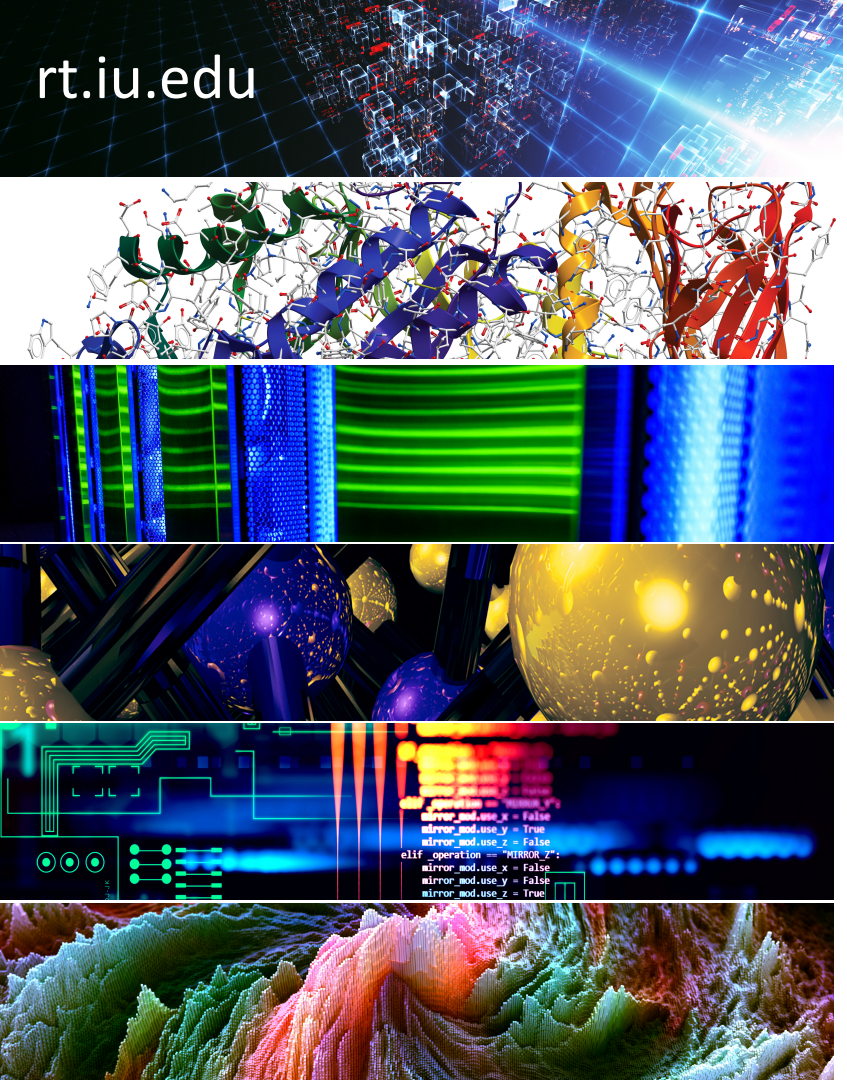


RESEARCH TECHNOLOGIES
PERVASIVE TECHNOLOGY INSTITUTE

RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES

Jetstream Overview: A national research and education cloud

UC Professional Development Workshop

January 29, 2021 – Bloomington, IN (Webinar)

Sanjana Sudarshan, PhD – ssudarsh@iu.edu

Senior Technical Advisor, Jetstream Cloud

UITS Research Technologies

Indiana University

Sudarshan, S. (2021). Jetstream: A national research and education cloud. 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 (Frontera awarded late 2018)
- 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 (extended to Jetstream2)

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

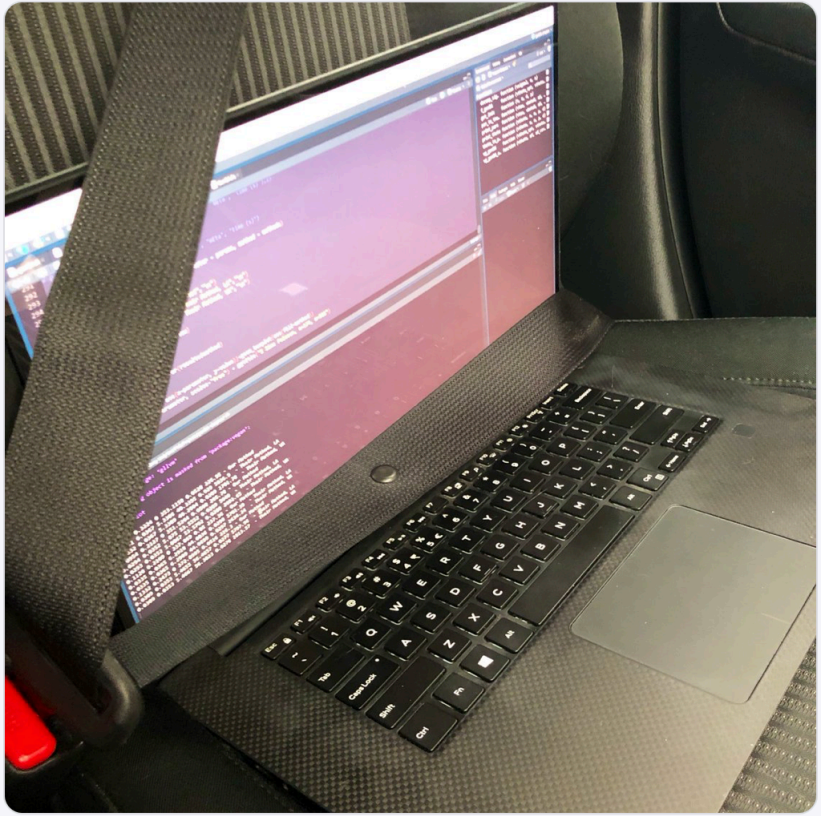
*XSEDE Cloud Survey Report - http://hdl.handle.net/2142/45766

“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

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

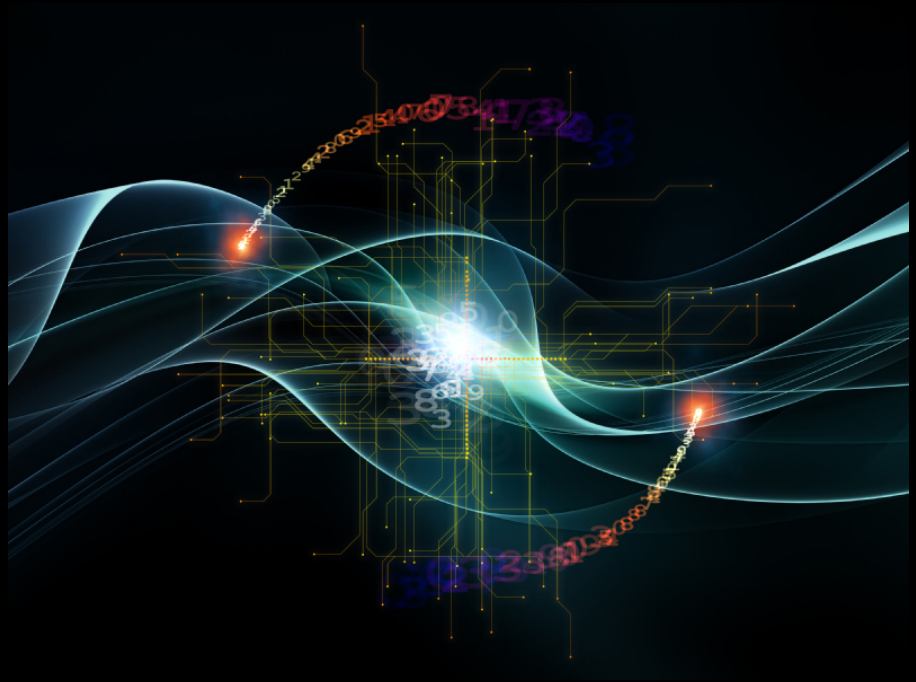
- Provides on-demand *interactive* computing and analysis or persistent services such as gateways
- Enables *configurable* environments; *programmable cyberinfrastructure*
- Reproducibility

Who uses Jetstream?

- The researcher needing a handful of cores (1 to 44 vCPUs)
- 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!)
- It isn't Amazon, Azure, or GCE (similar, but…)



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**

** Here there be dragons…

Jetstream and the 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 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…**

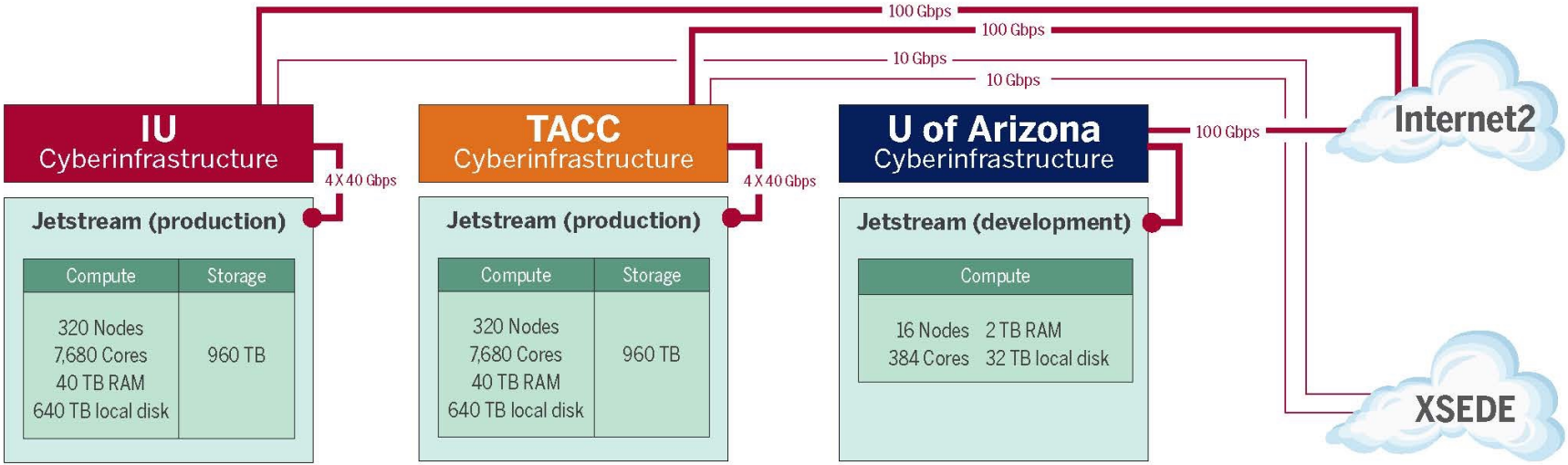
Thinking about VMs…

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.

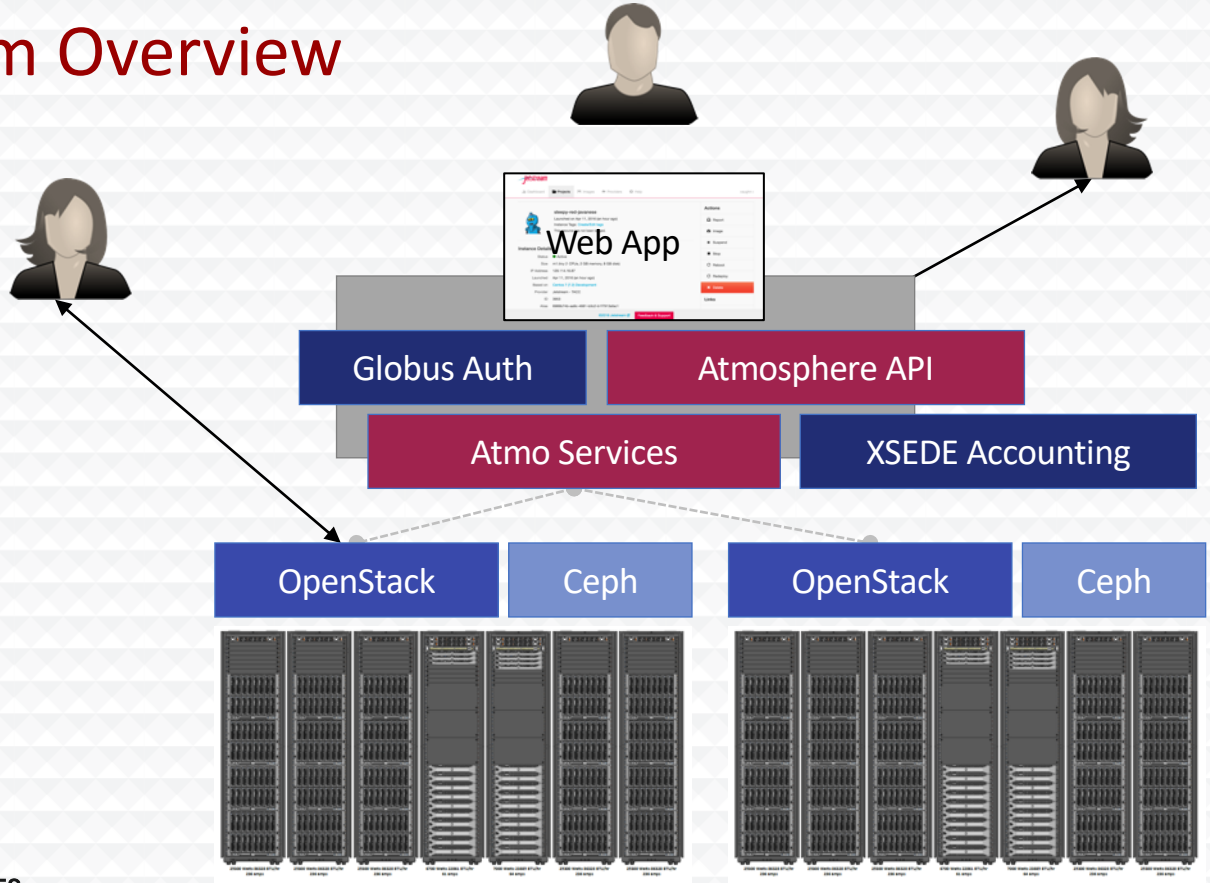
-- Mike Lowe (Jetstream architect)

**some caveats for gateways…

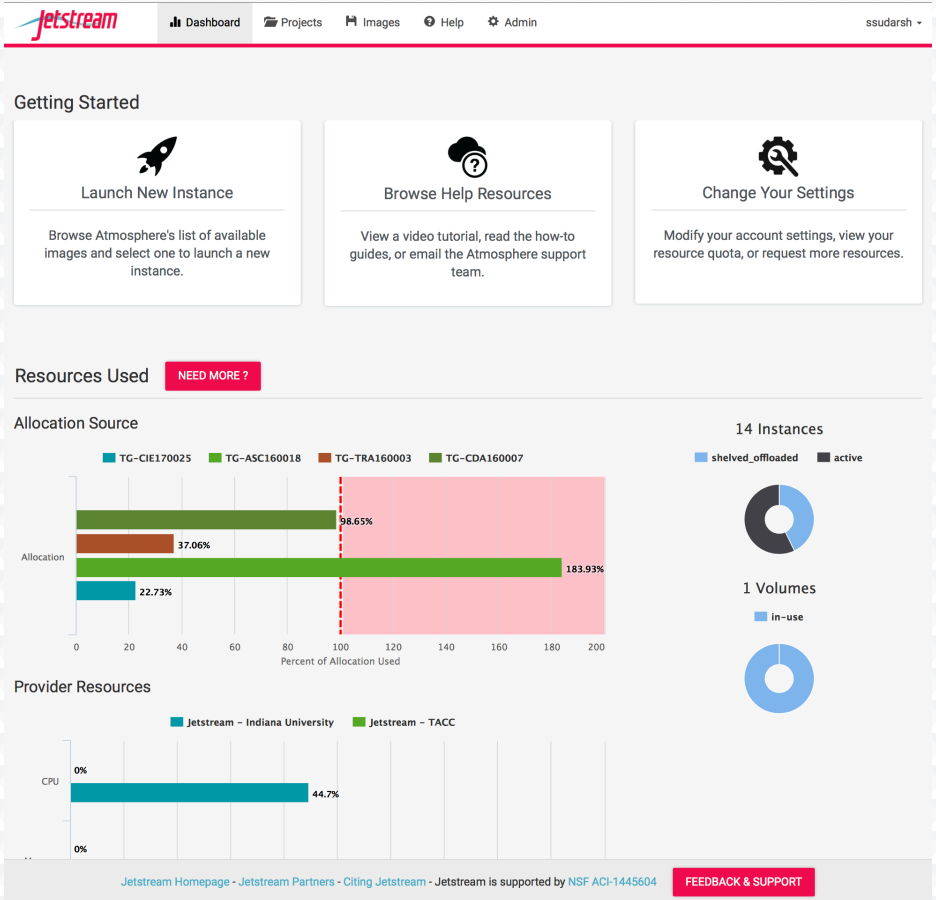
Jetstream System Overview



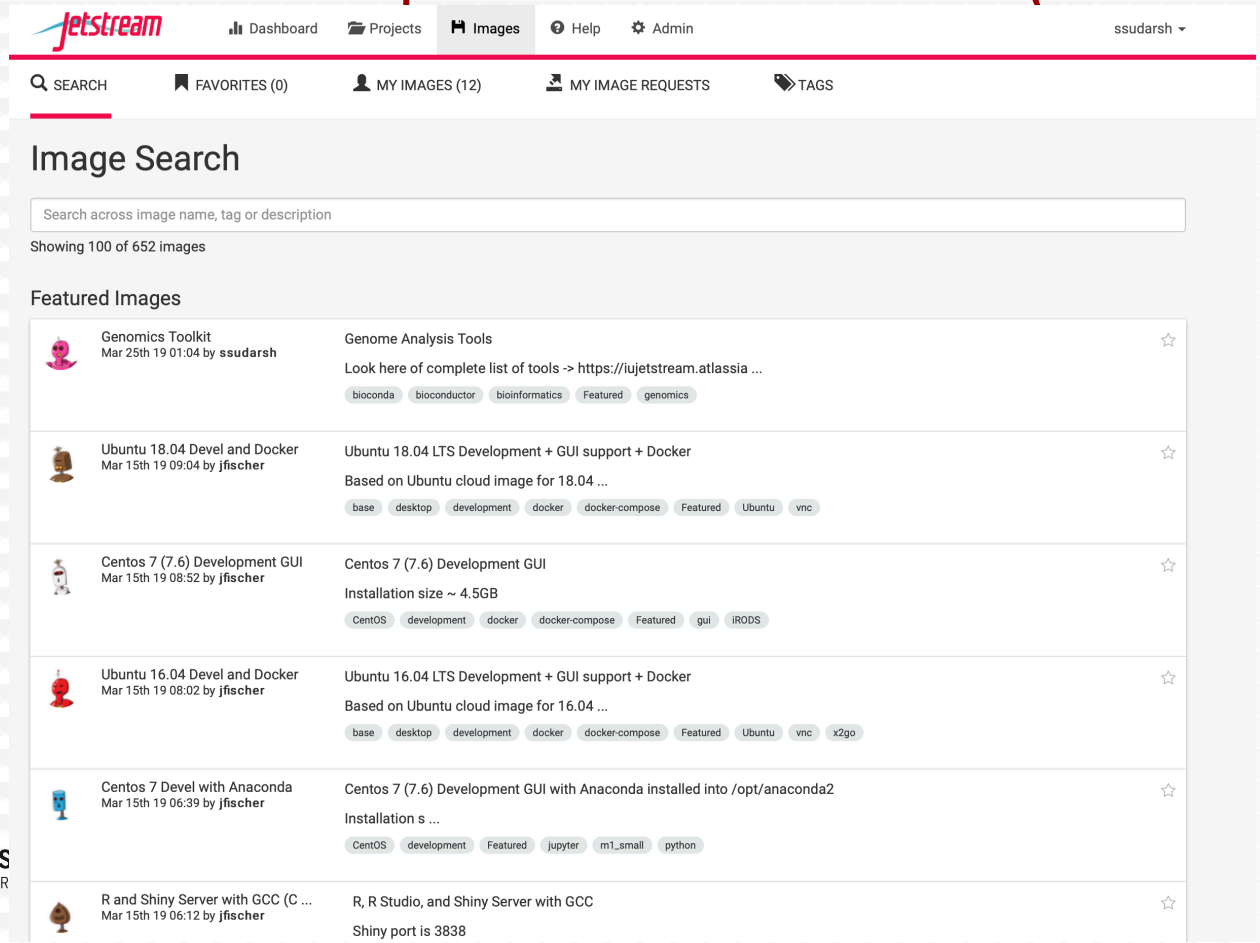
Platform Overview



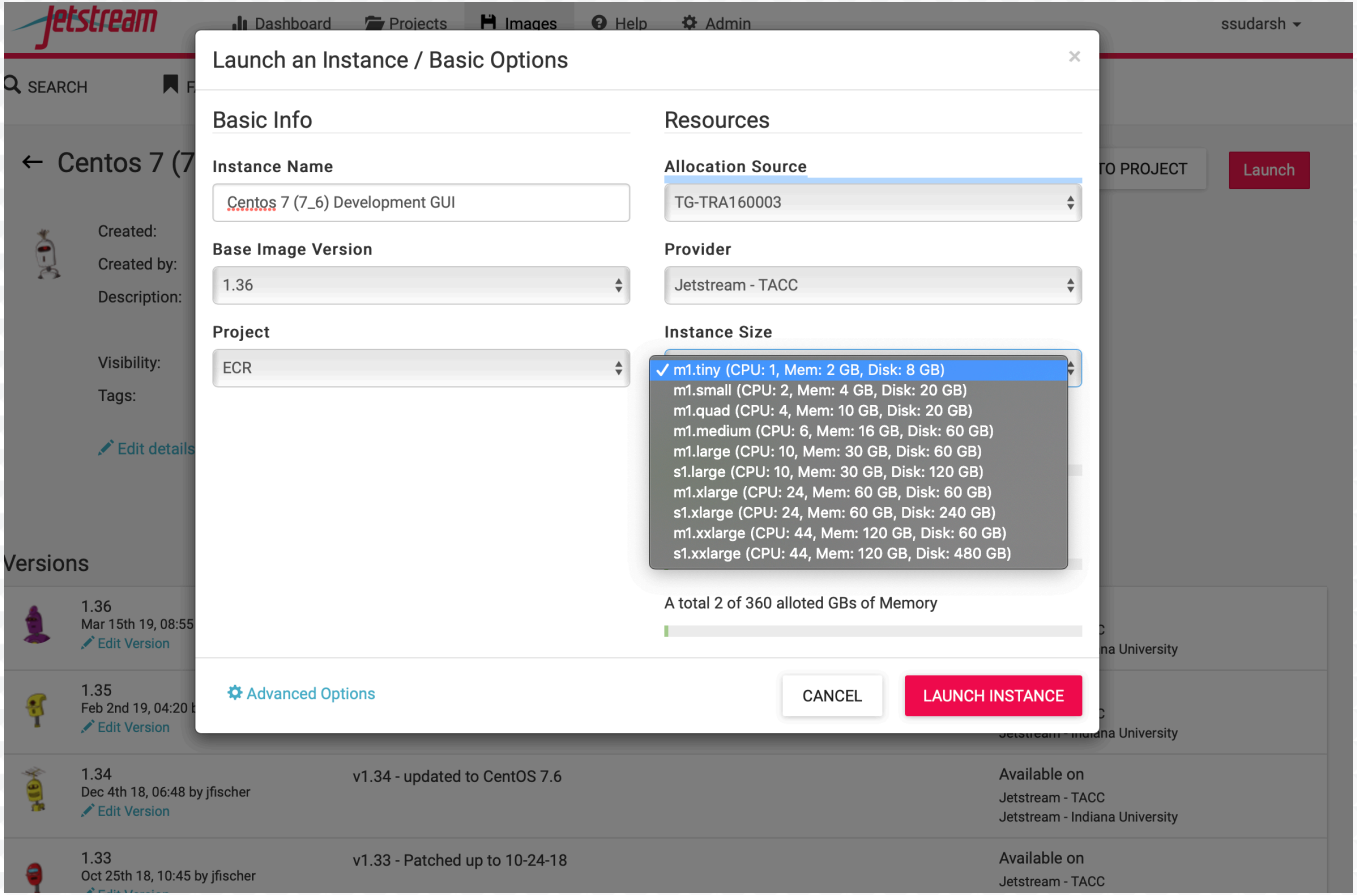
The Jetstream Atmosphere web interface



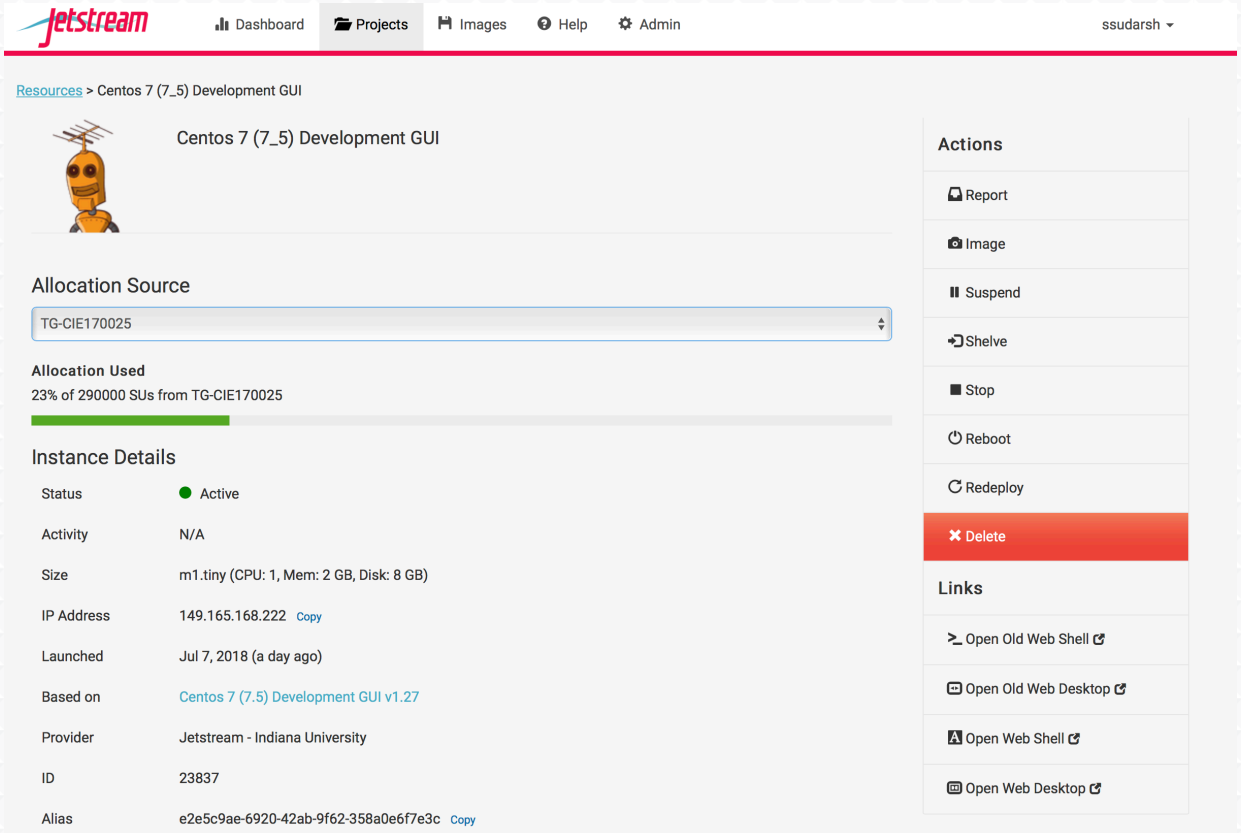
The Jetstream Atmosphere web interface (contd..)



Look! It’s more Jetstream web interface!



Even more Jetstream web interface…



Jetstream Web Shell

```
Welcome to Ubuntu 18.04.2 LTS (GNU/Linux 4.15.0-46-generic x86_64)

  System information as of Thu Mar 28 10:11:11 EDT 2019

  System load:  0.0               Processes:           172
  Usage of /:   52.1% of 7.58GB   Users logged in:     0
  Memory usage: 31%               IP address for ens3: 172.23.17.5
  Swap usage:   0%

 * Read about Ubuntu updates for L1 Terminal Fault Vulnerabilities
   (L1TF).

   - https://ubu.one/L1TF

 * Check out 6 great IDEs now available on Ubuntu. There may even be
   something worthwhile there for those crazy EMACS fans ;)

   - https://bit.ly/6-cool-IDEs

  Get cloud support with Ubuntu Advantage Cloud Guest:
    http://www.ubuntu.com/business/services/cloud

 * Canonical Livepatch is available for installation.
   - Reduce system reboots and improve kernel security. Activate at:
     https://ubuntu.com/livepatch

1 package can be updated.
1 update is a security update.


Welcome to
    _   _                                  _
   / \ | |_ _ __ ___   ___  ___ _ __ | |__   ___ _ __ ___
  / _ \| __| '_ ` _ \ / _ \/ __| '_ \| '_ \ / _ \ '__/ _ \
 / ___ \ |_| | | | | | (_) \__ \ |_) | | | |  __/ | |  __/
/_/   \_\__|_| |_| |_|\___/|___/ .__/|_| |_|\___|_|  \___|
                               |_|
ssudarsh@js-168-140:~$
```

Jetstream Web Desktop

Hardware and Instance “Flavors”

Flavor	vCPUs	RAM	Storage	Per Node
tiny	1	2	8	46
small	2	4	20	23
quad	4	10	20	11
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

Newly available on Jetstream:

- 6 Dell C4140 nodes with 4 NVIDIA 16GB V100 GPUs each
- GPUs are portioned into ½, ¼, and whole GPU and assigned to a vm
- CUDA enabled codes run accelerated and unmodified
- Card memory is divided and you have access to all the CUDA cores during your timeslices

Caveats

- Not accessible from Atmosphere
- Must use specialized drivers from us that match underlying hypervisor
- Live migration restrictions and limited numbers mean more interruptions than the rest of Jetstream

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

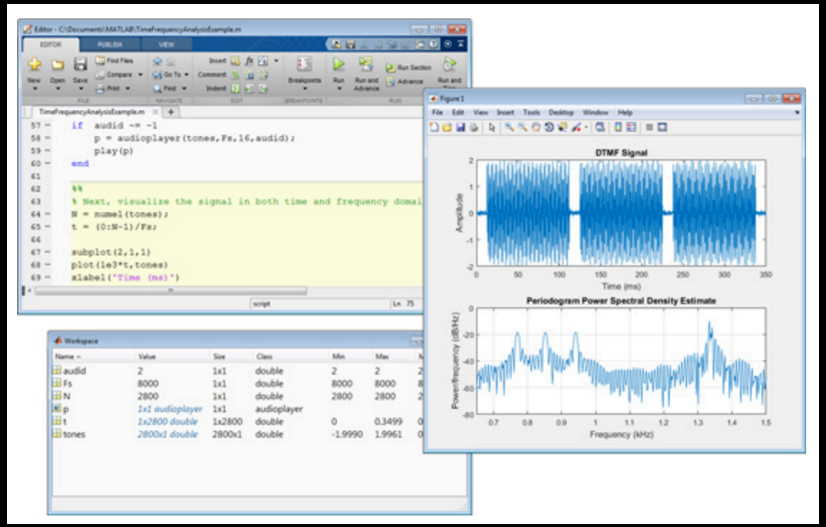
Primary methods of logging into Jetstream VMs to work

- Interactive user access via web interface with VNC/SSH
- Direct VNC/SSH to individual instances

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%	18.40%	3.59%
Biophysics	86	3,627,026	44.15%	4.44%	13.56%
Computer Science	72	6,883,269	32.28%	8.42%	2.98%
Earth Sciences	37	5,476,250	37.06%	6.70%	4.60%
Education and Training	128	16,599,512	2.62%	20.31%	4.66%
Engineering	1 3	520,690	71.21%	0.64%	1.75%
Materials Science	6	1,035,508	100%	1.27%	13.89%
Mathematics	13	688,505	150.37%	0.84%	0.90%
Molecular Science/Biochemistry	21	4,254,643	10.15%	5.20%	5.83%
Neuroscience	19	4,708,180	327.87%	5.76%	1.98%
Physics	10	2,440,581	15.58%	2.99%	8.65%
Social Sciences and Humanities	28	2,409, 633	192.27%	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 that uses 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...

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

- **Simulations of Nanoscale Biomolecular Systems -** Aleksei Aksimentiev, University of Illinois Champaign-Urbana
- **The Neuroscience Gateway** - Amitava Majumdar, University of California, San Diego
- **Parallelizing Development of Immunomics and Genomics Tools** - Ramy Arnaout, Beth Israel Deaconess Medical Center
- **Atmospheric Science in the Cloud: Enabling Data-Proximate Science** – Mohan Ramamurthy, UNIDATA (University Corporation for Atmospheric Research)
- **Science and Engineering Applications Grid (SEAGrid): A Gateway for Simulation of Molecular and Material Structures and Dynamics –** Sudhakar Pamidighantam, Indiana University

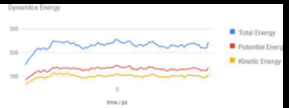
And others!

NEUROSCIENCE GATEWAY
A Portal for Computational Neuroscience

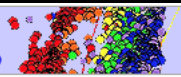
SEAGRID

TINKER

TINKER is a molecular dynamics package from the Jay Ponder Lab.



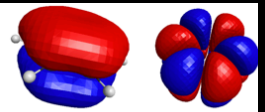
IRIS IEB Help



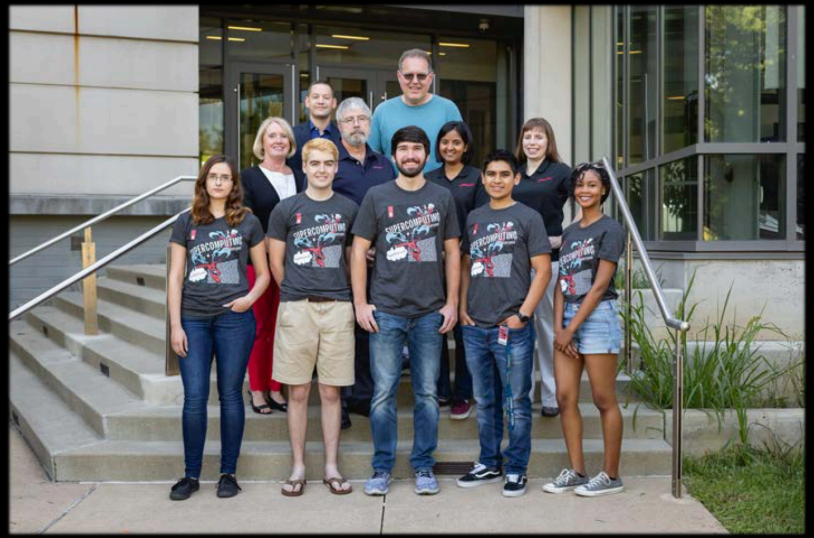
IRIS Earthquake Browser

GAMESS

(The General Atomic and Molecular Electronic Structure System) is a quantum chemistry package.



Jetstream REU Program



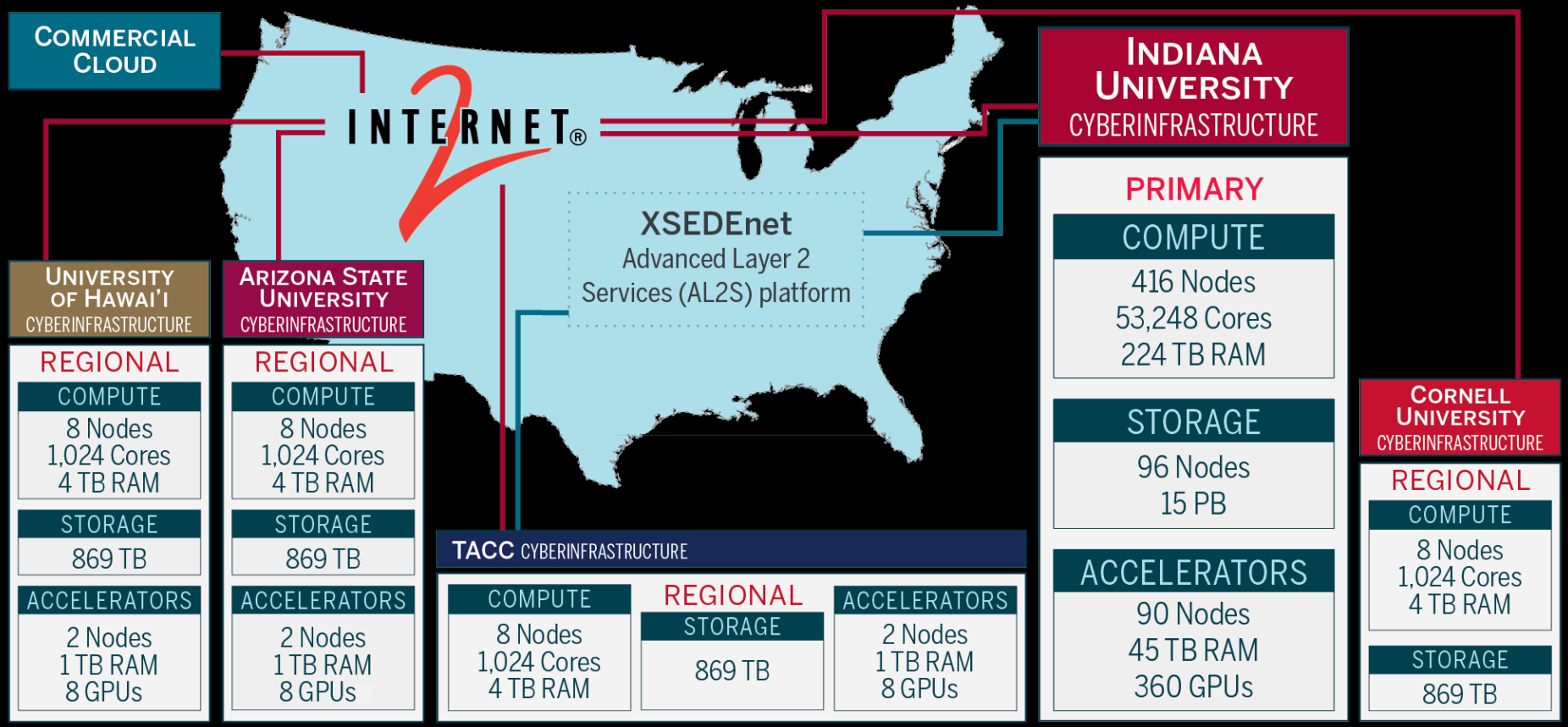
NSF Supplement for undergraduates

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

- REU student videos on YouTube
  https://www.youtube.com/user/IUPTI

Jetstream usage highlights – 1 Jan 2021

- 395 active XSEDE projects covering 56 fields of science and over 2400 **active users** representing **405 institutions**
- **80%** of Jetstream users have **not used any other XSEDE system**
- >373M CPU hours allocated to XSEDE projects since June 2016
- 49 active science gateways
- 46 education/teaching allocations serving over 850 students
- 1189 mean active VMs in previous qtr, 1632 peak active VM count
- **Highest** user satisfaction in most recent XSEDE survey



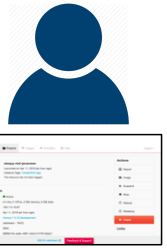
RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES

Jetstream2

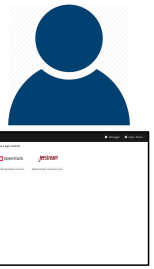
Platform Overview

Horizon

OpenStack CLI



Atmosphere



3rd Party Clients

Science Gateways

Authentication Service

XSEDE Accounting

OpenStack

Ceph

OpenStack

Jetstream2 Core System

Timeline

- Jetstream now in 5th year of operations
- Jetstream extension requested through November 2021
- Jetstream2
  - Early operations planned for August 2021
  - Production operations by October 2021

Flickr user Oiluj Samall Zeid - Lejos de Yulín

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 2020 – 15 Apr 2020
- Research allocation: Project desc (<10 pages) and Scaling doc (<5 pages)

We can help!

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

INDIANA UNIVERSITY
PERVASIVE TECHNOLOGY INSTITUTE

JOHNS HOPKINS
UNIVERSITY

THE UNIVERSITY OF ARIZONA

THE UNIVERSITY OF CHICAGO

NSIDC
National Snow and Ice Data Center

JSU 1877 JACKSON STATE UNIVERSITY

UNC
THE ODUM INSTITUTE

Jetstream2 Partners

http://jetstream-cloud.org/
National Science Foundation
Award #ACI-2005506