

Jetstream Overview

Programmable Cyberinfrastructure: Clusters in the Cloud

Gateways 2018 – September 25-27, 2018 – Austin, TX

Jeremy Fischer – jeremy@iu.edu – Jetstream Senior Technical Advisor, UITS Research Technologies

Sanjana Sudarshan – ssudarsh@iu.edu – Jetstream Senior Technical Advisor, UITS Research Technologies

Eric Coulter – jecoulte@iu.edu – XSEDE Campus Bridging Engineer, UITS Research Technologies

Fischer, J., Sudarshan, S., Coulter J.E. (2018). Jetstream Overview - Programmable Cyberinfrastructure: Clusters in the Cloud. Austin, TX. Retrieved from <https://jetstream-cloud.org/research/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
- User-selectable library of preconfigured virtual machines
- Provides on-demand *interactive* computing and analysis or persistent gateways (SEAGrid, Galaxy, GenApp NAMDRunner, CIPRES and others)
- Enables *configurable* environments and **programmable cyberinfrastructure**
- Reproducibility: Share VMs and then store, publish via IU Scholarworks (DOI)

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

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...)

HPC vs Cloud

Adapting to a different environment:

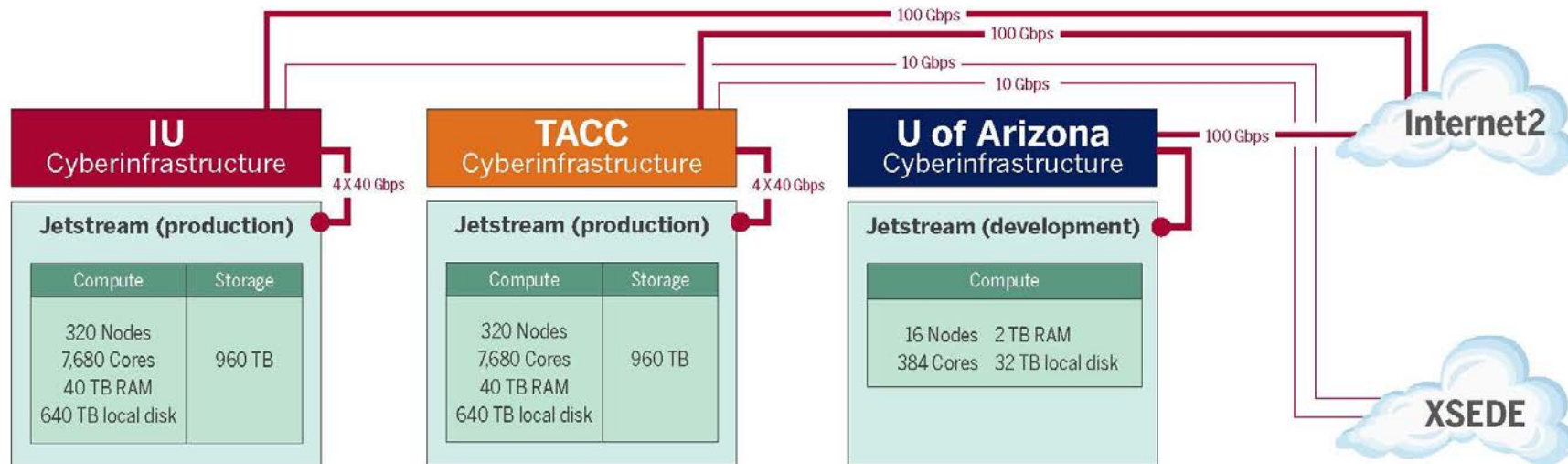
- No reservations, no queueing – more interactive usage
- Being your own admin – hey, we have root!**
- You really can have almost any (linux) software you want**
- Constantly getting new features (<https://www.openstack.org/software/project-navigator/>)

** Here there be dragons...

Jetstream and 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

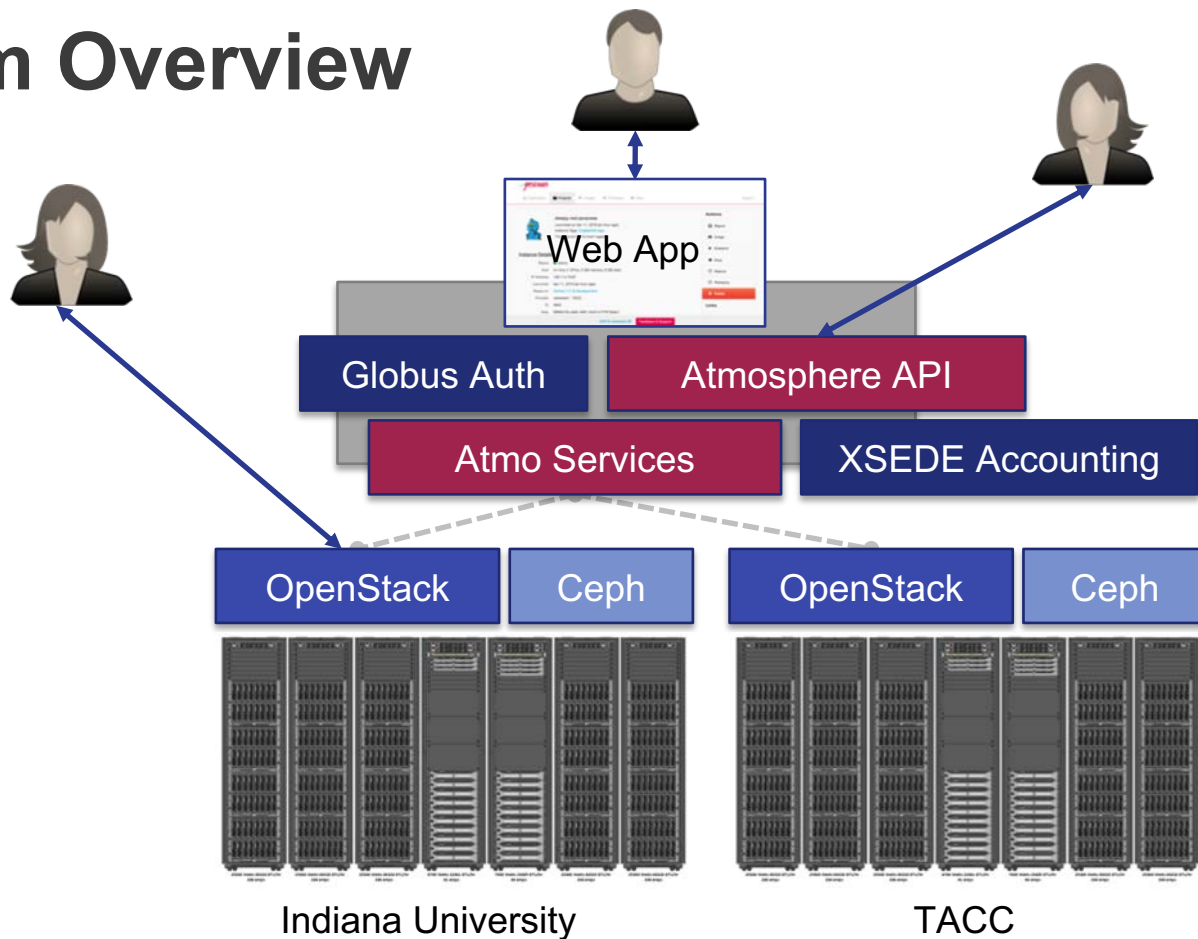
Jetstream System Overview



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

Platform Overview



Hardware and Instance "Flavors"

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

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/>



funded by the National Science Foundation
Award #ACI-1445604



Jetstream Partners



INDIANA UNIVERSITY
PERVASIVE TECHNOLOGY INSTITUTE



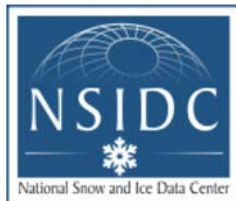
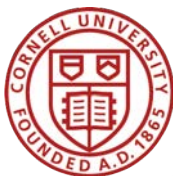
JOHNS HOPKINS
UNIVERSITY



THE UNIVERSITY
OF ARIZONA



THE UNIVERSITY OF
CHICAGO



UNC
THE ODUM INSTITUTE



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, Jeremy. September 25, 2018. Jetstream Overview: Programmable Cyberinfrastructure: Clusters in the Cloud. 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.



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

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

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.

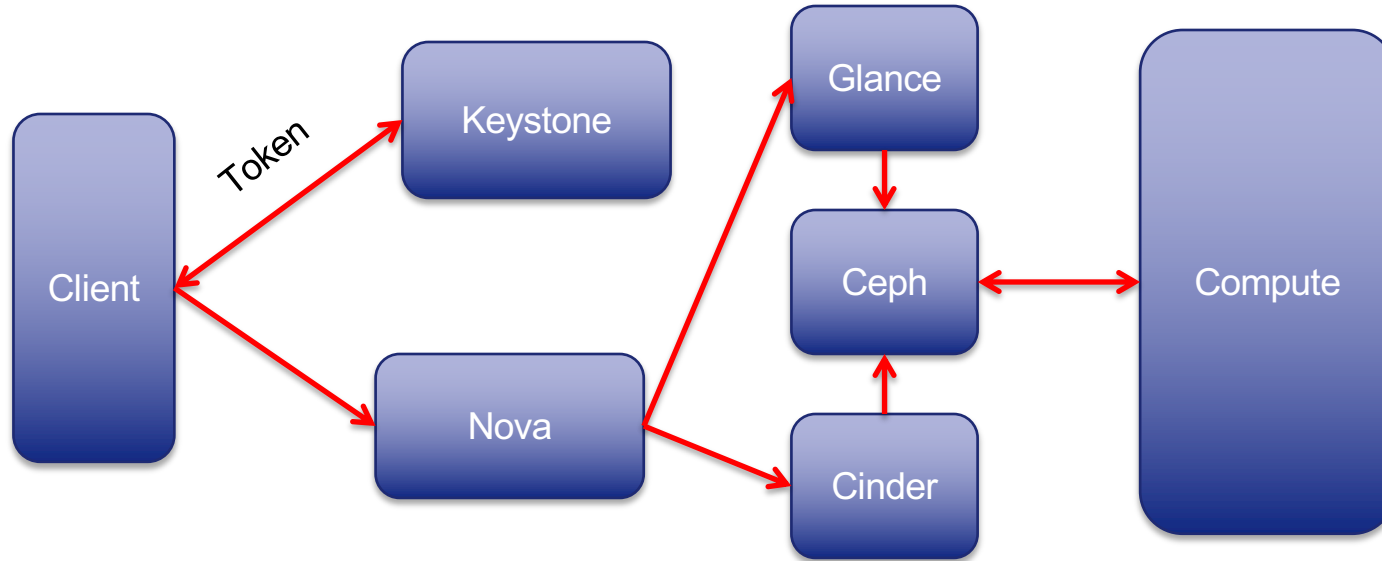
Block store: a software defined entity akin to an unformatted hardware disk drive.

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!!!)

OpenStack Overview



Getting into the hands on part -

Open <https://goo.gl/7X3HQz> 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