Jetstream Overview:
A national research and education cloud

Economic Research in High Performance Computing Environments Workshop 2018 – October 9-10, 2018 – Kansas City, MO

Jeremy Fischer – 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

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 2nd phase)
- Chameleon (renewed for 2nd phase)

Now funding clouds to do research

- Bridges (Hybrid system)
- Jetstream

Jetstream
http://jetstream-cloud.org/

NSF 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

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

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

Jetstream
http://jetstream-cloud.org/

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
- Focus on ease-of-use, broad accessibility
- Provides on-demand *interactive* computing and analysis or persistent gateways
- Enables *configurable* environments and ***programmable cyberinfrastructure***

Jetstream
http://jetstream-cloud.org/

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

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

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…)

NSF funded by the National Science Foundation
Award #ACI-1445604

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
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

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…**

Jetstream
http://jetstream-cloud.org/

NSF 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.

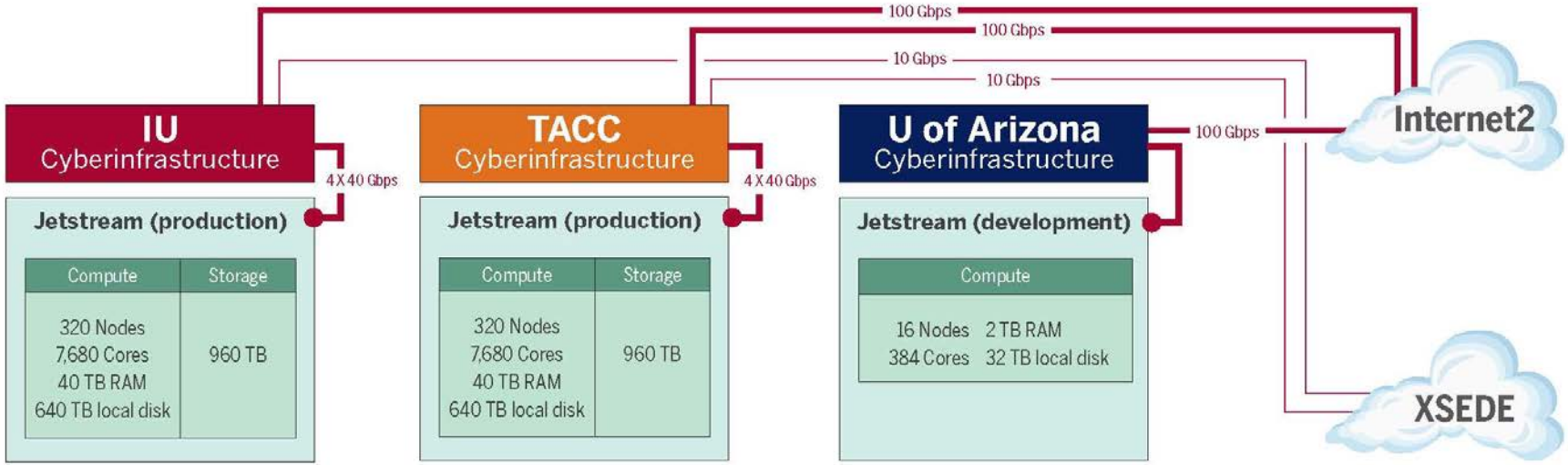
-- Mike Lowe (Jetstream architect)

**some caveats for gateways…

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

Jetstream System Overview



Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

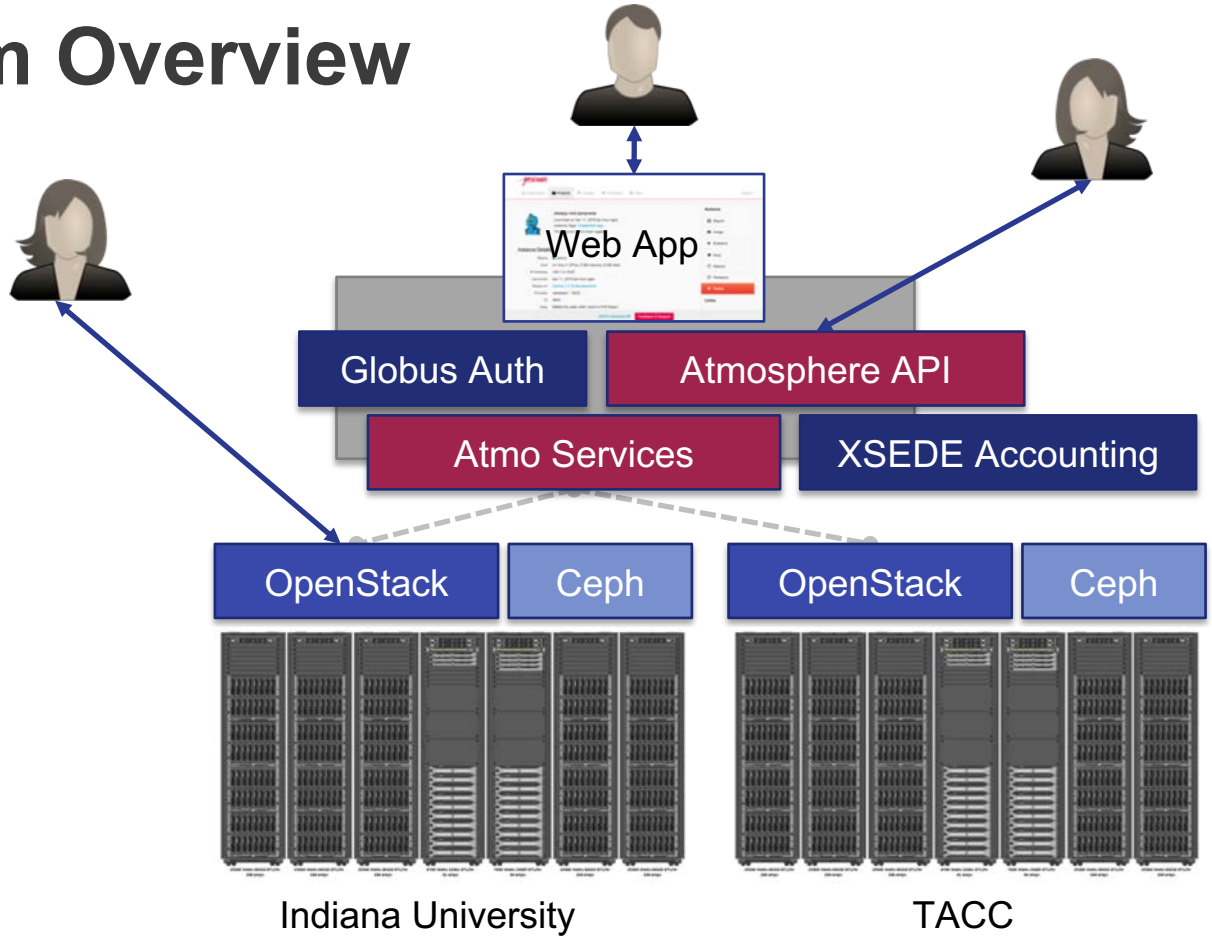
Production cloud hardware (per site)

Hardware	Number	Specifications	Function (IU)
Dell PowerEdge M630 blades	320	2x Intel E5-2680v3 "Haswell" 24 cores @ 2.5 GHz 128 GB RAM 2 TB local disk	Compute hosts OpenStack services
Dell PowerEdge R630 1U server	7	2x Intel E5-2680v3 "Haswell" 24 cores @ 2.5 GHz 128 GB RAM 2 TB local disk	Cluster management High Availability Databases RabbitMQ
Dell PowerEdge R730xd 2U servers	20	2x Intel E5-2680v3 "Haswell" 24 cores @ 2.5 GHz 64 GB RAM 48 TB storage for Ceph pool	~1 PB Ceph storage
Dell S6000-ON network switches	9	32+2 40 Gb/s ports	Top of Rack Spine

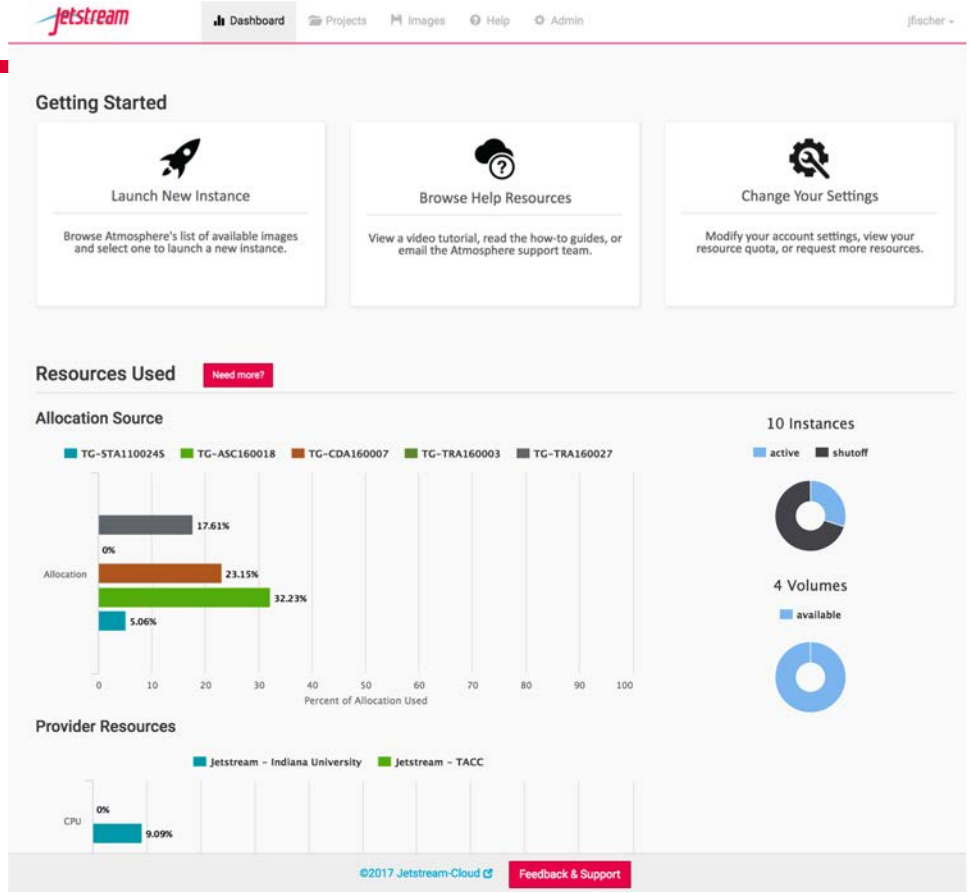
Jetstream
http://jetstream-cloud.org/

NSF
funded by the National Science Foundation
Award #ACI-1445604

Platform Overview

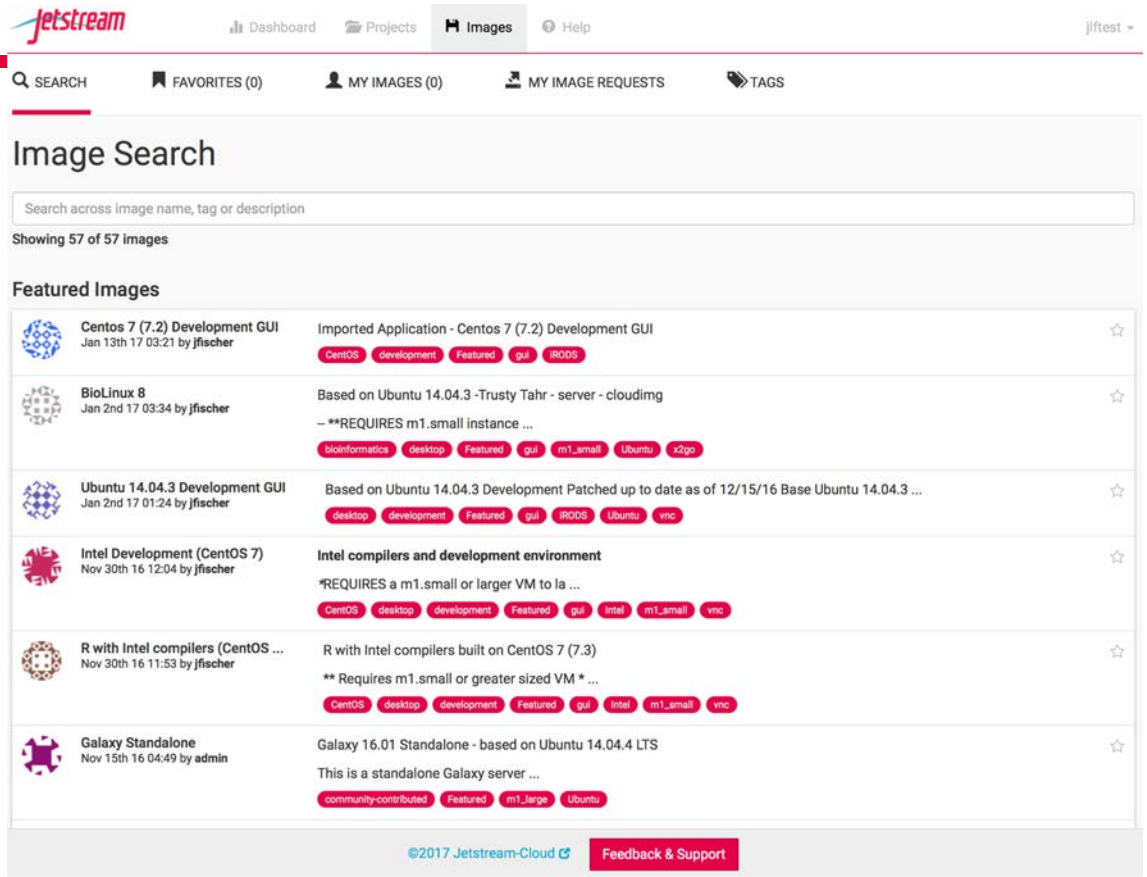


The Jetstream Atmosphere web interface



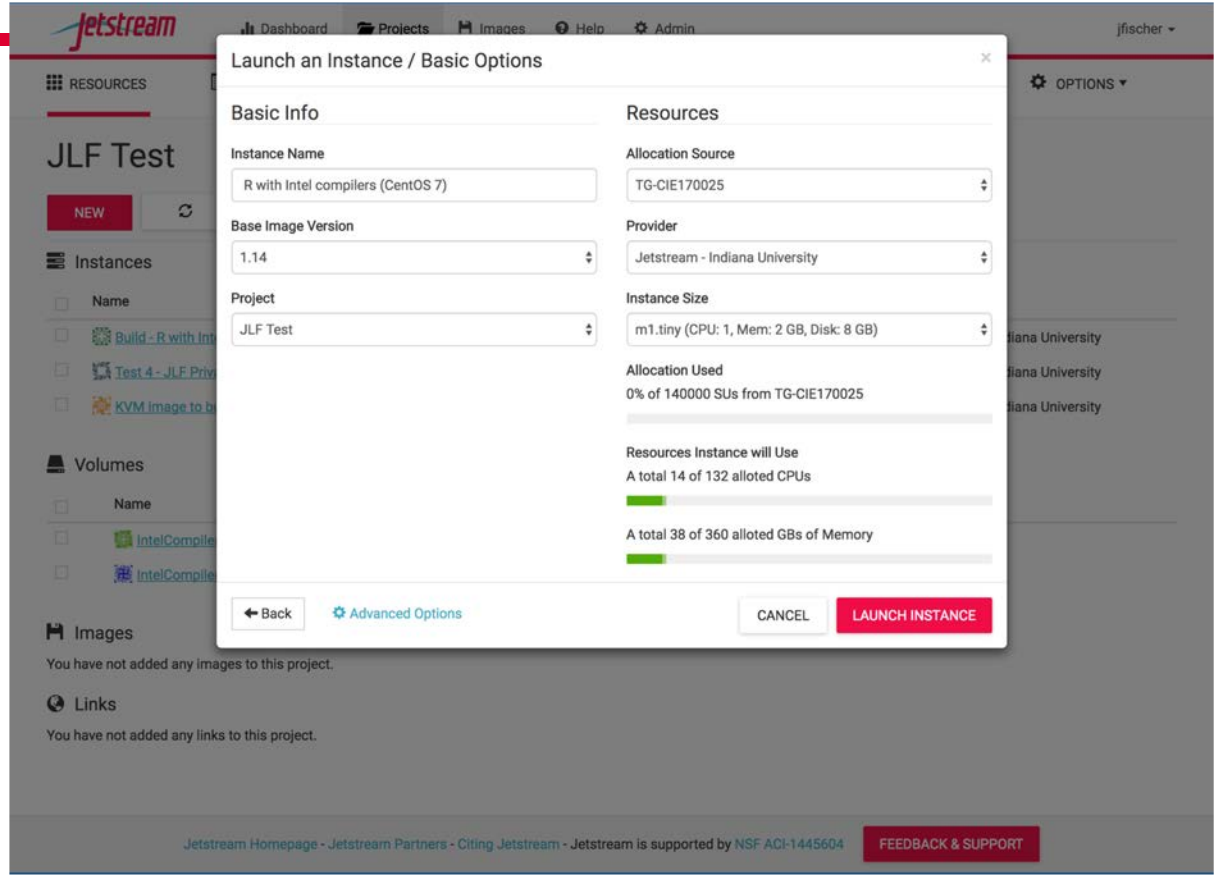
Jetstream
http://jetstream-cloud.org/

The Jetstream Atmosphere web interface



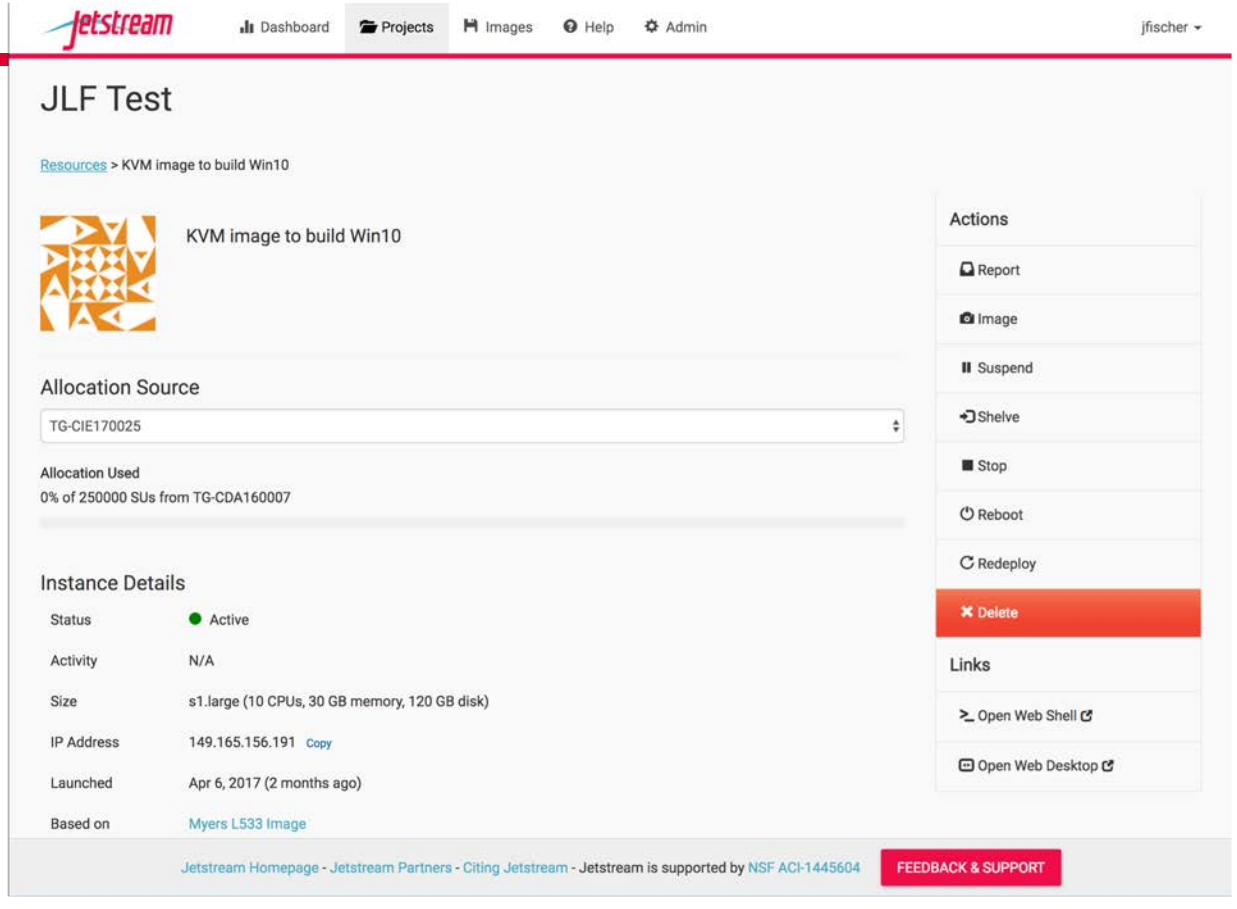
Jetstream
http://jetstream-cloud.org/

Look! It's more Jetstream web interface!



Jetstream
http://jetstream-cloud.org/

Even more Jetstream web interface…



Jetstream
http://jetstream-cloud.org/

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

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

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

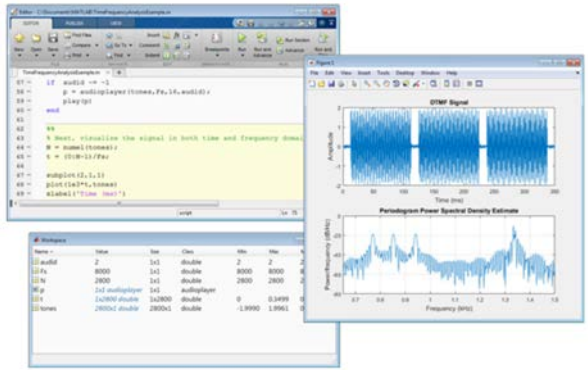
Primary methods of logging into Jetstream VMs to work

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

funded by the National Science Foundation
Award #ACI-1445604

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%
Atmospheric Sciences	4	2,752,400	7.55%	3.73%
Biological Sciences	57	5,199,000	14.27%	4.95%
Campus/Domain Champions	123	6,105,500	16.76%	0.09%
Computational Science	11	1,150,000	3.16%	0.92%
Computer Science	15	4,944,302	13.57%	1.8%
Education Allocations	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%
Training & Development	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

Jetstream
http://jetstream-cloud.org/

NSF 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

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!

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

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
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

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

NSF funded by the National Science Foundation
Award #ACI-1445604

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
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

*M&O PY2 to date 11/12 months

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

NSF funded by the National Science Foundation
Award #ACI-1445604

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!

NSF funded by the National Science Foundation
Award #ACI-1445604

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

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

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/

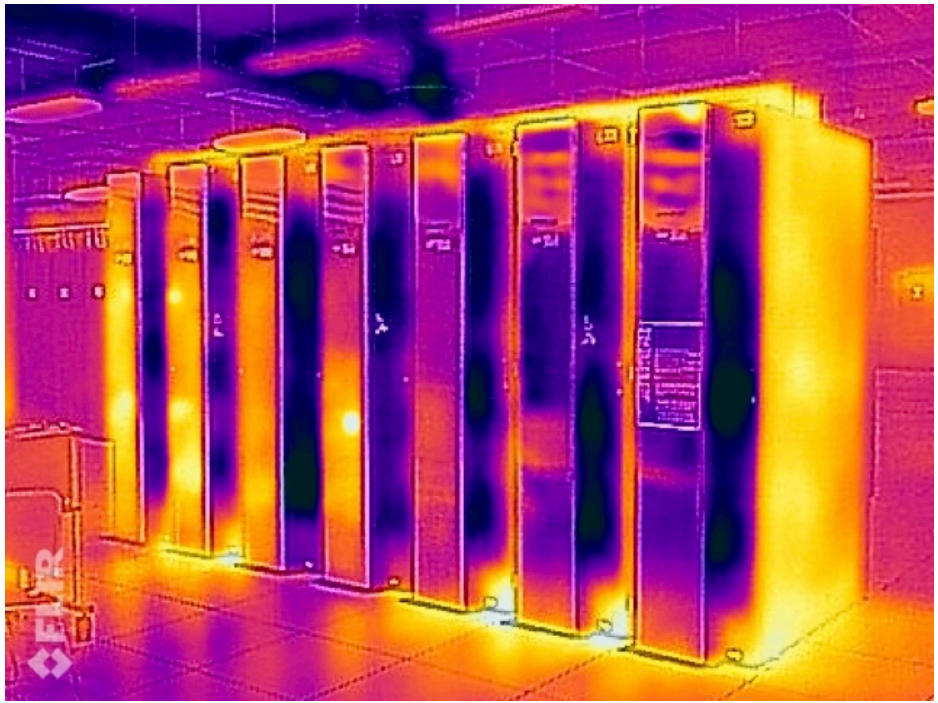
NSF funded by the National Science Foundation
Award #ACI-1445604

Jetstream Fun: Happy cluster / Angry Cluster

Jetstream
http://jetstream-cloud.org/

NSF
funded by the National Science Foundation
Award #ACI-1445604

Infrared image of Jetstream



ChilledDoor Rack Cooling System by motivair

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

Jetstream Partners

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

Questions?

Project website: http://jetstream-cloud.org/

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

License Terms

funded by the National Science Foundation
Award #ACI-1445604