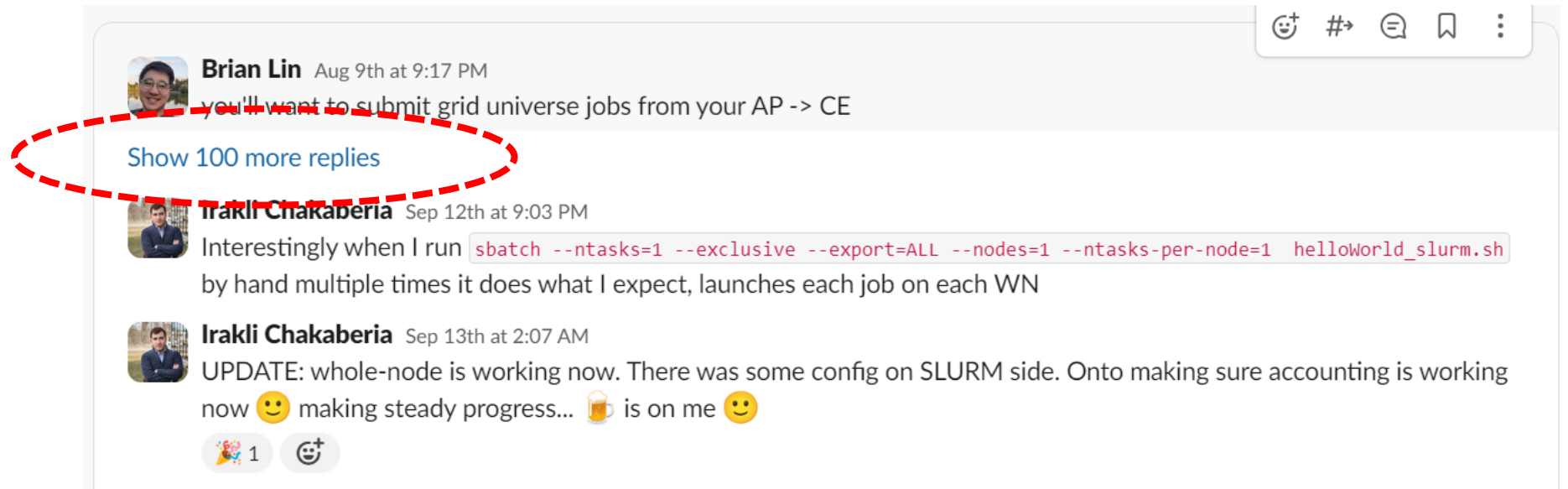
Conversation 18 Commits 22 Checks 0 Files changed 116

edquist commented on Feb 10, 2021 Contributor

No description provided.

Huge gratitude to...

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



Aug 9th at 9:17 PM
you'll want to submit grid universe jobs from your AP -> CE

[Show 100 more replies](#)

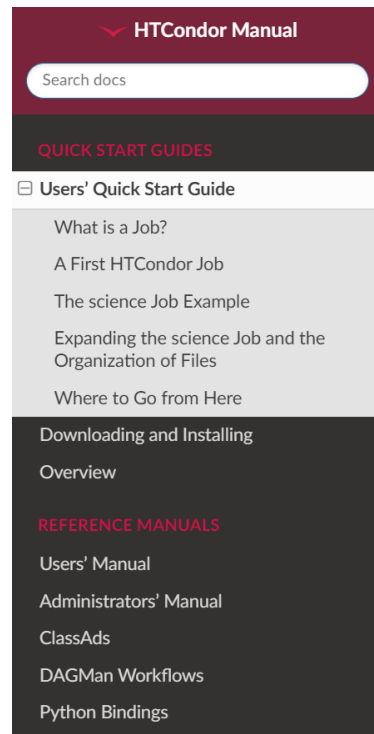
Sep 12th at 9:03 PM
Interestingly when I run `sbatch --ntasks=1 --exclusive --export=ALL --nodes=1 --ntasks-per-node=1 helloWorld_slurm.sh` by hand multiple times it does what I expect, launches each job on each WN

Sep 13th at 2:07 AM
UPDATE: whole-node is working now. There was some config on SLURM side. Onto making sure accounting is working now 😊 making steady progress... 🍺 is on me 😊

1 🧑🏻‍🤝‍🧑🏻

Documentation is good too

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



HTCondor Manual

Search docs

QUICK START GUIDES

Users' Quick Start Guide

- What is a Job?
- A First HTCondor Job
- The science Job Example
- Expanding the science Job and the Organization of Files
- Where to Go from Here

Downloading and Installing

Overview

REFERENCE MANUALS

- Users' Manual
- Administrators' Manual
- ClassAds
- DAGMan Workflows
- Python Bindings

» Users' Quick Start Guide

[Edit on GitHub](#)

Users' Quick Start Guide

HTCondor is a system for dynamically sharing computational resources between competing computational tasks. As an HTCondor user, you will describe your computational tasks as a series of independent, asynchronous "jobs." You access computational resources managed by HTCondor by submitting (or "placing") job descriptions at an HTCondor "access point" (AP), also known as a "submit node." HTCondor locates an appropriate machine for each job, packages up the job and ships it off to that machine for execution. Machines providing resources to HTCondor are therefore known as execution points (EP).

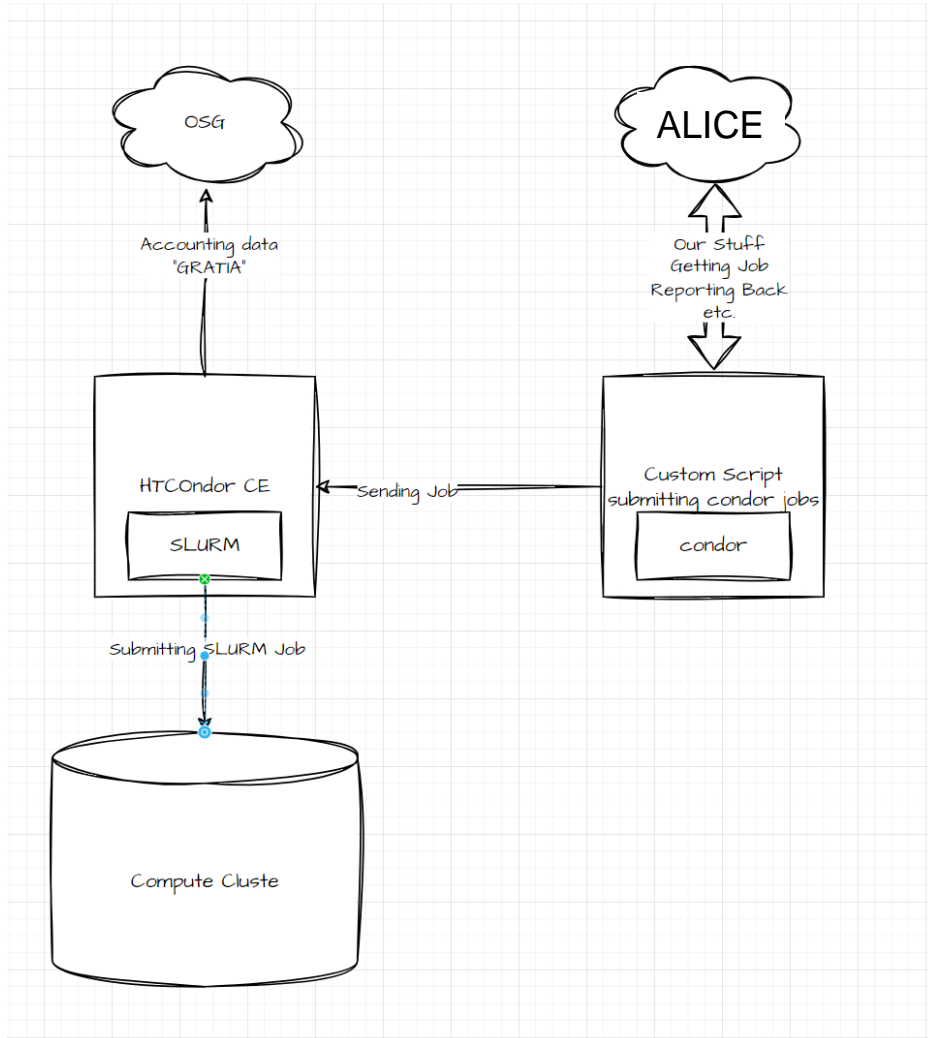
This guide covers submitting and observing the successful completion of a first, example job. It then suggests extensions that you can apply to your own jobs.

This guide presumes that

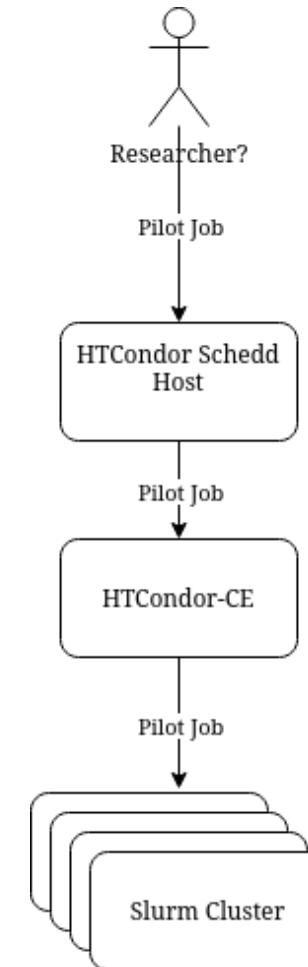
- HTCondor is running
- You have access to a machine within the pool that may submit jobs, termed an Access Point (AP).
- You are logged in to and working on the AP. (If you just finished [getting HTCondor](#), the one machine you just installed is this AP.)
- Your program executable, your submit description file, and any needed input files are all on the file system of the AP.
- Your job (the program executable) is able to run without any interactive input. Standard input (from the keyboard), standard output (seen on the display), and standard error (seen on the display) may still be used, but their contents will be redirected from/to files.

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/wlwg.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
```

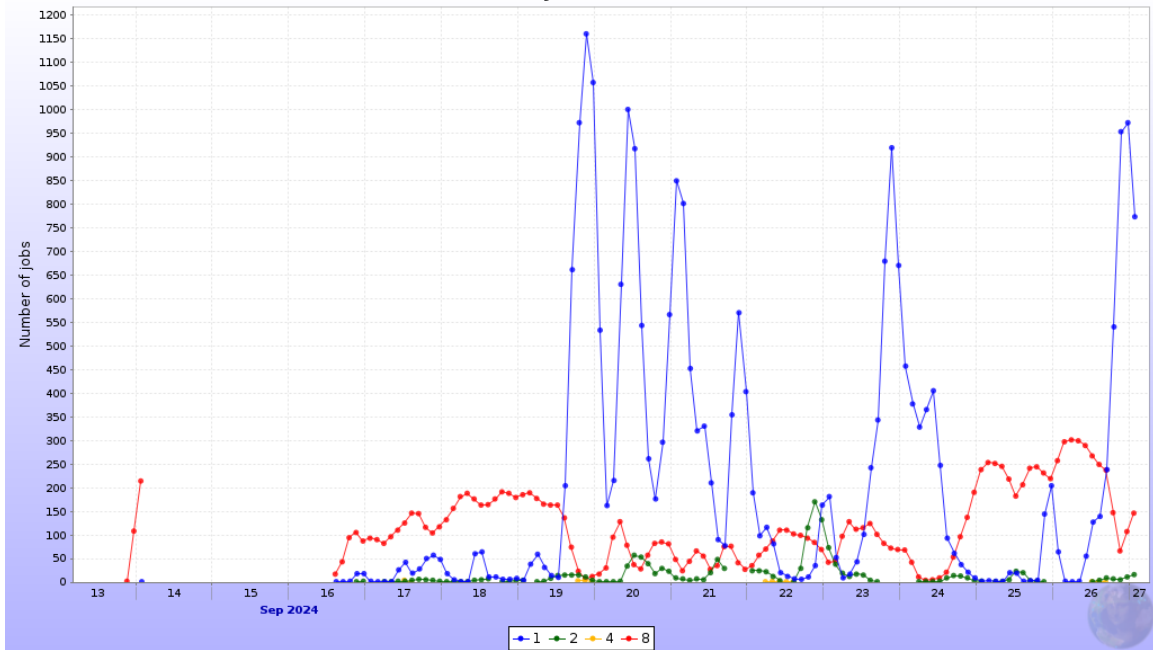
On the VOBox

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

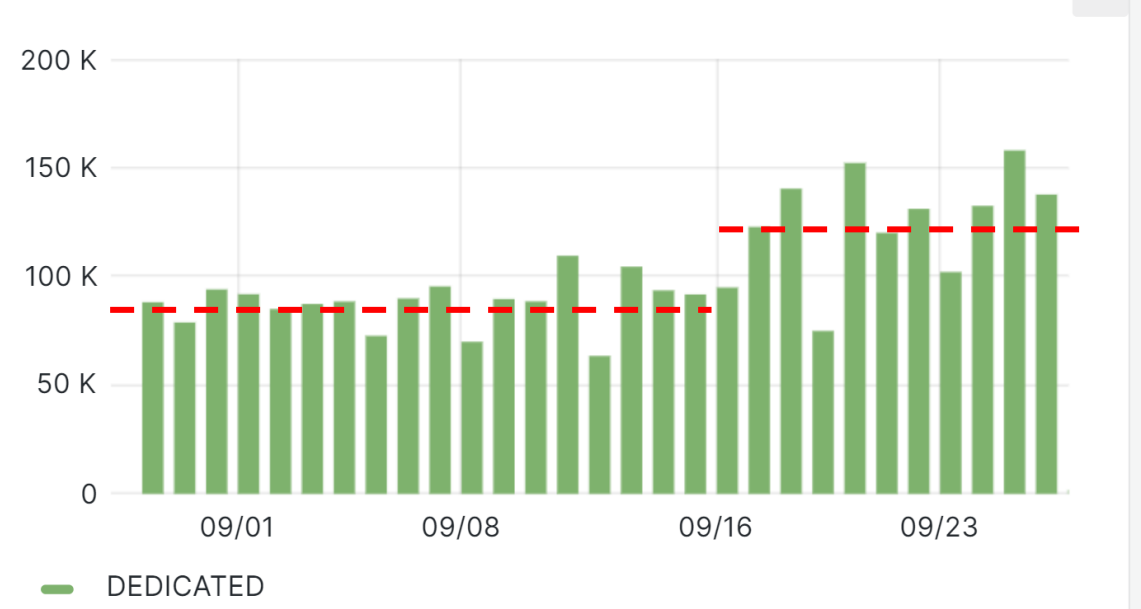
HTCondor @ ORNL

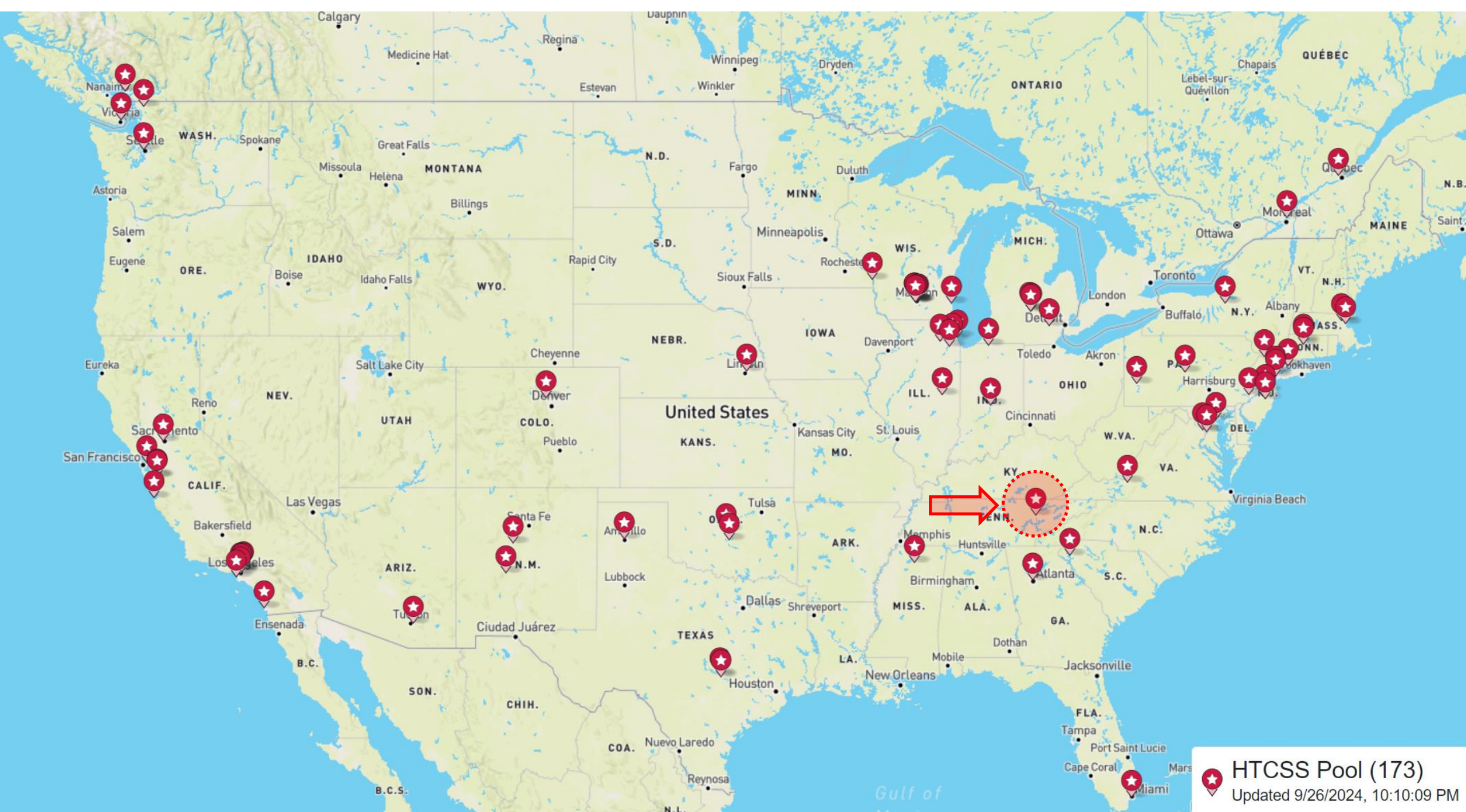
- It took a few weeks of configuring, debugging, reconfiguring, re-debugging, understanding, fixing, coffee, bothering HTCondor people, repeating
- But we got there...

Number of jobs at ORNL-test



Core Hours by Usage Model by 1d





 **HTCSS Pool (173)**
Updated 9/26/2024, 10:10:09 PM

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