

# Jetstream Overview

## A national research and education cloud

Purdue University - September 20, 2017 – West Lafayette, IN

Jeremy Fischer – [jeremy@iu.edu](mailto:jeremy@iu.edu)

Senior Technical Advisor,  
UITS Research Technologies

Fischer, J. (2017). Jetstream Overview: A national research and education cloud. Retrieved from <https://jetstream-cloud.org/archive/publications.php>



# NSF Funding Areas in HPC

---

Traditionally concentrated on enabling petascale capability

- Blue Waters – 13.3 petaflops, 2012
- Stampede – 9.6 petaflops, 2013
- Comet – ~2.0 petaflops, 2014

Has funded research into building clouds and computer science

- CloudLab
- Chameleon

Now funding clouds to do research

- Bridges (Hybrid system)
- Jetstream



funded by the National Science Foundation  
Award #ACI-1445604



# Jetstream - Expanding NSF XD's reach and impact

Lots of stats below –

**tl;dr summary: no one has enough computing resources. Ever. But they need easy access and use.**

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>



funded by the National Science Foundation  
Award #ACI-1445604



# What is Jetstream and why does it exist?

---

- NSF's first production cloud facility
- Part of the NSF eXtreme Digital (XD) program
- Provides on-demand *interactive* computing and analysis
- Enables *configurable* environments and *programmable cyberinfrastructure*
- User-friendly, widely accessible cloud environment
- User-selectable library of preconfigured virtual machines



funded by the National Science Foundation  
Award #ACI-1445604



# What is Jetstream, continued...

---

- Focus on ease-of-use, broad accessibility
- Command line access for those who want it and GUI access for those who don't
- Will support persistent gateways (SEAGrid, Galaxy, GenApp NAMDRunner, CIPRES and others)
- Reproducibility: Share VMs and then store, publish via IU Scholarworks (DOI)



funded by the National Science Foundation  
Award #ACI-1445604



# Who uses Jetstream?

---

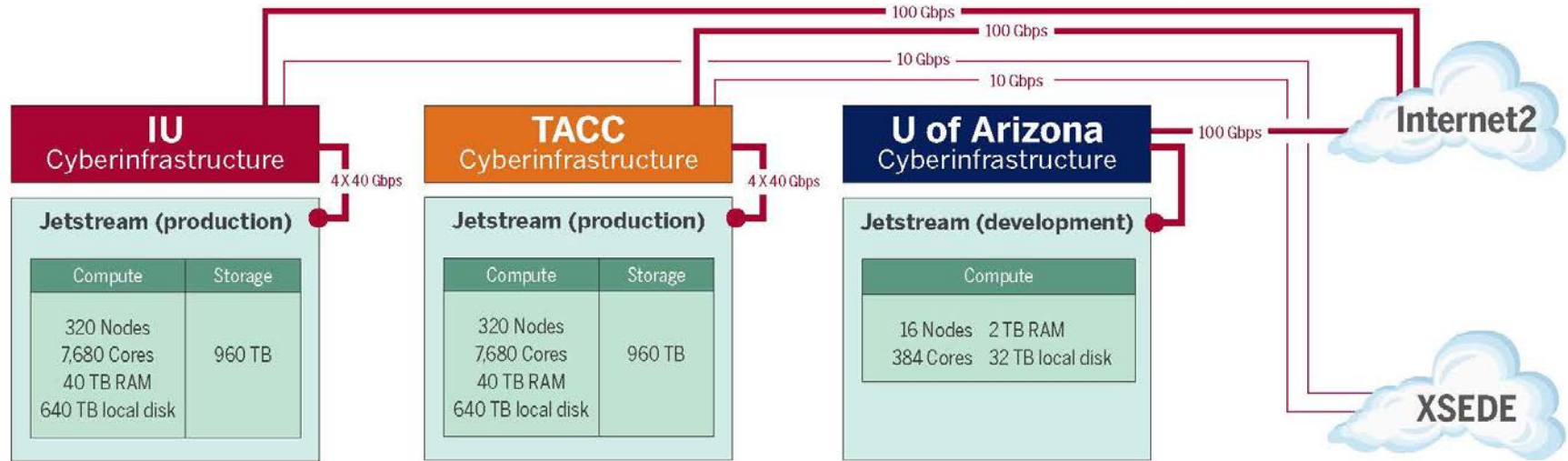
- The researcher needing a handful of cores (1 to 44/vCPU)
- Software creators and researchers needing to create their own customized virtual machines and workflows
- Science gateway creators using Jetstream as either the frontend or processor for scientific jobs
- STEM Educators teaching on a variety of subjects



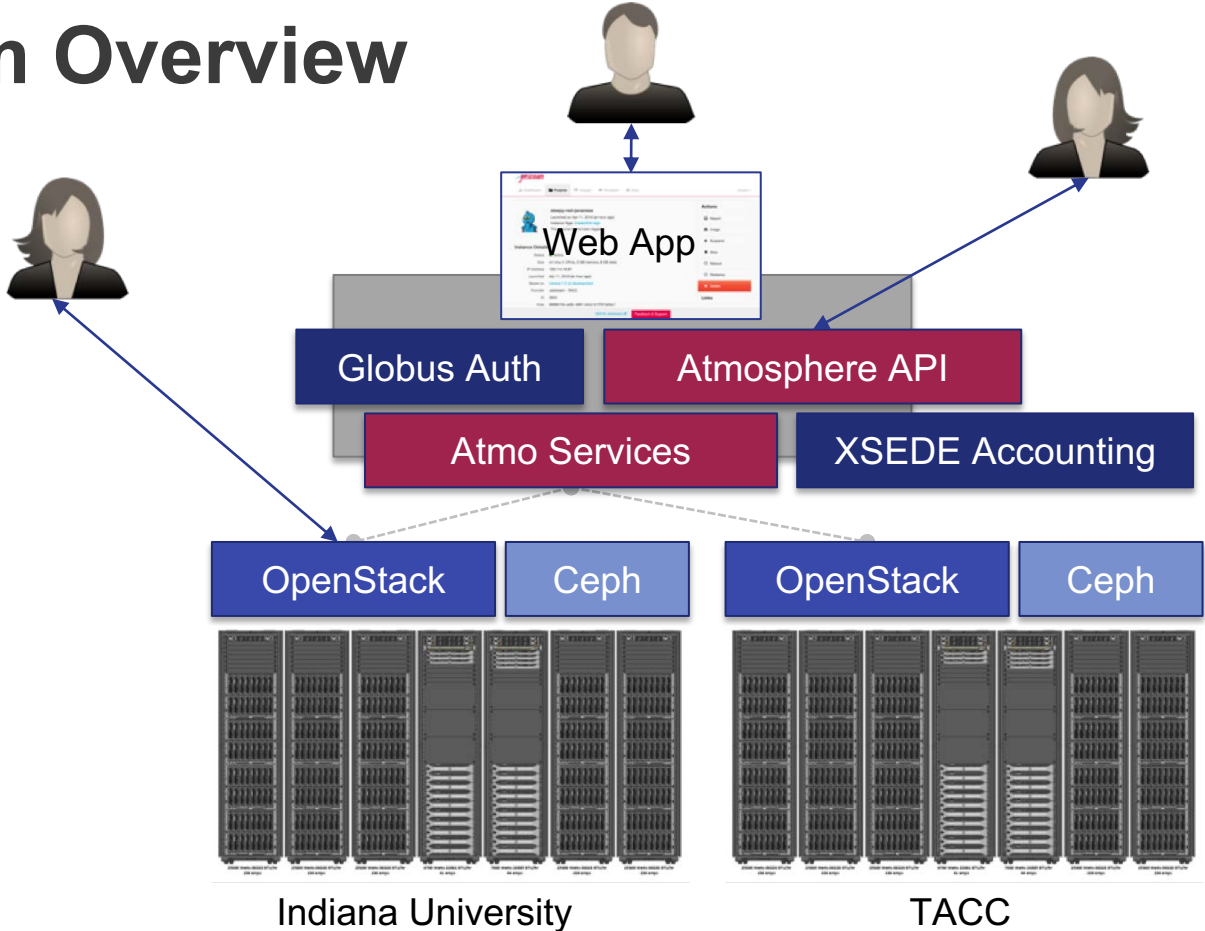
funded by the National Science Foundation  
Award #ACI-1445604



# Jetstream System Overview



# Platform Overview





# Hardware and Instance "Flavors"

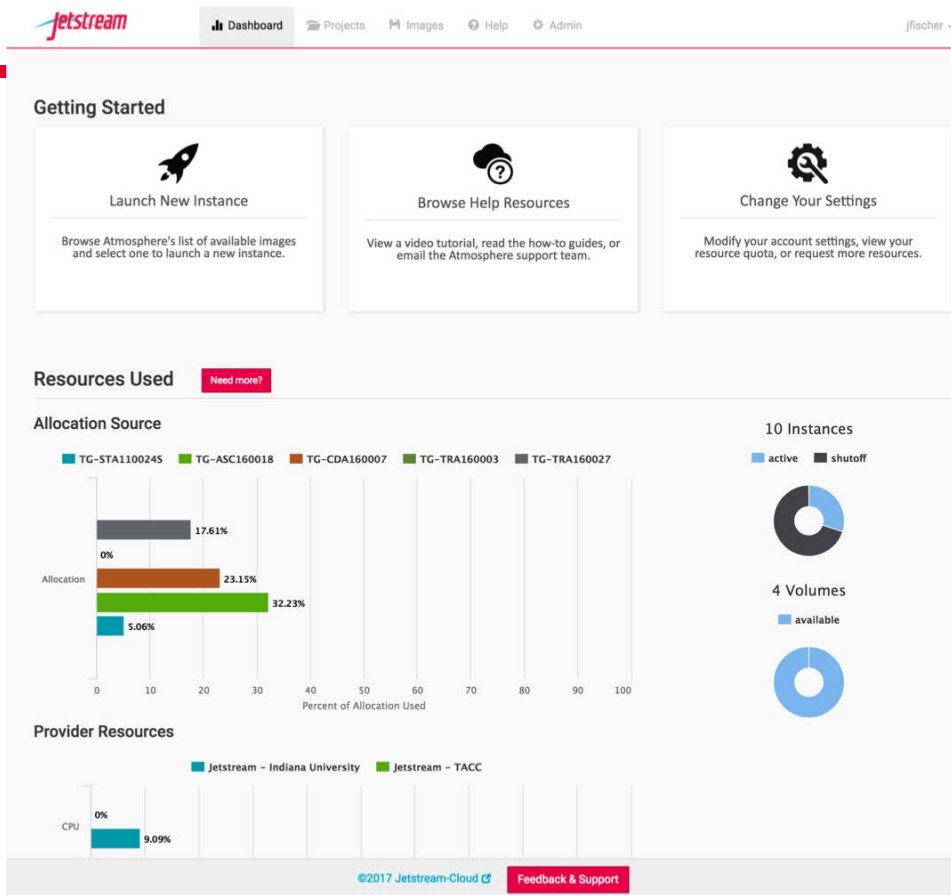
## VM Host Configuration

- Dual Intel E-2680v3 "Haswell"
- 24 physical cores/node @ 2.5 GHz (Hyperthreading on)
- 128 GB RAM
- Dual 1 TB local disks
- 10GB dual uplink NIC
- Running KVM Hypervisor
  
- Short-term *ephemeral* storage comes as part of launched instance
- Long-term storage is XSEDE-allocated
- Implemented as OpenStack Volumes
- Each user can get 10 volumes up to 500GB total storage\*

Flavor	vCPUs	RAM	Storage	Per Node
m1.tiny	1	2	8	46
m1.small	2	4	20	23
m1.medium	6	16	60	7
m1.large	10	30	60	4
m1.xlarge	24	60	60	2
m1.xxlarge	44	120	60	1
s1.large**	10	30	120	4
s1.xlarge**	24	60	240	2
s1.xxlarge**	44	120	480	1

\*\* s1.\* based instances are not eligible to be saved into a customized image

# The Jetstream Atmosphere web interface



# Using Jetstream VMs

---

Manipulating Jetstream VMs:

- Jetstream Atmosphere web interface
- Direct API access via OpenStack command line or Horizon access
  - API access enables Science Gateways and other always on services or on demand use cases; e.g. elastic compute techniques

Primary methods of logging into Jetstream VMs to work

- Interactive user access via web interface with VNC/SSH
- Direct VNC/SSH to individual instances



funded by the National Science Foundation  
Award #ACI-1445604



# HPC vs Cloud

---

Adapting to a different environment:

- No reservations, no queueing
- More interactive use and less/no batch queuing
- What? No parallel filesystem?!?
- Being your own admin – hey, we have root!
- You really can have almost any (linux) software you want\*\*
- Constantly getting new features (<https://www.openstack.org/software/project-navigator/>)

\*\* Here there be dragons...



funded by the National Science Foundation  
Award #ACI-1445604



# Thinking about VMs...



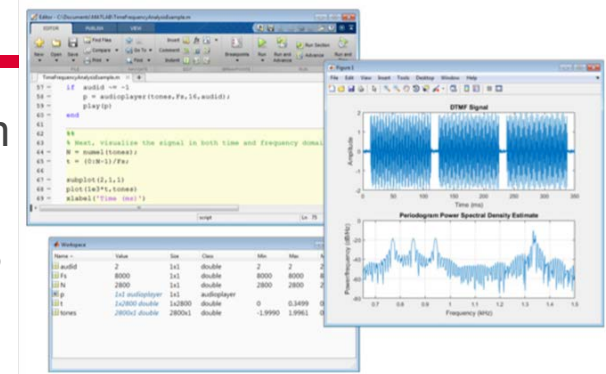
Cows, not pets: pets take great amount of care, feeding, and you name them; cows you intend to have high turnover and you give them numbers.

-- Mike Lowe (Jetstream architect)

# Jetstream for engineering researchers (and others)

Matlab and 52 standard toolkits are installed on Jetstream

You do NOT need to have a local license to use MATLAB on Jetstream



If you are a researcher, and MATLAB and one of the standard first 52 toolkits that come with MATLAB help you... you're ready to go!

If you are an engineering researcher, and you need other tools... Let us know – we are happy to consider other requests

# Jetstream for Education

---

Jetstream has been used in multiple IU Informatics Graduate Courses

- INFO 535 – Management, Access, and Use of Big and Complex Data
- INFO 590 – Topics in Informatics

BlueWaters Workflow Workshop

Multiple Research Data Alliance Workshops

Upcoming workshops/classes on Galaxy, data analysis in finance using R, security and intrusion detection, and principles in cloud computing



funded by the National Science Foundation  
Award #ACI-1445604



# Another Use Case: Galaxy riding Jetstream

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

- The main project instance (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 (e.g., new tools)

## Use Jetstream as a *bursting* platform

- From Galaxy Main, offload jobs onto a remote Slurm cluster running on Jetstream instances
- Run Galaxy Interactive Environments (i.e., Dockerized IPython/RStudio containers) in an isolated environment on 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



funded by the National Science Foundation  
Award #ACI-1445604





# 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



funded by the National Science Foundation  
Award #ACI-1445604



# Requesting access to Jetstream

---

- You can request startup allocations anytime. (Startups are simple!)
- You can request allocations for educational use anytime.
- Next submission period for large allocations is NOW- 15 Sept 2017– 15 Oct 2017.
- We are happy to help you prepare a request and create a successful proposal.
- You do not have to have prior use of Jetstream to be successful.

# Allocation types and docs needed for each

---

- Startup allocation (apply anytime)
  - Current CV for PI and any Co-Pis
  - Brief abstract/description of work
- Education allocation (apply anytime)
  - Current CV for PI and any Co-PIs
  - Syllabus/Class/Workshop description
  - Description of use --> justification of SUs requested
- Research allocation (quarterly allocation window)
  - Current CV for PI and any Co-PIs
  - Main project description (up to 10 pages unless > 15M SUs, then 15 pages)
  - Scaling doc (up to 5 pages)

# Jetstream Project 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 – over 1 million computing hours used to date for MD
- Coming gateways: CIPRES Gateway, The Neuroscience Gateway, ChemCompute gateway, UltraScan III

# Jetstream Overall Highlights

---

As of September 1, 2017:

- 322 active XSEDE projects covering 59 fields of science and 2000+ active users representing 189 institutions
- Over 70 million CPU hours allocated to XSEDE projects since June 2016
- 9 active science gateways
- 40 education/teaching allocations serving over 600 undergraduate and graduate students

# Jetstream Timeline...what comes next?

---

- Completed our first year of operations on September 1, 2017
- 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

# 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>

Training Videos / Virtual Workshops (TBD)



funded by the National Science Foundation  
Award #ACI-1445604



# Jetstream Fun: Happy cluster / Angry Cluster





# Jetstream Partners



INDIANA UNIVERSITY

PERVASIVE TECHNOLOGY INSTITUTE



JOHNS HOPKINS  
UNIVERSITY

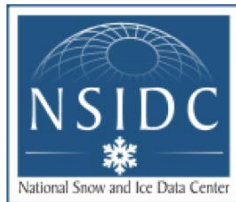
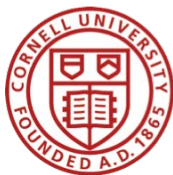


THE UNIVERSITY  
OF ARIZONA

TACC



THE UNIVERSITY OF  
CHICAGO



JSU | JACKSON  
STATE  
UNIVERSITY



funded by the National Science Foundation  
Award #ACI-1445604



# What are we going to cover today in the hands on portion? (The 40,000 foot view...)

---

Getting on to Jetstream's web interface

The Dashboard – features, settings, options

Images on Jetstream today and in the future

Projects and the resources they contain

Launching an instance

Accessing, using, and customizing your instance

Talking about custom images

Preserving your image for publication



funded by the National Science Foundation  
Award #ACI-1445604



# Questions?

---

Project website: <http://jetstream-cloud.org/>

Project email: [help@jetstream-cloud.org](mailto:help@jetstream-cloud.org) Direct email: [jeremy@iu.edu](mailto:jeremy@iu.edu)

## License Terms

- Fischer, Jeremy. September 20, 2017. Jetstream Overview – University of Arkansas - Fayetteville. Also available at: <http://jetstream-cloud.org/archive/publications.php>
- Jetstream is supported by NSF award 1445604 (Craig Stewart, IU, PI)
- XSEDE is supported by NSF award 1053575 (John Towns, UIUC, PI)
- This research was supported in part by the Indiana University Pervasive Technology Institute, which was established with the assistance of a major award from the Lilly Endowment, Inc. Opinions presented here are those of the author(s) and do not necessarily represent the views of the NSF, IUPTI, IU, or the Lilly Endowment, Inc.
- Items indicated with a © are under copyright and used here with permission. Such items may not be reused without permission from the holder of copyright except where license terms noted on a slide permit reuse.
- Except where otherwise noted, contents of this presentation are copyright 2015 by the Trustees of Indiana University.
- This document is released under the Creative Commons Attribution 3.0 Unported license (<http://creativecommons.org/licenses/by/3.0/>). This license includes the following terms: You are free to share – to copy, distribute and transmit the work and to remix – to adapt the work under the following conditions: attribution – you must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work). For any reuse or distribution, you must make clear to others the license terms of this work.



funded by the National Science Foundation  
Award #ACI-1445604

