

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

Indiana University Pervasive Technology Institute

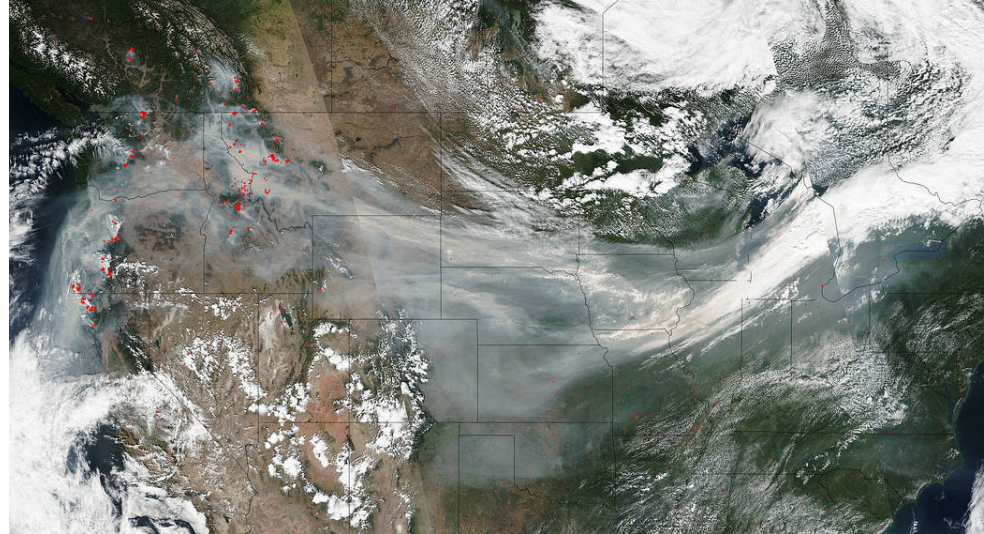
Director – Science, Community, Tools



What is “the” Jetstream?

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

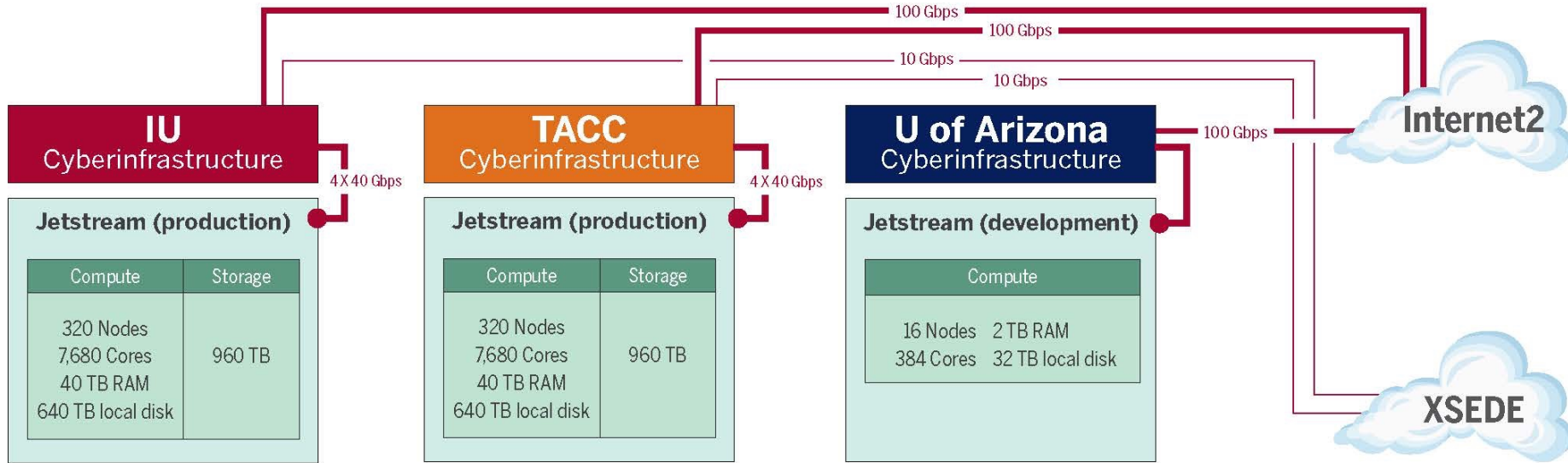
NASA's Suomi NPP satellite collected this natural-color image using the VIIRS (Visible Infrared Imaging Radiometer Suite) instrument on Sept. 4, 2017. Actively burning areas are outlined in red. NASA image courtesy Jeff Schmaltz LANCE/EOSDIS MODIS Rapid Response Team, GSFC



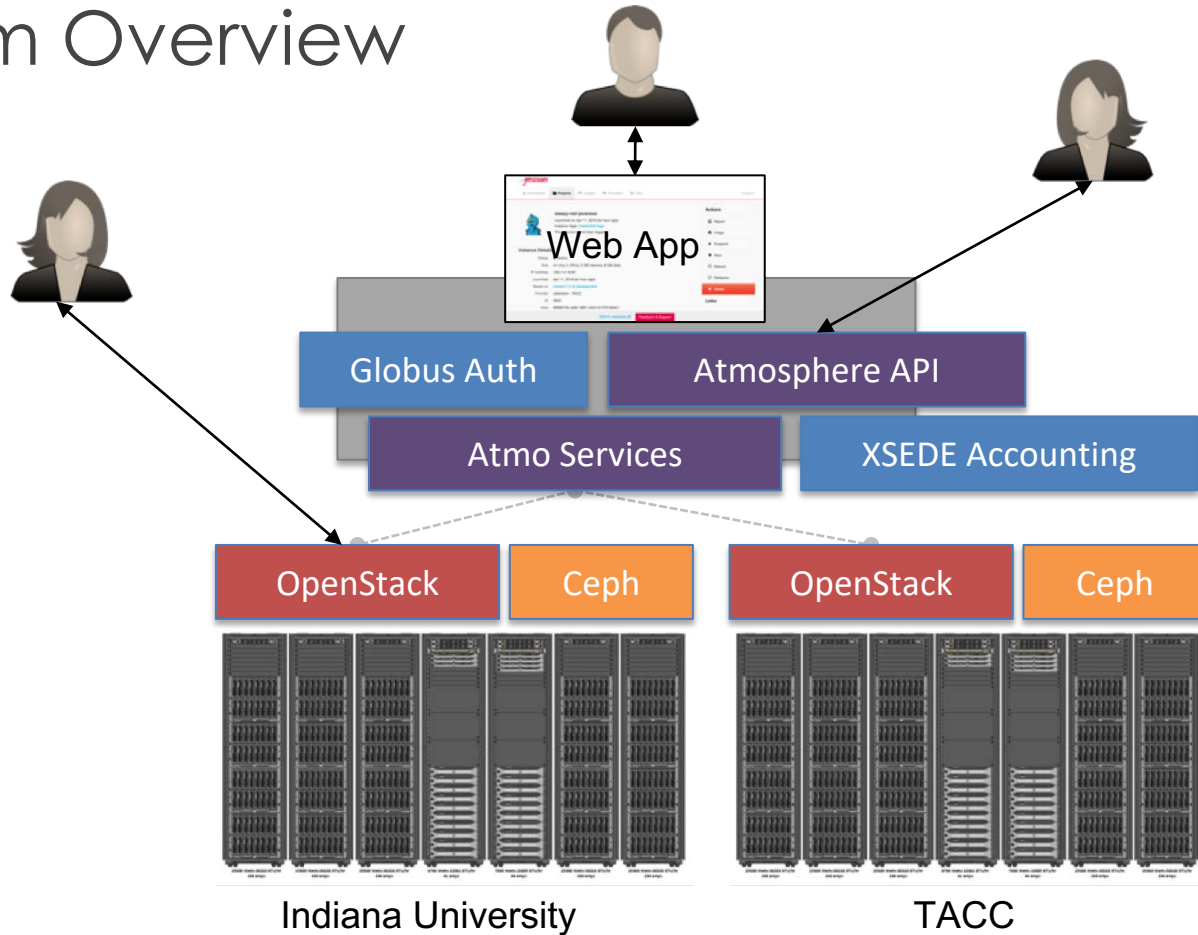
What is Jetstream and why does it exist?

- 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 System Overview



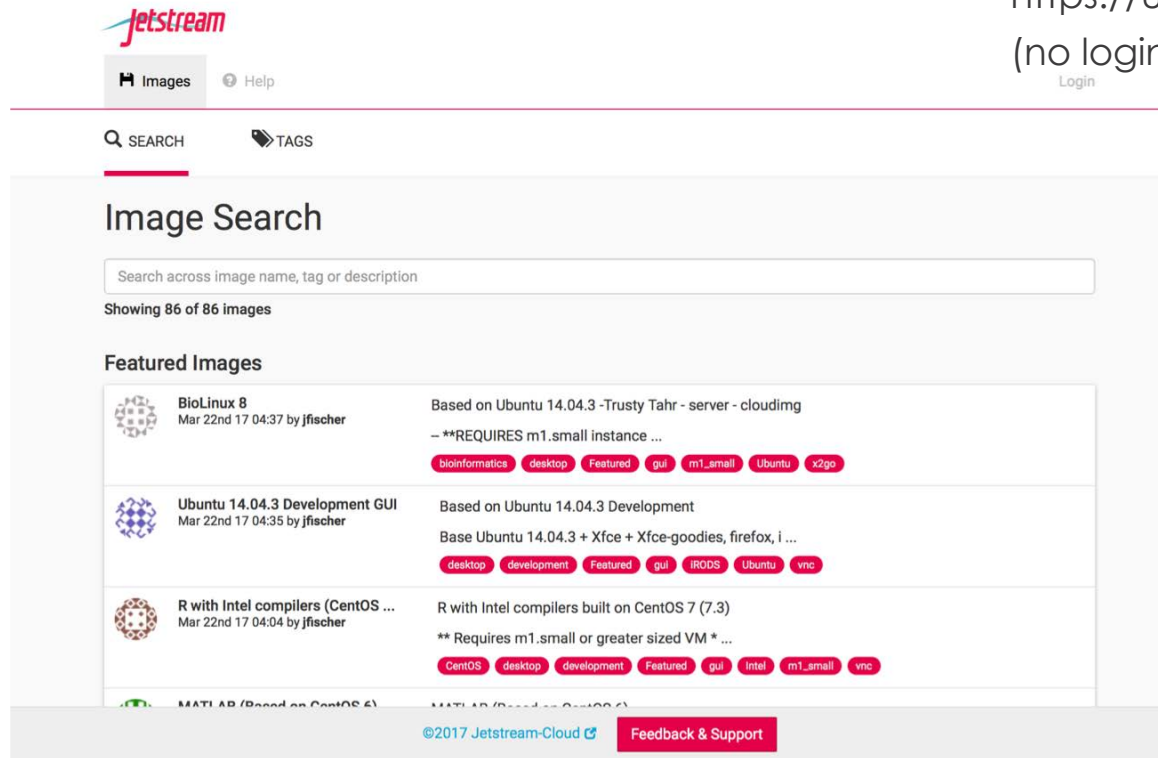
Platform Overview



Jetstream's Atmosphere interface

<https://use.jetstream-cloud.org/>

(no login required to this point)



The screenshot shows the Jetstream Atmosphere interface. At the top left is the Jetstream logo. Below it are navigation links for 'Images' and 'Help'. A search bar is labeled 'SEARCH' and 'TAGS'. The main heading is 'Image Search'. Below the heading is a search input field with the placeholder text 'Search across image name, tag or description'. Below the search bar, it says 'Showing 86 of 86 images'. The 'Featured Images' section lists three items:

- BioLinux 8**: Mar 22nd 17 04:37 by jfischer. Based on Ubuntu 14.04.3 -Trusty Tahr - server - cloudimg. --**REQUIRES m1.small instance ...
Tags: bioinformatics, desktop, Featured, gui, m1_small, Ubuntu, x2go
- Ubuntu 14.04.3 Development GUI**: Mar 22nd 17 04:35 by jfischer. Based on Ubuntu 14.04.3 Development. Base Ubuntu 14.04.3 + Xfce + Xfce-goodies, firefox, i ...
Tags: desktop, development, Featured, gui, iRODS, Ubuntu, vnc
- R with Intel compilers (CentOS ...)**: Mar 22nd 17 04:04 by jfischer. R with Intel compilers built on CentOS 7 (7.3). ** Requires m1.small or greater sized VM * ...
Tags: CentOS, desktop, development, Featured, gui, Intel, m1_small, vnc

At the bottom of the featured images section, there is a footer with the text '@2017 Jetstream-Cloud' and a 'Feedback & Support' button.

Jetstream's Atmosphere interface

(Select identity provider)



Log in to use Jetstream Web App

Use your existing organizational login

e.g., university, national lab, facility, project

Didn't find your organization? Then use [Globus ID](#) to sign in. (What's this?)

Continue

Or

 Sign in with Google



funded by the National Science Foundation
Award #ACI-1445604




Jetstream's Atmosphere interface




(user's home space)

Getting Started




Launch New Instance

Browse Atmosphere's list of available images and select one to launch a new instance.



Browse Help Resources

View a video tutorial, read the how-to guides, or email the Atmosphere support team.



Change Your Settings

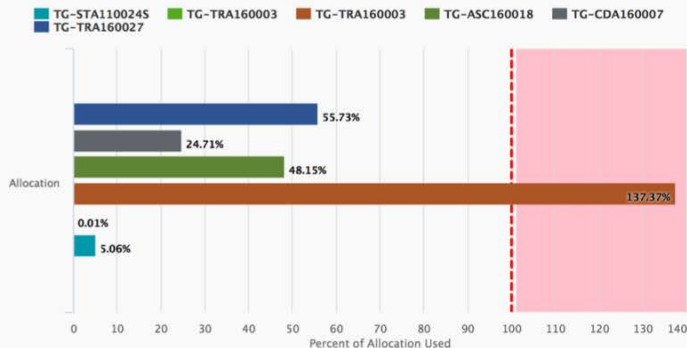
Modify your account settings, view your resource quota, or request more resources.

Resources Used

Need more?

Allocation Source

0 Instances



0 Volumes

Provider Resources

Jetstream - Indiana University Jetstream - TACC

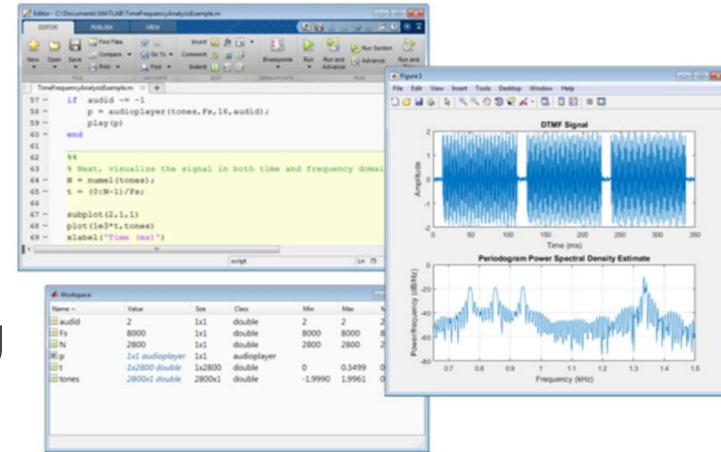


Jetstream usage highlights – March 2018

- 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)

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



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

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.”

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

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!

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

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

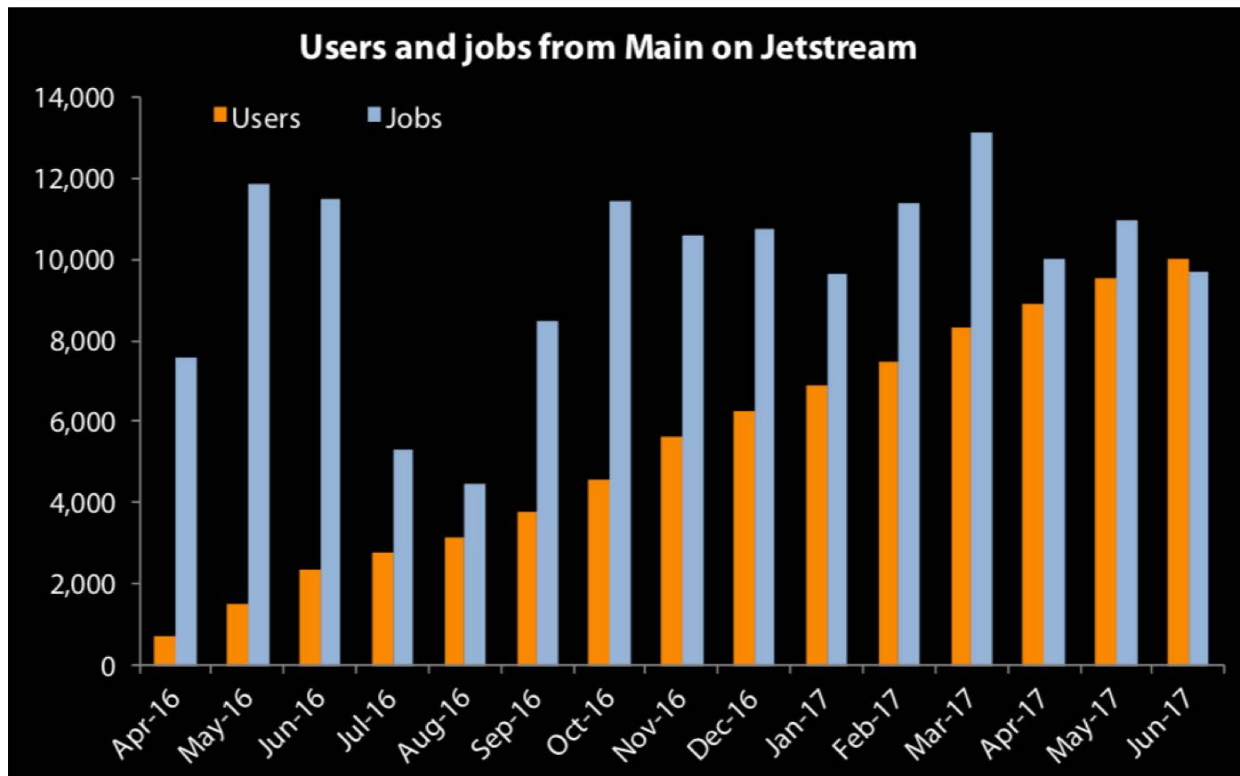
- Pre-built Galaxy image configured with hundreds of tools and access to TBs of genomic reference data, available via the self-launch model within minutes
- Allows users to acquire (free) resources, and gives them complete control



funded by the National Science Foundation
Award #ACI-1445604



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



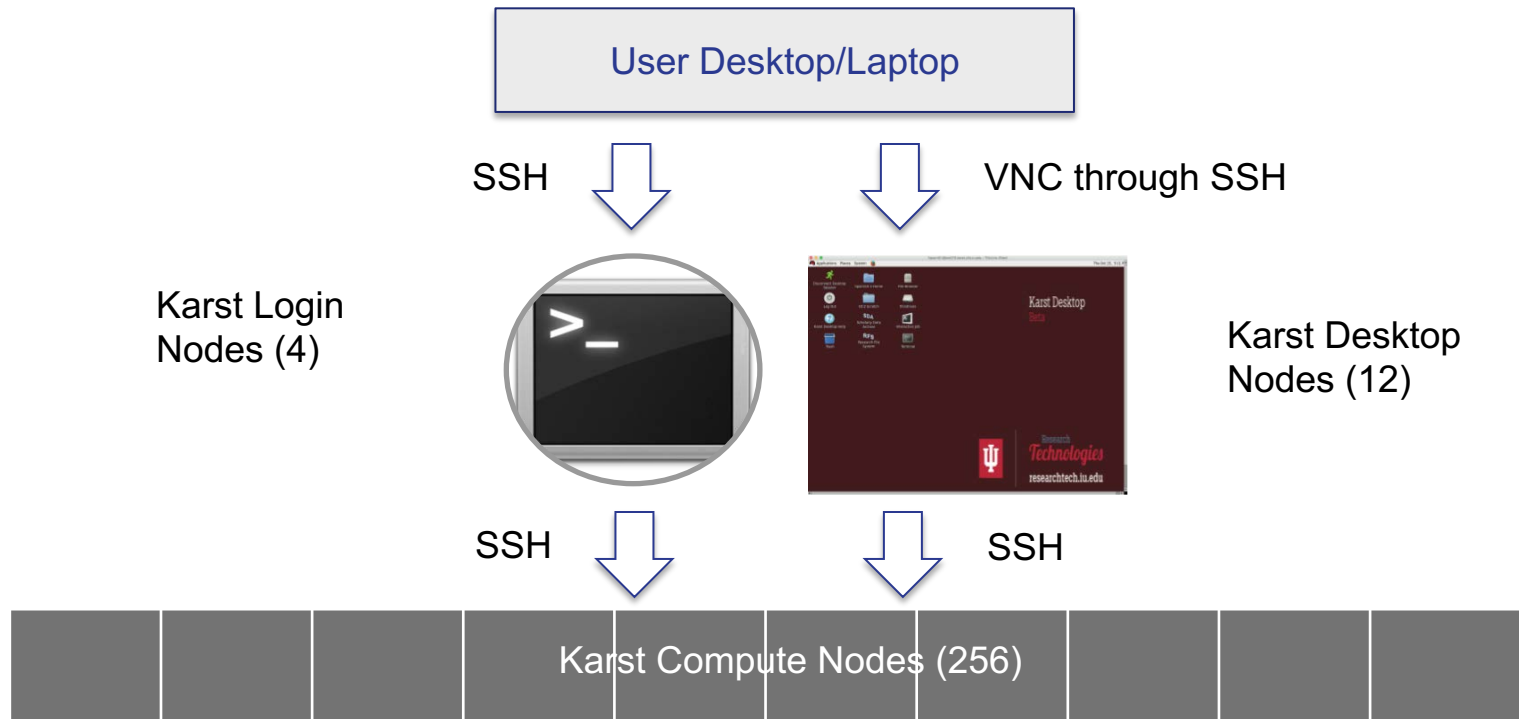
< 5% of the Jobs

15–20% of the CPU time

Disk throughput is important

Now leveraging containers and Jupyter Notebooks

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



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services



**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY



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>

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



funded by the National Science Foundation
Award #ACI-1445604



Jetstream partners



INDIANA UNIVERSITY
PERVASIVE TECHNOLOGY INSTITUTE



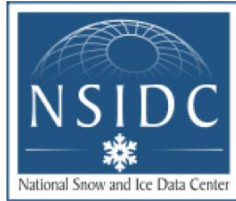
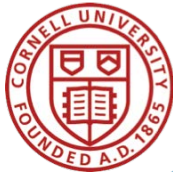
JOHNS HOPKINS
UNIVERSITY



THE UNIVERSITY
OF ARIZONA



THE UNIVERSITY OF
CHICAGO



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 science and engineering cloud environment](#)

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



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

- Hancock, David Y. March 7, 2018. Increasing User and Application Diversity through Programmable Cyberinfrastructure for SIAM PP18. Available at: <http://jetstream-cloud.org/publications.php>
- Jetstream is supported by NSF award 1445604 (David 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.



funded by the National Science Foundation
Award #ACI-1445604

