A (Brief) Jetstream Overview
A national research and education cloud

ESIP Summer Meeting 2017 - July 26, 2017 – Bloomington, IN

Jeremy Fischer – jeremy@iu.edu

Senior Technical Advisor, Jetstream
UITS Research Technologies

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

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

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

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

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)

Jetstream
http://jetstream-cloud.org/

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

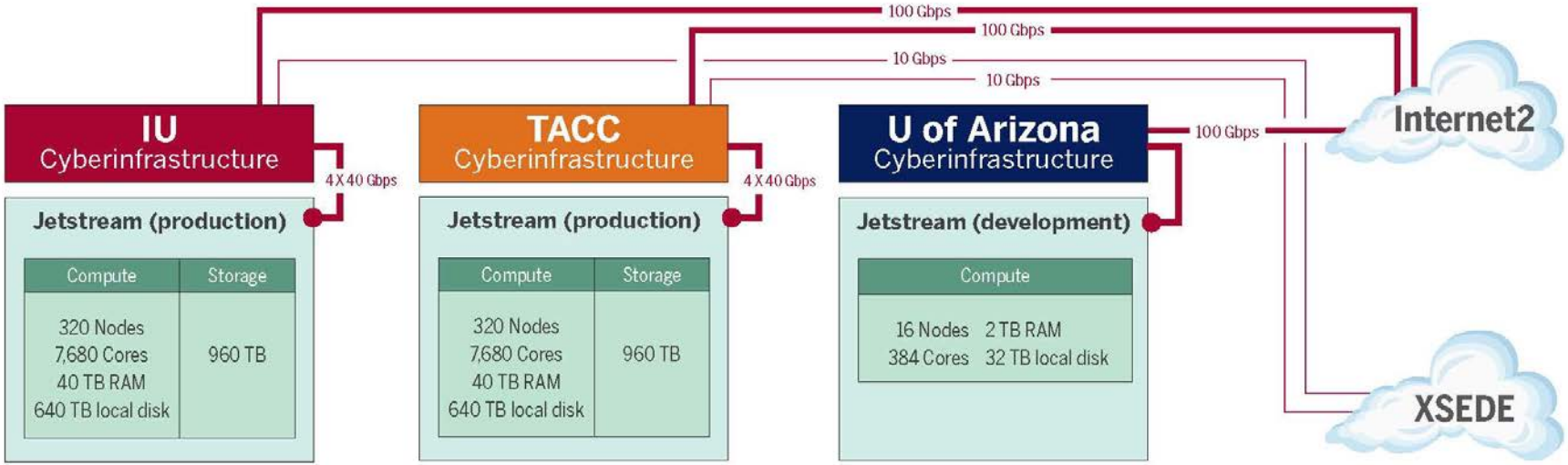
Who uses Jetstream?

- 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

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

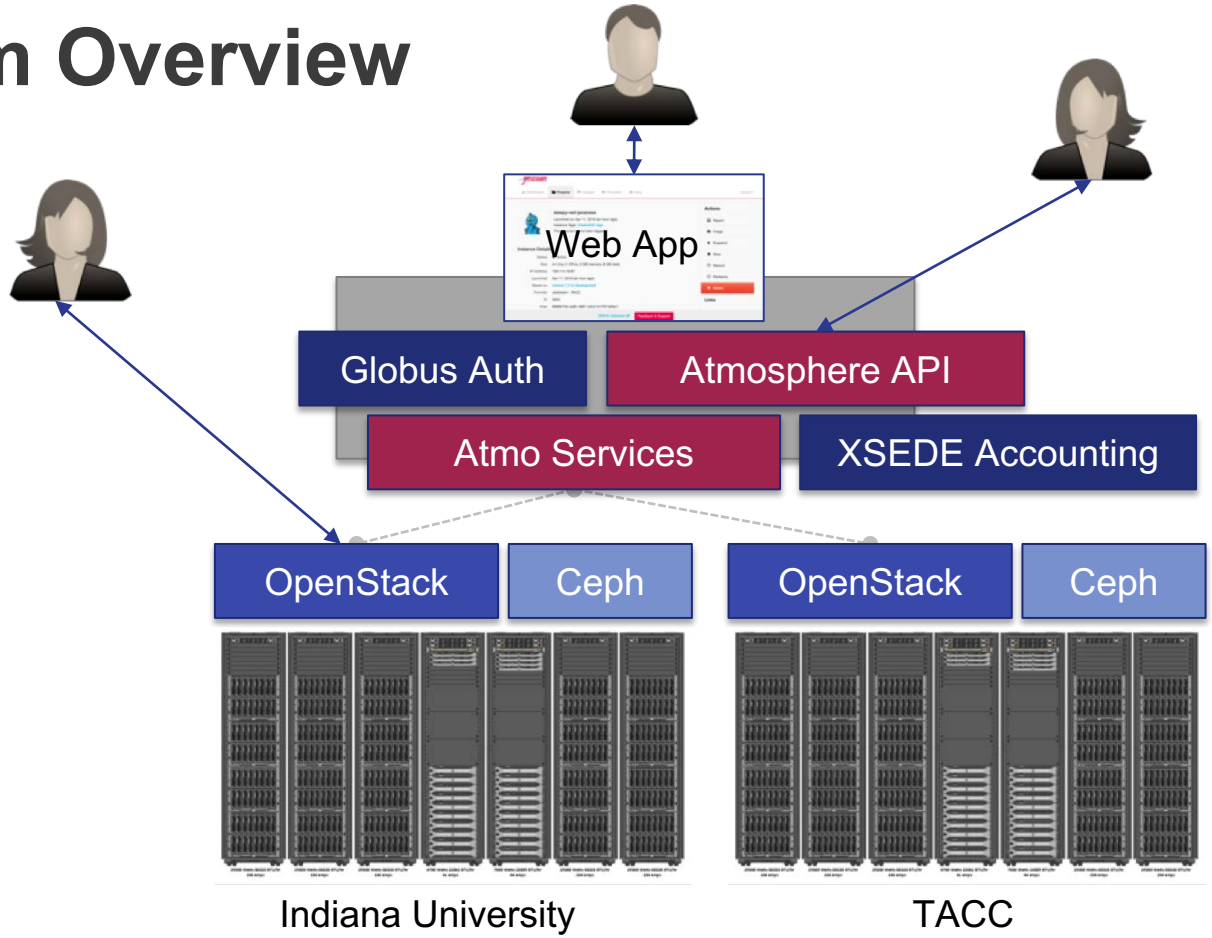
Jetstream System Overview



Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

Platform Overview



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*

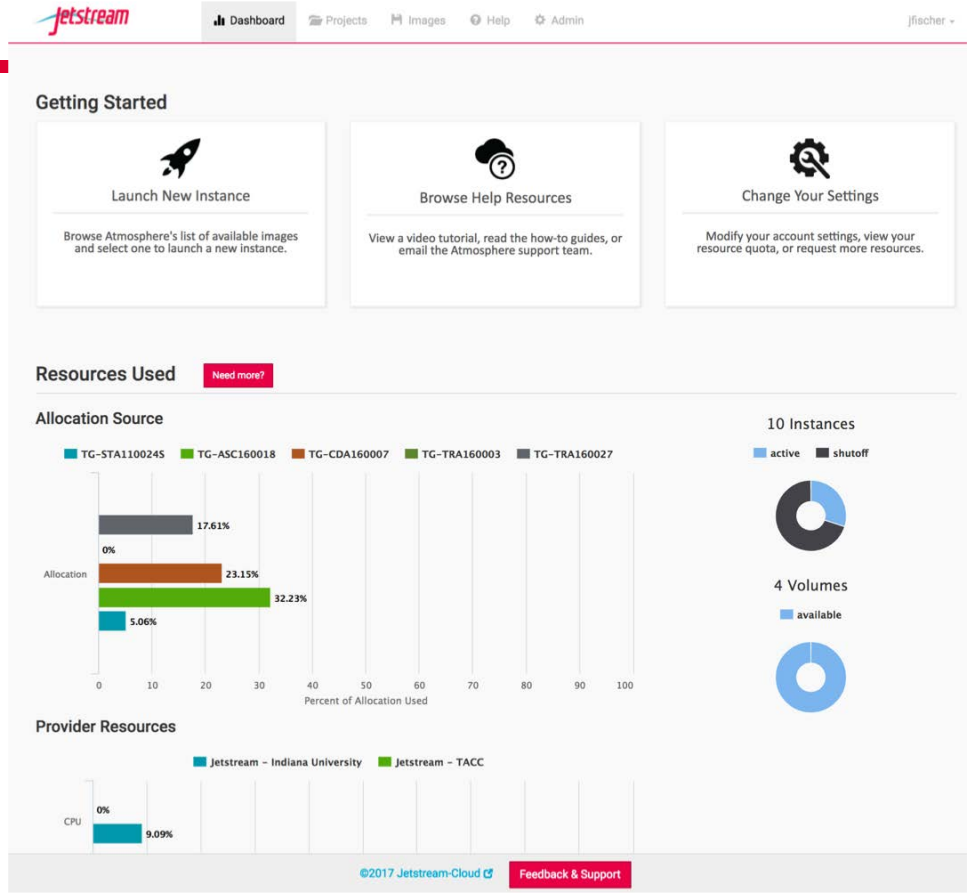
Flavor	vCPUs	RAM	Storage	Per Node
m1.tiny	1	2	8	46
m1.small	2	4	20	23
m1.medium	6	16	60	7
m1.large	10	30	60	4
m1.xlarge	24	60	60	2
m1.xxlarge	44	120	60	1
s1.large**	10	30	120	4
s1.xlarge**	24	60	240	2
s1.xxlarge**	44	120	480	1

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

Jetstream
http://jetstream-cloud.org/

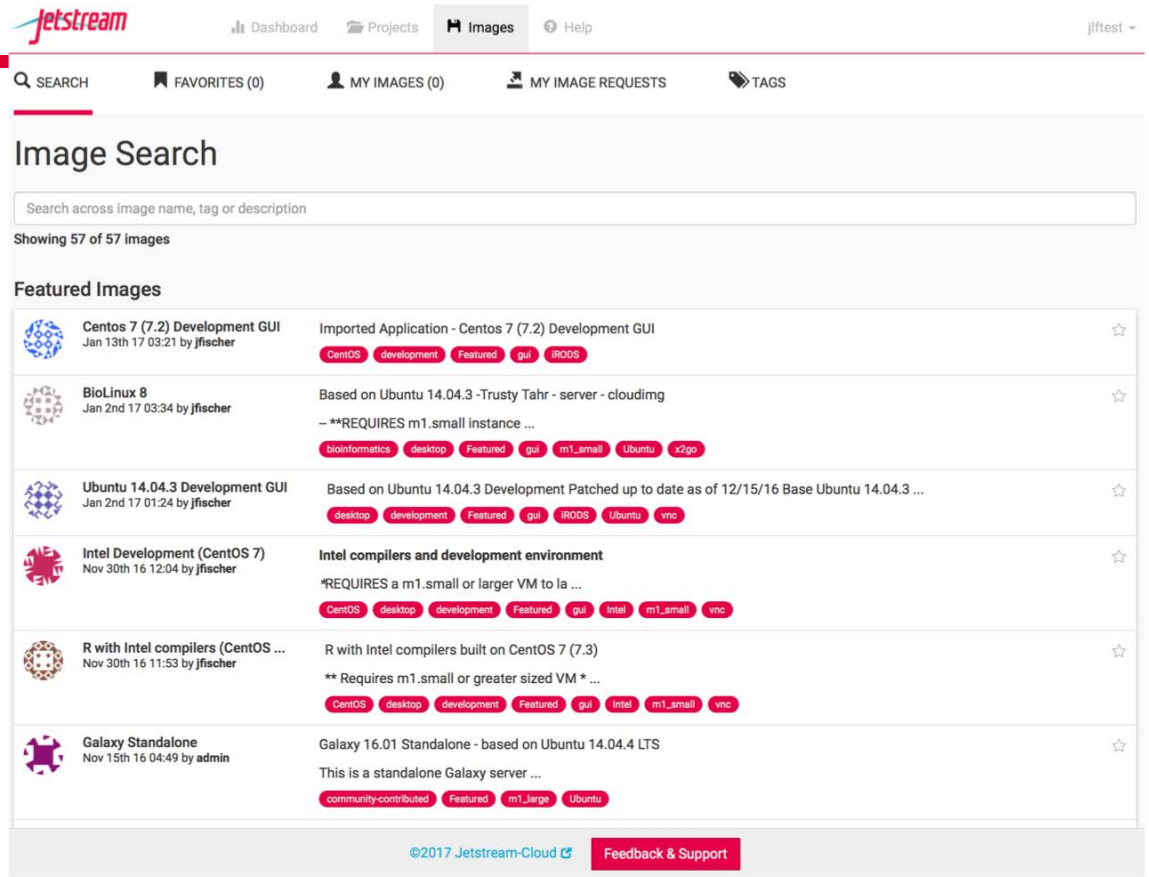
funded by the National Science Foundation
Award #ACI-1445604

The Jetstream Atmosphere web interface



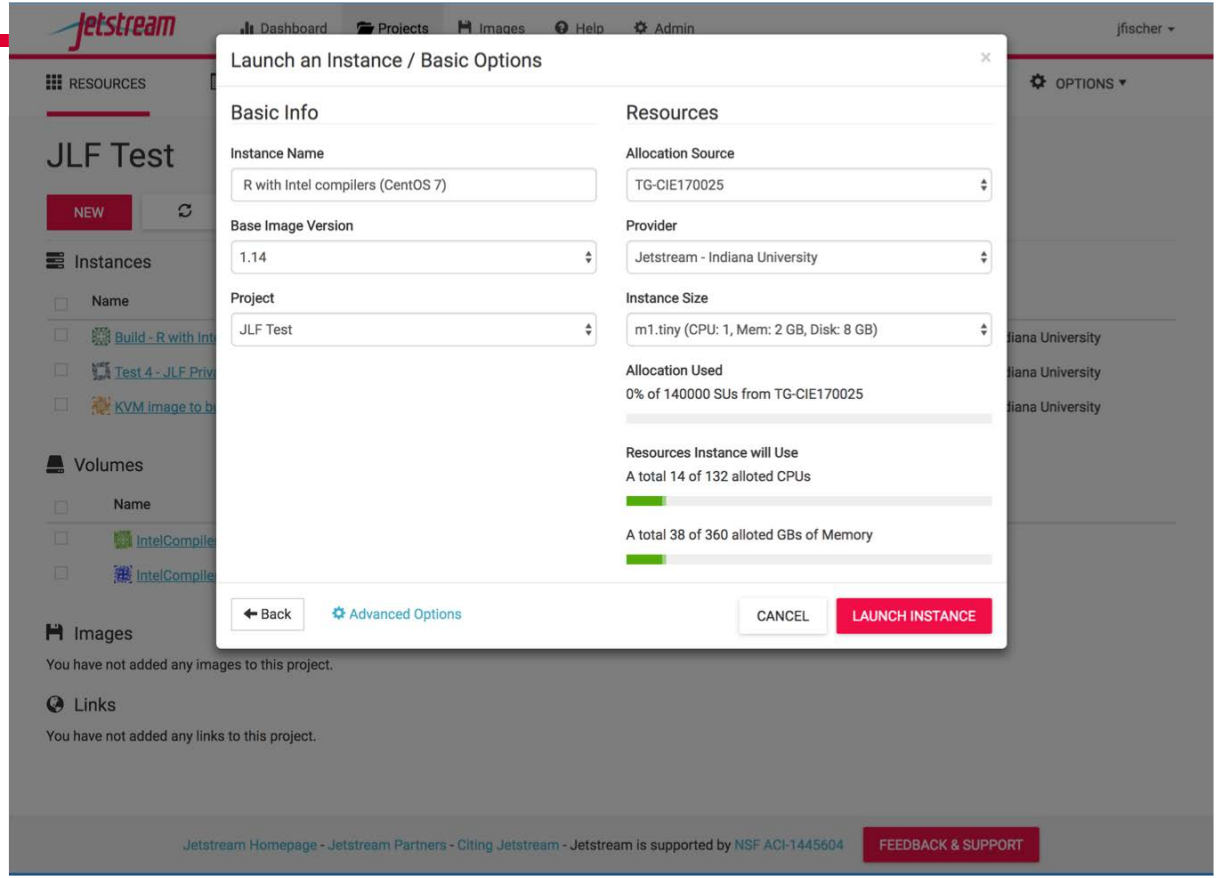
Jetstream
http://jetstream-cloud.org/

The Jetstream Atmosphere web interface



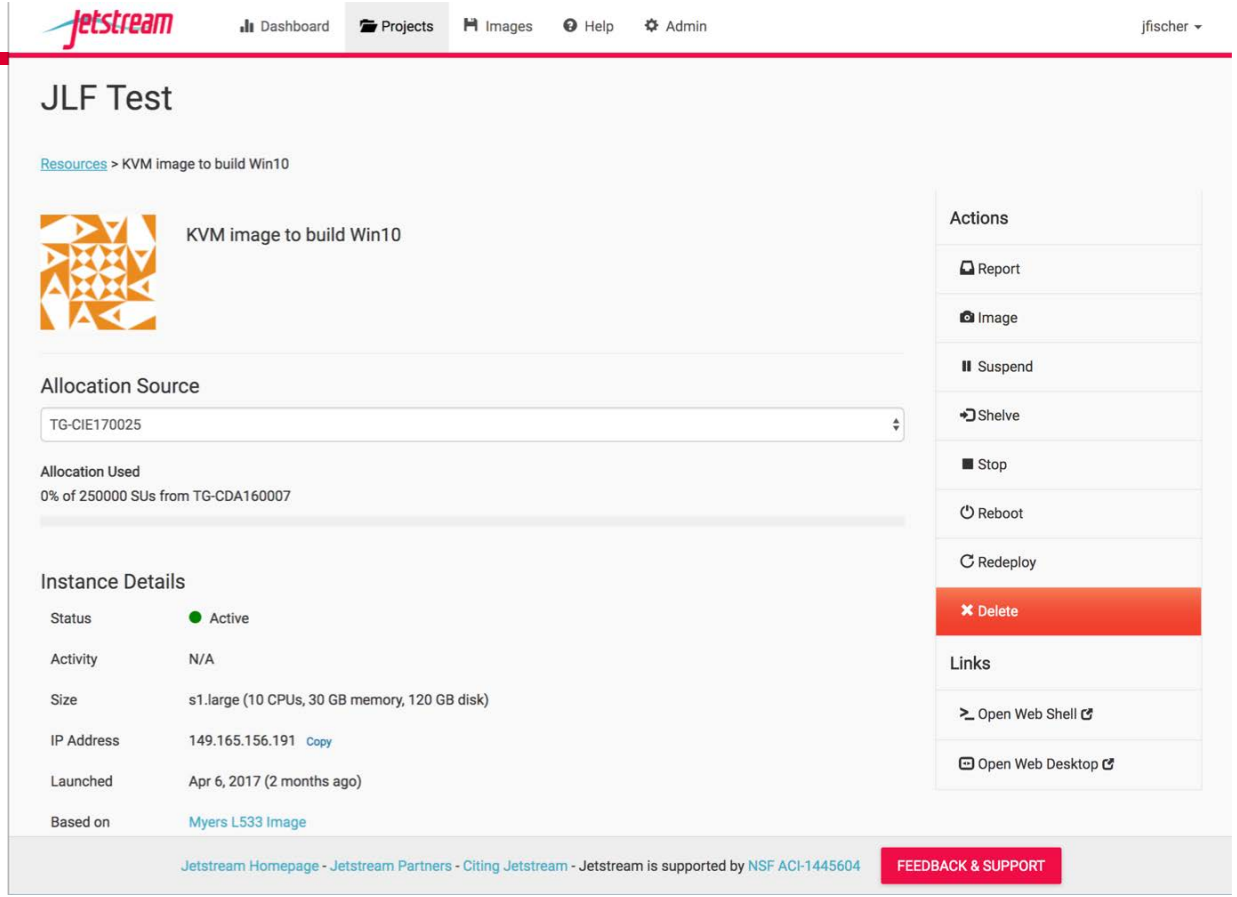
Jetstream
http://jetstream-cloud.org/

Look! It's more Jetstream web interface!



Jetstream
http://jetstream-cloud.org/

Even more Jetstream web interface…



Jetstream
http://jetstream-cloud.org/

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

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

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…

Jetstream
http://jetstream-cloud.org/

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

Thinking about VMs…

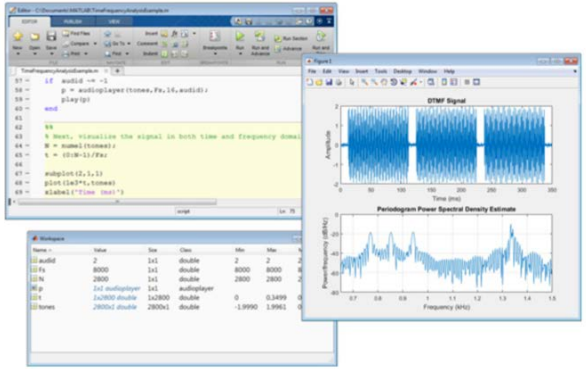
Cattle, not pets: pets take great amount of care, feeding, and you name them; cattle you intend to have high turnover and you give them numbers.

-- Mike Lowe (Jetstream architect)

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

Other Jetstream uses…



Matlab and 52 standard toolkits are installed on Jetstream

(You do NOT need to have a local license to use MATLAB on Jetstream)

Jetstream has been used for a number of courses and workshops to date. Ongoing classes/workshops on data analysis in finance using R, security and intrusion detection, bioinformatics, and principles in cloud computing

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

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

Jetstream
http://jetstream-cloud.org/

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

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)

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

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 September 2017–15 October 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.

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

Jetstream Overall Highlights

As of June 1, 2017:

- 298 active XSEDE projects covering 54 fields of science and 1900+ active users representing 165 institutions
- Over 65 million CPU hours allocated to XSEDE projects since June 2016
- 9 active science gateways
- 28 education/teaching allocations serving almost 600 undergraduate and graduate students

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

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
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

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/

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

Jetstream Partners

INDIANA UNIVERSITY
PERVASIVE TECHNOLOGY INSTITUTE

JOHNS HOPKINS UNIVERSITY

THE UNIVERSITY OF ARIZONA

TACC

THE UNIVERSITY OF CHICAGO

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

Questions?

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

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

License Terms

- Fischer, Jeremy. July 2017. Jetstream Overview – ESIP Summer Meeting 2017, Bloomington, IN.
- 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.

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604