

rt.iu.edu



RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES



RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES

Jetstream Overview: A national research and education cloud

ECSS Webinar
Jan 19, 2021 – Bloomington, IN.

*Jeremy Fischer – Jeremy@iu.edu - Indiana University
Manager, Jetstream Cloud, UITS Research Technologies*

*Fischer, J. (2021). Jetstream Overview: A national research and
education cloud. Bloomington, IN. Retrieved from
<https://jetstream-cloud.org/research/publications.php>*

Code of Conduct

XSEDE has an external code of conduct which represents our commitment to providing an inclusive and harassment-free environment in all interactions regardless of race, age, ethnicity, national origin, language, gender, gender identity, sexual orientation, disability, physical appearance, political views, military service, health status, or religion. The code of conduct extends to all XSEDE-sponsored events, services, and interactions.

Code of Conduct: <https://www.xsede.org/codeofconduct>

Contact:

- Event organizer: **Bob Sinkovits** (sinkovit@sdsc.edu)
- XSEDE ombudspersons:
 - Linda Akli, Southeastern Universities Research Association (akli@sura.org)
 - Lizanne Destefano, Georgia Tech (lizanne.destefano@ceismc.gatech.edu)
 - Ken Hackworth, Pittsburgh Supercomputing Center (hackworth@psc.edu)
 - Bryan Snead, Texas Advanced Computing Center (jbsnead@tacc.utexas.edu)
- Anonymous reporting form available at <https://www.xsede.org/codeofconduct>.



Terminology Statement

In line with XSEDE's Code of Conduct, XSEDE is committed to providing training events that foster inclusion and show respect for all. This commitment applies not only to how we interact during the event; it also applies to the training materials and presentation. It is not XSEDE's position to use, condone, or promote offensive terminology.

XSEDE instructors strive to keep inclusive language at the forefront. In the event that we have included inappropriate materials, verbal or written, please let us know at terminology@xsede.org

While XSEDE has no control over external third-party documentation, we are taking steps to effect change by contacting the relevant organizations; we hope this will be addressed by all third parties soon.

*If you see any terminology concerns in the following presentation or slides, we want to know!
Please contact the Terminology Task Force: terminology@xsede.org*



XSEDE

What is Jetstream(1) and why does it exist?

- *NSF's first production cloud facility*
- *Focus on ease-of-use, broad accessibility*
- *User-selectable library of preconfigured virtual machines*
- *Provides **on-demand** interactive computing and analysis or persistent services such as gateways*
- *Enables configurable environments; **programmable cyberinfrastructure***



Who uses Jetstream(1) today?

- *The researcher needing a handful of cores (1 to 44/vCPU)*
- *Software creators and researchers needing to create their own VMs and workflows*
- *Science gateway creators using Jetstream as either the frontend or processor for scientific jobs*
- *STEM Educators teaching on a variety of subjects*

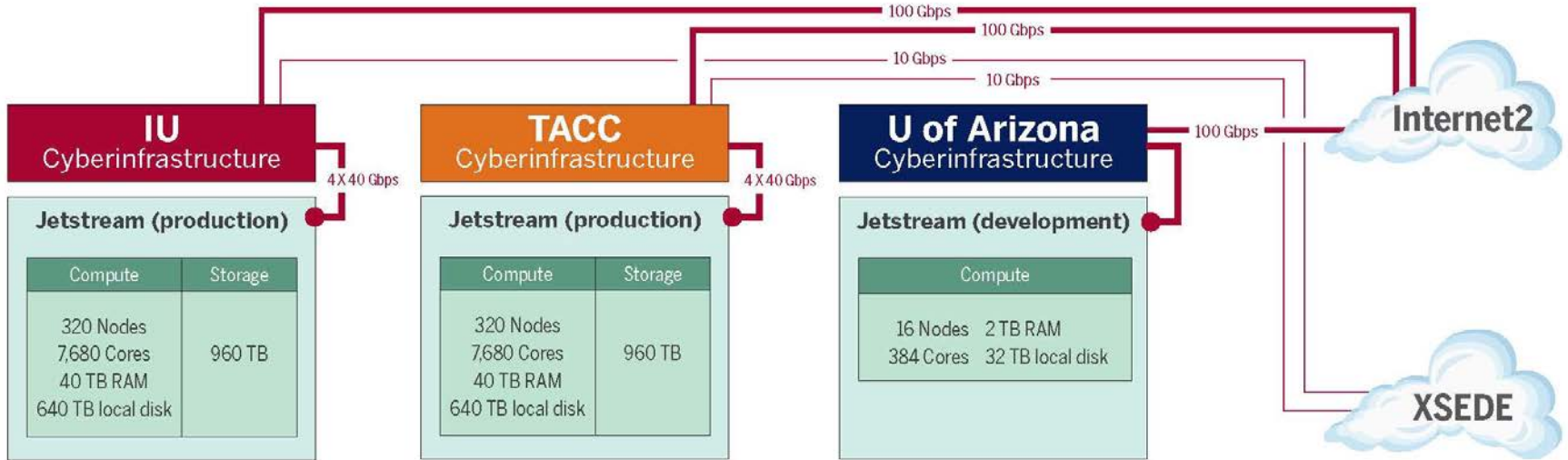


What Jetstream isn't...

- *It's not traditional HPC*
- *There's no shared filesystem (think cloudy!)*
- *There's no high-end interconnect fabric (keep thinking cloudy!)*
- *There aren't GPUs (well...)*
- *It isn't Amazon, Azure, or GCE (similar, but...)*



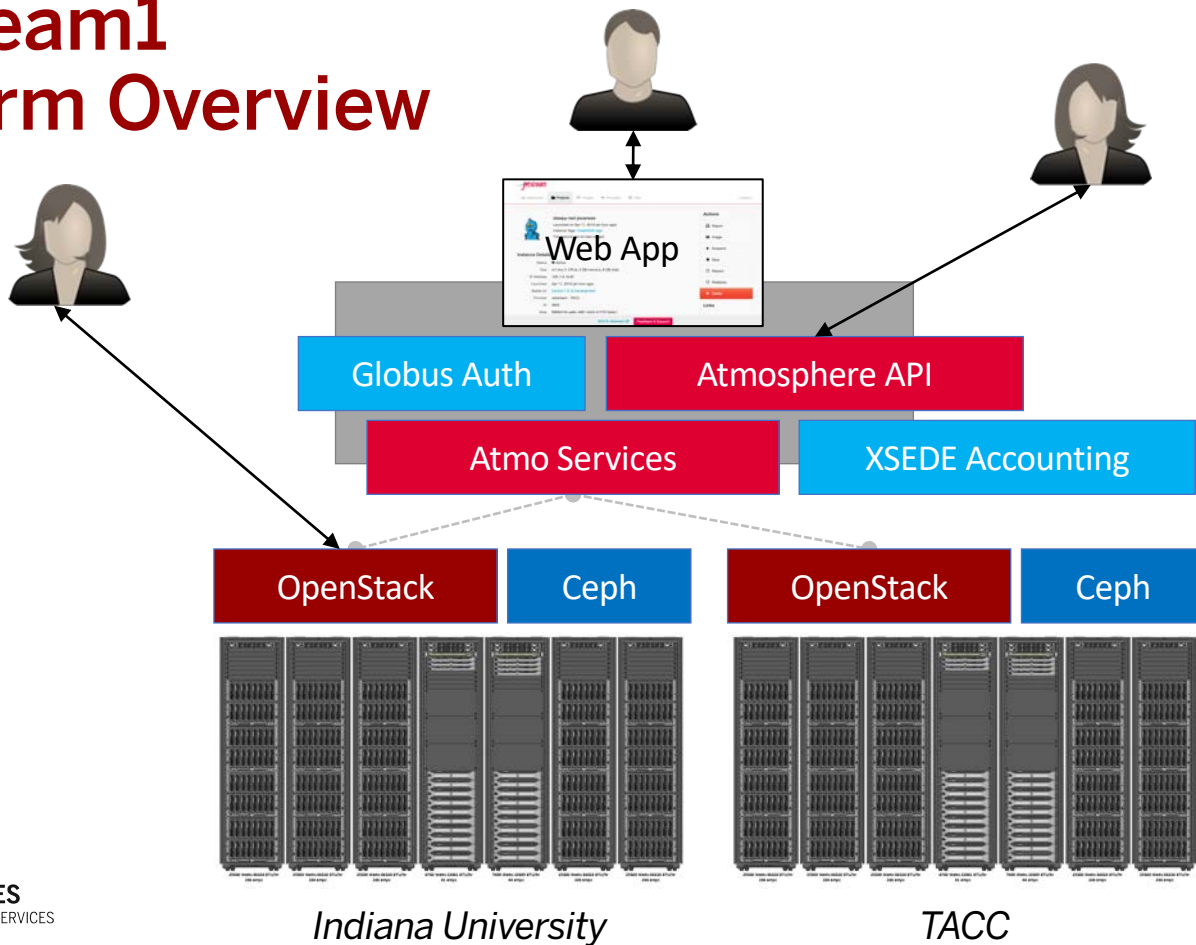
Jetstream1 System Overview



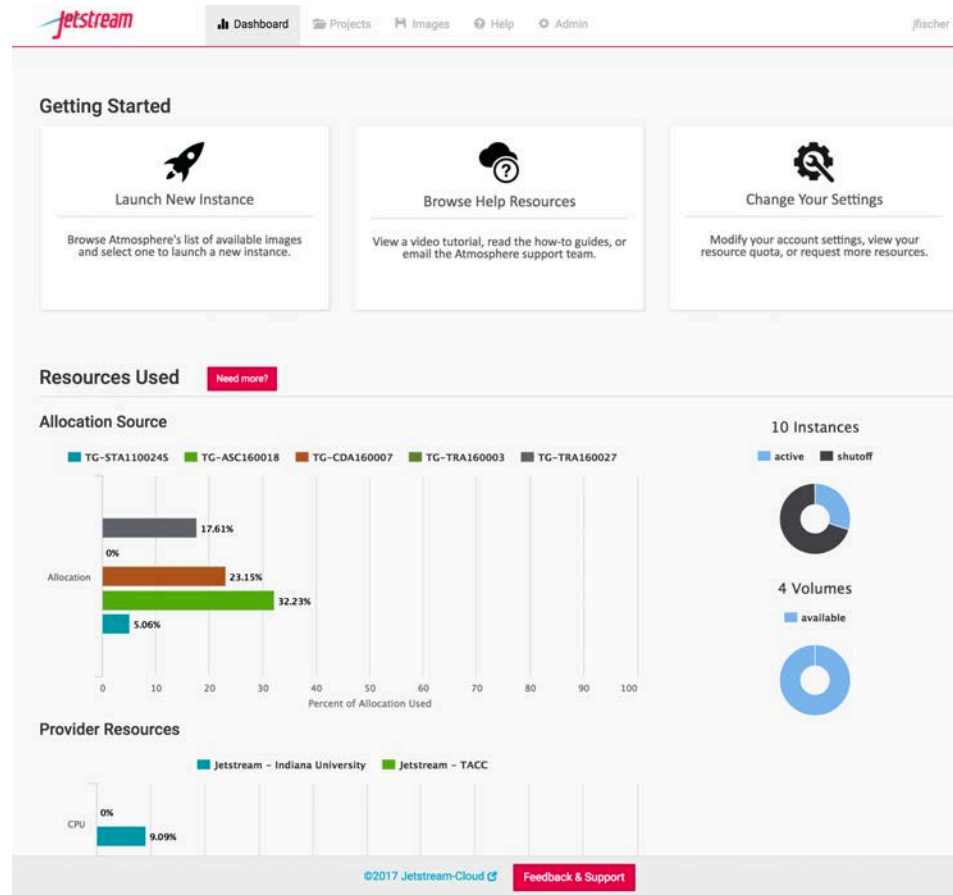
<http://wiki.jetstream-cloud.org/Network+configuration+and+policies>



Jetstream1 Platform Overview



The Jetstream1 Atmosphere web interface



API Access to Jetstream

- *What was unexpected*
 - *Demand for programmable cyberinfrastructure*
 - *Great platform for learning system administration skills*
 - *Great platform for teaching & learning cloudy technologies*
- *Command line clients*
- *Horizon dashboard very popular; but, incomplete*
- *Programmatic control; python is popular*
(<https://docs.openstack.org/openstacksdk/latest/>)
- *Slack channel for collaboration API users of Jetstream*
- *Paves the way for 3rd party interfaces like Exosphere*



Using the OpenStack CLI on Jetstream

What an openrc file looks like:

```
export OS_AUTH_URL=https://iu.jetstream-cloud.org:35357/v3
export OS_PROJECT_NAME="TG-ABC190028"
export OS_USER_DOMAIN_NAME="tacc"
export OS_USERNAME="taccusername"
export OS_IDENTITY_API_VERSION=3
# export OS_PASSWORD='string'
read -sr OS_PASSWORD_INPUT
export OS_PASSWORD=$OS_PASSWORD_INPUT
```

- CLI is python based – reads this information from the environment.
- Horizon can generate an openrc file for you (see the Wiki docs)
- **Common pitfall** – make sure you specify the correct Project (allocation) if you have more than one!



Horizon GUI interface

- *Allows most things you can do from the CLI*
- *Nice for some tasks*
 - *Network visualizer is something we tend to use as a troubleshooting tool*
 - *Easier to look at security groups on Horizon (IMHO)*
- *Downsides:*
 - *considerably slower than using CLI*
 - *not all features are present that are in CLI*
 - *can't do things programmatically*



The screenshot shows the OpenStack Dashboard login page. At the top, there is the OpenStack logo (a red cube) and the text "openstack" in a sans-serif font, with "DASHBOARD" in a smaller font below it. Below the logo is the heading "Log In". There are three input fields: "Domain" with the value "TACC", "User Name" (empty), and "Password" (empty). Each input field has a small icon on the right side. At the bottom right of the form is a blue button labeled "Connect".



M87 black hole: how cloud computing supports astronomy



Event Horizon Telescope

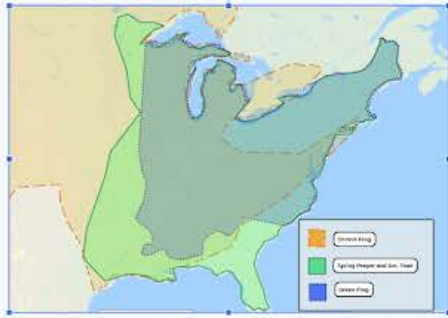


*M87 black hole image
generated by EHT*

- *Event Horizon Telescope (a telescope array consisting of a global network of radio telescopes), a large number of scientists, NASA spacecraft, and a variety of computing resources enabled the first image of a black hole.*
- *For the M87 black hole image, two critical steps were done in the cloud and piloted on Jetstream*
 - *correcting for anomalies, so that further image processing could occur, and*
 - *large survey study of how image reconstruction algorithms affect the final images.*
- *The team is also developing new methods to correlate data from multiple telescopes (to reduce data from petascale to terascale) in the cloud*



AI for Everyone – Recognizing Frog Calls



American Toad



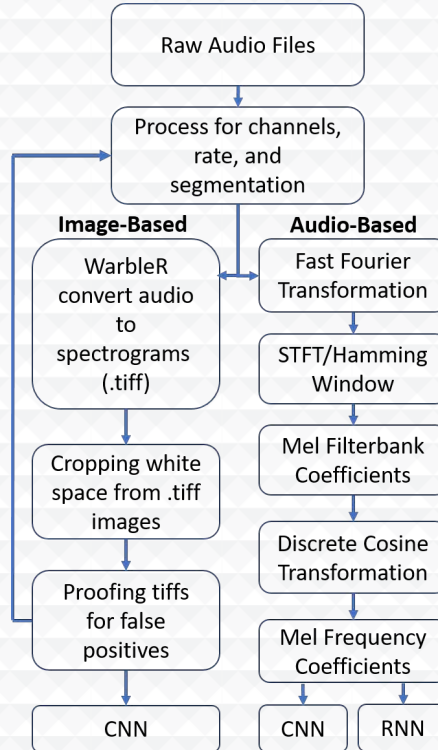
Spring Peeper



Chorus Frog



Green Frog



2019 Jetstream REU participants examined best-practices for supporting AI projects for field biologists



Brainlife.io is a science gateway for neuroscience analysis.

Allows creation of custom workflows that can be saved and shared

Began using only Jetstream and other XSEDE resources and has grown to use

Expanded to use Microsoft's Azure cloud via the Midwest Big Data Hub

Backed with multiple virtual Slurm clusters on IU and TACC clouds



Jetstream for education – in action at AMS2020

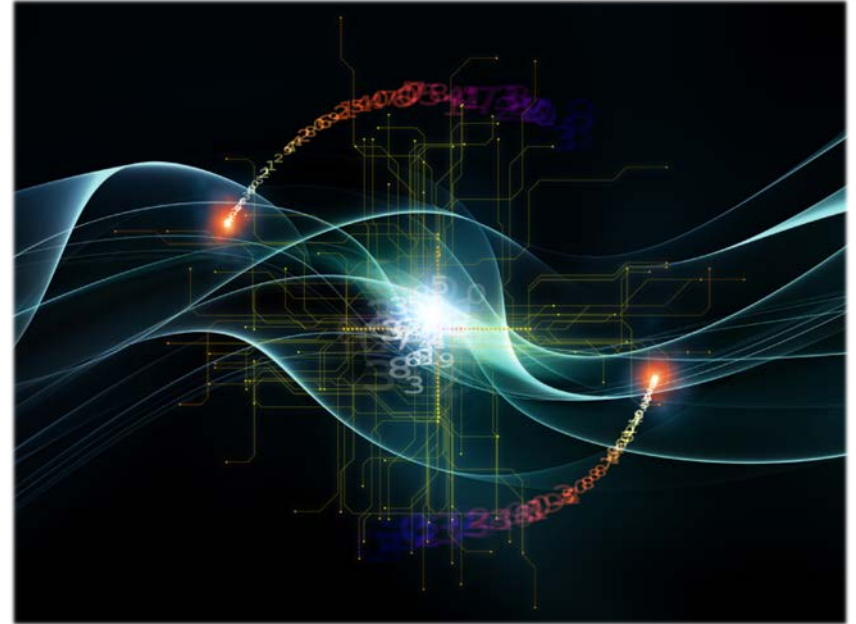
- Unidata-led workshop at American Meteorological Society (AMS) 2020 conference
- 127 users actively participating
- Participants used a JupyterHub running on Jetstream (40 node Kubernetes cluster of 6 core m1.medium VMs) for a 90 minute Unidata PyAOS (Python for the Atmospheric and Oceanic Sciences) workshop
- The students were successfully able to run their interactive Python code notebooks as the instructors presented their material



Not just the usual suspects...

Physics, chemistry, and other “usual” HPC suspects are represented, but Jetstream also is home to projects on:

- Financial analysis / Economics
- Political science
- Humanities / Text analysis
- Network analysis
- Computer Science / Machine learning
- Satellite data analysis



What worked?

- Allowing API access and full control (root privileges)
- Allowing allocations to run continuously – as long as the PI renewed – allowing workflows to run indefinitely
- Development of trial allocations



What didn't work?

- Forcing small allocations into the research allocation process
- Lack of multi-year allocations
- Lack of shared data set storage
- Trying to use accounting systems built to handle HPC accounting workloads



Jetstream usage highlights – 1 Jan 2020

- 395 XSEDE projects covering 56 fields of science and almost 2400 active users representing 405 institutions
- **80%** of Jetstream users have **not used any other XSEDE system**
- >373M CPU hours allocated to XSEDE projects since June 2016
- 49 active science gateways
- 46 education/teaching allocations serving over 850 students
- 1189 mean active VMs in previous qtr, 1632 peak active VM count
- **Highest** user satisfaction in most recent XSEDE survey

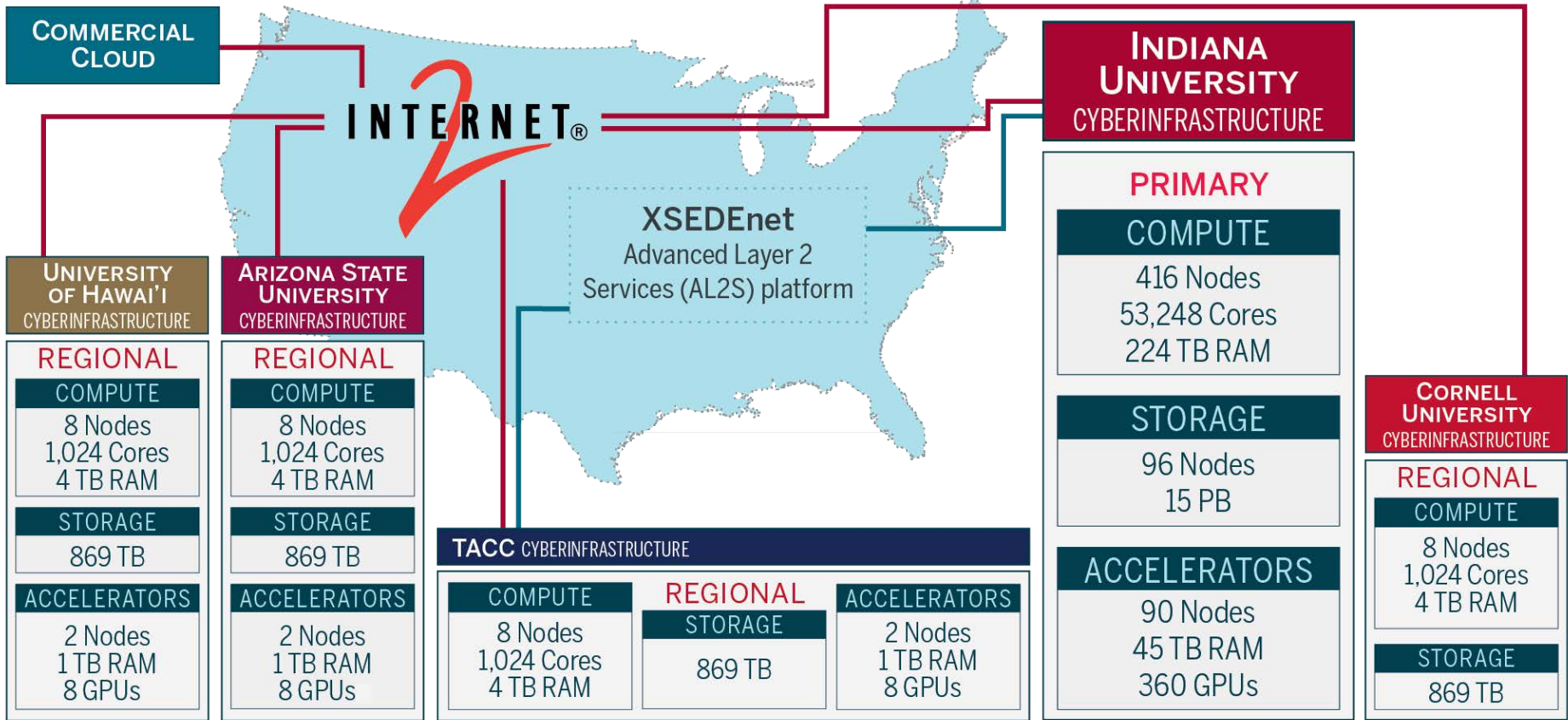


And introducing...

Jetstream2

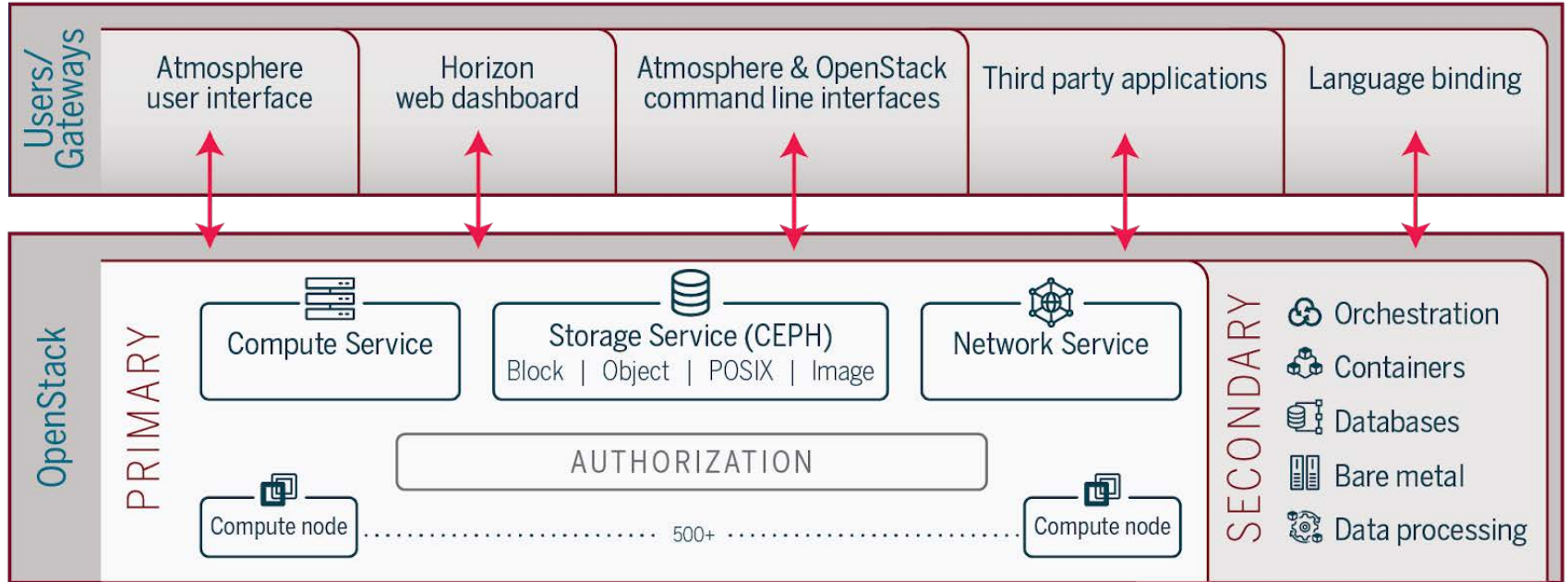


RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES



Jetstream2 Proposed Architecture

Conceptual Jetstream2 Architecture



What improvements are planned?

- *Improving access to higher level orchestration*
- *Improving documentation and training for orchestration*
- *Implementing “push button” virtual clusters*
- *Federating JupyterHubs and making the implementation of JupyterHubs a simple process*
- *Creating a shared application service for VMs to make common scientific software more accessible*
- *Improved storage access, including object storage and storage that is sharable between VMs in the same allocation*



Future Plans with Jetstream2

- Focusing on **programmable cyberinfrastructure** using technologies like Terraform to make creating infrastructure easy on Jetstream2, commercial clouds, or other private clouds
- Making enhanced container support for interoperability a priority
- Planned collaborations with commercial clouds:
- AWS to provide workshops on cloud interoperability
- Bursting to Azure via on-premises data gateway
- Implementation of Google's Cloud Service Platform (allowing management of hybrid cloud environments via gcloud CLI or Google GUI).
- Interactive GPU access and the ability to have long-running training for AI workloads



Where can I get help?

- **Wiki / Documentation:** <http://wiki.jetstream-cloud.org>
- *API CLI Tutorial:* <https://github.com/jlf599/JetstreamAPITutorial>
- *User guides:* <https://portal.xsede.org/user-guides>
- *XSEDE KB:* <https://portal.xsede.org/knowledge-base>
- *Email:* help@xsede.org



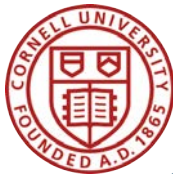
Acknowledgements

NSF Awards 1053575 & 1548562 (XSEDE), 1445604 (Jetstream) and 2005506 (Jetstream2)

This document was developed with support from the National Science Foundation. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.



Jetstream Partners



funded by the National Science Foundation
Award #ACI-1445604



Jetstream2 partners

