

Methodologies and practices for adoption of a novel national research environment

PEARC18 – July 22-26, 2018 – Pittsburgh, PA

Jeremy Fischer, Brian W. Beck, Sanjana Sudarshan, George Turner, Winona Snapp-Childs, Craig A. Stewart, David Y. Hancock

Presented by: Jeremy Fischer – jeremy@iu.edu
Senior Technical Advisor, UITS Research Technologies

Fischer, J. (2018). Methodologies and practices for adoption of a novel national research environment. Pittsburgh, PA. 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



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>



funded by the National Science Foundation
Award #ACI-1445604



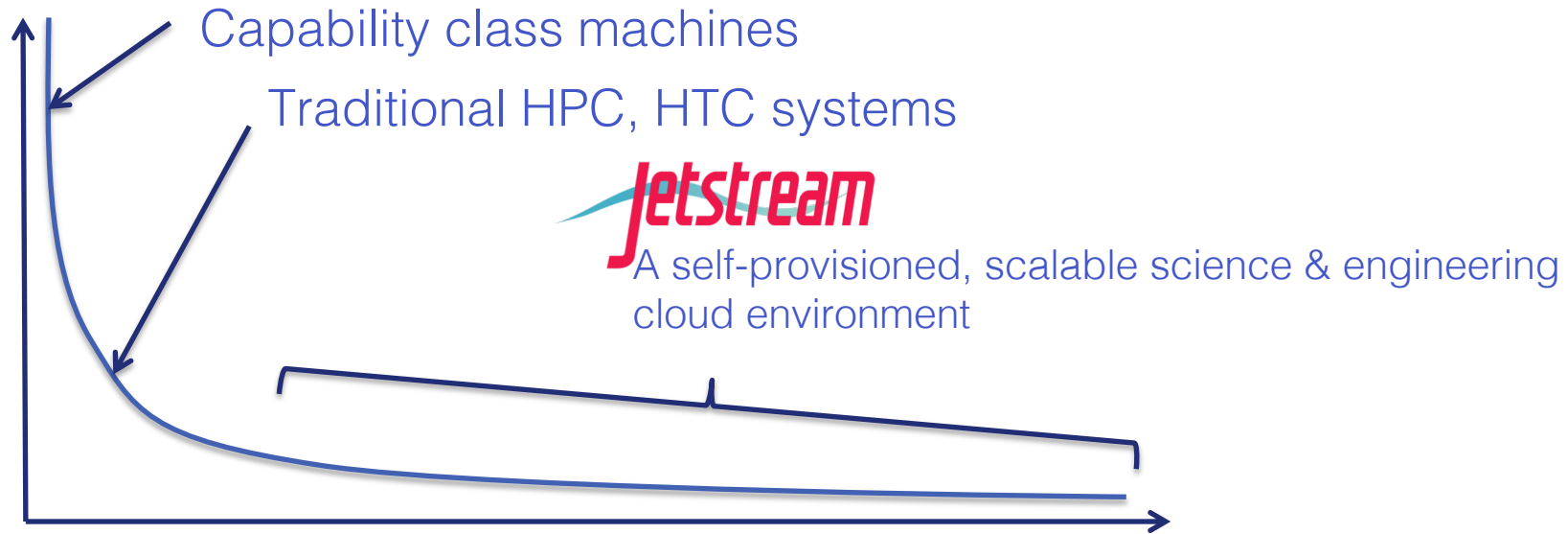
So we got funding...now what?

- The difference between knowing what you're going to build and knowing how researchers plan to use it
- Democratizing resources – the culture of inclusion
- Focus on the "long tail" users



What is Jetstream?

Particularly focused on researchers working in the “long tail” of science with born digital data



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)

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



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

Cultivating relationships for advancing research

- Finding the communities of users
- National resources like Champions, ACI-Ref Facilitators
- EPSCoR opportunities
- Regional and local resources – support specialists

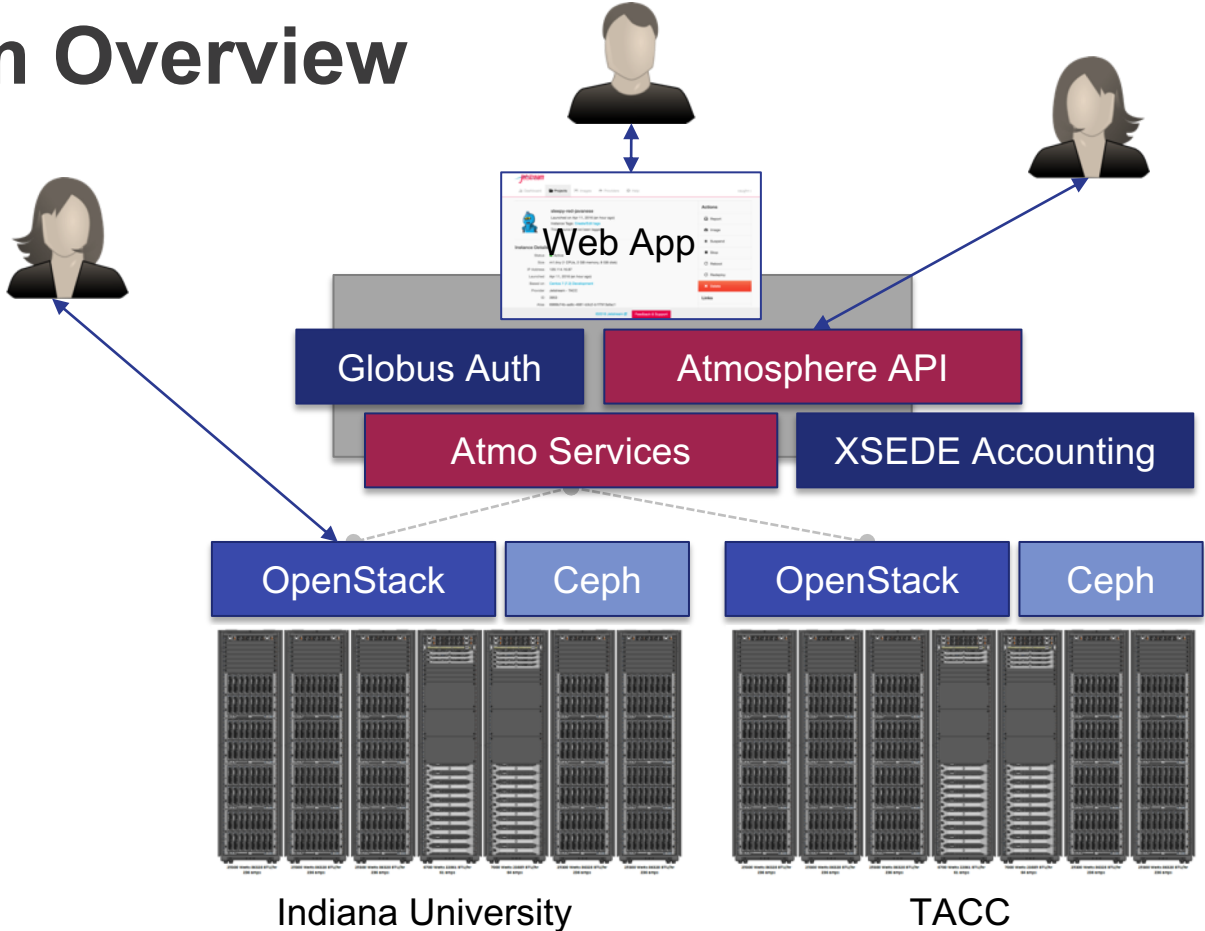
Start planning early... ...and know the plan is going to change...a lot.

- Finding someone to lead the efforts for EOT work
- Make a marketing plan
- The ultimate goals: communicating performance, ease of use, and value to the researcher/educator

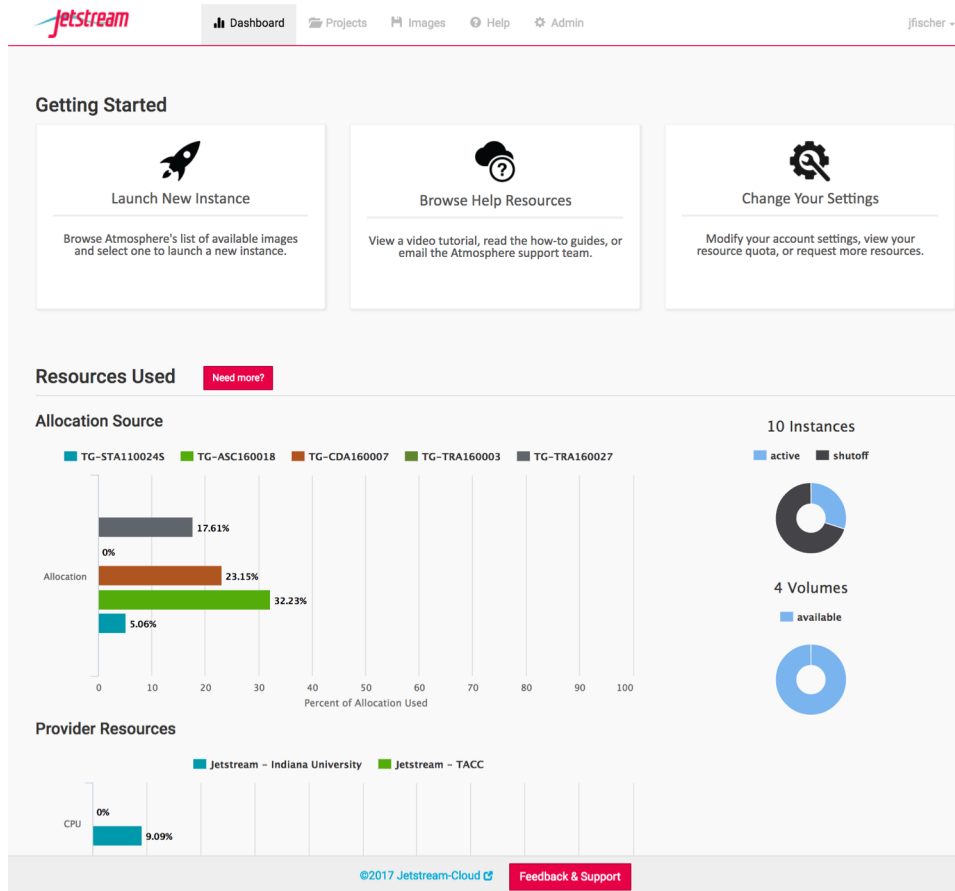
Early Engagements

- Aimed at XSEDE (since it was to be an XSEDE resource)
- Southern Partnership in Advanced Networking (SPAN) workshop series (Alabama, Florida, South Carolina workshops)
- Internet2 Tech Exchange
- Galaxy Community Conference
- HPC User Forum
- Earth Science Information Partners (ESIP)
- OpenStack Summit
- Regional conferences such as I-Light

Platform Overview



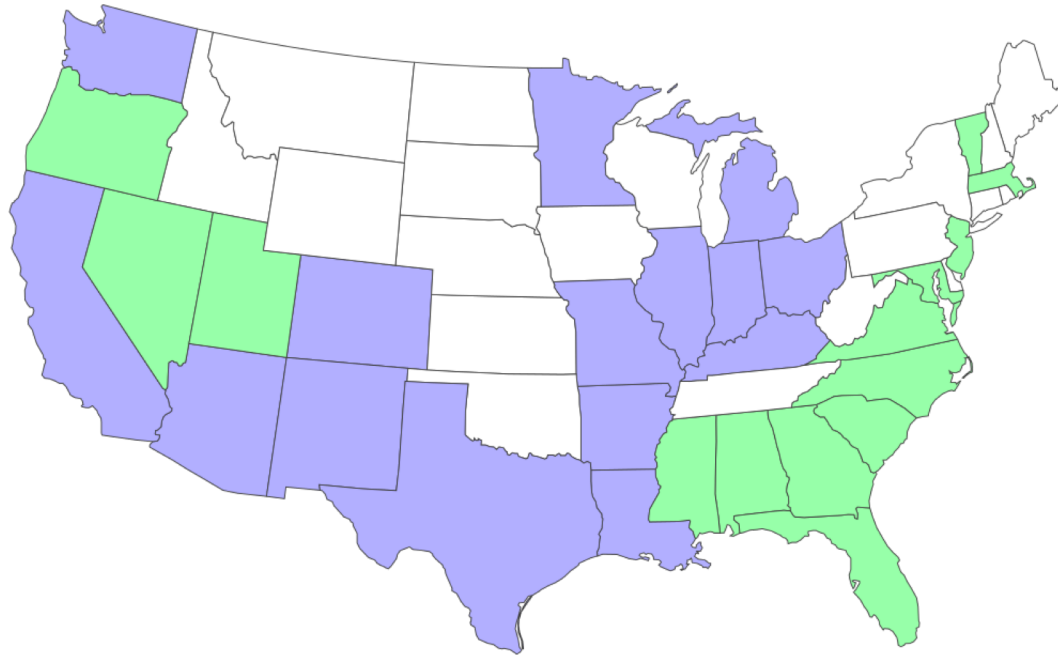
The Jetstream Atmosphere web interface



Moving forward to post-operations events

- Continuing to present on Jetstream wherever possible
- Getting researchers on Jetstream at events
- Non-Jetstream sponsored workshops
- Courses at universities and colleges
- Jetstream EOT staff directly worked with 762 researchers, staff, and students at 26 events from July 2016 to June 2017

Oh the places we've been...



■ States visited during Jetstream EOT efforts - June 2017 to May 2018

■ States visited during Jetstream EOT efforts - prior to Production Year 2

Lessons learned from the 1st year of EOT

- Unsurprisingly, hands on tutorials are more engaging
- Getting to all of the places that wanted infoshares or tutorials was almost impossible with just one person
- Even with everyone on the project being tapped at one point or another for EOT events, more staffing was needed, especially to grow outreach efforts

Lessons learned beyond year 1

- Still encountering many researchers that have never heard of XSEDE or Jetstream
- Not all events are the best for promoting Jetstream but even so, they're rarely a waste of time (even if we might not go back)
- Documentation and training materials need to be ever-evolving

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%

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



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>

Moving forward – beyond EOT events

- In PY2 the team traveled to more than 40 events held in more than 15 states
- 183 people were able to try out the system using trial allocations and more than 259 obtained startup allocations as a result of these efforts
- We needed to find out how satisfied these users are and if they would or would not be considering applying for research allocations going forward

Surveying the users

- Two groups of participants (trial and certain startup users)
- 313 people surveyed, 102 responded with informed consent – 32.6% response rate
- Goal of the survey was to better understand why those individuals who had obtained a small allocation (trial or startup) had not yet applied for a larger research allocation
- Survey instrument in paper

Survey results

- 81.63% were satisfied or extremely satisfied with Jetstream
- 73.47% were likely to apply for a research allocation
- One of the top 3 reasons for not planning to obtain a research allocation was “I was able to complete my work with the smaller (trial or startup allocation).”

Moving forward

- Future surveys/user feedback needed to help steer Jetstream's EOT and development efforts – allocation sizes, reproducibility/DOI needs
- Surveys of non-users (domain science or other groups) to find services and training topics we might not be focusing on well
- Approaching new groups and institutions - social/peer influences to increase adoption among communities
- that have not traditionally made extensive use of NSF-funded cyberinfrastructure.

Jetstream usage highlights – 1 June 2018

- 395 active XSEDE projects covering 70 fields of science and **2481 active users** representing **194 institutions**
- **80%** of Jetstream users have **not used any other XSEDE system**
- **>118M** CPU hours allocated to XSEDE projects since June 2016
- 14 active science gateways
- 43 education/teaching allocations serving over 850 students



funded by the National Science Foundation
Award #ACI-1445604



*M&O PY2 to date 11/12 months

Jetstream Partners



INDIANA UNIVERSITY
PERVASIVE TECHNOLOGY INSTITUTE



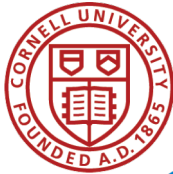
JOHNS HOPKINS
UNIVERSITY



THE UNIVERSITY
OF ARIZONA



THE UNIVERSITY OF
CHICAGO



funded by the National Science Foundation
Award #ACI-1445604



Questions?

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

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

License Terms

- Fischer, Jeremy. May 30, 2018. Methodologies and practices for adoption of a novel national research environment. Also available at: <http://jetstream-cloud.org/research/publications.php>
- Jetstream is supported by NSF award 1445604 (David Y. Hancock, IU, PI)
- XSEDE is supported by NSF award 1053575 (John Towns, UIUC, PI)
- Atmosphere is developed by CyVerse (formerly The iPlant Collaborative) under grant NSF DBI-1265383
- 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

