Hands on with Jetstream

Jetstream: A national research and education cloud

Jeremy Fischer (jeremy@iu.edu)
ORCID 0000-0001-7078-6609
Senior Technical Advisor
UITS Research Technologies

I2 Tech Exchange – September 27, 2016

NSF Funding Areas in HPC

- Traditionally concentrated on enabling petascale capability
  - Blue Waters – 13.3 petaflops, 2012
  - Stampede – 9.6 petaflops, 2013
  - Comet – ~2.0 petaflops, 2014
- Has funded research into building clouds and computer science
  - CloudLab
  - Chameleon
- Now funding clouds to do research
  - Bridges (Hybrid system)
  - 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 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 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

* https://www.xsede.org/xsede-nsf-release-cloud-survey-report

What is Jetstream?

- NSF's first production cloud facility
- Part of the NSF eXtreme Digital (XD) program
- Provide on-demand *interactive* computing and analysis
- Enable *configurable* environments and architectures
- User-friendly, widely accessible cloud environment
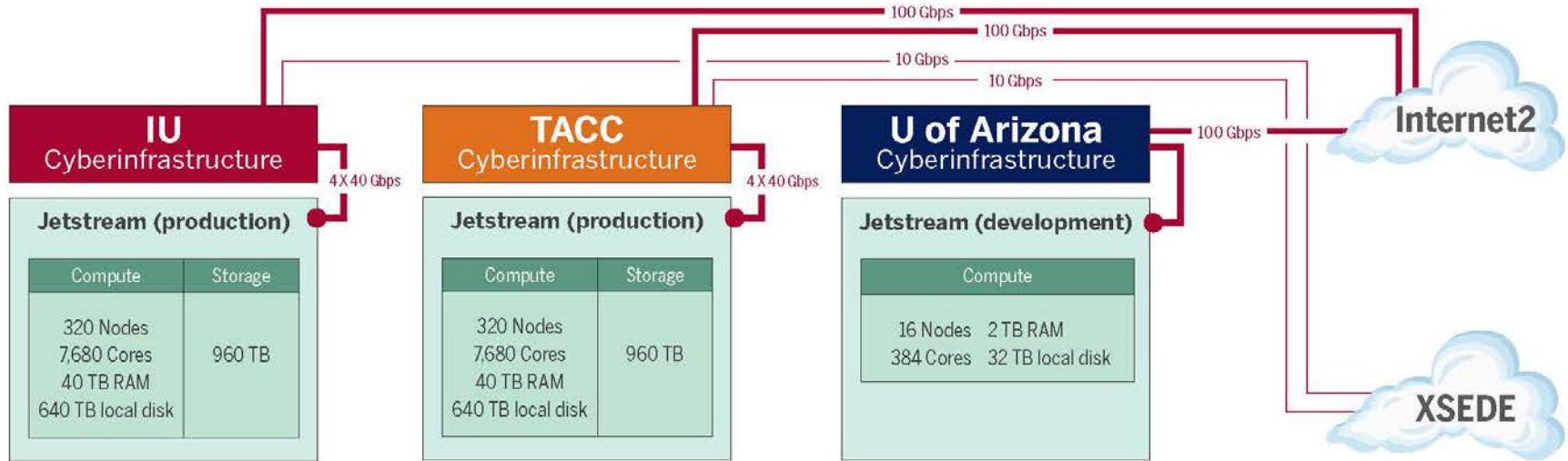- User-selectable library of preconfigured virtual machines
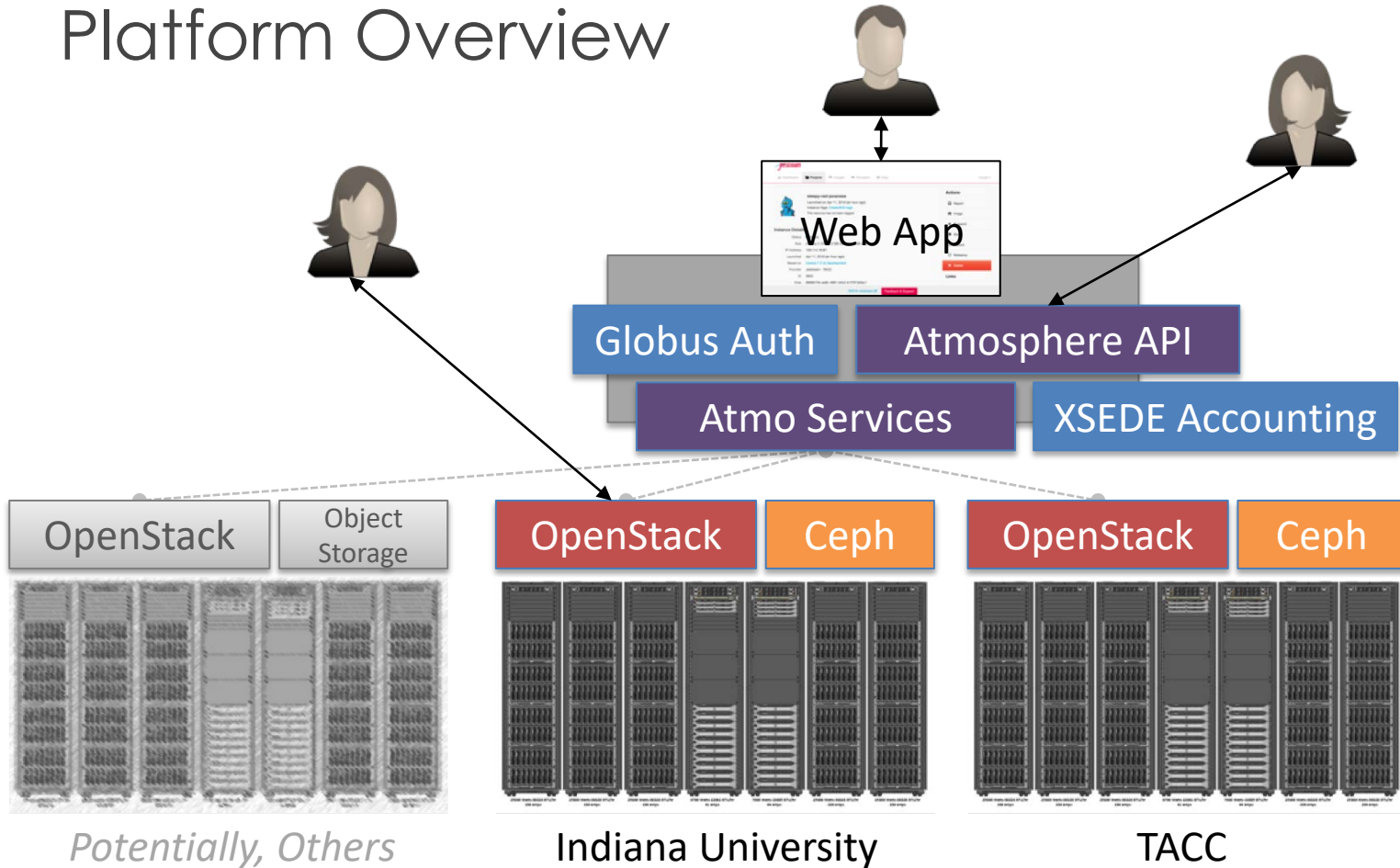
21st century workforce development

- Specialized virtual Linux desktops and applications to enable research and research education at small colleges and universities

- HBCUs (Historically Black Colleges and Universities)

- MSIs (Minority Serving Institutions)

- Tribal colleges

- Higher-education institutions in EPSCoR States

Jetstream System Overview

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 | Object Storage

OpenStack | Ceph

OpenStack | Ceph

Potentially, Others   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

Flavor	vCPUs	RAM	Storage	Per Node
m.tiny	1	2	8	46
m.small	2	4	20	23
m.medium	6	16	60	7
m.large	10	30	120	4
m.xlarge	24	60	240	2
m.xxlarge	44	120	480	1

- Short-term storage comes as part of launched instance
- Long-term storage is XSEDE-allocated
- Implemented on backend as OpenStack Volumes
- Each user gets 10 volumes up to 500GB total storage
- Piloting object storage as well after recent update

Who will use Jetstream?

- For the researcher 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 frontend or processor for scientific jobs
- STEM Educators teaching on a variety of subjects

Levels of access

Two levels of access

- Interactive user access via web interface and vnc/ssh
- Persistent access for Science Gateways and other "always on" services or services launched programmatically on demand; e.g. elastic compute techniques

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 HTC workloads!
- [Let's add yours here!]

What do you optimize for?

- HPC
  - Utilization
  - Capability or Capacity Science
  - Checkpoint/Restart I/O
  - Memory/Network Bandwidth & Latency
- Cloud
  - Availability
  - Multi-level API Interactions
  - On-demand/Interactive Use
  - Using Commodity Components

Reservations & Queueing

- HPC
  - Staples of the HPC world with powerful tools (e.g. Moab/Slurm)
  - Decades of expertise and tuning
  - Condo computing "anti-batch"
- Cloud
  - No reservations, no queueing, refocus
    - Some opposition to these concepts
  - Reserved instances "anti-cloud"
  - However… factions in OS community still pushing for do what AWS does

The Jetstream Interface

http://jetstream-cloud.org/

Supporting Jetstream Users

http://jetstream-cloud.org/
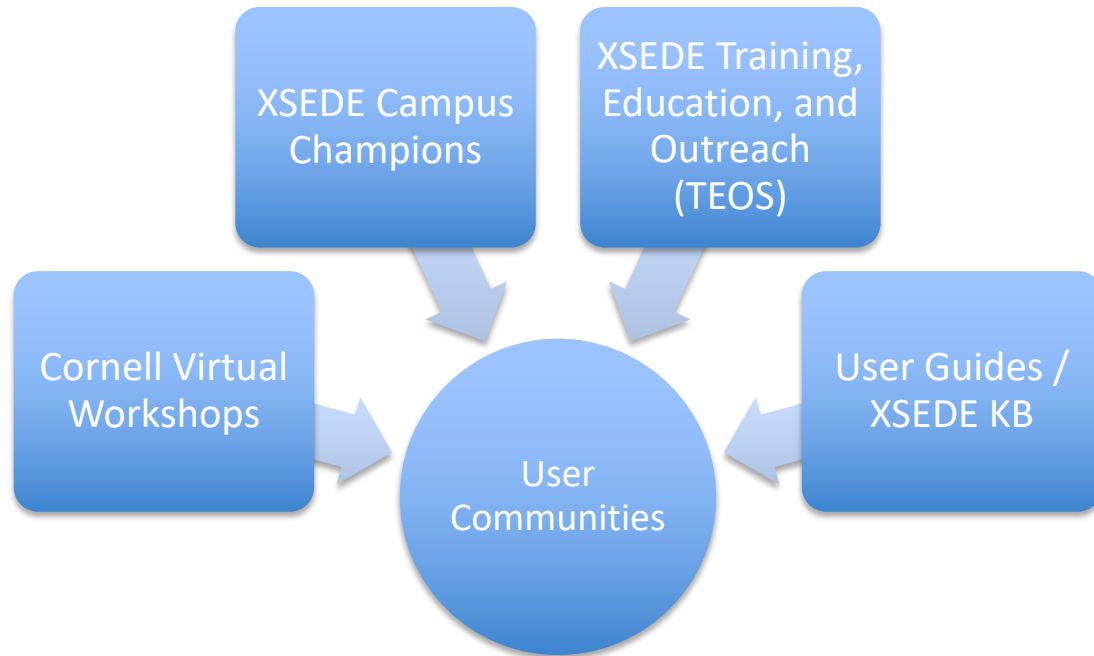
Jetstream Timeline…what comes next?

- Transitioned to full operations post-acceptance review on September 1, 2016
- September 2016: **125 XSEDE projects and 500+ users**
- Soliciting *Research* allocation requests NOW plus *Startup* and *Education* allocations – including Science Gateways!
- Adding services as deemed useful/mature (heat, ceilometer, magnum, trove, manila, etc)
- Atmosphere enhancements
- Working on partnerships with groups like HubZero

http://jetstream-cloud.org/

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)

Just for fun: Happy Cluster – Mad Cluster

http://jetstream-cloud.org/

Jetstream Partners

http://jetstream-cloud.org/
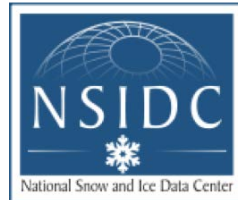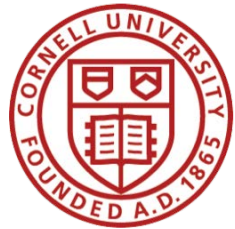
What are we going to cover today in the hands on portion? (The 40,000 foot view…)

- Getting on to Jetstream's web interface
- The Dashboard – features, settings, options
- Images on Jetstream today and in the future
- Projects and the resources they contain
- Launching an instance
- Accessing, using, and customizing your instance
- Talking about custom images
- Preserving your image for publication

Questions?

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

License Terms

- Fischer, Jeremy. September, 2016. Hands on with Jetstream – I2 Tech Exchange Conference. Available at: http://jetstream-cloud.org/publications.php
- 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.

http://jetstream-cloud.org/

NSF
funded by the National Science Foundation
Award #ACI-1445604