A national science & engineering cloud

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

- NSF’s first cloud dedicates top support 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 interactive computing and data analysis resources “on demand.”
- It will provide a user-selectable library of virtual machines that users can select from to do their research.
- Software creators and researchers will also be able 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.
- Two especially important biology platforms will be supported – iPlant and Galaxy.
What does the name mean? And 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. Software layers:
  – Atmosphere interface
  – KVM
  – OpenStack
  – CentOS Linux
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
21st century workforce development

- Jetstream will include virtual Linux desktops and applications specifically aimed to enable 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 States
- Jetstream will also support deployment of user-friendly Science Gateways
# Jetstream Hardware Components

<table>
<thead>
<tr>
<th>Jetstream Site</th>
<th>#CPUs</th>
<th># Physical Cores</th>
<th>PFLOPS</th>
<th>Total RAM (GB)</th>
<th>Node Local Storage (TB)</th>
<th>Secondary Storage (TB)</th>
<th>Connection to Internet2 (Gbps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Production Systems</td>
<td>IU</td>
<td>640</td>
<td>7,680</td>
<td>0.258</td>
<td>40,960</td>
<td>640</td>
<td>960</td>
</tr>
<tr>
<td></td>
<td>TACC</td>
<td>640</td>
<td>7,680</td>
<td>0.258</td>
<td>40,960</td>
<td>640</td>
<td>960</td>
</tr>
<tr>
<td>Test &amp; Development System</td>
<td>Arizona</td>
<td>32</td>
<td>384</td>
<td>0.013</td>
<td>2,048</td>
<td>32</td>
<td>192</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>1,312</td>
<td>15,744</td>
<td>0.529</td>
<td>83,968</td>
<td>1,312</td>
<td>2,112</td>
</tr>
</tbody>
</table>
News for software developers

- Jetstream is *enabling* cyberinfrastructure
- RESTful APIs
- You build, package, deploy
- Users run on NSF-funded hardware
- Can leverage Globus technology for data movement and authentication
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
- Penn State University
Indiana University – Lead

- Dr. Craig Stewart, Principal Investigator
- Site for Dell production hardware
- Primary data storage facility
- IU School of Informatics & Computing faculty participants:
  - Dr. Beth Plale – Hathi Trust Research Center applications; Big Data
  - Dr. Katy Börner – Network science applications
  - Dr. Volker Brendel – Bioinformatics, computational science education
- IUPUI Computer & Information Science faculty participants:
  - Dr. Fengguang Song – CPU- and node- level performance
University of Texas Austin – Texas Advanced Computing Center

- Dr. Matthew Vaughn, Co-Principal Investigator
- Site for Dell production hardware
- Primary data storage facility
University of Chicago – Computation Institute

- Dr. Ian Foster, Co-Principal Investigator
- Integrate Globus services into Jetstream
University of Arizona

- Dr. Nirav Merchant, Co-Principal Investigator
- Dell hardware test/development site
- Lead Atmosphere implementation
- Bryan Heidorn participating in VMs related to field research stations
The University of Texas at San Antonio

- Dr. Paul Rad, Open Cloud Lab
- OpenStack software integration for cloud environment
Johns Hopkins University & Penn State University

- Dr. James Taylor, Co-Principal Investigator (JHU)
- Dr. Anton Nekrutenko, PSU
- Lead implementation of Galaxy software
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
Questions?

• Questions can be e-mailed to pti@iu.edu
• Visit the project website - pti.iu.edu/jetstream