



# RESEARCH TECHNOLOGIES

UNIVERSITY INFORMATION TECHNOLOGY SERVICES

*rt.iu.edu*



# Jetstream Overview: A national research and education cloud

University of Hawai'i  
February 15, 2019 – Honolulu, HI.

*Jeremy Fischer – [Jeremy@iu.edu](mailto:Jeremy@iu.edu) - Indiana University*

*Senior Technical Advisor,*

*UITS Research Technologies*

*Fischer, J. (2019). Jetstream Overview: A national research and education cloud. Honolulu, IN. Retrieved from <https://jetstream-cloud.org/research/publications.php>*

# NSF Funding Areas in HPC

*Traditionally concentrated on enabling petascale capability*

- *Blue Waters – 13.3 petaflops, 2012 (Frontera awarded late 2018, coming in 2019)*
- *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*



# 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*



# 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***



# 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...they're coming!)*
- *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\*\*

\*\* 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...

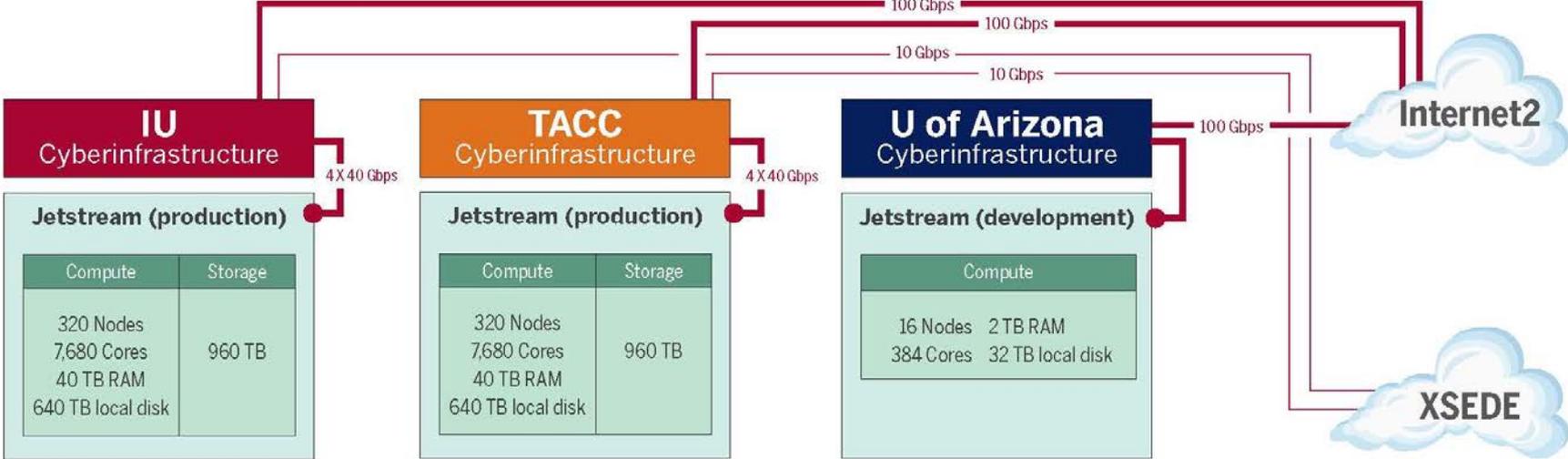
*Cows, not pets: pets take great amount of care, feeding, and you name them; cows you intend to have high turnover and you give them numbers.*

*-- Mike Lowe (Jetstream architect)*

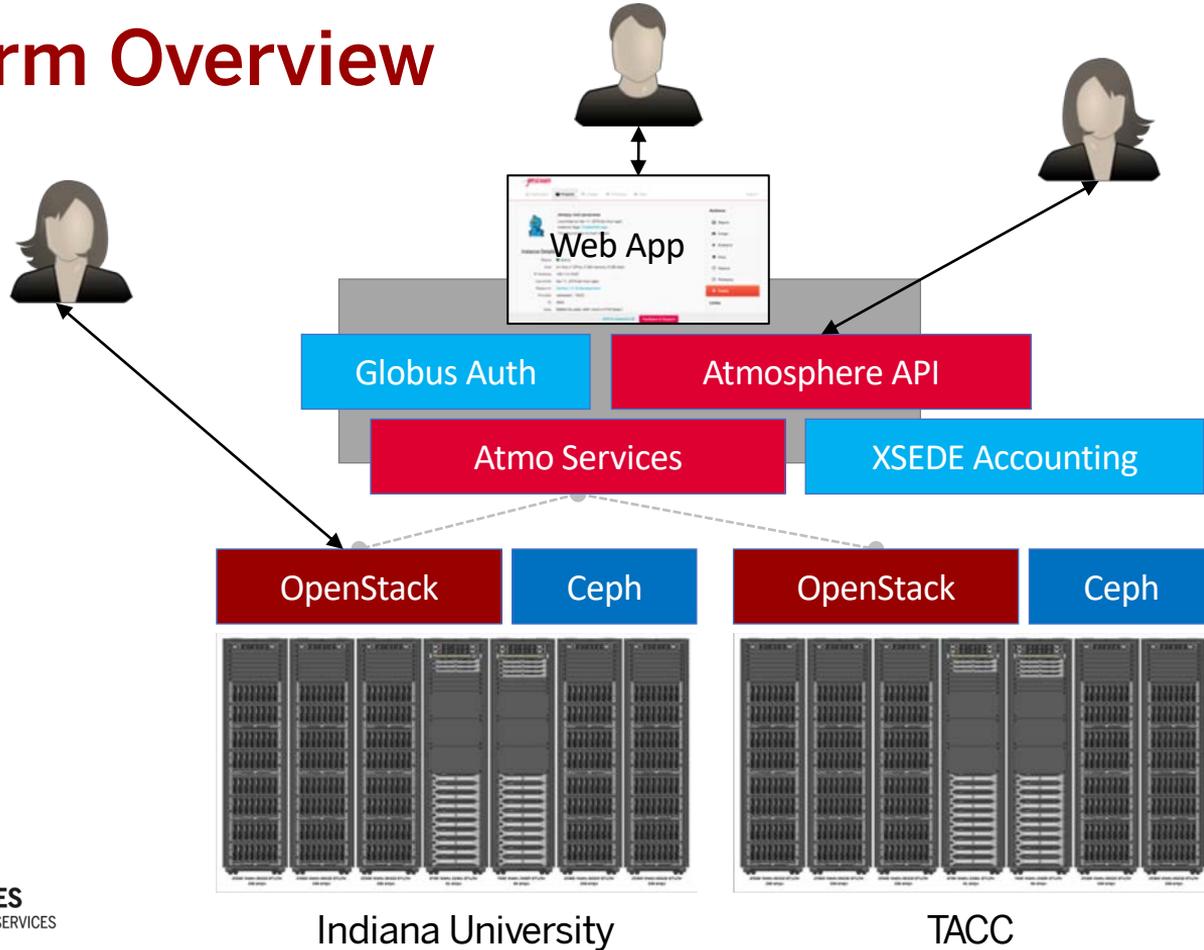
*\*\*some caveats for gateways...*



# Jetstream System Overview



# Platform Overview



# The Jetstream Atmosphere web interface

The screenshot displays the Jetstream Atmosphere web interface. At the top, there is a navigation bar with the Jetstream logo and links for Dashboard, Projects, Images, Help, and Admin. The user's name 'jlfacher' is visible in the top right corner.

The main content area is divided into several sections:

- Getting Started:** Three cards provide quick actions: 'Launch New Instance' (with a rocket icon), 'Browse Help Resources' (with a question mark icon), and 'Change Your Settings' (with a gear icon). Each card includes a brief description of the action.
- Resources Used:** A section with a 'Need more?' button. It contains two charts:
  - Allocation Source:** A horizontal bar chart showing the percentage of allocation used by different sources. The data is as follows:

Allocation Source	Percent of Allocation Used
TG-STA1100245	5.06%
TG-ASC160018	32.23%
TG-CDA160007	23.15%
TG-TRA160003	0%
TG-TRA160027	17.61%
  - Provider Resources:** A horizontal bar chart showing the percentage of provider resources used. The data is as follows:

Provider	Percent of Provider Resources Used
Jetstream - Indiana University	9.09%
Jetstream - TACC	0%
- Summary Metrics:** On the right side, there are three donut charts:
  - 10 Instances:** A donut chart showing the distribution of 10 instances between 'active' (blue) and 'shutoff' (black) states.
  - 4 Volumes:** A donut chart showing the distribution of 4 volumes between 'available' (blue) and 'shutoff' (black) states.

At the bottom of the interface, there is a footer with the copyright notice '©2017 Jetstream-Cloud' and a 'Feedback & Support' button.

# 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 is logged in as 'jlfest'. Below the navigation bar, there is a search bar and several user-specific links: SEARCH, FAVORITES (0), MY IMAGES (0), MY IMAGE REQUESTS, and 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</b> <b>development</b> <b>Featured</b> <b>gui</b> <b>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</b> <b>desktop</b> <b>Featured</b> <b>gui</b> <b>m1_small</b> <b>Ubuntu</b> <b>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</b> <b>development</b> <b>Featured</b> <b>gui</b> <b>iRODS</b> <b>Ubuntu</b> <b>vmc</b>	☆
 <b>Intel Development (CentOS 7)</b> Nov 30th 16 12:04 by <b>jfischer</b>	<b>Intel compilers and development environment</b> *REQUIRES a m1.small or larger VM to la ... <b>CentOS</b> <b>desktop</b> <b>development</b> <b>Featured</b> <b>gui</b> <b>intel</b> <b>m1_small</b> <b>vmc</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</b> <b>desktop</b> <b>development</b> <b>Featured</b> <b>gui</b> <b>intel</b> <b>m1_small</b> <b>vmc</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</b> <b>Featured</b> <b>m1_large</b> <b>Ubuntu</b>	☆

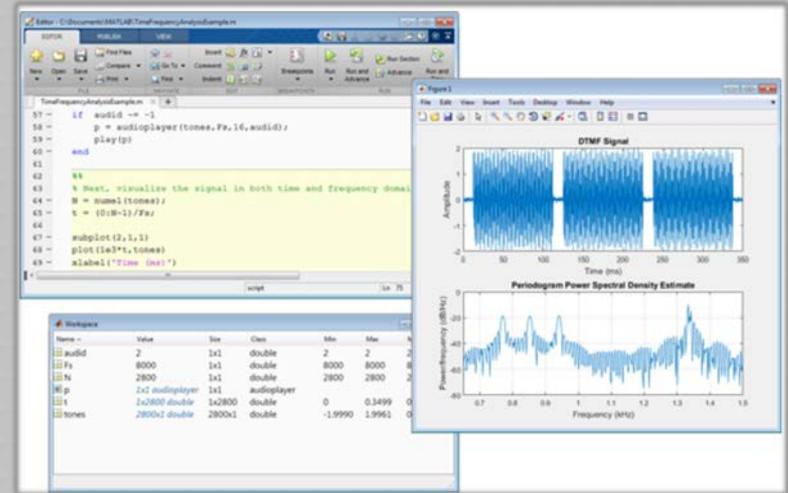
©2017 Jetstream-Cloud [Feedback & Support](#)



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% Increase	4.24%	0.61%
Biological Sciences	89	15,041,928	72.18% Increase	18.40%	3.59%
Biophysics	86	3,627,026	44.15% Decrease	4.44%	13.56%
Computer Science	72	6,883,269	33.28% Increase	8.42%	2.98%
Earth Sciences	37	5,476,250	3706% Increase	6.70%	4.60%
Education and Training	128	16,599,512	2.62% Increase	20.31%	4.66%
Engineering	13	520,690	71.21% Increase	0.64%	1.75%
Materials Science	6	1,035,508	100% Increase	1.27%	13.89%
Mathematics	13	688,505	150.37% Increase	0.84%	0.90%
Molecular Science/Biochemistry	21	4,254,643	10.15% Decrease	5.20%	5.83%
Neuroscience	19	4,708,180	327.87% Increase	5.76%	1.98%
Physics	10	2,440,581	15.58%	2.99%	8.65%

# Jetstream for engineering researchers (and others)

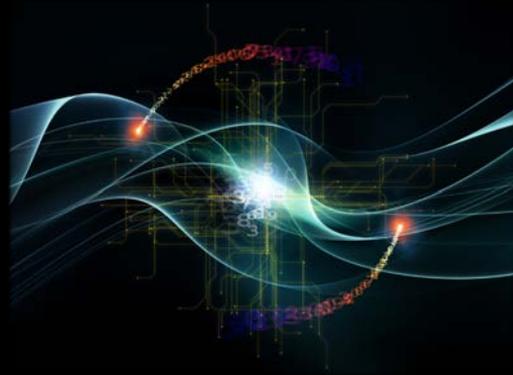
- *Matlab and SimuLink and additional toolkits are installed on Jetstream*
- *You do NOT need to have a local license to use MATLAB on Jetstream*
- *If you are a researcher, and MATLAB or SimuLink... you're ready to go!*
- *If you are an engineering researcher, and you need other tools... let us know!*



# 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

- **Simulations of Nanoscale Biomolecular Systems** - Aleksei Aksimentiev, University of Illinois Champaign-Urbana
- **The Neuroscience Gateway** - Amitava Majumdar, University of California, San Diego
- **Parallelizing Development of Immunomics and Genomics Tools** - Ramy Arnaout, Beth Israel Deaconess Medical Center
- **Atmospheric Science in the Cloud: Enabling Data-Proximate Science** – Mohan Ramamurthy, UNIDATA (University Corporation for Atmospheric Research)
- **Science and Engineering Applications Grid (SEAGrid): A Gateway for Simulation of Molecular and Material Structures and Dynamics** – Sudhakar Pamidighantam, Indiana University



# Jetstream usage highlights – 1 February 2019

- 409 active XSEDE projects covering 72 fields of science and over 2250 active users representing 196 institutions
- **80%** of Jetstream users have **not used any other XSEDE system**
- >180M CPU hours allocated to XSEDE projects since June 2016
- 19 active science gateways
- 47 education/teaching allocations serving over 975 students
- 1288 mean active VMs in previous qtr, 1530 peak active VM count
- **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 Mar 2019 – 15 Apr 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>

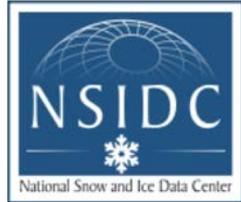


# Where can I get help?

- *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

