Jetstream Overview
Champions Edition

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

What we’ll cover today -

• What Jetstream is
• What Jetstream ISN’T
• What Atmosphere is – target audience
• What the API is and why gateway developers want to be there
• The basics we provide (images, networks/ips, etc)
• Ephemeral storage vs persistent storage on Jetstream
• Cattle not pets
• Some of the possibilities -- true elastic computing, virtual clusters, other things
• Utilizing the two distinct clouds to your benefit
• Using resources like Wrangler for maximizing use of Jetstream for large data sets and big data problems
• Discussion: What images/features would your constituents like to see? (Please don’t say Windows.) 😊
What is Jetstream and why does it exist?

• NSF’s first production cloud facility
• Part of the NSF eXtreme Digital (XD) program
• Provides on-demand *interactive* computing and analysis or persistent services such as gateways
• Enables *configurable* environments and *programmable cyberinfrastructure*
• User-selectable library of preconfigured virtual machines
• Focus on ease-of-use, broad accessibility
• Will support persistent gateways (SEAGrid, Galaxy, GenApp, and others)
• 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…)

funded by the National Science Foundation
Award #ACI-1445604
Jetstream System Overview

Jetstream (production)
- IU Cyberinfrastructure
- TACC Cyberinfrastructure
- U of Arizona Cyberinfrastructure

Jetstream (development)

Computes:
- IU: 320 Nodes, 7,680 Cores, 40 TB RAM, 640 TB local disk
- TACC: 320 Nodes, 7,680 Cores, 40 TB RAM, 640 TB local disk
- U of Arizona: 16 Nodes, 2 TB RAM, 384 Cores, 32 TB local disk

Storage:
- IU: 960 TB
- TACC: 960 TB
- U of Arizona: 32 TB local disk

Network Connections:
- 100 Gbps
- 10 Gbps
- 4 x 40 Gbps

Funded by the National Science Foundation
Award #ACI-1445604

http://jetstream-cloud.org/
Platform Overview

Web App

Globus Auth
Atmosphere API
Atmo Services
XSEDE Accounting
OpenStack
Ceph
OpenStack
Ceph

Indiana University
TACC
# 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*

## Hardware and Instance "Flavors"

<table>
<thead>
<tr>
<th>Flavor</th>
<th>vCPUs</th>
<th>RAM</th>
<th>Storage</th>
<th>Per Node</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1.tiny</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>46</td>
</tr>
<tr>
<td>m1.small</td>
<td>2</td>
<td>4</td>
<td>20</td>
<td>23</td>
</tr>
<tr>
<td>m1.medium</td>
<td>6</td>
<td>16</td>
<td>60</td>
<td>7</td>
</tr>
<tr>
<td>m1.large</td>
<td>10</td>
<td>30</td>
<td>60</td>
<td>4</td>
</tr>
<tr>
<td>m1.xlarge</td>
<td>24</td>
<td>60</td>
<td>60</td>
<td>2</td>
</tr>
<tr>
<td>m1.xxlarge</td>
<td>44</td>
<td>120</td>
<td>60</td>
<td>1</td>
</tr>
<tr>
<td>s1.large**</td>
<td>10</td>
<td>30</td>
<td>120</td>
<td>4</td>
</tr>
<tr>
<td>s1.xlarge**</td>
<td>24</td>
<td>60</td>
<td>240</td>
<td>2</td>
</tr>
<tr>
<td>s1.xxlarge**</td>
<td>44</td>
<td>120</td>
<td>480</td>
<td>1</td>
</tr>
</tbody>
</table>

** s1.* based instances are not eligible to be saved into a customized image
The Jetstream Atmosphere web interface
The Jetstream Atmosphere web interface

[Image of the Jetstream Atmosphere web interface]

- Centos 7 (7.2) Development GUI
- BioLinux 8
- Ubuntu 14.04.3 Development GUI
- Intel Development (CentOS 7)
- R with Intel compilers (CentOS 7)
- Galaxy Standalone

[Image of the Jetstream logo and website]

http://jetstream-cloud.org/
Look! It’s more Jetstream web interface!
Even more Jetstream web interface...
Using Jetstream as an infrastructure 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
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 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)
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.

-- George Turner (Jetstream architect)
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 PEARC18 – more info soon!

- Other possibilities
  - https://www.openstack.org/software/project-navigator/
  - Mistral (OSG) – cron as a service
  - Senlin (a coming attraction for making virtual clusters easier)
  - other additions like Manila (filesystems as a service), etc
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)
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 Fun: Happy cluster / Angry Cluster
Where can I get help?


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 Partners

funded by the National Science Foundation
Award #ACI-1445604

http://jetstream-cloud.org/
Discussion -

• What images would you like to see?
• Is there a distribution and standard set of packages that would benefit many developers/researchers outside of what’s provided?
• What features would developers/researchers like to see?
• Other things?
• Can send feedback to help@jetstream-cloud.org or to jeremy@iu.edu
Questions?

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

License Terms

- 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.