Jetstream2: Accelerating science and engineering on-demand

David Y. Hancock – Indiana University
Director for Advanced Cyberinfrastructure
Jetstream2 Primary Investigator

NITRD MAGIC Subcommittee – 1 February 2023
NSF Vision and Blueprint

U.S. National Science Foundation (NSF) envisions an agile, integrated, robust, trustworthy and sustainable CI ecosystem that drives new thinking and transformative discoveries in all areas of S&E research and education.

• View CI more holistically…
• Recognize and support the translational research continuum…
• Develop a strategy that balances innovations with stability and continuity…
• Work closely with the diverse S&E communities to tightly couple discovery and innovation…
• Achieve new levels of usability by easing the pathways for discovering, accessing, understanding, and utilizing powerful CI capabilities…

From: OAC Vision & Blueprint: Overview and Computational Ecosystem (Apr 2019)
What is “the” Jetstream(2)?

- NSF-funded production cloud environment
- Ease-of-use focus, rapid on-ramp to XSEDE/ACCESS
- On-demand interactive computing and persistent services for science gateways
- Enables configurable environments; programmable cyberinfrastructure

Now with GPUs, large-memory, more faster PB!

By Maria Morris: JS2 rear doors (lower) Banksy adaptation [non-commercial] (right)
Platform Overview

- Caco
- XSEDE Accounting
- OpenStack
- Ceph
- Jetstream2
- Core System
- Horizon
- OpenStack CLI
- Exosphere
- Authentication Service
- Science Gateways
Exosphere

https://exosphere.Jetstream-cloud.org or try.exosphere.app
Differences from Jetstream[1] GUI

- Co-exists with other OpenStack interfaces, other research clouds
- Much easier collaboration between users on same allocation
- Choose an operating system instead of browsing a list of images
- Multi-instance create and delete
- Volume-backed instances ➡ large, persistent root disks
- Live instance load graphs (CPU, GPU, RAM, storage)
- Live instance resize
- Push-button virtual clusters with elastic scaling
- Reproducible workbenches with Binder-compatible repositories
CACAO

https://cacao.jetstream-cloud.org
Deep dive into DL

- First-of-a-kind workshop to apply DL techniques to agricultural data sets in April 2022
- AIIRA, AI Institute for Resilient Agriculture, intends to distribute the digital twin built on JS2 for community re-use
- Allows community training and inference
- Provided via Terraform templates and customized UI through CACAO

...we were able to easily provide so many students with a GPU-enabled container so quickly. Normally, getting GPU resources on an HPC scheduler, like OnDemand, takes time, and the high demand for GPUs makes finding 40 or more unoccupied resources an impossibility.

- Tyson Swetnam, CyVerse Co-PI and workshop instructor

https://cyverse.org/deep_learning_workshop
Gateways use JS2 in several ways

- Gateway web hosting
- Datasets and Database hosting
- Gateway Security Services
- Integrated JupyterHub
- Interactive Computing
- Elastic Virtual Clusters
Elastic Virtual Clusters

• One Click OnDemand Cluster Augmenting the cloud Capabilities
  – Bundled lightweight HPC Stack, including SLURM.
  – Users deploy scientific software with complete OS control.
  – Dedicated and Responsive scheduler for rapid testing and development like workloads.
  – Mounted persistent storage.
Virtual Clusters Architecture

- All these steps are bundled into the Ansible orchestrion.
- The entire system is a single click invoked from Exosphere.
Early Operations Projects & Activity

- First PI invitations and projects added in February 2022
  - Remaining project migrations in May – July
- Retired Jetstream[1] in July/August 2022
- Full production in September 2022 after NSF approval
- Dec 2022: 357 projects and 1753 individuals (565 students)
- Approximately 900 unique people have created JS2 instances to date via Exosphere
- Includes multiple science gateways and education/training allocations

"Bike Exchange - 2009 IU Women's Little 500" by Indiana Public Media
Flickr CC BY-NC 2.0
Vision for Jetstream2 is that it functions as a production system yet does not cede our pilot roots.

- Obsolescence vs Maturity & graceful aging
- Carry new lessons into the future

*Imitation is the sincerest form of flattery*

- Influenced design of many other systems
- Distinct utility, focus, and inclusion
- Reflecting on Why?

"Metamorphosis" by h.koppdelaney
Flickr CC BY-ND 2.0
Dynamic Connections

Importance of leveraging other projects
• XSEDE -> ACCESS
• Exosphere
• CyVerse – CACAO
• Globus
• Custos / CI Logon
• Open Source
Operations highlights

- OpenStack upgrades x4 Wallaby -> Zed
- Shared storage availability (Manila)
- DNSaaS for instances (Designate)
- Only 17.2 hrs downtime (<0.2%) 4Q22
- 96.4% of instances started in <5 min
- Using CI/CD for image build pipeline
  - Weekly updates (vs periodic)
  - Allows more flavors (currently 7)
  - Allows reuse of our pipeline for others

"The Highlights" by Modesto del Río
Flickr CC BY 2.0
Developments to come

- LBaaS (load balancing)
- Secret storage
- Managed Kubernetes (via OpenStack)
- IPV6
- Addition of a new partner
- Continued/evolving outreach
- Increased Cacao use / features

Exosphere specific
- Shared storage integration
- GPU-accelerated desktops
- Education / workshop features

"Work in progress" by Alexander Baxevanis
Flickr CC BY 2.0
What’s next?

• Midway into YR 1 operations
• Prepare for panel review (~March 2023)
• Integrate new partners
• Survey JS2 community
• Grow the community, focus on new tools and approaches
• Support hybrid science gateways
• Upgrade, share, and evolve
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.

Special thanks to contributors & Jetstream2 partners
- Vendors, particularly Dell and NVIDIA, also deserve recognition for their efforts