Prepared for GlobusWorld 2015

14 April 2015

Kurt A. Seiffert
Enterprise Architect, Indiana University – Research Technologies

funded by the National Science Foundation
Award #ACI-1445604
What is Jetstream?

• NSF’s first cloud for science and engineering research across all areas of activity supported by the NSF
• Jetstream will be a user-friendly cloud environment designed to give researchers and research students access to self-provisioned computing and data analysis resources.
• Jetstream will leverage Globus tools for data movement and authentication
• It will provide a user-selectable library of domain-oriented virtual machines that users can select from to do their research.
Who uses Jetstream

• For the researcher needing a handful of cores TODAY rather than thousands next week.
• Software creators and researchers needing to create their own customized virtual machines -or- their own “private computing system” within Jetstream.
• It will enable countless discoveries across disciplines such as biology, atmospheric science, economics, network science, observational astronomy, and social sciences.
What does the name mean? And it is it really a cloud?

- **Name**
  - In the atmosphere the Jetstream lies at the border of two different air masses
  - The Jetstream system stands at the border of the existing NSF-funded XD program and advanced cyberinfrastructure resources and users who have not previously used such NSF funded infrastructure before.

- Yep, it’s really a cloud, or at least a cloud environment (one could quibble over the definition of cloud vis-à-vis expansability). Software layers:
  - Atmosphere interface
  - KVM
  - OpenStack
  - CentOS Linux
### VM Instance Sizes

<table>
<thead>
<tr>
<th>Instance Type</th>
<th>vCPUs (48 total)</th>
<th>RAM (128 GB total)</th>
<th>Storage (2048 GB total)</th>
<th>Instances/Node</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tiny</td>
<td>1</td>
<td>2</td>
<td>20</td>
<td>46</td>
</tr>
<tr>
<td>Small</td>
<td>2</td>
<td>4</td>
<td>40</td>
<td>23</td>
</tr>
<tr>
<td>Medium</td>
<td>6</td>
<td>16</td>
<td>130</td>
<td>7</td>
</tr>
<tr>
<td>Large</td>
<td>10</td>
<td>30</td>
<td>230</td>
<td>4</td>
</tr>
<tr>
<td>X-Large</td>
<td>22</td>
<td>60</td>
<td>460</td>
<td>2</td>
</tr>
<tr>
<td>XX-Large</td>
<td>44</td>
<td>120</td>
<td>920</td>
<td>1</td>
</tr>
</tbody>
</table>

Node config: 2 Intel 2680 v3 “Haswell.” 2.5 GHz base frequency. Floating point intensive operations utilizing the AVX instruction set run at 2.1 GHz.
Science Domains and Users

- Biology
- Earth Science/Polar Science
- Field Station Research
- Geographical Information Systems
- Network Science
- Observational Astronomy
- Social Sciences
- Jetstream will be particularly focused on researchers working in the “long tail” of science with born-digital data
- Enabling analysis of field-collected empirical data on the impact and effects of global climate change will be one of the specific foci of Jetstream
- Whatever you do
Types of applications supported

- Interactive, VM-based work
- Persistent science gateways
- Hadoop at modest scale
21st century workforce development

- Jetstream will include virtual Linux desktops and applications specifically aimed at enabling research and research education at small colleges and universities including HBCUs (Historically Black Colleges and Universities), MSIs (Minority Serving Institutions), Tribal colleges, and higher-ed institutions in EPSCoR (Experimental Program to Stimulate Competitive Research) States
- Jetstream will also support deployment of user-friendly Science Gateways
Jetstream System Diagram

Jetstream (production)
- Compute: 320 Nodes, 7680 Cores, 40 TB RAM, 640 TB local disk
- Storage: 960 TB

TACC Cyberinfrastructure
- Compute: 320 Nodes, 7680 Cores, 40 TB RAM, 640 TB local disk
- Storage: 960 TB

U of Arizona Cyberinfrastructure
- Compute: 16 Nodes, 2 TB RAM, 384 Cores, 32 TB local disk

Connections:
- 4x40 Gbps
- 100 Gbps
- 10 Gbps
Jetstream Deployment Partner Organizations

A seasoned team of organizations and experts:

- University of Texas Austin (TACC)
- University of Chicago (Argonne National Lab)
- University of Arizona
- University of Texas at San Antonio (Open Cloud Lab)
- Johns Hopkins University
- Pennsylvania State University
Jetstream Deployment Partner Organizations

- Initial construction (funded partners):
  - University of Texas Austin (TACC)
  - University of Chicago (Argonne National Lab)
  - University of Arizona
  - Johns Hopkins University

- Planned funded partners for O&M phase:
  - University of Texas at San Antonio (Open Cloud Lab)
  - Penn State University
  - Cornell University

- Several unfunded collaborating partners:
  - University of Hawaii
  - University of Arkansas Pine Bluff
  - National Snow and Ice Data Center
  - University of North Carolina Odum Center
Jetstream Application & Outreach Collaborators

- Cornell University – Ms. Susan Mehringer, Lead. Cornell® Virtual Workshops about Jetstream and applications running on jetstream.
- University of Arkansas at Pine Bluff – Dr. Jesse Walker, lead. cybersecurity education, Minority Serving Education outreach.
- University of Hawaii – Dr. Gwen Jacobs, lead. EPSCoR early adopter/user. Jacobs will chair Science Advisory Board.
- National Snow and Ice Data Center (NSIDC) – Dr. Ron Weaver, lead. Data retrieval from NSIDC, application integration with ice sheet analysis applications.
- University of North Carolina, Odum Center – Dr. Thomas Carsey, lead. Data retrieval from Dataverse Network.
- National Center for Genome Analysis at Indiana University – providing genome analysis software. Includes TACC, PSC, and SDSC as partners.
Timeline

- Test gear arriving this quarter
- Production gear in mid-summer
- Friendly user mode prior to SC
- Production operation status by January 2016
Questions

• Follow-up questions may be sent via email to stewart@iu.edu