
RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES
rt.iu.edu

RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES

Jetstream Overview
A national research and education cloud

David Y. Hancock – Indiana University

Director for Advanced Cyberinfrastructure

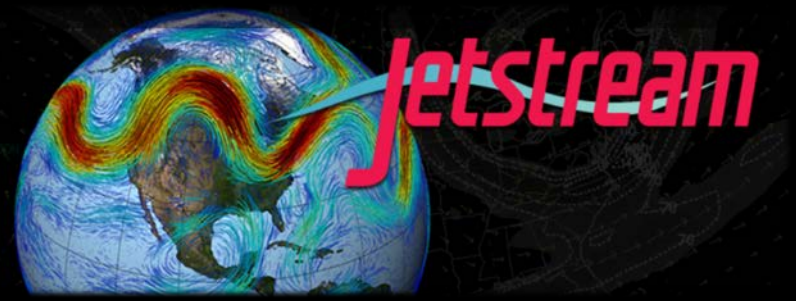
Jetstream Primary Investigator

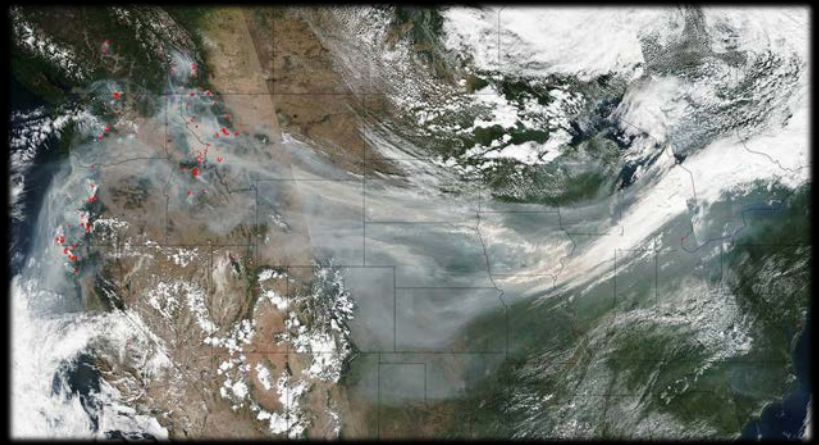
Affiliated with IU Pervasive Technology Institute

April 12, 2019 – EarthCube Council of Data Facilities

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

What is Jetstream and why does it exist?

- NSF's first production cloud facility
- Focus on ease-of-use, broad accessibility
- User-selectable library of preconfigured virtual machines
- Provides on-demand *interactive* computing and analysis or persistent services such as gateways
- Enables *configurable* environments; *programmable cyberinfrastructure*

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

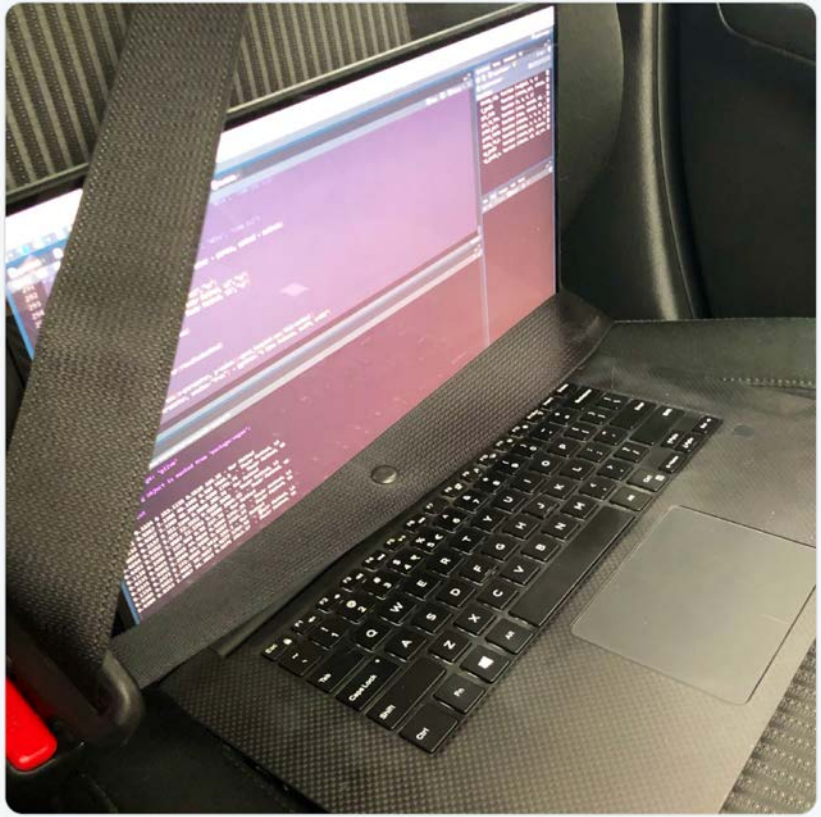
*XSEDE Cloud Survey Report - http://hdl.handle.net/2142/45766

“But I really don’t have research needs… I don’t need the national research cyberinfrastructure.”

– multiple researchers at small colleges and universities

Sarah Romanes @sarah_romanes · Mar 17

When you have to drive to work - but R has also only completed 10hrs of 24hrs worth of simulations. Can I go in the T2 lane with this thing? 😂 #rstats #sydneytraffic



40 74 1.1K

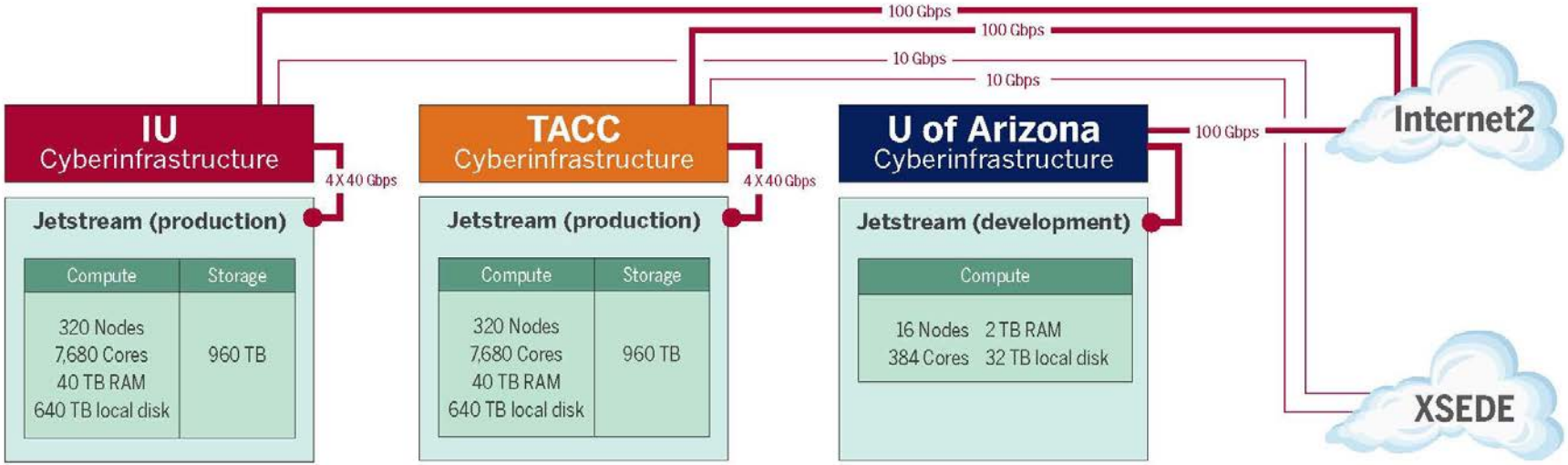
Who uses Jetstream?

- The researcher needing a handful of cores (1 to 44/vCPU)
- Software creators and researchers needing to create their own VMs 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...)

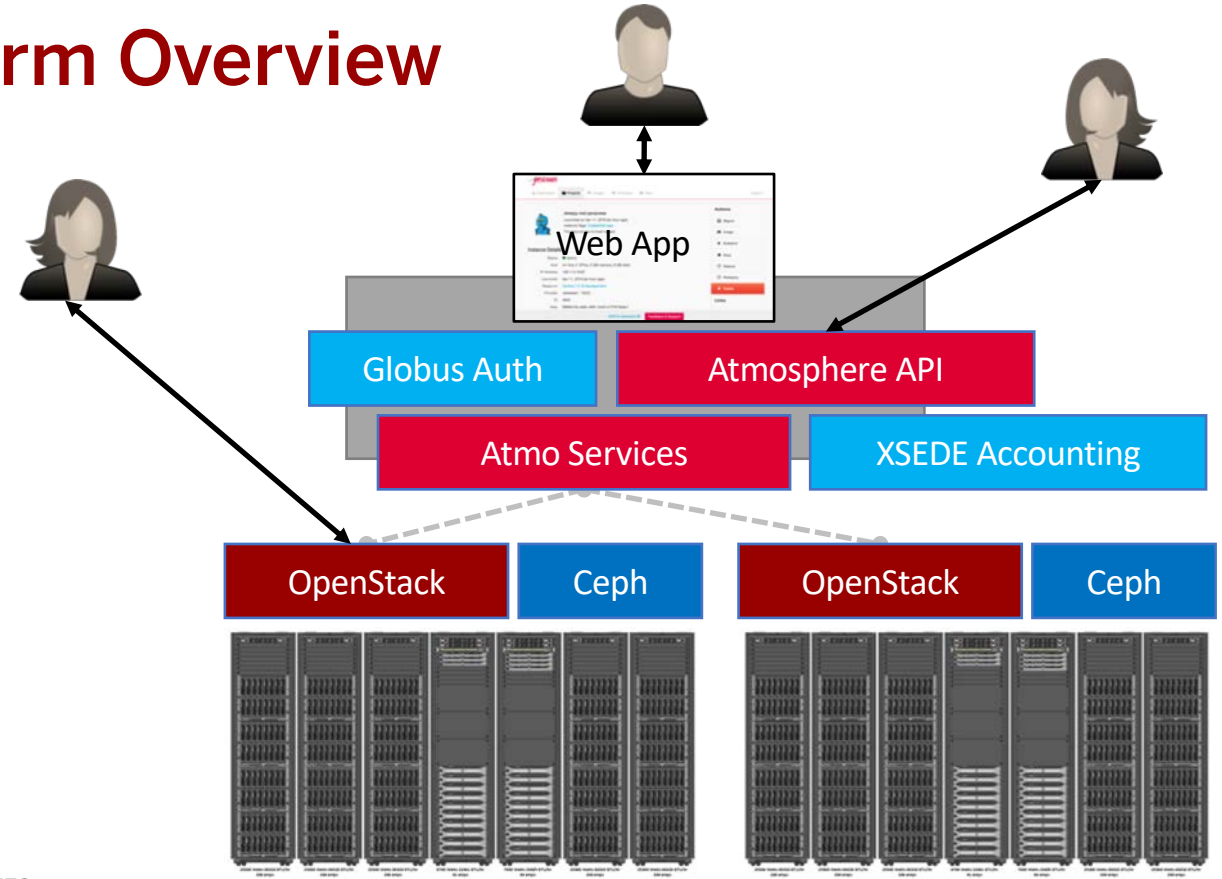
Jetstream System Overview



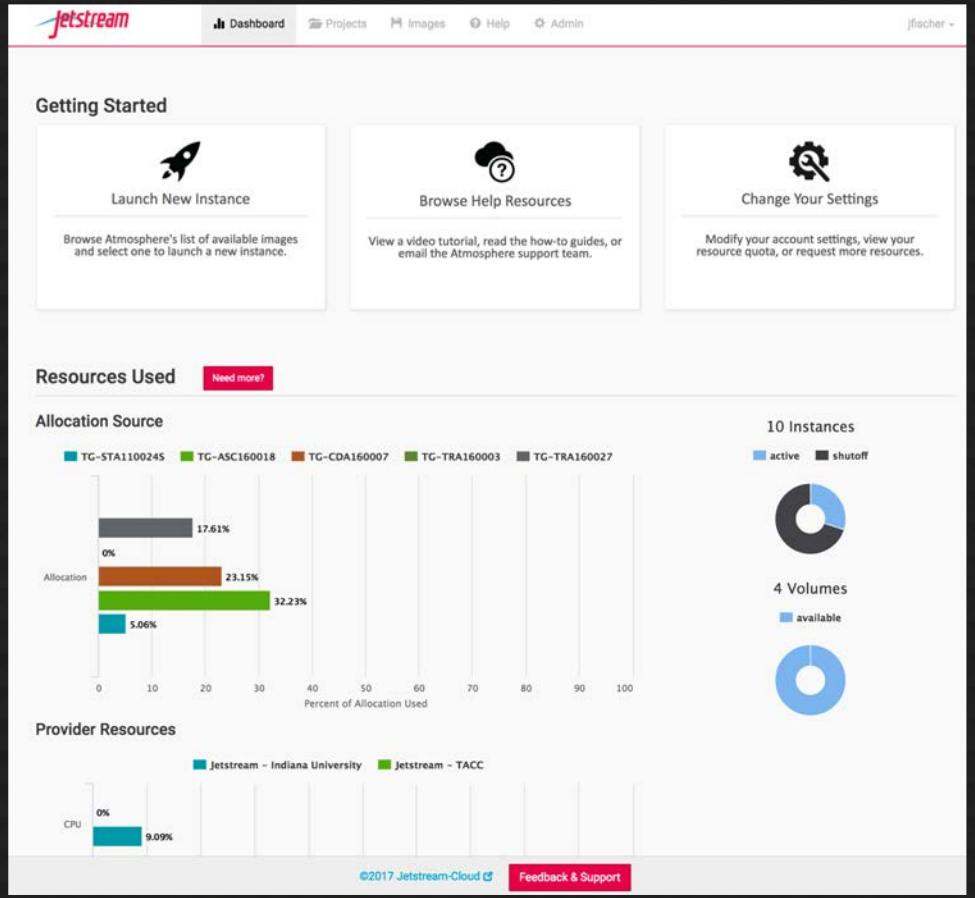
Production cloud hardware (per site)

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

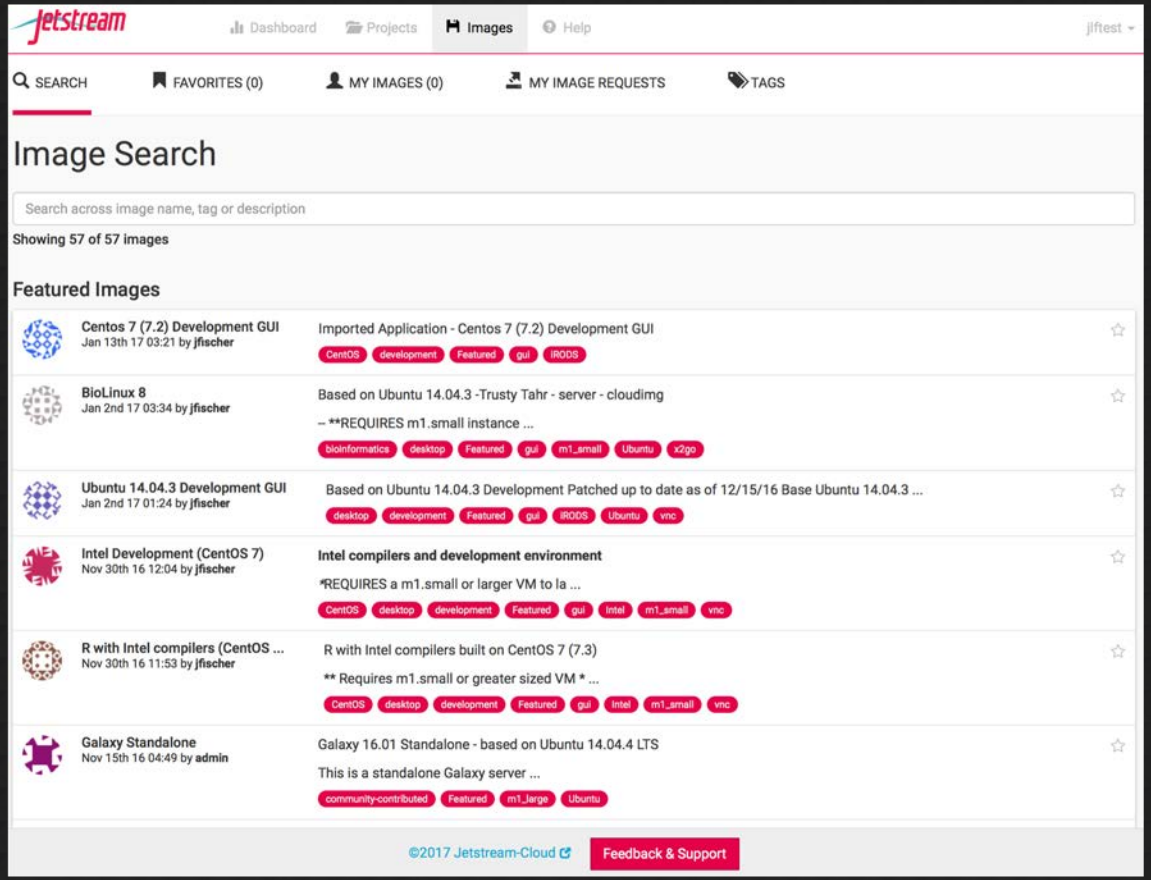
Platform Overview



Jetstream Atmosphere web interface



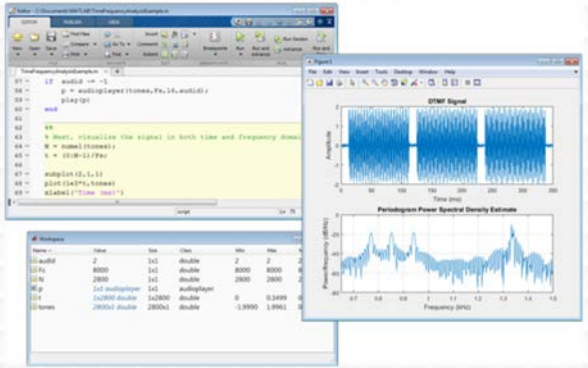
Jetstream Atmosphere web interface



Discipline or area of interest	#of Jetstream allocations	SUs allocated on Jetstream	SU increase/Decrease on Jetstream over previous year	% of SUs allocated on Jetstream	% of all SUs allocated on other XSEDE-supported systems
Behavioral Sciences	6	3,465,516	100%	4.24%	0.61%
Biological Sciences	89	15,041,928	72%	**18.40%**	3.59%
Biophysics	86	3,627,026	(44%)	4.44%	**13.56%**
Computer Science	72	6,883,269	32%	**8.42%**	2.98%
Earth Sciences	37	5,476,250	37%	6.70%	4.60%
Education and Training	128	16,599,512	3%	**20.31%**	4.66%
Engineering	1 3	520,690	71%	0.64%	1.75%
Materials Science	6	1,035,508	100%	1.27%	**13.89%**
Mathematics	13	688,505	150%	0.84%	0.90%
Molecular Science/Biochemistry	21	4,254,643	(10%)	5.20%	5.83%
Neuroscience	19	4,708,180	328%	5.76%	1.98%
Physics	10	2,440,581	(16%)	2.99%	**8.65%**
Social Sciences and Humanities	28	2,409, 633	192%	2.95%	0.81%

Jetstream for engineering researchers (and others)

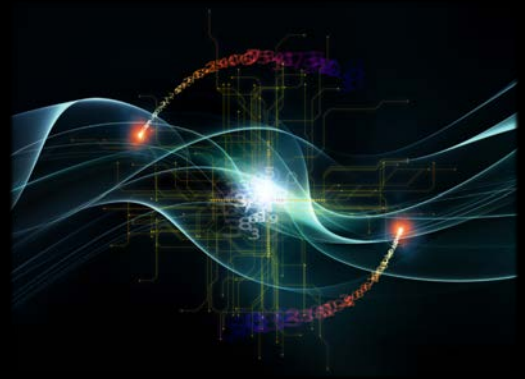
- Matlab and SimuLink and additional toolkits are installed on Jetstream
- You do NOT need to have a local license to use MATLAB on Jetstream
- If you are a researcher, and MATLAB or SimuLink… you're ready to go!
- If you are an engineering researcher, and you need other tools… let us know!



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



Jetstream for Education

Jetstream has been used in multiple graduate and undergraduate courses

- Management, Access, and Use of Big and Complex Data
- Multiple informatics and general bioinformatics courses
- Business Intelligence (big data and analysis)
- Research Topics in Music
- Multiple genetics and sequencing courses
- Multiple information security and assurance courses

Research Data Alliance workshops, Galaxy workshops, data analysis in finance using R, security and intrusion detection, and principles in cloud computing…

Another Use Case: Galaxy riding Jetstream

Galaxy, a platform for biomedical research, focused on accessibility, transparency and reproducibility

- 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

Use Jetstream as a bursting platform

- From Galaxy Main, send jobs to a Slurm cluster running on Jetstream
- Run Galaxy Interactive Environments (Jupyter/RStudio containers) via 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

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

- ChemCompute Gateway

Providing a computational chemistry gateway for educational use

Coming gateways: The Neuroscience Gateway, UltraScan III, and others

Jetstream usage highlights – 1 March 2019

- 415 active XSEDE projects covering 71 fields of science and over 2250 **active users** representing **205 institutions**
- **80%** of Jetstream users have **not used any other XSEDE system**
- >193M CPU hours allocated to XSEDE projects since June 2016
- 19 active science gateways
- 52 education/teaching allocations serving over 975 students
- 1288 mean active VMs in previous qtr, 1530 peak active VM count
- **Highest** user satisfaction in most recent XSEDE survey

Jetstream Timeline…what comes next?

- Completed our second year of operations with extension to November 2020
- 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
  - http://wiki.jetstream-cloud.org/Jetstream+Trial+Access+Allocation
- You can request startup allocations anytime. (Startups are simple!)
- http://wiki.jetstream-cloud.org/Jetstream+Allocations
- You can request allocations for educational use anytime.
- Next submission period for large allocations is 15 Mar 2019 – 15 Apr 2019
- Research allocation: Project desc (≤10 pages) and Scaling doc (≤5 pages)
  - We can help!

Expanding the reach: Jetstream REU Program

NSF Supplement for undergraduates

- 4 students participated in 2017
- 6 students participated in 2018
- Hosting 6 students in 2019

- 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/

Jetstream partners

Jetstream
http://jetstream-cloud.org/

NSF
funded by the National Science Foundation
Award #ACI-1445604