Jetstream - In production & supporting science
Available to everyone (including you, engineers! And you too, anyone with a biological collection)

Craig A. Stewart, Executive Director, IU Pervasive Technology Institute, Jetstream P.I.
David Y. Hancock, Program Director, Advanced Cyberinfrastructure, IU Pervasive Technology Institute
Jennifer Laherty, Associate Librarian, Head of Sciences, IU Libraries
Outline

- Overview of Jetstream including current status
- Info on orchestrated use of Jetstream and new cool tools (e.g. Docker, Mesos, Kubernetes, etc.)
- Info of interest to researchers working in engineering
- Example of Jetstream aiding a research area not typically supported by XSEDE: Imago, a tool for managing biological collections
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 aren’t they using XD/XSEDE systems?
• Activation energy is pretty high
• HPC resources are scarce and not well-matched to their needs
• They just don’t need that much capability

What is Jetstream?

- NSF’s first production cloud facility
- Part of the NSF eXtreme Digital (XD) program
- Provides on-demand interactive computing and analysis
- Enables configurable environments and architectures
- User-friendly, widely accessible cloud environment
- User-selectable library of preconfigured virtual machines

http://jetstream-cloud.org/
Jetstream System Overview

Jetstream (production) at IU, TACC, and U of Arizona Cyberinfrastructure

- **IU Cyberinfrastructure**
  - Jetstream (production)
  - Compute: 320 Nodes, 7,680 Cores, 40 TB RAM, 640 TB local disk
  - Storage: 960 TB

- **TACC Cyberinfrastructure**
  - Jetstream (production)
  - Compute: 320 Nodes, 7,680 Cores, 40 TB RAM, 640 TB local disk
  - Storage: 960 TB

- **U of Arizona Cyberinfrastructure**
  - Jetstream (development)
  - Compute: 16 Nodes, 384 Cores, 2 TB RAM, 32 TB local disk
  - Storage: 960 TB

Funded by the National Science Foundation Award ACI-1445604

http://jetstream-cloud.org/
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
- Using KVM Hypervisor

<table>
<thead>
<tr>
<th>Flavor</th>
<th>vCPUs</th>
<th>RAM</th>
<th>Storage</th>
<th>SU Per Hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>m.tiny</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>m.small</td>
<td>2</td>
<td>4</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>m.medium</td>
<td>6</td>
<td>16</td>
<td>60</td>
<td>6</td>
</tr>
<tr>
<td>m.large</td>
<td>10</td>
<td>30</td>
<td>120</td>
<td>10</td>
</tr>
<tr>
<td>m.xlarge</td>
<td>24</td>
<td>60</td>
<td>240</td>
<td>24</td>
</tr>
<tr>
<td>m.xxlarge</td>
<td>44</td>
<td>120</td>
<td>480</td>
<td>44</td>
</tr>
</tbody>
</table>

- Short-term storage comes as part of a launched instance
- Long-term storage mounted though block volumes
- Each user gets 10 volumes up to 500GB total storage
- Piloting object storage
Who uses Jetstream?

Researchers needing a handful of cores (1 to 44/vCPU)

Software creators and researchers needing to create their own customized virtual machines, containers, and workflows

Science gateway creators using Jetstream as either the front end or processor for scientific workflows

Educators teaching on a variety of subjects
The Jetstream Interface

funded by the National Science Foundation
Award #ACI-1445604

http://jetstream-cloud.org/
Methods of access

Primary methods of access

• Interactive user access via web interface with VNC/SSH
• Direct VNC/SSH to individual instances
• Direct OpenStack or Atmosphere API access
• API access enables Science Gateways and other *always on* services or *on demand* use cases; e.g. elastic compute techniques
You want to use Mesos, Docker, Kubernetes, whatever the latest and greatest new thing is? You want to run a science gateway?

• Two things keep happening to us:
  o We get asked “Do you support <new cool thing>?”
  o And then someone builds a VM image that makes use of <new cool thing>
• Several researchers are using Docker and Mesos already, and we are experimenting with Kubernetes.
• Orchestration engines make it easier to provision Jetstream resources behind gateways
• Supporting science gateways and encouraging orchestration engine use facilitates scaling applications and greater CPU usage
Science Domains Examples

Biology: iPlant and Galaxy VMs

Earth Science: NSIDC data analysis, EarthCube ECITE/CHORDS

Field Station Research: data collection and analysis tools to support data sharing and collaboration

Network Science: Network Workbench gateway and VMs

Social Sciences: VMs utilizing data from the Odum Institute (and others)

Computer Science/Cyberinfrastructure: RADICAL Tools, several education allocations

Whatever you do, probably… unless you run large scale MPI codes or need accelerators
Jetstream for engineering researchers

Matlab and the 52 standard toolkits are installed on Jetstream

You do **NOT** need to have a local license to use MATLAB on Jetstream

If you are a researcher, and MATLAB and one of the standard first 52 toolkits that come with MATLAB help you… **have at it!**

If you are an engineering researcher, and you need other tools... Let us know – we are happy to consider other requests

© Mathworks
Science/Engineering Applications

- Using OpenStack Heat to create a PBS cluster
- Ansible is used to dynamically install/configure applications
- Leveraging Apache Airavata GFac to submit workloads
- Supports apps such as:
  - Quantum Espresso
  - Gromacs
  - LAMMPS
  - NWChem

funded by the National Science Foundation
Award #ACI-1445604
• Shared-use research infrastructure funded by NSF to enable transformative research in natural hazards engineering
• Jobs run (via Agave) on HPC and Cloud systems
• VM sized applications packaged as Docker containers where appropriate for portability
• Using Jupyter with bursting capabilities to Jetstream to support high usage training
Supporting Education

Jetstream has been used in multiple IU Informatics Graduate Courses

• INFO 535 – Management, Access, and Use of Big and Complex Data
• INFO 590 – Topics in Informatics

BlueWaters Workflow Workshop

Multiple Research Data Alliance Workshops
An example of Jetstream enabling services for groups of scientists and citizens who never before would have thought of using an XSEDE-supported resource.
Introducing Imago

For Physical Specimen Biocollections

- Objects:
  Digital Library Preservation Repository
  2-D & 3-D images

- Metadata

2017
- Available on a VM in Jetstream

- Import/Export functions to facilitate workflows with NSF supported Specify and Symbiota

http://imago.indiana.edu/catalog
### Descriptions

<table>
<thead>
<tr>
<th>Attribute Name</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identifier</td>
<td><a href="http://purl.dlb.indiana.edu/udi/herbarium/IND-0010840">http://purl.dlb.indiana.edu/udi/herbarium/IND-0010840</a></td>
</tr>
<tr>
<td>Rights</td>
<td>Attribution-NonCommercial 3.0 United States</td>
</tr>
<tr>
<td>Collection code</td>
<td>herbarium</td>
</tr>
<tr>
<td>Catalog number</td>
<td>IND-0010840</td>
</tr>
<tr>
<td>Country</td>
<td>United States</td>
</tr>
<tr>
<td>State province</td>
<td>Indiana</td>
</tr>
<tr>
<td>County</td>
<td>Shelby</td>
</tr>
<tr>
<td>Kingdom</td>
<td>Plantae</td>
</tr>
<tr>
<td>Phylum</td>
<td>Magnoliophyta</td>
</tr>
<tr>
<td>Order</td>
<td>Liliales</td>
</tr>
<tr>
<td>Family</td>
<td>Melanthiaceae</td>
</tr>
<tr>
<td>Genus</td>
<td>Trillium</td>
</tr>
<tr>
<td>Specific epithet</td>
<td>recurvatum</td>
</tr>
<tr>
<td>Scientific name</td>
<td>Trillium recurvatum</td>
</tr>
<tr>
<td>Scientific name authorship</td>
<td>Beck</td>
</tr>
<tr>
<td>Basis of record</td>
<td>Preserved Specimen</td>
</tr>
</tbody>
</table>

*Image of a specimen of *Trillium recurvatum* from Indiana, United States.*
Journey: Metadata + Digital Images

Symbiota - Promoting Bio-Collaboration

GBIF - Global Biodiversity Information Facility

Imago

Specify - Software Project

iDigBio - Integrated Digitized Biocollections

EarthCube

USGS - Science for a Changing World

Biodiversity Information Serving Our Nation (BISON) - Explore & download U.S. species occurrence data & maps
Exploration of marine mammal paleogeography in the Northern Hemisphere over the Cenozoic using beta diversity

Fig. 1.
Geographical distribution of all 3535 marine mammal collections and division of ocean regions for this study. Each dot represents one collection. Solid lines represent ocean divisions by the International
You can put YOUR collection online
Via Imago and Jetstream
Broader Impacts

Help us train the next generation of Merlin Bird Photo ID

The Merlin team is hard at work on the next generation of photo ID tools. You can help us train the computer models that will help identify bird photos right on your phone. All you need to do is draw a box around the bird in each photo. Click Get Started to give it a try.
Current Partners

- Indiana University Bloomington (IU System)
- Indiana University Southeast (State)
- Huntington University (National)
- University of Wyoming

Jennifer Laherty, IU Libraries, jlaherty@indiana.edu
Jetstream Timeline...what comes next?

Transitioned to full operations on September 1, 2016

October 2016: **225 XSEDE projects and 650+ users**

Soliciting *Research* allocation requests quarterly plus *Startup* and *Education* allocations *NOW* – including Science Gateways!

Adding services as deemed useful/mature (heat, magnum, trove, manila, etc)

Atmosphere enhancements

Working on partnerships with groups like HubZero
Resource requests

You can request startup allocations anytime.

You can request allocations for educational use anytime.


We are happy to help you prepare a request and create a successful proposal.

You do not have to have prior use of Jetstream to be successful.
Where can I get help or learn more?

Production:

Wiki: 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 Partners

funded by the National Science Foundation
Award #ACI-1445604
Questions?

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

Project email:  help@jetstream-cloud.org or help@xsede.org

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.