CYVERSE™

Transforming Science Through Data-driven Discovery

Fall in Love with High Performance Computing at CyVerse

John Fonner, Texas Advanced Computing Center

February 14th, 2020

The Vision of CyVerse

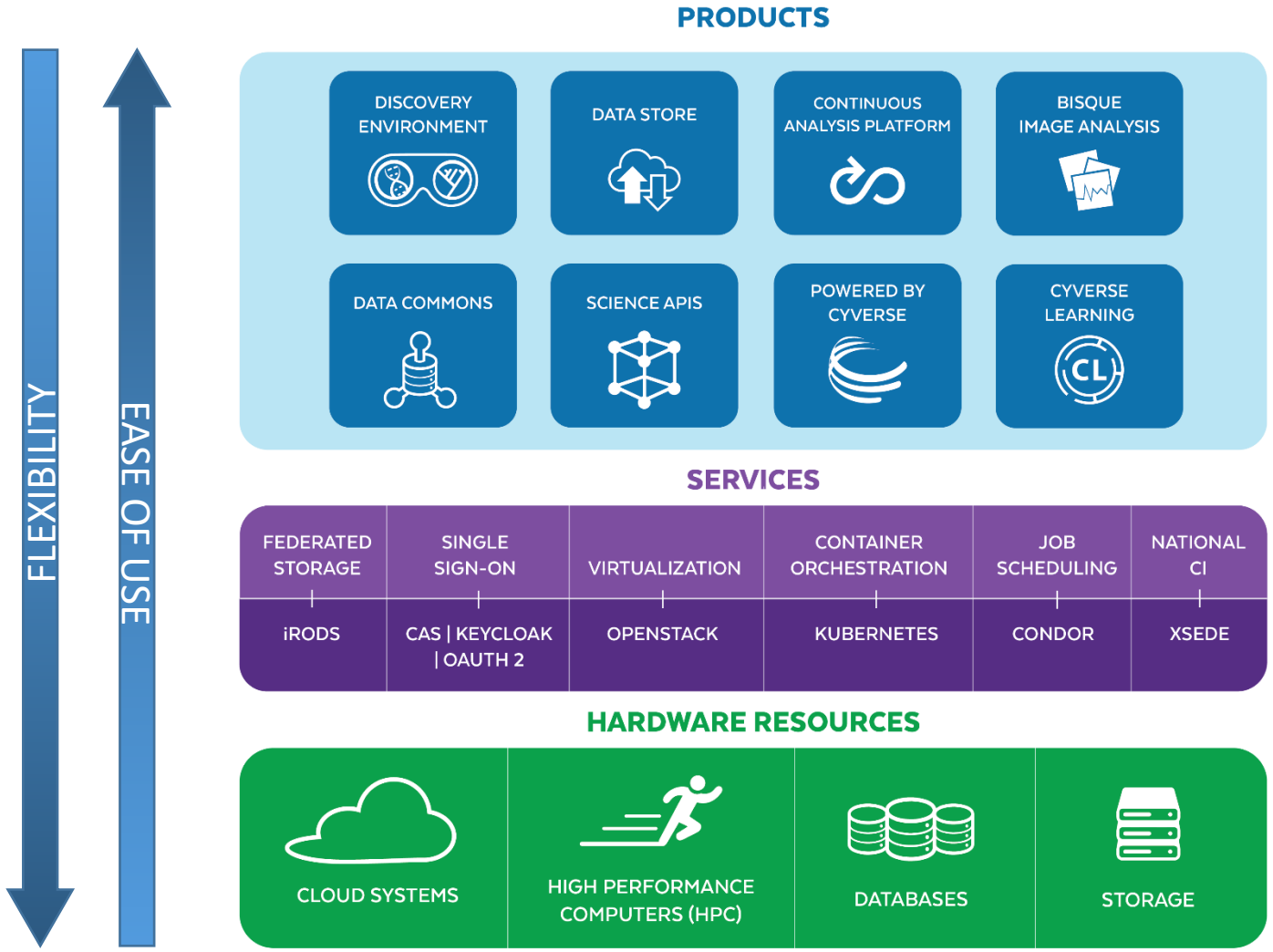
"Transforming Science through Data-Driven Discovery"

Deliver cyberinfrastructure for life sciences research…

(i.e. tools, training, and compute resources)

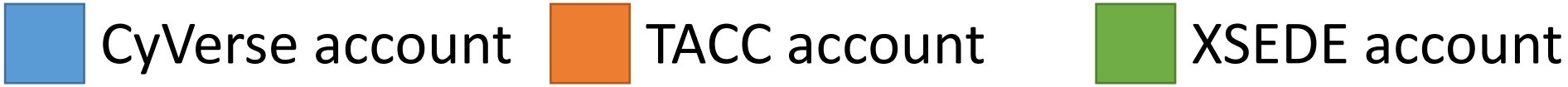
…to make computational research more accessible, scalable, reproducible, and automated

The Anatomy of Cyberinfrastructure



Hardware Resources

- Data Storage – CyVerse Datastore
- High Throughput Computing –CyVerse cluster (U. of Arizona)
- Cloud Computing – Atmosphere, Jetstream
- High Performance Computing – TACC, XSEDE, Tapis
- Functions as a Service - Tapis



What is the Texas Advanced Computing Center?

TACC at a Glance

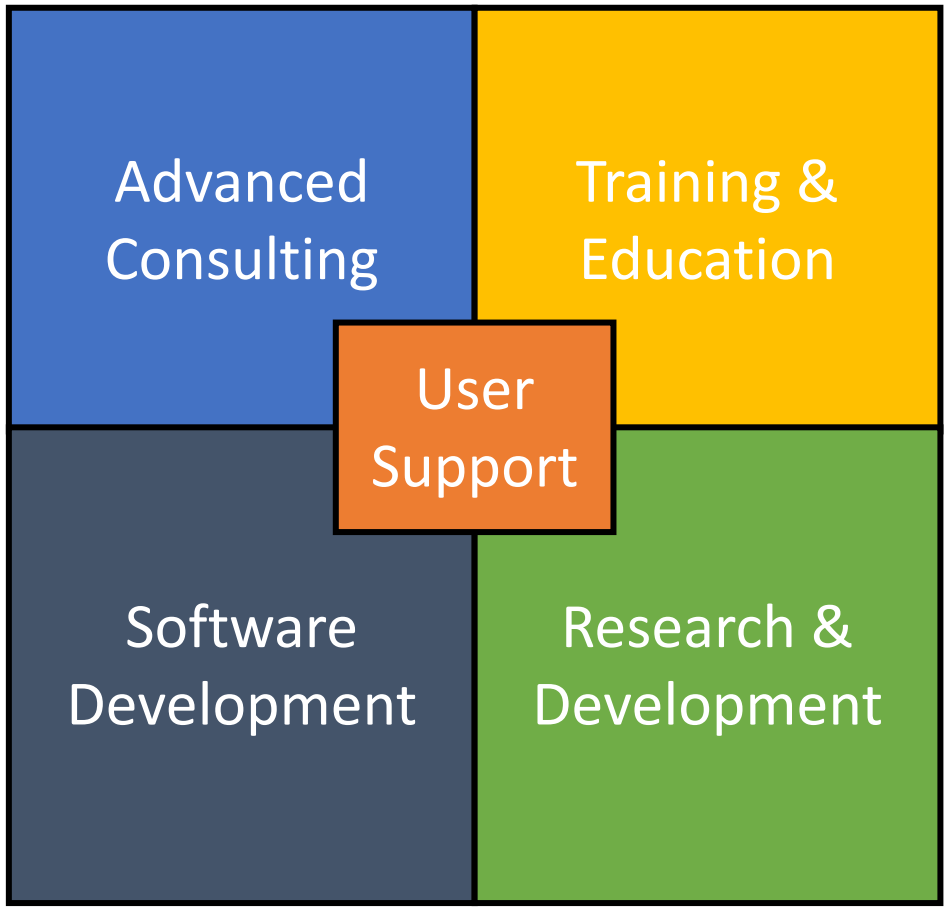
185 staff members
- 70 PhDs, ~25 students
- 1 in 3 have 10+ years exp

Over 50 open source software codes

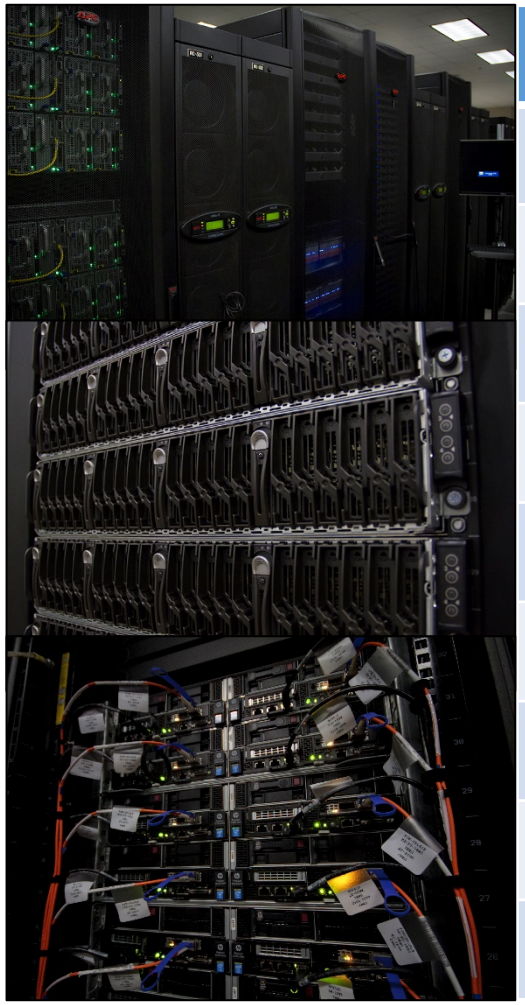
30+ active research projects

Training – from Data Science to MPI to UNIX 101
- 500+ trainees annually

300+ K12 students engaged each year



Hardware Organized Around CAPABILITIES



Computing System	Specialization
Frontera	38PF Intel Cascade Lake – Leadership HPC
Stampede 2	18PF Intel Skylake and KNL – Capability HPC
Lonestar 5	2PF Intel Haswell – HPC & HTC
Longhorn	3PF NVIDIA Tesla V100 GPU – ML, Analytics, Visualization
Jetstream & Chameleon	OpenStack clusters – Cloud, Programmable infrastructure
Rodeo	VMware/OpenStack – Production hosting
Stockyard	20PB Lustre – Compute-optimized filesystem
Corral	15PB GPFS – High-integrity, performant HDD
Ranch	160PB SAMFS – Long-term archival storage

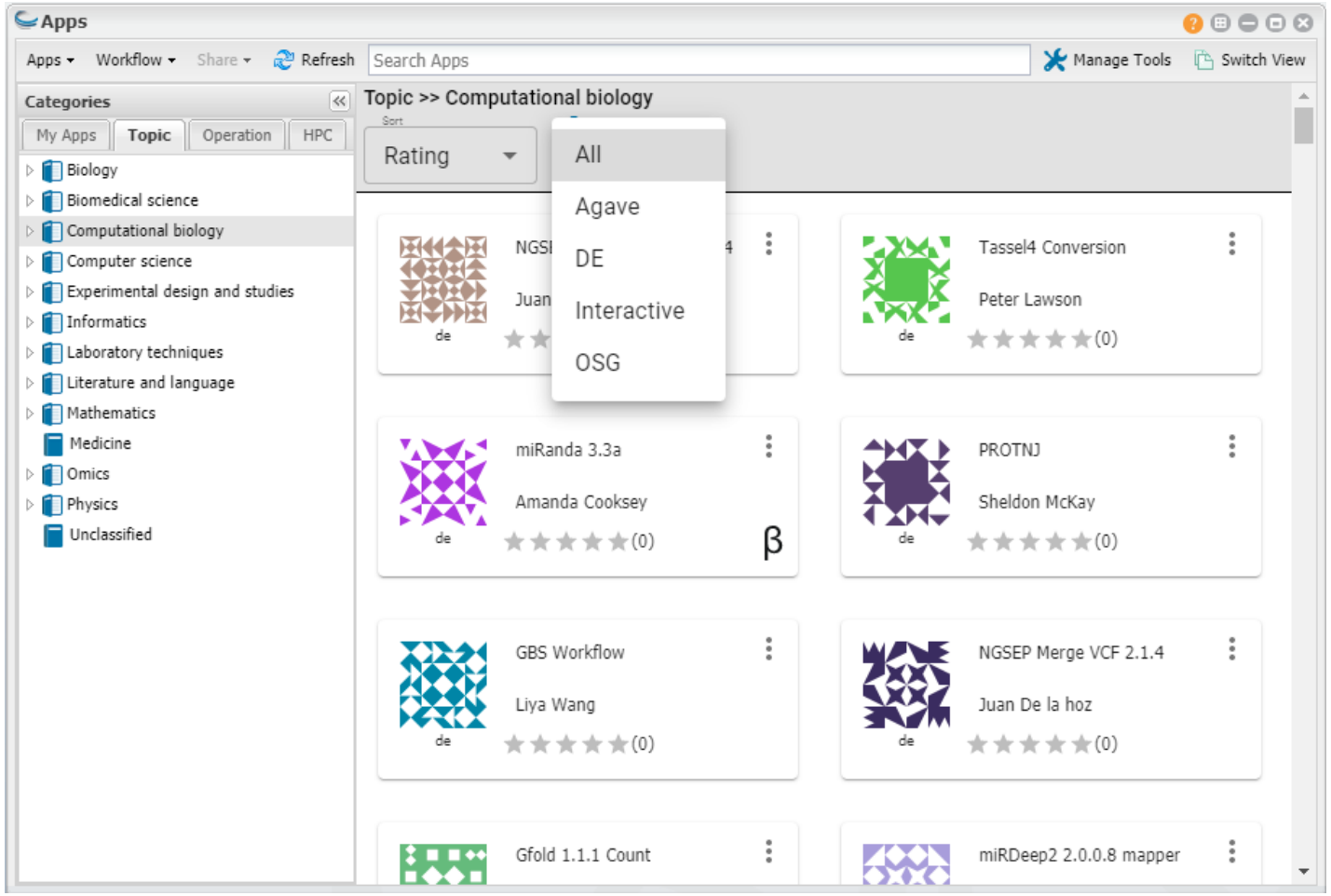
Tech specs at https://portal.tacc.utexas.edu/user-guides

Getting Access to HPC Resources

- 300+ public “HPC” apps already available through the DE or Tapis APIs

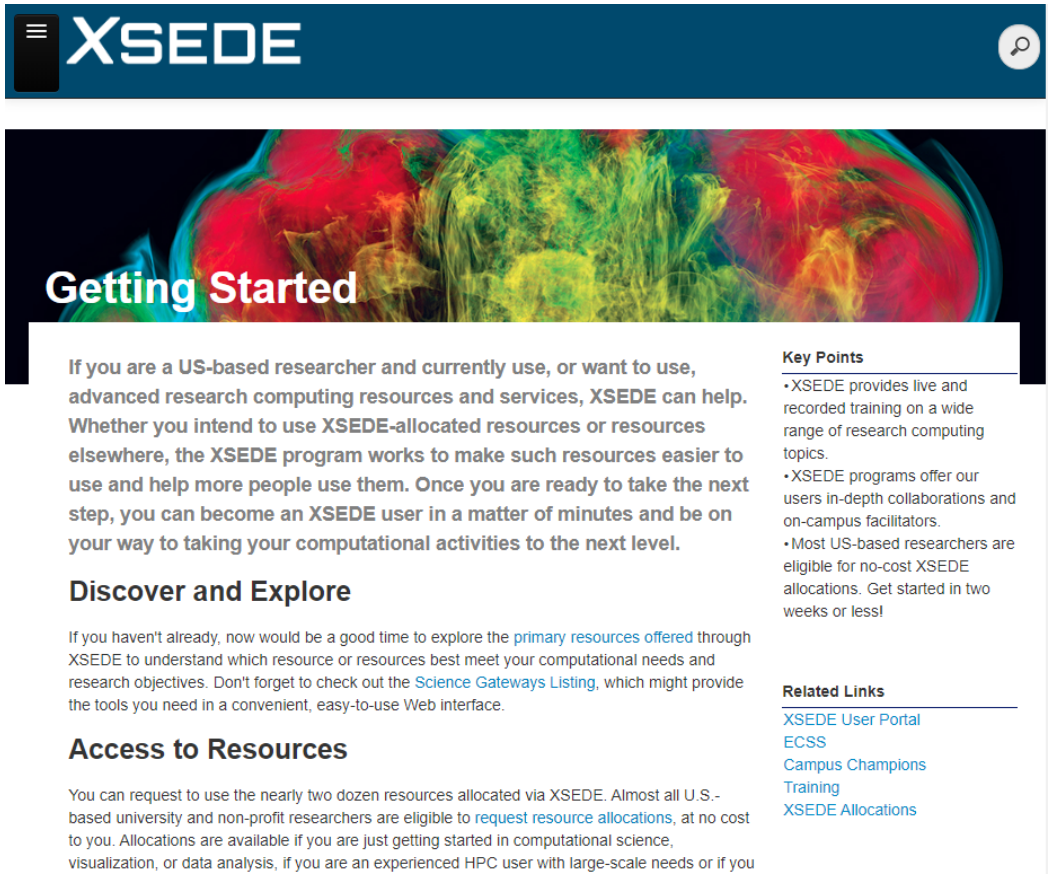
  (more on Tapis later)

- CyVerse □s its community developers!



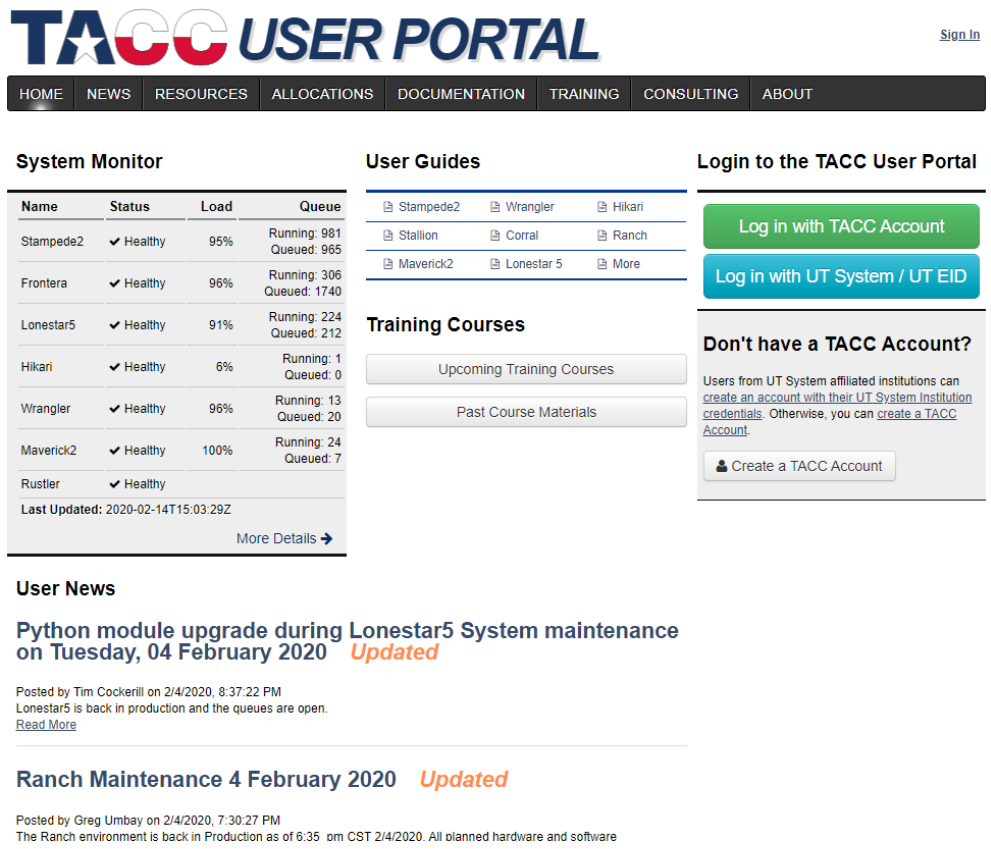
Getting Access to HPC Resources

- **Stampede2** and **Jetstream** access allocated through XSEDE
- Freely available to all US-based researchers and their collaborators
- “Startup” allocations are simple to apply for



https://portal.xsede.org/allocations/startup

Getting Access to HPC Resources



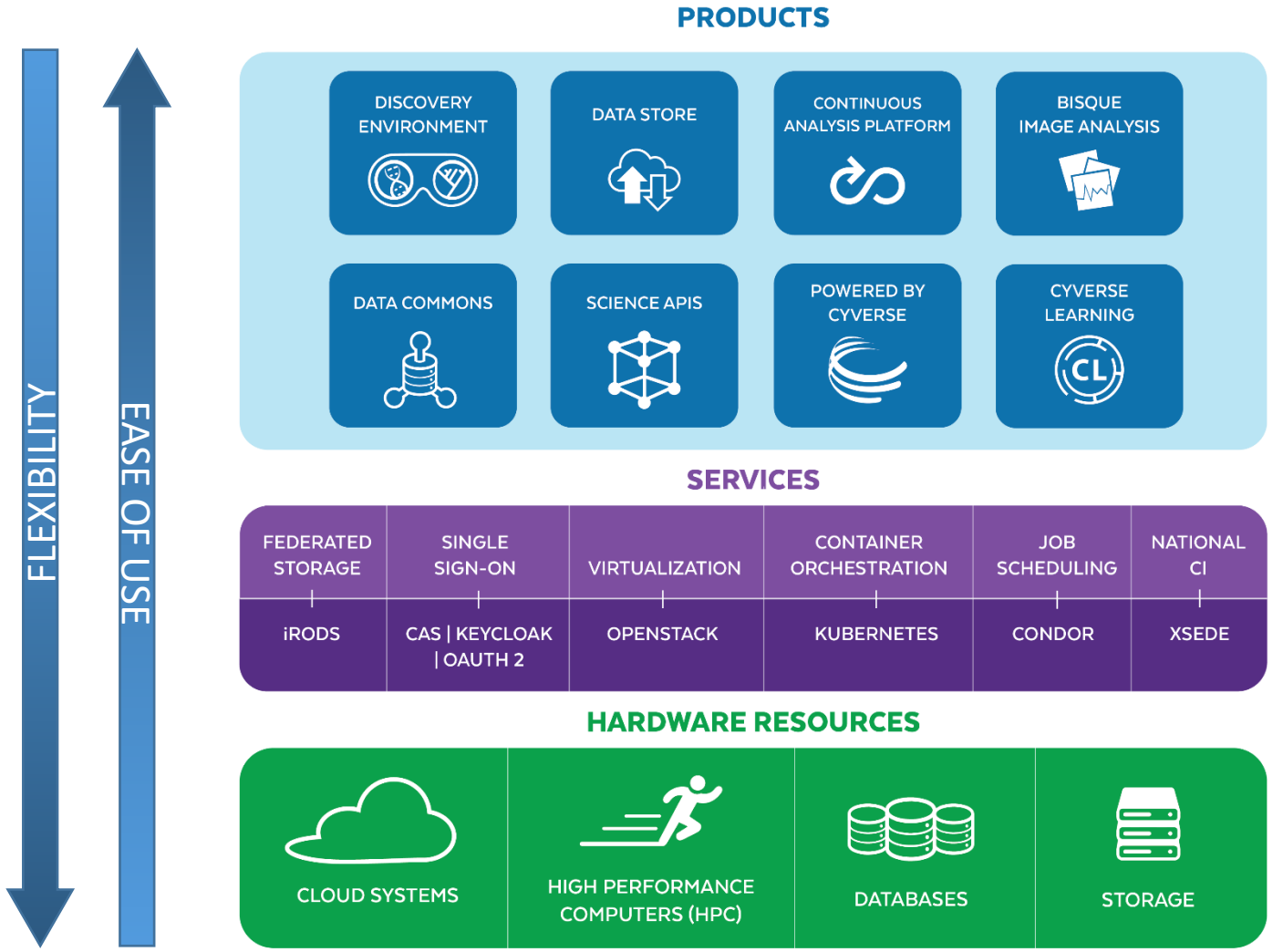
- Most Texas-based researchers can also apply for allocations directly on the TACC Portal for **Stampede2**, **Lonestar5**, **Wrangler**, and other systems

https://portal.tacc.utexas.edu/

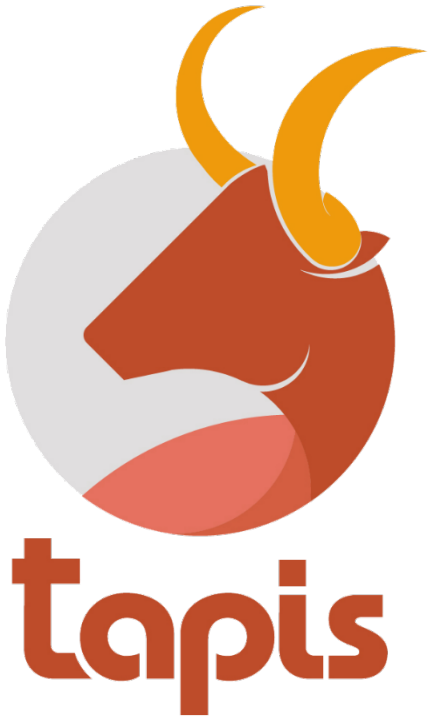
Getting Access to HPC Resources

- Direct command-line access
  - Maximum flexibility…
  - …and the steepest learning curve
- Best for developers and advanced users
- Lots of learning resources available!
  - https://learn.tacc.utexas.edu/
  - https://portal.tacc.utexas.edu/

The Anatomy of Cyberinfrastructure



Science APIs



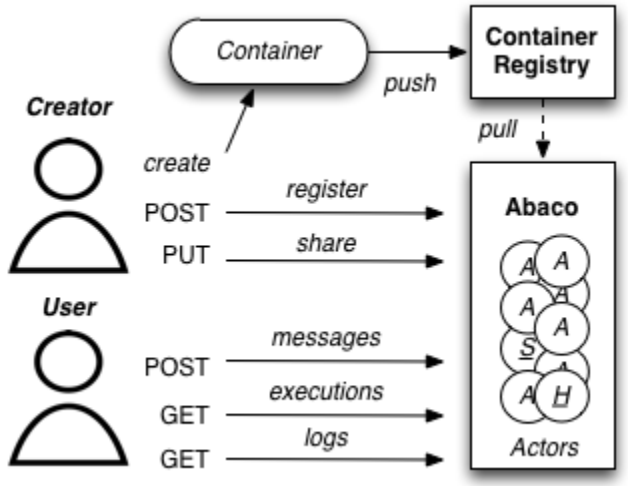
- “Application Programming Interfaces” are machine-readable services (that humans can mostly read too)
- Connects the Discovery Environment and other Powered by CyVerse sites to HPC, data storage, and private clusters
- Platform for deploying and sharing Apps
- Functions as a Service APIs for automation
- https://www.tacc.utexas.edu/tapis

Custom Apps

- The essential components you need to create your own app are:
  1. An app bundle directory containing definitions and assets for the app
  2. A Docker/Singularity image containing the executable and all runtime dependencies
  3. A wrapper script (generally written in bash) that runs the container or executable
- The result is a new tool that is versioned, shareable, and available to run in the Discovery Environment or elsewhere

Deploying Actors

- Functions as a Service model
- Messages can be sent to actors, and the content of that message is passed into the container as an environment variable
- Actors can call other actors, Tapis apps, or third party services
- Intended for automation, orchestration, and small compute tasks

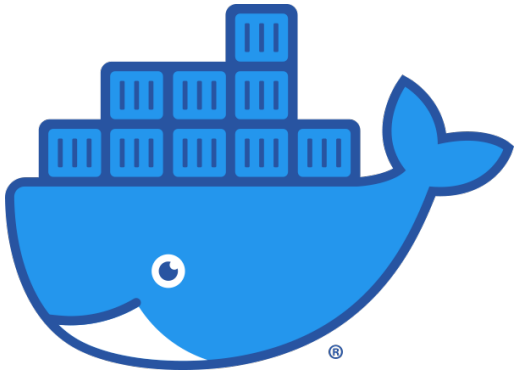


Developing with Science APIs

- Automation, integration, and extensibility for advanced users and developers
- Command line toolkit: https://github.com/TACC-Cloud/tapis-cli-ng
- Command line docs: https://tapis-cli.readthedocs.io/
- API documentation: https://tacc-cloud.readthedocs.io/
- Slack community: https://tacc-cloud.slack.com/
- Slack invite link: https://bit.ly/join-tapis

Containers

- PSA: Containers are one of the core building blocks for scientific code
- TACC hosts and runs 5,000+ containers on our resources (just staff!)
- Docker for building and deploying containers to the cloud
- Singularity for deploying containers to HPC
- The container runtime becomes the only dependency



CyVerse Container Camp

Join us for an intensive three-day hands-on workshop to learn how to create, use, deploy and share software containers (e.g., Docker, Singularity) across a variety of computational systems from your laptop to cloud and HPC.

https://cyverse.org/cc

CyVerse is supported by the National Science Foundation under Grants No. DBI-0735191, DBI-1265383 and DBI-1743442.