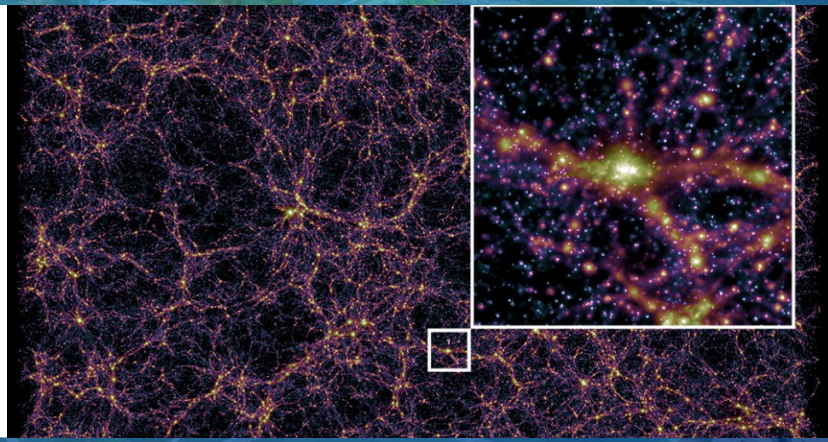
OpenCosmo

COLLABORATIVE COSMOLOGY AT SCALE AND SCIENCE AS A SERVICE



PATRICK WELLS
Argonne National Laboratory

Workflows Community Talks
Feb. 18th 2026

U.S. DEPARTMENT of ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.

Argonne NATIONAL LABORATORY

A brief story…

Frontier-E Simulation

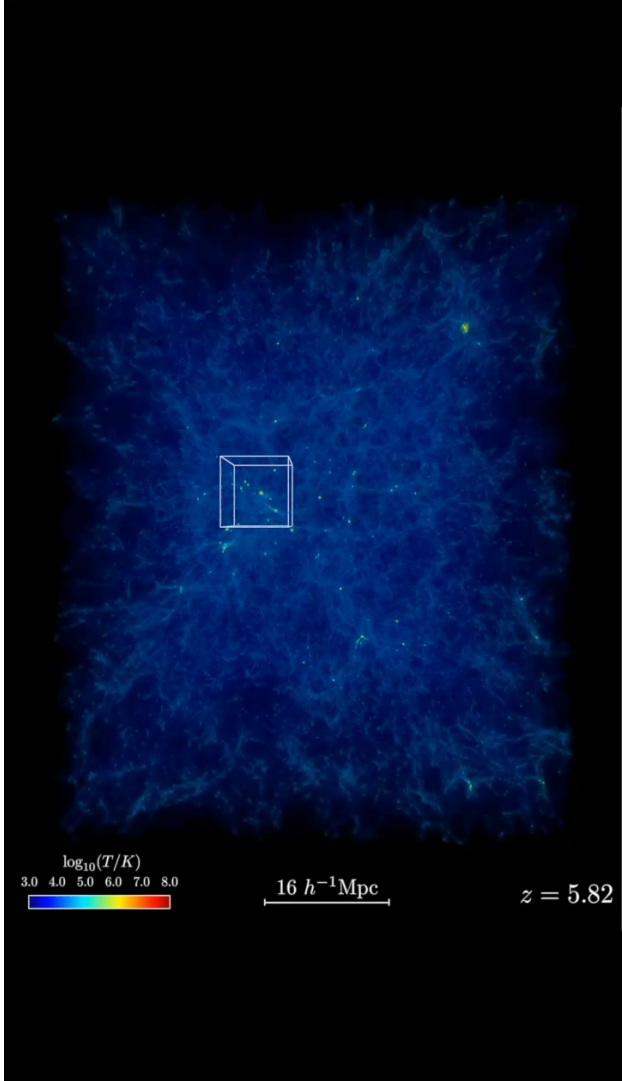
THE CHALLENGE

Extreme-Scale Cosmological Simulations

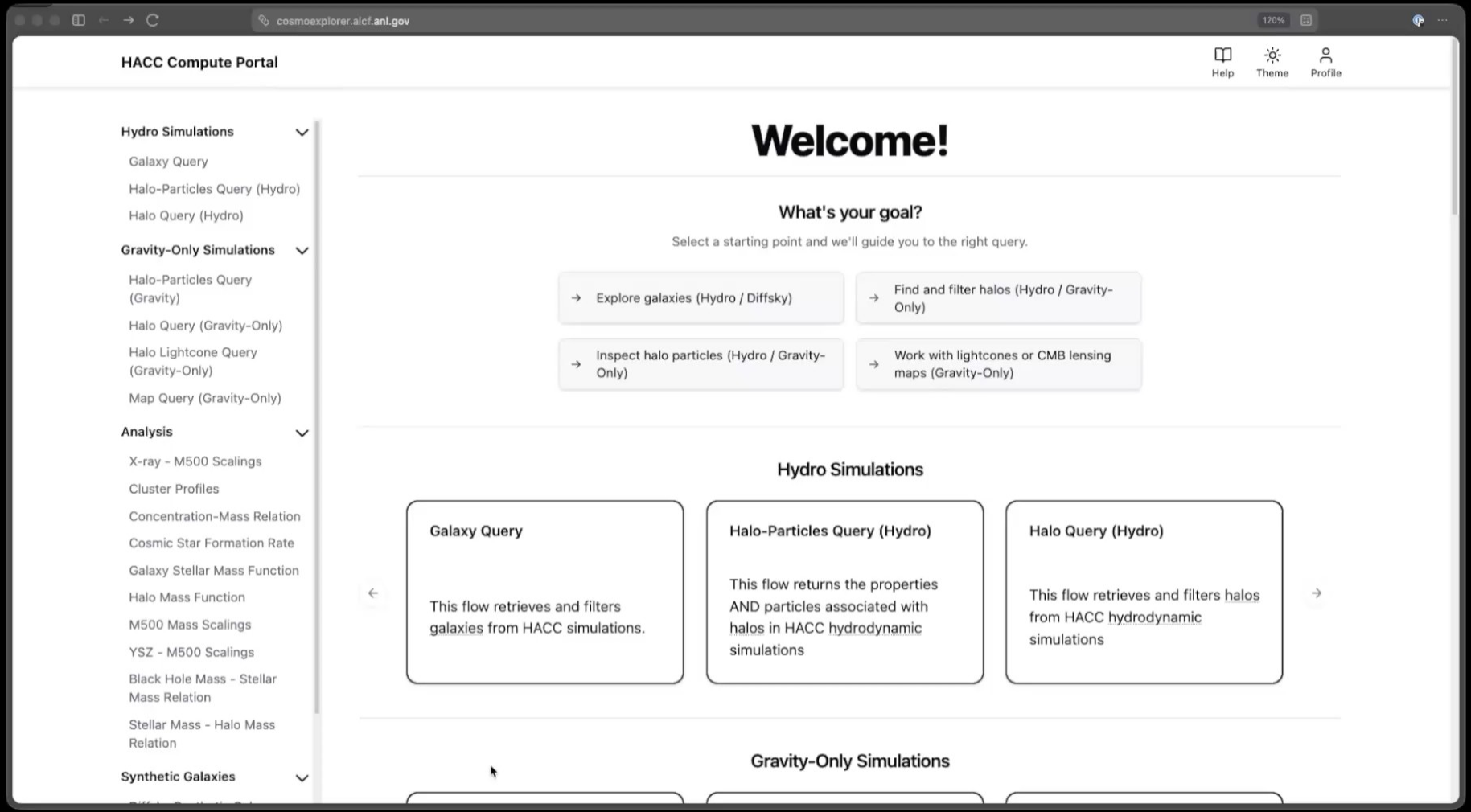
- Petabytes of data
- On several different machines
- In a custom format

First Goal:

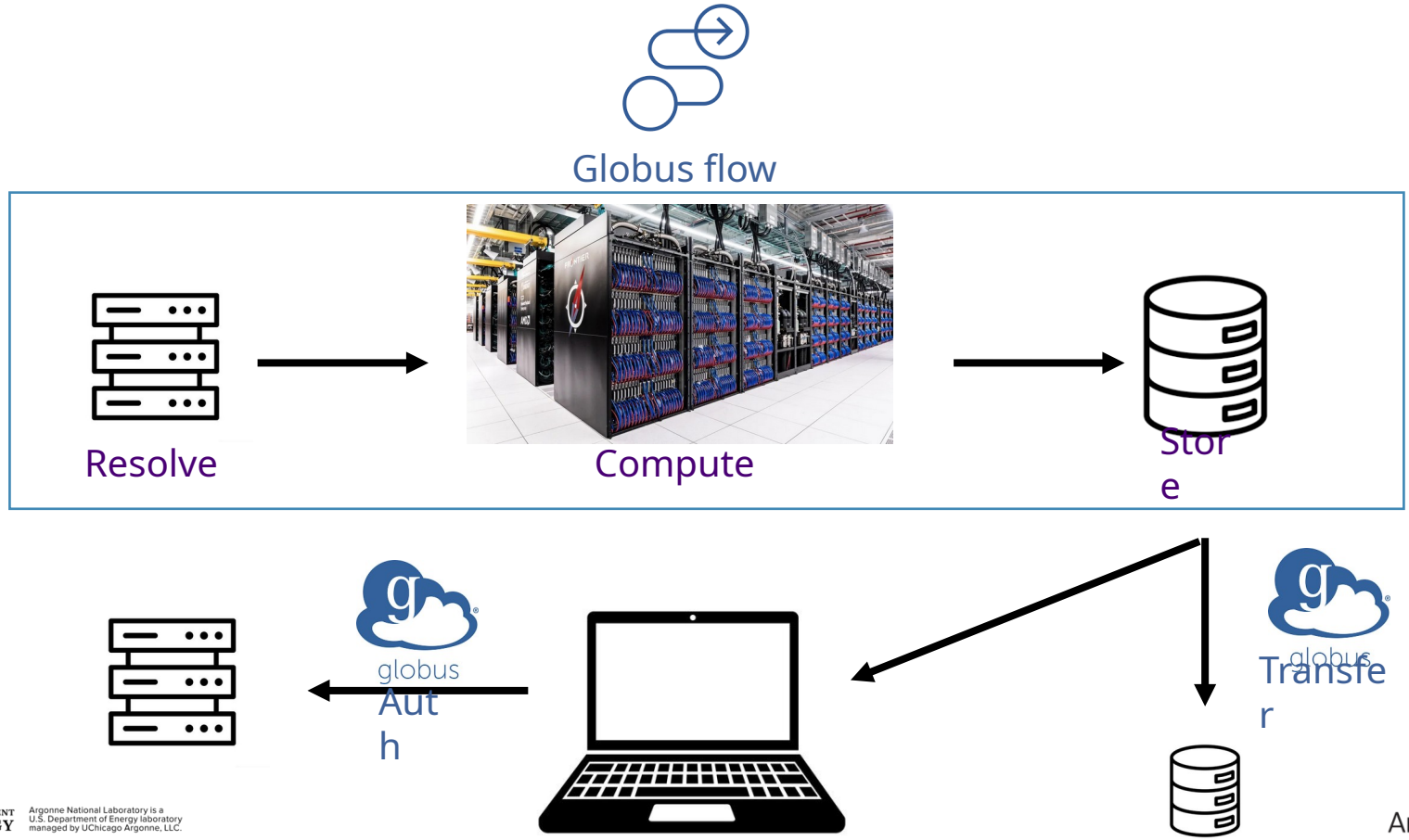
Provide a unified, user-friendly interface to access and query this cosmological data while keeping the user “out of HPC space”*



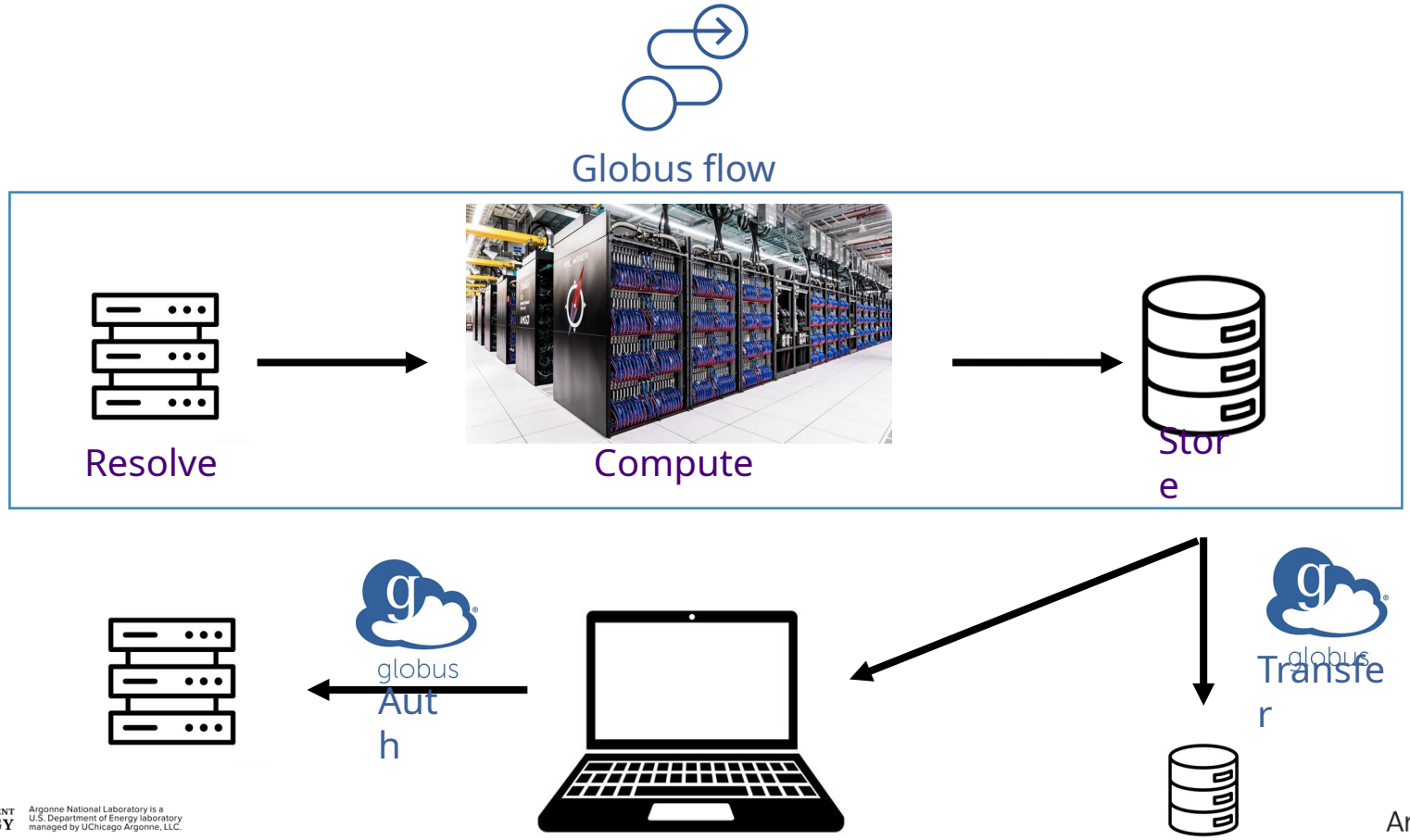
*whenever possible


cosmoexplorer.alcf.anl.gov
120%
HACC Compute Portal
Help
Theme
Profile
Hydro Simulations
Galaxy Query
Halo-Particles Query (Hydro)
Halo Query (Hydro)
Gravity-Only Simulations
Halo-Particles Query (Gravity)
Halo Query (Gravity-Only)
Halo Lightcone Query (Gravity-Only)
Map Query (Gravity-Only)
Analysis
X-ray - M500 Scalings
Cluster Profiles
Concentration-Mass Relation
Cosmic Star Formation Rate
Galaxy Stellar Mass Function
Halo Mass Function
M500 Mass Scalings
YSZ - M500 Scalings
Black Hole Mass - Stellar Mass Relation
Stellar Mass - Halo Mass Relation
Synthetic Galaxies
Welcome!
What's your goal?
Select a starting point and we'll guide you to the right query.
Explore galaxies (Hydro / Diffsky)
Find and filter halos (Hydro / Gravity-Only)
Inspect halo particles (Hydro / Gravity-Only)
Work with lightcones or CMB lensing maps (Gravity-Only)
Hydro Simulations
Galaxy Query
This flow retrieves and filters galaxies from HACC simulations.
Halo-Particles Query (Hydro)
This flow returns the properties AND particles associated with halos in HACC hydrodynamic simulations
Halo Query (Hydro)
This flow retrieves and filters halos from HACC hydrodynamic simulations
Gravity-Only Simulations

LIFECYCLE OF A REQUEST

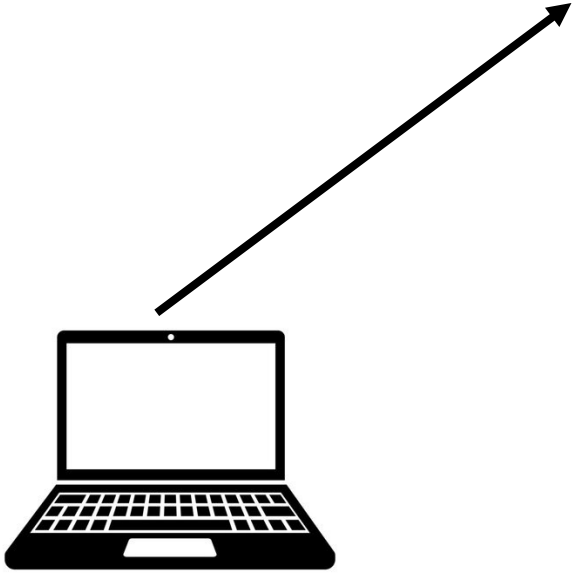


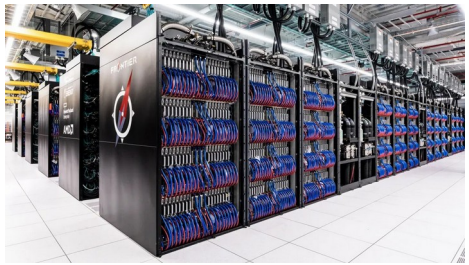
IS THERE ANYTHING SPECIAL ABOUT THIS?

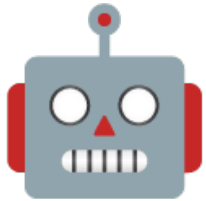


```python
import opencosmo as oc

dataset = oc.open("haloproperties.hdf5", "galaxyproperties.hdf5")
dataset = dataset
            .filter(oc.col("fof_halo_mass") > 1e14)
            .take(10000, at="random")
            .evaluate(
                compute_richness,
                insert=True,
                format="numpy"
            )
```







Argonne NATIONAL LABORATORY

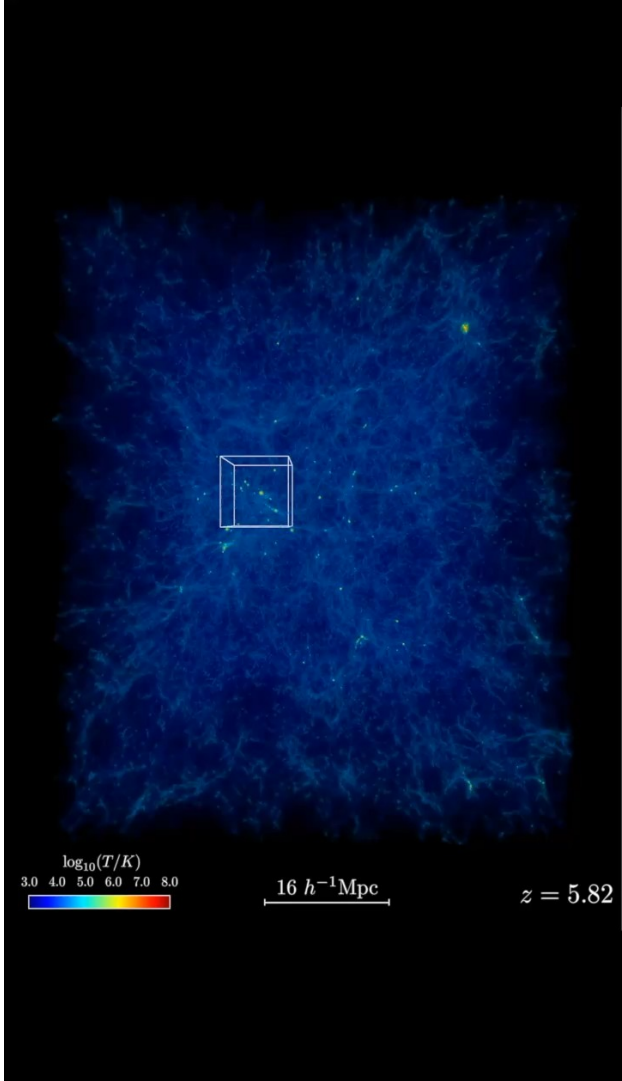
THE REAL CHALLENGE

Extreme-Scale Scientific Data

- Petabytes of tabular data
- On several different machines
- Need to support analytics-oriented queries

Real Goal:

Provide easy-to-use infrastructure tools
that can be used across many domains of science



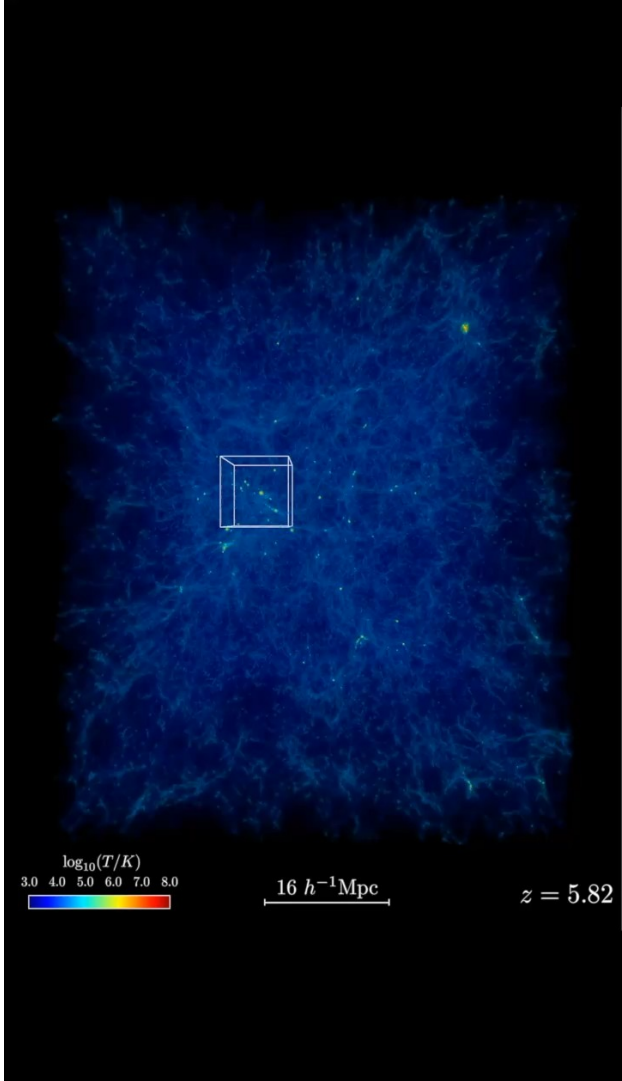
SCIENCE AS A SERVICE

Extreme-Scale Cosmological Simulations

- Petabytes of mixed data
- Across the entire HPC ecosystem
- Complicated analytics-based workloads

Ultimate Goal

Build a layer of abstraction on top of traditional HPC and allow domain scientists to focus on what they do best



U.S. DEPARTMENT of ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.
OpenCosmo

SCIENCE AS A SERVICE

`

Today

- Automated querying at scale across facilities
- Web frontend for user interaction
- Common data format and analysis tooling

Tomorrow

- Automated *analysis* at scale across facilities
- APIs for remote execution of complex, user-defined workloads
- Seamless transitions between remote and local analysis workloads

SCIENCE AS A SERVICE

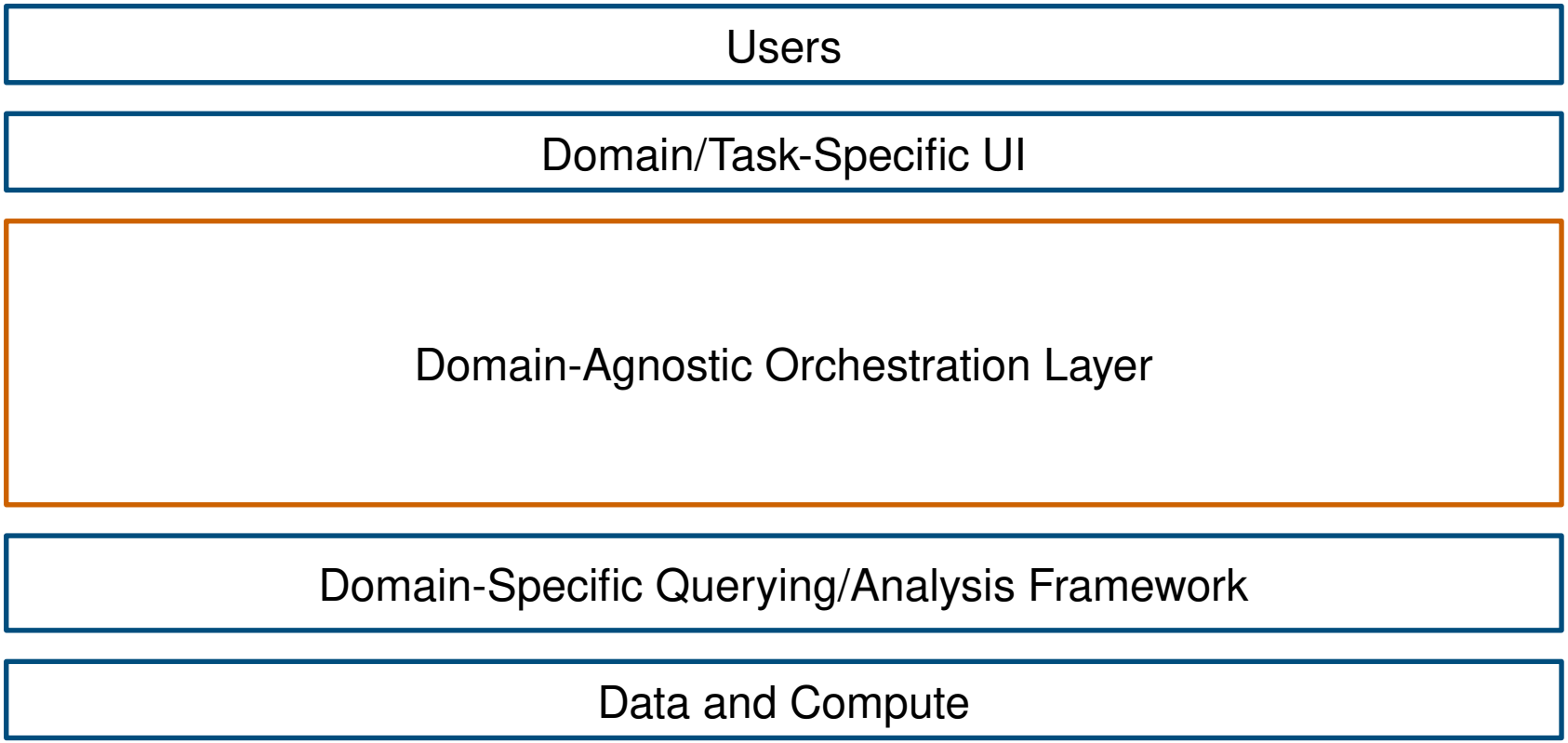
Challenges and the Role of AmSC

While running a *single* workflow is technically straightforward, management of workflows at scale across several HPC-oriented facilities is a much more challenging issue

- Logging
- Automated deployment and CI/CD
- Automated testing
- Security

AmSC?

IS THERE ANYTHING SPECIAL ABOUT THIS?



Workflow: An abstract computational pattern wrapped up in a a nice, easy-to-use framework that allows scientists to focus on science.

Revisiting my Question from the Beginning

Michael Buehlmann

Patricia Larsen

Will Hicks

Manpreet Dhillon

Katrin Heitmann

Salman Habib

Benoit Cote

Tom Uram

Funding support by the U.S. Department of Energy, Office of High Energy Physics