Jetstream and the future of cloud computing in HPC

David Hancock – dyh@iu.edu

Acting Principal Investigator – Jetstream
Program Director – Advanced Cyberinfrastructure
Indiana University Pervasive Technology Institute

SC17 Student Program

November 12, 2017

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
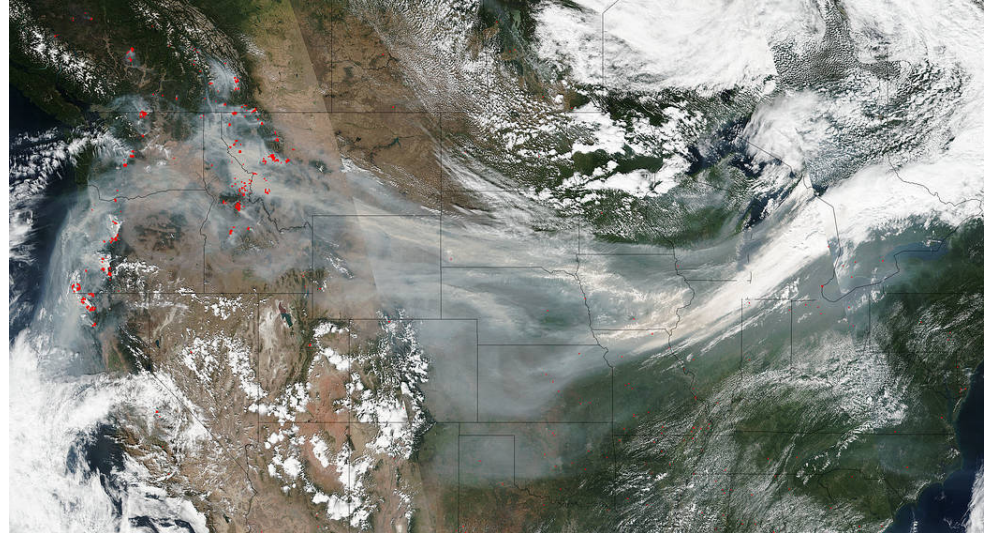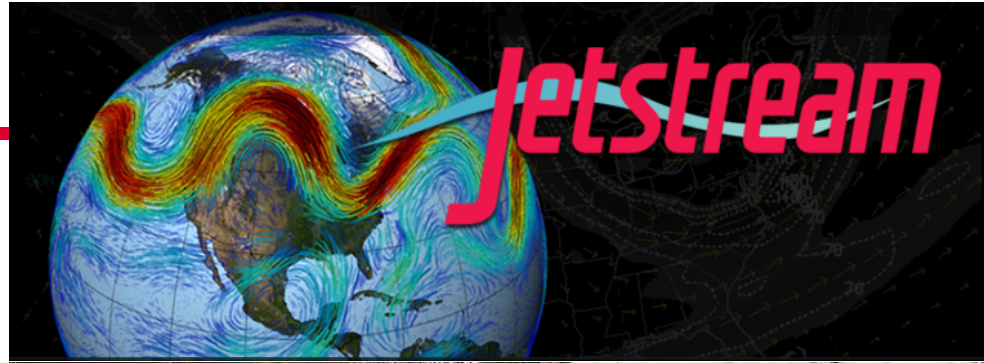
National Science Foundation – Funding in HPC

- Traditionally concentrated on enabling peta-scale capability via track I/II programs
  - Blue Waters – 13.3 petaflops, 2012 (under re-compete)
  - Stampede – 9.6 petaflops, 2013 (extended to Stampede 2)
  - Comet – ~2.0 petaflops, 2014
- Have funded research into building clouds and computer science testbeds
  - CloudLab
  - Chameleon
- Now funding clouds to do research
  - Bridges (Hybrid system)
  - Jetstream

Jetstream
http://jetstream-cloud.org/

What is Jetstream and why does it exist?

- NSF's first production cloud facility

- Focus on ease-of-use, broad accessibility

- Encourage collaboration and community development

- User-selectable library of preconfigured virtual machines

- Provides on-demand *interactive* computing and analysis or persistent services such as gateways (e.g. SEAGrid, Galaxy, GenApp, and others)

- Enables *configurable* environments and **programmable cyberinfrastructure**

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

http://jetstream-cloud.org/

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/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/

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

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…)

http://jetstream-cloud.org/

Jetstream and way of the cloud…

- **Cloudy Technologies**: clouds are more the just virtual machines (VM)
  - **Old way**: robust (expensive) infrastructure, weak (cheap) software
    - You expect the hardware to not fail
    - State in maintained in volatile data structures
  - **Cloudy way**: commodity infrastructure, robust software
    - Expect & plan for infrastructure to fail
    - Put intelligence into the software to handle infrastructure failure
  - **And my favorite…**

http://jetstream-cloud.org/

Thinking about VMs...

Flickr user Nanak26 - Normandie

- 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.
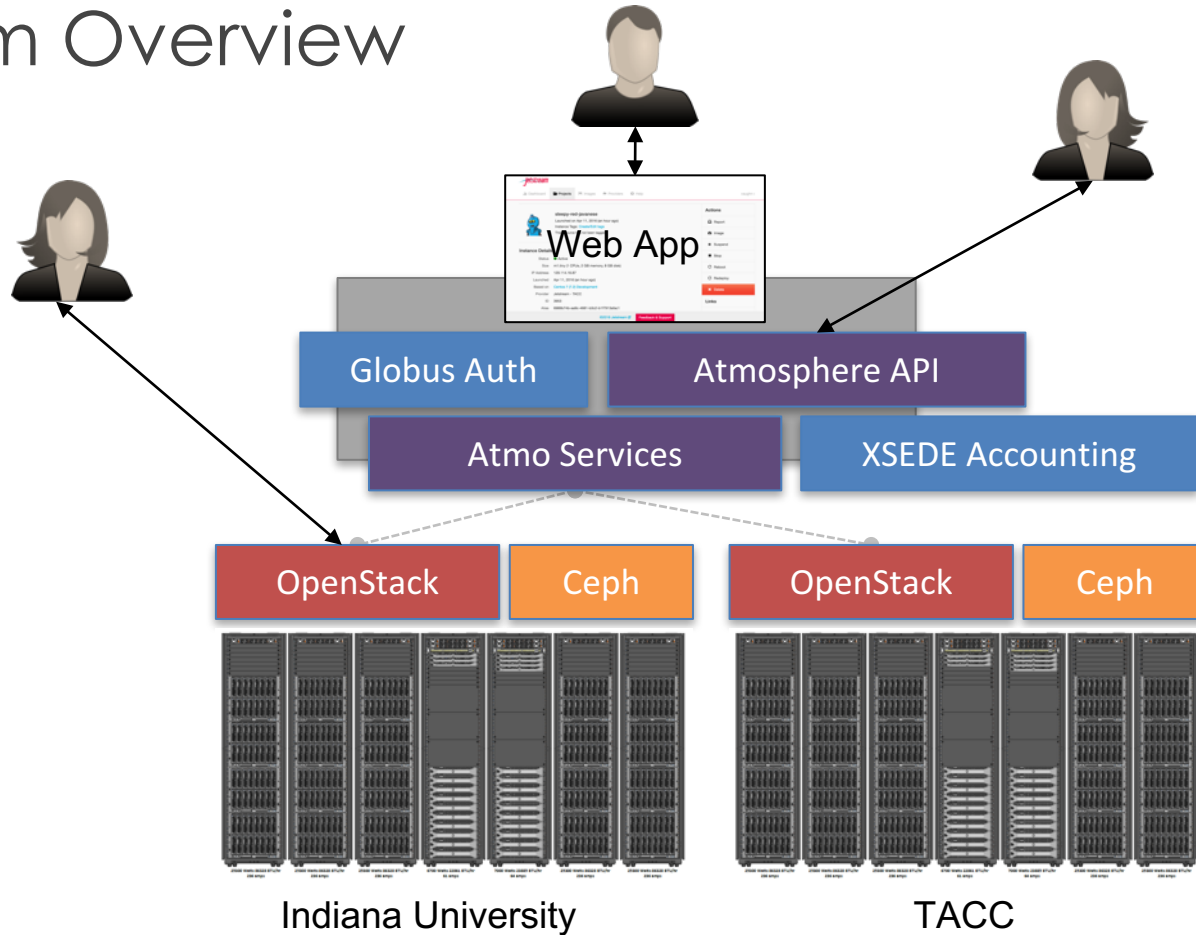
-- Mike Lowe (Jetstream architect)

** Some caveats for gateways, but...

jetstream
http://jetstream-cloud.org/

Jetstream System Overview
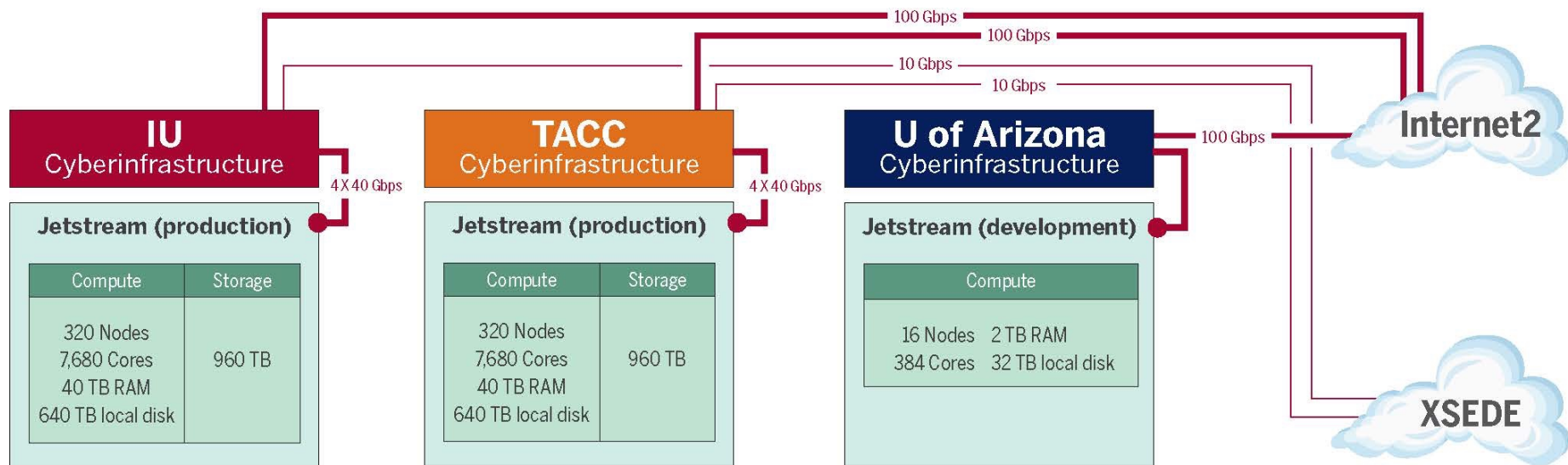
Platform Overview

Jetstream usage highlights

As of October 2017:

- 360 active XSEDE projects covering 66 fields of science and **2180 active users** representing **191 institutions**
- **86%** of Jetstream users new to XSEDE (at end of PY1)
- >76 million CPU hours allocated to XSEDE projects since June 2016
- 9 science gateways
- 42 education/teaching allocations serving almost 800 students
- Averaging 816 concurrent VMs
- **99.4%** cap availability, **97.7%** "job" completion (at end PY1)

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

Requesting access to Jetstream

- Trial allocations available **TODAY**
- You can request startup allocations anytime.
- You can request allocations for educational use anytime.
- Next submission period for large allocations is 15 December 2017 – 15 January 2018.
- 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.
- You **do** need a US-based collaborator

http://jetstream-cloud.org/

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

** Here there be dragons…

Flickr user José Silenzi - dragon
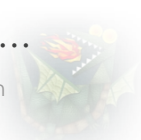
Flickr user Haz – No Reservation

jetstream
http://jetstream-cloud.org/

HPC in the Cloud

- Examples today

Cray in Azure, AWS with NVIDIA V100, IBM Cloud (SoftLayer)

- Customer-driven desires + Vendor Opportunities
- Who's the audience?
- Does it make financial sense?
- Does it make policy sense?

Flickr: Daniel Kulinski – Failure of Cloud Computing

Jetstream timeline…what comes next?

- Year 2 operations review ~Spring 2018
- Refining EOT destinations, targeting underserved communities
- Second summer REU (Research Experiences for Undergraduates)
- Continuously soliciting research allocation requests plus Startup and Education allocations – including Science Gateways!
- Adding services as deemed useful/mature
- Atmosphere enhancements on a regular cycle ~monthly
- Working with XSEDE on cloud image build repo

http://jetstream-cloud.org/

Jetstream partners

http://jetstream-cloud.org/
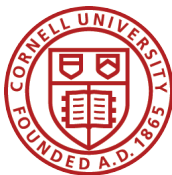
Help / References

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

Paper describing Jetstream: Jetstream: A self-provisioned, scalable scince and engineering cloud environment

Configuration management: https://github.com/jetstream-cloud/Jetstream-Salt-States

http://jetstream-cloud.org/

Questions?

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

License Terms

Jetstream
http://jetstream-cloud.org/

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