



**CYVERSE™**

Transforming Science Through Data-driven Discovery

# Fall in Love with High Performance Computing at CyVerse

John Fonner, Texas Advanced Computing Center

February 14<sup>th</sup>, 2020 □



# The Vision of CyVerse

“Transforming Science through Data-Driven Discovery”

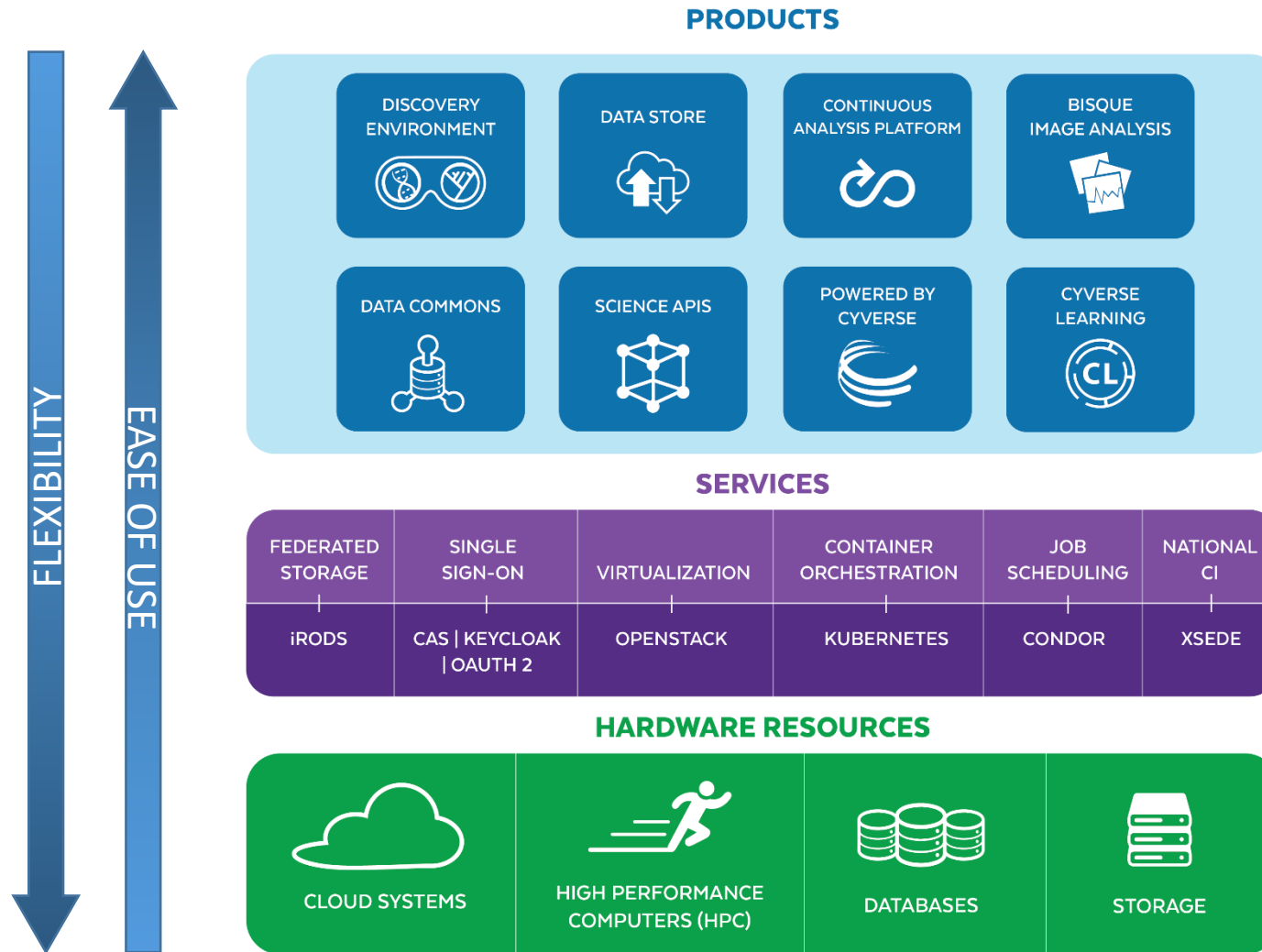
Deliver cyberinfrastructure for life sciences research...

(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](#)



CyVerse account



TACC account



XSEDE account



# What is the Texas Advanced Computing Center?



# TACC at a Glance

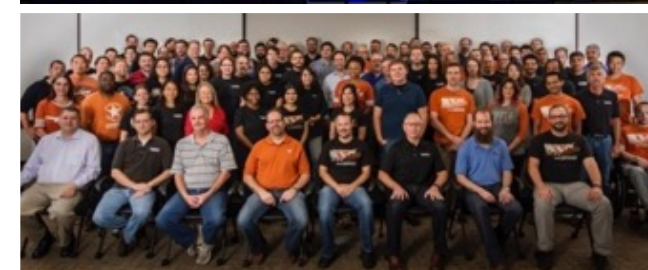
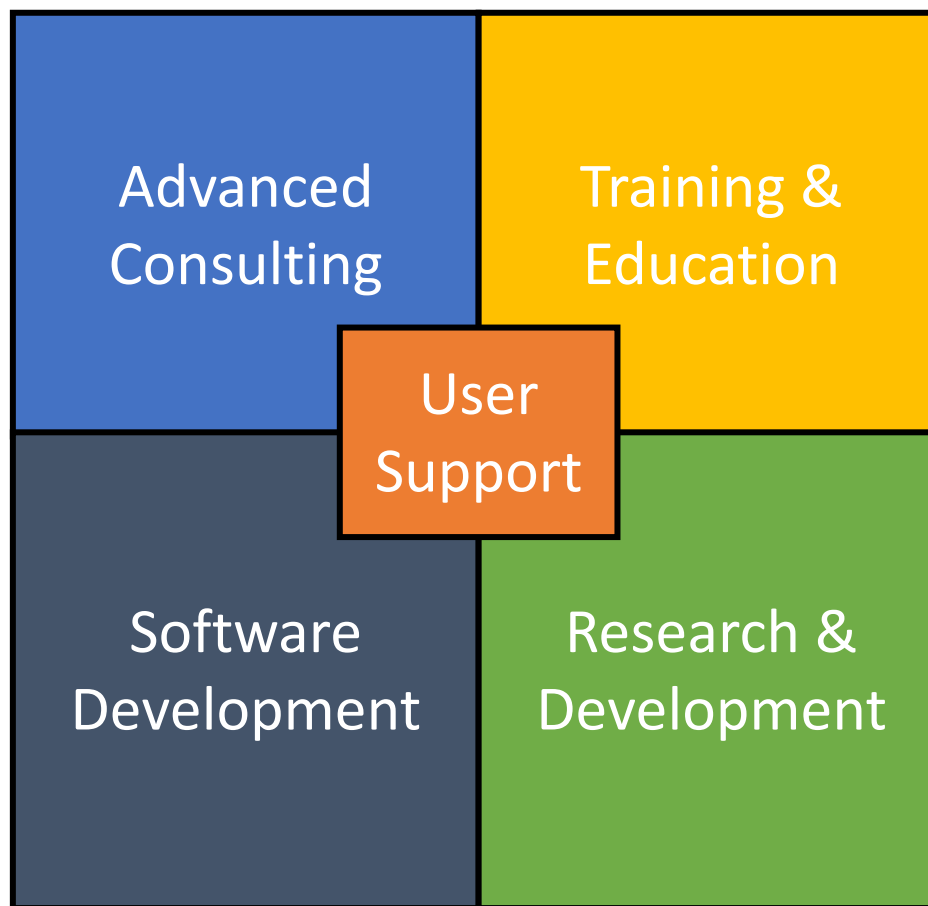
185 staff members  
- 70 PhDs, ~25 students  
- 1 in 3 have 10+ years exp

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>

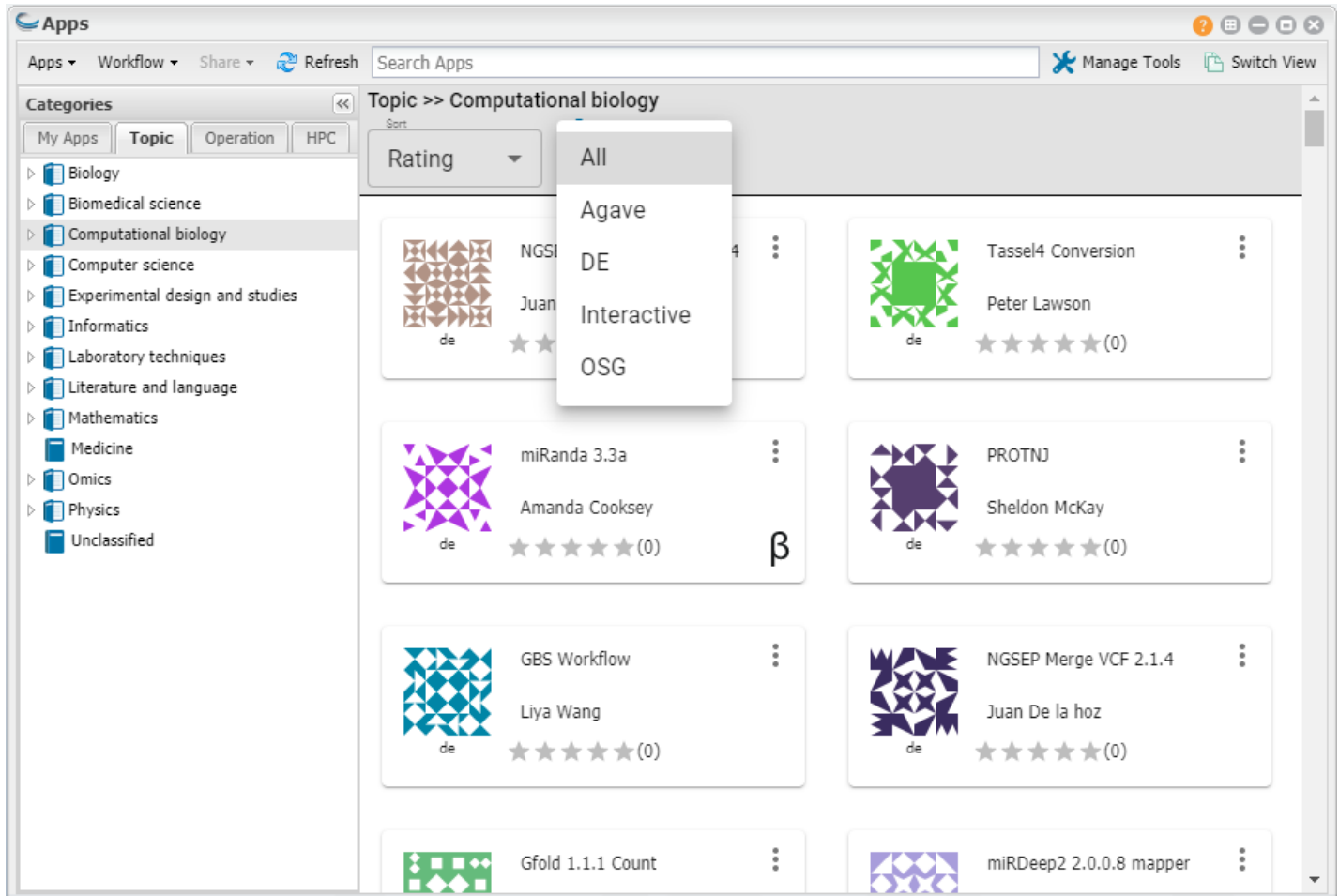


# Getting Access to HPC Resources

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

(more on Tapis later)

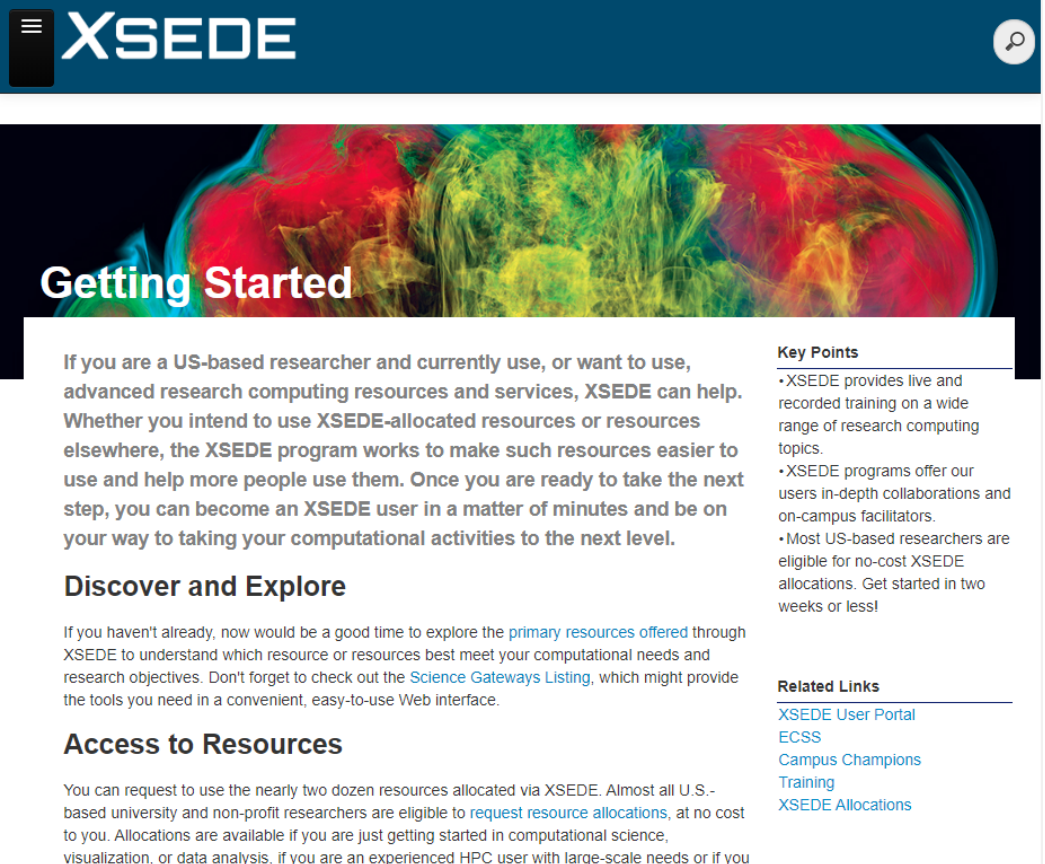
- CyVerse  its community developers!





# 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



The screenshot shows the XSEDE website header with the logo and a search icon. Below the header is a banner image with the text "Getting Started". The main content area is divided into three columns: a large introductory paragraph, a "Discover and Explore" section, and an "Access to Resources" section. On the right side, there are two sidebars: "Key Points" and "Related Links".

**XSEDE**

## Getting Started

If you are a US-based researcher and currently use, or want to use, advanced research computing resources and services, XSEDE can help. Whether you intend to use XSEDE-allocated resources or resources elsewhere, the XSEDE program works to make such resources easier to use and help more people use them. Once you are ready to take the next step, you can become an XSEDE user in a matter of minutes and be on your way to taking your computational activities to the next level.

### Discover and Explore

If you haven't already, now would be a good time to explore the [primary resources offered](#) through XSEDE to understand which resource or resources best meet your computational needs and research objectives. Don't forget to check out the [Science Gateways Listing](#), which might provide the tools you need in a convenient, easy-to-use Web interface.

### Access to Resources

You can request to use the nearly two dozen resources allocated via XSEDE. Almost all U.S.-based university and non-profit researchers are eligible to [request resource allocations](#), at no cost to you. Allocations are available if you are just getting started in computational science, visualization, or data analysis, if you are an experienced HPC user with large-scale needs or if you

#### Key Points

- XSEDE provides live and recorded training on a wide range of research computing topics.
- XSEDE programs offer our users in-depth collaborations and on-campus facilitators.
- Most US-based researchers are eligible for no-cost XSEDE allocations. Get started in two weeks or less!

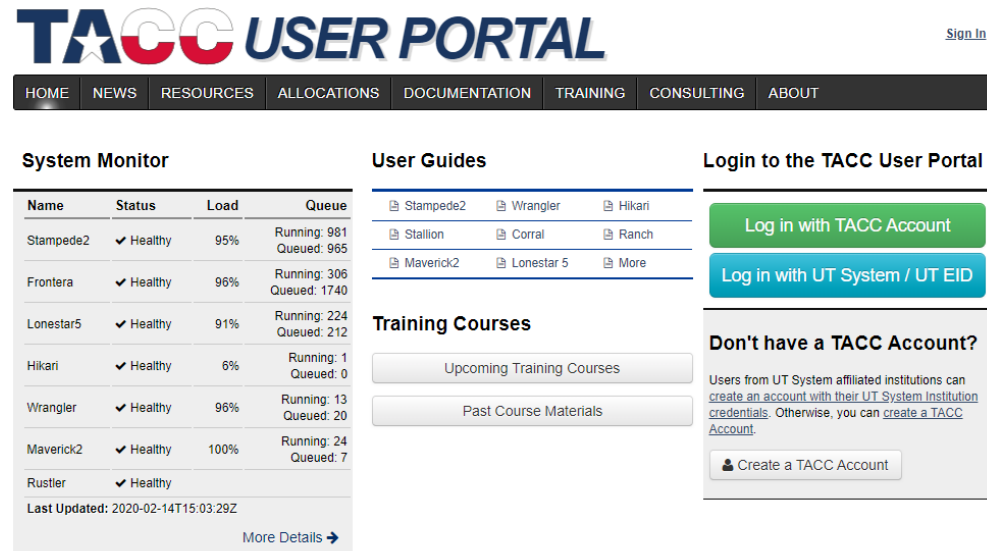
#### Related Links

- [XSEDE User Portal](#)
- [ECSS](#)
- [Campus Champions Training](#)
- [XSEDE Allocations](#)

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



# Getting Access to HPC Resources



The screenshot shows the TACC User Portal interface. At the top left is the TACC logo and the text 'USER PORTAL'. To the right is a 'Sign In' link. Below this is a navigation menu with links for HOME, NEWS, RESOURCES, ALLOCATIONS, DOCUMENTATION, TRAINING, CONSULTING, and ABOUT. The main content area is divided into three columns: 'System Monitor', 'User Guides', and 'Login to the TACC User Portal'. The 'System Monitor' column contains a table with columns for Name, Status, Load, and Queue. The 'User Guides' column has links for Stampede2, Wrangler, Hikari, Stallion, Corral, Ranch, Maverick2, Lonestar 5, and More. The 'Login to the TACC User Portal' column has buttons for 'Log in with TACC Account' and 'Log in with UT System / UT EID', and a section for 'Don't have a TACC Account?' with a 'Create a TACC Account' button. Below the main content is a 'User News' section with two news items: 'Python module upgrade during Lonestar5 System maintenance on Tuesday, 04 February 2020' and 'Ranch Maintenance 4 February 2020'.

Name	Status	Load	Queue
Stampede2	✓ Healthy	95%	Running: 981 Queued: 965
Frontera	✓ Healthy	96%	Running: 306 Queued: 1740
Lonestar5	✓ Healthy	91%	Running: 224 Queued: 212
Hikari	✓ Healthy	6%	Running: 1 Queued: 0
Wrangler	✓ Healthy	96%	Running: 13 Queued: 20
Maverick2	✓ Healthy	100%	Running: 24 Queued: 7
Rustler	✓ Healthy		

- 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

```
1 stampede2:login1 +
-----
Welcome to the Stampede2 Supercomputer
Texas Advanced Computing Center, The University of Texas at Austin
-----

** Unauthorized use/access is prohibited. **

If you log on to this computer system, you acknowledge your awareness
of and concurrence with the UT Austin Acceptable Use Policy. The
University will prosecute violators to the full extent of the law.

TACC Usage Policies:
http://www.tacc.utexas.edu/user-services/usage-policies/

-----
Welcome to Stampede2, *please* read these important system notes:

--> Stampede2 user documentation is available at:
https://portal.tacc.utexas.edu/user-guides/stampede2

----- Project balances for user jfonner -----
| Name      Avail SUs  Expires | Name      Avail SUs  Expires |
| RepServer    2500  2020-03-31 | DARPA-ProteinDesign  17131  2020-12-31 |
| UT-2015-05-18  1977  2020-03-31 | CSRA-CDC-AMD      500  2020-09-28 |
| DARPA-SB4D    2484  2020-09-30 | SD2E-Community    21284  2020-09-30 |
| iPlant-Collabs 12190  2020-03-31 | DARPA-SNCS        2000  2020-09-30 |
| iPlant-Master  1868  2023-08-31 | DARPA-TERRA       1000  2020-09-30 |
----- Disk quotas for user jfonner -----
| Disk      Usage (GB)  Limit  %Used  File Usage  Limit  %Used |
| /home1    6.0         10.0   59.59  35045      200000  17.52 |
| /work     592.8      1024.0 57.89  997298    3000000 33.24 |
| /scratch  7.9         0.0    0.00   33680     0        0.00 |
-----

Tip 113 (See "module help tacc_tips" for features or how to disable)

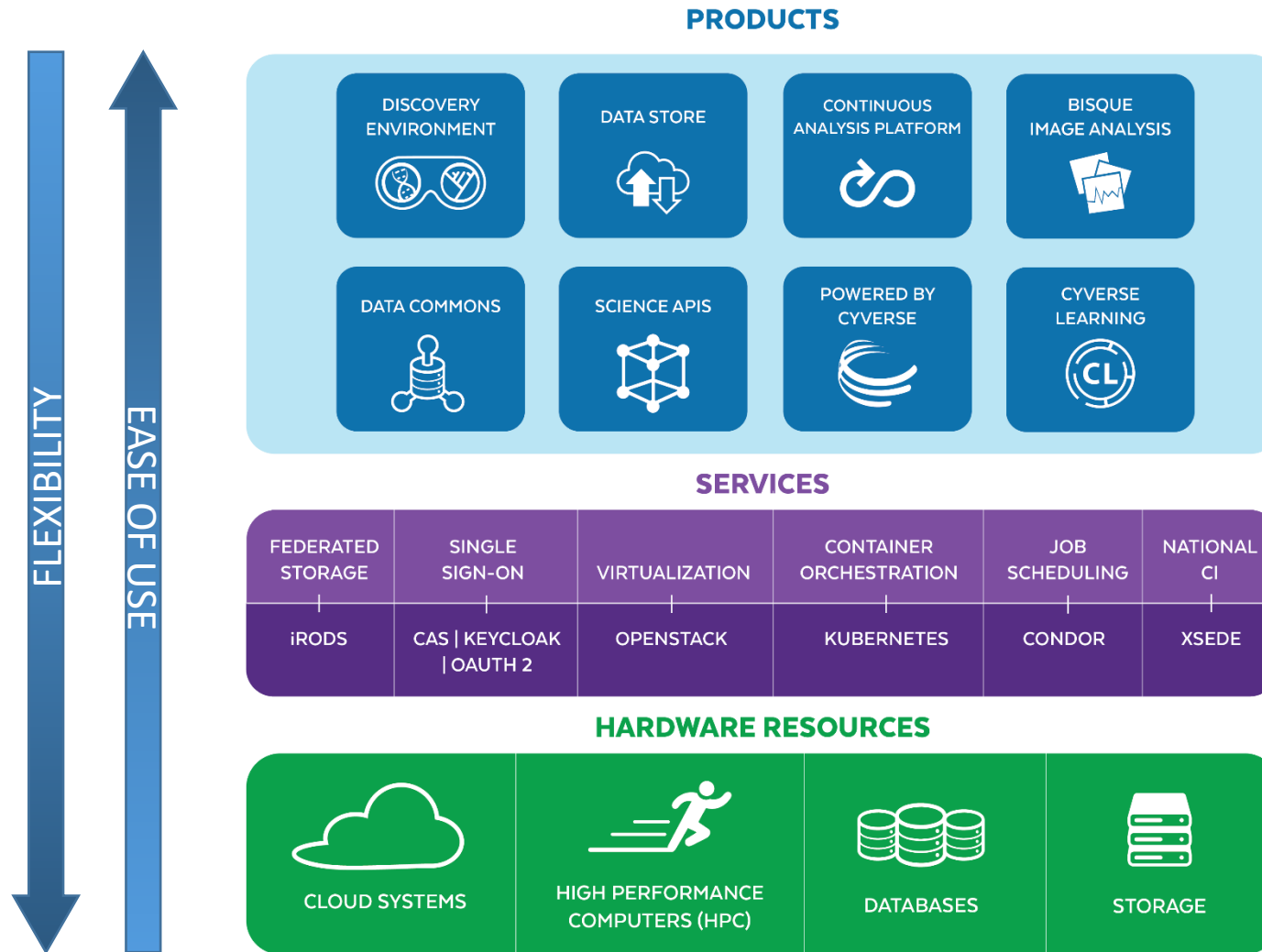
To diff two variables do: "diff <(echo $a) <(echo $b)"

jfonner@login1:~$ █
```

- 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>

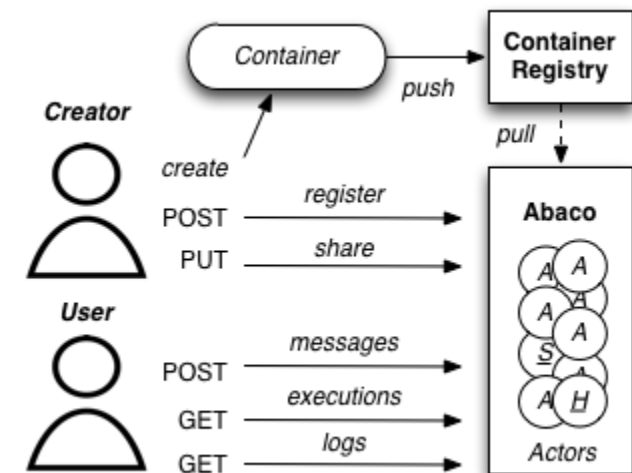
# Custom Apps

- 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

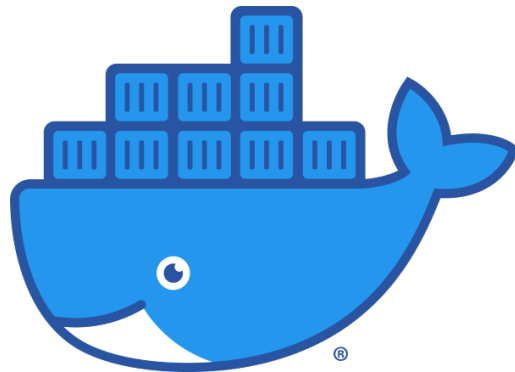
- Automation, integration, and extensibility for advanced users and developers
- Command line toolkit: <https://github.com/TACC-Cloud/tapis-cli-ng>
- Command line docs: <https://tapis-cli.readthedocs.io/>
- API documentation: <https://tacc-cloud.readthedocs.io/>
- Slack community: <https://tacc-cloud.slack.com/>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

