



RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES

rt.iu.edu



RESEARCH TECHNOLOGIES

UNIVERSITY INFORMATION TECHNOLOGY SERVICES

National and Local Impacts of Programmable Cyberinfrastructure using Jetstream

David Y. Hancock – Indiana University

Director for Advanced Cyberinfrastructure

Jetstream Primary Investigator

Affiliated with IU Pervasive Technology Institute

SC19

Denver, CO – November 19, 2019



7 University Campuses
9 Medical Education Centers
2 Academic Centers

Indiana University, *est. 1820*

- **\$3.7B** enterprise
- Partnered with **\$6.4B** IU Health system
- **94,000** Students
- **1.1M** credit hours per semester
- **>20,000** degrees per year
- **\$950M** in financial aid
- **\$604M** in research grants
- **20,000** faculty and staff
- **7,873** Acres
- **949** owned and leased buildings, **36M** square feet
- **>690,000** *living* alumni

Almost completed
2009 Network Master
Plan for IUB/IUPUI 361
buildings <\$100M
(Orig Budget \$172M)

Multi-institution Partnerships

Networks



National network collaboration of 400+ Universities and national labs. IU acts as network operations center for the organization.



Indiana state network for higher education with over 100 members. IU acts as its network operations center.



Collaboration of I-Light, IU, Notre Dame, and Purdue providing high-speed, high-availability, feature-rich network to Indiana's higher education institutions.

Libraries



Began as collaboration between the Big Ten Academic Alliance and the University of California system to digitize library print holdings. Now includes over 150 universities. IU serves as primary backup site for repository.

Research Technologies



IU-led collaboration with Texas Advanced Computing Center, University of Chicago, University of Arizona, and University of Texas (SA) to develop cloud based tools for scientific research funded by National Science Foundation.



Collaboration with Texas Advanced Computing Center, Pittsburgh Supercomputing Center and San Diego Supercomputing Center to analyze massive genomic data.



Collaboration of IU, Purdue, and Notre Dame, with public and private partners to facilitate the translation of scientific discoveries in the lab into clinical trials and new patient treatments.

Learning Technologies



Coalition of universities dedicated to collaborative digital education. Founded by IU, Michigan, Colorado State, and Florida. Now includes 25 institutions.

Security



Research and Education Networks Information Sharing and Analysis Center. Includes 620 member universities. One of 19 nationally recognized industry-specific ISACs.



Center for Applied Cybersecurity Research, Est. 2003, integrates applied research in cybersecurity technology, education, and policy guidance.



Shared Cybersecurity Operations Center. Founded by IU, Nebraska, Northwestern, Purdue, and Rutgers.



ResearchSOC provides cybersecurity services to NSF-funded facilities and projects, such as Gemini Observatory, UNAVCO and GAGE, and the National Radio Astronomy Observatory.

Enterprise Systems



Community sourced software suite driving down cost of enterprise systems. IU co-founded and currently serves on board of directors. Includes over 50 university members.



Portal replacement service discovery application developed at IU. Currently used by 85 university campuses.

Service & Support



Application developed at IU to integrate cloud storage systems into a single place. Currently has 8 university subscribers and 6 universities testing.



System to manage software licenses developed at IU. Subscribers include University of Maryland and University of Alabama-Huntsville.

What is “the” Jetstream?

- Fast moving air currents
- Hot/Cold air boundaries
- An NSF-funded cloud environment



Suomi NPP satellite collected this natural color image using the VIIRS (Visible Infrared Imaging Radiometer Suite) instrument on Sept. 4, 2017. Actively burning areas are outlined in red. NASA image courtesy Jeff Schmaltz LANCE/EOSDIS MODIS Rapid Response Team, GSFC



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
- XSEDE-supported CI serves **more than just the NSF**
- Provides **on-demand interactive** computing and analysis or persistent services such as gateways
- Enables *configurable* environments; *programmable cyberinfrastructure*

NIH (21%) + DOE (20%) + DOD (8%) > NSF (34%) usage over first 6 years of XSEDE



Expanding NSF CI's reach and impact

Around 350,000 researchers, educators, & learners received NSF support in 2015

- **Less than 2%** completed a computation, data analysis, or visualization task on XSEDE-supported resources
- Less than 4% had an XSEDE Portal account
- **70%** of researchers surveyed* claimed to be **resource constrained**

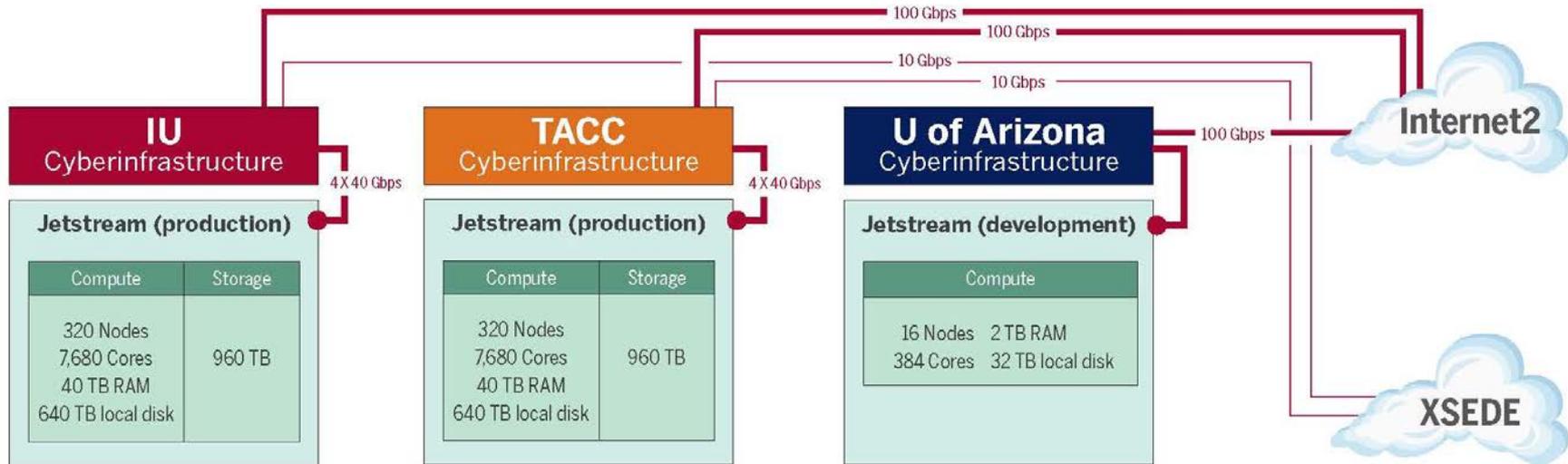
Why are the people not using XSEDE-supported systems not using them?

- Perceived **ease of access** and use
- HPC resources – the traditional view of what XSEDE offers - are often **not well-matched** to their needs
- They just don't need *that much* capability

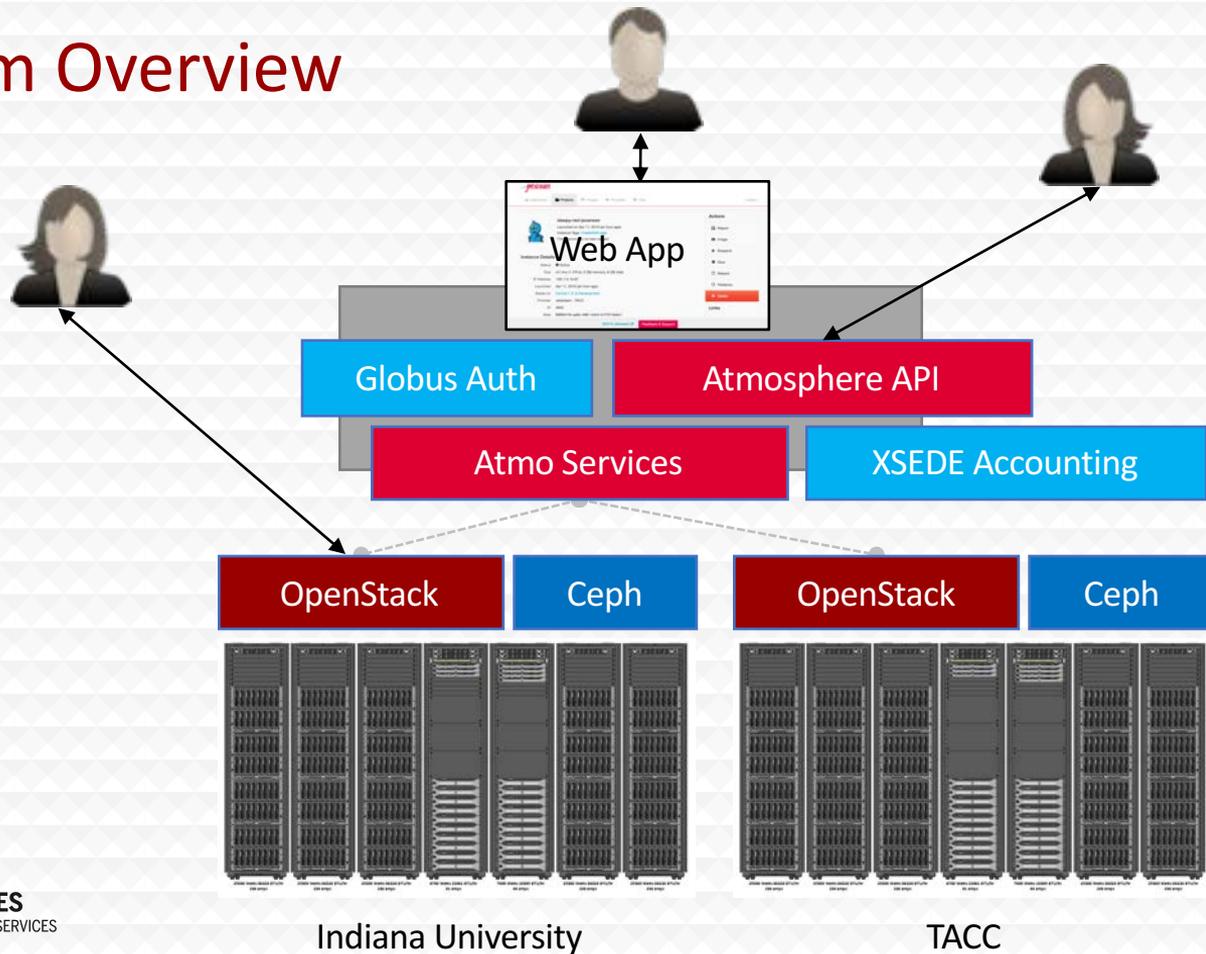
*XSEDE Cloud Survey Report - <http://hdl.handle.net/2142/45766>



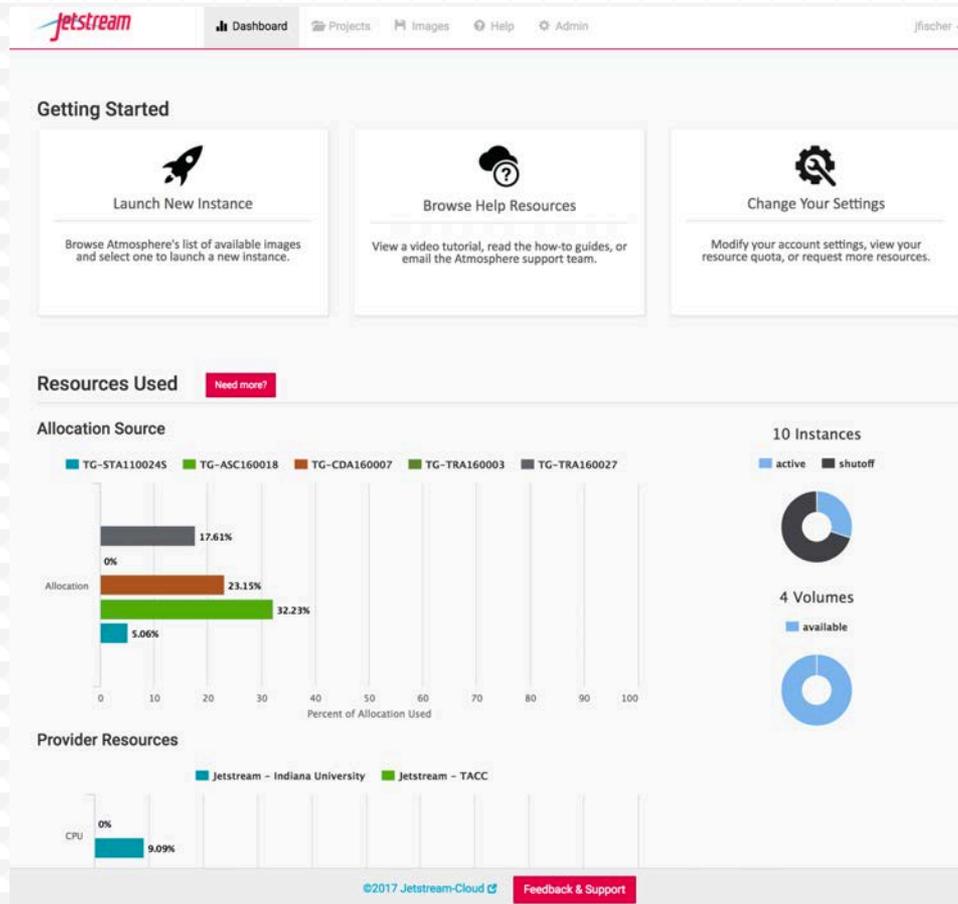
Jetstream System Overview



Platform Overview



Jetstream Atmosphere web interface



Jetstream Atmosphere web interface

The screenshot displays the Jetstream Atmosphere web interface. At the top, there is a navigation bar with the Jetstream logo and menu items: Dashboard, Projects, Images (selected), and Help. A user profile 'jlfest' is visible in the top right. Below the navigation bar is a search bar and several tabs: SEARCH, FAVORITES (0), MY IMAGES (0), MY IMAGE REQUESTS, and TAGS. The main heading is 'Image Search'. A search input field contains the placeholder text 'Search across image name, tag or description'. Below this, it says 'Showing 57 of 57 Images'. The 'Featured Images' section lists six images, each with a thumbnail, title, creation date, author, and a list of tags. The tags are color-coded: CentOS (red), development (orange), Featured (green), gui (blue), IRDS (purple), Ubuntu (yellow), x2go (pink), desktop (light blue), intel (light green), m1_small (light orange), m1_large (light purple), and community-contributed (light yellow).

Image Thumbnail	Image Title	Image Description	Tags
	CentOS 7 (7.2) Development GUI Jan 13th 17 03:21 by jfischer	Imported Application - Centos 7 (7.2) Development GUI	CentOS, development, Featured, gui, IRDS
	BioLinux 8 Jan 2nd 17 03:34 by jfischer	Based on Ubuntu 14.04.3 -Trusty Tahr - server - cloudimg --**REQUIRES m1.small instance ...	bioinformatics, desktop, Featured, gui, m1_small, Ubuntu, x2go
	Ubuntu 14.04.3 Development GUI Jan 2nd 17 01:24 by jfischer	Based on Ubuntu 14.04.3 Development Patched up to date as of 12/15/16 Base Ubuntu 14.04.3 ...	desktop, development, Featured, gui, IRDS, Ubuntu, vnc
	Intel Development (CentOS 7) Nov 30th 16 12:04 by jfischer	Intel compilers and development environment *REQUIRES a m1.small or larger VM to la ...	CentOS, desktop, development, Featured, gui, intel, m1_small, vnc
	R with Intel compilers (CentOS ...) Nov 30th 16 11:53 by jfischer	R with Intel compilers built on CentOS 7 (7.3) ** Requires m1.small or greater sized VM * ...	CentOS, desktop, development, Featured, gui, intel, m1_small, vnc
	Galaxy Standalone Nov 15th 16 04:49 by admin	Galaxy 16.01 Standalone - based on Ubuntu 14.04.4 LTS This is a standalone Galaxy server ...	community-contributed, Featured, m1_large, Ubuntu

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



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



Discipline or area of interest	#of Jetstream allocations	SUs allocated on Jetstream	SU increase/Decrease on Jetstream over previous year	% of SUs allocated on Jetstream	% of all SUs allocated on other XSEDE-supported systems
Behavioral Sciences	6	3,465,516	100%	4.24%	0.61%
Biological Sciences	89	15,041,928	72%	18.40%	3.59%
Biophysics	86	3,627,026	(44%)	4.44%	13.56%
Computer Science	72	6,883,269	32%	8.42%	2.98%
Earth Sciences	37	5,476,250	37%	6.70%	4.60%
Education and Training	128	16,599,512	3%	20.31%	4.66%
Engineering	13	520,690	71%	0.64%	1.75%
Materials Science	6	1,035,508	100%	1.27%	13.89%
Mathematics	13	688,505	150%	0.84%	0.90%
Molecular Science/Biochemistry	21	4,254,643	(10%)	5.20%	5.83%
Neuroscience	19	4,708,180	328%	5.76%	1.98%
Physics	10	2,440,581	(16%)	2.99%	8.65%
Social Sciences and Humanities	28	2,409,633	192%	2.95%	0.81%



“But I really don’t have research needs... I don’t need the national research cyberinfrastructure.”

– multiple researchers at small colleges and universities



Sarah Romanes @sarah_romanesh · Mar 17

When you have to drive to work - but R has also only completed 10hrs of 24hrs worth of simulations. Can I go in the T2 lane with this thing? 😂 #rstats #sydneytraffic



40 74 1.1K



RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES

What Jetstream isn't...

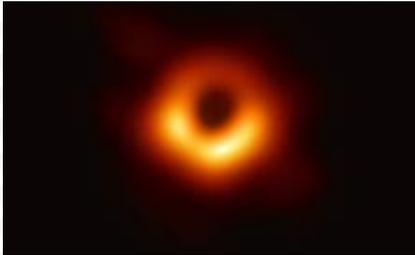
- It's not traditional HPC
- There's no shared filesystem (think cloudy!)
- There's no high-end interconnect fabric (keep thinking cloudy!)
- There aren't many GPUs
- It isn't Amazon, Azure, or GCE (similar, but...)



M87 black hole: how cloud computing supports astronomy



Event Horizon Telescope

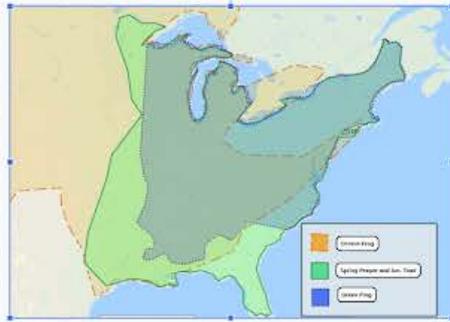


*M87 black hole image
generated by EHT*

- 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



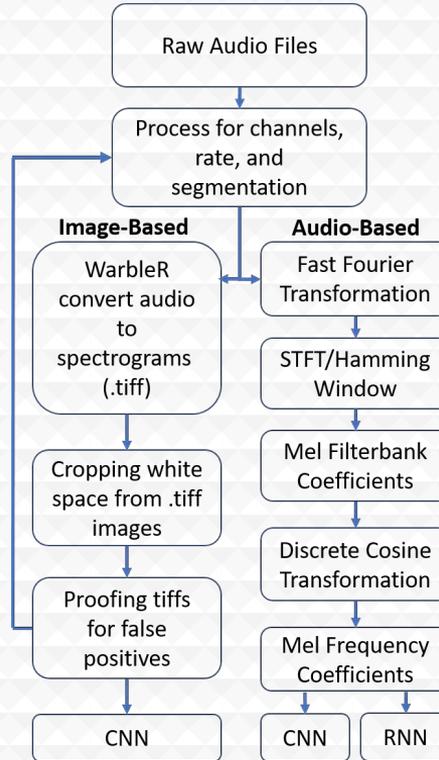
Spring Peeper



Chorus Frog



Green Frog



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



Project TERRIER: modernizing political science research



Journalists, through their words, provide an invaluable service, sharing information about global events to which many of us would not otherwise have access.

They send missives directly from event sites, recording during protests, summits, and speeches. For political scientists, these articles offer a rich mine of data about these events.

Project Temporally Extended, Regular, Reproducible International Event Records ([TERRIER](#)) extracts event data from roughly 300M news articles and puts it into a form usable by researchers.

Jetstream provides the storage and structure to launch the pipeline and process news articles for TERRIER.



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...



Another Use Case: Galaxy riding Jetstream

Galaxy, a platform for biomedical research, focused on accessibility, transparency and reproducibility

- usegalaxy.org has more than 100,000 registered users executing 300,000+ jobs each month
- Many users need more capacity than the public quota, or other customizations

Use Jetstream as a bursting platform

- From Galaxy Main, send jobs to a Slurm cluster running on Jetstream
- Run Galaxy Interactive Environments (Jupyter/RStudio containers) via a Swarm cluster running on Jetstream

Use Jetstream as a self-service platform

- Pre-built Galaxy image configured with hundreds of tools and access to TBs of genomic reference data, available via the self-launch model within minutes
- Allows users to acquire (free) resources, and gives them complete control



Jetstream Gateway Highlights

- IRIS

Serving large scale earthquake and geographical data for analysis

- Unidata

Providing distribution and analysis of meteorological data

- OpenMRS

Providing medical records systems for the resource-constrained

- SEAGrid

Computational chemistry, molecular and fluid dynamics, and structural mechanics gateway

- NAMDRunner

Based on the GenApp gateway

- ChemCompute Gateway

Providing a computational chemistry gateway for educational use

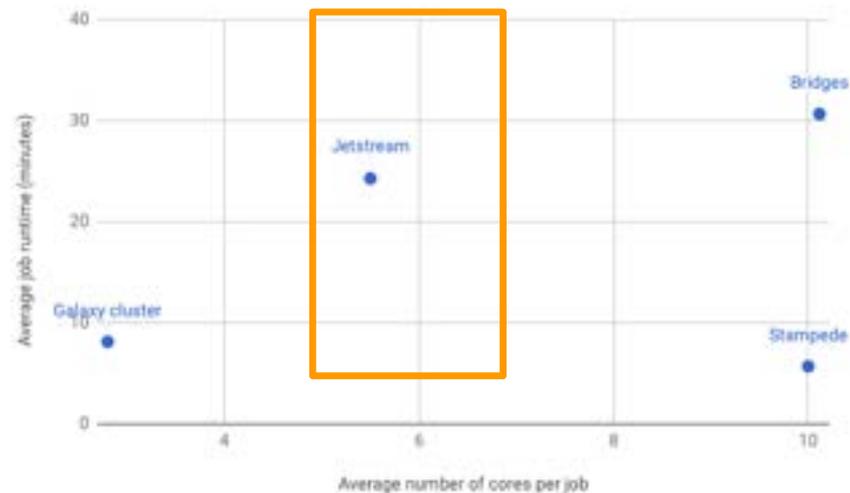
Coming gateways: The Neuroscience Gateway, UltraScan III, and others



Science Gateway: Galaxy

Galaxy Main gateway usage stats for Jetstream

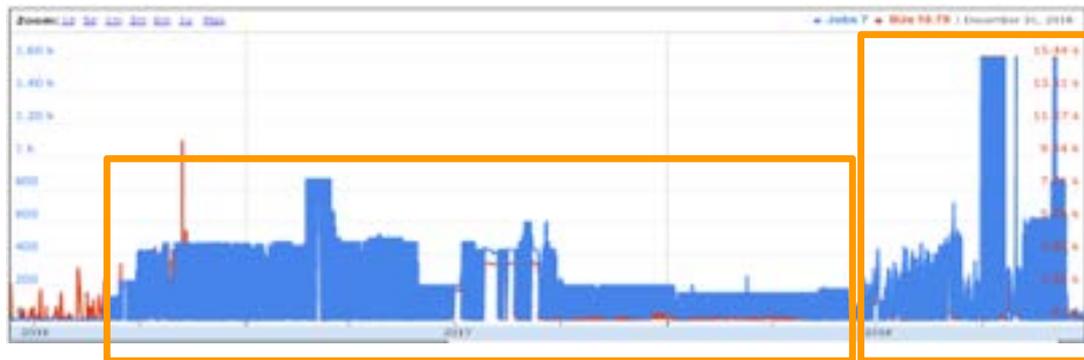
- Galaxy Main (*usegalaxy.org*) submits a portion of its workload to remote Jetstream nodes while it automatically handles data staging.
- Jetstream processed jobs for **26,184 individual users**
- Galaxy submitted **> 200K jobs**
- Jetstream is **the only resource handling medium-size jobs for Galaxy Main**



Science Gateway: Galaxy

In 2018, Galaxy implemented elastic node scaling

SlurmScale Ansible role monitors Slurm node usage and adapts VM state, making more efficient use of the Jetstream allocation.



Manually set number of job nodes.

Elastic scaling.



The CPU hours consumption remains relatively constant yet allocation utilization is bursty. Galaxy is leveraging Jetstream elasticity to reduce job wait.

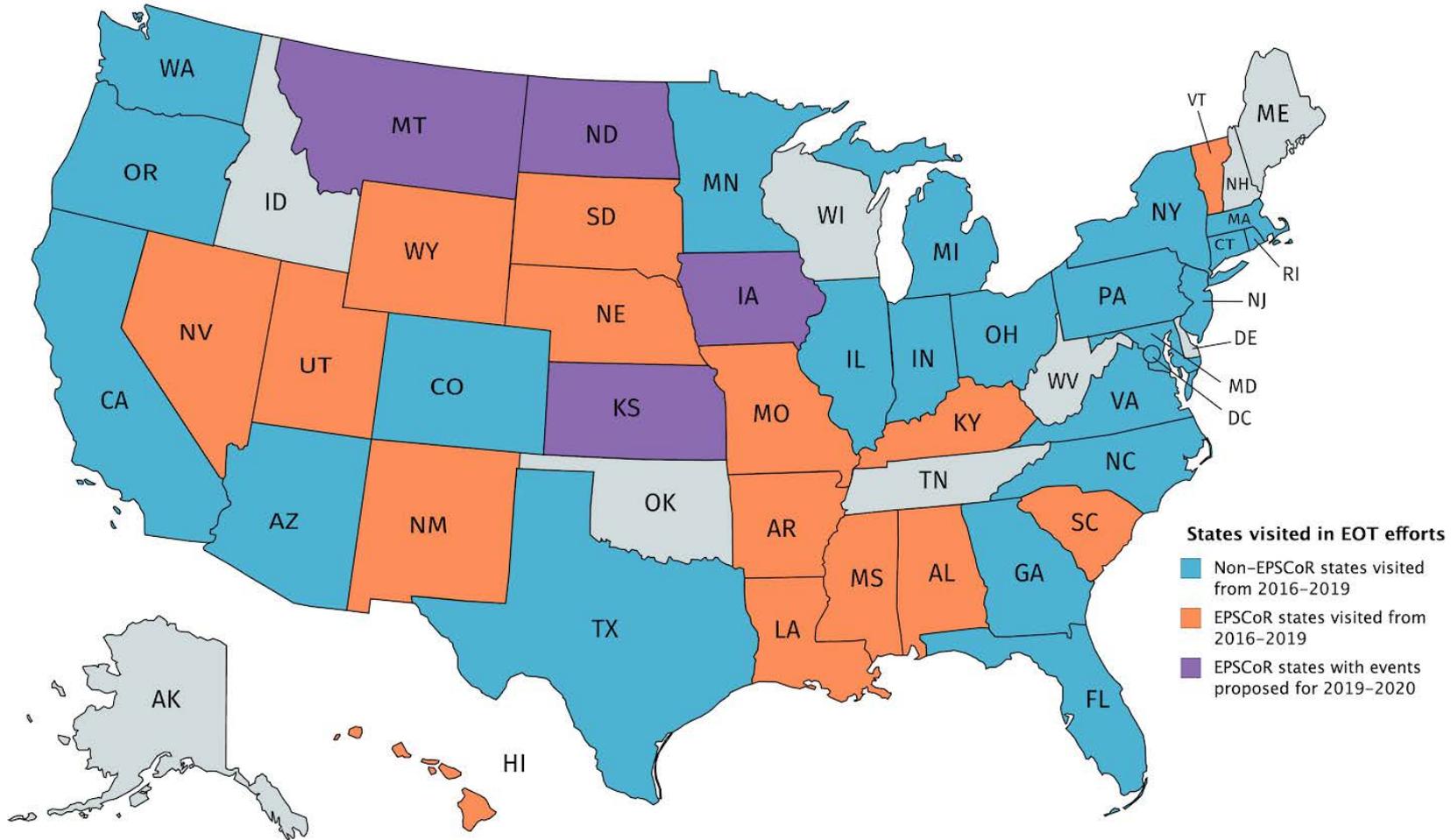


Jetstream usage highlights – 1 Sep 2019

- 398 XSEDE projects covering 72 fields of science and over 2300 **active users** representing **207 institutions**
- **80%** of Jetstream users have **not used any other XSEDE system**
- 2,218 students to date
- 27 active science gateways
 - **26,184** users have run jobs on Jetstream from Galaxy
- 53 education/teaching allocations serving almost 1,000 students
- **Highest** user satisfaction in most recent XSEDE survey



States visited in Jetstream Outreach Efforts - 2016 to 2019



Expanding the reach: Jetstream REU Program



NSF Supplement for undergraduates

- 4 students in 2017
 - 6 students in 2018
 - 7 students in 2019
 - **Planned supplement request for 2020**
-
- REU student videos on YouTube
<https://www.youtube.com/user/IUPTI>



Requesting access to Jetstream

- Trial allocations available **TODAY**
 - <http://wiki.jetstream-cloud.org/Jetstream+Trial+Access+Allocation>
- You can request **startup** allocations **anytime**. (Startups are simple!)
- <http://wiki.jetstream-cloud.org/Jetstream+Allocations>
- You can request allocations for **educational** use **anytime**.
- Next submission period for large allocations is 15 Mar 2019 – 15 Apr 2019
- Research allocation: Project desc (≤ 10 pages) and Scaling doc (≤ 5 pages)
 - **We can help!**



Where can I get help?

- **Wiki / Documentation:** <http://wiki.jetstream-cloud.org>
- User guides: <https://portal.xsede.org/user-guides>
- XSEDE KB: <https://portal.xsede.org/knowledge-base>
- Email: help@xsede.org
- Campus Champions: <https://www.xsede.org/campus-champions>
- Introduction to Jetstream Virtual Workshop: <https://cvw.cac.cornell.edu/jetstream/>
- Jetstream Allocations Virtual Workshop: <https://cvw.cac.cornell.edu/JetstreamReq/>



Jetstream Timeline... looking ahead?

- Completed 3rd year of operations with extension to November 2020
- Soliciting Research allocation requests plus Startup and Education allocations – including Science Gateways!
- Services added (Heat, Magnum, Trove, Manila, etc)
- Atmosphere enhancements
- Belief that the NSF should continue investments for the community



Flickr user Oiluj Samall Zeid - Lejos de Yulín





Acknowledgements

NSF Awards 1053575 & 1548562 (XSEDE), 1341711 (Wrangler), and 1445604 (Jetstream).

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 & Jetstream partners:

- Craig Stewart, Therese Miller, Jeremy Fischer, J. Michael Lowe, Winona Snapp-Childs, George Turner, Brad Wheeler, Dan Calarco, and Maria Morris.
- TACC, University of Arizona, Johns Hopkins, University of Chicago, Cornell, UT San Antonio

Jetstream partners



INDIANA UNIVERSITY
PERVASIVE TECHNOLOGY INSTITUTE



THE UNIVERSITY
OF ARIZONA

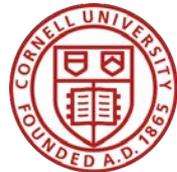
TACC



JOHNS HOPKINS
UNIVERSITY



THE UNIVERSITY OF
CHICAGO



JSU | JACKSON
STATE
UNIVERSITY



UNC
THE ODUM INSTITUTE

Jetstream
<http://jetstream-cloud.org/>



funded by the National Science Foundation
Award #ACI-1445604



A PARTNER
in new
POSSIBILITIES



RESEARCH TECHNOLOGIES

UNIVERSITY INFORMATION TECHNOLOGY SERVICES

RT.IU.EDU