

Jetstream

Introduction to Jetstream: Accessible cloud computing for the national science and engineering communities

Brian W. Beck, PhD

ORCID 0000-0002-7731-3457

Life Science Computing & DFW Liaison

Texas Advanced Computing Center, University of Texas at Austin

bbeck@tacc.utexas.edu

(940) 488-4703



funded by the National Science Foundation
Award #ACI-1445604

Jetstream - Expanding NSF's eXtreme Digital (XD) reach and impact

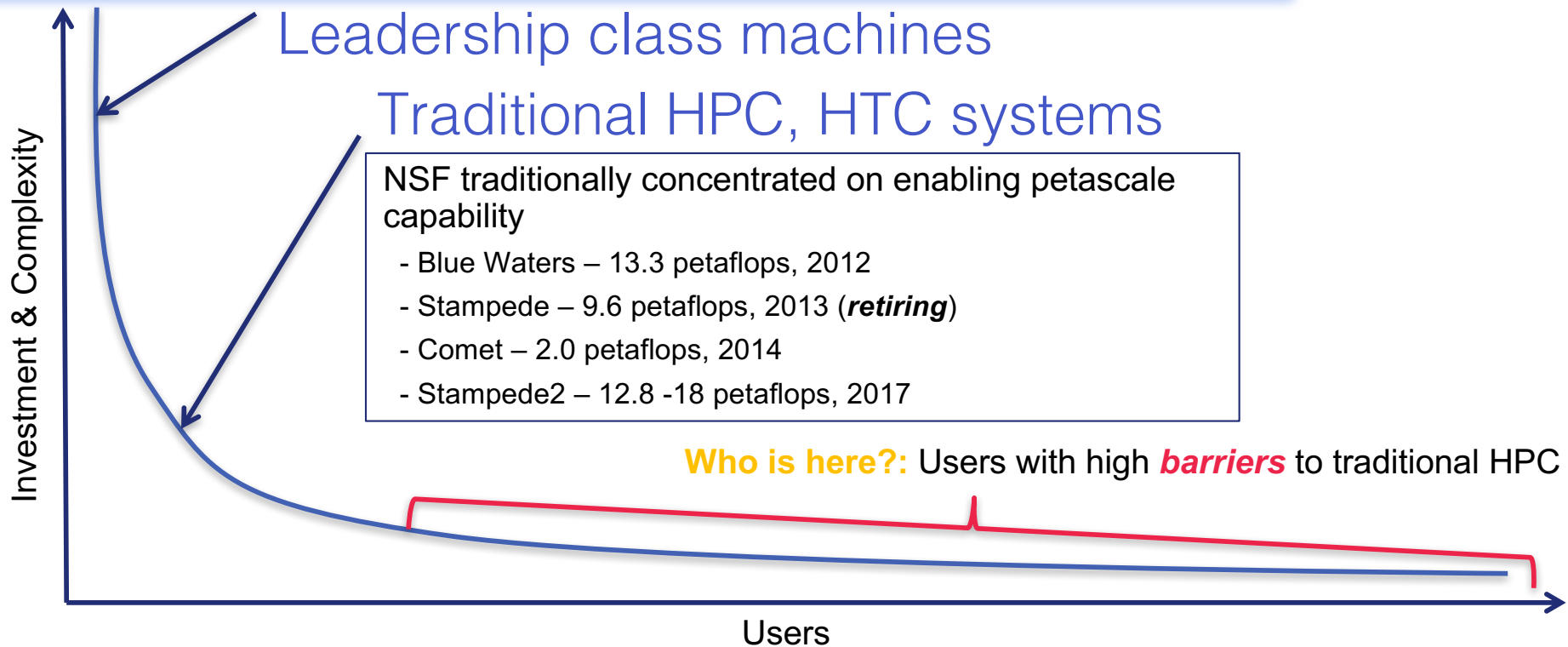
- Around **350,000** researchers, educators, & learners received NSF support in 2015
 - **Less than 2%** completed a computation, data analysis, or visualization task on XD/XSEDE program resources
 - **Less than 4%** had an XSEDE Portal account (<http://portal.xsede.org>)
 - **70% of researchers surveyed** 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** or **limits are too restrictive**.



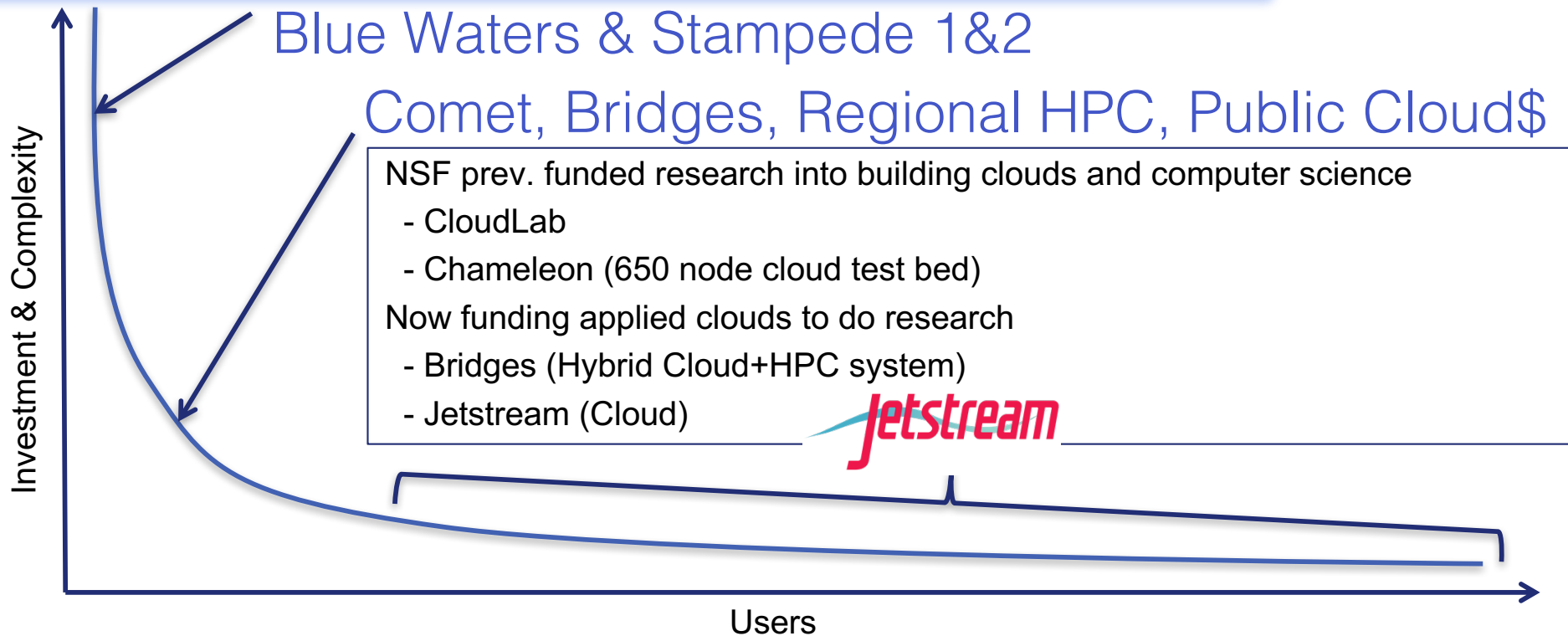
funded by the National Science Foundation
Award #ACI-1445604



Jetstream serves the Long Tail of Science, Engineering, and Education



Jetstream serves the Long Tail of Science, Engineering, and Education



Jetstream serves the Long Tail of Science, Engineering, and Education

Researchers, developers, and scientists who:

- Need between 1 and a few hundred cores
 - **RIGHT NOW**
 - For an extended period of time
 - Not forever
- Want to fully **customize** the OS and configuration for **their** research computing environment
- Are working with **cloud-native** applications & workflows
- Use **interactive mode** for their computing & analytics



funded by the National Science Foundation
Award #ACI-1445604



What is Jetstream?

- First **production cloud platform** for NSF-sponsored researchers
- Provides **on-demand interactive computing** and analysis
 - Will support persistent gateways (CyVerse, Galaxy, SEAGrid, GenApp, NAMDRunner, CIPRES, *et al.*)
- Enables **configurable environments** and architectures
 - User-selectable library of preconfigured **virtual machines (VM)**
- Supports **computational reproducibility** and sharing
 - Share VMs and then store, publish via IU Scholarworks (DOI)
- Democratizes access to **cloud-native software**
- Focused on **ease of use** for all adopters
- **Expands the community of users benefiting from NSF cyberinfrastructure**



funded by the National Science Foundation
Award #ACI-1445604



Jetstream serves the Long Tail of Science, Engineering, and Education

Jetstream is useful to

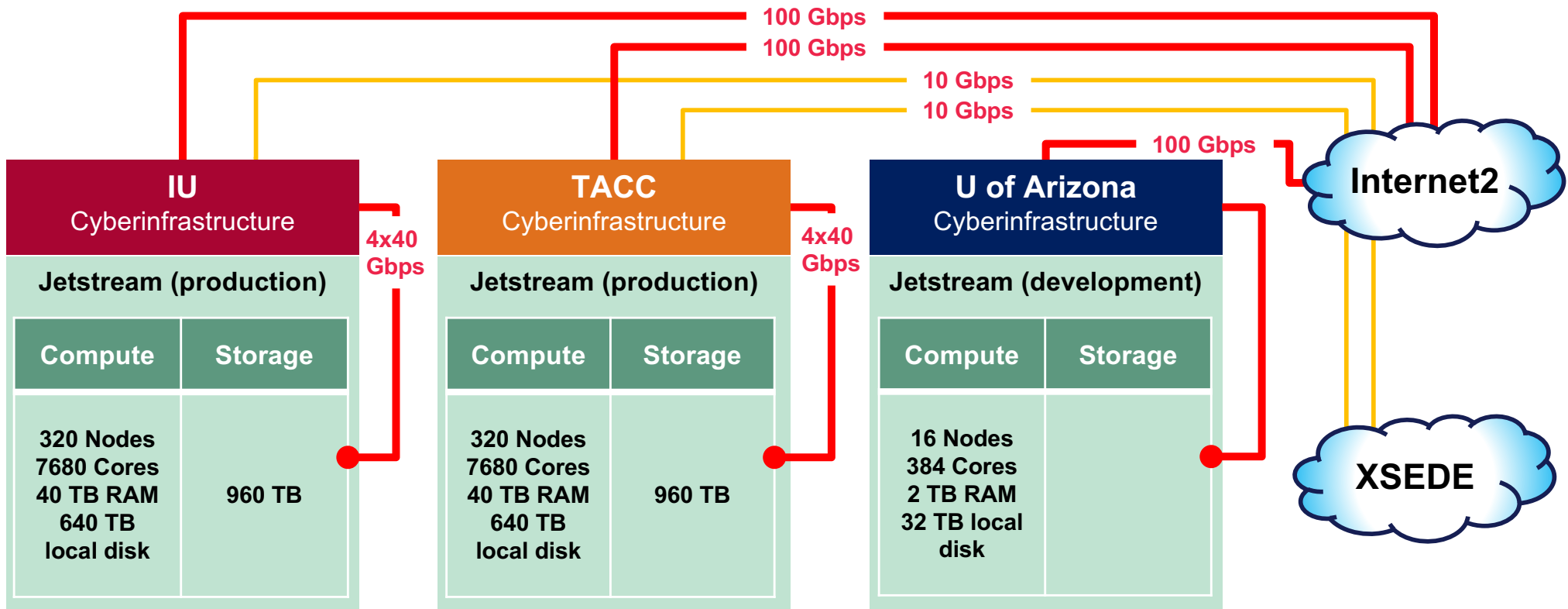
- Science gateway operators
 - Run web applications, databases, and services (front-end)
 - Use it for on-demand provisioned compute capacity
- STEM educators teaching a variety of subjects
 - Create a reference *Virtual Machine* (VM) appliance
 - Provision an entire classroom
 - Minimize need for local IT support
 - Onboard students into XSEDE ecosystem



funded by the National Science Foundation
Award #ACI-1445604



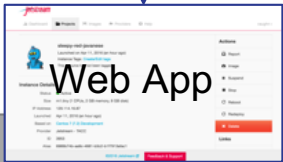
Jetstream System Overview



Platform Overview



Novice-to-Advanced User:
Easy-to-Use frontend



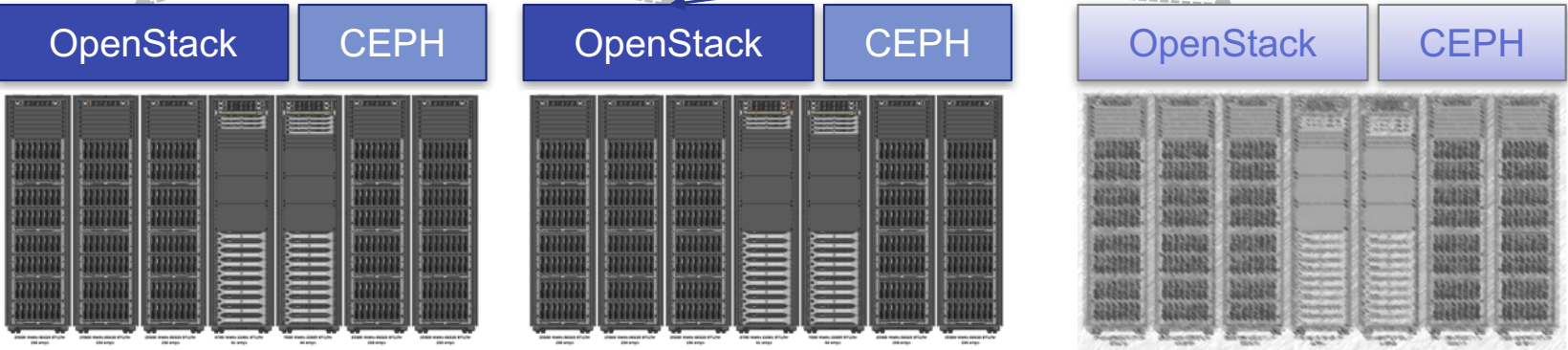
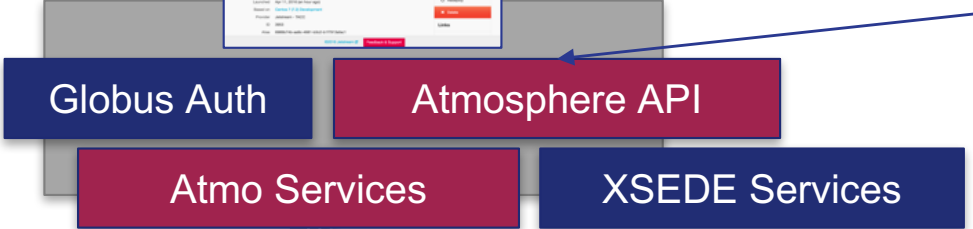
Web App



Developer: Creating Services with Atmosphere



Developer: Creating Services with OpenStack



Indiana University

TACC

Potentially, Others

Seeking New
Cloud Partners!!

Hardware and VM Instance "Flavors"

VM Host Configuration

- Dual Intel E-2680v3 "Haswell"
- 24 physical cores/node @ 2.5 GHz (Hyperthreading on)
- 128 GB RAM
- Dual 1 TB local disks
- 10GB dual uplink NIC
- Running KVM Hypervisor

Flavor	vCPUs	RAM	Storage	Per Node
m1.tiny	1	2	8	46
m1.small	2	4	20	23
m1.medium	6	16	60	7
m1.large	10	30	60	4
m1.xlarge	24	60	60	2
m1.xxlarge	44	120	60	1
s1.large**	10	30	120	4
s1.xlarge**	24	60	240	2
s1.xxlarge**	44	120	480	1

****Cannot be used for imaging**

- Short-term **ephemeral storage** comes as part of launched instance
- Long-term storage is XSEDE-allocated
 - Implemented on backend as OpenStack Volumes
 - Each user gets 10 volumes up to 500GB total storage
- Piloting **object storage** as well after recent update

The Jetstream Atmosphere web interface

The screenshot shows the Jetstream Atmosphere web interface dashboard. At the top, there is a navigation bar with the Jetstream logo, a 'Dashboard' menu item, and links for 'Projects', 'Images', 'Help', and 'Admin'. The user 'jfischer' is logged in. The main content area is divided into several sections:

- Getting Started:** Three cards with icons and text:
 - 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:** A section with a 'Need more?' button. It contains:
 - Allocation Source:** A horizontal bar chart showing the percentage of allocation used by five sources: TG-STA1100245 (5.06%), TG-ASC160018 (32.23%), TG-CDA160007 (23.15%), TG-TRA160003 (17.61%), and TG-TRA160027 (0%).
 - 10 Instances:** A donut chart showing the status of 10 instances, with a legend for 'active' (blue) and 'shutoff' (black).
 - 4 Volumes:** A donut chart showing the status of 4 volumes, with a legend for 'available' (blue).
- Provider Resources:** A horizontal bar chart showing CPU usage for two providers: Jetstream - Indiana University (9.09%) and Jetstream - TACC (0%).

At the bottom of the dashboard, there is a footer with the copyright notice '©2017 Jetstream-Cloud' and a 'Feedback & Support' button.

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



Using Jetstream VMs

- Managing Jetstream VMs:
 1. Jetstream Atmosphere web interface: <http://use.jetstream-cloud.org>
 2. 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
 1. Interactive web-interface with VNC (desktop) or SSH (CLI)
 2. Direct VNC/SSH to individual instances using 3rd party tools



funded by the National Science Foundation
Award #ACI-1445604



Jetstream for Education/Training

- Jetstream has been used in multiple Informatics Graduate Courses
 - Management, Access, and Use of Big and Complex Data
 - Topics in Informatics
- BlueWaters Workflow Workshop
- Multiple Research Data Alliance Workshops
- TACC Cloud Computing Summer Institute
- 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



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 Timeline: Overall Highlights

- As of *September 1, 2017*:
 - **322** active XSEDE projects covering **59** fields of science and **2000+** active users representing **189** institutions
 - Over **70 million CPU hours** allocated to XSEDE projects since June 2016
 - **9** active science gateways
 - **40** education/teaching allocations serving over **600** undergraduate and graduate students



funded by the National Science Foundation
Award #ACI-1445604



Jetstream Timeline: what comes next?

- **Soliciting** Research allocation requests plus Startup and Education allocations – including Science Gateways!
- **Adding services** as deemed useful/mature (heat, ceilometer, magnum, trove, manila, etc)
- Making Atmosphere-interface **enhancements**
- Working on **partnerships** with groups like HubZero and others to extend the value of Jetstream



funded by the National Science Foundation
Award #ACI-1445604



Requesting access to Jetstream

- You can request startup allocations **anytime**. (Startups are simple!)
- You can request allocations for educational use **anytime**.
- Next **submission period for large** allocations is 15 Sept 2017– 15 Oct 2017.
- 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.



funded by the National Science Foundation
Award #ACI-1445604



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>
- Training Videos / Virtual Workshops (**In development**)



funded by the National Science Foundation
Award #ACI-1445604



Jetstream Fun: Happy cluster / Angry Cluster



Partners

Construction



THE UNIVERSITY OF ARIZONA

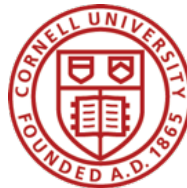


THE UNIVERSITY OF CHICAGO



JOHNS HOPKINS UNIVERSITY

Management & Operations



UNIVERSITY of ARKANSAS AT PINE BLUFF



UNC THE ODUM INSTITUTE



funded by the National Science Foundation
Award #ACI-1445604

vendors



QUESTIONS?

Jetstream help: help@jetstream-cloud.org

Direct to Brian Beck: bbeck@tacc.utexas.edu

Twitter: @jetstream_cloud

Facebook: <https://www.facebook.com/jetstreamcloud>

Jetstream Web UI: <http://use.jetstream-cloud.org>

XSEDE Portal: <http://portal.xsede.org>

Github Source: <https://github.com/jetstream-cloud>

- Jetstream is supported by NSF award 1445604 (Craig Stewart, IU, PI)
- XSEDE is supported by NSF award 1053575 (John Towns, UIUC, PI)