

Jetstream - Early Operations Performance, Adoption, and Impacts

David Hancock – dyh@iu.edu

SC17 Presentation

Acting Principal Investigator – Jetstream
Program Director – Advanced Cyberinfrastructure
Indiana University Pervasive Technology Institute

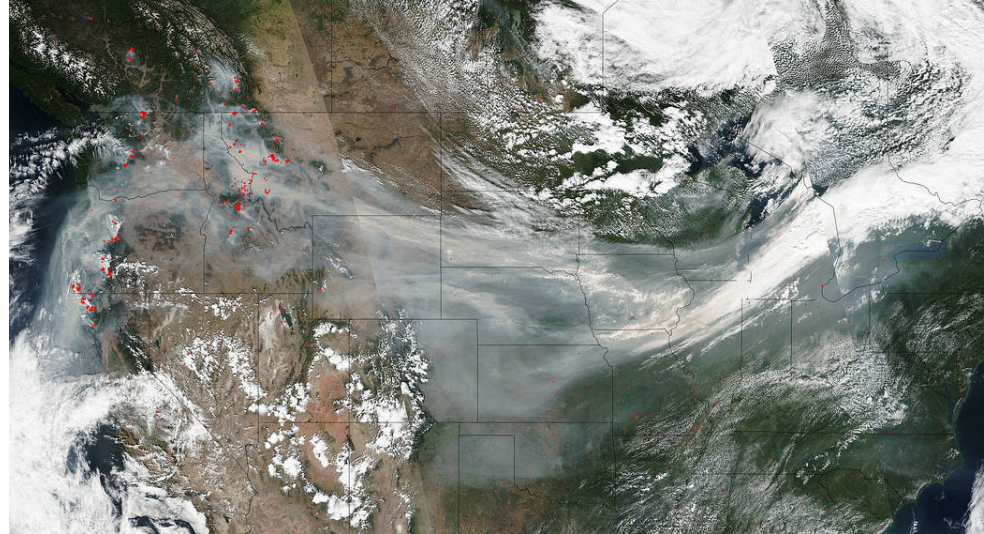
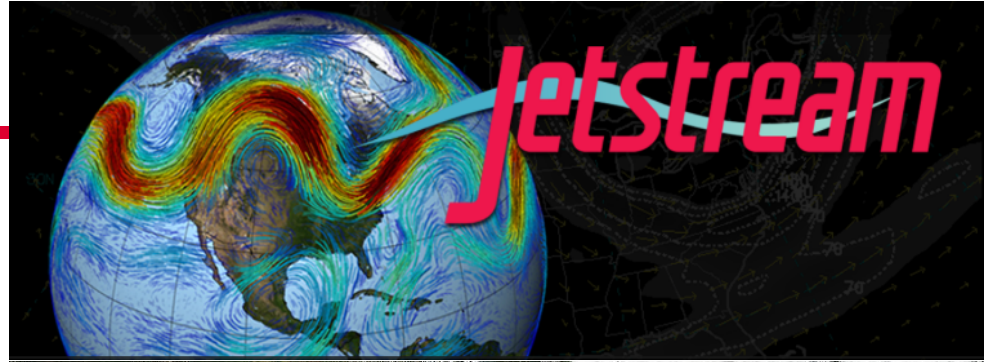
November 14, 2017



What is “the” Jetstream?

- Fast moving air currents
- Hot/Cold air boundaries
- An NSF-funded cloud environment

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



National Science Foundation – Funding in HPC

- Traditionally concentrated on enabling peta-scale capability via track I/II programs
 - Blue Waters – 13.3 petaflops, 2012 (under re-compete)
 - Stampede – 9.6 petaflops, 2013 (extended to Stampede 2)
 - Comet – ~2.0 petaflops, 2014
- Have funded research into building clouds and computer science testbeds
 - CloudLab
 - Chameleon
- Now funding clouds to do research
 - Bridges (Hybrid system)
 - Jetstream



funded by the National Science Foundation
Award #ACI-1445604



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)



funded by the National Science Foundation
Award #ACI-1445604



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>



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

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 and way of the cloud...

- **Cloudy Technologies:** clouds are more the 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...



Flickr user Nanak26 - Normandie

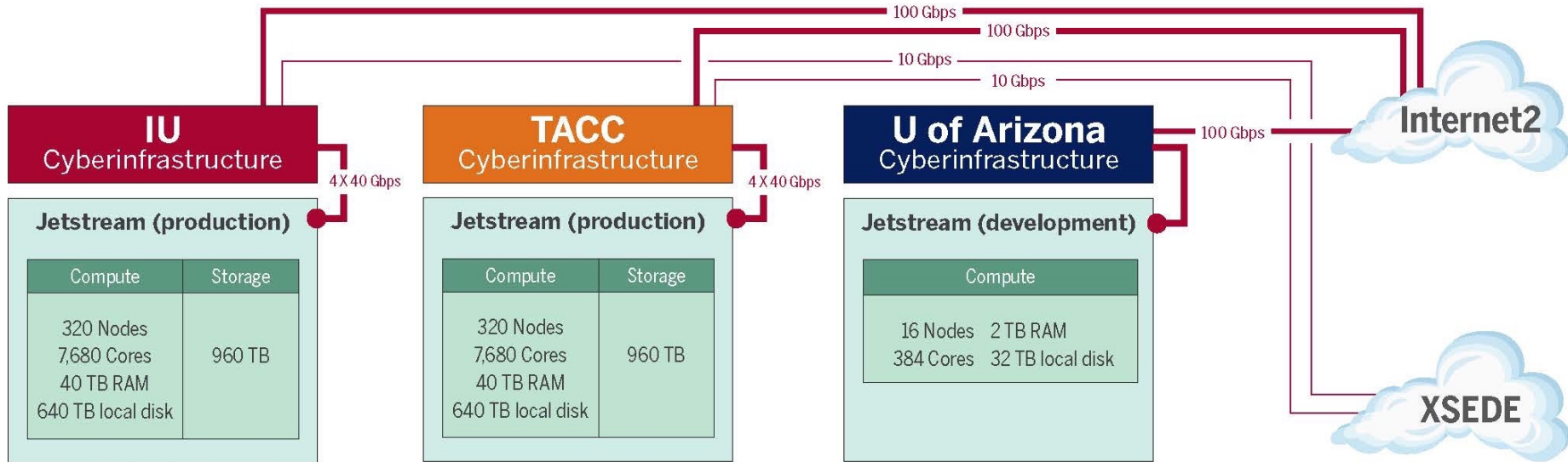


- Cattle, not pets: pets take great amount of care, feeding, and you name them; cattle you intend to have high turnover and you give them numbers.

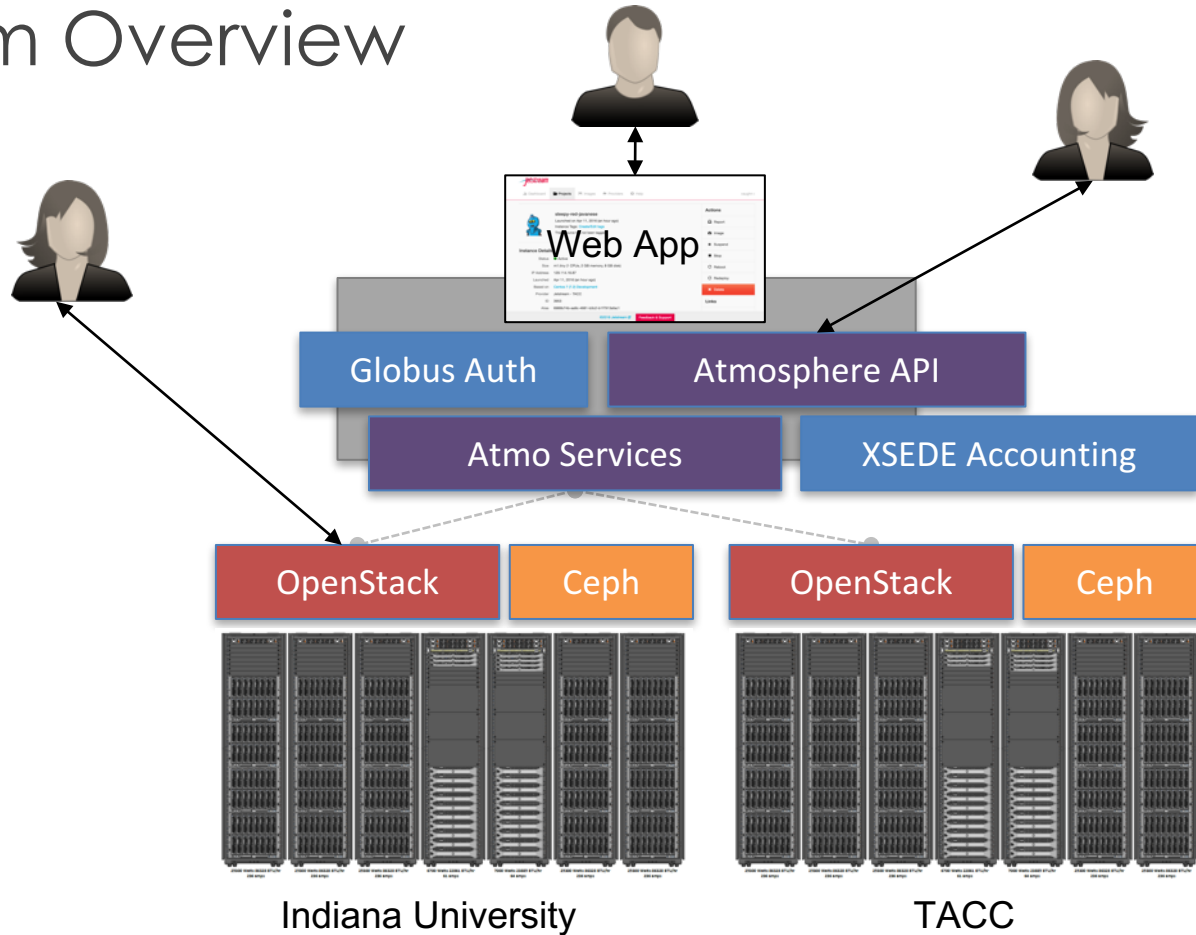
-- Mike Lowe (Jetstream architect)

** Some caveats for gateways, but...

Jetstream System Overview



Platform Overview



Jetstream usage highlights

As of October 2017:

- 360 active XSEDE projects covering 66 fields of science and **2180 active users** representing **191 institutions**
- **86%** of Jetstream users new to XSEDE (at end of PY1)
- >76 million CPU hours allocated to XSEDE projects since June 2016
- 9 science gateways
- 42 education/teaching allocations serving almost 800 students
- Averaging 816 concurrent VMs
- **100%** system availability, **99.4%** cap availability
- **97.7%** “job” completion (at end PY1)



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%

PY1 Challenges

- Outreach efforts largely funded by IU and partners
- Barriers to XSEDE adoption in general
- XRAC process cumbersome to intended audience
- User support needs are significant
- Can be intimidating to port traditional HPC workflows

Jetstream REU Program 2017



- NSF Supplement for 4 undergraduates
- Looking to recruit 4-6 students for 2018
- 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 December 2017 – 15 January 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



PY2 Plans

- Enabling better Jupyter deployments for training and research
- Continue improving trial allocations
- More videos/training
- Encouraging orchestration for more communities
- Image build repository (proposed to XSEDE for funding)



PY2 plans continued...

- Partner with XSEDE Campus Champions and ACI-REF Facilitators to do tutorials for interested under-served researchers
- Explore better ways to communicate with the user community and to allow them to communicate with each other
- Develop additional domain science images with input from the Jetstream and XSEDE community
- Begin work on Windows VM adaptation for Jetstream API if licensing can be resolved

Jetstream partners



INDIANA UNIVERSITY

PERVASIVE TECHNOLOGY INSTITUTE



JOHNS HOPKINS
UNIVERSITY

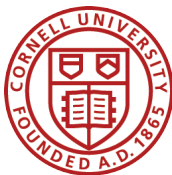


THE UNIVERSITY
OF ARIZONA.

TACC



THE UNIVERSITY OF
CHICAGO



JSU | JACKSON
STATE
UNIVERSITY®



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. November 14, 2017. Jetstream Early Operations Performance, Adoption, and Impacts for SC17 IU Booth Presentation. Will be available at: <http://jetstream-cloud.org/publications.php>
- Jetstream is supported by NSF award 1445604 (Craig Stewart, 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



Things left behind...



Flickr user Oiluj Samall Zeid - Lejos de Yulín



funded by the National Science Foundation
Award #ACI-1445604

