
PERVASIVE
TECHNOLOGY INSTITUTE
RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES

RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES

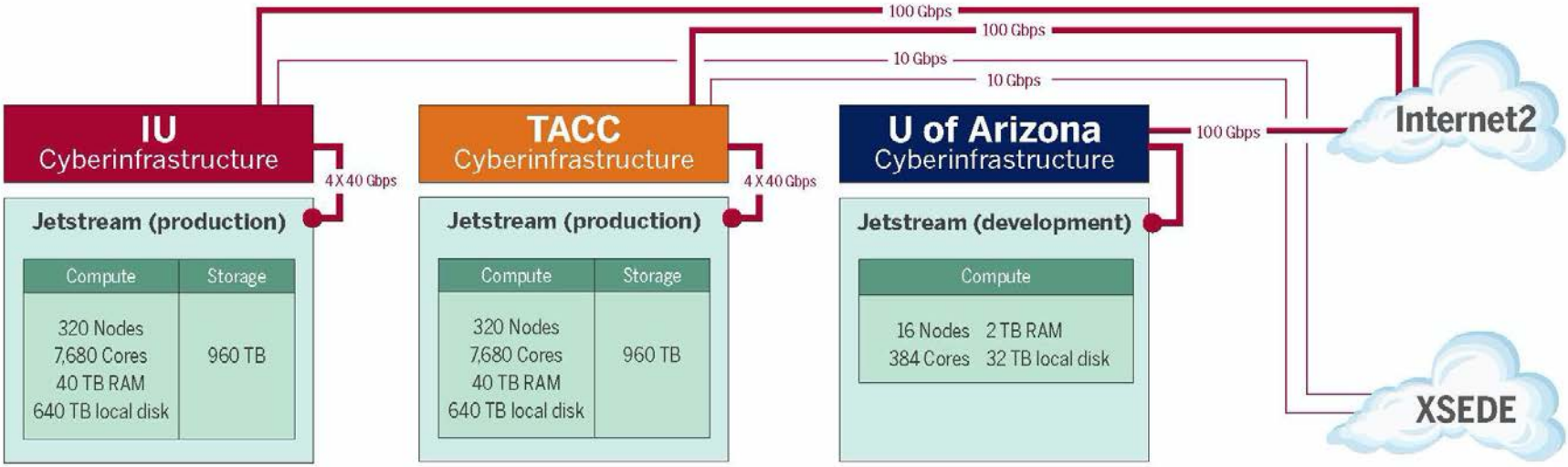
Jetstream & Jetstream2: Hybrid cloud in the past and future

Jeremy Fischer – Jeremy@iu.edu - Indiana University

Manager, Jetstream Cloud, UITS Research Technologies

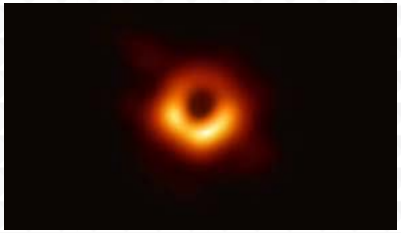
CASC Hybrid Cloud Panel – Oct 16, 2020

Jetstream 1 Architecture



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 back 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.

After completing the pilot project, transitioned to Google Cloud as their needs exceeded what Jetstream could offer

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

Brainlife.io

- *Deployments in Jetstream and commercial cloud (mostly AWS)*
- *Placed depending on need*
  - *Scaling educational efforts and some data processing on Jetstream*
  - *Large data collections on commercial cloud*
- *Partner in Jetstream2 to expand efforts going forward*

What worked?

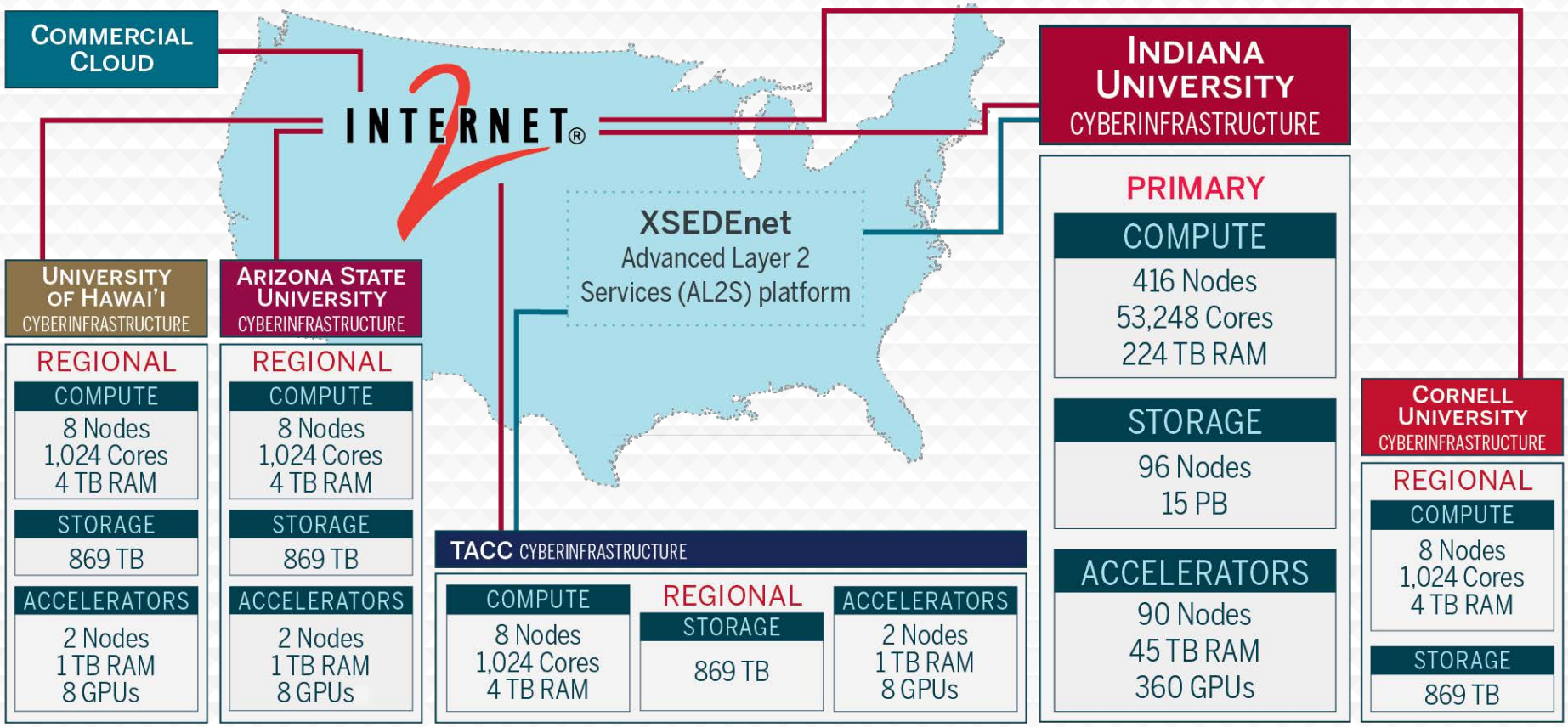
- *Allowing API access and full control (root privileges)*
- *Allowing allocations to run continuously – as long as the PI renewed – allowing workflows to run indefinitely*
- *Development of trial allocations*

What didn't work?

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

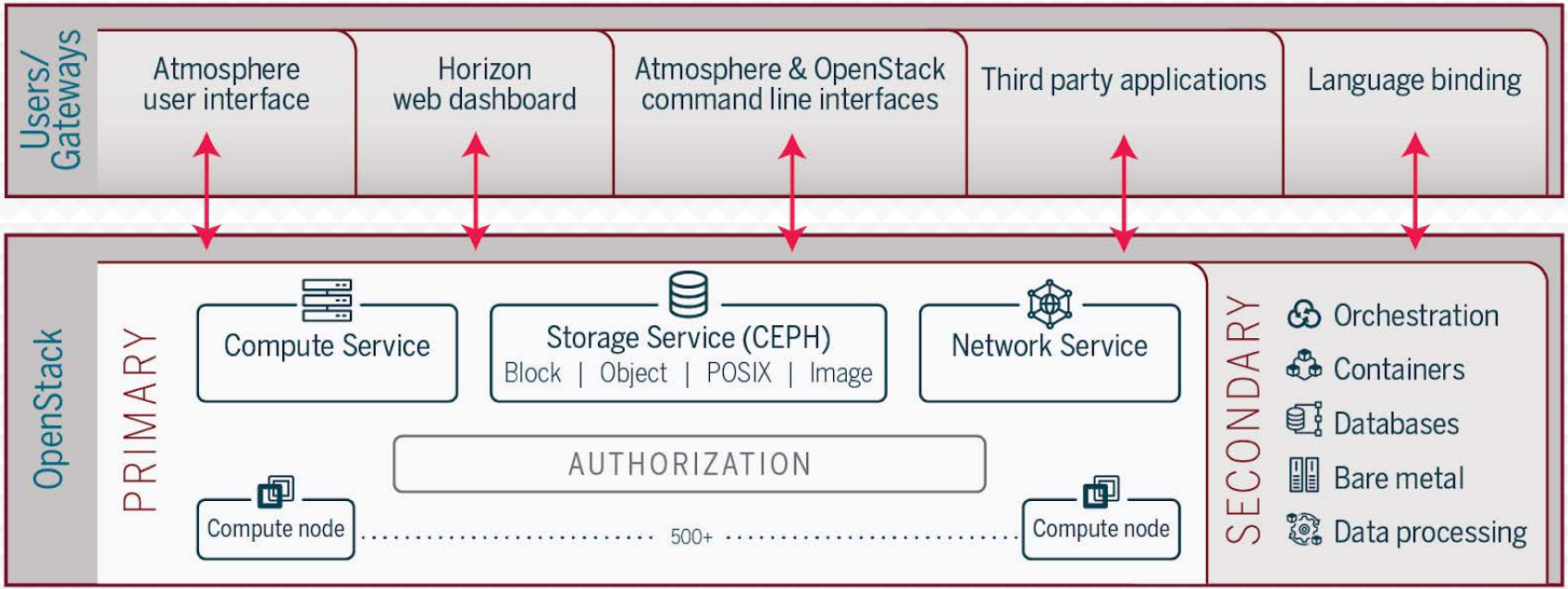
What improvements are planned?

- *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*



Jetstream2 Proposed Architecture

Conceptual Jetstream2 Architecture



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

Jetstream2

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

Jetstream2

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