OSG Fabric of Distributed Services

Miron Livny

John P. Morgridge Professor of Computer Science

UW Center for High Throughput Computing

Morgridge Institute for Research

OSG Statement of Purpose

OSG is a consortium dedicated to the advancement of open science via the practice of distributed High Throughput Computing (dHTC), and the advancement of its state of the art.

OSG Consortium

- Established in 2005, the OSG is a consortium governed by a **council**
- Consortium Members (Stakeholder) include campuses, research collaborations, software providers and compute, storage, networking providers
- The OSG provides a **fabric of dHTC Services** to the consortium members and to the broader Science and Engineering (S&E) community
- While **members own** and operate resources, the consortium does not own or operate any resources
- Council elects the OSG **Executive Director** who appoints an **Executive team**. Together they steer and manage available effort

 project - a Partnership launched by the NSF in 2020 between UW-Madison Center for High Throughput Computing (**CHTC**) and the **OSG** Consortium to advance **Throughput Computing.**

 Software Suite (HTCSS) manages High Throughput workloads across all forms of research computing resources from campus clusters to commercial clouds and HPC facilities

 services enable science collaborations and campuses to build and operate private distributed computing environments across >70 sites
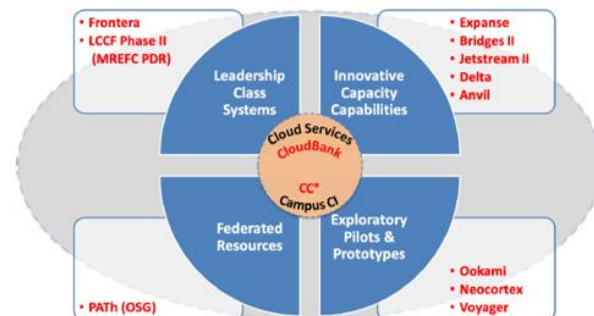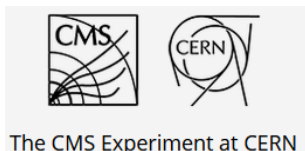
Aligned with NSF CI Eco-system

IEEE COMPUTER SOCIETY

Technical Committee on Distributed Processing
2020 Outstanding Technical Achievement Award

Professor Miron Livny

For Influential Contributions of the Condor System to Distributed and High Throughput Computing

CENTER FOR HIGH THROUGHPUT COMPUTING

OSG

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

Claims for "benefits" provided by
Distributed Processing Systems

P.H. Enslow *"What is a Distributed Data Processing System?"* IEEE Computer, January 1978

- – **High Availability and Reliability**
- – **High System Performance**
- – **Ease of Modular and Incremental Growth**
- – **Automatic Load and Resource Sharing**
- – **Good Response to Temporary Overloads**
- – **Easy Expansion in Capacity and/or Function**

Unity of Control

All the component of the system should be **unified** in their desire to achieve a **common goal**. This goal will determine the rules according to which each of these elements will be controlled.

Component autonomy

The components of the system, both the logical and physical, should be **autonomous** and are thus afforded the ability to refuse a request of service made by another element. However, in order to achieve the system's goals they have to interact in a **cooperative** manner and thus adhere to a common set of policies. These policies should be carried out by the control schemes of each element.

Campus Cyberinfrastructure (CC*)

National Science Foundation

Directorate for Computer and Information Science and Engineering
Office of Advanced Cyberinfrastructure
Division of Computer and Network Systems

Proposals are required to commit to a minimum of 20% shared time on the cluster and describe their approach to making the cluster available as a shared resource external to the campus, with access and authorization according to local administrative policy. Conversely, the proposal should describe the approach to providing **on-demand** access to additional external computing resources for its targeted on-campus users and projects. One possible approach to implementing such a federated distributed computing solution is joining a multi-campus or national federated system such as the Open Science Grid. Whatever opportunistic, federated, scalable, distributed computing platform is

CENTER FOR HIGH THROUGHPUT COMPUTING

OSG

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

OSG Fabric of Services

- Organized under three main thrusts – *Community Building, Research Computing Facilitation, and Operation*

- Designed and operated to assure, **scalability, trustworthiness, reproducibility**.

- OSG claims its services enabled in the past **12** month more than **2B** core hours across more than **130** clusters located at more than **70** sites and more than **200TB** of data cached across **17** caches worldwide.

HTC Access Points (OSG-Connect)

An HTCSS Technology and OSG service provided and operated by PATh

- Manage High Throughput Computing workloads

- Can execute jobs on HTCondor pools

- Can delegate jobs to Batch Systems

- Supports Bring Your Own Resources (BYOR)

- Supports Workflow Management Systems

- Integrated with the Open Science Data Federation

- Can be dedicated to an organization
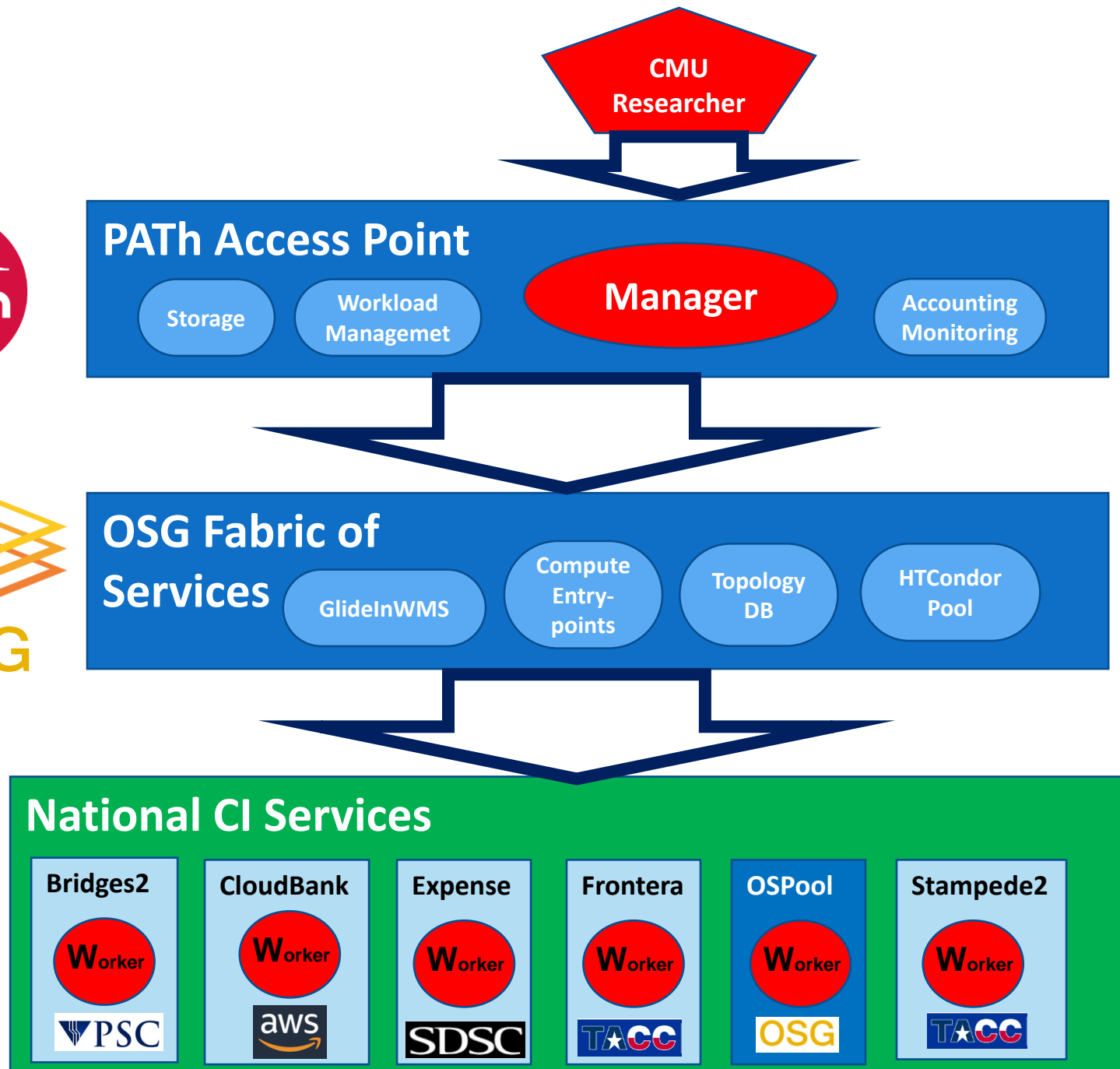
The Open Science Pool (OSPool)

An OSG service operated by PATh

- A means for campus/organization to make (following autonomous policies) capacity available to the US open science community

- A HTCondor pool that serves as fair-share source of High Throughput Capacity (**HTC**) to researchers

- Integrated with the OSG Data Federation

OSPool (typical) Weekly Numbers:

- runs more then **2M** jobs
- placed by more than **70** users
- from more than **60** projects
- at more than **35** Access Points
- on more than **35K** cores
- at more than **50** sites
- that consume more than **4M** core hours

How do sites contribute to the OSPool?

When a site in the OSG Compute Federation wants to contribute the capacity of one of its server to the OSPool, it runs an Execution Point on the server

- – Site can start and stop the Execution Point at any time
- – Execution Point needs to establish trust with HTCondor Pool
- – OSG provides services to remotely activate Execution Points through a Compute EntryPoint (CE) that submits activation requests to the batch system of the cluster
- – OSG provides services to automate remote activation of Execution Points
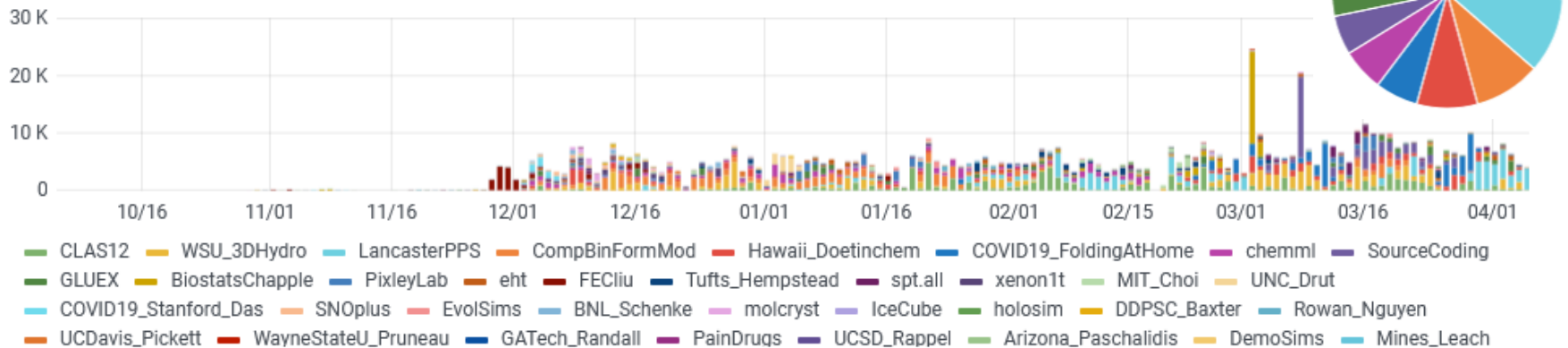- – Execution Point prefers to have out going network connectivity

OpenStack Environment

The **Jetstream** cloud is an OpenStack, NSF-funded academic cloud operated by the Indiana University and TACC.

Operators decide when to start and terminate Virtual Machines (or containers) that run OSPool Execution Points.

750K core hours

Core Wall Hours by Project

Work with NSF funded **Chameleon** is progressing nicely

k8s Cluster

The NSF funded Pacific Research Platform **PRP** operates a k8s cluster across more than 35 location worldwide.
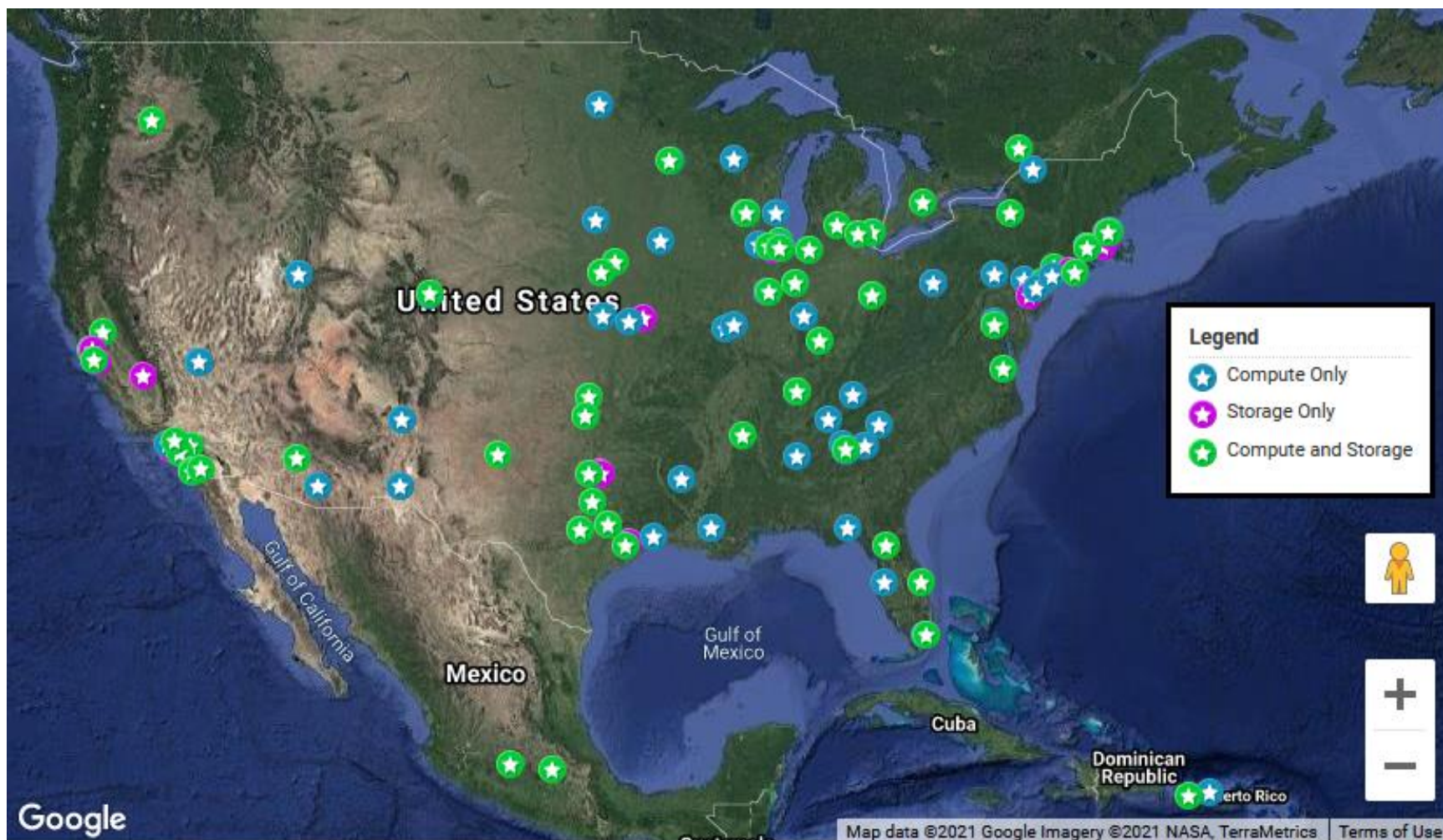
A PRP HTCondor pool grows and shrinks under the control of the k8S cluster pod scheduler

A CE submits requests to HTCondor to deploy OSPool XPoints.

800K GPU hours

GPU Wall Hours by Project

Legend
- Compute Only
- Storage Only
- Compute and Storage
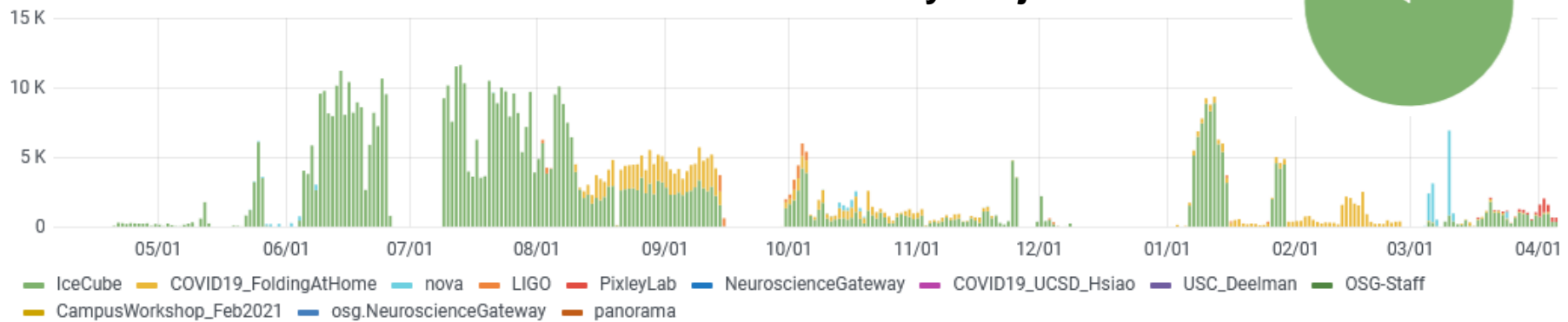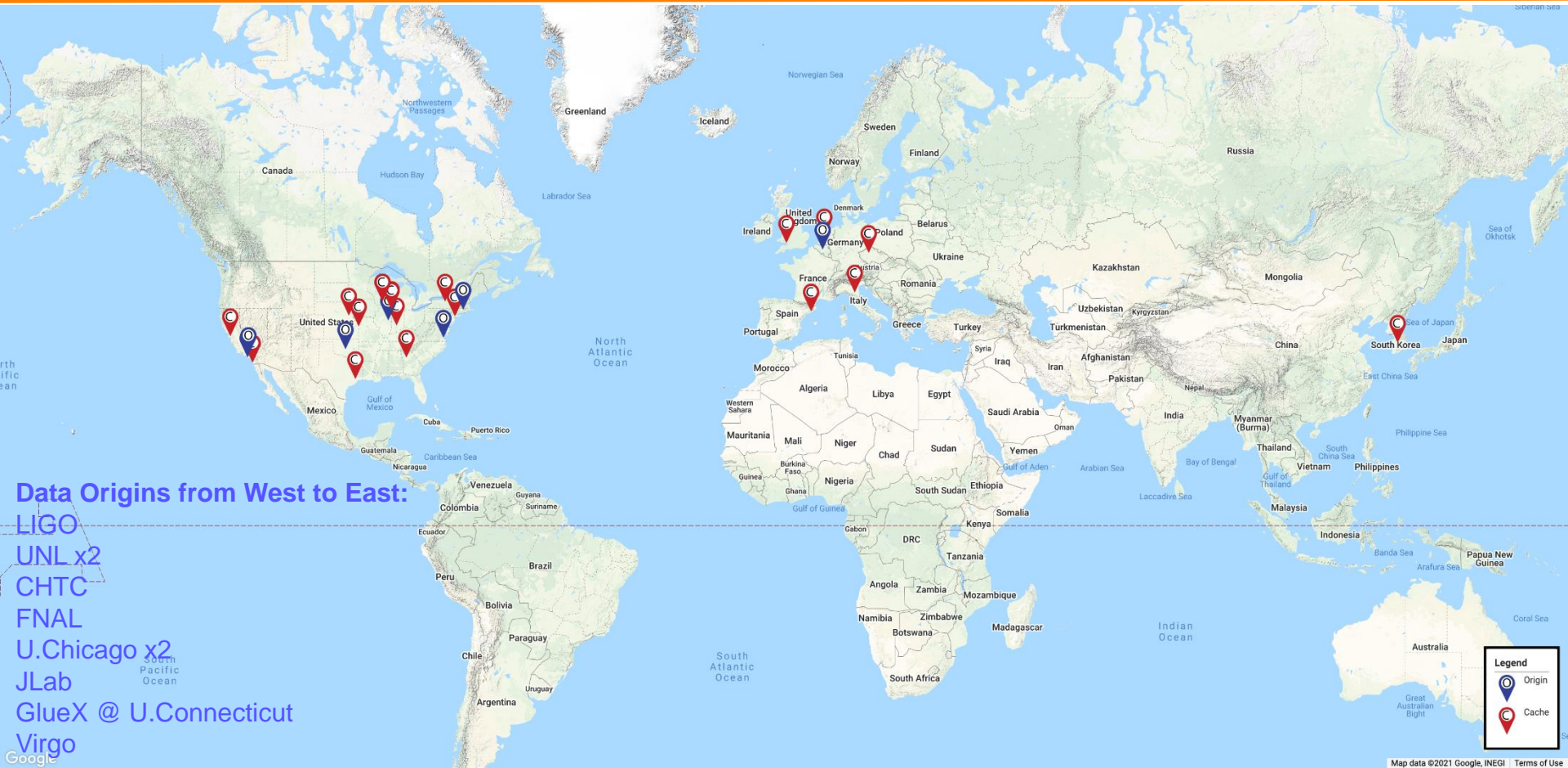
Open Science Data Federation (OSDF)

An OSG service operated by PATh

- Enables via **Data Origins** researchers on campuss to remotely process their local data

- Enables via a network of **Data Caches** effective, sustained access to remote data across large numbers of jobs

OSG Data Federation

Data Origins from West to East:
LIGO
UNL x2
CHTC
FNAL
U.Chicago x2
JLab
GlueX @ U.Connecticut
Virgo

Legend
Origin
Cache

Map data ©2021 Google, INEGI | Terms of Use

17 Caches … 6 of which are in R&E network backbone

10 Data Origins … one of which is for all open science

OSPool is not the only pool!

Organizations like science collaborations (CMS, LIGO, IceCube) and campuses (UCSD, UNL, UW-Madison, JLab) leverage OSG services to deploy and operate private distributed HTCondor pools

- Access Points are private
- Deployment of Execution Points negotiated by the organization with sites
- Resource acquisition and allocation policies defined by the organization

* A site can contribute to different HTCondor pools

HTCondor dHTC CI
of >51K GPUs in the Cloud @ SC19
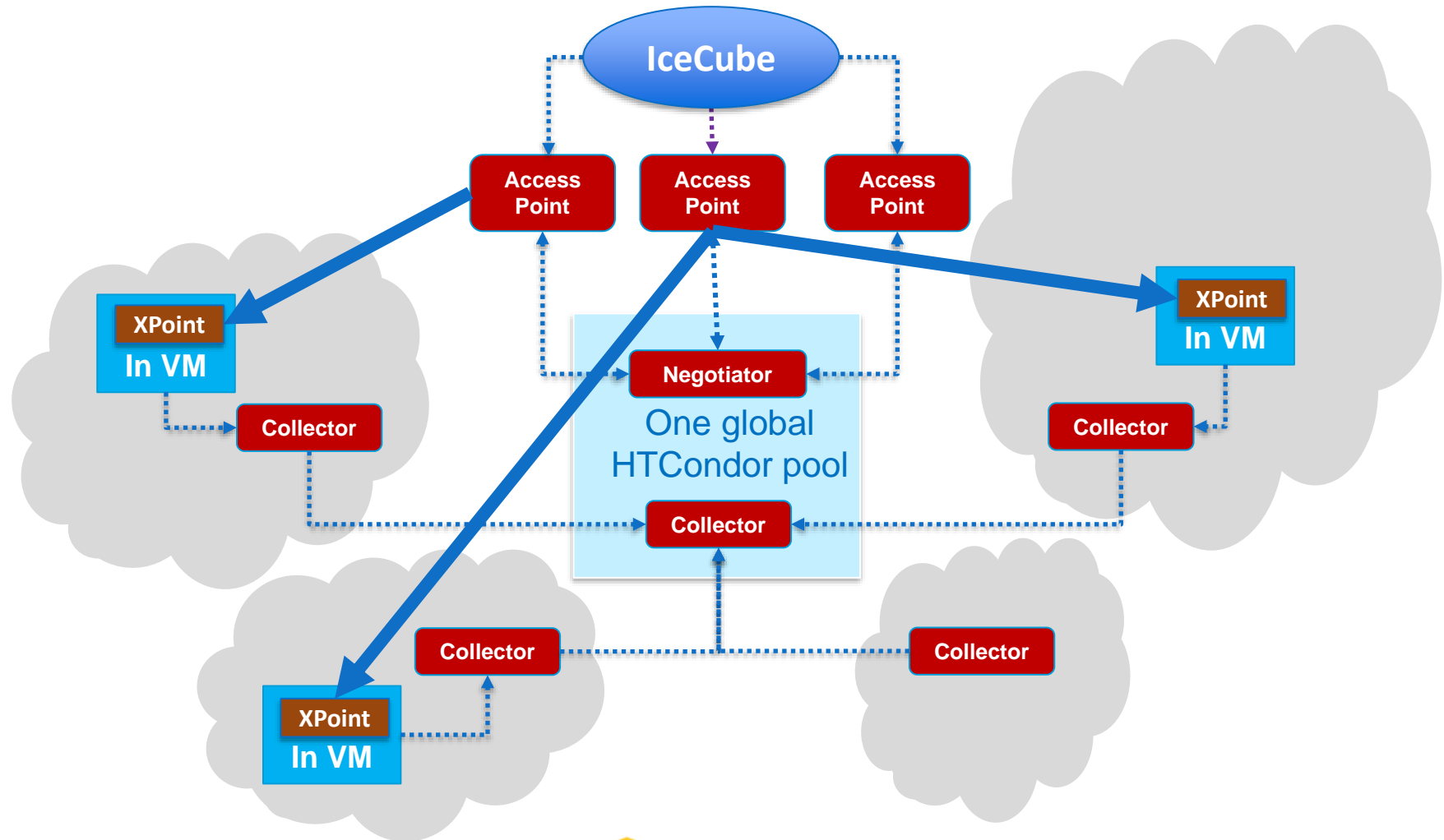
Give me a place to stand, and I shall move the world.

Archimedes of Syracuse was a Greek mathematician, philosopher, scientist and engineer.

Give me a place to run an Execution Point and I shall run your job

Frank Würthwein is a Physics professor at UCSD and the Executive Director of the OSG