

**RESEARCH TECHNOLOGIES**  
UNIVERSITY INFORMATION TECHNOLOGY SERVICES

[rt.iu.edu](http://rt.iu.edu)



**RESEARCH TECHNOLOGIES**

UNIVERSITY INFORMATION TECHNOLOGY SERVICES

# Programmable CI using Jetstream: National & Local Impacts

**David Y. Hancock – Indiana University**

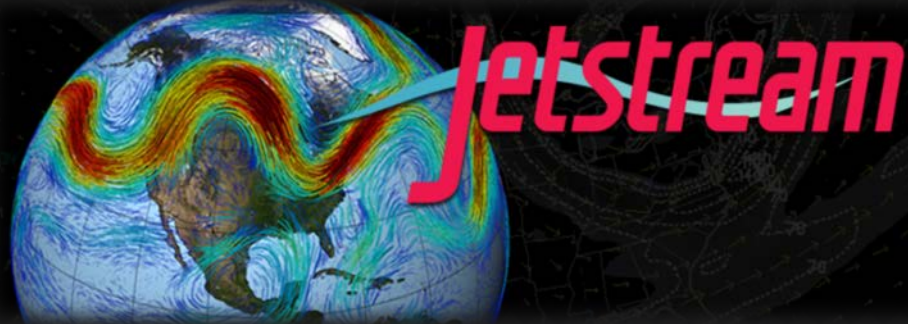
Director for Advanced Cyberinfrastructure

Jetstream Primary Investigator

Affiliated with IU Pervasive Technology Institute

SOS23 – March 29, 2019

# What is Jetstream and why's it relevant?

- NSF's first production cloud facility
  - Provides **on-demand** interactive computing and analysis or persistent services such as gateways
  - Enables configurable environments; **programmable cyberinfrastructure**
  - 409 active XSEDE projects covering 72 fields of science and **>2250 active users** representing **196 institutions**
- 
- **80%** of Jetstream users have not used any other XSEDE system
  - **Highest** user satisfaction in most recent XSEDE survey



# Chair Questions

*How do you define HPC?*

- "P" for Performance or Productivity?
- Powerful vs Persistent
- What about "P" as in Price?
- Where's the Plentiful user growth?
  - In the long-tail?
- Recent XSEDE usage (2017-2018)
  - **51%** SUs  $\leq$  128 cores
  - **84%** SUs  $\leq$  1024 cores
- NIH (21%) + DOE (20%) + DOD (8%) > NSF (34%) usage over first 6 years of XSEDE

Did someone say long-tail?



Flickr User: The Wasp Factory



# Chair Questions

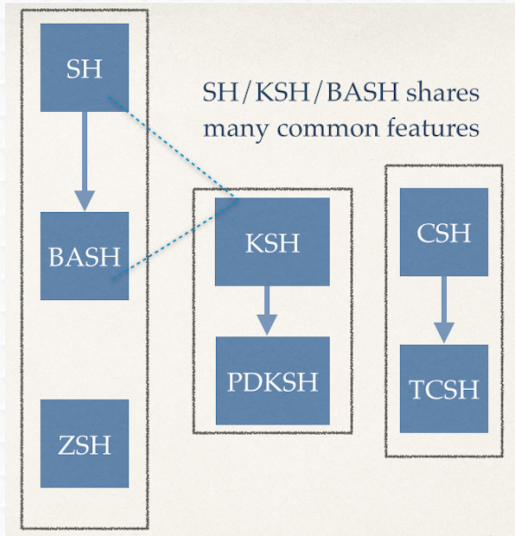
*Is your proposed cloud solution going to benefit the existing user base or will be at an expense of the existing user base?*

- Complementary
- Interactive by design
- Workflow complexity
- Mandates portability
- Spectrum of resources
- Private or Public Clouds
  - Streaming data ingest & analysis
  - Interactive on-demand
  - Platform for training & workforce development
  - Containers as packaging means cloud as dev



# Local view – observations & misconceptions

\*nix **UX** enhancement chart



Cyberciti.biz shell primer

- Evolution beyond “\$” -> “#”
- Availability
  - Users **like** no schedule downtime!
- Runtime
  - 2 day runs vs **2 years**



# Local view – observations & misconceptions

Flickr: Daniel Kulinski – Failure of Cloud Computing



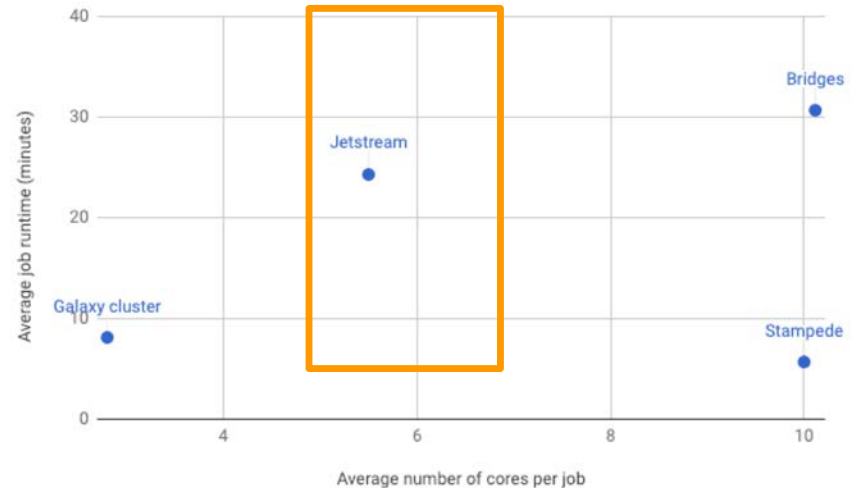
- ~~Video~~ Cloud killed the ~~radio star~~ HPC system
- The cloud is **always** available
- HPC is **always** faster
- "I don't need it."
- Pretend this doesn't affect you, two words...



# Science Gateway: Galaxy

## Galaxy Main gateway usage stats for Jetstream

- Galaxy Main (*usegalaxy.org*) submits a portion of its workload to remote Jetstream nodes while it automatically handles data staging.
- Jetstream processed jobs for **19,560 individual users**
- Galaxy submitted **199,401 user jobs**
- Those jobs consumed **1.15M CPU hours**
- Jetstream is **the only resource handling medium-size jobs for Galaxy Main**

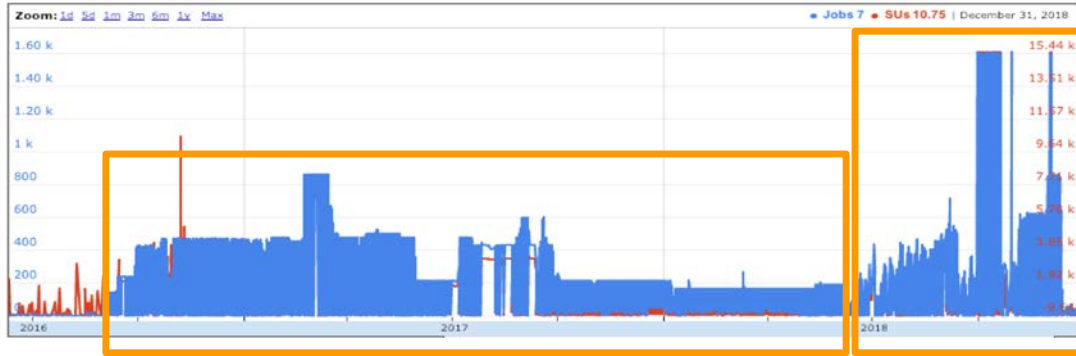




# Science Gateway: Galaxy

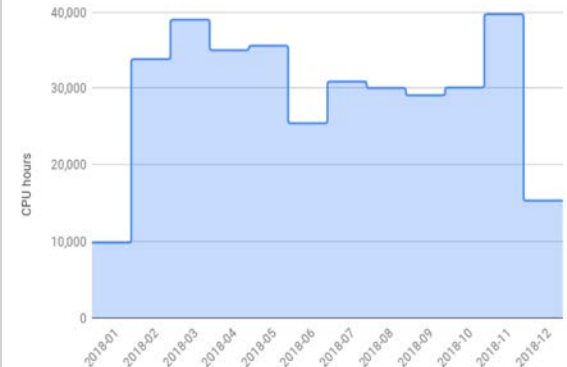
## In 2018, Galaxy implemented elastic node scaling

SlurmScale Ansible role monitors Slurm node usage and adapts VM state, making more efficient usage of the Jetstream allocation.



Manually set number of job nodes.

Elastic scaling.



The CPU hours consumption remains relatively constant yet allocation utilization is bursty. [Galaxy is leveraging Jetstream elasticity to reduce job wait.](#)



# Workforce development & Training

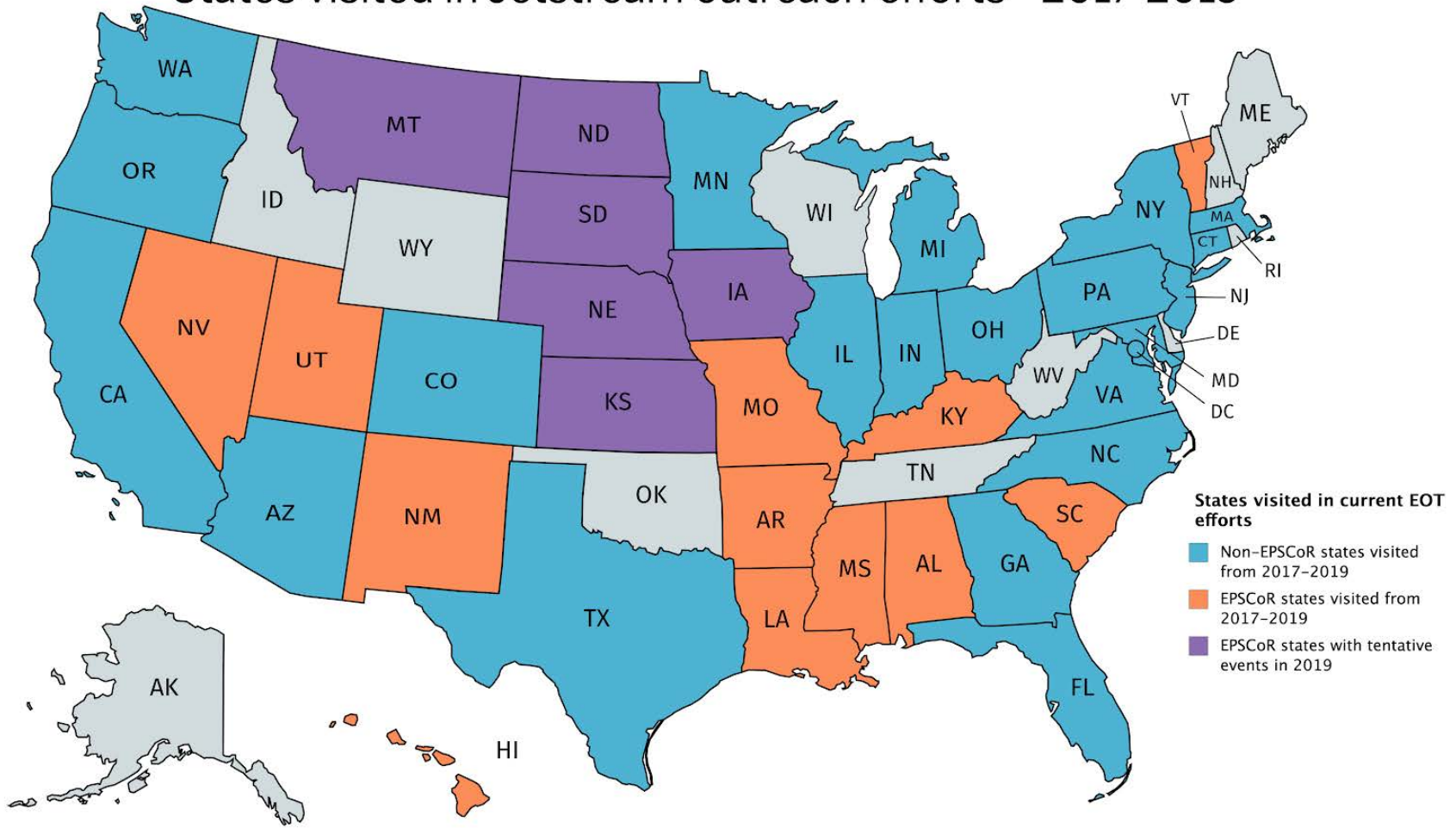
Jetstream REU Program 2018 – courtesy UITS ITCO



- Commitment **required** (March 31)
- <https://jetstream-cloud.org/research/reu.php>
- Requires many to exit comfort zone



# States visited in Jetstream outreach efforts - 2017-2019



# Conclusions, looking ahead...

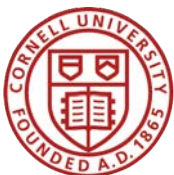


Flickr user Oiluj Samall Zeid - Lejos de Yulín

- Jetstream proposed in 2014
- Future is more ambitious
- Continued campus convergence
  - 5 HPC systems
  - 3 virtual desktop platforms
  - 2 private clouds
- Today's large-scale -> tomorrow's long-tail



# Jetstream partners



funded by the National Science Foundation  
Award #ACI-1445604

