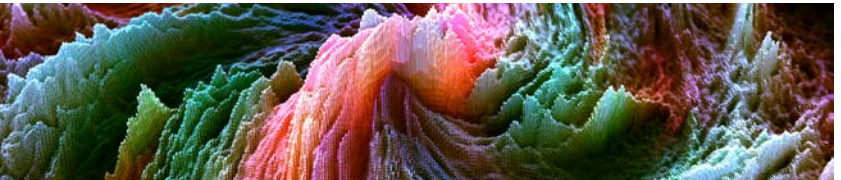
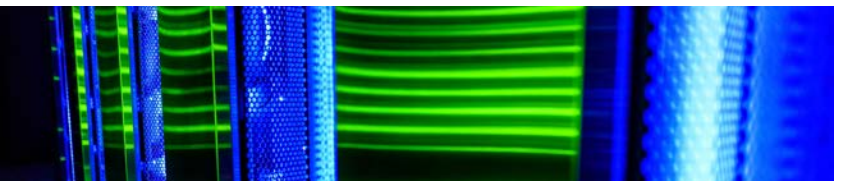
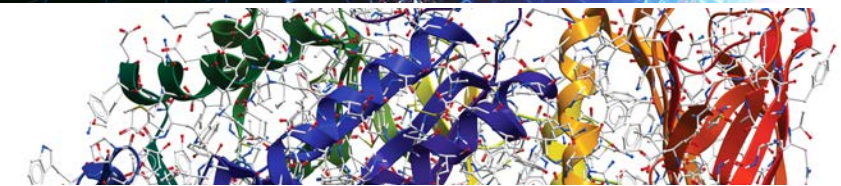


rt.iu.edu



RESEARCH TECHNOLOGIES
PERVASIVE TECHNOLOGY INSTITUTE



RESEARCH TECHNOLOGIES

UNIVERSITY INFORMATION TECHNOLOGY SERVICES

Jetstream Overview: A national research and education cloud

**Consortium for Computing Sciences in Colleges –
Northwestern Conference**

October 3, 2020 – Bloomington, IN (Webinar)

Sanjana Sudarshan, PhD – ssudarsh@iu.edu

Senior Technical Advisor, Jetstream Cloud

UITS Research Technologies

Indiana University

Sudarshan, S. (2020). 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>



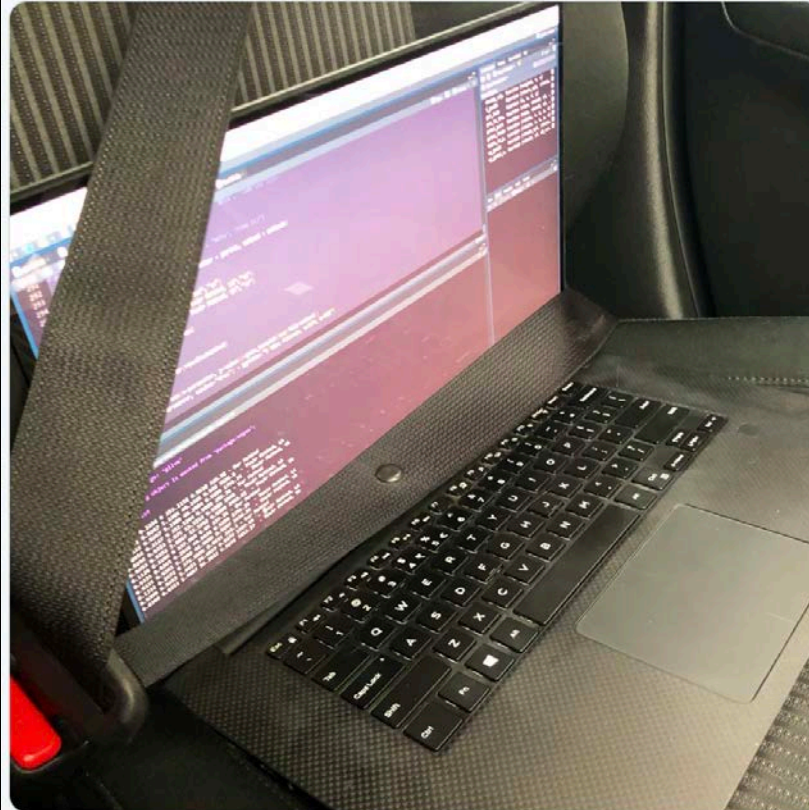
“But I really don’t have research needs...I don’t need the national research cyberinfrastructure.”

– multiple researchers at small colleges and universities



Sarah Romanes @sarah_romanesh · 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



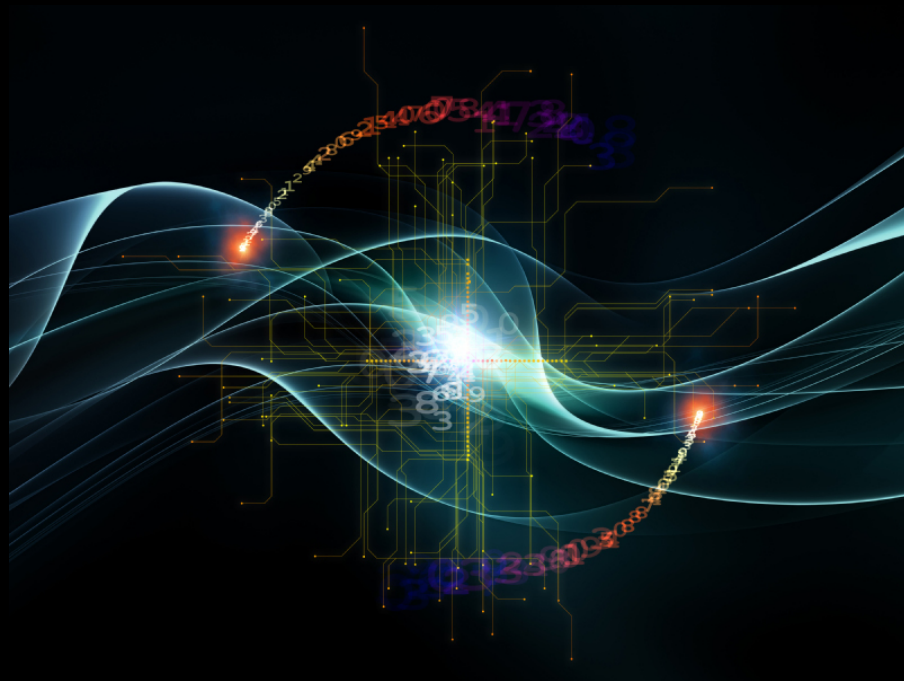
Who uses Jetstream?

- 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 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...



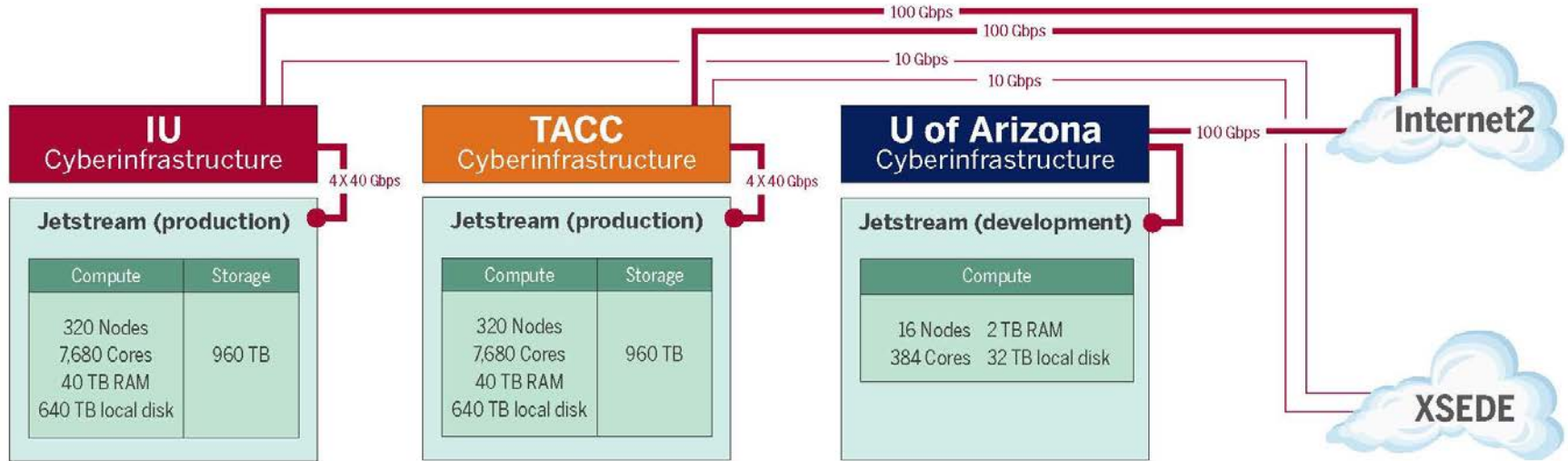
Cows, not pets: pets take great amount of care, feeding, and you name them; cows you intend to have high turnover and you give them numbers.

-- Mike Lowe (Jetstream architect)

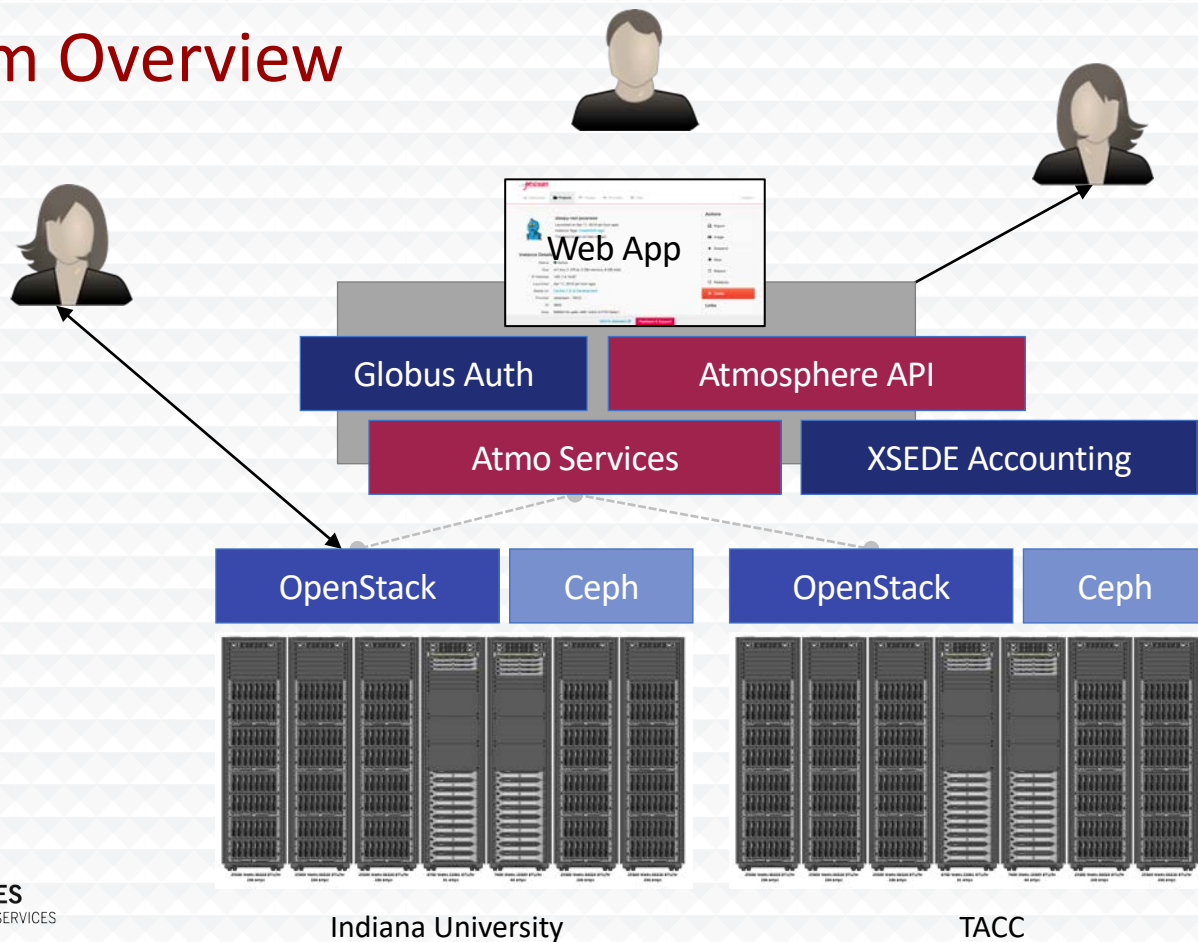
**some caveats for gateways...



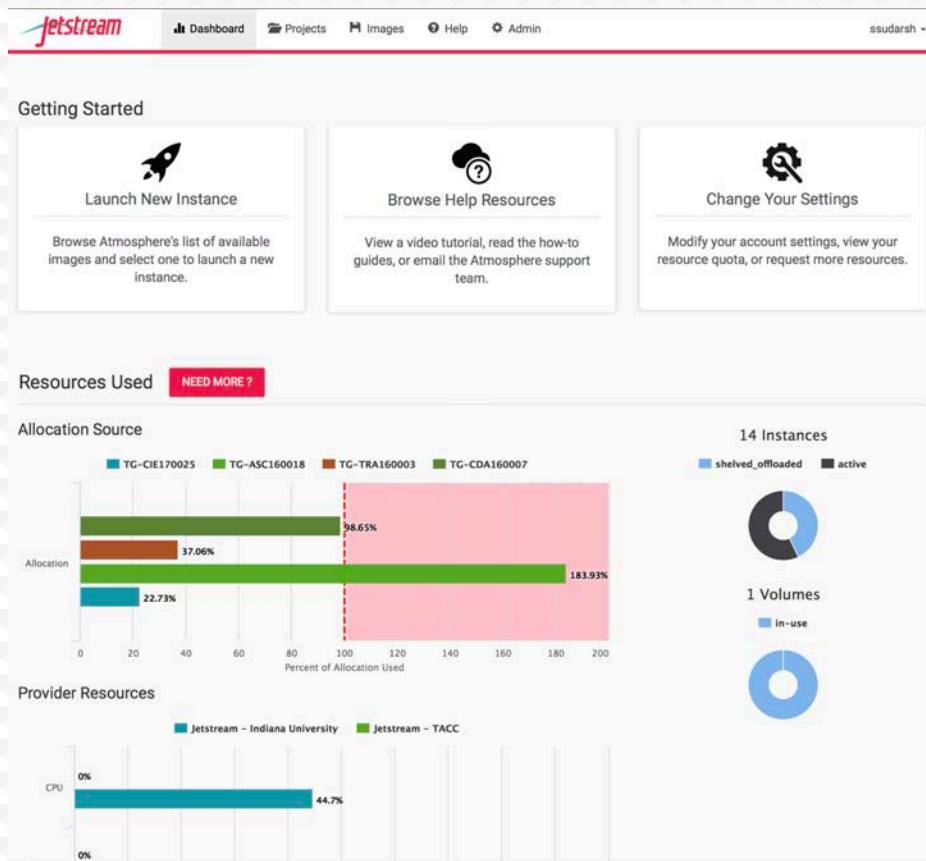
Jetstream System Overview



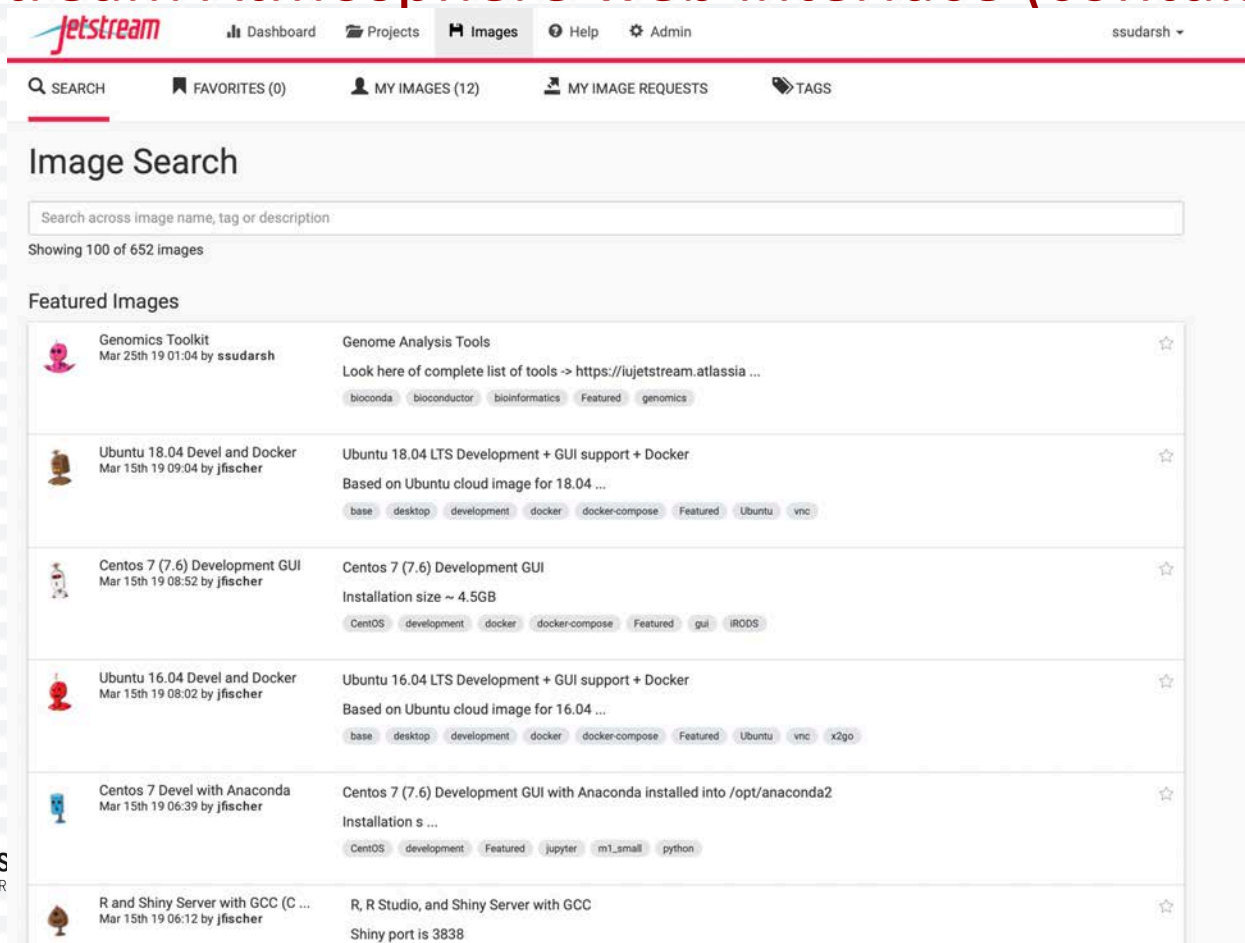
Platform Overview



The Jetstream Atmosphere web interface



The Jetstream Atmosphere web interface (contd..)















The screenshot displays the Jetstream Atmosphere web interface. At the top, there is a navigation bar with the Jetstream logo, a dashboard icon, and menu items for Dashboard, Projects, Images, Help, and Admin. The user 'ssudarsh' is logged in. Below the navigation bar is a search bar and several filters: SEARCH, FAVORITES (0), MY IMAGES (12), MY IMAGE REQUESTS, and TAGS.

Image Search

Search across image name, tag or description

Showing 100 of 652 images

Featured Images

 Genomics Toolkit Mar 25th 19 01:04 by ssudarsh	Genome Analysis Tools Look here of complete list of tools -> https://iujetstream.atlassa... bioconda bioconductor bioinformatics Featured genomics	
 Ubuntu 18.04 Devel and Docker Mar 15th 19 09:04 by jfischer	Ubuntu 18.04 LTS Development + GUI support + Docker Based on Ubuntu cloud image for 18.04 ... base desktop development docker docker-compose Featured Ubuntu vnc	
 Centos 7 (7.6) Development GUI Mar 15th 19 08:52 by jfischer	Centos 7 (7.6) Development GUI Installation size ~ 4.5GB CentOS development docker docker-compose Featured gui iRODS	
 Ubuntu 16.04 Devel and Docker Mar 15th 19 08:02 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	
 Centos 7 Devel with Anaconda Mar 15th 19 06:39 by jfischer	Centos 7 (7.6) Development GUI with Anaconda installed into /opt/anaconda2 Installation s ... CentOS development Featured jupyter m1_small python	
 R and Shiny Server with GCC (C ... Mar 15th 19 06:12 by jfischer	R, R Studio, and Shiny Server with GCC Shiny port is 3838	



Look! It's more Jetstream web interface!

Launch an Instance / Basic Options

Basic Info

Instance Name:

Base Image Version:

Project:

[Advanced Options](#)

Resources

Allocation Source:

Provider:

Instance Size

- m1.tiny (CPU: 1, Mem: 2 GB, Disk: 8 GB)
- m1.small (CPU: 2, Mem: 4 GB, Disk: 20 GB)
- m1.quad (CPU: 4, Mem: 10 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 of 2 of 360 allotted GBs of Memory

Centos 7 (7.6) Development GUI

Created:

Created by:

Description:

Visibility:

Tags:

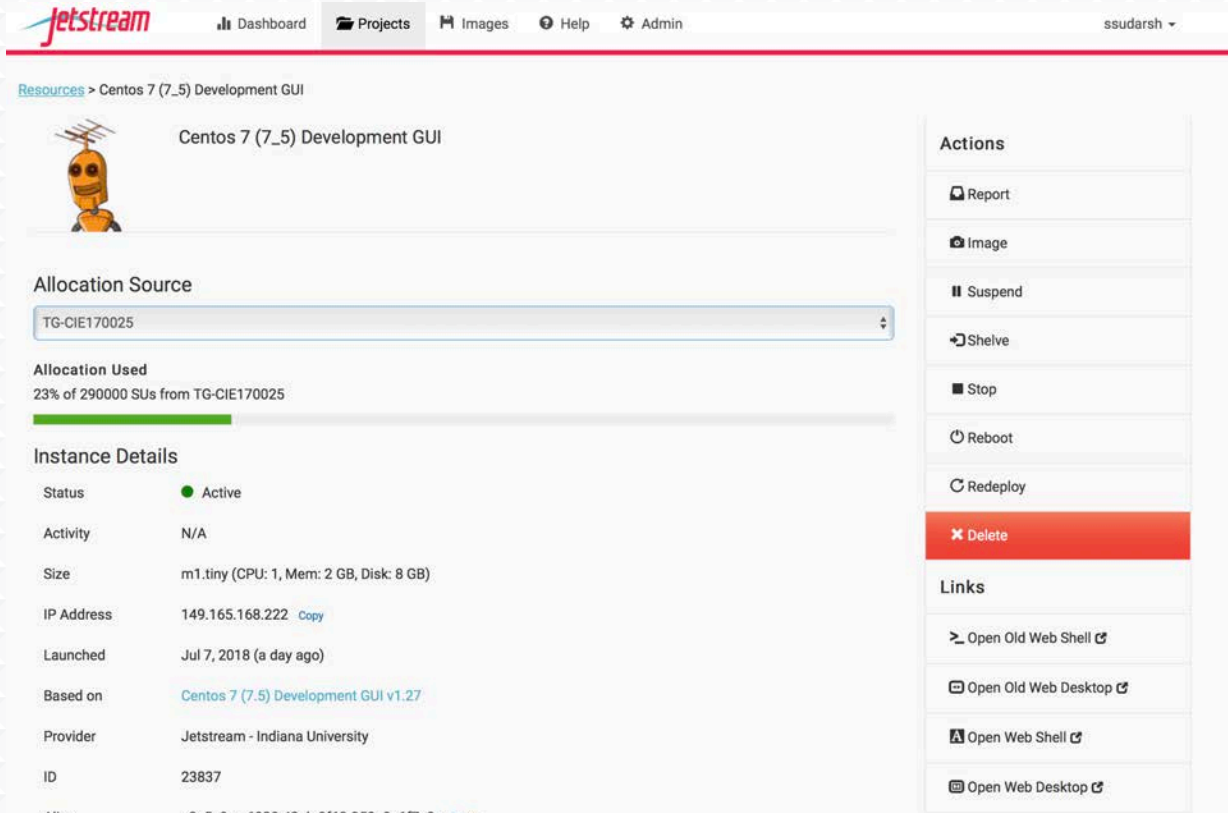
[Edit details](#)

Versions

	1.36 Mar 15th 19, 08:55 Edit Version			
	1.35 Feb 2nd 19, 04:20 Edit Version			
	1.34 Dec 4th 18, 06:48 by jfscher Edit Version	v1.34 - updated to CentOS 7.6		Available on Jetstream - TACC Jetstream - Indiana University
	1.33 Oct 25th 18, 10:45 by jfscher Edit Version	v1.33 - Patched up to 10-24-18		Available on Jetstream - TACC



Even more Jetstream web interface...



The screenshot displays the Jetstream web interface. At the top, there is a navigation bar with the Jetstream logo and menu items: Dashboard, Projects, Images, Help, and Admin. The user's name, ssudarsh, is visible in the top right corner. Below the navigation bar, the breadcrumb trail reads "Resources > Centos 7 (7_5) Development GUI".

The main content area features a header for the instance "Centos 7 (7_5) Development GUI" with a small robot icon. Below this, the "Allocation Source" is shown as a dropdown menu with "TG-CIE170025" selected. The "Allocation Used" section indicates that 23% of 290,000 SUs are used from TG-CIE170025, accompanied by a green progress bar.

The "Instance Details" section provides the following information:

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

On the right side, there is an "Actions" panel with the following options: Report, Image, Suspend, Shelve, Stop, Reboot, Redeploy, and Delete (highlighted in red). Below the actions is a "Links" panel with the following options: Open Old Web Shell, Open Old Web Desktop, Open Web Shell, and Open Web Desktop.



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:~$ █
```



RESEARCH TECHNOLOGIES

UNIVERSITY INFORMATION TECHNOLOGY SERVICES

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 $\frac{1}{2}$, $\frac{1}{4}$, 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

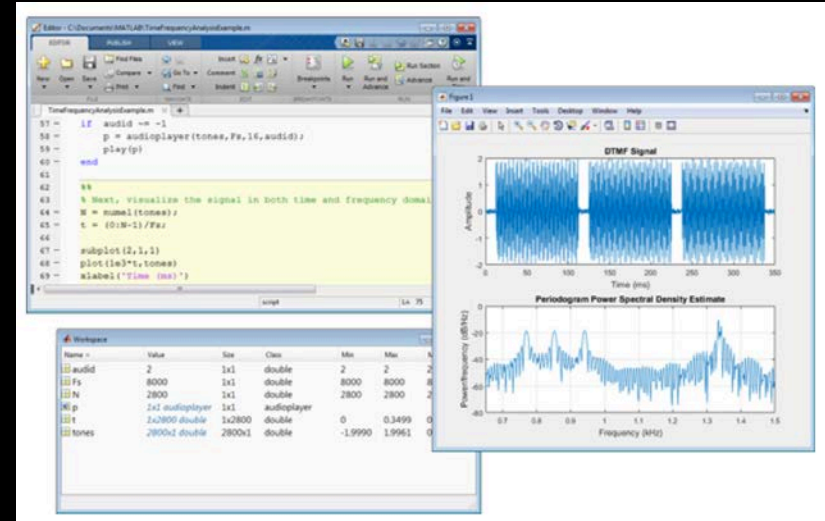
- Interactive user access via web interface with VNC/SSH
- Direct VNC/SSH to individual instances



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	13	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

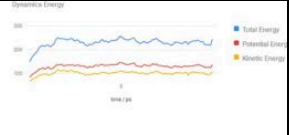
- **Simulations of Nanoscale Biomolecular Systems** - Aleksei Aksimentiev, University of Illinois Champaign-Urbana
- **The Neuroscience Gateway** - Amitava Majumdar, University of California, San Diego
- **Parallelizing Development of Immunomics and Genomics Tools** - Ramy Arnaout, Beth Israel Deaconess 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!



TINKER

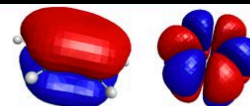
TINKER is a molecular dynamics package from the Jay Ponder Lab.

A line graph showing energy components over time. The y-axis is labeled 'Component Energy' and ranges from 0 to 1000. The x-axis is labeled 'Time (ps)' and ranges from 0 to 1000. The legend indicates three energy components: Total Energy (blue line), Potential Energy (red line), and Kinetic Energy (yellow line). The Total Energy line fluctuates around a mean value of approximately 500. The Potential Energy line fluctuates around a mean value of approximately 200. The Kinetic Energy line fluctuates around a mean value of approximately 300.

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 July 2020

- 406 active XSEDE projects covering 78 fields of science and over 2100 **active users** representing **206 institutions**
- **80%** of Jetstream users have **not used any other XSEDE system**
- >337M CPU hours allocated to XSEDE projects since June 2016
- 38 active science gateways
- 43 education/teaching allocations serving over 700 students currently and >4900 through July 2020
- 1189 mean active VMs in previous qtr, 1632 peak active VM count
- **Highest** user satisfaction in most recent XSEDE survey



**COMMERCIAL
CLOUD**



XSEDEnet
Advanced Layer 2
Services (AL2S) platform

**INDIANA
UNIVERSITY
CYBERINFRASTRUCTURE**

**PRIMARY
COMPUTE**

416 Nodes
53,248 Cores
224 TB RAM

STORAGE

96 Nodes
15 PB

ACCELERATORS

90 Nodes
45 TB RAM
360 GPUs

**CORNELL
UNIVERSITY
CYBERINFRASTRUCTURE**

REGIONAL

COMPUTE

8 Nodes
1,024 Cores
4 TB RAM

STORAGE

869 TB

**UNIVERSITY
OF HAWAI'I
CYBERINFRASTRUCTURE**

REGIONAL

COMPUTE

8 Nodes
1,024 Cores
4 TB RAM

STORAGE

869 TB

ACCELERATORS

2 Nodes
1 TB RAM
8 GPUs

**ARIZONA STATE
UNIVERSITY
CYBERINFRASTRUCTURE**

REGIONAL

COMPUTE

8 Nodes
1,024 Cores
4 TB RAM

STORAGE

869 TB

ACCELERATORS

2 Nodes
1 TB RAM
8 GPUs

TACC CYBERINFRASTRUCTURE

COMPUTE

8 Nodes
1,024 Cores
4 TB RAM

REGIONAL

STORAGE

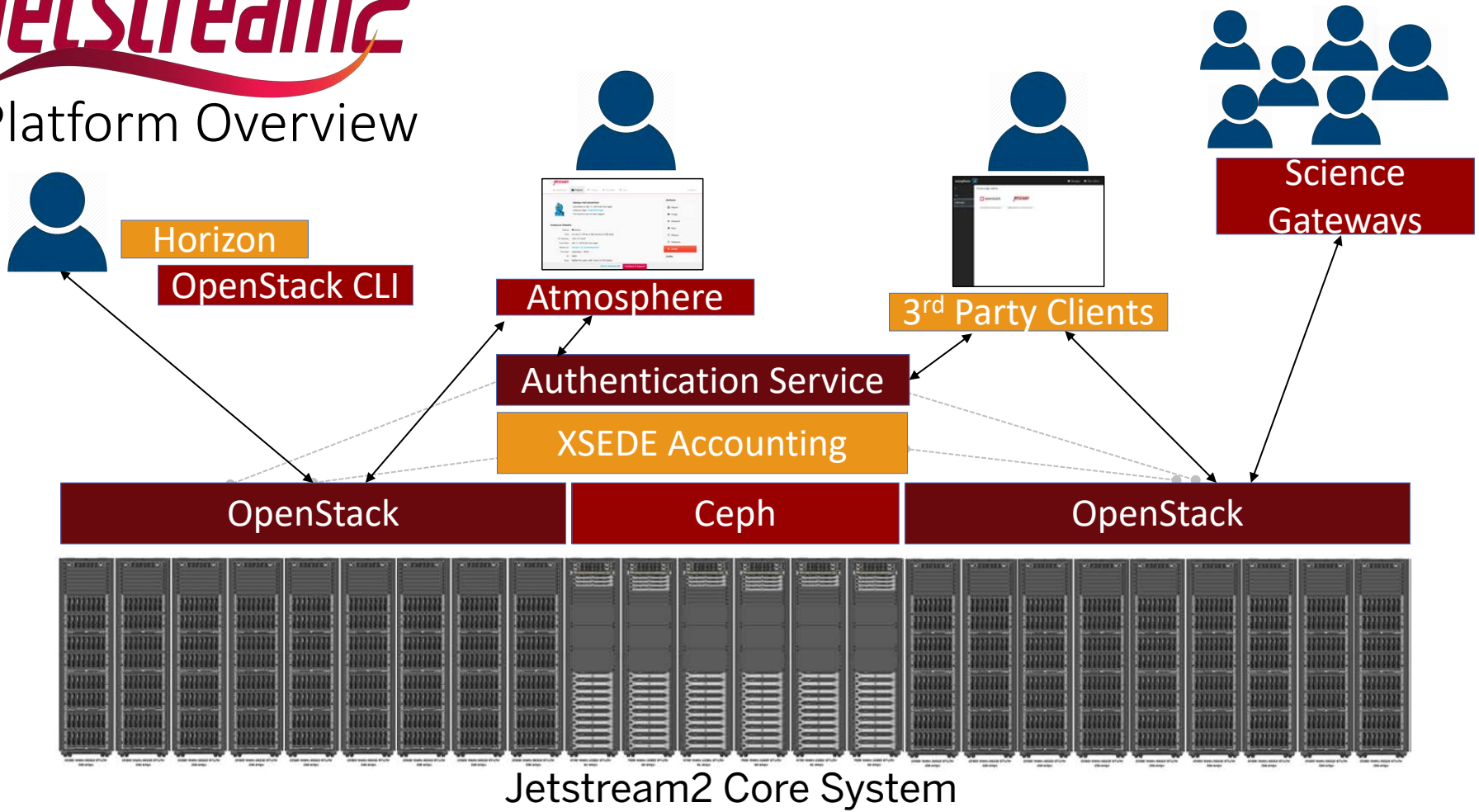
869 TB

ACCELERATORS

2 Nodes
1 TB RAM
8 GPUs

Jetstream2

Platform Overview



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 Sep 2020 – 15 Oct 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



THE UNIVERSITY
OF ARIZONA



THE UNIVERSITY OF
CHICAGO



JOHNS HOPKINS
UNIVERSITY



UNC
THE ODUM INSTITUTE



RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES

Jetstream² Partners

