Production Services Update

Brian Bockelman,

HTC23, Madison, WI

What are Production Services anyway?

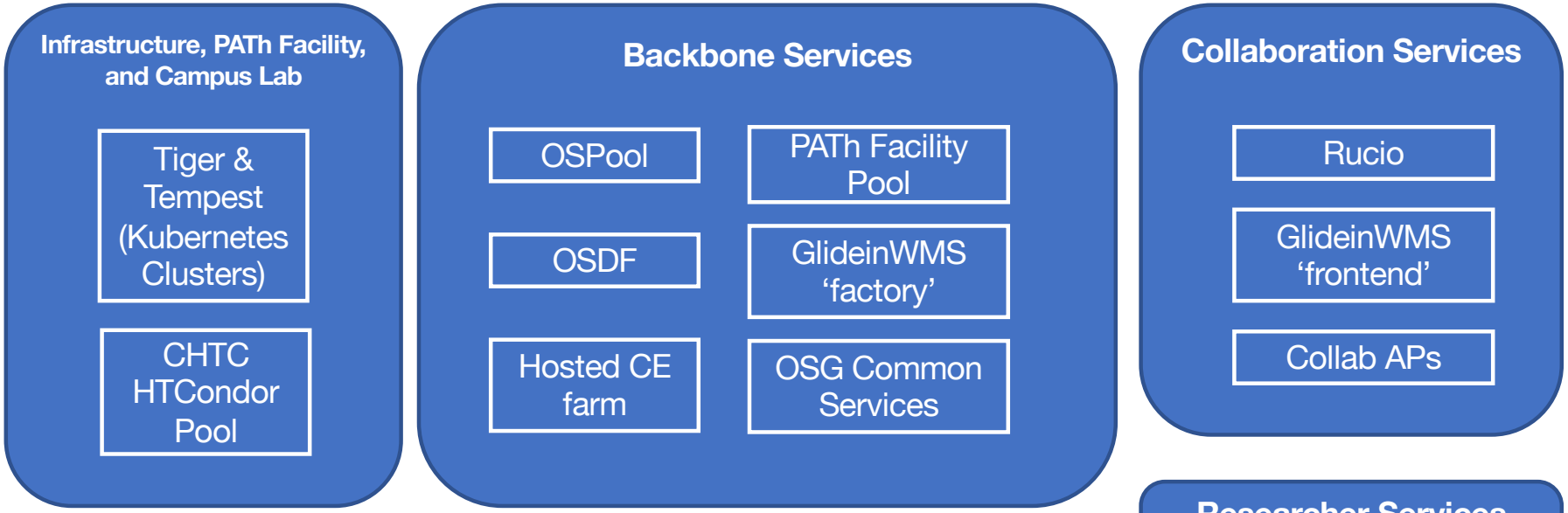
- The ‘OSG Fabric of Services’ is broadly defined:
  - The OSG Software team provides Software Services, an integrated software stack that institutions can download, run, and operate.
  - The Facilitation & Collaboration Services team provides Consultation Services, helping groups small to large utilize dHTC.
  - And finally, we have Production Services, centrally run network services that allow the OSG federations to work together.

The OSG Consortium

Established in 2005, the OSG Consortium operates a fabric of distributed High Throughput Computing (dHTC) services in support of the National Science & Engineering community. The research collaborations, campuses, national laboratories, and software providers that form the consortium are unified in their commitment to advance open science via these services.

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING iris hep OSG CHTC Center For High Throughput Computing

The Task

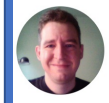


- **Backbone services** – Common production services used by all.
- **Collaboration Services** – services run in support of a specific collaboration.

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING iris hep OSG CHTC Center For High Throughput Computing

The Team

Infrastructure, PATh Facility, and Campus Lab


Aaron Moate
Systems Administrator
University of Wisconsin–Madison


Theng Vang
System Administrator
Center For High Throughput Computing


Jeff Peterson
System Administrator
Morgridge Institute


Jeronimo Bezerra
Senior Systems Administrator
Florida International University


Farnaz Golnaraghi
Systems Administrator
University of Chicago

Backbone Services


Colby Walsworth
Software Integration Developer
University of California - San Diego


Mats Rynge
Systems Integrator
University of Southern California - Information Sciences Institute


Ashton Graves
DevOps Engineer
University of Lincoln-Nebraska


Jeff Dost
Program Analyst
University of California San Diego


John Thiltges
Systems Administrator
University of Nebraska-Lincoln


Fabio Andrijauska
Senior Software Developer
University of California San Diego


Derek Weitzel
Software Integration Developer
University of Nebraska-Lincoln


Mátyás Selmeci
Software Integration Developer
University of Wisconsin–Madison

Collaboration & Researcher Services


Pascal Paschos
OSG Collaborations Area Coordinator/Sr. Computational Scientist
University of Chicago


Jason Patton
Software Integration Developer
University of Wisconsin-Madison


Judith Stephen
Systems Administrator
University of Chicago


David Jordan
Systems Administrator
University of Chicago

Note: Some individuals contribute to multiple teams

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

iris hep

OSG

CHTC Center For High Throughput Computing

Infrastructure

Infrastructure Agility is Key

- We must advance throughput computing to succeed. This implies a consistently evolving set of services.
- Since the start of PATh, we have rebuilt almost all possible OSG services to be based on Kubernetes
- Services are run inside container images rebuilt weekly.
- Services are run at the "Tiger" cluster in Madison and "River" cluster in Chicago (currently migrating to the new "Tempest" cluster).

Why Kubernetes?

- Developers write high-level descriptions of the infrastructure needed, Kubernetes deploys them – including network / DNS / filesystem integration.
- Kubernetes provides a uniform layer across our distributed operations team.
- Deploy the hardware according to local policies but provide the same abstraction.
- Provides a mechanism to migrate between sites.

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING iris hep OSG CHTC Center For High Throughput Computing

The backbone services include over **200 containers in 3 Kubernetes clusters**.

Our Git-based Kubernetes infrastructure has deployed over **7,500 changes**.

We build **100 container images a week**.

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

iris hep

OSG

CHTC Center For High Throughput Computing

Backbone Services

The Backbone is OSCF and OSDF

- The Open Science Compute Federation (OSCF): the interoperable services (CEs, Glidein factories, HTCondor pools) used to make compute capacity available for science.

- The Open Science Data Federation (OSDF): The OSDF service federates datasets in disparate repositories into a coherent namespace and to deliver their objects to computational capacity through a network of caches.

- Many other services (accounting, registration, software distribution) which we won't touch on today.

OSCF Integrates Institutions Across 6 Continents

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

iris hep

OSG

CHTC Center For High Throughput Computing

The OSPool

- The OSPool is the HTCondor pool operated for Open Science.
  - Policies are managed by the OSG Executive Director on behalf of the OSG Consortium
  - Projects receive fair share, not allocations.
  - Projects can be either a PI-led group (professor + students) or a multi-institutional collaboration (1 CI professional representing 1,000 scientists).
- The OSPool is a popular place to put 'opportunistic' capacity:
  - Campuses integrate their capacity through a 'Compute Entrypoint' (CE) service that accepts glideins, contributing otherwise-idle cores.
  - One way for CC* awardees to integrate with national cyberinfrastructure services.

PATh THROUGHPUT COMPUTING

iris hep

OSG

CHTC Center For High Throughput Computing

Connecting to the OSPool

Two ways to add capacity to the OSPool:

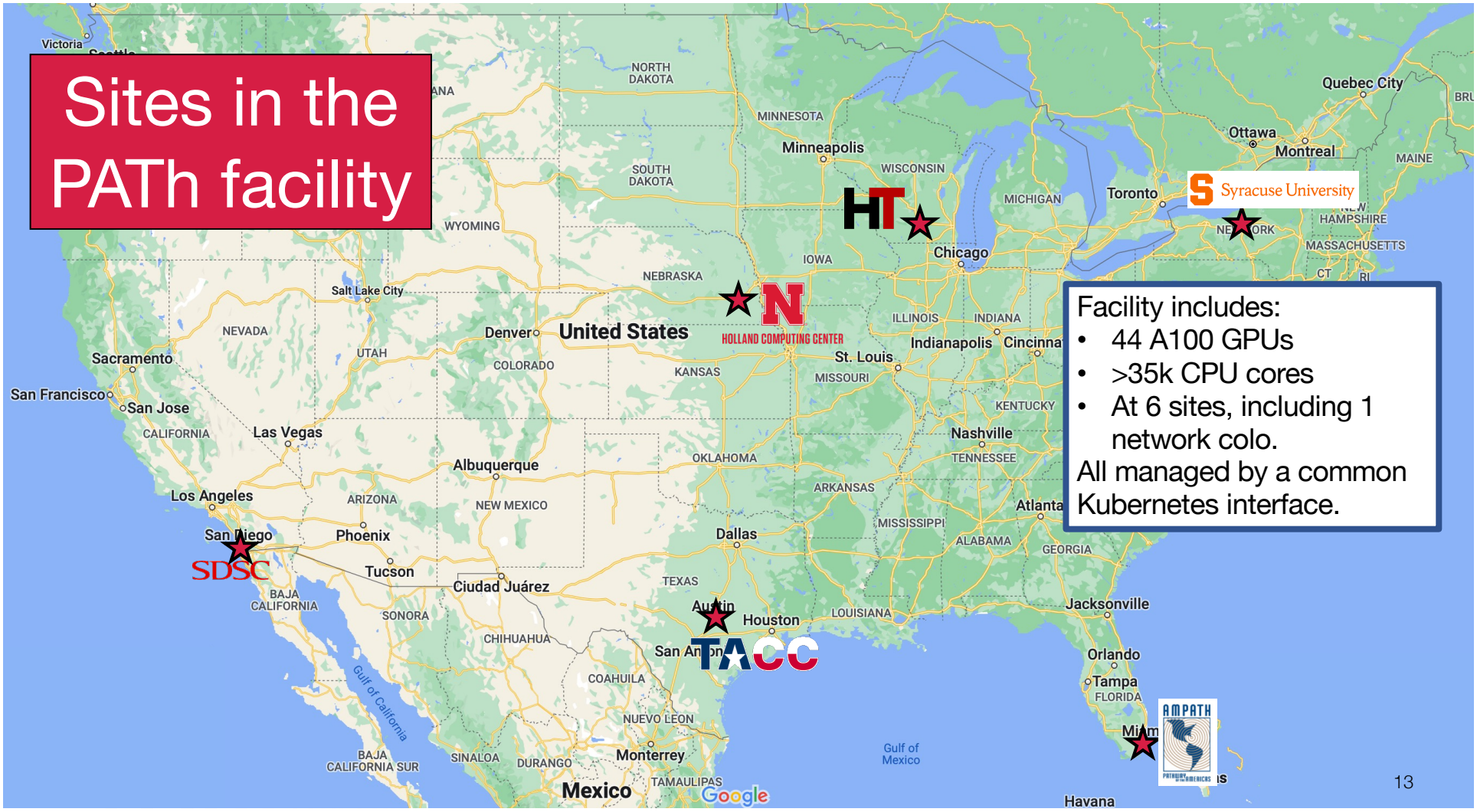
1. CEs. Based on a measurement of the workloads at each AP, a central service submits glideins to site batch systems through the CE.
   - Roughly 2/3 of the capacity.
2. Containers. The site starts containers on available local capacity (typically, via Kubernetes).

Rank	Institution	Core hours	Jobs
1	Syracuse University	58,345,077	17,290,423
★ 2	University of California San Diego	28,345,997	15,901,291
★ 3	University of Wisconsin	25,434,442	11,313,212
★ 4	Lancium	23,790,038	8,817,627
★ 5	Great Plains Network	11,546,421	9,171,016
6	Fermi National Accelerator Laboratory	8,730,168	4,309,296
7	University of Connecticut	7,768,870	8,379,684
8	University of Chicago	6,229,758	3,141,145
9	Indiana University	5,099,281	4,372,638
10	University of Colorado	3,677,453	2,758,990
11	American Museum of Natural History	3,116,648	2,087,056
12	University of Nebraska	2,276,761	1,697,565
13	University of Notre Dame	2,086,139	1,751,513
14	Lehigh University	2,026,829	1,469,439
15	University of Washington	1,914,389	2,424,714

★ = Majority/all capacity contributed via containers

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING iris hep OSG CHTC Center For High Throughput Computing

Sites in the PATh facility



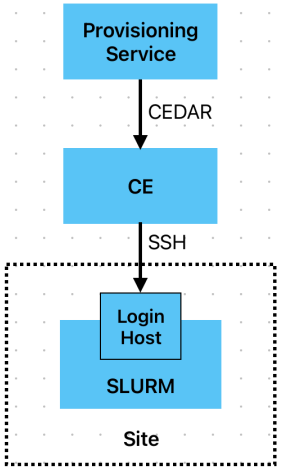
Facility includes:

- 44 A100 GPUs
- >35k CPU cores
- At 6 sites, including 1 network colo.

All managed by a common Kubernetes interface.

Backbone Services: Hosted CE farm

55 hosted CEs



- The CE is the entrypoint into a cluster.
- Most clusters already have an entrypoint: SSH.
- Hosted CE: The team operates the CE service on behalf of the site, submitting batch jobs via SSH.

A Diversity of Sites:

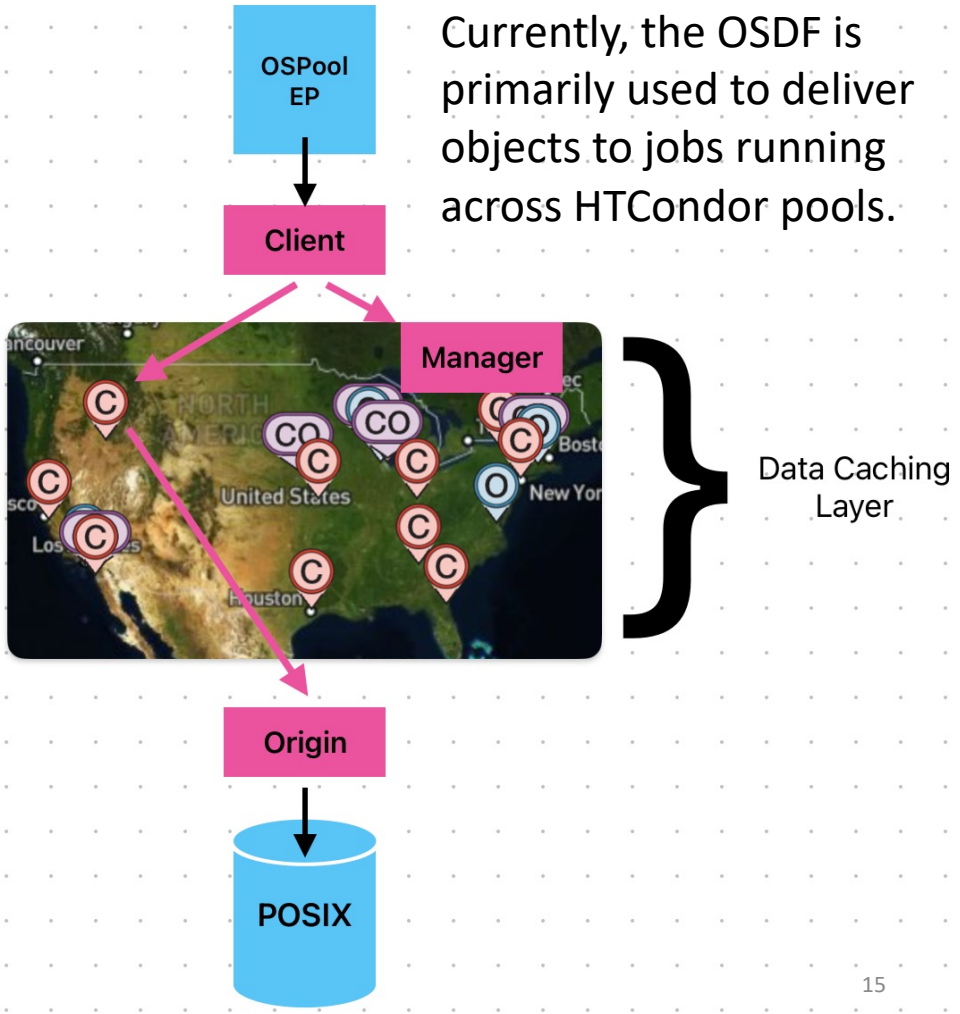
We work with an enormous range of local expertise.
Some sites need help configuring SLURM.
Some sites have unexpectedly long maintenances or outages.

On any given week, ~10% of sites aren't accepting jobs.
On any given week, ~5% of CEs are changed by our operators.
(new SSH cred., debug quota issues, tune glideins, new OS, etc)

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING iris hep OSG CHTC Center For High Throughput Computing

OSDF Architecture

- An origin service integrates the object store into the OSDF in the same way a CE integrates a batch system into the OSPool. Interfaces to move data and map authorizations.
- The cache service stores and forwards objects, providing scalability to the data access.
- The manager selects a source/sink of an object for clients and maintains the namespace.



Currently, the OSDF is primarily used to deliver objects to jobs running across HTCondor pools.

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

iris hep

OSDF at a Glance

In one month,

45 OSG Projects

Moved

2.8PB

Through

28 caches

The OSDF provides origins with a scalable, uniform, way to share data with the world.

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

iris hep

OSG

CHTC Center For High Throughput Computing

Collaboration & Researcher Services

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

iris hep

OSG

CHTC Center For High Throughput Computing

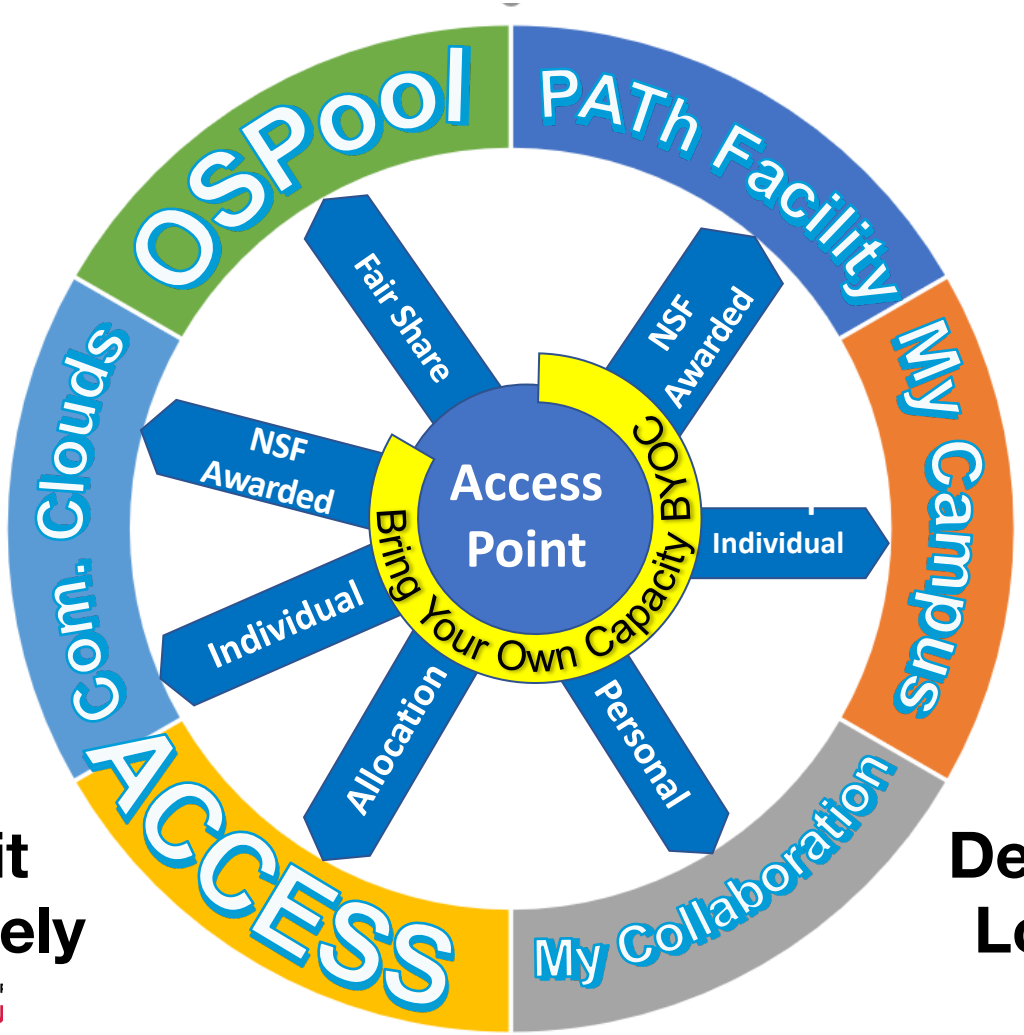
The Access Point (AP)

AP Services	Technology
Login / Terminal access	SSH
Job Database	HTCSS
Workflow Execution Engine	DAGMan, Pegasus
Job Execution	HTCSS
Filesystems, Object Stores	(Various) CephFS
Transfer Services	OSDF

The Access Point is a service where users can place their workloads and delegate their execution.

- It is not merely the "submit node" but **all the corresponding services**.
  - It is more a toolkit than single monolith. E.g., for filesystems, many possible technologies.
- The AP can connect to multiple pools or users can **BYOC** – Bring Your Own Capacity.
- The AP is an independent service, not "part of a cluster".

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING iris hep OSG CHTC Center For High Throughput Computing



Use it Remotely

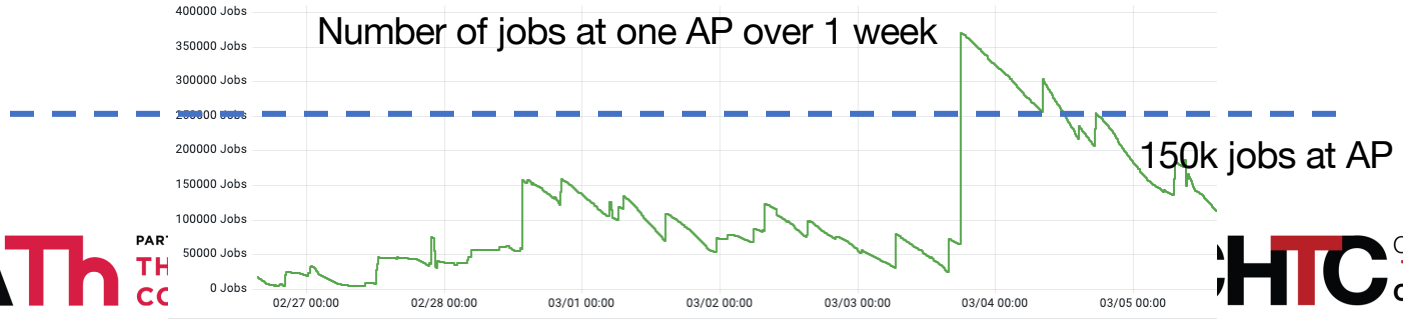
Deploy it Locally

PATh PARTNERSHII THROU COMPU ING

Center For High Throughput Computing

Example: OSG-Connect

- The OSG-Connect APs at UChicago are general purpose services for PI-led research groups.
  - Each AP can manage about 1M jobs.
  - Primarily connected to the OSPool but allows researchers to BYOC from ACCESS or other sources.
  - Data can be brought to the AP through traditional file transfer (scp, rsync) or Globus. Workflows really start at the researcher's laptop!
  - Given the diversity of capacity that can be connected to the AP, most researchers use either containers (~75% of jobs) or other portable runtimes (conda).



Collaboration Services

- For collaborations, we provide a wide range of services, some bespoke, to help them effectively utilize dHTC.
  - Think of this as “wholesale”: supporting a single collaboration may empower a thousand scientists!
- The ultimate goal is to “graduate” collaborations: have them utilize shared services and support channels to advance their science.

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING iris hep OSG CHTC Center For High Throughput Computing

Looking Forward

From a rebuilt foundation

- In the first half of PATh, the OSG Fabric of Services transitioned to a new operational model, overhauled most of the backbone services, and added major new infrastructure into the project.
- What's next for the team?
  - Support the new CC* Storage awards: 9 new collaborations! Quite different task compared to the CC* Compute collaborations.
  - Migrating collaborations to use more backbone services.
  - Add new ways for groups to connect to the PATh Facility.
- And, of course, adapt to the changing ecosystem internal and external to the OSG Consortium!

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING iris hep OSG CHTC Center For High Throughput Computing

Acknowledgements

This work is supported by NSF under Grant Nos. 2030508 and 1836650. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.