



**PERVASIVE
TECHNOLOGY INSTITUTE**



RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES



RESEARCH TECHNOLOGIES

UNIVERSITY INFORMATION TECHNOLOGY SERVICES

Jetstream Overview: A national research and education cloud

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

Unidata User Committee Meeting – Nov 16, 2020

What is Jetstream 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?

- *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*



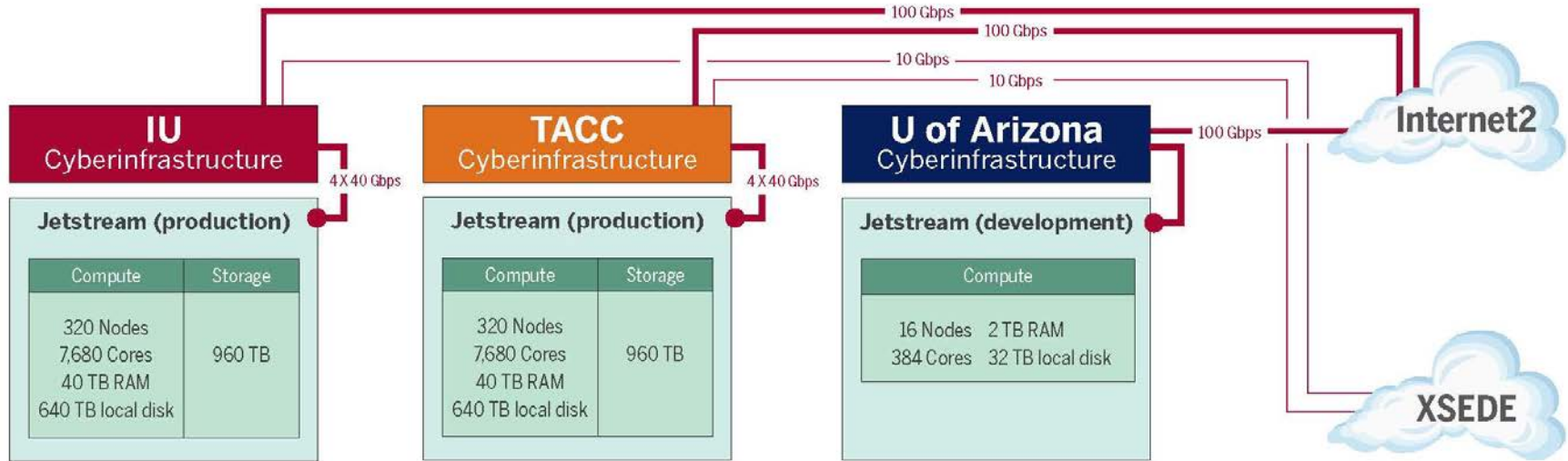
What Jetstream isn't...

- *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 (yet...they're coming!)*
- *It isn't Amazon, Azure, or GCE (similar, but...)*



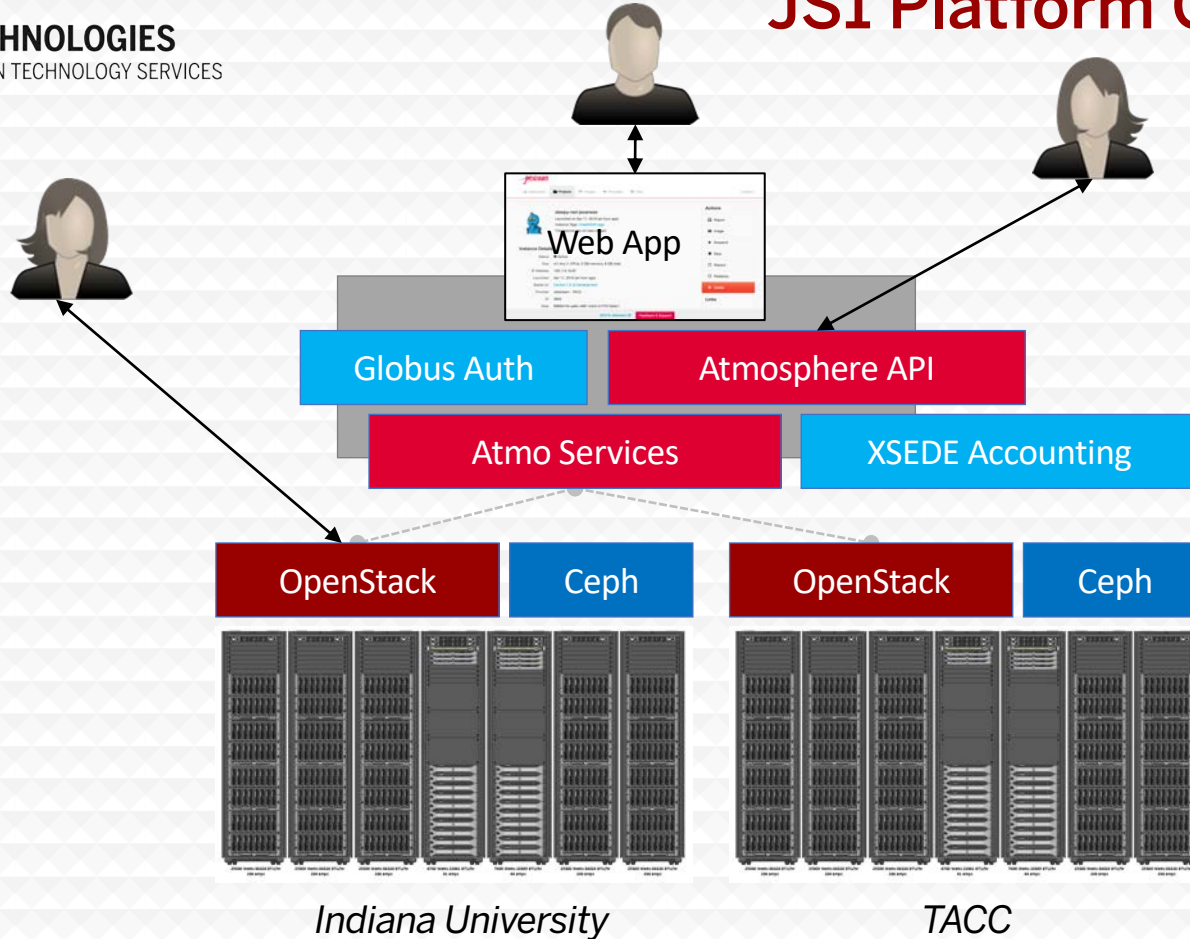


Jetstream 1 Architecture





JS1 Platform Overview





The Jetstream Atmosphere web interface

The screenshot shows the Jetstream Atmosphere web interface dashboard. At the top, there is a navigation bar with the Jetstream logo and menu items: Dashboard, Projects, Images, Help, and Admin. The user's name 'jfischer' is visible in the top right corner. The main content area is titled 'Getting Started' and contains three cards: 'Launch New Instance' (with a rocket icon), 'Browse Help Resources' (with a question mark icon), and 'Change Your Settings' (with a gear icon). Below this is the 'Resources Used' section, which includes a 'Need more?' button and an 'Allocation Source' bar chart. The bar chart shows the percentage of allocation used by five different sources: TG-STA1100245 (5.06%), TG-ASC160018 (32.23%), TG-CDA160007 (23.15%), TG-TRA160003 (17.61%), and TG-TRA160027 (0%). To the right of the bar chart are two donut charts: '10 Instances' (with 'active' and 'shutoff' categories) and '4 Volumes' (with 'available' category). Below the bar chart is the 'Provider Resources' section, which shows a bar chart for CPU usage: Jetstream - Indiana University (9.09%) and Jetstream - TACC (0%). At the bottom of the page, there is a footer with '©2017 Jetstream-Cloud' and a 'Feedback & Support' button.

Jetstream

Dashboard Projects Images Help Admin jfischer

Getting Started

- Launch New Instance**
Browse Atmosphere's list of available images and select one to launch a new instance.
- Browse Help Resources**
View a video tutorial, read the how-to guides, or email the Atmosphere support team.
- Change Your Settings**
Modify your account settings, view your resource quota, or request more resources.

Resources Used

[Need more?](#)

Allocation Source

Allocation Source	Percent of Allocation Used
TG-STA1100245	5.06%
TG-ASC160018	32.23%
TG-CDA160007	23.15%
TG-TRA160003	17.61%
TG-TRA160027	0%

10 Instances

- active
- shutoff

4 Volumes

- available

Provider Resources

Provider	CPU
Jetstream - Indiana University	9.09%
Jetstream - TACC	0%

©2017 Jetstream-Cloud [Feedback & Support](#)



RESEARCH TECHNOLOGIES

UNIVERSITY INFORMATION TECHNOLOGY SERVICES

Pause...as we look at the Atmosphere Interface

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*



Jetstream for Education

Jetstream has been used in multiple graduate and undergraduate courses

- *Management, Access, and Use of Big and Complex Data*
- *Multiple informatics and general bioinformatics courses*
- *Business Intelligence (big data and analysis)*
- *Research Topics in Music*
- *Multiple genetics and sequencing courses*
- *Multiple information security and assurance courses*

Research Data Alliance workshops, Galaxy workshops, data analysis in finance using R, security and intrusion detection, and principles in cloud computing...





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





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/>)

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?

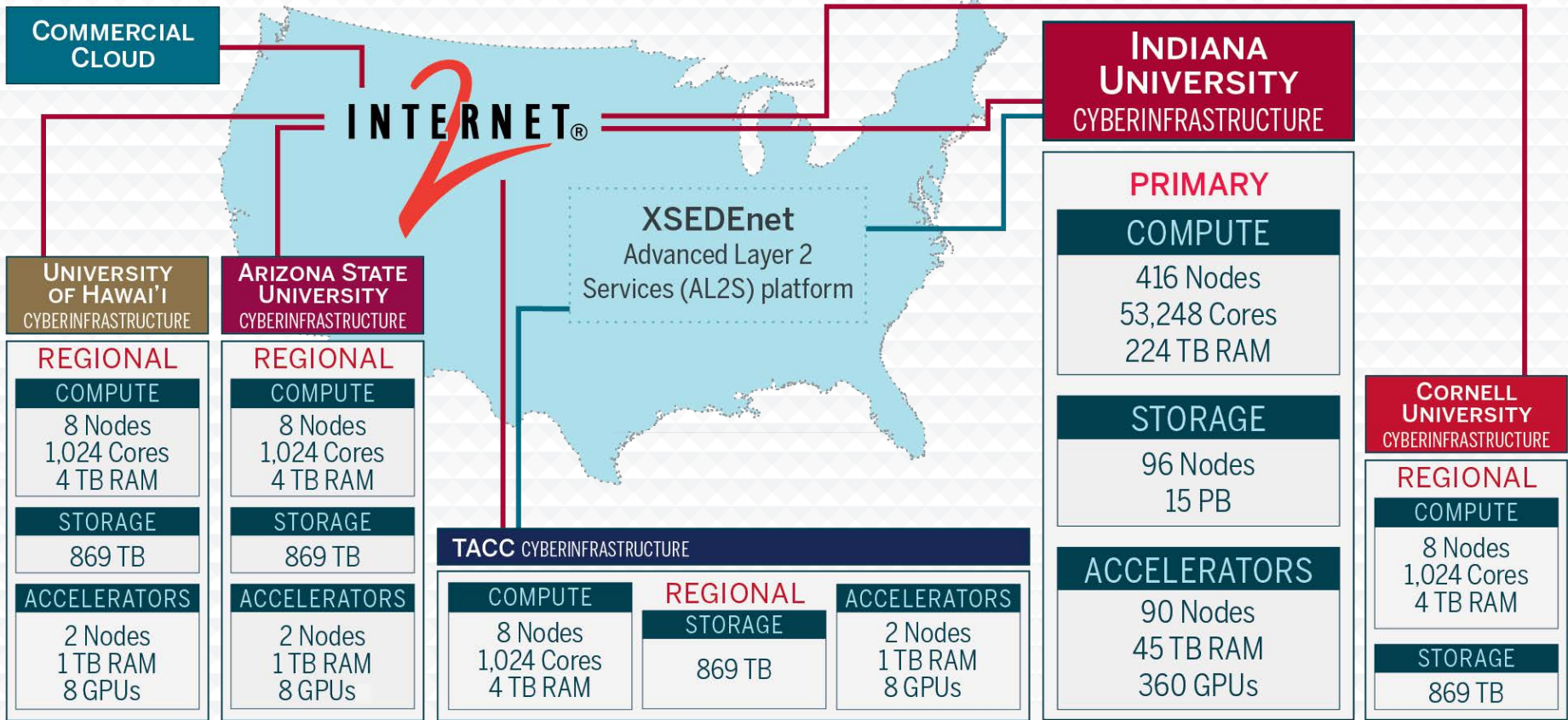
- *Forcing small allocations into the research allocation process*
- *Lack of multi-year allocations*
- *Lack of shared data set storage*



Jetstream usage highlights – 1 Nov 2020

- 410 XSEDE projects covering 75 fields of science and almost 2400 active users representing 407 institutions
- **80%** of Jetstream users have **not used any other XSEDE system**
- >362M CPU hours allocated to XSEDE projects since June 2016
- 48 active science gateways
- 49 education/teaching allocations serving almost 900 students
- 1189 mean active VMs in previous qtr, 1632 peak active VM count
- **Highest** user satisfaction in most recent XSEDE survey

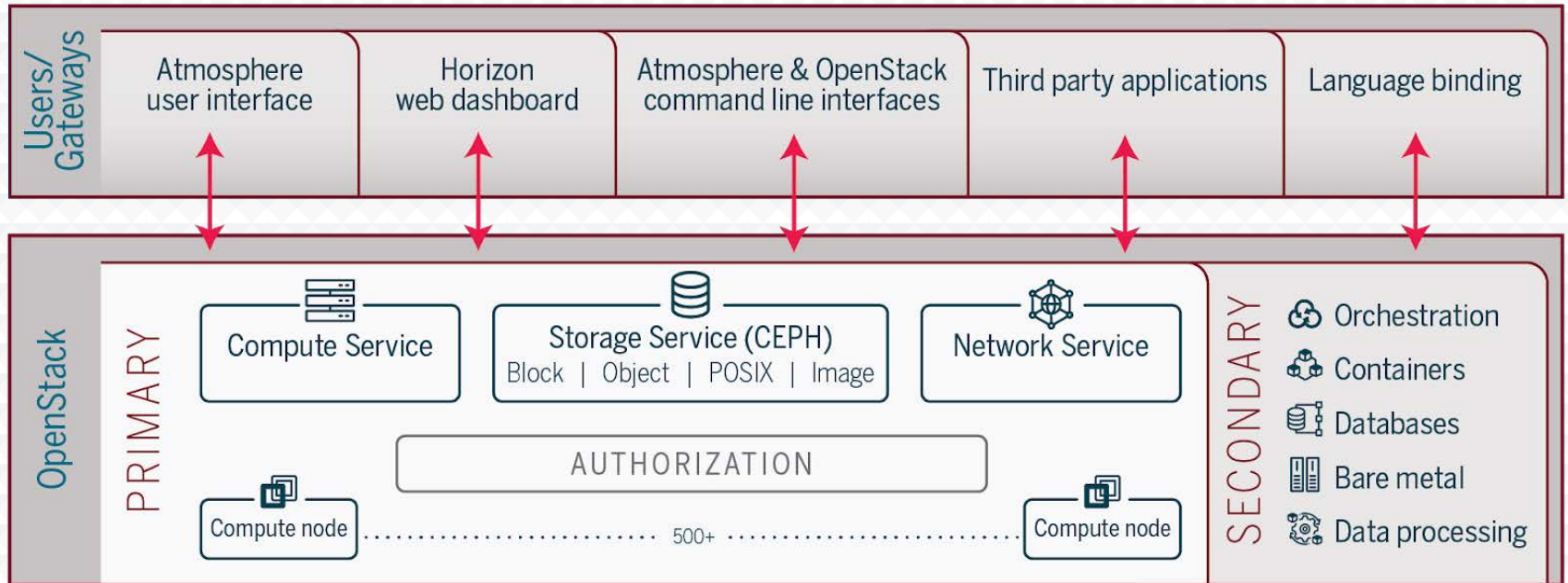




Jetstream2 Proposed Architecture



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

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.



Jetstream2 partners



<http://jetstream-cloud.org/>
National Science Foundation
Award #ACI-2005506