

# Jetstream A Cloud System Enabling Learning in Higher Education Communities

SIGUCCS17 – October 3, 2017 – Seattle, WA

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

Senior Technical Advisor,  
UITS Research Technologies

Fischer, J. (2017). Jetstream A Cloud System Enabling Learning in Higher Education Communities. 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?

---

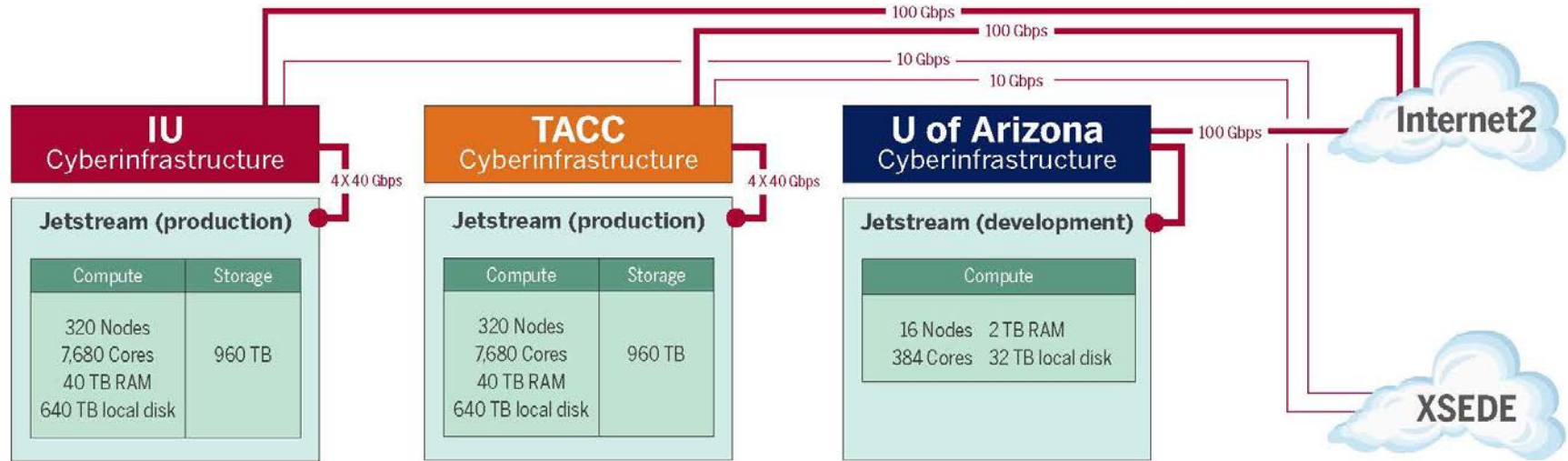
- The researcher needing a handful of cores (1 to 44/vCPU)
- Software creators and researchers needing to create their own customized virtual machines and workflows
- Science gateway creators using Jetstream as either the frontend or processor for scientific jobs
- STEM Educators teaching on a variety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



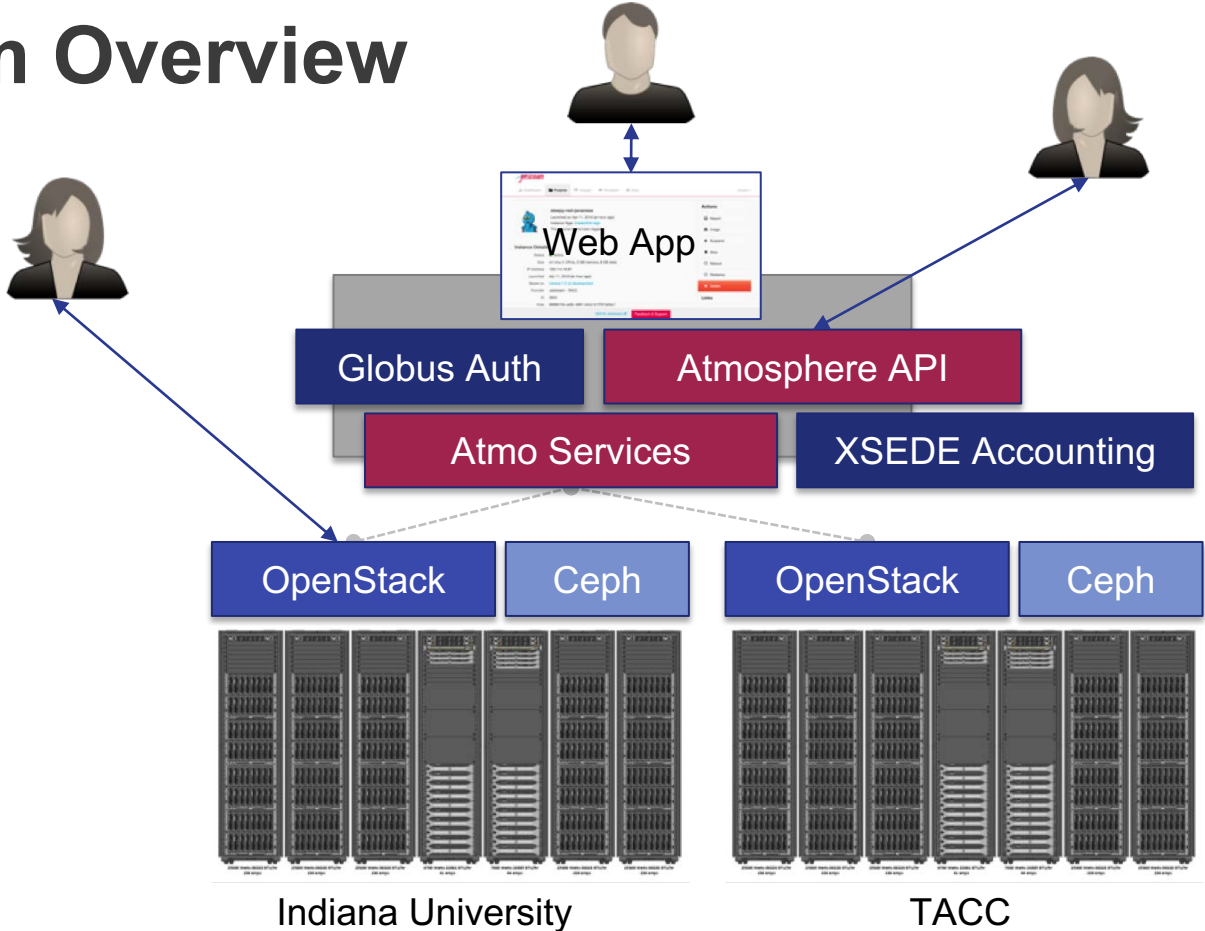
funded by the National Science Foundation  
Award #ACI-1445604



# 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

# 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

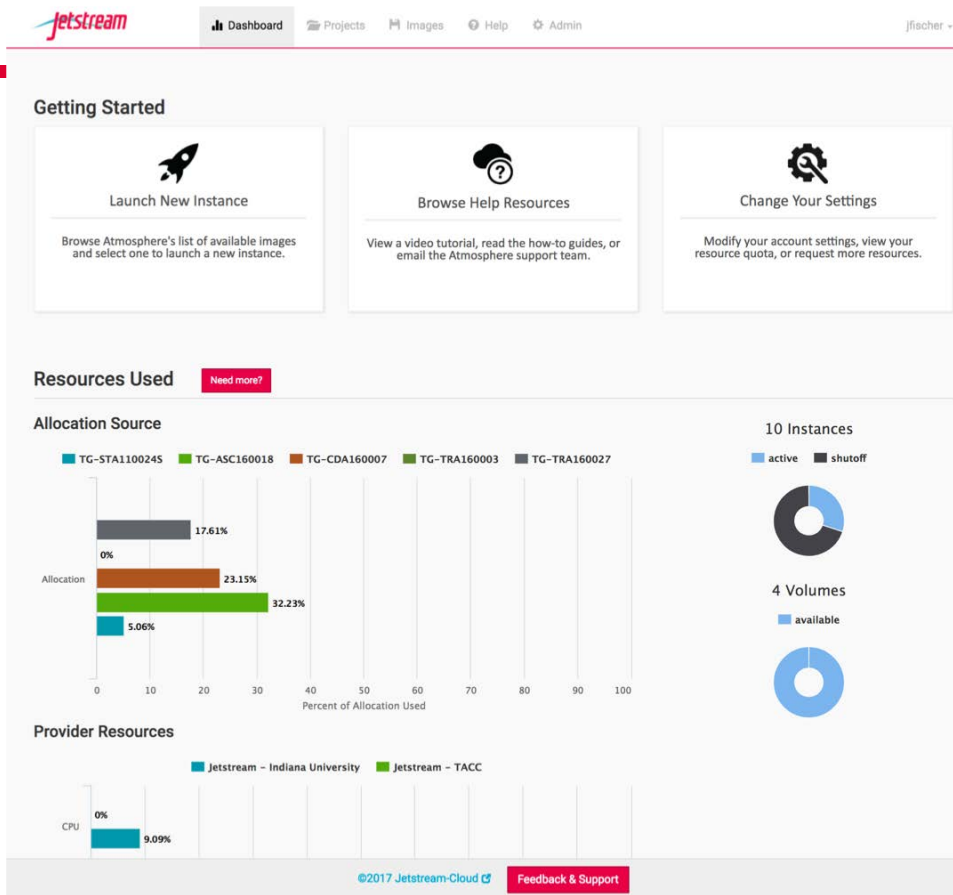
- Interactive user access via web interface with VNC/SSH
- Direct VNC/SSH to individual instances



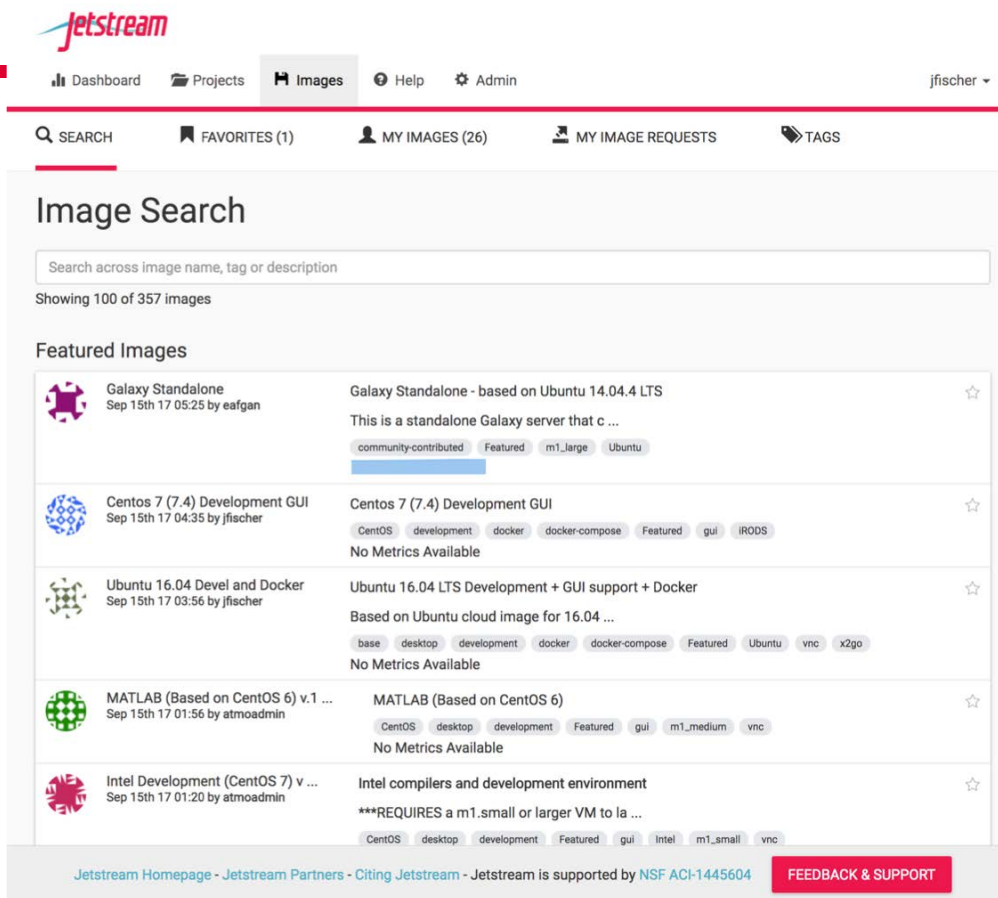
funded by the National Science Foundation  
Award #ACI-1445604








# The Jetstream Atmosphere web interface



# More Atmosphere...



The screenshot displays the Jetstream web interface. At the top, the Jetstream logo is on the left, and navigation links for Dashboard, Projects, Images, Help, and Admin are in the center. A user profile 'jfischer' is on the right. Below the navigation is a search bar and links for Favorites (1), My Images (26), My Image Requests, and Tags. The main content area is titled 'Image Search' and contains a search input field with the placeholder 'Search across image name, tag or description'. Below the search bar, it says 'Showing 100 of 357 images'. The 'Featured Images' section lists five items:

Image Icon	Image Name & Author	Description & Tags	Star
	Galaxy Standalone Sep 15th 17 05:25 by eafgan	Galaxy Standalone - based on Ubuntu 14.04.4 LTS This is a standalone Galaxy server that c ... community-contributed Featured m1_large Ubuntu	☆
	Centos 7 (7.4) Development GUI Sep 15th 17 04:35 by jfischer	Centos 7 (7.4) Development GUI CentOS development docker docker-compose Featured gui IRODS No Metrics Available	☆
	Ubuntu 16.04 Devel and Docker Sep 15th 17 03:56 by jfischer	Ubuntu 16.04 LTS Development + GUI support + Docker Based on Ubuntu cloud image for 16.04 ... base desktop development docker docker-compose Featured Ubuntu vnc x2go No Metrics Available	☆
	MATLAB (Based on CentOS 6) v.1 ... Sep 15th 17 01:56 by atmoadmin	MATLAB (Based on CentOS 6) CentOS desktop development Featured gui m1_medium vnc No Metrics Available	☆
	Intel Development (CentOS 7) v ... Sep 15th 17 01:20 by atmoadmin	Intel compilers and development environment ***REQUIRES a m1.small or larger VM to la ... CentOS desktop development Featured gui Intel m1_small vnc	☆

At the bottom of the page, there is a footer with the Jetstream logo and URL (<http://jetstream-cloud.org/>), navigation links for Jetstream Homepage, Jetstream Partners, and Citing Jetstream, a statement that Jetstream is supported by NSF ACI-1445604, and a red button labeled 'FEEDBACK & SUPPORT'.

# Atmosphere...launching!

use.jetstream-cloud.org

Dashboard Projects Images Help

mmacmane

RESOURCES

gen711\_test

NEW

Instances

You have not added any instances

Volumes

You have not added any volumes

Images

You have not added any images

Links

You have not added any links

Launch an Instance / Basic Options

Basic Info

Instance Name

ADMIN BOOT TEST - Ubuntu 16\_04 Devel and Docker

Base Image Version

1.7

Project

gen711\_test

Resources

Allocation Source

TG-ASC160018

Provider

Jetstream - Indiana University

Instance Size

m1.tiny (CPU: 1, Mem: 2 GB, Disk: 8 GB)

Allocation Used

57% of 1000 SUs from TG-ASC160018

Resources Instance will Use

A total 1 of 132 alloted CPUs

A total 2 of 360 alloted GBs of Memory

Back Advanced Options CANCEL LAUNCH INSTANCE

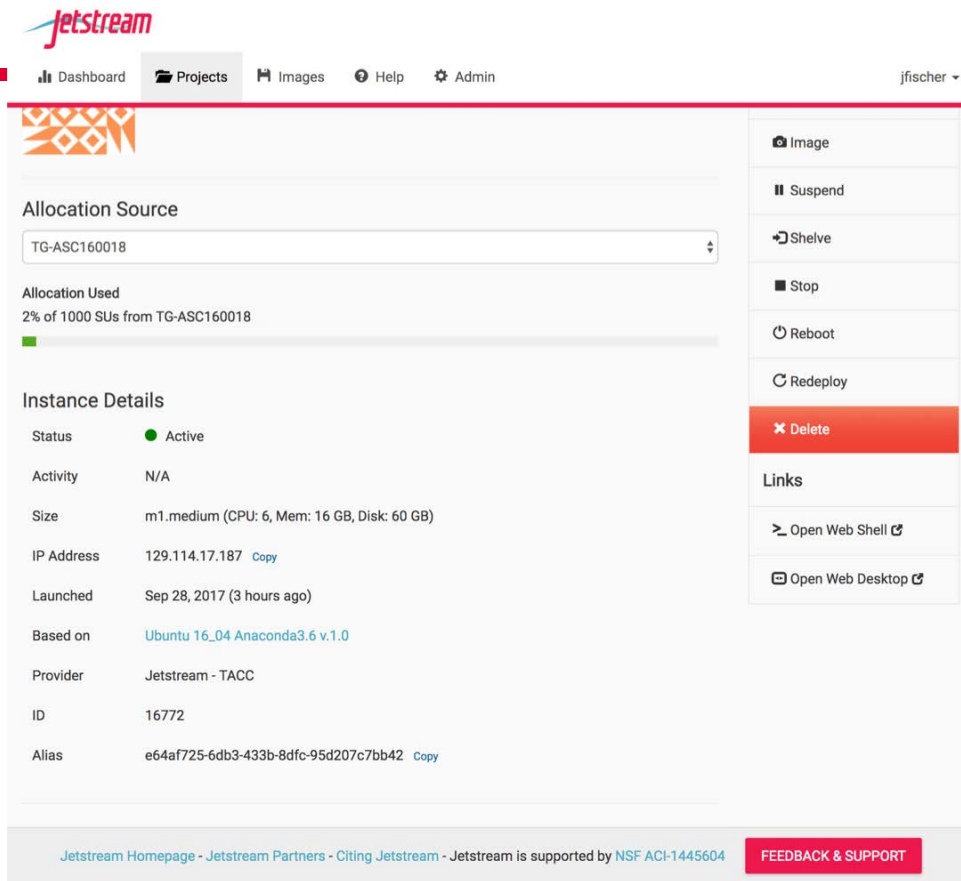
Jetstream

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

Jetstream Homepage - Jetstream Partners - Citing Jetstream - Jetstream is supported by NSF ACI-1445604

FEEDBACK & SUPPORT

# The instance information screen



The screenshot displays the Jetstream web interface for an instance. At the top, the Jetstream logo is on the left, and navigation links for Dashboard, Projects, Images, Help, and Admin are in the center. The user 'jfischer' is logged in on the right. The main content area shows the instance's allocation source as 'TG-ASC160018' and that it is using 2% of 1000 SUs. The instance details table lists its status as Active, size as m1.medium, IP address as 129.114.17.187, and other metadata. A sidebar on the right contains action buttons: Image, Suspend, Shelve, Stop, Reboot, Redeploy, and Delete (highlighted in red). Below these are links for 'Open Web Shell' and 'Open Web Desktop'. The footer includes the Jetstream logo and URL, navigation links, and a 'FEEDBACK & SUPPORT' button.

**Jetstream**

Dashboard Projects Images Help Admin jfischer

**Allocation Source**

TG-ASC160018

**Allocation Used**

2% of 1000 SUs from TG-ASC160018

**Instance Details**

Status	Active
Activity	N/A
Size	m1.medium (CPU: 6, Mem: 16 GB, Disk: 60 GB)
IP Address	129.114.17.187 <a href="#">Copy</a>
Launched	Sep 28, 2017 (3 hours ago)
Based on	<a href="#">Ubuntu 16_04 Anaconda3.6 v.1.0</a>
Provider	Jetstream - TACC
ID	16772
Alias	e64af725-6db3-433b-8dfc-95d207c7bb42 <a href="#">Copy</a>

Image

Suspend

Shelve

Stop

Reboot

Redeploy

Delete

Links

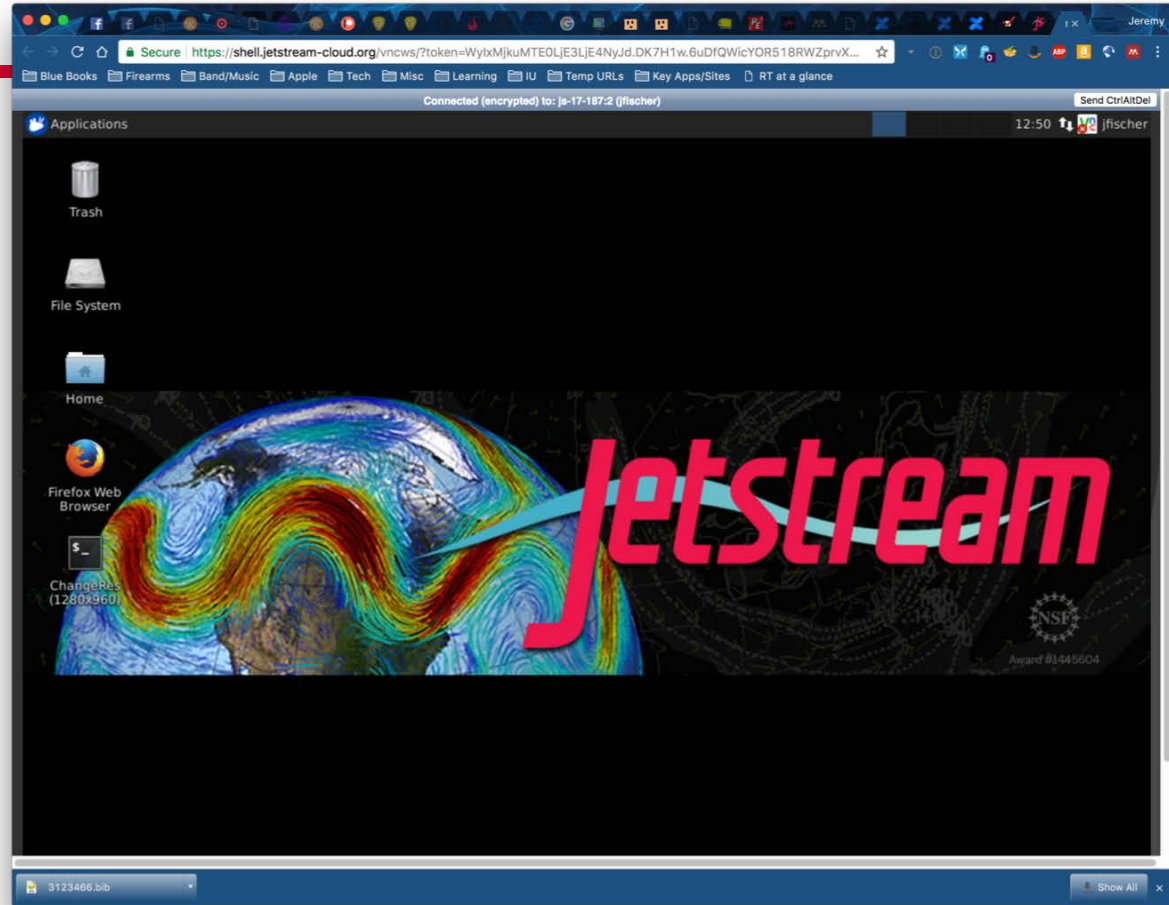
Open Web Shell

Open Web Desktop

Jetstream Homepage - Jetstream Partners - Citing Jetstream - Jetstream is supported by NSF ACI-1445604 **FEEDBACK & SUPPORT**



# The web desktop





# 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...



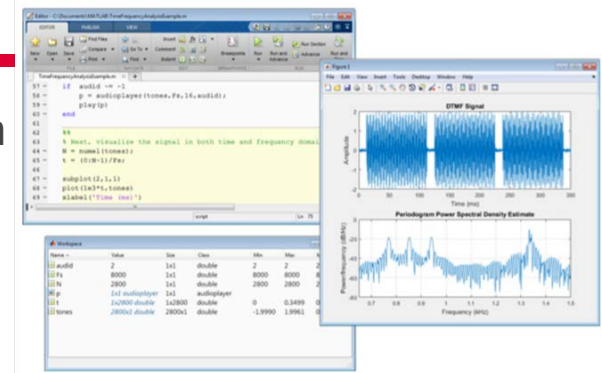
Cows, not pets: pets take great amount of care, feeding, and you name them; cows you intend to have high turnover and you give them numbers.

-- 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

# 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



# Jetstream for Education

---

- Research is a primary function
  - Training current and future researchers is crucial
    - Consistent environments
    - Customizable environments
    - Easy to access (Location independent, Common tools)
    - Easy to use (Familiar environments)

# Surveying Educators on Jetstream

---

- Selected from \*active\* education allocations (mostly or entirely used)
- Selection of university courses (graduate and undergraduate) and workshops
- IRB-approved instrument for contacting PIs and key personnel
- 9 individuals (64.29%) agreed to participate with informed consent and have their results disclosed
- Only four cases are discussed due to individuals consenting to participate but requesting no disclosure, and/or allowing disclosure but not providing enough detail either through survey or follow-up to detail.

# Courses covered in survey results

Event Title	Type	Number of Students	Location	Audience
Analysis of large, complex biobehavioral, bioinformatic, and genomic data sets	Workshop	36	Brandeis University	Graduate students
Bioinformatics: Tools for Genome Analysis	Course	9	Online (via Johns Hopkins University)	Graduate students
'Digital Pedagogy' / 'Introduction to Text Analysis'	Course	30	University of Pittsburgh / Carnegie Mellon University	Graduate students
Management, Access, and Uses of Big and Complex Data	Course	200	Indiana University	Undergraduate and graduate students
Data-Driven Neuroimaging	Workshop	30	University of California San Francisco	Researchers and graduate students
2017 Metagenomics Workshop	Workshop	30	University of California Santa Cruz	Researchers, graduate students, and undergraduates



# Key Points of Success

---

- 100% of Jetstream education PIs surveyed reported their use of Jetstream to be a success in their opinion
- “Jetstream went more smoothly than [commercial provider] in almost every way and seems to perfectly meet our need for training!”
- “fast and flexible access to computational resources at no cost”
- Beneficial to researchers to use in workshops because they could apply for their own allocations afterward and continue using the system that they learned on
- “giving students and instructors the ability to modify, install software on their own and maintain access on demand later on to those images is something of great value, which would not have been easily achieved on our cluster.”



funded by the National Science Foundation  
Award #ACI-1445604



# Where can Jetstream improve?

---

Constructive criticism is necessary!

- Better monitoring of usage by students (workarounds now, but better options coming)
- Improve training materials
- Add videos to training materials
- Persistent IPs in Atmosphere GUI environment
- Improved base images for areas of domain science and instruction

# Engaging educators

---

Common message of “This is great! More people should be using it! Get the word out!” (but no ideas on how to best do that)

- Engage and train IT support staff
- Presenting at conferences (PAG, SC, PEARC, SIGUCCS, targeting domain science conferences)
- Educause and regionals
- Site visits for tutorials
- Other suggestions certainly welcome!

# More usage after the survey...

---

- Supported a month-long workshop in Summer 2017 with 70 users (second workshop for PI!)
- Currently supporting multiple semester-long courses
- Education allocation requests are rising consistently (not necessarily quickly, but consistently – numbers coming up!)

# Conclusions - Jetstream for education

---

- Successful so far, though some hiccups
- Valuable for creating custom, uniform environments
- Must find ways to increase adoption and user base
- More accessible than commercial providers for the long run, so maybe more of a value to resource-constrained facilities

# 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 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

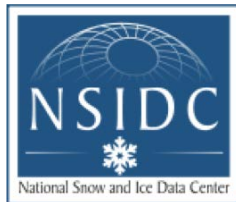
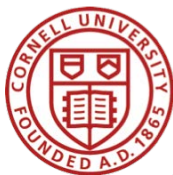


THE UNIVERSITY  
OF ARIZONA

TACC



THE UNIVERSITY OF  
CHICAGO



JSU | JACKSON  
STATE  
UNIVERSITY



funded by the National Science Foundation  
Award #ACI-1445604



# 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. October 3, 2017. Jetstream A Cloud System Enabling Learning in Higher Education Communities. SIGUCCS17. Seattle, WA. 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

