

# Jetstream Overview: A national research and education cloud

UTSA Outreach Event – October 29, 2018 – San Antonio, TX

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

Senior Technical Advisor,  
UITS Research Technologies

Fischer, J. (2018). Jetstream Overview: A national research and education cloud. Kansas City, MO. Retrieved from <https://jetstream-cloud.org/research/publications.php>



# What is XSEDE

---

- Virtual organization
- Distributed cyberinfrastructure
- Support
- Expertise
- Funded by the NSF



# XSEDE offers a variety of resources

---

- Leading-edge distributed memory systems
- Very large shared memory systems
- High throughput systems, e.g. OSG
- Visualization servers
- Accelerators/co-processors including NVIDIA GPUs and Intel ManyCore
- Cloud services

*Many scientific problems have components that call for use of more than one architecture.*

# XSEDE Training and Outreach

---

- Student Programs
- Under-represented Community Engagement
- Champions Program
- Annual XSEDE/PEARC Conference
  
- Training available for XSEDE resources and selected research topics via the XSEDE User Portal

# Stay Connected with XSEDE

---

- XSEDE's public web site is [www.xsede.org](http://www.xsede.org)
- Create an XSEDE User Portal sign on and receive news and notices  
- [portal.xsede.org](http://portal.xsede.org)
- Training events are announced via the public web site; and registrations via the XSEDE User Portal
- For access to additional training and educational resources  
[www.hpcuniversity.org](http://www.hpcuniversity.org)

# NSF Funding Areas in HPC

---

Traditionally concentrated on enabling petascale capability

- Blue Waters – 13.3 petaflops, 2012 (under re-compete)
- Stampede – 9.6 petaflops, 2013 (extended to Stampede2 in 2017 – 18 petaflops)
- Comet – ~2.0 petaflops, 2014

Has funded research into building clouds and computer science

- CloudLab (renewed for 2<sup>nd</sup> phase)
- Chameleon (renewed for 2<sup>nd</sup> phase)

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...but most aren't using XSEDE in any capacity at all.**

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>

# Identifying the potential users

---

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

--- multiple researchers at a number of small colleges and universities



# What is Jetstream and why does it exist?

---

- NSF's first production cloud facility
- Part of the NSF eXtreme Digital (XD) program
- Focus on ease-of-use, broad accessibility
- Provides on-demand *interactive* computing and analysis or persistent gateways
- Enables *configurable* environments and ***programmable cyberinfrastructure***

# 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

# 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...)



# HPC vs Cloud

---

Adapting to a different environment:

- No reservations, no queueing – more interactive usage
- 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...

# Jetstream and way of the cloud...

---

- **Cloudy Technologies:** clouds are more than just virtual machines (VM)
  - **Old way:** robust (expensive) infrastructure, weak (cheap) software
    - You expect the hardware to not fail
    - State is maintained in volatile data structures
  - **Cloudy way:** commodity infrastructure, robust software
    - Expect & plan for infrastructure to fail
    - Put intelligence into the software to handle infrastructure failure
- **And my favorite...**

# Thinking about VMs...

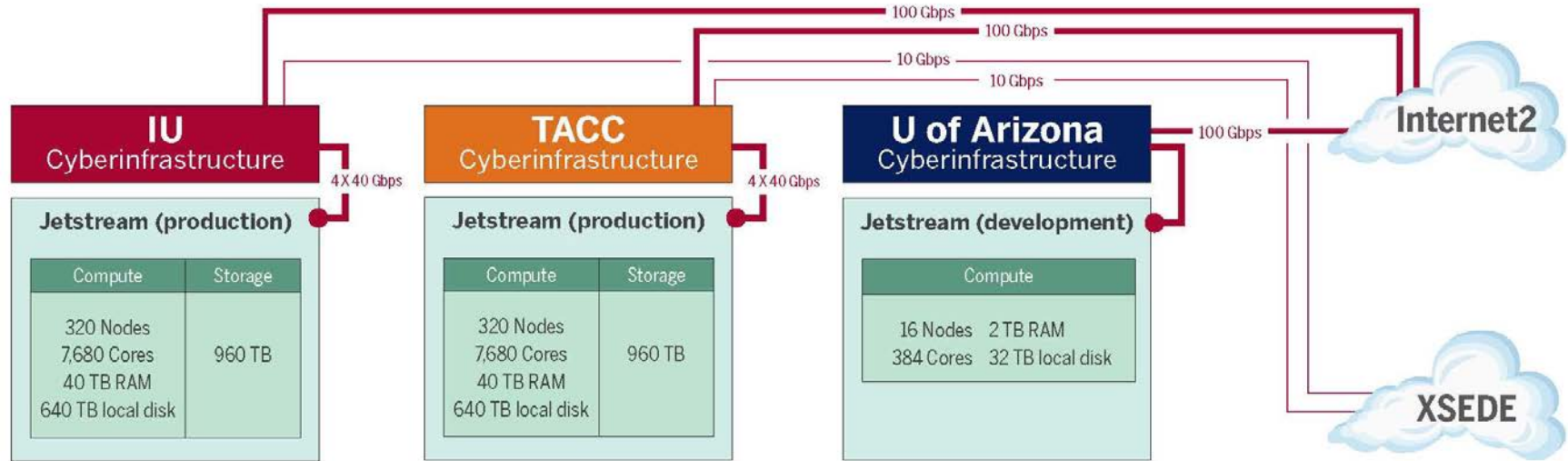


Cows, not pets: pets take great amount of care, feeding, and you name them; cows you intend to have high turnover and you give them numbers.

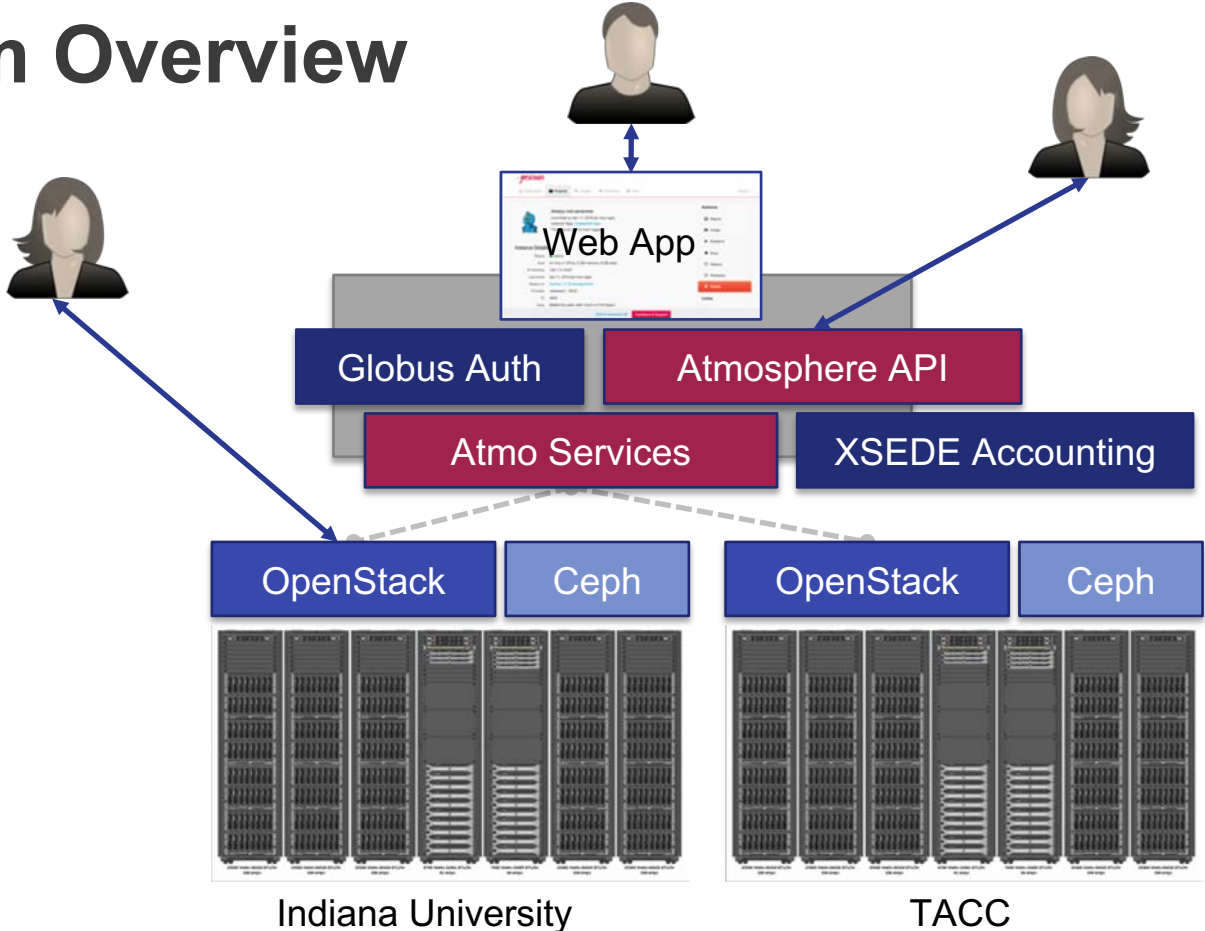
-- Mike Lowe (Jetstream architect)

\*\*some caveats for gateways...

# Jetstream System Overview

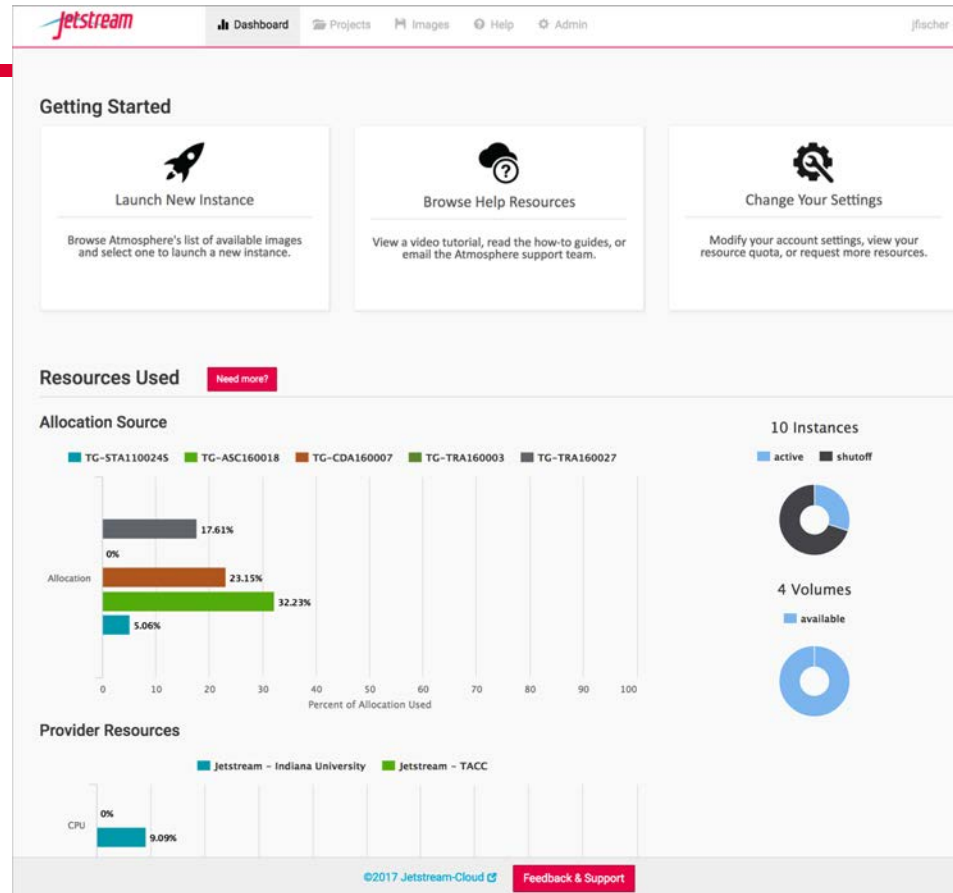


# Platform Overview

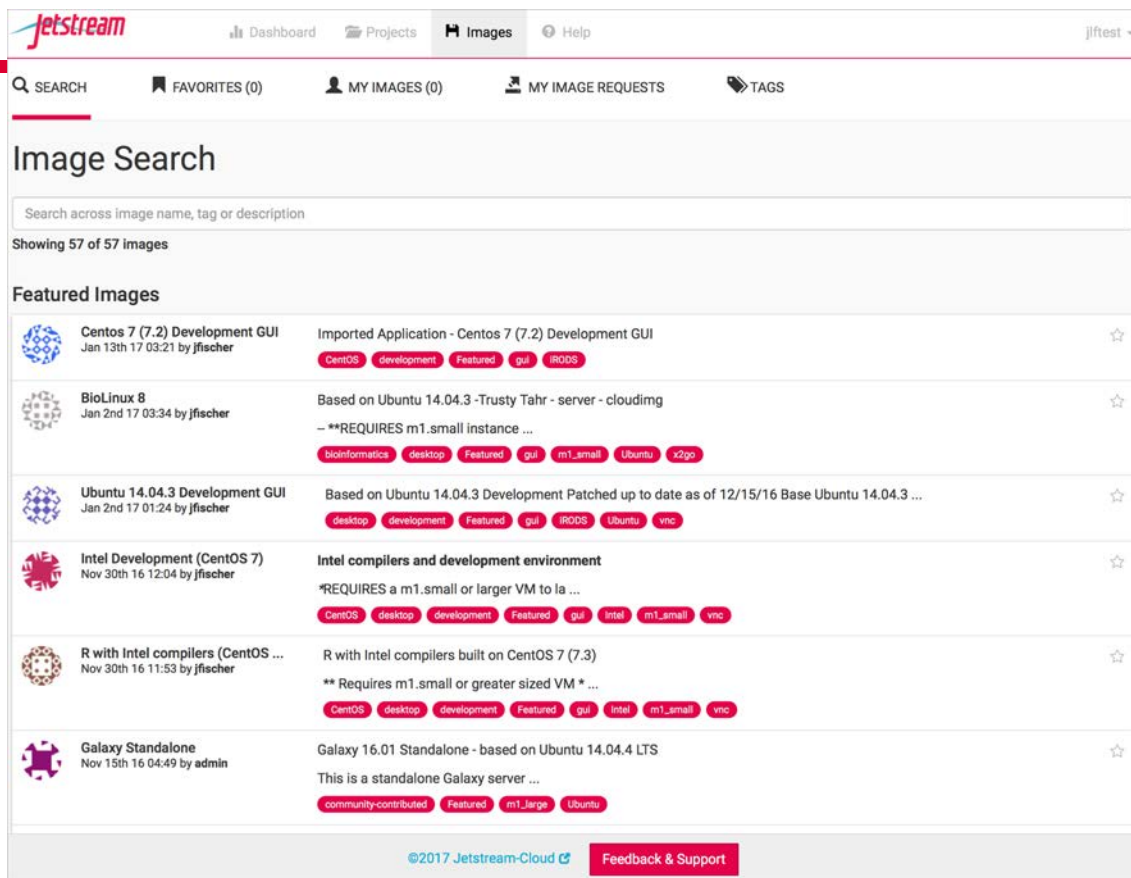




# The Jetstream Atmosphere web interface



# The Jetstream Atmosphere web interface



The screenshot displays the Jetstream Atmosphere web interface. At the top, there is a navigation bar with the Jetstream logo, a dashboard icon, and menu items for 'Dashboard', 'Projects', 'Images', and 'Help'. The user 'jftest' is logged in. Below the navigation bar is a search bar and several filters: 'SEARCH', 'FAVORITES (0)', 'MY IMAGES (0)', 'MY IMAGE REQUESTS', and 'TAGS'. The main content area is titled 'Image Search' and contains a search input field with the placeholder text 'Search across image name, tag or description'. Below the search bar, it indicates 'Showing 57 of 57 Images'. A section titled 'Featured Images' lists six image entries, each with a thumbnail icon, a title, a date, the creator's name, a description, and a set of tags. The tags are displayed in red rounded rectangles. The footer of the interface includes a copyright notice '©2017 Jetstream-Cloud' and a 'Feedback & Support' button.

Image Icon	Title	Date	Creator	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, @RODS
	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, @RODS, 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

# Hardware and Instance "Flavors"

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

\*\* s1.\* storage-rich instances are not eligible to be saved into a customized image

- Short-term *ephemeral* storage comes as part of launched instance
- Long-term storage is XSEDE-allocated
- Implemented as OpenStack Volumes and object storage
- Default storage is modest, but more is available via allocation

# 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

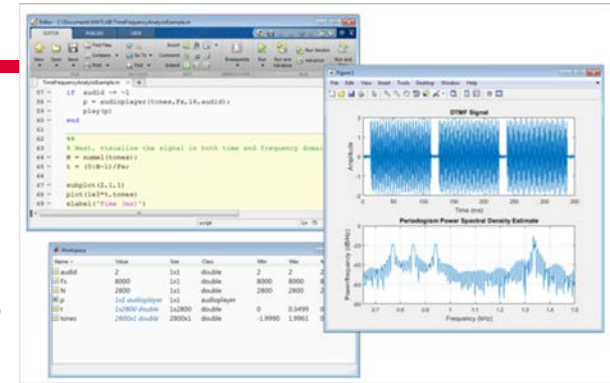
- Interactive user access via web interface with VNC/SSH
- Direct VNC/SSH to individual instances

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%

# Jetstream for engineering researchers (and others)

Matlab and SimuLink and additional 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 or SimuLink... 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

# 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
- ...and others...

Multiple Research Data Alliance Workshops, multiple workshops/classes on Galaxy, data analysis in finance using R, security and intrusion detection, and principles in cloud computing and more!



# 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

# 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 – over 1 million computing hours used to date for MD
- 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\*
- **Highest** user satisfaction in most recent XSEDE survey

# Jetstream Timeline...what comes next?

---

- Completed our second year of operations on September 1, 2018
- 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 Sept - 15 Oct 2018.
- Research allocation: Main project description (up to 10 pages) and Scaling doc (up to 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
- 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/>



funded by the National Science Foundation  
Award #ACI-1445604

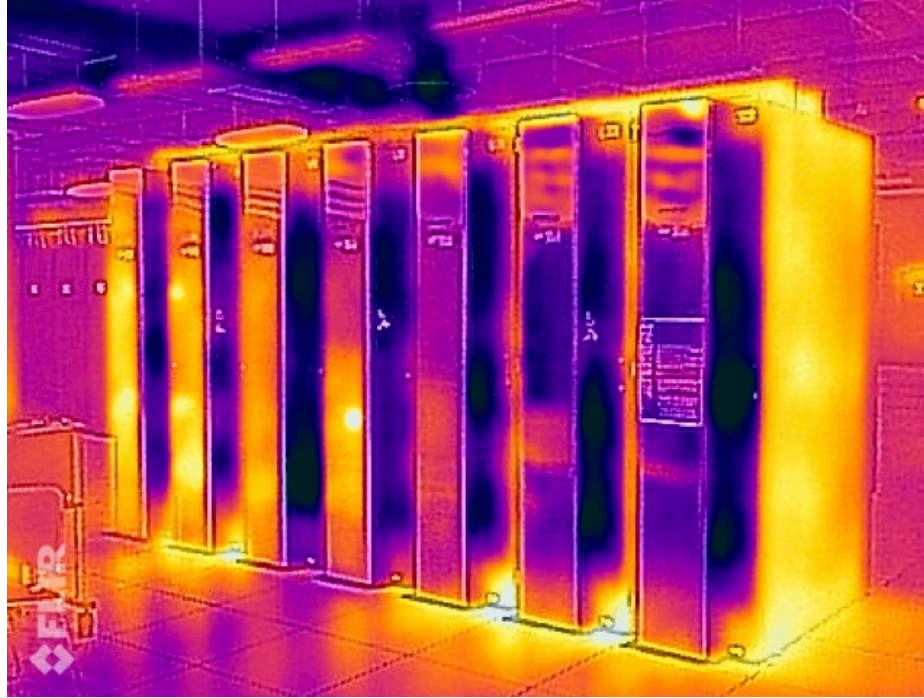


# Jetstream Fun: Happy cluster / Angry Cluster





# Infrared image of Jetstream



**ChilledDoor™**  
Rack Cooling System  
**motivair™**

# Jetstream Partners



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 29, 2018. Jetstream Overview – UTSA Outreach Event . Also available at: <http://Jetstream-cloud.org/research/publications.php>
- Jetstream is supported by NSF award 1445604 (David Y. Hancock, 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.

# Cloud Computing Terms...simplified

---

**Image:** a file on a disk. It will be booted to create an...

**Instance:** a running virtual server; i.e. something you can log into.

**Running:** the *instance* is up & running

**Suspended:** the *instance* is memory resident but not running

**Stopped:** the *instance* is shutdown akin to powering down

**Shelved:** the *instance* is shutdown, backedup, and stored

# Cloud Computing Terms...simplified

---

**Flavor:** the size of a running instance; i.e. #core, RAM, disk

**Hypervisor:** the thing the instance runs on; something akin to a software defined hardware compute server.

**Snapshot:** the process of taking an instance and turning it to an image.

**State:** something worth remembering; i.e. the state of the system

# Cloud Computing Terms...simplified (Cont.)

---

**Object store:** a blob of bits; it has a starting address & a size. There may be metadata associated with the object. The data is consumed in a streaming manner.

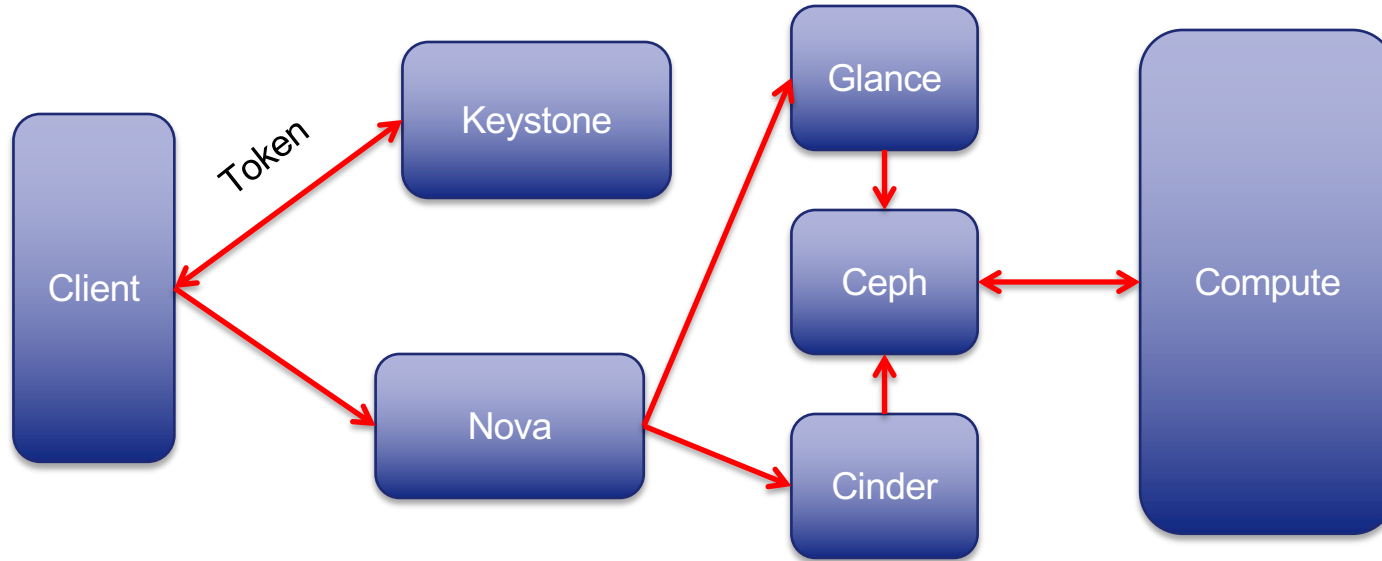
**Block store:** a software defined entity akin to an unformatted hardware disk drive.

**Filesystem:** hierarchical in nature, directories & files, ability to open, seek, read, write.

**Persistent storage:** If you pull the plug, it will still exist when power is restored. Safe to store data or state here.

**Ephemeral storage:** If you pull the plug, it no longer exists. (Don't put your data here!!!)

# OpenStack Overview



# Getting into the hands on part -

---

Open <https://use.Jetstream-cloud.org> in your browser

Login slips will be distributed momentarily!