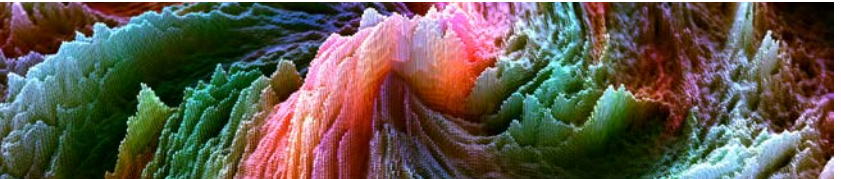
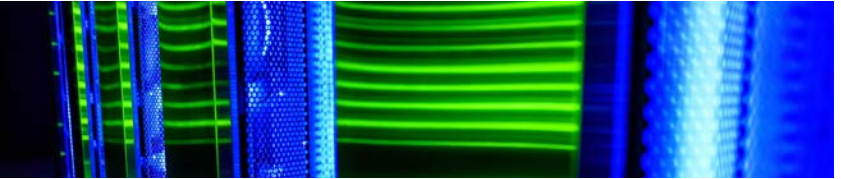
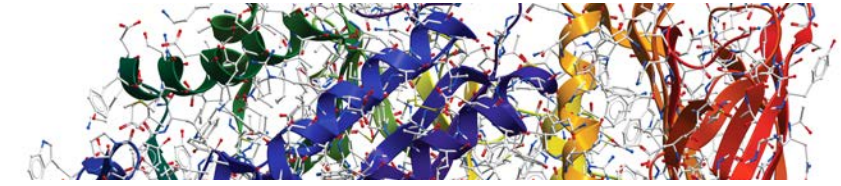


rt.iu.edu



**RESEARCH TECHNOLOGIES**  
PERVASIVE TECHNOLOGY INSTITUTE



**RESEARCH TECHNOLOGIES**  
UNIVERSITY INFORMATION TECHNOLOGY SERVICES

# Jetstream Overview: A national research and education cloud

**University of North Carolina Asheville**

**March 28, 2019**

**Sanjana Sudarshan, PhD – [ssudarsh@iu.edu](mailto:ssudarsh@iu.edu)**

Senior Technical Advisor,

UITS Research Technologies

Sudarshan, S. (2019). Jetstream: A national research and education cloud. UNCA. 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, coming in 2019)
- 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



# 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.”

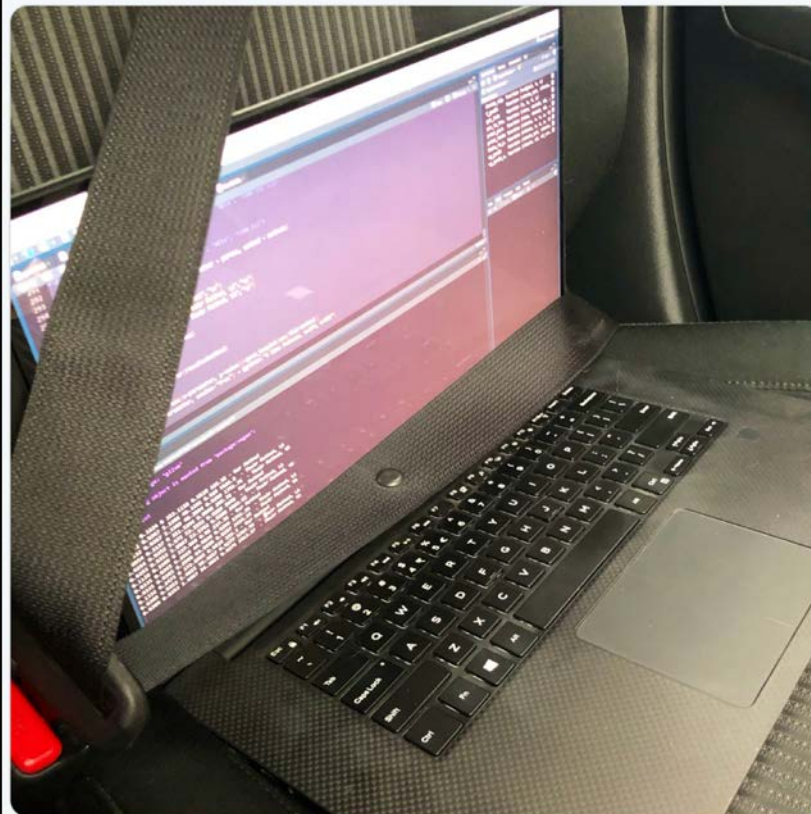
– multiple researchers at small colleges and universities



Sarah Romanes @sarah\_romanec · 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



RESEARCH TECHNOLOGIES

UNIVERSITY INFORMATION TECHNOLOGY SERVICES

# 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*



# Who uses Jetstream?

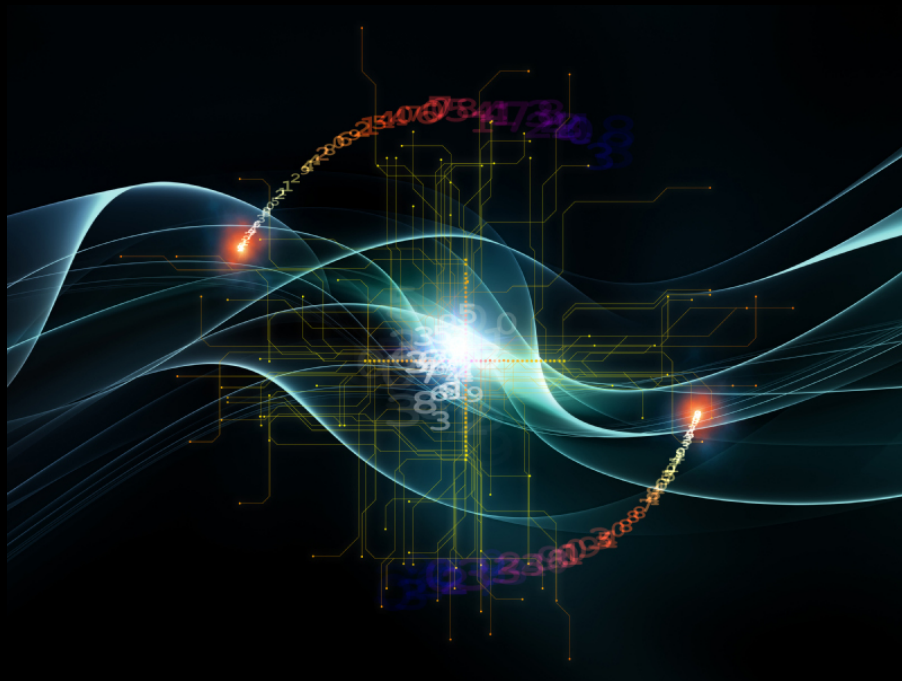
- The researcher needing a handful of cores (1 to 44 vCPUs)
- Software creators and researchers needing to create their own VMs and workflows
- Science gateway creators using Jetstream as either the frontend or processor for scientific jobs
- STEM Educators teaching on a variety of subjects





# What Jetstream isn't...

- It's not traditional HPC
- There's no shared filesystem (think cloudy!)
- There aren't GPUs (yet...they're coming!)
- 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...



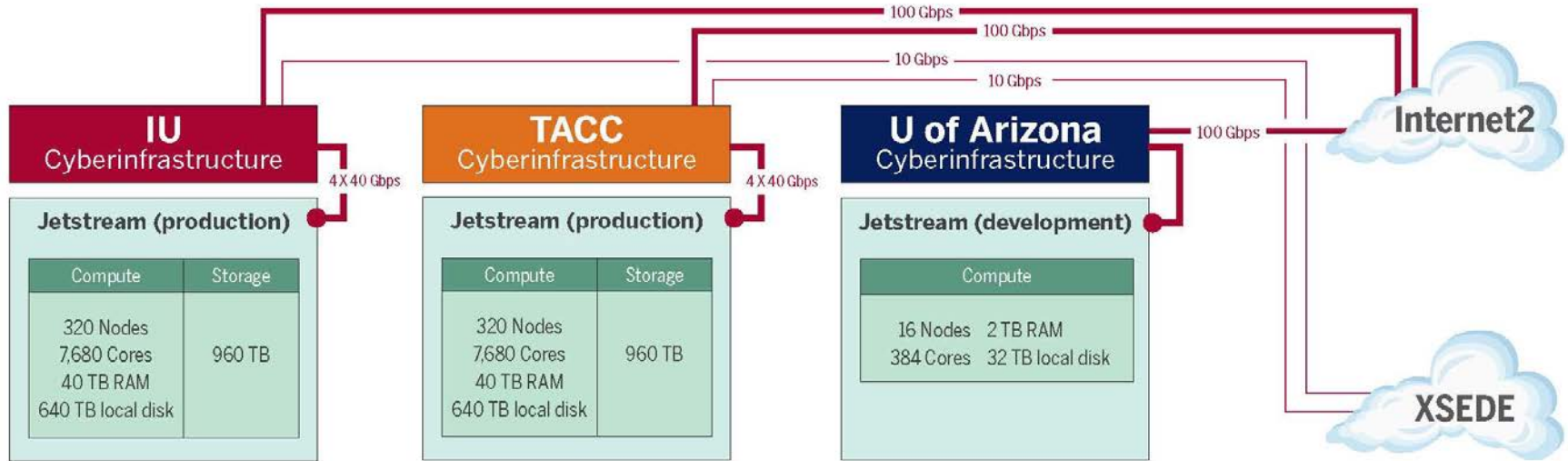
Cows, not pets: pets take great amount of care, feeding, and you name them; cows you intend to have high turnover and you give them numbers.

-- Mike Lowe (Jetstream architect)

\*\*some caveats for gateways...



# Jetstream System Overview

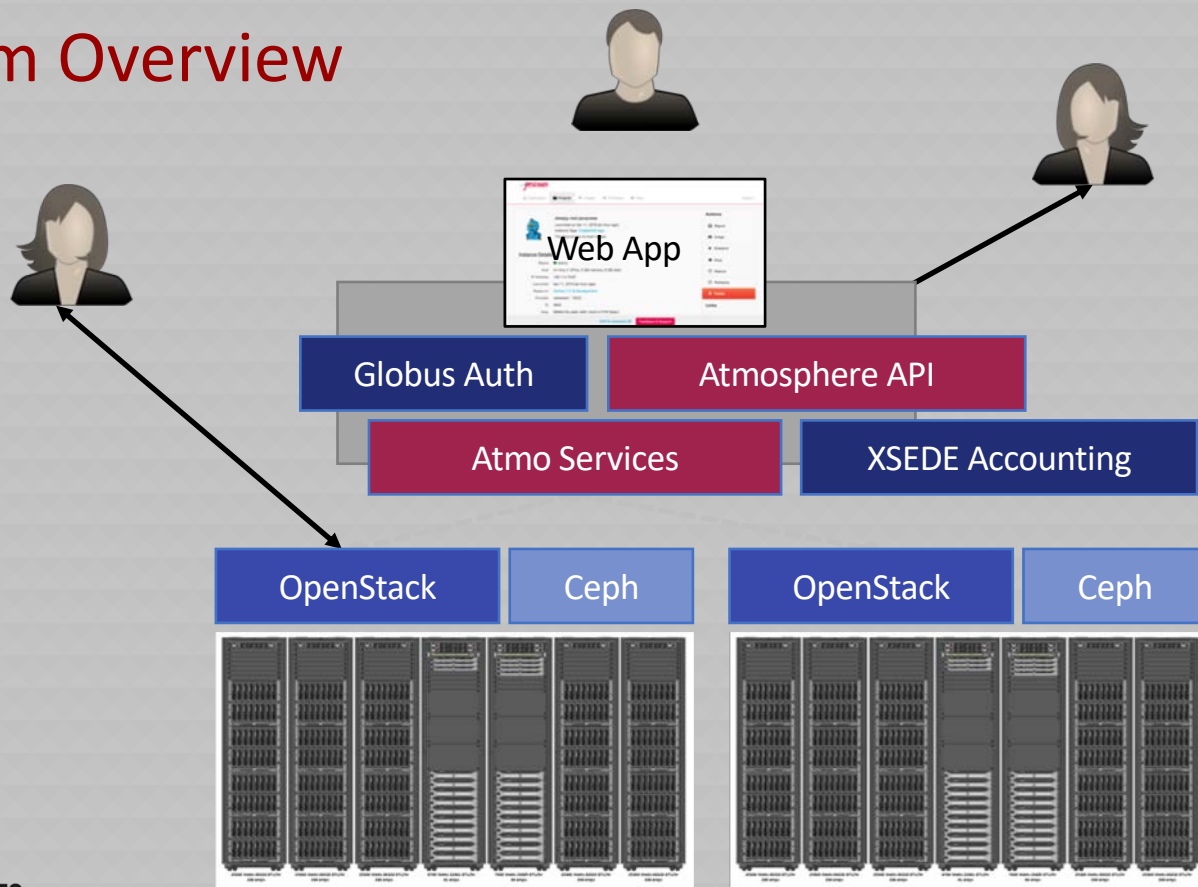


# Production cloud hardware (per site)

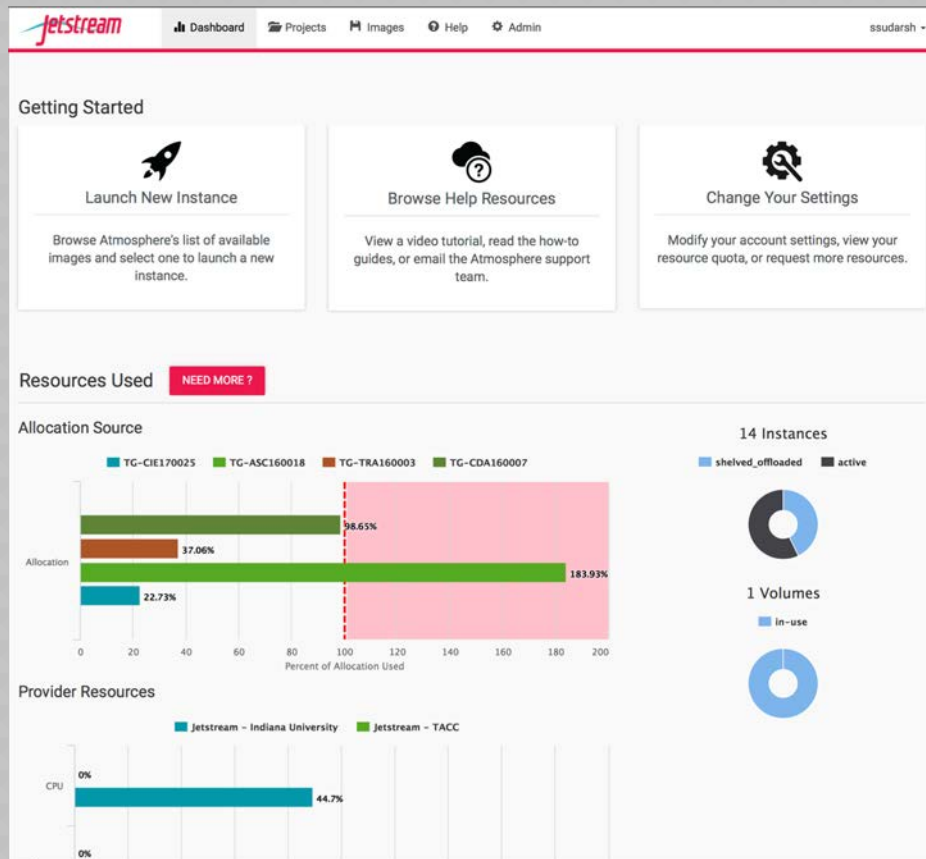
| Hardware                         | Number | Specifications   | Function (IU)  |
|----------------------------------|--------|--|--|
| Dell PowerEdge M630 blades       | 320    | 2x Intel E5-2680v3 "Haswell"<br>24 cores @ 2.5 GHz<br>128 GB RAM<br>2 TB local disk            | Compute hosts<br>OpenStack services                              |
| Dell PowerEdge R630 1U server    | 7      | 2x Intel E5-2680v3 "Haswell"<br>24 cores @ 2.5 GHz<br>128 GB RAM<br>2 TB local disk            | Cluster management<br>High Availability<br>Databases<br>RabbitMQ |
| Dell PowerEdge R730xd 2U servers | 20     | 2x Intel E5-2680v3 "Haswell"<br>24 cores @ 2.5 GHz<br>64 GB RAM<br>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<br>Spine   |



# 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 links for Dashboard, Projects, Images, Help, and Admin. The user's name, ssudarsh, is visible in the top right corner. Below the navigation bar is a search bar and several filters: SEARCH, FAVORITES (0), MY IMAGES (12), MY IMAGE REQUESTS, and TAGS. The main content area is titled "Image Search" and features a search input field with the placeholder text "Search across image name, tag or description". Below the search bar, it indicates "Showing 100 of 652 images".

The "Featured Images" section lists several image entries, each with a thumbnail, title, date, author, and a list of tags. The entries are:

- Genomics Toolkit**: Mar 25th 19 01:04 by ssudarsh. Description: Genome Analysis Tools. Look here of complete list of tools -> <https://iujetstream.atlassa...>. Tags: bioconda, bioconductor, bioinformatics, Featured, genomics.
- Ubuntu 18.04 Devel and Docker**: Mar 15th 19 09:04 by jfischer. Description: Ubuntu 18.04 LTS Development + GUI support + Docker. Based on Ubuntu cloud image for 18.04 ... Tags: base, desktop, development, docker, docker-compose, Featured, Ubuntu, vnc.
- Centos 7 (7.6) Development GUI**: Mar 15th 19 08:52 by jfischer. Description: Centos 7 (7.6) Development GUI. Installation size ~ 4.5GB. Tags: CentOS, development, docker, docker-compose, Featured, gui, IRODS.
- Ubuntu 16.04 Devel and Docker**: Mar 15th 19 08:02 by jfischer. Description: Ubuntu 16.04 LTS Development + GUI support + Docker. Based on Ubuntu cloud image for 16.04 ... Tags: base, desktop, development, docker, docker-compose, Featured, Ubuntu, vnc, x2go.
- Centos 7 Devel with Anaconda**: Mar 15th 19 06:39 by jfischer. Description: Centos 7 (7.6) Development GUI with Anaconda installed into /opt/anaconda2. Installation s ... Tags: CentOS, development, Featured, jupyter, m1\_small, python.
- R and Shiny Server with GCC (C ...)**: Mar 15th 19 06:12 by jfischer. Description: 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:

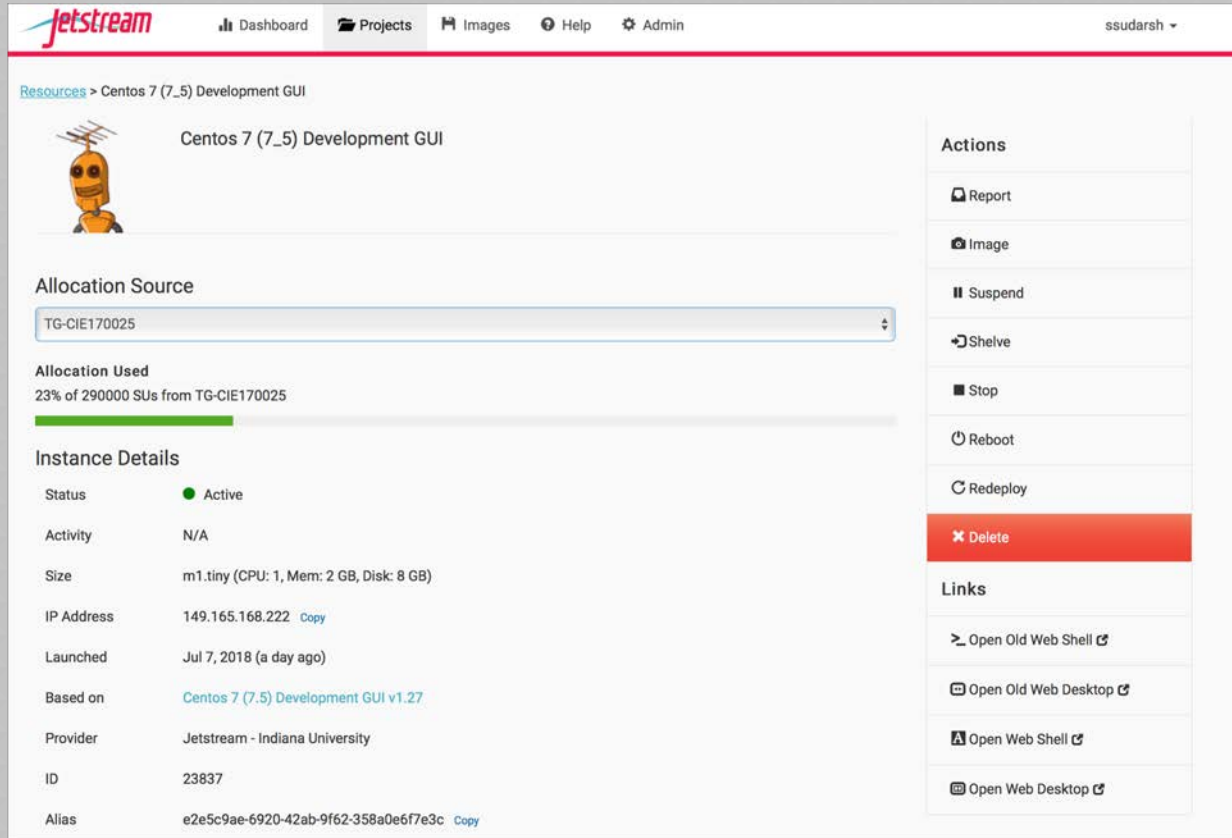
[Edit details](#)

**Versions**

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



# Even more Jetstream web interface...



The screenshot displays the Jetstream web interface for a specific instance. The top navigation bar 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 small robot icon and the instance name 'Centos 7 (7\_5) Development GUI'. Below this, the 'Allocation Source' is shown as 'TG-CIE170025'. The 'Allocation Used' section indicates that 23% of 290,000 SUs are used from the source, represented by a green progress bar. The 'Instance Details' table provides the following information:

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

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 options: Open Old Web Shell, Open Old Web Desktop, Open Web Shell, and Open Web Desktop.





# 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



# 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



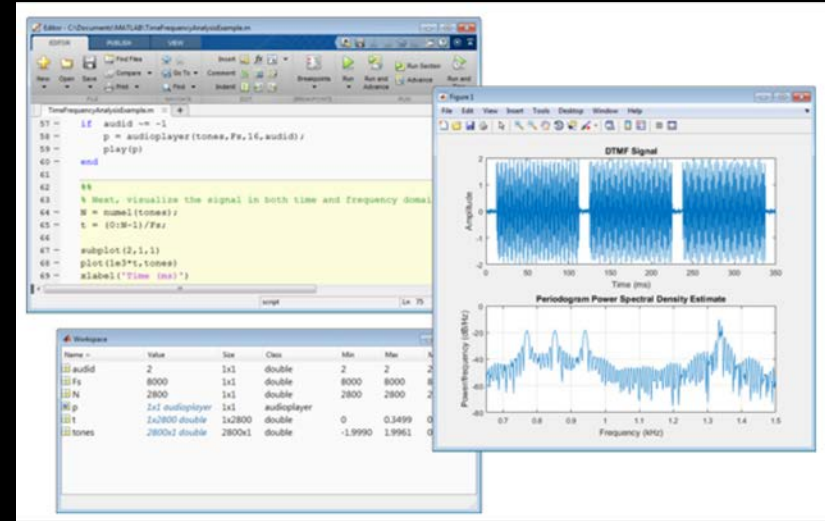
| Discipline or area of interest | #of Jetstream allocations | SUs allocated on Jetstream | SU increase/Decrease on Jetstream over previous year | % of SUs allocated on Jetstream | % of all SUs allocated on other XSEDE-supported systems |
|--------------------------------|---------------------------|----------------------------|--|---------------------------------|---|
| 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

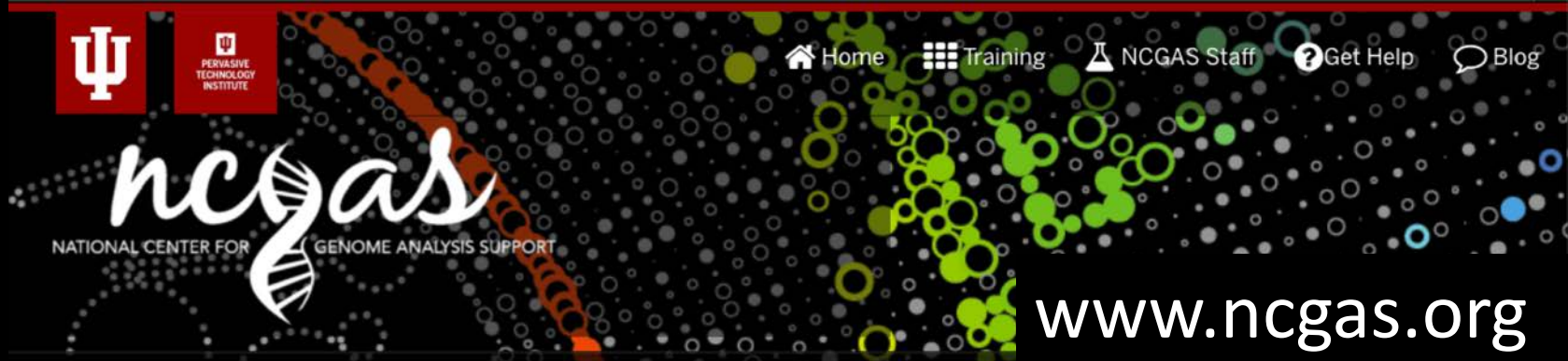


# Other images..

- Genomics toolkit
- PacBio
- NCGAS popup browser
- R
- R Studio
- TensorFlow



# The National Center for Genome Analysis Support



## Who NCGAS serves

- IU students and researchers
- NSF researchers - as an NSF-funded service center
- "For pay" work on others' grants (NSF, USDA, *etc.*)

## Goals

- Provide bioinformatics expertise through consultation
- Maintain a curated set of genomics applications
- Provide access to HPC resources
- Pursue outreach to biologists

# Jetstream Gateway Highlights

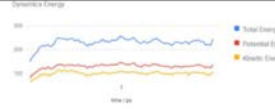
- **Simulations of Nanoscale Biomolecular Systems** - Aleksei Aksimentiev, University of Illinois Champaign-Urbana
- **The Neuroscience Gateway** - Amitava Majumdar, University of California, San Diego
- **Parallelizing Development of Immunomics and Genomics Tools** - Ramy Arnaout, Beth Israel Deaconess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

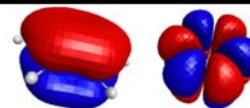
TINKER is a molecular dynamics package from the Jay Ponder Lab.

A line graph showing energy components over time. The y-axis is labeled 'Energy (kJ/mol)' and ranges from 0 to 100. The x-axis is labeled 'Time (ps)' and ranges from 0 to 100. Three lines are plotted: Total Energy (blue), Potential Energy (red), and Kinetic Energy (yellow). The Total Energy line fluctuates around a mean value of approximately 50 kJ/mol. The Potential Energy line fluctuates around a mean value of approximately 20 kJ/mol. The Kinetic Energy line fluctuates around a mean value of approximately 30 kJ/mol.

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
- **6 positions available in 2019 - APPLY NOW!!**
- REU student videos on YouTube  
<https://www.youtube.com/user/IUPTI>





# Jetstream usage highlights – 1 March 2019

- 418 active XSEDE projects covering 71 fields of science and over 2250 **active users** representing **198 institutions**
- **80%** of Jetstream users have **not used any other XSEDE system**
- >182M CPU hours allocated to XSEDE projects since June 2016
- 19 active science gateways
- 47 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
- 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!**



# Where can I get help?

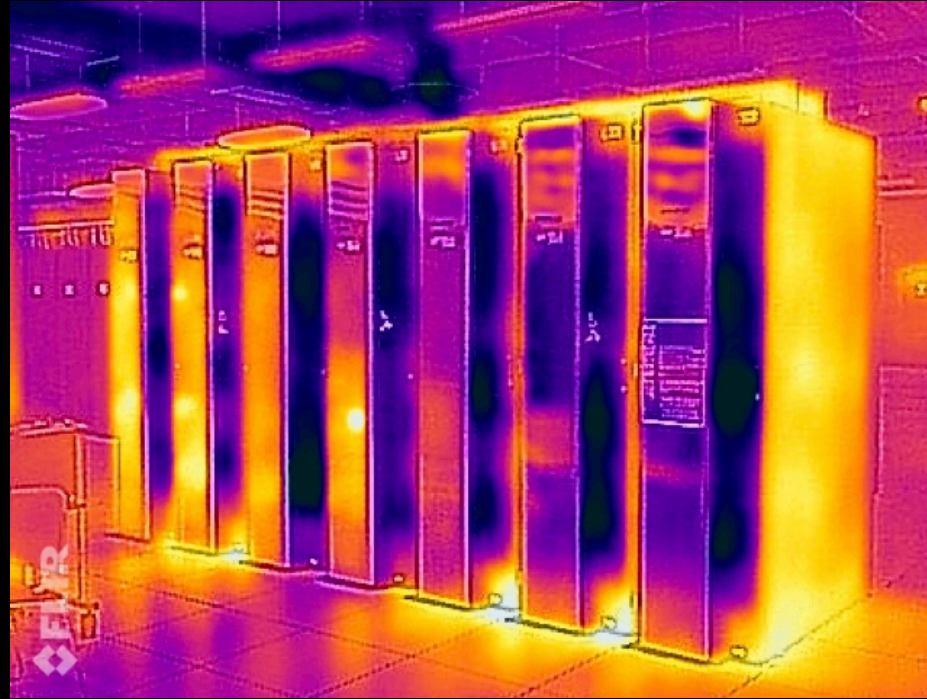
- Wiki / Documentation: <http://wiki.jetstream-cloud.org>
- User guides: <https://portal.xsede.org/user-guides>
- XSEDE KB: <https://portal.xsede.org/knowledge-base>
- Email: [help@xsede.org](mailto:help@xsede.org)
- Campus Champions: <https://www.xsede.org/campus-champions>
- Introduction to Jetstream Virtual Workshop: <https://cvw.cac.cornell.edu/jetstream/>
- Jetstream Allocations Virtual Workshop: <https://cvw.cac.cornell.edu/JetstreamReq/>



# Jetstream Partners



# Infrared image of Jetstream



# Jetstream Fun: Happy cluster / Angry Cluster



# Questions?

Project website: <http://jetstream-cloud.org/>

Project email: [help@jetstream-cloud.org](mailto:help@jetstream-cloud.org) Direct email: [ssudarsh@iu.edu](mailto:ssudarsh@iu.edu)

## License Terms

- Sudarshan, Sanjana. November 3, 2018. Jetstream: A national research and education cloud. CCSCSE – Salem, VA.  
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.







**RESEARCH TECHNOLOGIES**  
UNIVERSITY INFORMATION TECHNOLOGY SERVICES

# Jetstream Overview: Hands-on with Atmosphere

**University of North Carolina Asheville**

**March 28, 2019**

**Sanjana Sudarshan, PhD – [ssudarsh@iu.edu](mailto:ssudarsh@iu.edu)**

Senior Technical Advisor,

UITS Research Technologies

Sudarshan, S. (2019). Jetstream: A national research and education cloud. UNCA. Retrieved from <https://jetstream-cloud.org/research/publications.php>

# Cloud Computing Terms...simplified

**Image:** a file on a disk. It will be booted to create an...

**Instance:** a running virtual server; i.e. something you can log into

**State:** something worth remembering; i.e. the state of the system



# Cloud Computing Terms...simplified

- **Flavor:** the size of a running instance; i.e. #core, RAM, disk
- **Hypervisor:** the thing the instance runs on; something akin to a software defined hardware compute server.
- **Snapshot:** the process of taking an instance and turning it to an image.



# Cloud Computing Terms...simplified

## States:

- **Running:** the *instance* is up & running
- **Suspended:** the *instance* is memory resident but not running
- **Stopped:** the *instance* is shutdown akin to powering down
- **Shelved:** the *instance* is shutdown, written to disk, and stored



# Getting into the hands on part

Open <https://use.jetstream-cloud.org> in your browser

Login slips will be distributed momentarily!

