

Jetstream: A national research and education cloud

ISCB 2018– July 9, 2018 – Chicago, IL

Sanjana Sudarshan, PhD - ssudarsh@iu.edu

Senior Technical Advisor,
UITS Research Technologies

Sudarshan, S. (2018). Jetstream: A national research and education cloud. ISCB Chicago . 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 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



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>



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 (SEAGrid, Galaxy, GenApp NAMDRunner, CIPRES 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



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



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...)



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...



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 is 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...

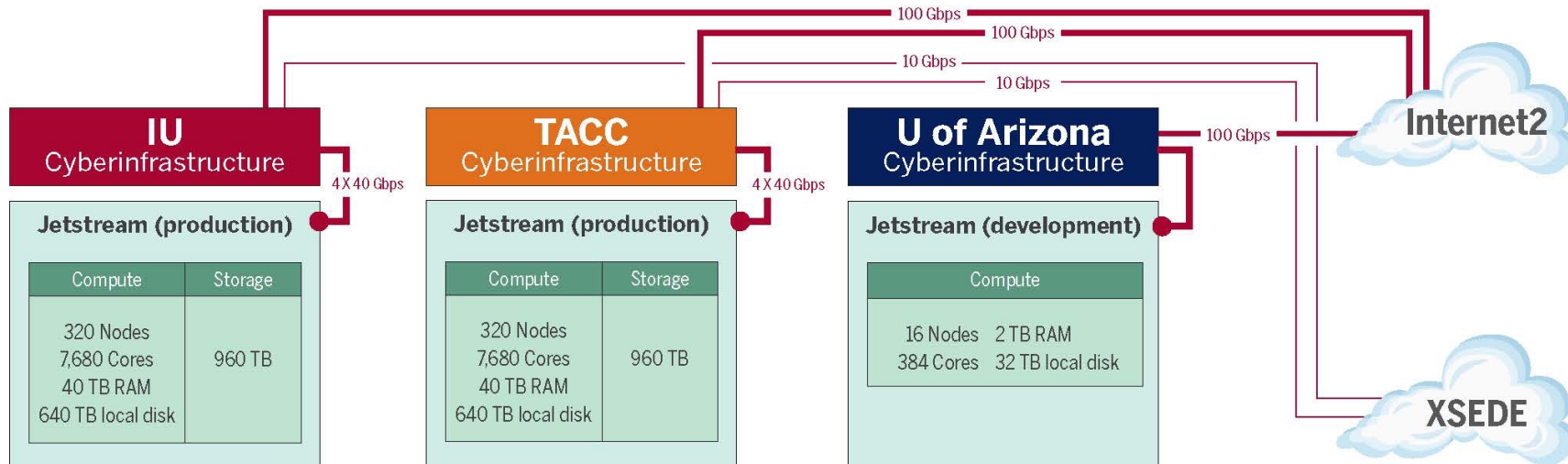


Cows, not pets: pets take great amount of care, feeding, and you name them; cows you intend to have high turnover and you give them numbers.

-- Mike Lowe (Jetstream architect)

**some caveats for gateways...

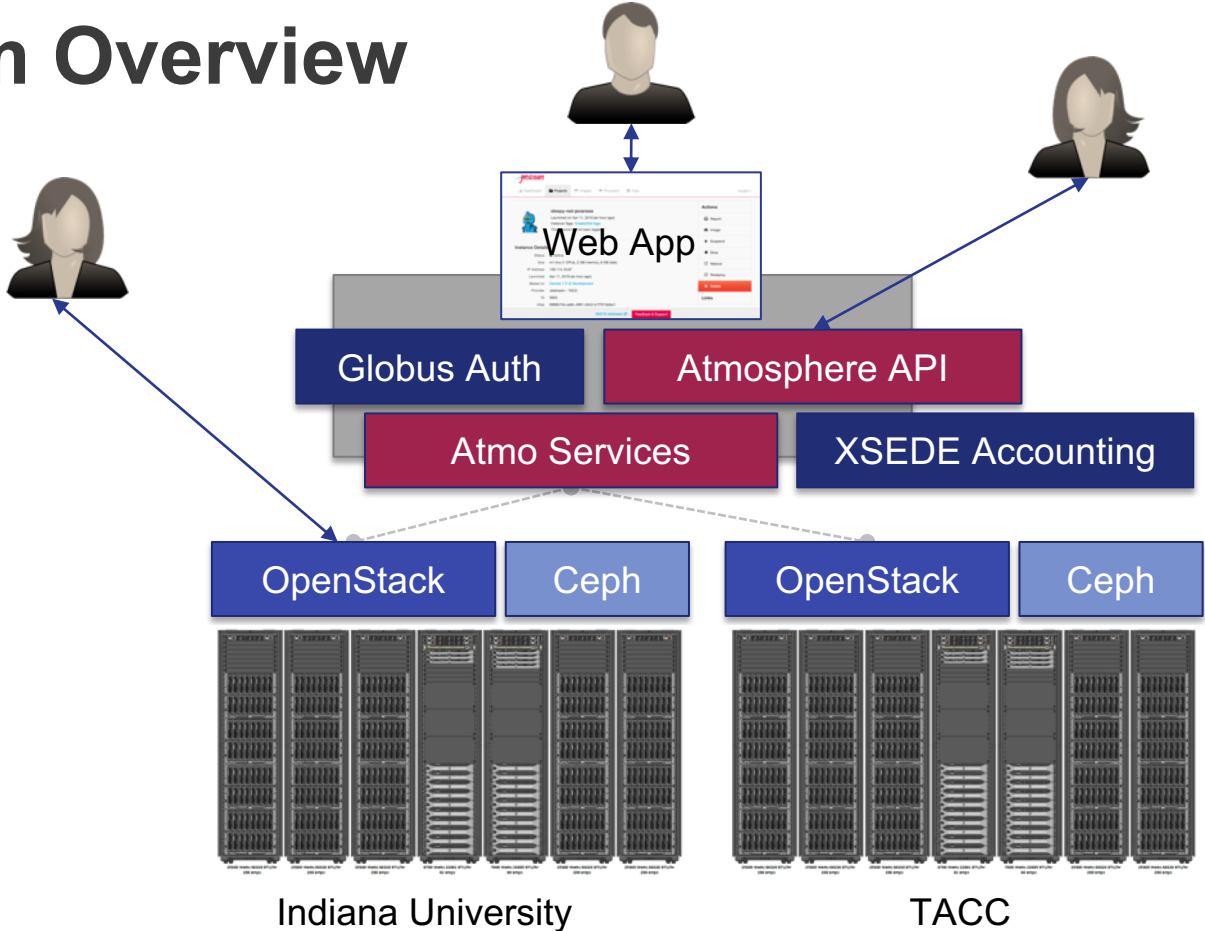
Jetstream System Overview



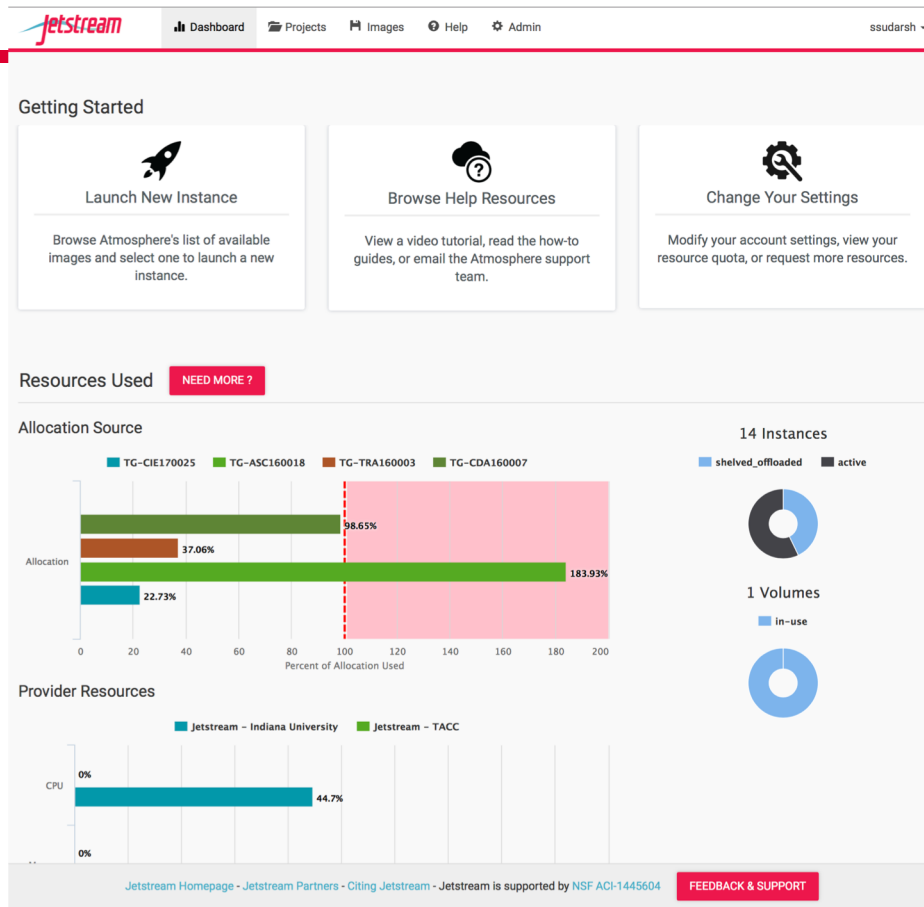
Production cloud hardware (per site)

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

Platform Overview



The Jetstream Atmosphere web interface



The Jetstream Atmosphere web interface



Dashboard Projects Images Help Admin

ssudarsh






SEARCH FAVORITES (0) MY IMAGES (7) MY IMAGE REQUESTS TAGS

Image Search

Search across image name, tag or description

Showing 100 of 563 images

Featured Images

-  **Ubuntu 18.04 Devel and Docker**
Jun 19th 18 03:40 by jfischer
Ubuntu 18.04 LTS Development + GUI support + Docker
Based on Ubuntu cloud image for 16. ...
Featured gui Ubuntu
-  **Centos 7 (7.5) Development GUI**
Jun 19th 18 03:12 by jfischer
Centos 7 (7.5) Development GUI
Installation size ~ 4.5GB
CentOS development docker docker-compose Featured gui iRODS
-  **Ubuntu 16.04 Devel and Docker**
Jun 19th 18 01:43 by jfischer
Ubuntu 16.04 LTS Development + GUI support + Docker
Based on Ubuntu cloud image for 16.04 ...
base desktop development docker docker-compose Featured Ubuntu vnc x2go
-  **Matlab (R2017b - Ubuntu 16.04)**
Jun 19th 18 11:26 by jfischer
Matlab R2107b install - requires an m1.medium or larger
Based on Ubuntu 16.04 LTS Dev ...
development Featured gui m1_medium Mathematics MATLAB Simulink Ubuntu
-  **Intel Development (CentOS 7)**
Jun 19th 18 10:59 by jfischer
Intel compilers and development environment
***REQUIRES a m1.small or larger VM to la ...
CentOS desktop development Featured gui Intel m1_small vnc



Look! It's more Jetstream web interface!

Launch an Instance / Basic Options

Basic Info

Instance Name: Centos 7 (7_5) Development GUI

Base Image Version: 1.27

Project: Iscb

Resources

Allocation Source: TG-CIE170025

Provider: Jetstream - Indiana University

Instance Size: m1.tiny (CPU: 1, Mem: 2 GB, Disk: 8 GB)

- m1.tiny (CPU: 1, Mem: 2 GB, Disk: 8 GB)
- m1.small (CPU: 2, Mem: 4 GB, Disk: 20 GB)
- m1.medium (CPU: 6, Mem: 16 GB, Disk: 60 GB)
- m1.large (CPU: 10, Mem: 30 GB, Disk: 60 GB)
- s1.large (CPU: 10, Mem: 30 GB, Disk: 120 GB)
- m1.xlarge (CPU: 24, Mem: 60 GB, Disk: 60 GB)
- s1.xlarge (CPU: 24, Mem: 60 GB, Disk: 240 GB)
- m1.xxlarge (CPU: 44, Mem: 120 GB, Disk: 60 GB)
- s1.xxlarge (CPU: 44, Mem: 120 GB, Disk: 480 GB)

A total 156 of 360 allotted GBs of Memory

← Back ⚙️ Advanced Options CANCEL LAUNCH INSTANCE

Even more Jetstream web interface...

The screenshot displays the Jetstream web interface for a specific instance. The navigation bar at the top includes 'Dashboard', 'Projects', 'Images', 'Help', and 'Admin', with the user 'ssudarsh' logged in. The main content area is titled 'Resources > Centos 7 (7_5) Development GUI'. It features a robot icon, the instance name, and an 'Allocation Source' dropdown menu set to 'TG-CIE170025'. Below this, a progress bar indicates 'Allocation Used' as '23% of 290000 SUs from TG-CIE170025'. The 'Instance Details' section lists various attributes: Status (Active), Activity (N/A), Size (m1.tiny), IP Address (149.165.168.222), Launched date (Jul 7, 2018), Based on image (Centos 7 (7.5) Development GUI v1.27), Provider (Jetstream - Indiana University), ID (23837), and Alias (e2e5c9ae-6920-42ab-9f62-358a0e6f7e3c). On the right, an 'Actions' panel offers options like Report, Image, Suspend, Shelve, Stop, Reboot, Redeploy, and Delete (highlighted in red). A 'Links' panel provides shortcuts for 'Open Old Web Shell', 'Open Old Web Desktop', 'Open Web Shell', and 'Open Web Desktop'.

Resources > Centos 7 (7_5) Development GUI

Centos 7 (7_5) Development GUI

Allocation Source

TG-CIE170025

Allocation Used

23% of 290000 SUs from TG-CIE170025

Instance Details

Status	Active
Activity	N/A
Size	m1.tiny (CPU: 1, Mem: 2 GB, Disk: 8 GB)
IP Address	149.165.168.222 Copy
Launched	Jul 7, 2018 (a day ago)
Based on	Centos 7 (7.5) Development GUI v1.27
Provider	Jetstream - Indiana University
ID	23837
Alias	e2e5c9ae-6920-42ab-9f62-358a0e6f7e3c Copy

Actions

- Report
- Image
- Suspend
- Shelve
- Stop
- Reboot
- Redeploy
- Delete

Links

- Open Old Web Shell
- Open Old Web Desktop
- Open Web Shell
- Open Web Desktop

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



funded by the National Science Foundation
Award #ACI-1445604



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



funded by the National Science Foundation
Award #ACI-1445604

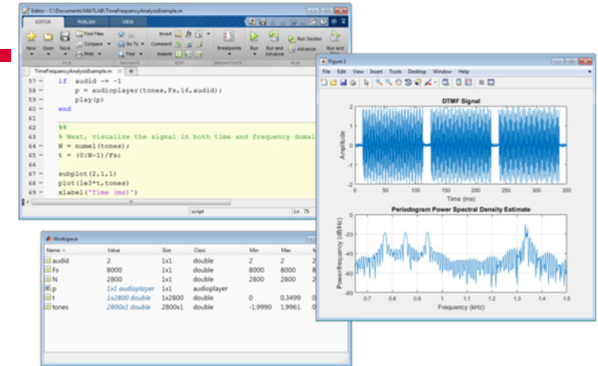


Discipline or area of interest	#of Jetstream allocations	SUs allocated on Jetstream	% of SUs allocated on Jetstream	% of all SUs allocated on other XSEDE-supported systems
Astronomy	2	1,108,096	3.04%	8.61%
Atmospheric Sciences	4	2,752,400	7.55%	3.73%
Biological Sciences	57	5,199,000	14.27%	4.95%
Campus/Domain 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

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



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



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



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 usage highlights – 1 May 2018

- 313 active XSEDE projects covering 71 fields of science and **2411 active users** representing **205 institutions**
- **80%** of Jetstream users new to XSEDE
- >117 million CPU hours allocated to XSEDE projects since June 2016
- 14 active science gateways
- 41 education/teaching allocations serving almost 850 students
- Averaging 1151 concurrent active VMs in current quarter, 955 in PY2*
- **100%** system availability, **99.2%** cap availability in PY2*
- **98.8%** “job” completion in PY2*



funded by the National Science Foundation
Award #ACI-1445604



*M&O PY2 to date 11/12

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

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 June 2018 - 15 July 2018.
- Research allocation: Main project description (up to 10 pages) and Scaling doc (up to 5 pages) – We can help!

Jetstream REU Program



- NSF Supplement for undergraduates
- 6 students for 2018
- REU student videos on YouTube
<https://www.youtube.com/user/IUPTI>

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/>



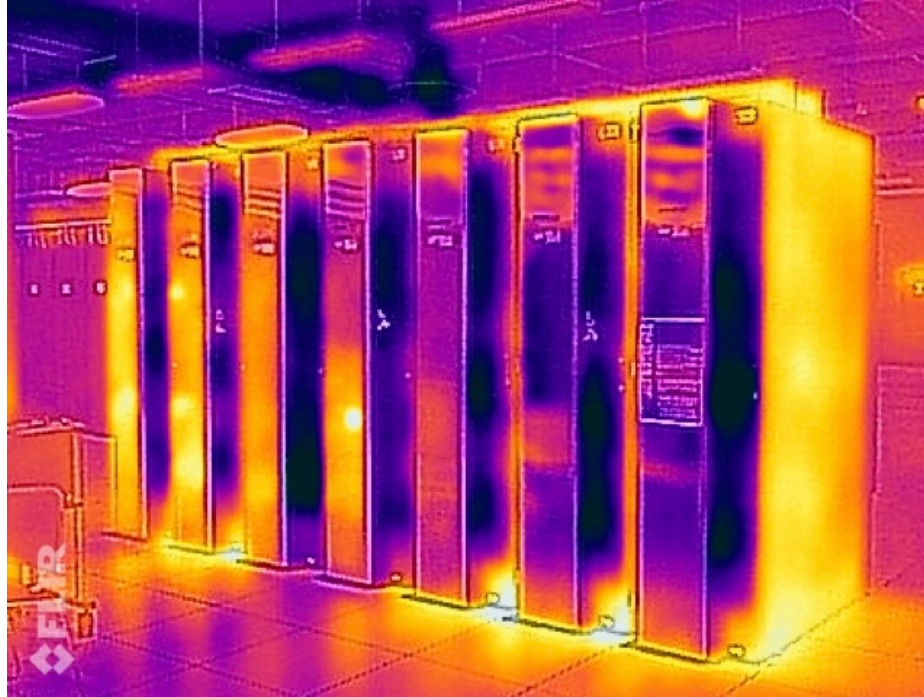
funded by the National Science Foundation
Award #ACI-1445604



Jetstream Fun: Happy cluster / Angry Cluster



Infrared image of Jetstream



ChilledDoor™
Rack Cooling System
motivair™

Jetstream Partners



INDIANA UNIVERSITY

PERVASIVE TECHNOLOGY INSTITUTE



JOHNS HOPKINS
UNIVERSITY

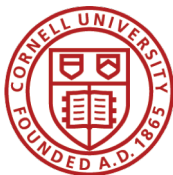


THE UNIVERSITY
OF ARIZONA

TACC



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: ssudarsh@iu.edu

License Terms

- Fischer, Jeremy and Sudarshan, Sanjana. July 9, 2018. Jetstream Overview, ISCB Chicago. 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)
- 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

