# Jetstream Overview
# Gateways Conference

Jeremy Fischer – jeremy@iu.edu

Senior Technical Advisor,
UITS Research Technologies

Fischer, J. (2017). Jetstream Overview: Gateways Conference. Retrieved from https://jetstream-cloud.org/publications.php

# What is Jetstream and why does it exist?

- NSF's first production cloud facility, part of the NSF eXtreme Digital (XD) program
- Focus on ease-of-use, broad accessibility
- Encourage collaboration and community development
- User-selectable library of preconfigured virtual machines
- Provides on-demand *interactive* computing and analysis or persistent services such as gateways (e.g. SEAGrid, Galaxy, GenApp, and others)
- Enables *configurable* environments and *programmable cyberinfrastructure*
- Reproducibility: Share VMs and then store, publish via IU Scholarworks (DOI)

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

# Who uses Jetstream?

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

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

# 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
http://jetstream-cloud.org/

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

# Jetstream and way of the cloud…

- **Cloudy Technologies**: clouds are more the 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…**

Jetstream
http://jetstream-cloud.org/

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

# Thinking about VMs…

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

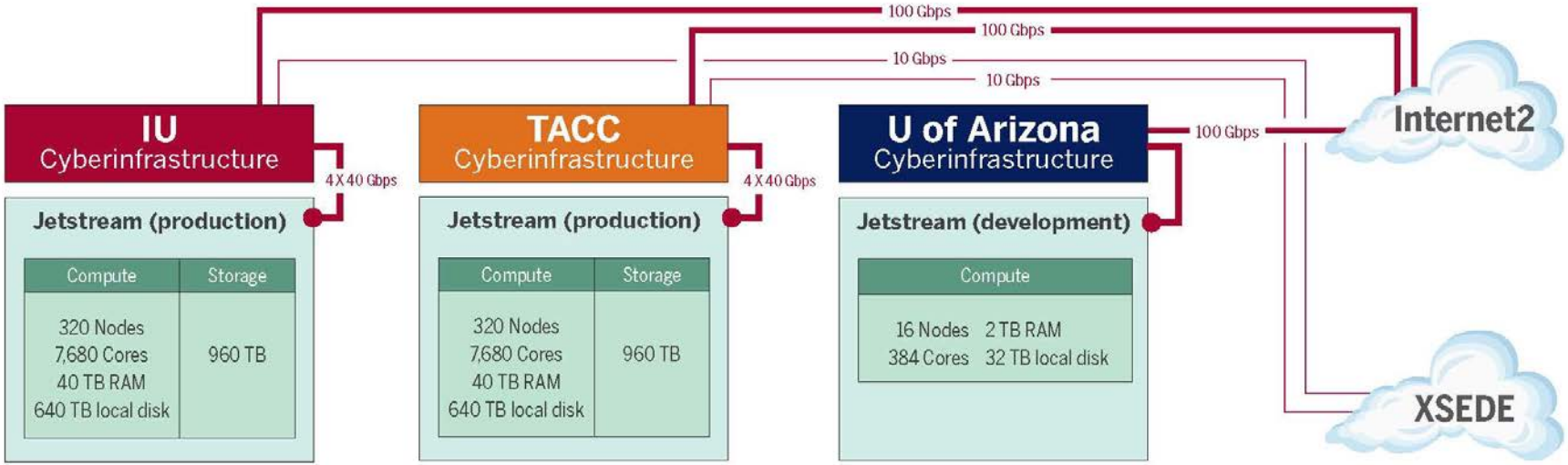
-- Mike Lowe (Jetstream architect)

** Some caveats for gateways, but…

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

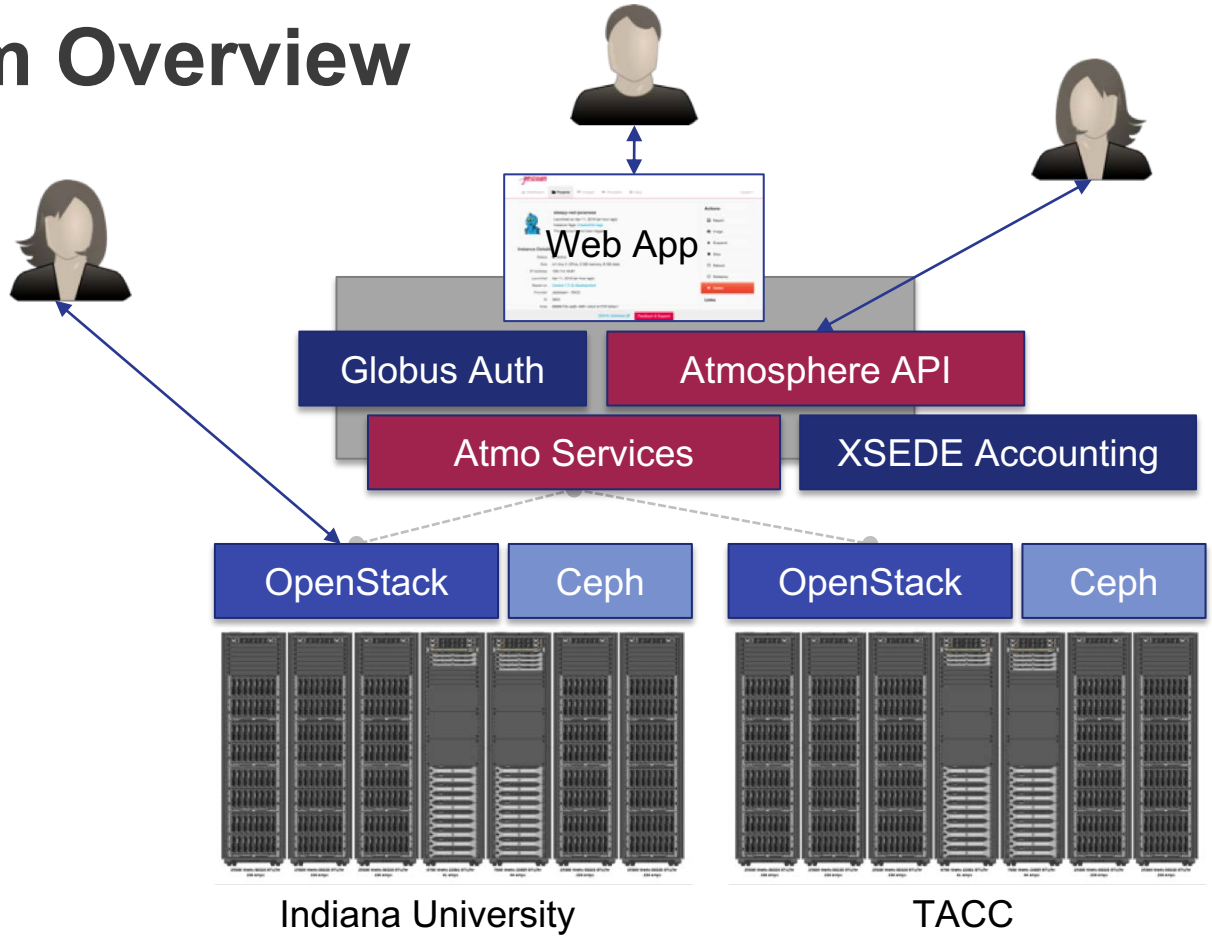
# Jetstream System Overview




Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

# Platform Overview



# Hardware and Instance ”Flavors”

**VM Host Configuration**

- Dual Intel E-2680v3 “Haswell”
- 24 physical cores/node @ 2.5 GHz (Hyperthreading on)
- 128 GB RAM
- Dual 1 TB local disks
- 10GB dual uplink NIC
- Running KVM Hypervisor

- Short-term *ephemeral* storage comes as part of launched instance
- Long-term storage is XSEDE-allocated
- Implemented as OpenStack Volumes
- Each user can get 10 volumes up to 500GB total storage*

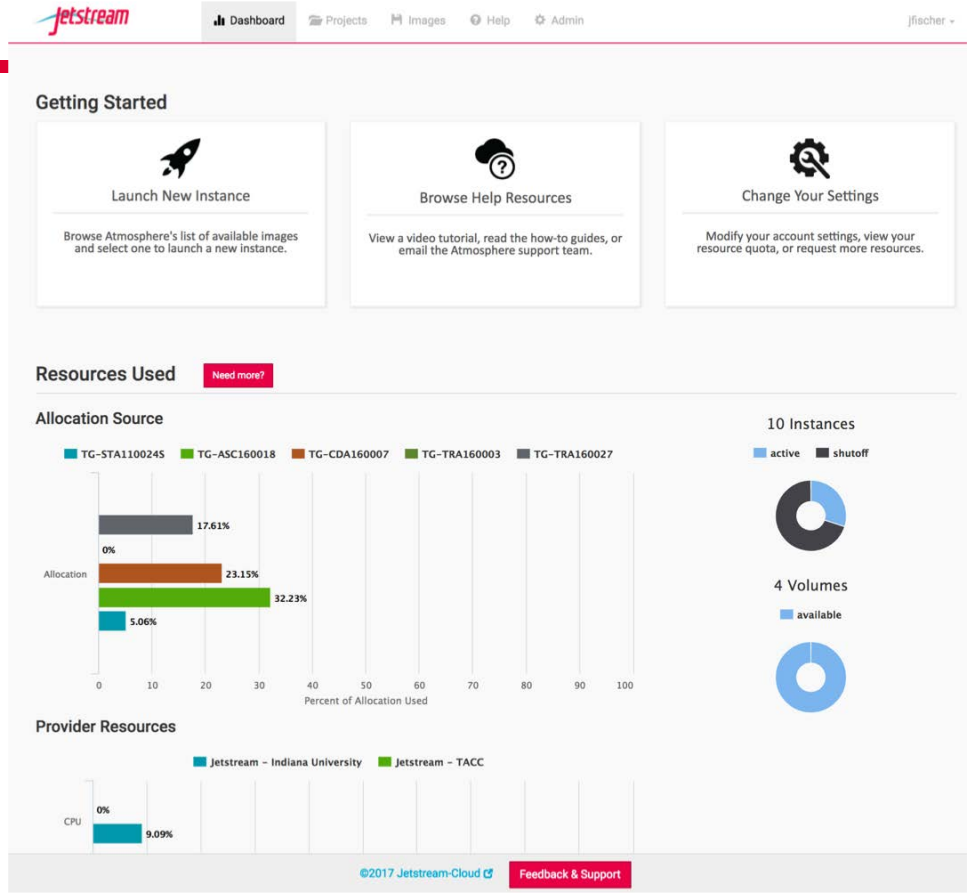
| Flavor | vCPUs | RAM | Storage | Per Node |
|---|---|---|---|---|
| m1.tiny | 1 | 2 | 8 | 46 |
| m1.small | 2 | 4 | 20 | 23 |
| m1.medium | 6 | 16 | 60 | 7 |
| m1.large | 10 | 30 | 60 | 4 |
| m1.xlarge | 24 | 60 | 60 | 2 |
| m1.xxlarge | 44 | 120 | 60 | 1 |
| s1.large** | 10 | 30 | 120 | 4 |
| s1.xlarge** | 24 | 60 | 240 | 2 |
| s1.xxlarge** | 44 | 120 | 480 | 1 |

** s1.* based instances are not eligible to be saved into a customized image

Jetstream
http://jetstream-cloud.org/

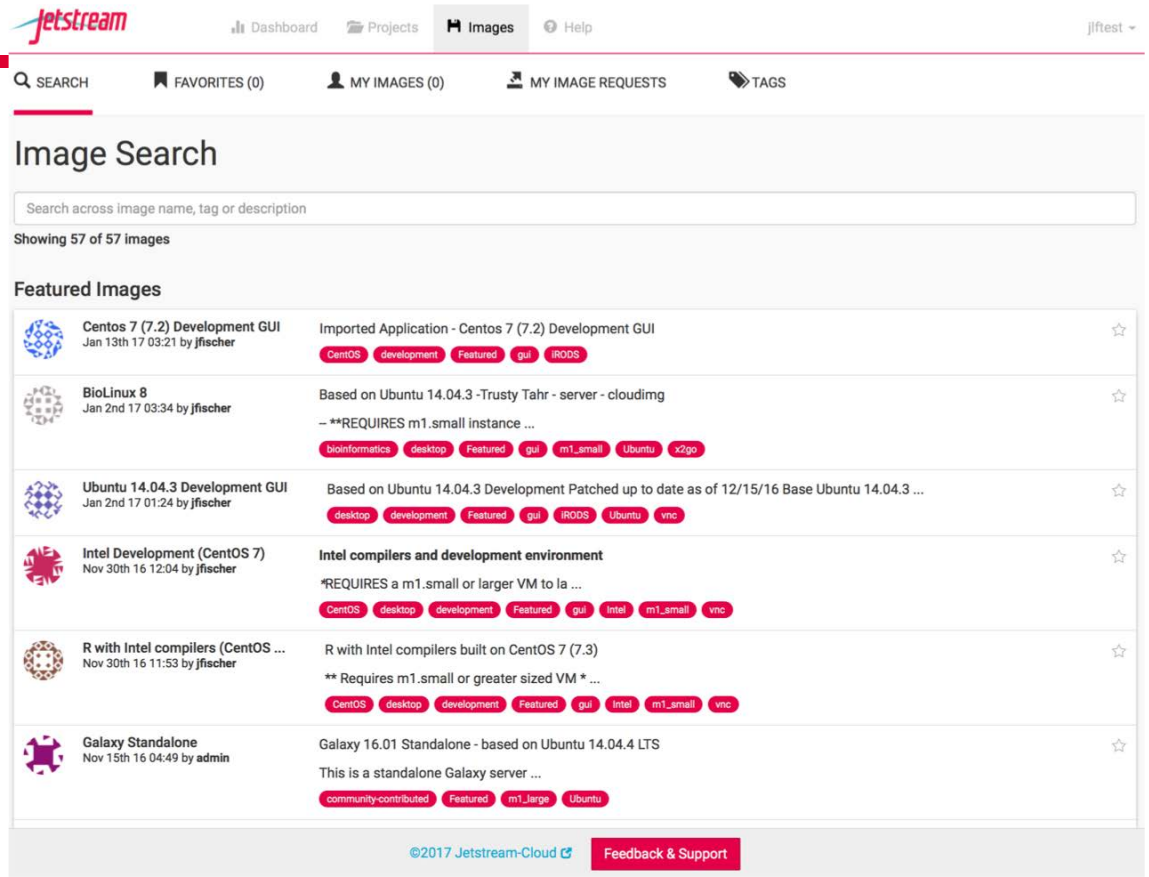
funded by the National Science Foundation
Award #ACI-1445604

# The Jetstream Atmosphere web interface



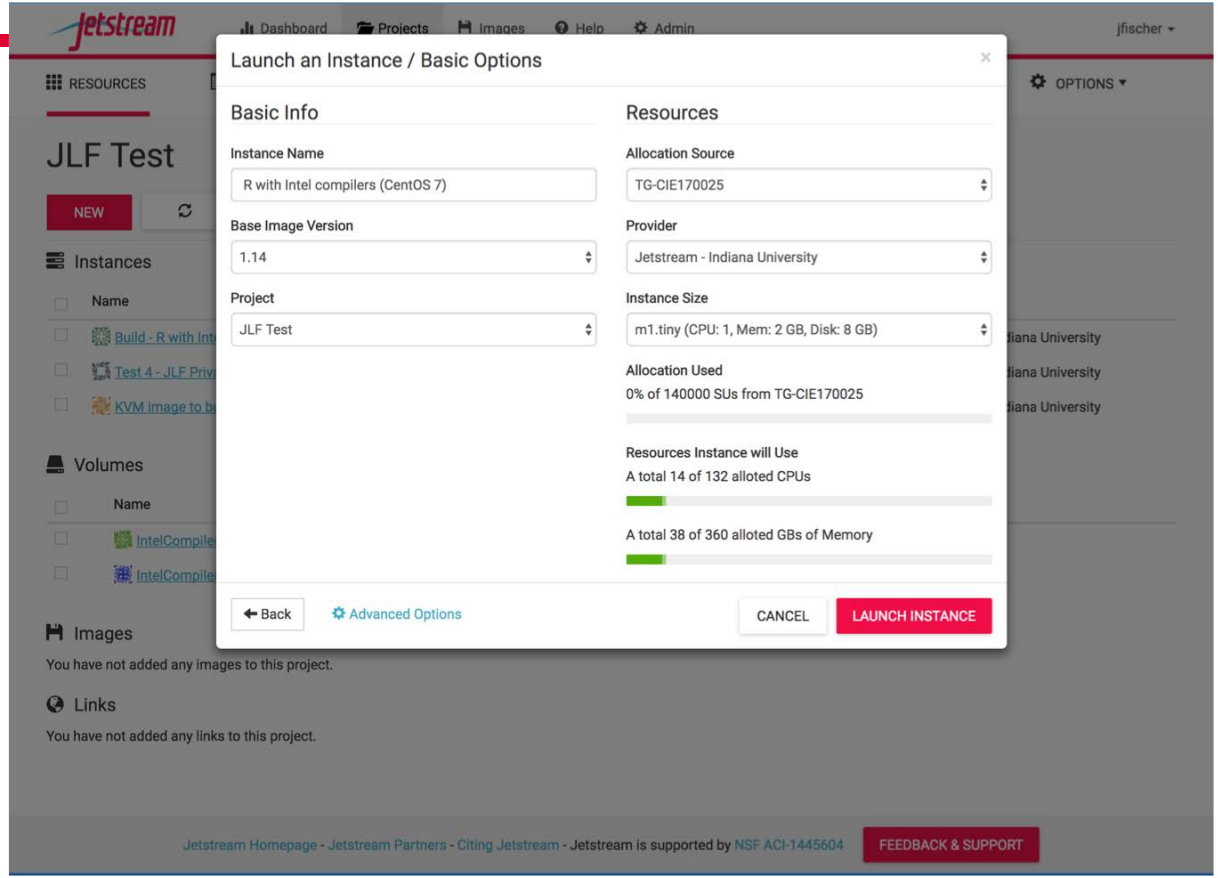

Jetstream
http://jetstream-cloud.org/

# The Jetstream Atmosphere web interface



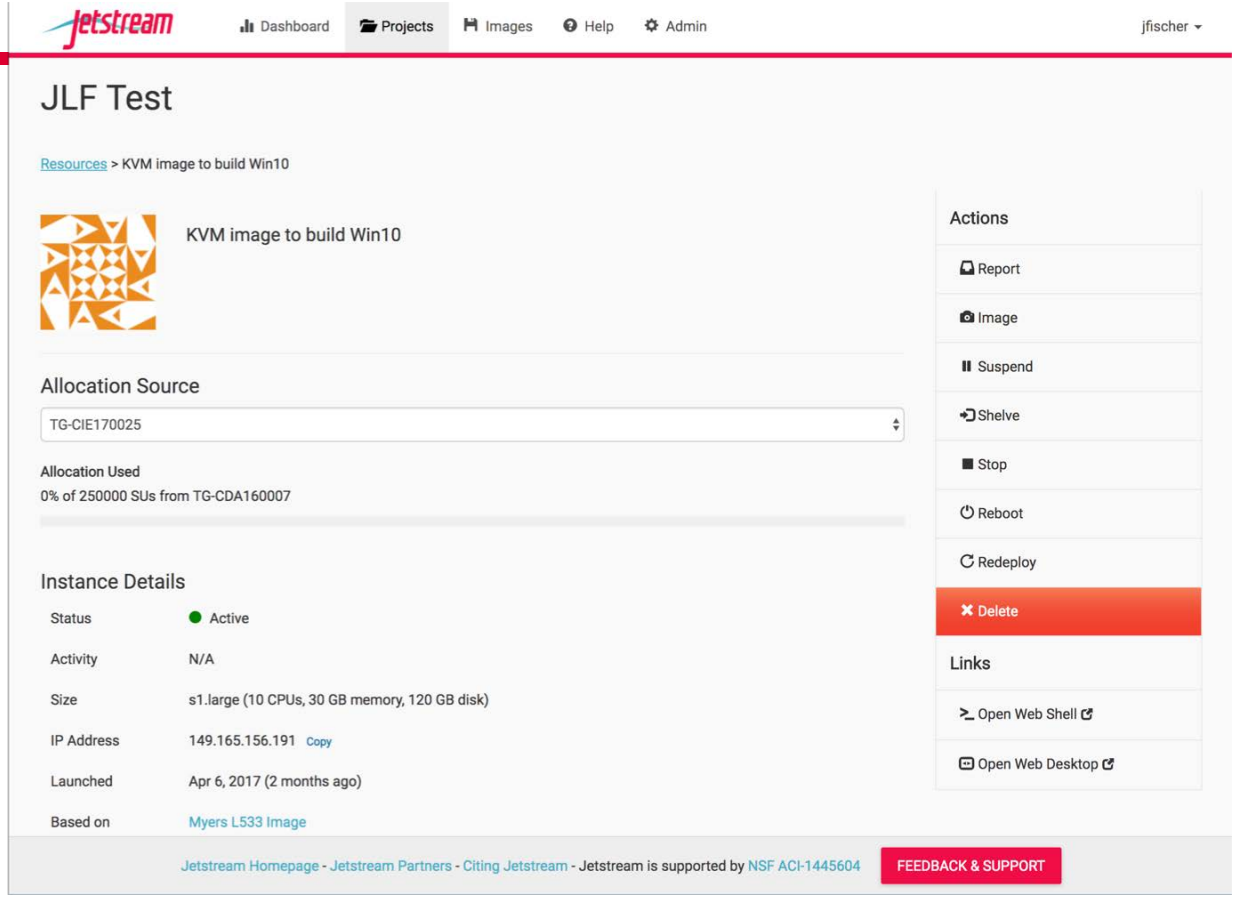

Jetstream
http://jetstream-cloud.org/

# Look! It's more Jetstream web interface!




Jetstream
http://jetstream-cloud.org/

# Even more Jetstream web interface…




Jetstream
http://jetstream-cloud.org/

# Using Jetstream as a gateway developer

Manipulating Jetstream VMs:

- Direct API access via OpenStack CLI or Horizon access
- Log in via ssh (or gui if you install X and a VNC server– but that's up to you)

Why:

- Programmatic access – Programmable cyberinfrastructure
- Reserved IP pools
- Ability for true cloud benefits like elastic computing are available via the API

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

# The basics that Jetstream provides

- Base images to start from
- A basic network space (some DIY required)
- An IP pool
- Isolated project space for your team (on two clouds)
- A blank canvas (to some degree) – you can truly install just about anything you want – and you can BYOLicense if needed

Jetstream
http://jetstream-cloud.org/

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

# Jetstream storage

Storage built into the VM flavors = ephemeral (replicated, but…)

Volume storage = persistent (erasure coded, 4 data 2 recovery)

Valuable data should be on volumes (and backed up elsewhere)



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

# Some of the possibilities on Jetstream…

- True elastic computing
  - OpenStack Heat
  - OpenStack Magnum
  - Your own creation?
- Virtual clusters
  - Several gateways using virtual clusters
  - Working on a bigger and better long-term solution
  - Workshop at PEARC17 – (link in the tutorial walkthrough)
- Other possibilities
  - https://www.openstack.org/software/project-navigator/
  - Mistral (OSG) – cron as a service
  - Senlin (a coming attraction for making virtual clusters elastic-ness easier)
  - other additions like Manila (filesystems as a service), etc

Jetstream
http://jetstream-cloud.org/

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

# Left Twix, Right Twix

- Having two clouds = some semblance of fault tolerance
- Making that work for your gateway
    - Fail over possibilities
    - Load distribution/performance
    - Maximizing simultaneous VMs (VM/IP limits per cloud)

Jetstream
http://jetstream-cloud.org/

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

# But what about big data sets and such?

- At IU, Wrangler is nearby
- Jetstream wired to Wrangler's switch and vice versa
- Dedicated NFS node(s) on Wrangler for Jetstream
- Dedicated vlan built for projects that need access (still in beta, but working!)
- Hoping to replicate at TACC soon-ish

Jetstream
http://jetstream-cloud.org/

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

# Jetstream Fun: Happy cluster / Angry Cluster



Jetstream
http://jetstream-cloud.org/

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

# 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

Training Videos / Virtual Workshops (TBD)

Jetstream
http://jetstream-cloud.org/

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

# Jetstream Partners

INDIANA UNIVERSITY
PERVASIVE TECHNOLOGY INSTITUTE

JOHNS HOPKINS UNIVERSITY

THE UNIVERSITY OF ARIZONA

TACC

THE UNIVERSITY OF CHICAGO

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

# Discussion -

- What images would gateway developers like to see?

  - Is there a distribution and standard set of packages that would benefit many gateway developers?

- What features would gateway developers like to see?

- Other things?

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

# Questions?

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

Project email: help@jetstream-cloud.org Direct email: jeremy@iu.edu

## License Terms

- Fischer, J. (2017). Jetstream Overview: Gateways Conference. Retrieved from https://jetstream-cloud.org/publications.php
- Jetstream is supported by NSF award 1445604 (Craig Stewart, 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.

Jetstream
http://jetstream-cloud.org/

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

# And onward…



funded by the National Science Foundation
Award #ACI-1445604

# 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.

**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, backedup, and stored

Jetstream
http://jetstream-cloud.org/

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

# 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.

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



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

# Cloud Computing Terms…simplified (Cont.)

**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.

**Filesystem**: hierarchical in nature, directories & files, ability to open, seek, read, write.

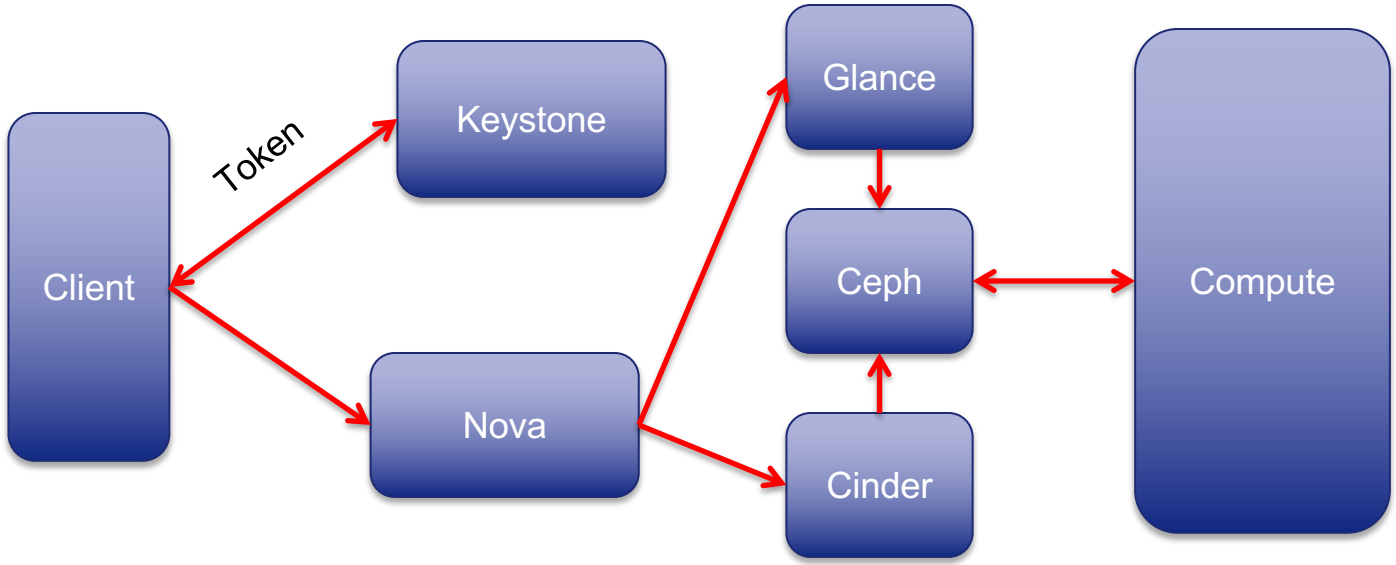
**Persistent storage**: If you pull the plug, it will still exist when power is restored. Safe to store data or state here.

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

Jetstream
http://jetstream-cloud.org/

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

# OpenStack Overview




Jetstream
http://jetstream-cloud.org/

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

# Getting into the hands on part -

Open https://goo.gl/8ke2fu 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*



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