SIAM PP18 – March 7, 2018

Increasing User and Application Diversity through Programmable Cyberinfrastructure

David Y. Hancock – dyh@iu.edu

Robert Henschel – henschel@iu.edu

Principal Investigator – Jetstream

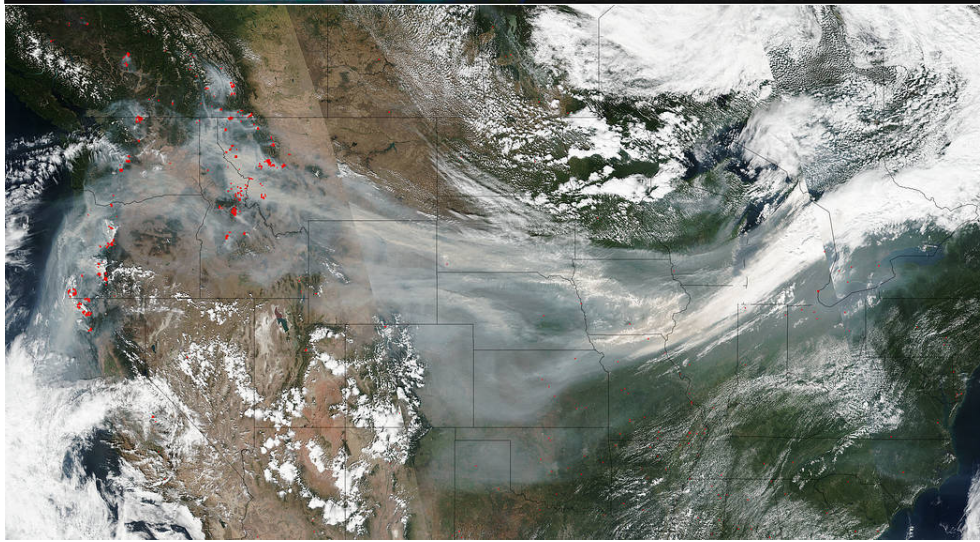
Program Director – Advanced Cyberinfrastructure

Director – Science, Community, Tools

Indiana University Pervasive Technology Institute

What is “the” Jetstream?

- Fast moving air currents
- Hot/Cold air boundaries
- Cloud resource funded by National Science Foundation



NASA's Suomi NPP satellite collected this natural-color image using the VIIRS (Visible Infrared Imaging Radiometer Suite) instrument on Sept. 4, 2017. Actively burning areas are outlined in red. NASA image courtesy Jeff Schmaltz LANCE/EOSDIS MODIS Rapid Response Team, GSFC

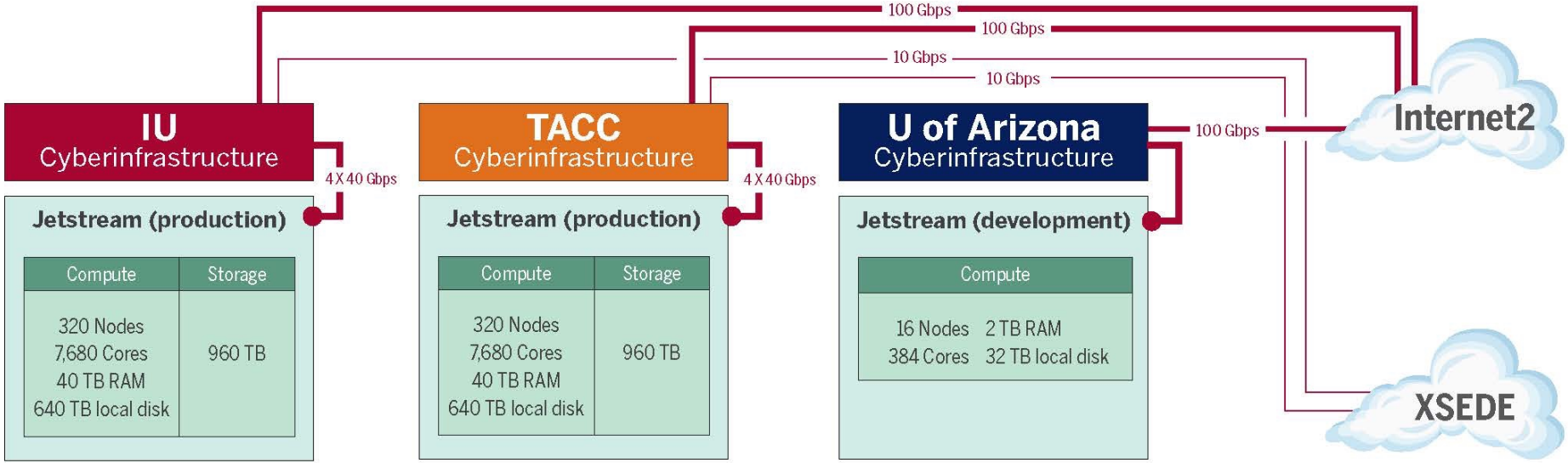
What is Jetstream and why does it exist?

- NSF's first production cloud facility
- Focus on ease-of-use, broad accessibility
- Encourage collaboration and community development
- User-selectable library of preconfigured virtual machines
- Provides on-demand *interactive* computing and analysis or persistent services such as gateways (e.g. SEAGrid, Galaxy, GenApp, and others)
- 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

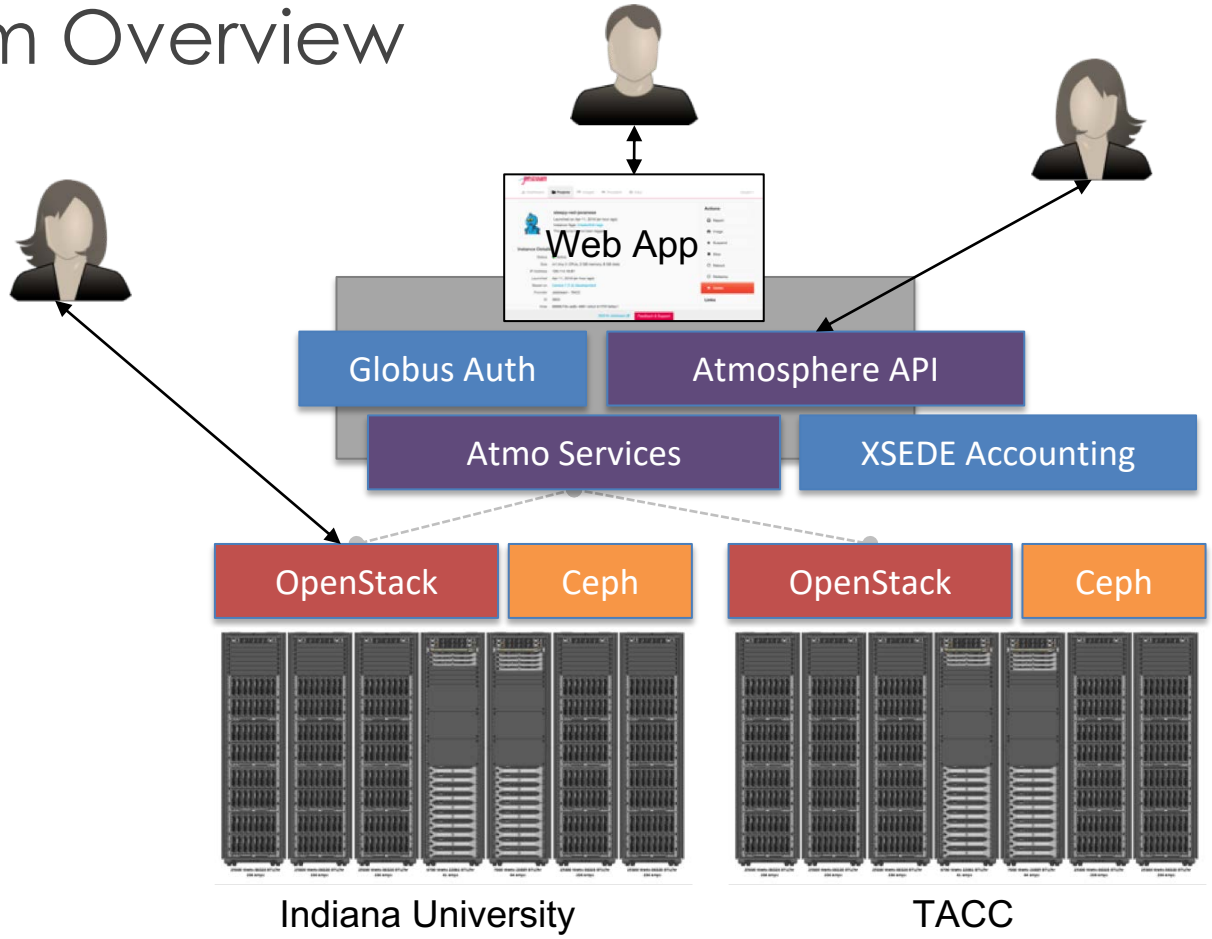
Jetstream System Overview



Jetstream
http://jetstream-cloud.org/

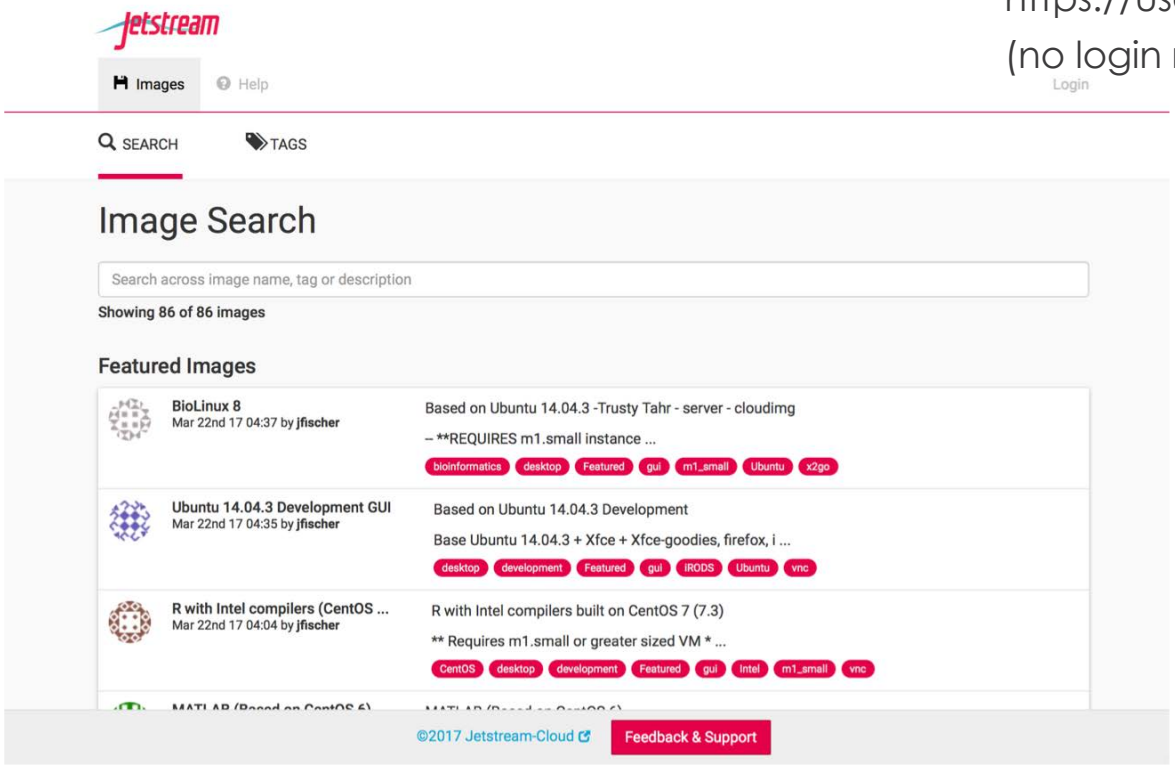
NSF
funded by the National Science Foundation
Award #ACI-1445604

Platform Overview



Jetstream's Atmosphere interface

https://use.jetstream-cloud.org/
(no login required to this point)

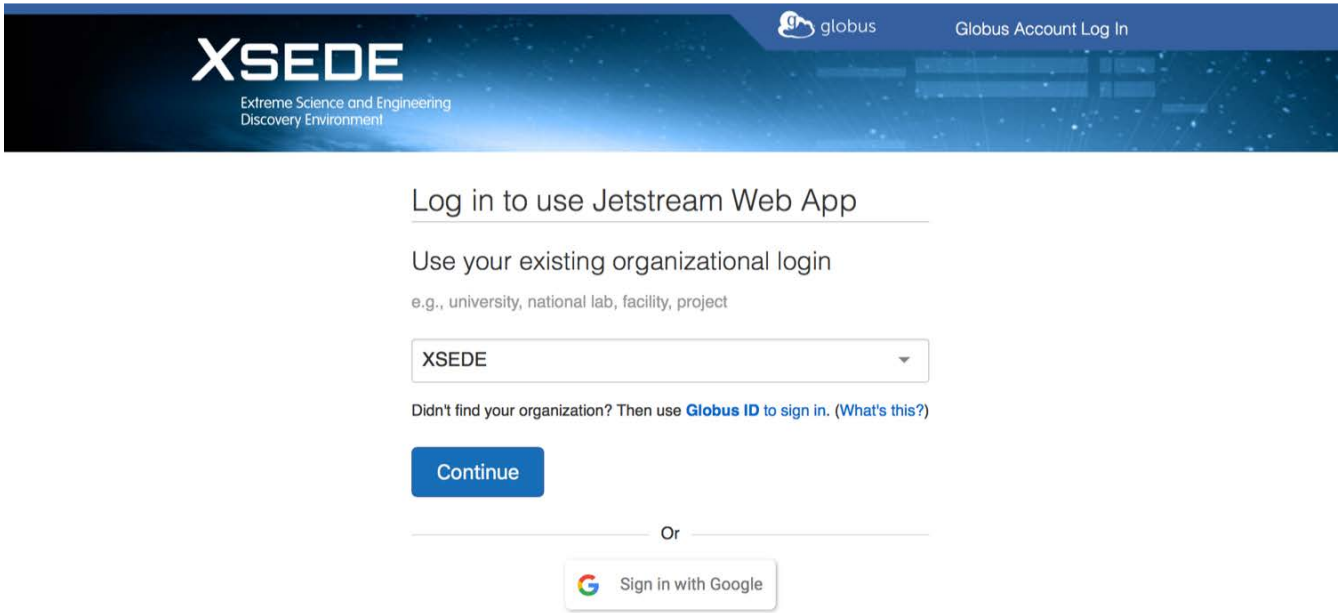


http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

Jetstream's Atmosphere interface

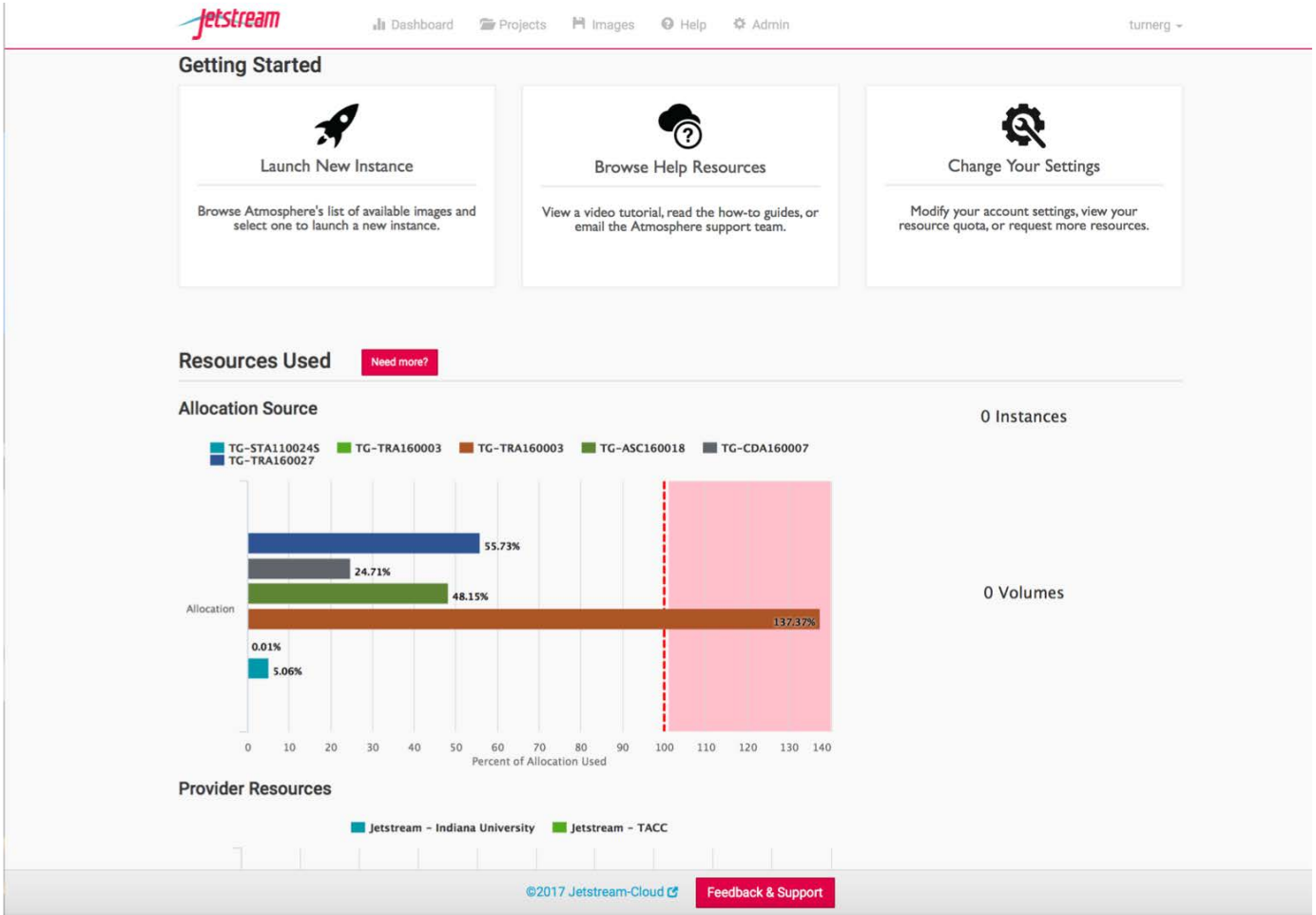
(Select identity provider)



Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

Jetstream's Atmosphere interface



(user's home space)

Jetstream usage highlights – March 2018

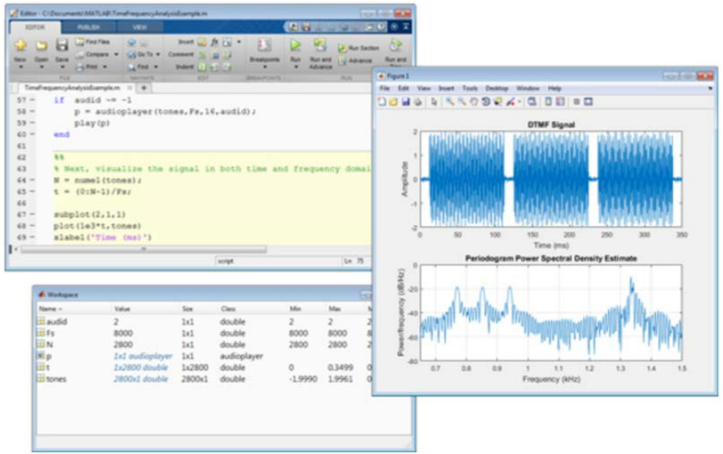
- 314 active XSEDE projects covering 72 fields of science and **2039 active users** representing **201 institutions**
- **86%** of Jetstream users new to XSEDE (at end of PY1 Ops)
- >104 million CPU hours allocated to XSEDE projects since June 2016
- 12 active science gateways
- 43 education/teaching allocations serving almost 700 students
- Averaging 1230 concurrent Active VMs in February 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

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/

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	3	978,406	1.99%	7.74%
Atmospheric Sciences	3	2,148,503	4.37%	4.08%
Biological Sciences	73	5,981,574	12.15%	2.15%
Chemistry	7	1,651,065	3.35%	6.58%
Computational Science	19	5,402,400	10.98%	0.56%
Computer Science	39	5,060,402	10.28%	2.14%
Education Allocations	45	3,261,971	6.63%	0.83%
Engineering	6	209,999	0.43%	2.78%
Geosciences	10	1,492,905	3.03%	0.52%
Humanities/Social Sciences	6	527,293	1.07%	4.13%
Molecular Biosciences	6	10,831,625	22.01%	0.59%
Network Science	4	488,261	0.99%	0.15%
Neuroscience	10	691,547	1.41%	1.29%
Physics	8	2,287,182	4.65%	9.58%
Training & Development	106	8,206,899	16.67%	2.72%

Jetstream for Education – SIGUCCS17 Paper

- Research is a primary function
- Training current and future researchers is crucial
  - Consistent environments
  - Customizable environments
  - Easy to access (Location independent, Common tools)
  - Easy to use (Familiar environments)
- Formal survey conducted in 2017

Jetstream
http://jetstream-cloud.org/

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

Key Points of Success

- **100%** of Jetstream education PIs surveyed reported their use of Jetstream to be a **success** in their opinion
- "Jetstream went more smoothly than [commercial provider] in almost every way and seems to perfectly meet our need for training!"
- "fast and flexible access to computational resources at no cost"
- Beneficial to researchers to use in workshops because they could apply for their own allocations afterward and **continue using the system that they learned on**
- "giving students and instructors the ability to modify, install software on their own and maintain access on demand later on to those images is something of great value, which would not have been easily achieved on our cluster."

Jetstream
http://jetstream-cloud.org/

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

Where can Jetstream improve?

- Constructive criticism is necessary!
- Better monitoring of usage by students (workarounds now, but better options coming)
- Improve training materials
- Add videos to training materials
- Persistent IPs in Atmosphere GUI environment
- Improved base images for areas of domain science and instruction

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

Engaging educators

- Common message of "This is great! More people should be using it! Get the word out!" (but no ideas on how to best do that)
- Engage and train IT support staff
- Presenting at conferences (PAG, SC, PEARC, SIGUCCS, targeting domain science conferences)
- Educause and regionals
- Site visits for tutorials
- Other suggestions certainly welcome!

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

About those gateways…

- 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
- ChemCompute – A gateway for computational chemistry instruction
- Coming gateways: CIPRES Gateway, The Neuroscience Gateway, UltraScan III

Jetstream
http://jetstream-cloud.org/

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

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

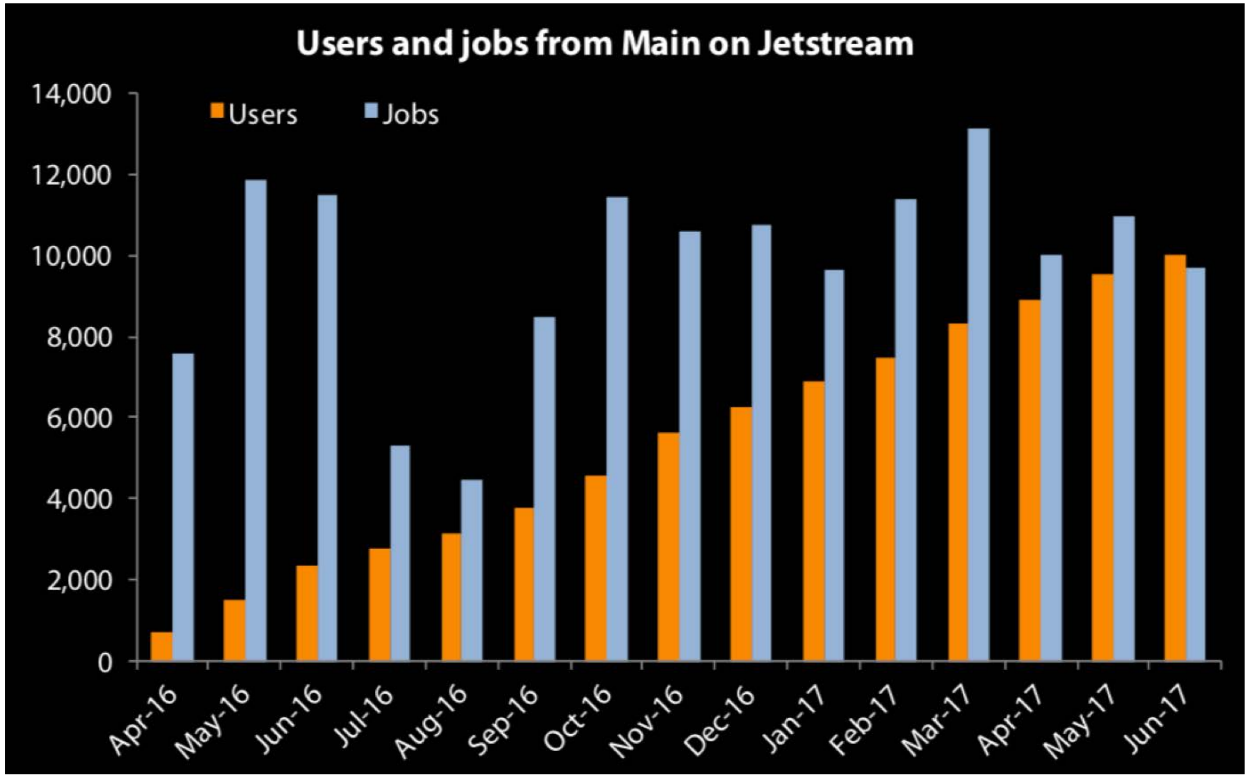
Use Jetstream as a ***self-service*** platform

- 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 Galaxy bursting: ~10K unique users,115K jobs



< 5% of the Jobs

15–20% of the CPU time

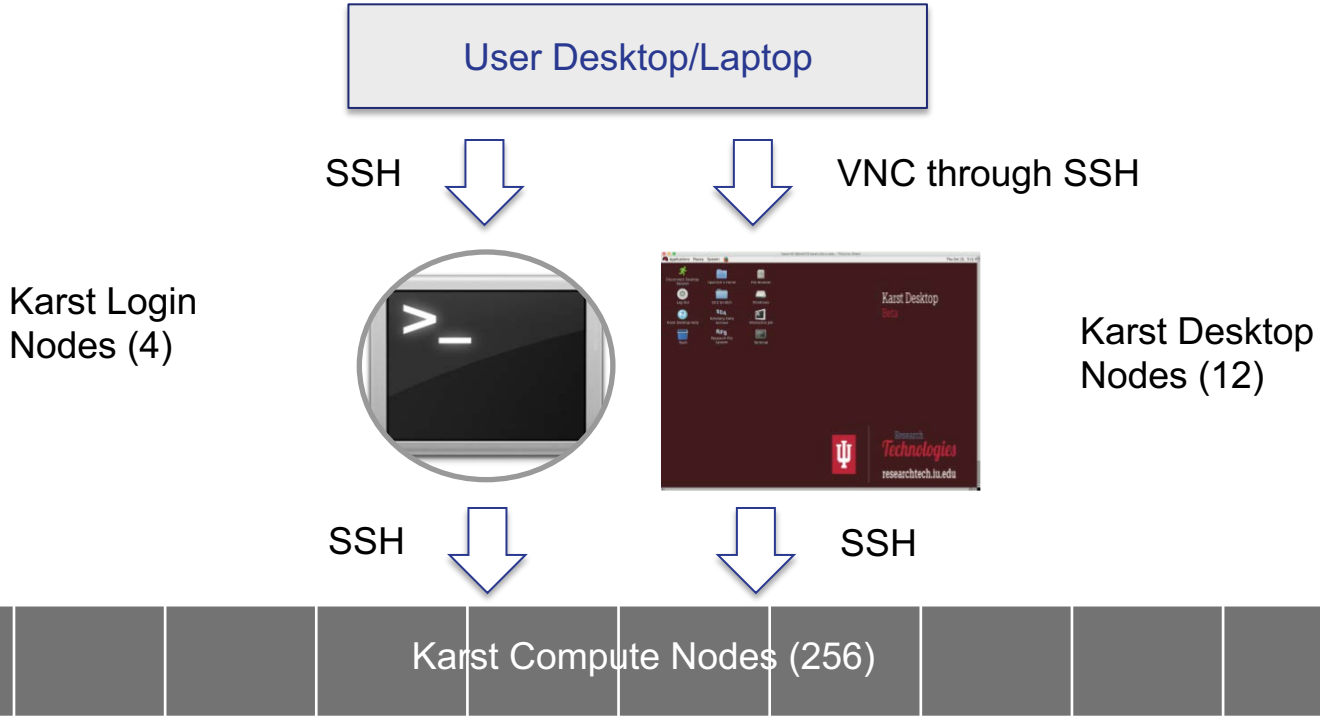
Disk throughput is important

Now leveraging containers and Jupyter Notebooks

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

Local View – Karst Desktop Architecture



Karst Desktop – Use Cases

- Running interactive mathematical and statistical applications
- GUIs of performance analysis applications such as Vampir, Allinea MAP, TotalView
- Visualization
- COMSOL Multiphysics Client/Server
- Data Enclave
- Desktop environment for crystallography tool suite
- Easy access to compilers for classes
- Long running data movement jobs
- Facilitates collaboration

Jetstream REU Program

- Recruitment for for 2018 has started (≤6 slots)
- March 31 Deadline
- Students can apply: reu@jetstream-cloud.org
- https://jetstream-cloud.org/forms/reu-application.php
- REU student videos on YouTube https://www.youtube.com/user/IUPTI
- News release describing their experiences

https://itconnections.iu.edu/2017-august/jetstream.php

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

Requesting access to Jetstream

- Trial allocations available **TODAY**
- You can request startup allocations anytime.
- You can request allocations for educational use anytime.
- Next submission period for large allocations is 15 March 2018 – 15 April 2018.
- We are happy to help you prepare a request and create a successful proposal.
- You do **not** have to have prior use of Jetstream to be successful.
- You **do** need a US-based collaborator

Jetstream
http://jetstream-cloud.org/

NSF 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

Help / References

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

Paper describing Jetstream: Jetstream: A self-provisioned, scalable scince and engineering cloud environment

Configuration management: https://github.com/jetstream-cloud/Jetstream-Salt-States

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: dyh@iu.edu

License Terms

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604