

rt.iu.edu



**RESEARCH TECHNOLOGIES**  
UNIVERSITY INFORMATION TECHNOLOGY SERVICES



**RESEARCH TECHNOLOGIES**

UNIVERSITY INFORMATION TECHNOLOGY SERVICES

# Jetstream Overview: A national research and education cloud

RMACC HPC Symposium 2020

May 21, 2020 – Boulder, CO / Bloomington, IN / Webinar

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

Manager, Jetstream Cloud, UITS Research Technologies

Fischer, J. (2020). Jetstream Overview: A national research and education cloud. Webinar. 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 (Expanse awarded mid 2019, coming in 2020)

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) – (Bridges2 awarded mid 2019, coming in 2020)
- 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

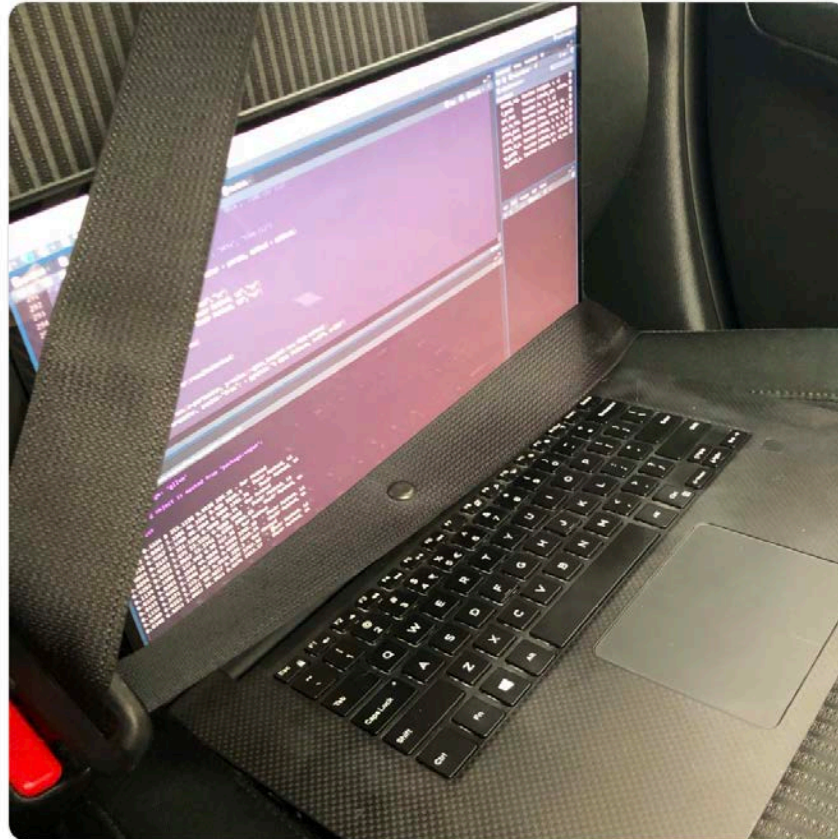




**Sarah Romanes** @sarah\_romanec · Mar 17



When you have to drive to work - but R has also only completed 10hrs of 24hrs worth of simulations. Can I go in the T2 lane with this thing? 😂 #rstats #sydneytraffic



# 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...

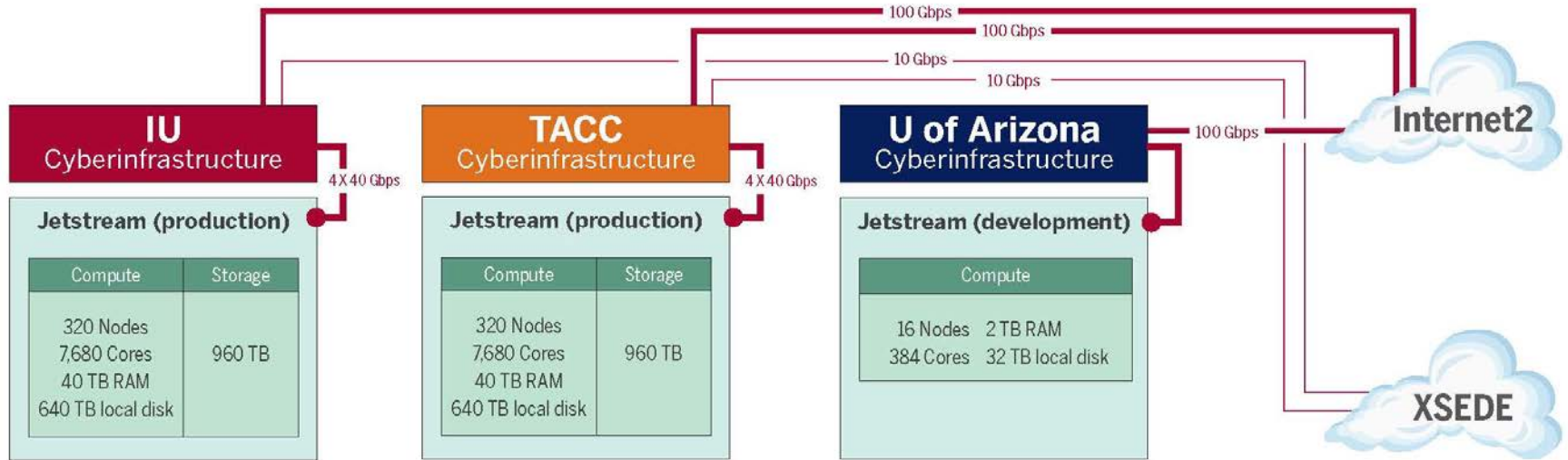
Cows, not pets: pets take great amount of care, feeding, and you name them; cows you intend to have high turnover and you give them numbers.

-- Mike Lowe (Jetstream architect)

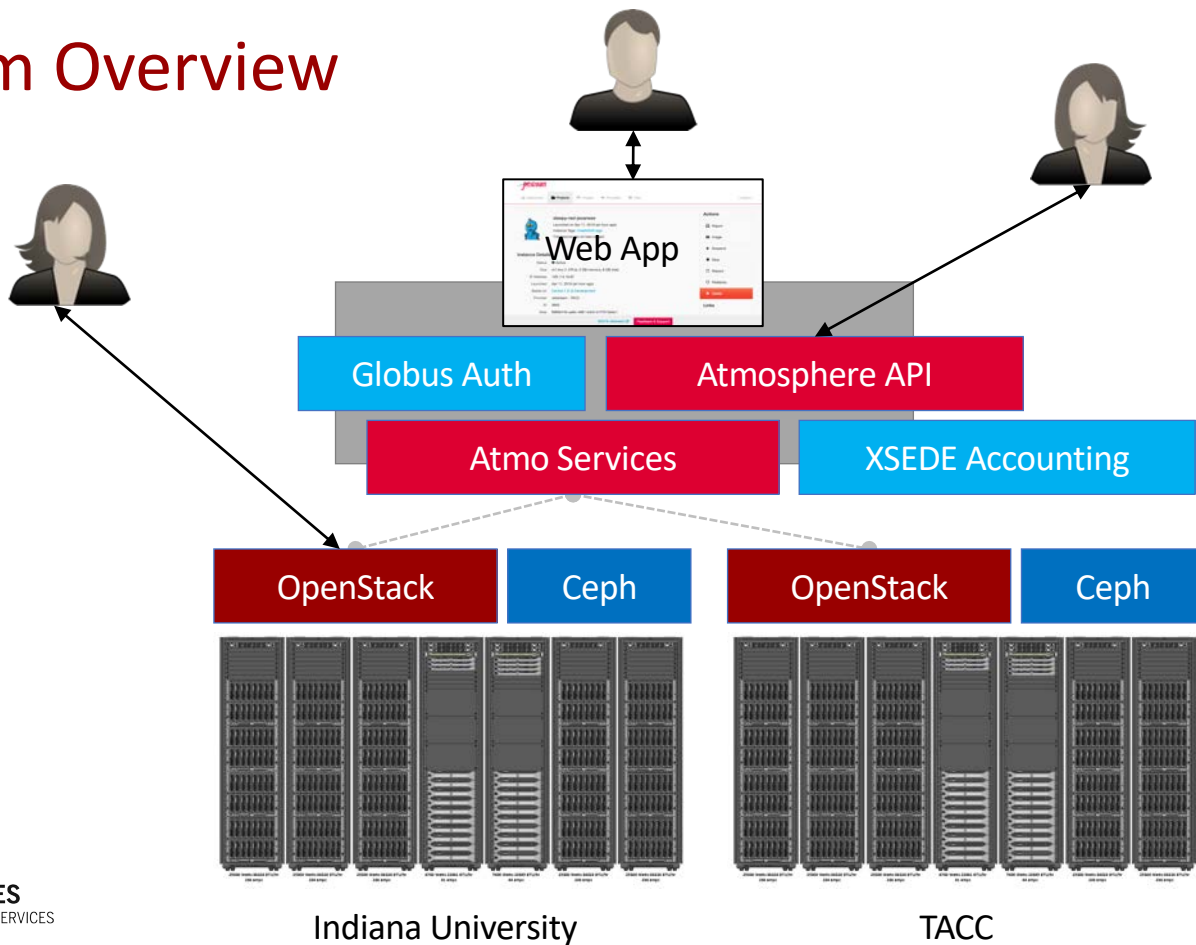
\*\*some caveats for gateways...



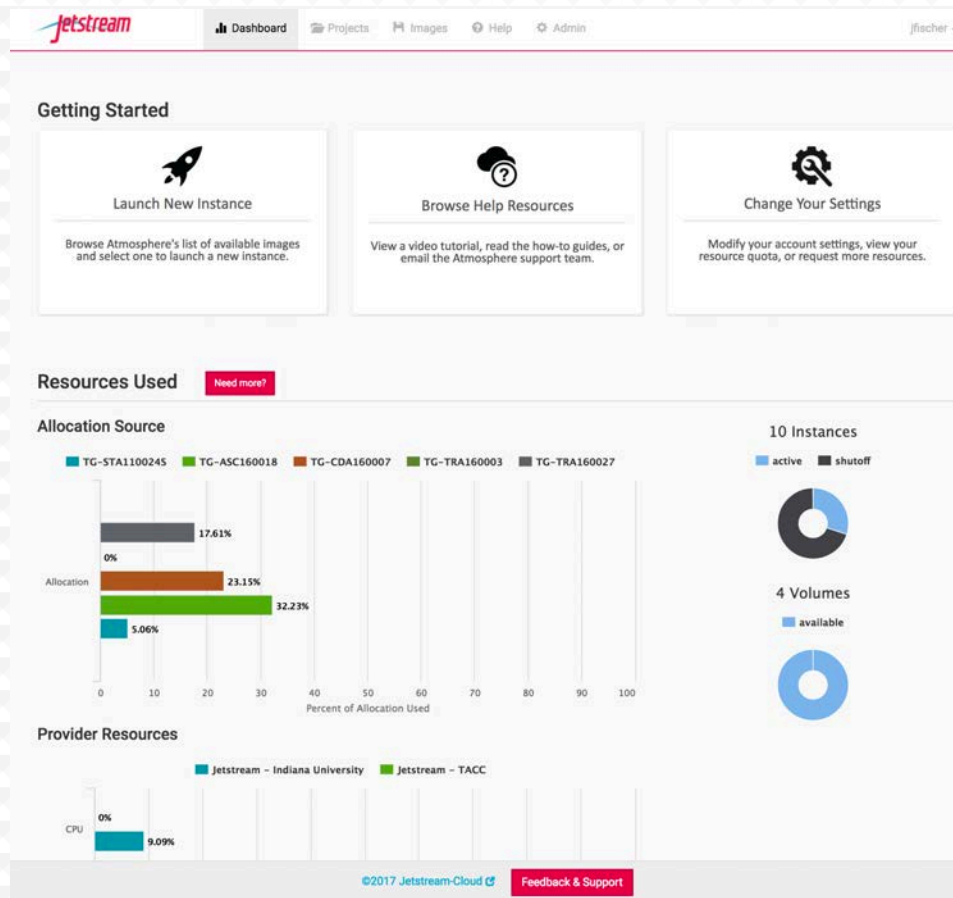
# Jetstream System Overview



# Platform Overview



# The Jetstream Atmosphere web interface



# The 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, and Help. Below this is a search bar and navigation links for FAVORITES (0), MY IMAGES (0), MY IMAGE REQUESTS, and TAGS. The main content area is titled "Image Search" and contains a search input field. Below the search field, it indicates "Showing 57 of 57 Images". A section titled "Featured Images" lists several image entries, each with a thumbnail icon, title, creation date, author, description, and a set of tags.

Thumbnail	Title	Date	Author	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, iRODS
	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, iRODS, 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

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

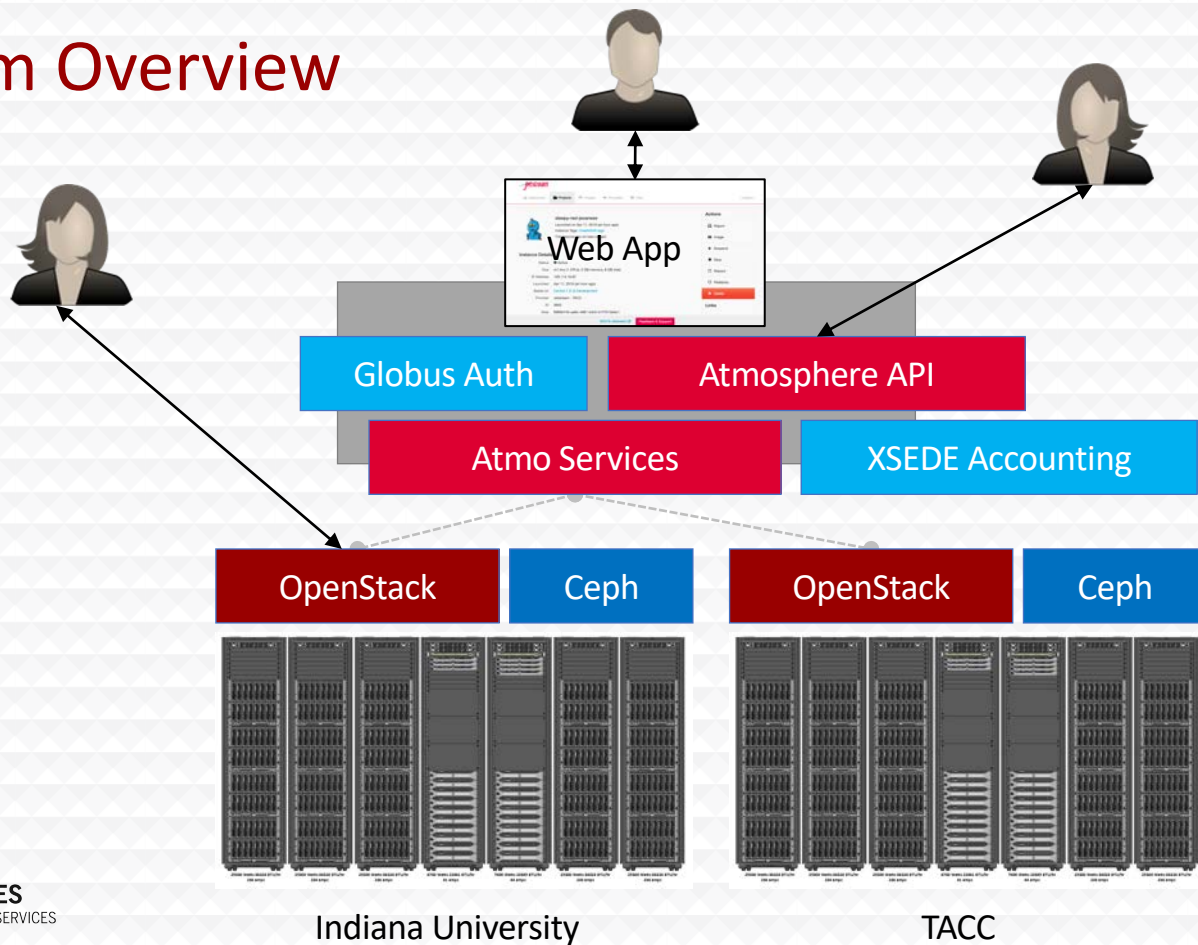




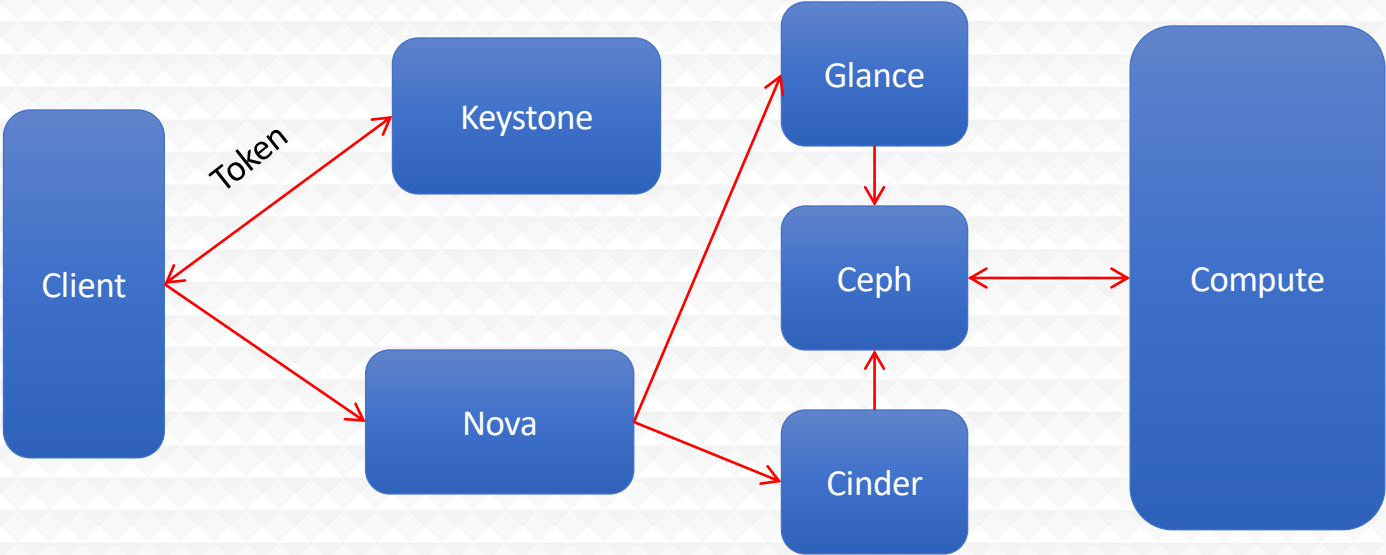
Pause...as we look at the Atmosphere Interface



# Platform Overview



# OpenStack Overview



# API Access to Jetstream

- What was **unexpected**
  - **Demand for programmable cyberinfrastructure**
  - Great platform for learning **system administration skills**
  - Great platform for **teaching & learning cloudy technologies**
- **Command line clients**
- **Horizon dashboard** very popular; but, incomplete
- **Programmatic control**; python is popular  
(<https://docs.openstack.org/openstacksdk/latest/>)



# Installing the client

- Simple on most Mac OS X and Linux hosts (a single pip command)
- Less simple, but still reasonably easy on Windows
  - Once you have a python installed, becomes a simple pip install
- Latest python-openstackclient (> 4.0.0) works with Python 3
- **Best practice – use a virtual environment like venv for your install**
- Docs on the wiki for this!
- Other CLI clients are available – e.g. python-swiftclient (Swift and S3), python-heatclient (Heat templates), etc
  - These are optional and not necessary for basic operations!



# Horizon GUI interface

- Allows most things you can do from the CLI
- Nice for some tasks
  - Network visualizer is something we tend to use as a troubleshooting tool
  - Easier to look at security groups on Horizon (IMHO)
- Downsides:
  - considerably slower than using CLI
  - not all features are present that are in CLI
  - can't do things programmatically



The screenshot shows the OpenStack Dashboard login page. At the top center is the OpenStack logo, which consists of a red 3D cube with a white square in the center, and the text "openstack" in a lowercase, sans-serif font below it. Underneath the logo is a light blue bar with the word "DASHBOARD" in small, uppercase letters. Below the logo is the heading "Log In" in a bold, black font. There are three input fields: "Domain" with the value "TACC", "User Name" (empty), and "Password" (empty). Each input field has a small icon on the right side. At the bottom right of the form is a blue button with the text "Connect" in white.



# Getting started with the API

Things you'll set up once (hopefully):

- SSH keys
- Security groups (though you'll build on the basics as you do more advanced things)
- Create a network
- Create a subnet
- Create a router

Things you'll likely do many times:

- Create and launch instances
- Screw up and delete instances
- Launch more instances
- Expand security groups

API CLI Tutorial walkthrough: <https://github.com/jlf599/JetstreamAPITutorial>

API Horizon walkthrough: <http://wiki.jetstream-cloud.org/Using+the+OpenStack+Horizon+GUI+Interface>



Wrapping up the API side...



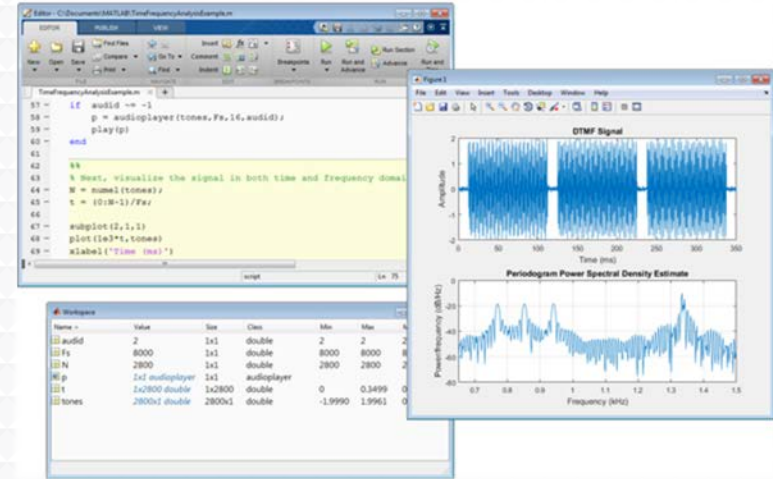


# Key Allocation Areas on Jetstream (Dec 2018 – Nov 2019)

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 all XSEDE-supported systems
Biological Sciences	75	7,267,083	54.15% Decrease	7.60%	4.66%
Computer Science / Advance Scientific Computing	141	17,034,146	22.28% Increase	17.82%	6.11%
Earth Sciences/ Geosciences	27	4,255,441	138.65% Increase	4.45%	1.37%
Electrical/ Mechanical/ Chemical Systems	13	7,516,478	560.71% Increase	3.85%	10.84%
Engineering	9	316,406	32.56% Decrease	0.33%	0.32%
Information, Robotics, and Intelligent Systems	15	3,265,334	994.80% Increase	3.42%	0.77%
Molecular Biosciences	82	15,561,599	58.61% Increase	16.28%	19.94%
Neuroscience and Behavioral Sciences	33	9,161,826	0.53% Increase	9.59%	2.98%
Ocean and Atmospheric Sciences	12	5,484,897	49.81% Increase	5.74%	5.31%
Physics	16	6,517,597	160.97% Increase	6.82%	11.11%
Social Sciences and Humanities	32	4,465,626	170.11% Increase	4.67%	1.16%
Training and Education	110	15,337,763	8.65% Increase	16.05%	4.29%

# Matlab for researchers and educators

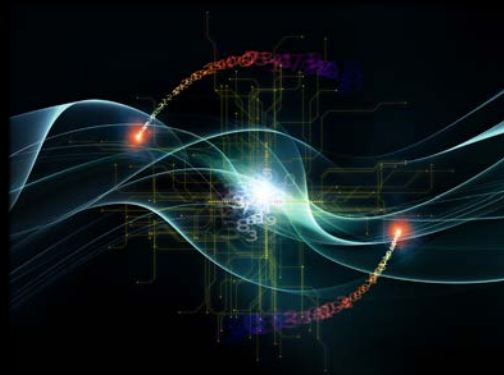
- Matlab and SimuLink and additional toolkits are installed on Jetstream
- You do NOT need to have a local license to use MATLAB on Jetstream
- If you are a researcher, and MATLAB or SimuLink... you're ready to go!
- If you are an engineering researcher, and you need other tools... let us know!



# 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

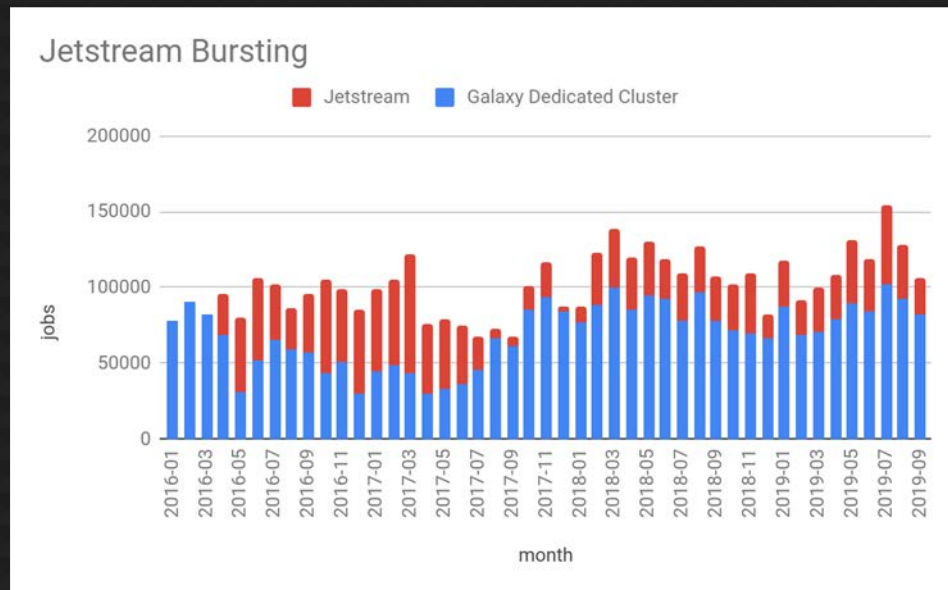
Research Data Alliance workshops, Galaxy workshops, data analysis in finance using R, security and intrusion detection, and principles in cloud computing...



# Update on Previous Key Project: Galaxy

## Jetstream usage from *usegalaxy.org*

- Bursting to Jetstream increased *usegalaxy.org*'s under-load capacity and throughput
- Jetstream well suited for near-immediate execution (no queue waiting)
- **27,902** distinct users have executed jobs on Jetstream via *usegalaxy.org* (as of January 14, 2020)



# Jetstream for education – in action at AMS2020

- Unidata-led workshop at American Meteorological Society (AMS) 2020 conference
- 127 users actively participating
- Participants used a JupyterHub running on Jetstream (40 node Kubernetes cluster of 6 core m1.medium VMs) for a 90 minute Unidata PyAOS (Python for the Atmospheric and Oceanic Sciences) workshop
- The students were successfully able to run their interactive Python code notebooks as the instructors presented their material



# Other top allocations on Jetstream

- **BrainLife (brainlife.io)** – Franco Pestilli, Indiana University
- **Cloudbased Cyberinfrastructure for Large Facility Operations and Natural Hazards Research**- Charles Meertens, UNAVCO
- **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



# Expanding the reach: Jetstream REU Program



NSF Supplement for undergraduates

- 4 students participated in 2017
- 6 students participated in 2018
- 7 students participating presently
- REU student videos on YouTube  
<https://www.youtube.com/user/IUPTI>





# Jetstream usage highlights – 1 Apr 2020

- 392 XSEDE projects covering 77 fields of science and over 2100 **active users** representing **203 institutions**
- **80%** of Jetstream users have **not used any other XSEDE system**
- >310M CPU hours allocated to XSEDE projects since June 2016
- 34 active science gateways
- 46 education/teaching allocations serving over 800 students
- 1189 mean active VMs in previous qtr, 1632 peak active VM count
- **Highest** user satisfaction in most recent XSEDE survey



# Jetstream Timeline...what comes next?

- Wrapping up our third year of operations with extension to November 2020 (and beyond, into 2021)
- 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 June 2020 – 15 July 2020
- Research allocation: Project desc ( $\leq 10$  pages) and Scaling doc ( $\leq 5$  pages)
  - **We can help!**

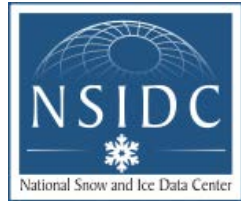
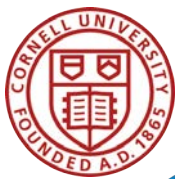


# Where can I get help?

- **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



# 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. May 21, 2020. Jetstream Overview – RMACC HPC Symposium. 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.

