RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES

rt.iu.edu

Jetstream Overview:
A national research and education cloud

Woods Hole Oceanographic Institution
June 20, 2019 – Woods Hole, MA.

Jeremy Fischer – Jeremy@iu.edu - Indiana University

Manager, Jetstream Cloud, UITS Research Technologies

Fischer, J. (2019). Jetstream Overview: A national research and education cloud. Woods Hole, MA. 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 2nd phase)
- Chameleon (renewed for 2nd 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

RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES

"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 in 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.
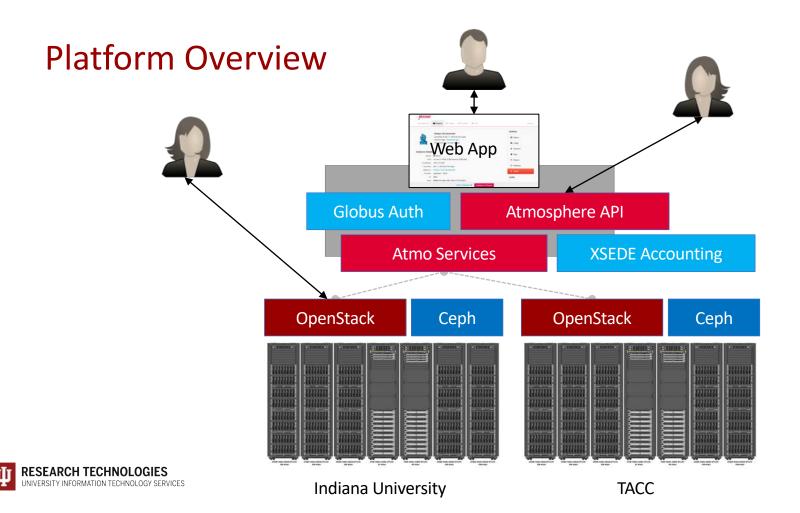
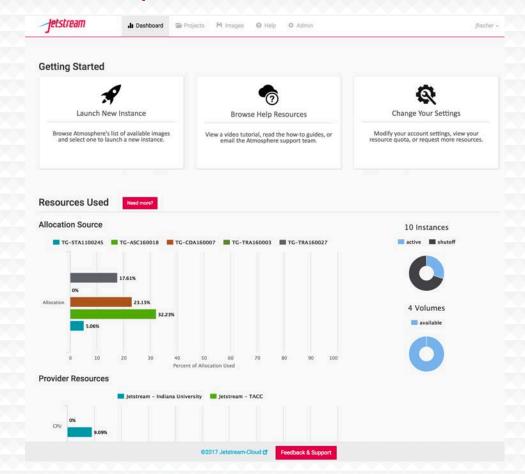-- Mike Lowe (Jetstream architect)

**some caveats for gateways…

Jetstream System Overview
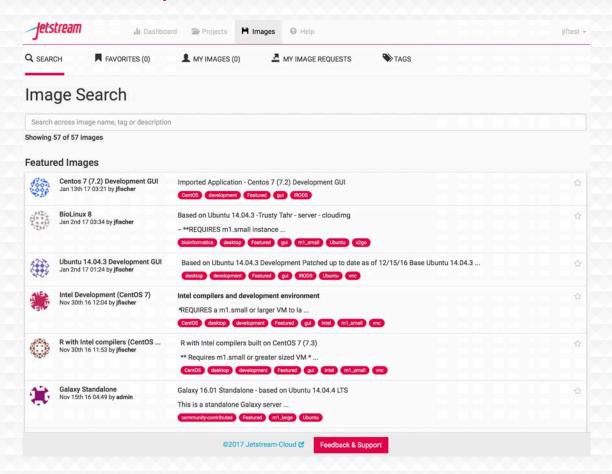
Platform Overview

Web App

Globus Auth

Atmosphere API

Atmo Services

XSEDE Accounting

OpenStack

Ceph

OpenStack

Ceph

Indiana University

TACC

RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES

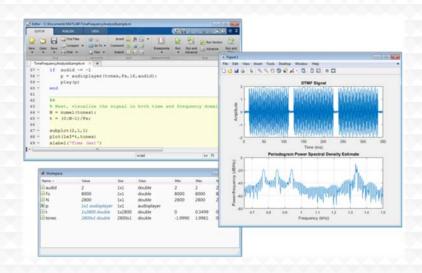# The Jetstream Atmosphere web interface

The Jetstream Atmosphere web interface

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	37.06% 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% Decrease	2.99%	8.65%
Social Sciences and Humanities	28	2,409,633	192.27% Increase	2.95%	0.81%

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!

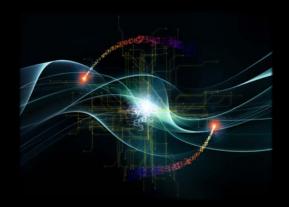- If you are an engineering researcher, and you need other tools… let us know!

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
- **GIS Gateway** (Coming Soon!) – Eric Shook, University of Minnesota

And others!

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

RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES

Jetstream usage highlights – 1 May 2019

- 418 XSEDE projects covering 71 fields of science and over 2375 **active users** representing **208 institutions**

- **80%** of Jetstream users have **not used any other XSEDE system**

- >200M CPU hours allocated to XSEDE projects since June 2016

- 25 active science gateways

- 58 education/teaching allocations serving over 1000 students

- 1438 mean active VMs in previous qtr, 1614 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 TODO

  - 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 2019 – 15 July 2019

- Research allocation: Project desc (<10 pages) and Scaling doc (<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

- 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/

RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES

Jetstream Partners

Questions?

- Project website: http://jetstream-cloud.org/
- Project email: help@jetstream-cloud.org Direct email: jeremy@iu.edu

RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES