National and Local Impacts of Programmable Cyberinfrastructure using Jetstream

David Y. Hancock – Indiana University
Director for Advanced Cyberinfrastructure
Jetstream Primary Investigator
Affiliated with IU Pervasive Technology Institute

November 14, 2018
What is “the” Jetstream?

- Fast moving air currents
- Hot/Cold air boundaries
- An NSF-funded cloud environment

NASA's Suomi NPP satellite collected this natural-color image using the VIIRS (Visible Infrared Imaging Radiometer Suite) instrument on Sept. 4, 2017. Actively burning areas are outlined in red. NASA image courtesy Jeff Schmaltz LANCE/EOSDIS MODIS Rapid Response Team, GSFC
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
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
Who uses Jetstream?

- The researcher needing a handful of cores (1 to 44/vCPU)
- 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’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...)
Platform Overview

Web App

Globus Auth
Atmosphere API
Atmo Services
XSEDE Accounting

OpenStack
Ceph
OpenStack
Ceph

Indiana University
TACC
The Jetstream Atmosphere web interface
The Jetstream Atmosphere web interface
<table>
<thead>
<tr>
<th>Discipline or area of interest</th>
<th># of Jetstream allocations</th>
<th>SUs allocated on Jetstream</th>
<th>% of SUs allocated on Jetstream</th>
<th>% of all SUs allocated on other XSEDE-supported systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Astronomy</td>
<td>2</td>
<td>1,108,096</td>
<td>3.04%</td>
<td>8.61%</td>
</tr>
<tr>
<td>Atmospheric Sciences</td>
<td>4</td>
<td>2,752,400</td>
<td>7.55%</td>
<td>3.73%</td>
</tr>
<tr>
<td>Biological Sciences</td>
<td>57</td>
<td>5,199,000</td>
<td>14.27%</td>
<td>4.95%</td>
</tr>
<tr>
<td>Campus/Domain Champions</td>
<td>123</td>
<td>6,105,500</td>
<td>16.76%</td>
<td>0.09%</td>
</tr>
<tr>
<td>Computational Science</td>
<td>11</td>
<td>1,150,000</td>
<td>3.16%</td>
<td>0.92%</td>
</tr>
<tr>
<td>Computer Science</td>
<td>15</td>
<td>4,944,302</td>
<td>13.57%</td>
<td>1.8%</td>
</tr>
<tr>
<td>Education Allocations</td>
<td>24</td>
<td>2,847,600</td>
<td>7.82%</td>
<td>0.01%</td>
</tr>
<tr>
<td>Engineering</td>
<td>1</td>
<td>100,000</td>
<td>0.27%</td>
<td>3.81%</td>
</tr>
<tr>
<td>Geosciences</td>
<td>10</td>
<td>1,978,400</td>
<td>5.43%</td>
<td>2.87%</td>
</tr>
<tr>
<td>Humanities/Social Sciences</td>
<td>10</td>
<td>560,000</td>
<td>1.54%</td>
<td>0.45%</td>
</tr>
<tr>
<td>Molecular Biosciences</td>
<td>8</td>
<td>4,647,520</td>
<td>12.75%</td>
<td>17.65%</td>
</tr>
<tr>
<td>Network Science</td>
<td>3</td>
<td>200,000</td>
<td>0.55%</td>
<td>0.06%</td>
</tr>
<tr>
<td>Ocean Science</td>
<td>3</td>
<td>230,000</td>
<td>0.63%</td>
<td>1.30%</td>
</tr>
<tr>
<td>Physics</td>
<td>4</td>
<td>2,252,400</td>
<td>6.18%</td>
<td>16.43%</td>
</tr>
<tr>
<td>Training &amp; Development</td>
<td>11</td>
<td>2,362,000</td>
<td>6.48%</td>
<td>0.16%</td>
</tr>
</tbody>
</table>
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, and 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...
Another Use Case: 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
Jetstream Gateway Highlights

• IRIS
  Serving large scale earthquake and geographical data for analysis

• Unidata
  Providing distribution and analysis of meteorological data

• OpenMRS
  Providing medical records systems for the resource-constrained

• SEAGrid
  Computational chemistry, molecular and fluid dynamics, and structural mechanics gateway

• NAMDRunner
  Based on the GenApp gateway

• ChemCompute Gateway
  Providing a computational chemistry gateway for educational use

Coming gateways: The Neuroscience Gateway, UltraScan III, and others
Jetstream usage highlights – 1 October 2018

• 413 active XSEDE projects covering 75 fields of science and **2558 active users** representing **190 institutions**
• **80%** of Jetstream users have **not used any other XSEDE system**
• >143M CPU hours allocated to XSEDE projects since June 2016

• 15 active science gateways
• 47 education/teaching allocations serving over 904 students
• 1151 (avg concurrent) active VMs in previous qtr, 955 in PY2 to date
• **Highest** user satisfaction in most recent XSEDE survey

*M&O PY2 to date 11/12 months*
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**
• You can request **startup** allocations **anytime**. (Startups are simple!)
• You can request allocations for **educational** use **anytime**.
• Next submission period for large allocations is 15 Dec 2018 – 15 Jan 2019
• Research allocation: Project desc (<10 pages) and Scaling doc (<5 pages)
  • We can help!
Expanding the reach: Jetstream REU Program

NSF Supplement for undergraduates
• 4 students participated in 2017
• 6 students participated in 2018
• Planning for 6 in 2019

• REU student videos on YouTube
  https://www.youtube.com/user/IUPTI
Where can I get help?

- User guides: [https://portal.xsede.org/user-guides](https://portal.xsede.org/user-guides)
- XSEDE KB: [https://portal.xsede.org/knowledge-base](https://portal.xsede.org/knowledge-base)
- Email: help@xsede.org
- Campus Champions: [https://www.xsede.org/campus-champions](https://www.xsede.org/campus-champions)
- Introduction to Jetstream Virtual Workshop: [https://cvw.cac.cornell.edu/jetstream/](https://cvw.cac.cornell.edu/jetstream/)
- Jetstream Allocations Virtual Workshop: [https://cvw.cac.cornell.edu/JetstreamReq/](https://cvw.cac.cornell.edu/JetstreamReq/)
Jetstream partners

funded by the National Science Foundation
Award #ACI-1445604

http://jetstream-cloud.org/