Jetstream Overview
A national research and education cloud

ESIP Tech Dive Webinar – April 12, 2018

Jeremy Fischer – jeremy@iu.edu
Senior Technical Advisors,
UITS Research Technologies

Fischer, J. (2018). 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 (under re-compete)
- Stampede – 9.6 petaflops, 2013 (extended to Stampede2, going into service now)
- Comet – ~2.0 petaflops, 2014

Has funded research into building clouds and computer science

- CloudLab
- Chameleon (renewed for 2nd phase)

Now funding clouds to do research

- Bridges (Hybrid system)
- Jetstream

Jetstream
http://jetstream-cloud.org/

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. Ever. But they need easy access and use.

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

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
- User-selectable library of preconfigured virtual machines
- Provides on-demand *interactive* computing and analysis or persistent gateways
- Enables *configurable* environments and ***programmable cyberinfrastructure***
- Reproducibility: Share VMs and then store, publish via IU Scholarworks (DOI)

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…)

Jetstream
http://jetstream-cloud.org/

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.

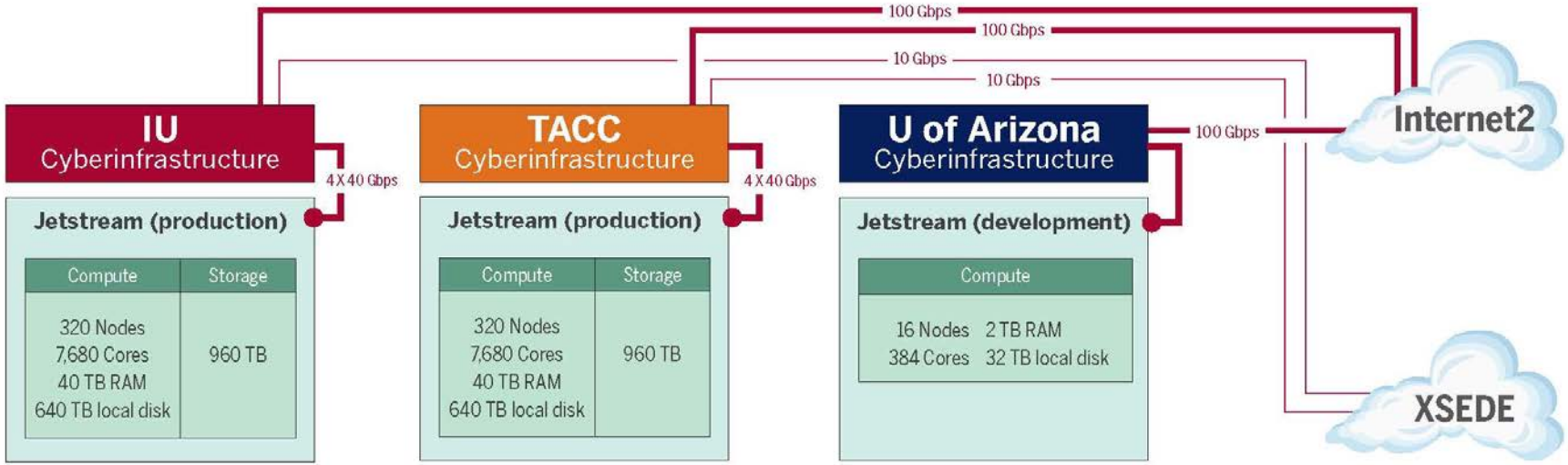
-- Mike Lowe (Jetstream architect)

**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

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

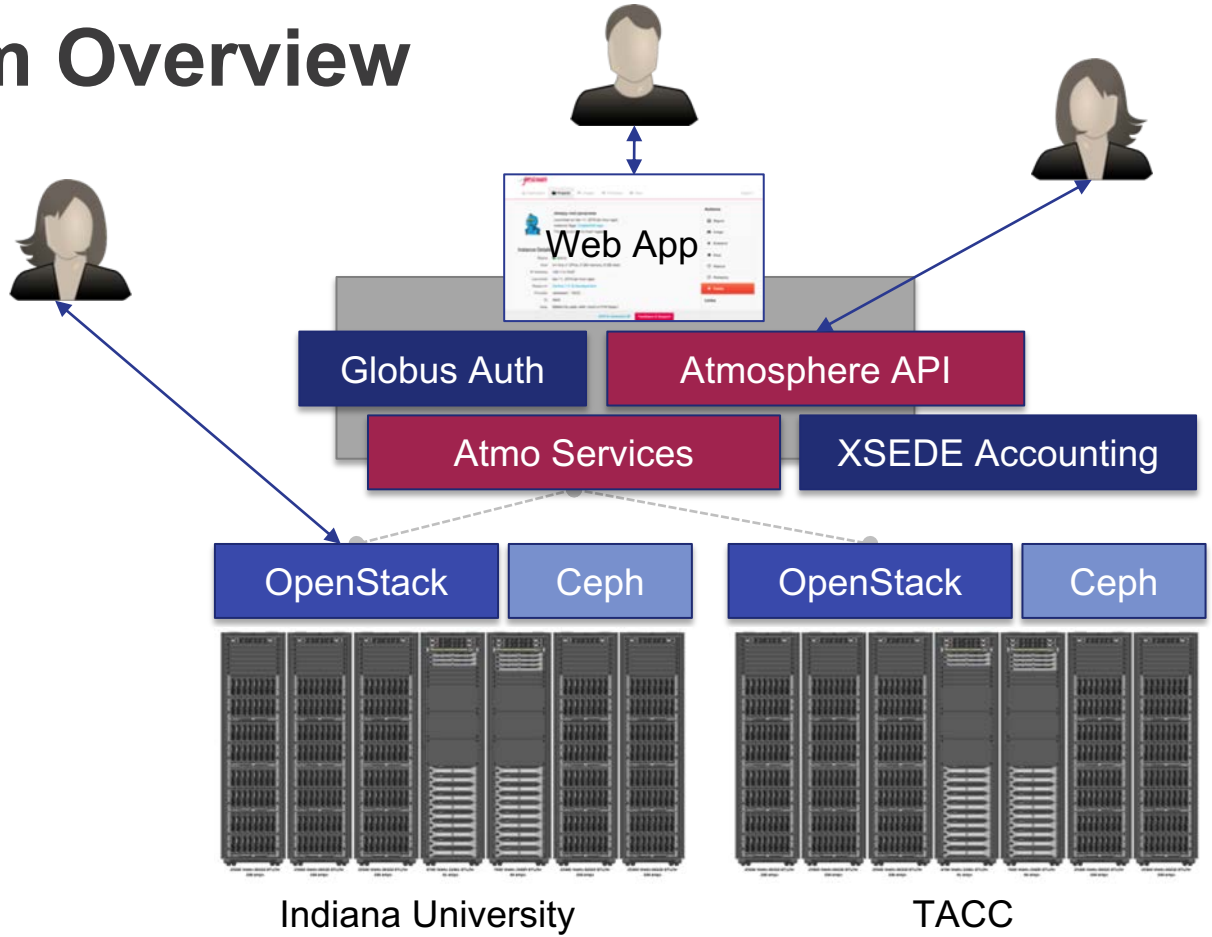
** s1.* storage-rich instances are not eligible to be saved into a customized image

- 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

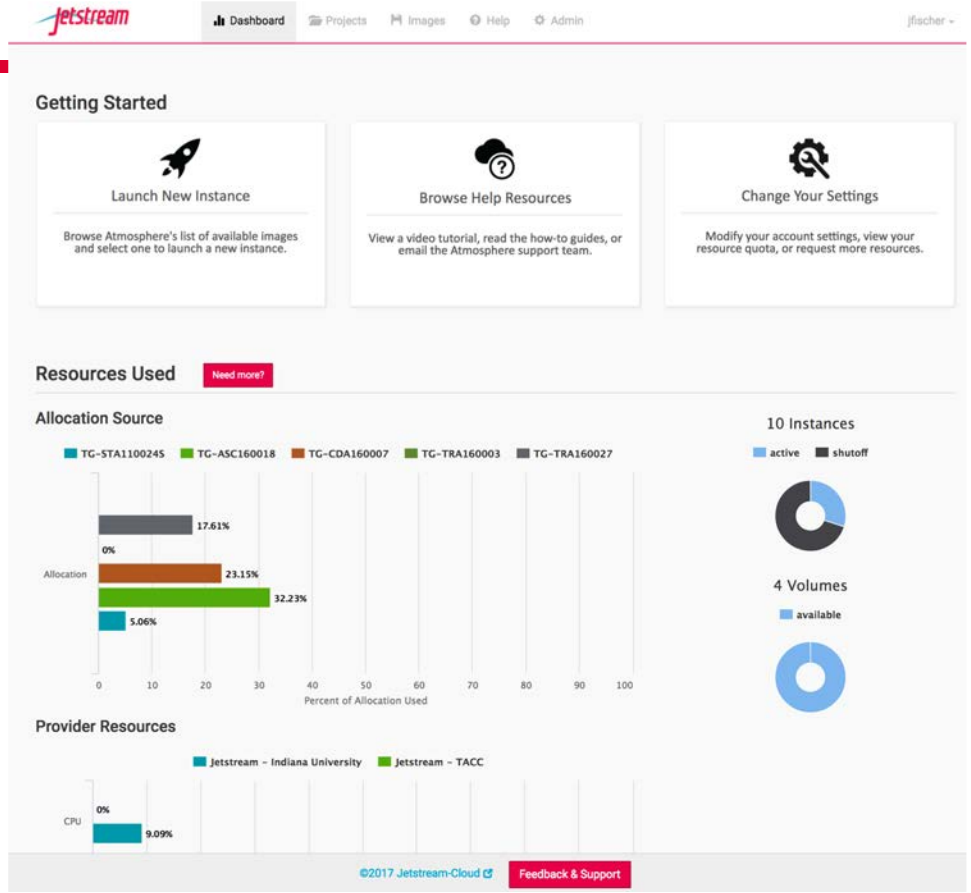
Jetstream
http://jetstream-cloud.org/

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

Platform Overview

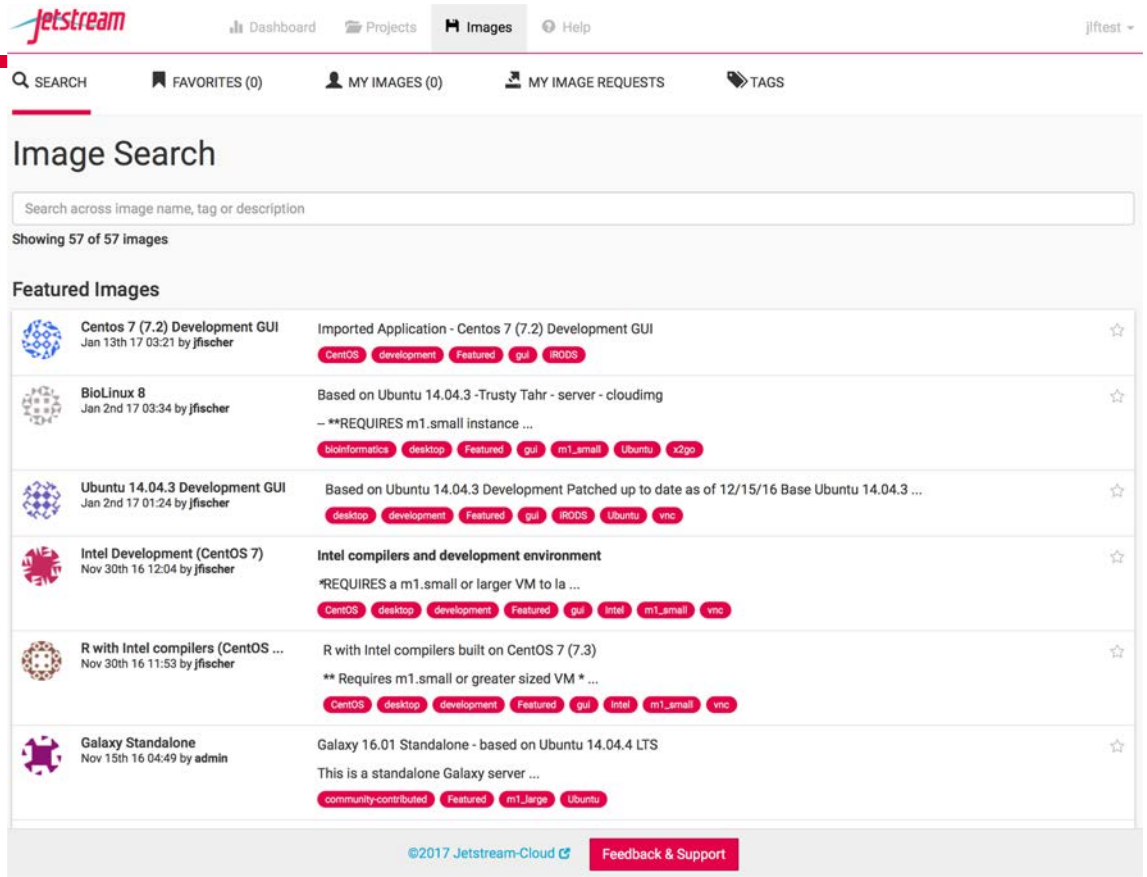


The Jetstream Atmosphere web interface



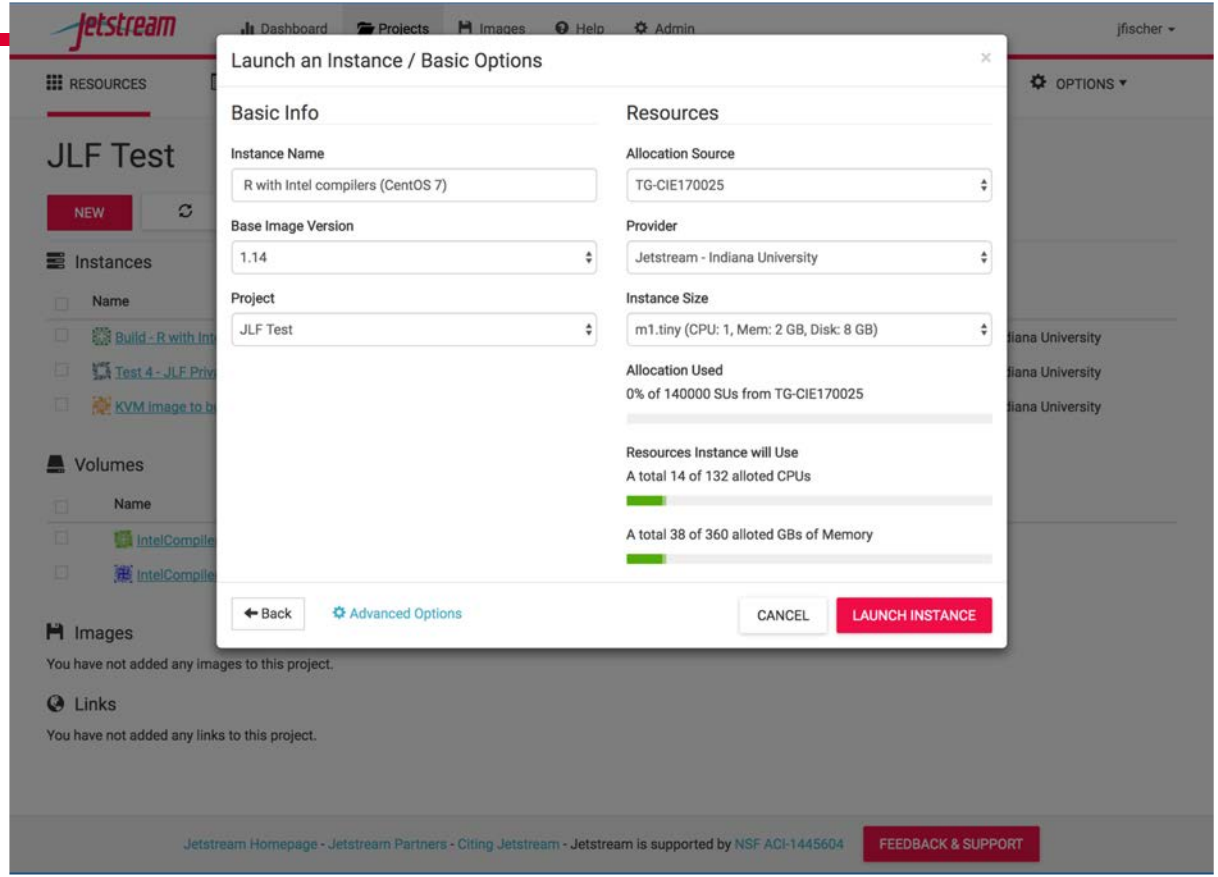
Jetstream
http://jetstream-cloud.org/

The Jetstream Atmosphere web interface



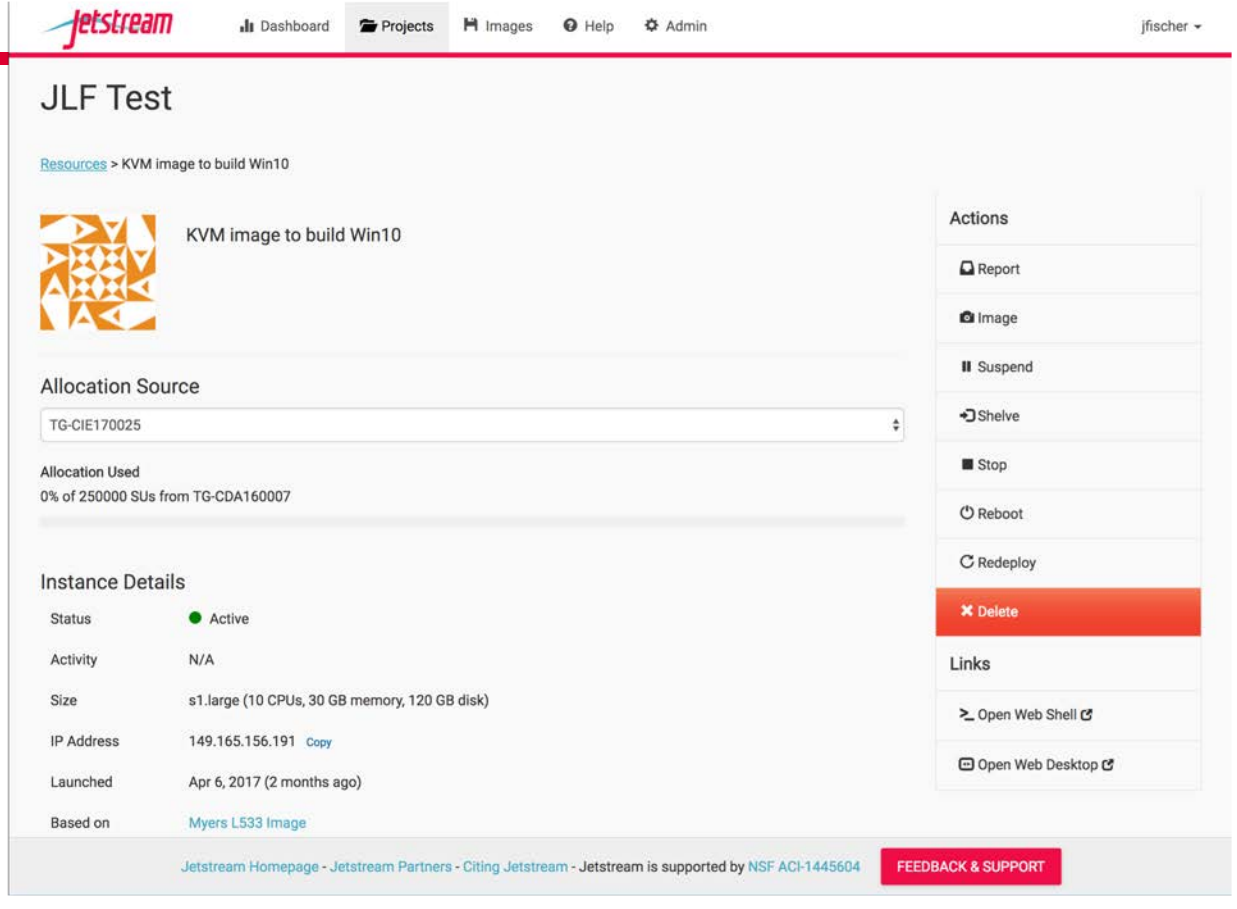
Jetstream
http://jetstream-cloud.org/

Look! It's more Jetstream web interface!



Jetstream
http://jetstream-cloud.org/

Even more Jetstream web interface…



Jetstream
http://jetstream-cloud.org/

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

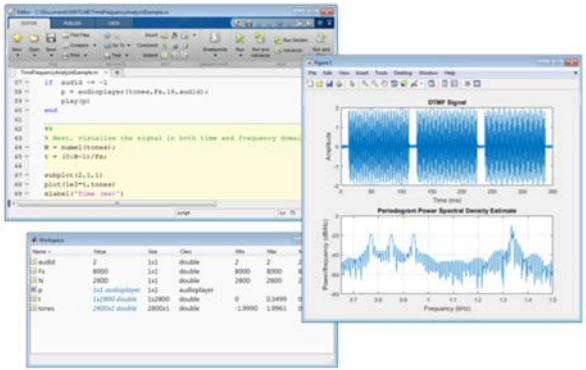
- Interactive user access via web interface with VNC/SSH
- Direct VNC/SSH to individual instances

Jetstream
http://jetstream-cloud.org/

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 52 standard 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 and one of the standard first 52 toolkits that come with MATLAB help you… 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/

funded by the National Science Foundation
Award #ACI-1445604

Jetstream for Education

Jetstream has been used in multiple IU Informatics Graduate Courses

- INFO 535 – Management, Access, and Use of Big and Complex Data
- INFO 590 – Topics in Informatics

BlueWaters Workflow Workshop

Multiple Research Data Alliance Workshops

Upcoming workshops/classes on Galaxy, data analysis in finance using R, security and intrusion detection, and principles in cloud computing

Jetstream
http://jetstream-cloud.org/

NSF 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

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
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: CIPRES Gateway, The Neuroscience Gateway, UltraScan III

Jetstream
http://jetstream-cloud.org/

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

The basics that Jetstream provides from the programmable cyberinfrastructure (API) perspective

- Base images to start from
- A basic network space (some DIY required)
- An IP pool
- Isolated project space for your team (on two clouds)
- A blank canvas (to some degree) – you can truly install just about anything you want – and you can BYOLicense if needed

Jetstream
http://jetstream-cloud.org/

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

Left Twix, Right Twix

- Having two clouds = some semblance of fault tolerance
- Making that work for your gateway or infrastructure service
    - Fail over possibilities
    - Load distribution/performance
    - Maximizing simultaneous VMs (VM/IP limits per cloud)

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

Jetstream storage

Storage built into the VM flavors = ephemeral (replicated, but…)

Volume storage = persistent (erasure coded, 4 data 2 recovery)

Valuable data should be on volumes (and backed up elsewhere)

Object storage available – still in early phases of availability

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

Some of the advanced possibilities on Jetstream…

- True elastic computing
  - OpenStack Heat
  - OpenStack Magnum
  - Your own creation?
- Virtual clusters
  - Several gateways using virtual clusters
  - Working on a bigger and better long-term solution
  - Workshop at PEARC18 – more info soon!
- Other possibilities
  - https://www.openstack.org/software/project-navigator/
  - Mistral (OSG) – cron as a service
  - Senlin (a coming attraction for making virtual clusters easier)
  - other additions like Manila (filesystems as a service), etc

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

But what about big data sets and such?

- At IU, Wrangler is nearby
- Jetstream wired to Wrangler's switch and vice versa
- Dedicated NFS node(s) on Wrangler for Jetstream
- Dedicated vlan built for projects that need access (still in beta, but working!)
- TACC options will be different – primarily S3/Swift access to Wrangler TACC

Jetstream
http://jetstream-cloud.org/

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

Jetstream usage highlights

As of April 1, 2018:

- 313 active XSEDE projects covering 69 fields of science and **2344 active users** representing **200 institutions**
- **86%** of Jetstream users new to XSEDE (at end of PY1)
- >108 million CPU hours allocated to XSEDE projects since June 2016
- 12 active science gateways
- 42 education/teaching allocations serving almost 700 students
- Averaging ~1150 concurrent Active VMs in March 2018
- **100%** system availability**, 99.4%** cap availability
- **97.7%** “job” completion (at end PY1)

Jetstream
http://jetstream-cloud.org/

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

Jetstream Timeline…what comes next?

- Completed our first year of operations on September 1, 2017
- 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

Jetstream
http://jetstream-cloud.org/

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

Requesting access to Jetstream

- Trial allocations available TODAY
- You can request startup allocations anytime. (Startups are simple!)
- You can request allocations for educational use anytime.
- Next submission period for large allocations is NOW - 15 March 2017– 15 April 2017.
- 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

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/

Jetstream
http://jetstream-cloud.org/

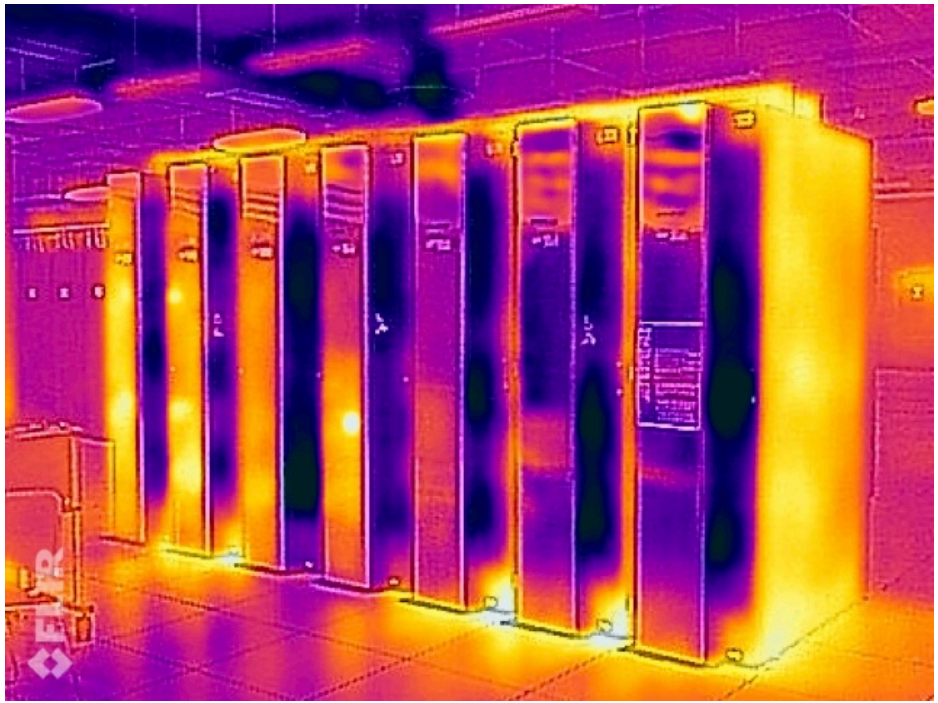
NSF
funded by the National Science Foundation
Award #ACI-1445604

Jetstream Fun: Happy cluster / Angry Cluster

Jetstream
http://jetstream-cloud.org/

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/

NSF
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 / ssudarsh@iu.edu

License Terms

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604