

# RESEARCH TECHNOLOGIES

UNIVERSITY INFORMATION TECHNOLOGY SERVICES

[rt.iu.edu](http://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

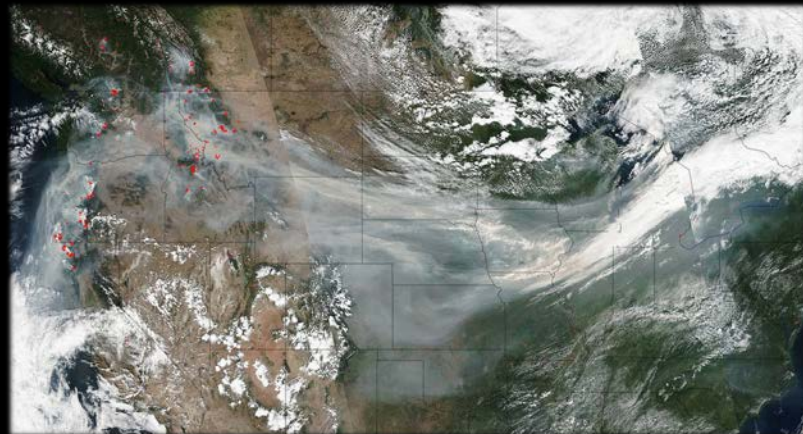
November 14, 2018

# What is “the” Jetstream?

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



NASA's 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
- Provides **on-demand** *interactive* computing and analysis or persistent services such as gateways
- Enables *configurable* environments; *programmable cyberinfrastructure*



# Expanding NSF XD'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 XD/XSEDE program resources
- Less than 4% had an XSEDE Portal account
- **70%** of researchers surveyed\* claimed to be **resource constrained**

Why are the people not using XD/XSEDE 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>





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

– multiple researchers at small colleges and universities



# 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



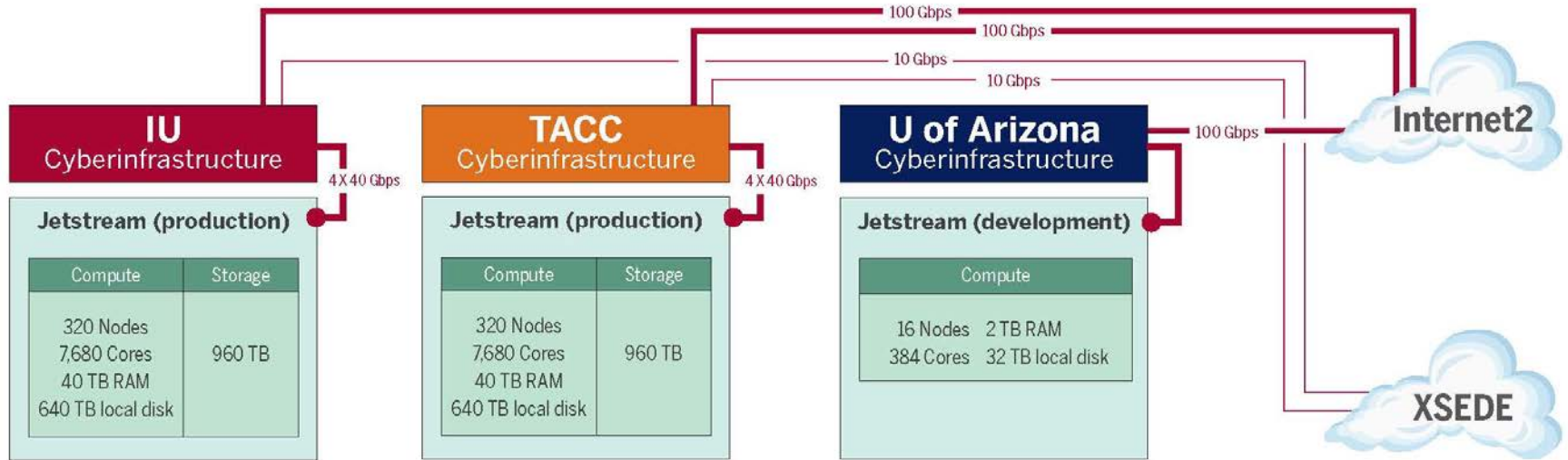
# What Jetstream isn't...

- 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...stay tuned)
- It isn't Amazon, Azure, or GCE (similar, but...)

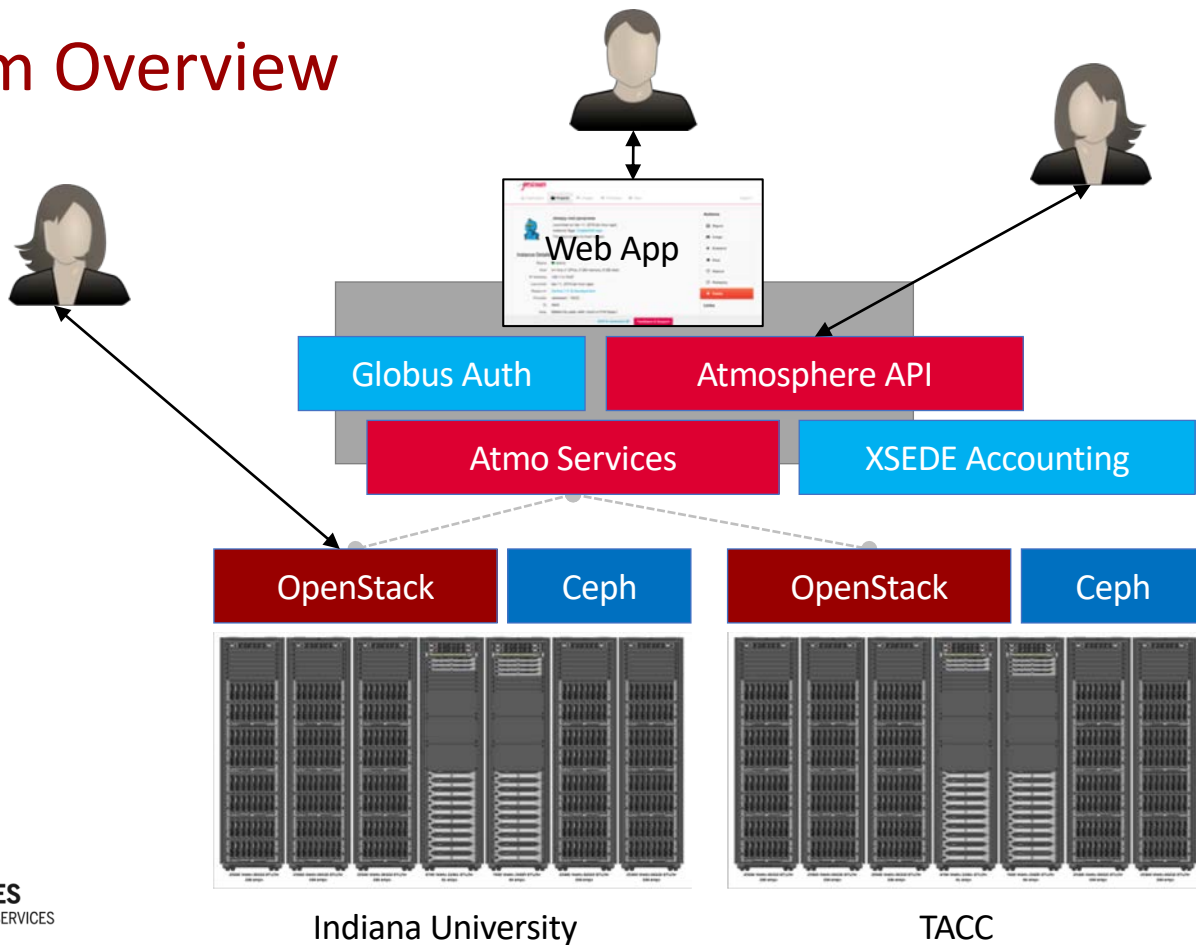




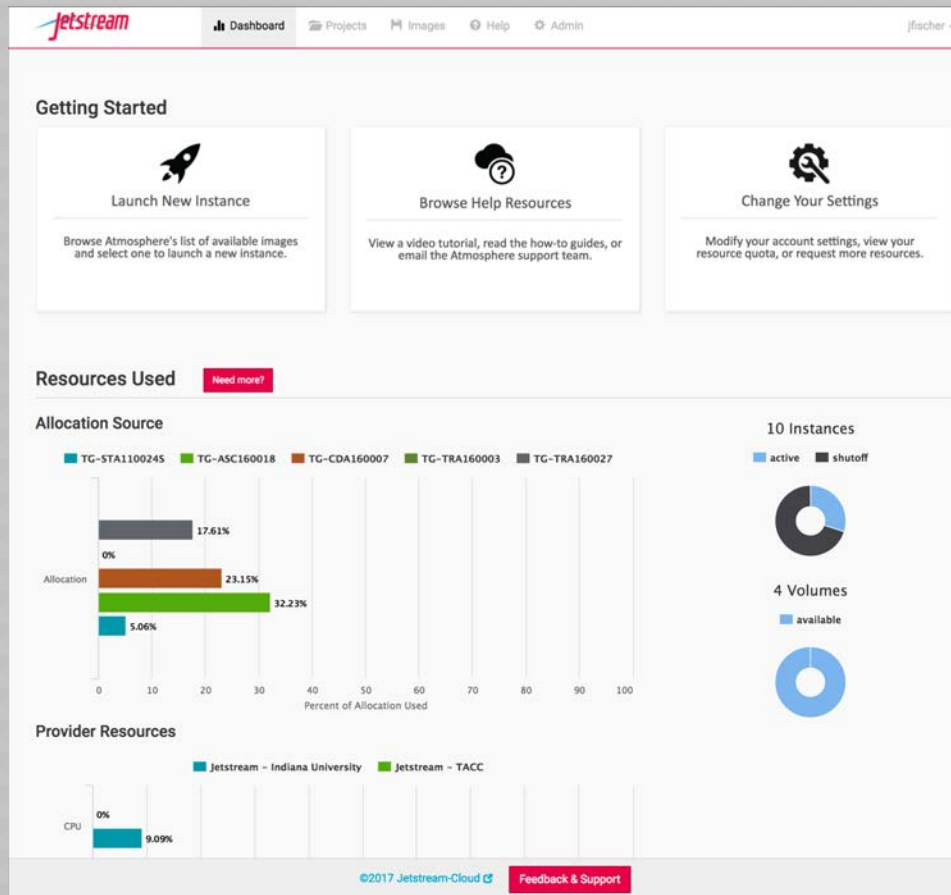
# Jetstream System Overview



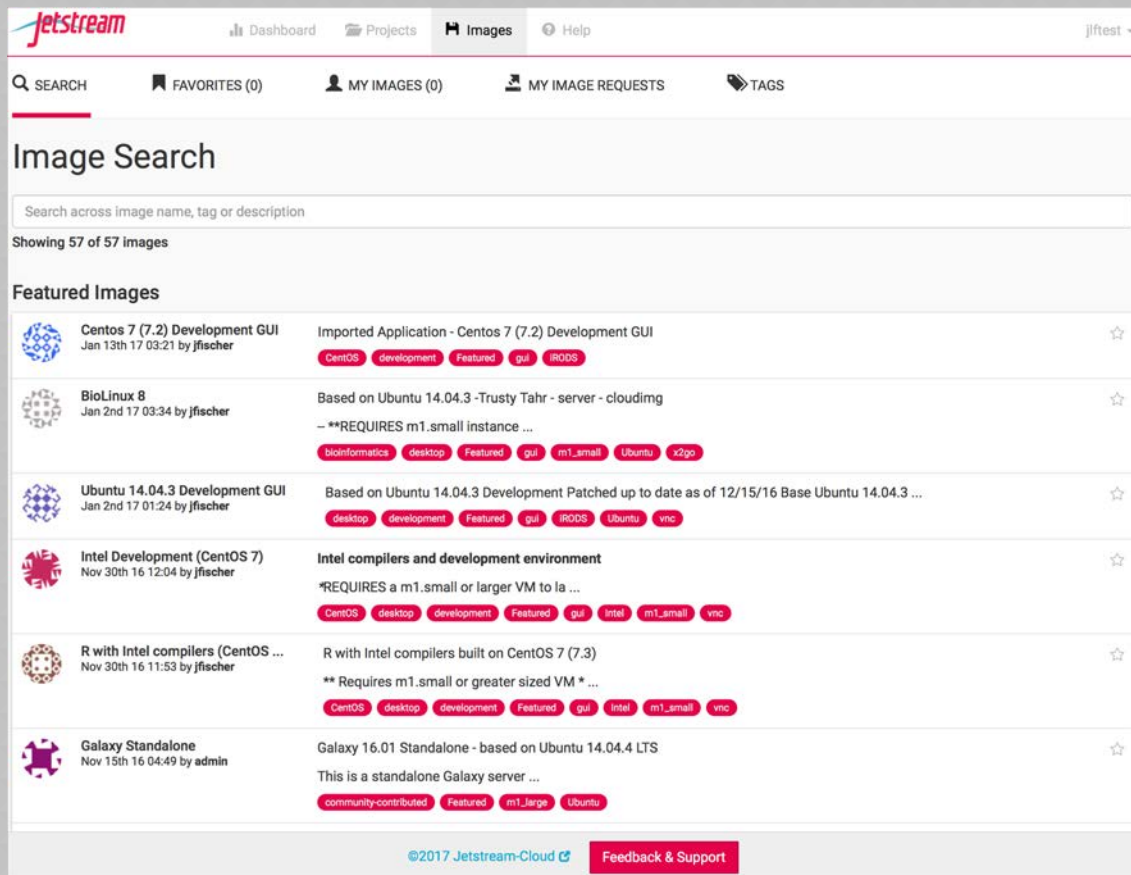
# Platform Overview



# The Jetstream Atmosphere web interface



# The Jetstream Atmosphere web interface



**Jetstream** Dashboard Projects Images Help jlfest







SEARCH FAVORITES (0) MY IMAGES (0) MY IMAGE REQUESTS TAGS

## Image Search

Search across image name, tag or description

Showing 57 of 57 Images

### Featured Images

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

©2017 Jetstream-Cloud Feedback & Support



Discipline or area of interest	#of Jetstream allocations	SUs allocated on Jetstream	% of SUs allocated on Jetstream	% of all SUs allocated on other XSEDE-supported systems
Astronomy	2	1,108,096	3.04%	8.61%
<b>Atmospheric Sciences</b>	4	2,752,400	7.55%	3.73%
<b>Biological Sciences</b>	57	5,199,000	14.27%	4.95%
<b>Campus/Domain Champions</b>	123	6,105,500	16.76%	0.09%
Computational Science	11	1,150,000	3.16%	0.92%
<b>Computer Science</b>	15	4,944,302	13.57%	1.8%
<b>Education Allocations</b>	24	2,847,600	7.82%	0.01%
Engineering	1	100,000	0.27%	3.81%
Geosciences	10	1,978,400	5.43%	2.87%
Humanities/Social Sciences	10	560,000	1.54%	0.45%
Molecular Biosciences	8	4,647,520	12.75%	17.65%
Network Science	3	200,000	0.55%	0.06%
Ocean Science	3	230,000	0.63%	1.30%
Physics	4	2,252,400	6.18%	16.43%
<b>Training &amp; Development</b>	11	2,362,000	6.48%	0.16%

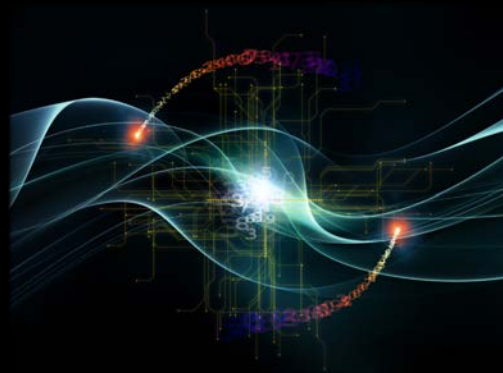




# Not just the usual suspects...

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...



# 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





# Jetstream usage highlights – 1 October 2018

- 413 active XSEDE projects covering 75 fields of science and **2558 active users** representing **190 institutions**
- **80%** of Jetstream users have **not used any other XSEDE system**
- >143M CPU hours allocated to XSEDE projects since June 2016
- 15 active science gateways
- 47 education/teaching allocations serving over 904 students
- 1151 (avg concurrent) active VMs in previous qtr, 955 in PY2 to date
- **Highest** user satisfaction in most recent XSEDE survey



# Jetstream Timeline...what comes next?

- Completed our second year of operations with extension to November 2020
- Soliciting Research allocation requests plus Startup and Education allocations – including Science Gateways!
- Adding services as deemed useful/mature (Heat, Magnum, Trove, Manila, etc)
- Atmosphere enhancements on a regular cycle
- Working on partnerships with groups like HubZero and others to extend the value of Jetstream



# 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 Dec 2018 – 15 Jan 2019
- Research allocation: Project desc ( $\leq 10$  pages) and Scaling doc ( $\leq 5$  pages)
  - **We can help!**



# Expanding the reach: Jetstream REU Program



NSF Supplement for undergraduates

- 4 students participated in 2017
- 6 students participated in 2018
- **Planning for 6 in 2019**

- REU student videos on YouTube  
<https://www.youtube.com/user/IUPTI>



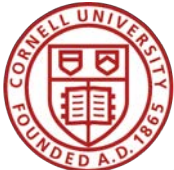
# Where can I get help?

- **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](mailto: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 partners



funded by the National Science Foundation  
Award #ACI-1445604

