Jetstream Overview
A national research and education cloud
Great Plains Networks Annual Meeting

Jeremy Fischer – jeremy@iu.edu

Senior Technical Advisor,
UITS Research Technologies

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

http://jetstream-cloud.org/

Jetstream - Expanding NSF XD's reach and impact

Lots of stats below –

tl;dr summary: no one has enough computing resources. Ever. But they need easy access and use.

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

http://jetstream-cloud.org/

What is Jetstream and why does it exist?

- 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 *programmable cyberinfrastructure*

- User-friendly, widely accessible cloud environment

- User-selectable library of preconfigured virtual machines

What is Jetstream, continued…

- Focus on ease-of-use, broad accessibility

- Command line access for those who want it and GUI access for those who don't

- Will support persistent gateways (SEAGrid, Galaxy, GenApp NAMDRunner, CIPRES and others)

- Reproducibility: Share VMs and then store, publish via IU Scholarworks (DOI)

http://jetstream-cloud.org/

Who uses Jetstream?

- The researcher needing a handful of cores (1 to 44/vCPU)

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

- Science gateway creators using Jetstream as either the frontend or processor for scientific jobs

- STEM Educators teaching on a variety of subjects

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

http://jetstream-cloud.org/

Platform Overview

Web App

Globus Auth

Atmosphere API

Atmo Services

XSEDE Accounting

OpenStack

Ceph

OpenStack

Ceph

Indiana University

TACC
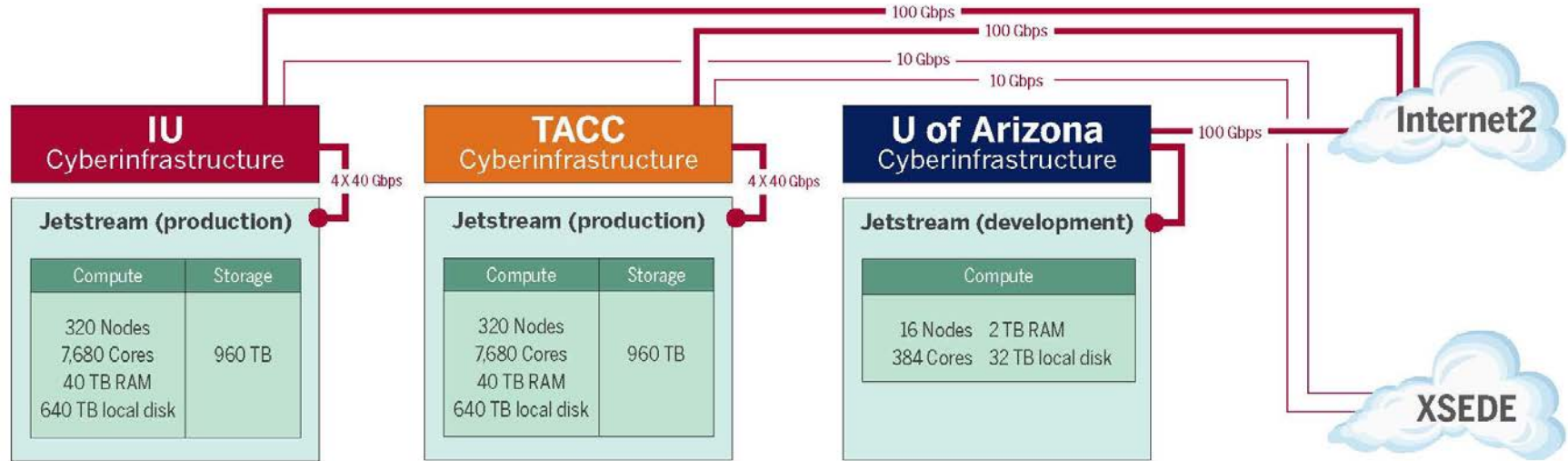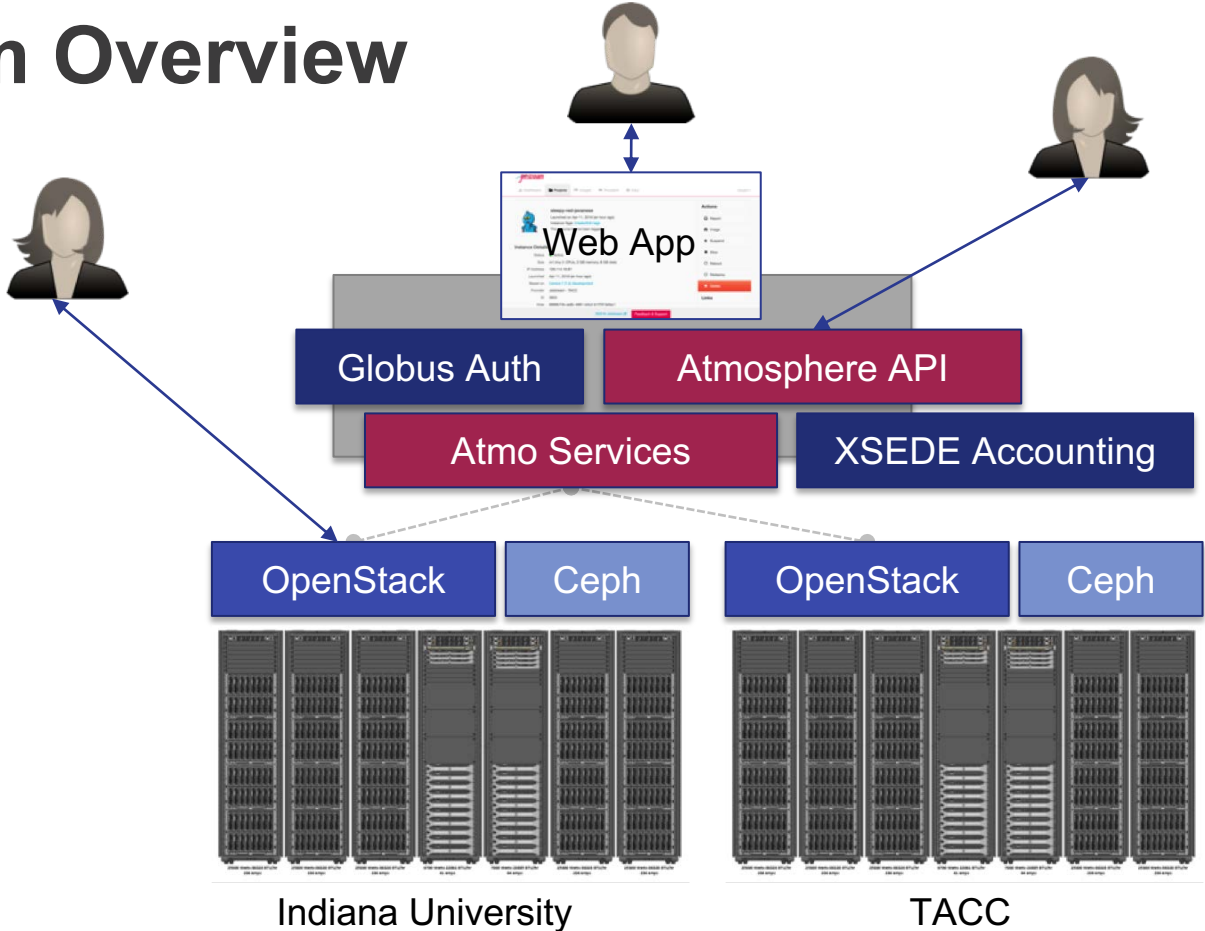
Hardware and Instance "Flavors"

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

- Short-term *ephemeral* storage comes as part of launched instance
- Long-term storage is XSEDE-allocated
- Implemented as OpenStack Volumes
- Each user can get 10 volumes up to 500GB total storage*

Flavor	vCPUs	RAM	Storage	Per Node
m1.tiny	1	2	8	46
m1.small	2	4	20	23
m1.medium	6	16	60	7
m1.large	10	30	60	4
m1.xlarge	24	60	60	2
m1.xxlarge	44	120	60	1
s1.large**	10	30	120	4
s1.xlarge**	24	60	240	2
s1.xxlarge**	44	120	480	1

** s1.* based instances are not eligible to be saved into a customized image

jetstream
http://jetstream-cloud.org/

NSF

IU

The Jetstream Atmosphere web interface

The Jetstream Atmosphere web interface

Look! It's more Jetstream web interface!

http://jetstream-cloud.org/

Even more Jetstream web interface…

Using Jetstream VMs

Manipulating Jetstream VMs:
- Jetstream Atmosphere web interface
- Direct API access via OpenStack command line or Horizon access
  - API access enables Science Gateways and other always on services or on demand use cases; e.g. elastic compute techniques

Primary methods of logging into Jetstream VMs to work
- Interactive user access via web interface with VNC/SSH
- Direct VNC/SSH to individual instances

HPC vs Cloud

Adapting to a different environment:

• No reservations, no queueing

• More interactive use and less/no batch queuing

• What? No parallel filesystem?!?

• Being your own admin – hey, we have root!

• You really can have almost any (linux) software you want**

• Constantly getting new features (https://www.openstack.org/software/project-navigator/)

** Here there be dragons…

Thinking about VMs…

Cattle, not pets: pets take great amount of care, feeding, and you name them; cattle you intend to have high turnover and you give them numbers.

-- Mike Lowe (Jetstream architect)

Jetstream for engineering researchers (and others)

Matlab and 52 standard toolkits are installed on Jetstream

You do NOT need to have a local license to use MATLAB on Jetstream
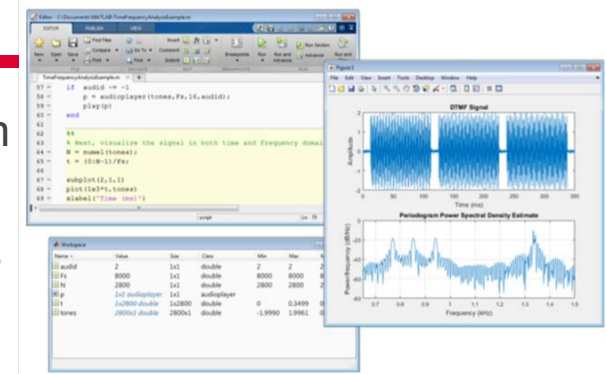
If you are a researcher, and MATLAB and one of the standard first 52 toolkits that come with MATLAB help you… you're ready to go!

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

jetstream
http://jetstream-cloud.org/

Jetstream for Education

Jetstream has been used in multiple IU Informatics Graduate Courses

- INFO 535 – Management, Access, and Use of Big and Complex Data (200 students!)
- INFO 590 – Topics in Informatics

BlueWaters Workflow Workshop

Multiple Research Data Alliance Workshops
Multiple workshops/classes on Galaxy

Ongoing classes/workshops on data analysis in finance using R, security and intrusion detection, and principles in cloud computing

Another Use Case: Galaxy riding Jetstream

Galaxy is a platform for biomedical research, focused on accessibility, transparency and reproducibility
- The main project instance (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 (e.g., new tools)

Use Jetstream as a *bursting* platform

- From Galaxy Main, offload jobs onto a remote Slurm cluster running on Jetstream instances

- Run Galaxy Interactive Environments (i.e., Dockerized IPython/RStudio containers) in an isolated environment on 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

Allocation types and docs needed for each

- Startup allocation (apply anytime) – "Fog a mirror"

  - Current CV for PI and any Co-Pis
  - Brief abstract/description of work

- Education allocation (apply anytime) – "Fog a mirror +"

  - Current CV for PI and any Co-PIs
  - Syllabus/Class/Workshop description
  - Description of use --> justification of SUs requested

- Research allocation (quarterly allocation window) – "A wee bit more than fog a mirror"

  - Current CV for PI and any Co-PIs
  - Main project description (up to 10 pages unless > 15M SUs, then 15 pages)
  - Scaling doc (up to 5 pages)

Requesting access to Jetstream

- 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 June 2017– 15 July 2017.
- 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.

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

http://jetstream-cloud.org/

Jetstream Project 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 – over 1 million computing hours used to date for MD

- Coming gateways: CIPRES Gateway, The Neuroscience Gateway, ChemCompute gateway, UltraScan III

Jetstream Overall Highlights

As of May 1, 2017:

- 290 active XSEDE projects covering 55 fields of science and 1700+ active users representing 167 institutions
- Over 56 million CPU hours allocated to XSEDE projects since June 2016
- 13 science gateways
- 24 education/teaching allocations
- Almost 600 undergraduate and graduate students using Jetstream

Jetstream Timeline…what comes next?

- Transitioned to full operations on September 1, 2016
- 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 YOURS!

Jetstream Fun: Happy cluster / Angry Cluster

http://jetstream-cloud.org/

Where can I get help?

Wiki / Documentation: 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
http://jetstream-cloud.org/

Jetstream Partners
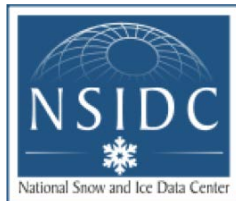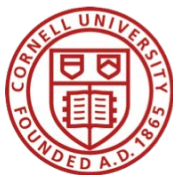
funded by the National Science Foundation
Award #ACI-1445604

http://jetstream-cloud.org/

Questions?

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

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

License Terms

Jetstream
http://jetstream-cloud.org/

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