





# Jetstream Overview: A national research and education cloud

CyberCarpentry UNC

July 16, 2019

Sanjana Sudarshan, PhD – ssudarsh@iu.edu

Senior Technical Advisor,

**UITS Research Technologies** 

Sudarshan, S. (2019). Jetstream: A national research and education cloud.

CyberCarpentry Workshop UNC. 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\_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



# 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?

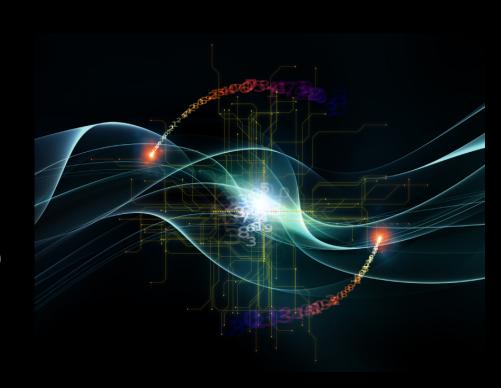
- The researcher needing a handful of cores (1 to 44 vCPUs)
- Software creators and researchers needing to create their own VMs and workflows

- 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 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...

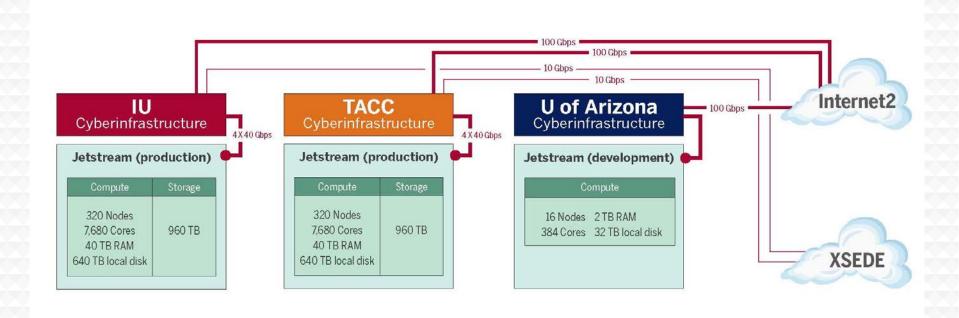


Cows, not pets: pets take great amount of care, feeding, and you name them; cows you intend to have high turnover and you give them numbers.

-- Mike Lowe (Jetstream architect)

\*\*some caveats for gateways...

# **Jetstream System Overview**

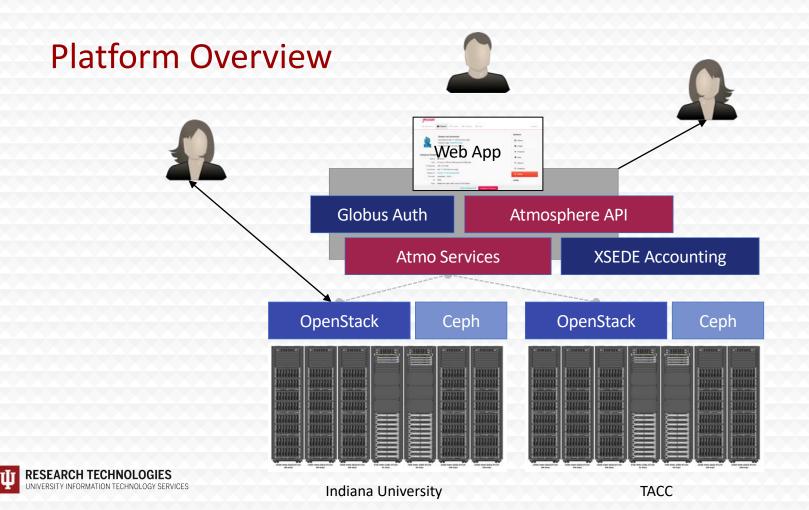




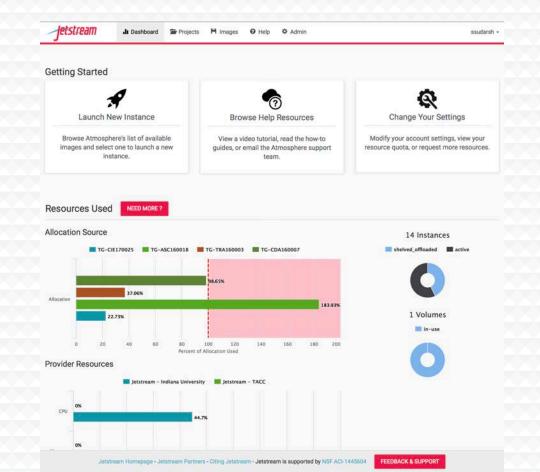
# Production cloud hardware (per site)

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	



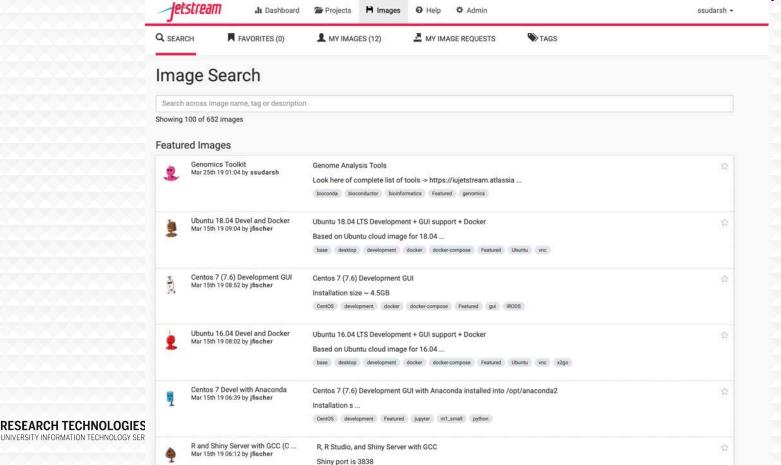


# The Jetstream Atmosphere web interface

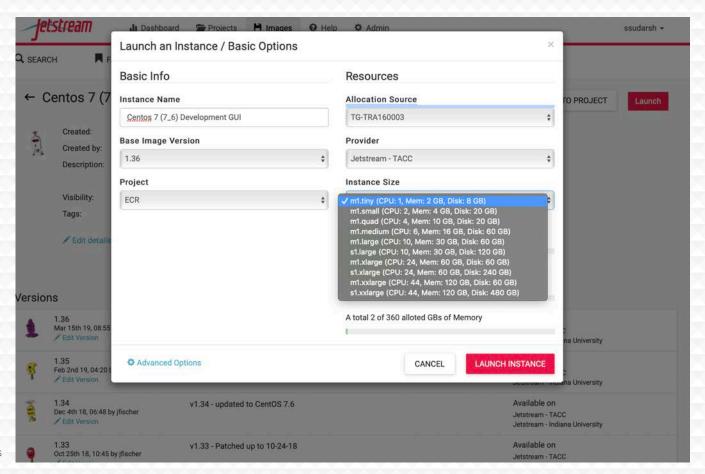




# The Jetstream Atmosphere web interface (contd..)

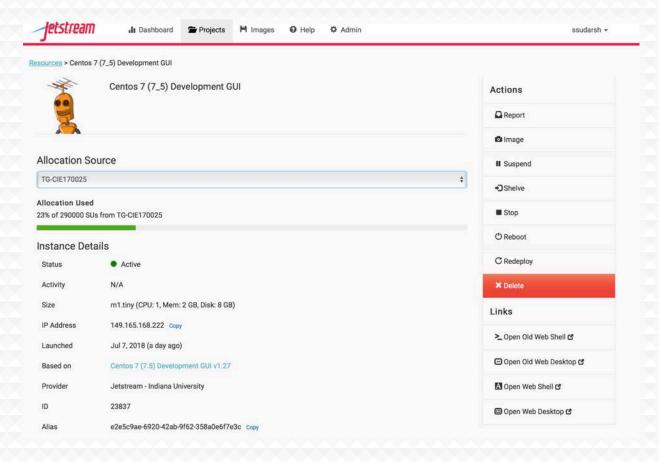


#### Look! It's more Jetstream web interface!





### 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: 0.0 172 Processes: Usage of /: 52.1% of 7.58GB Users logged in: 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*	11

<sup>\*\*</sup> 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



interest	allocations	Jetstream	on Jetstream over previous year	% of SUS allocated on Jetstream	on other XSEDE- supported systems
Behavioral Sciences	6	3,465,516	100%	4.24%	0.61%
<b>Biological Sciences</b>	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%

SUs allocated on

2,440,581

2,409,633

SU

increase/Decrease

15.58%

192.27%

% of all SUs allocated

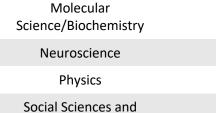
8.65%

0.81%

% of SUs allocated

2.99%

2.95%



Humanities

Discipline or area of

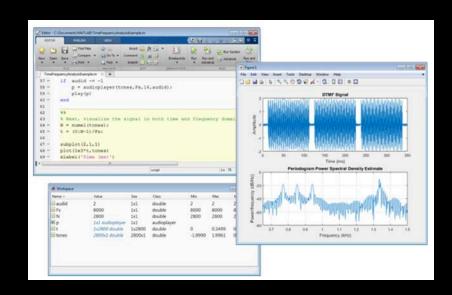
**#of Jetstream** 

10

28

# 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

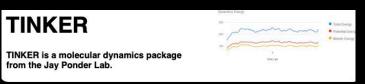
# Jetstream Gateway Highlights

- 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!











# 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 May 2019

- 418 active XSEDE projects covering 71 fields of science and over 2375 active users representing 208 institutions
- 80% of Jetstream users have not used any other XSEDE system
- >200M CPU hours allocated to XSEDE projects since June 2016

- 25 active science gateways
- 58 education/teaching allocations
- serving over 1000 students
- 1438 mean active VMs in previous qtr, 1614 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 Sep 2019 15 Oct 2019
- Research allocation: Project desc (<10 pages) and Scaling doc (<5 pages)</li>

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**















# Infrared image of Jetstream



# Jetstream Fun: Happy cluster / Angry Cluster





#### Questions?

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

Project email: <a href="mailto:help@jetstream-cloud.org">help@jetstream-cloud.org</a> Direct email: <a href="mailto:ssudarsh@iu.edu">ssudarsh@iu.edu</a>

#### License Terms

• Sudarshan, Sanjana. July 16, 2019. Jetstream: A national research and education cloud. Cybercarpentry – UNC, NC. Also available at: http://ietstream.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 (<a href="http://creativecommons.org/licenses/by/3.0/">http://creativecommons.org/licenses/by/3.0/</a>). 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.





## Jetstream Overview: Hands-on with Atmosphere

CyberCarpentry UNC

July 16, 2019

Sanjana Sudarshan, PhD – ssudarsh@iu.edu

Senior Technical Advisor,

**UITS Research Technologies** 

Sudarshan, S. (2019). Jetstream: A national research and education cloud.

CyberCarpentry Workshop UNC. 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 (Contd.)**

**Object store**: a blob of bits; it has a starting address & a size. There may be metadata associated with the object. The data is consumed in a streaming manner.

**Block store**: a software defined entity akin to an unformatted hardware disk drive, data is stored in blocks and has no metadata associated at the lowest levels

**Filesystem**: hierarchical in nature, directories & files, ability to open, seek, read, write. **Persistent storage**: If you pull the plug, it will still exist when power is restored. Safe to store data or state here.

**Ephemeral storage**: If you pull the plug, it no longer exists. (Don't leave your data here!!!)



# **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 - GUI**

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

Login slips will be distributed momentarily!

https://iu.jetstream-cloud.org



# **Getting into the hands on part - API**

- Open <a href="https://is.gd/jYsnyJ">https://is.gd/jYsnyJ</a> in your browser
- You'll be connecting to a prepped host...but you might want to wait until I tell you something key first.

ssh trainXX@tutorial.jetstream-cloud.org

