UPON IMPACT

RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES

RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES

IU Cyberinfrastructure:
National & Local Impacts

David Y. Hancock – Indiana University

Director for Advanced Cyberinfrastructure

Jetstream Primary Investigator

Affiliated with IU Pervasive Technology Institute

Workshop on Next Generation Cloud Resources

Princeton, NJ – November 11, 2019

7 University Campuses
9 Medical Education Centers
2 Academic Centers

Indiana University, *est. 1820*

- **$3.7B** enterprise
- Partnered with **$6.4B** IU Health system

- **94,000** Students
- **1.1M** credit hours per semester
- **>20,000** degrees per year
- **$950M** in financial aid
- **$604M** in research grants

- **20,000** faculty and staff
- **7,873** Acres
- **949** owned and leased buildings, **36M** square feet
- **>690,000** *living* alumni

Almost completed 2009 Network Master Plan for IUB/IUPUI 361 buildings <$100M (Orig Budget $172M)

Multi-institution Partnerships

Networks

National network collaboration of 400+ Universities and national labs. IU acts as network operations center for the organization.

Indiana state network for higher education with over 100 members. IU acts as its network operations center.

Collaboration of I-Light, IU, Notre Dame, and Purdue providing high-speed, high-availability, feature-rich network to Indiana's higher education institutions.

Libraries

Began as collaboration between the Big Ten Academic Alliance and the University of California system to digitize library print holdings. Now includes over 150 universities. IU serves as primary backup site for repository.

Research Technologies

IU-led collaboration with Texas Advanced Computing Center, University of Chicago, University of Arizona, and University of Texas (SA) to develop cloud based tools for scientific research funded by National Science Foundation.

Collaboration with Texas Advanced Computing Center, Pittsburgh Supercomputing Center and San Diego Super-computing Center to analyze massive genomic data.

Collaboration of IU, Purdue, and Notre Dame, with public and private partners to facilitate the translation of scientific discoveries in the lab into clinical trials and new patient treatments.

Learning Technologies

Coalition of universities dedicated to collaborative digital education. Founded by IU, Michigan, Colorado State, and Florida. Now includes 25 institutions.

Security

Research and Education Networks Information Sharing and Analysis Center. Includes 620 member universities. One of 19 nationally recognized industry-specific ISACs.

Center for Applied Cybersecurity Research, Est. 2003, integrates applied research in cybersecurity technology, education, and policy guidance.

Shared Cybersecurity Operations Center. Founded by IU, Nebraska, Northwestern, Purdue, and Rutgers.

ResearchSOC provides cybersecurity services to NSF-funded facilities and projects, such as Gemini Observatory, UNAVCO and GAGE, and the National Radio Astronomy Observatory.

Enterprise Systems

Community sourced software suite driving down cost of enterprise systems. IU co-founded and currently serves on board of directors. Includes over 50 university members.

Portal replacement service discovery application developed at IU. Currently used by 85 university campuses.

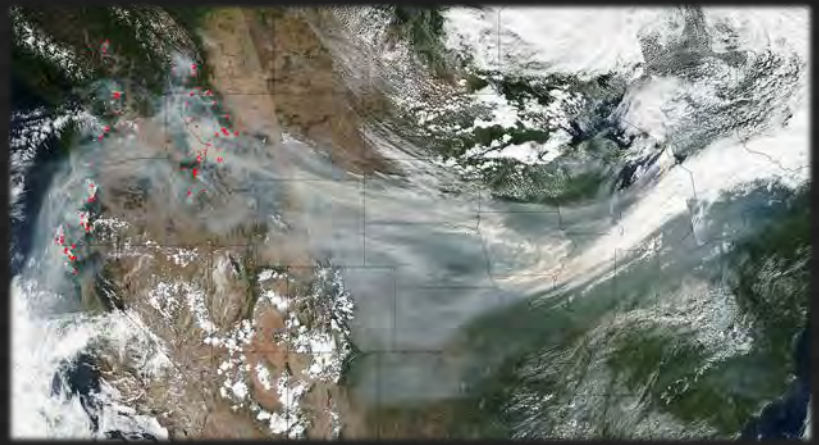
Service & Support

Application developed at IU to integrate cloud storage systems into a single place. Currently has 8 university subscribers and 6 universities testing.

System to manage software licenses developed at IU. Subscribers include University of Maryland and University of Alabama-Huntsville.

What is “the” Jetstream?

- Fast moving air currents
- Hot/Cold air boundaries
- An NSF-funded cloud environment



Suomi NPP satellite collected this naturaNASA'sl-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

What is Jetstream and why does it exist?

- NSF's first production cloud facility
- Focus on ease-of-use, broad accessibility
- User-selectable library of preconfigured virtual machines
- Provides on-demand *interactive* computing and analysis or persistent services such as gateways
- Enables *configurable* environments; *programmable cyberinfrastructure*

Expanding NSF CI'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 XSEDE-supported resources
- Less than 4% had an XSEDE Portal account
- **70%** of researchers surveyed* claimed to be **resource constrained**

Why are the people not using XSEDE-supported 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

“But I really don’t have research needs… I don’t need the national research cyberinfrastructure.”

– multiple researchers at small colleges and universities

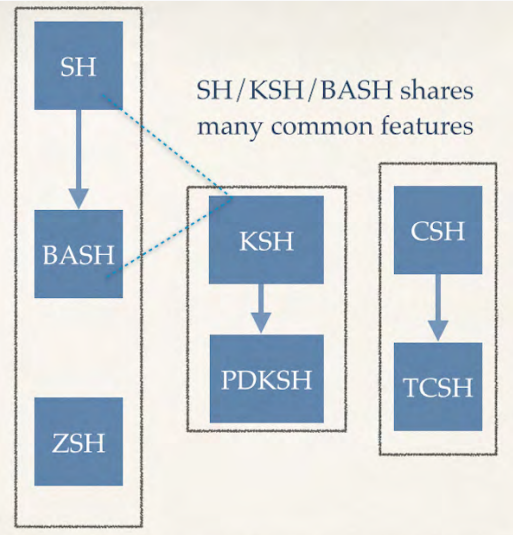
Sarah Romanes @sarah_romanes · Mar 17

When you have to drive to work - but R has also only completed 10hrs of 24hrs worth of simulations. Can I go in the T2 lane with this thing? 😂 #rstats #sydneytraffic

40 74 1.1K

Local view – observations & misconceptions

*nix **UX** enhancement chart



Cyberciti.biz shell primer

- Evolution beyond "**$**" -> "**#**"
- Availability
  - Users **like** no scheduled downtime!
- Runtime
  - 2 day runs vs **2 years**

Local view – observations & misconceptions

Flickr: Daniel Kulinski – Failure of Cloud Computing

Misconceptions

- ~~Video~~ Cloud killed the ~~radio star~~ HPC system
- The cloud is **always** available
- HPC is **always** faster

Local Observations

- "AI for Everyone" w/GPU virtualization
  - SPEC ACCEL Benchmarks on vGPUs
- Research Desktop on Jetstream

How do you define *HPC?*

- "**P**" for Performance or Productivity?
- **P**owerful and **P**ersistent
- What about "**P**" as in Price?
- Where's the **P**lentiful user growth?
  - In the long-tail?
- Recent XSEDE usage (2017-2018)
  - **51%** SUs < 128 cores
  - **84%** SUs < 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

Jetstream usage highlights – 1 Sep 2019

- 398 XSEDE projects covering 72 fields of science and over 2300 **active users** representing **207 institutions**
- **80%** of Jetstream users have **not used any other XSEDE system**
- 2,218 students to date

- 27 active science gateways
  - **26,184** users have run jobs on Jetstream from Galaxy
- 53 education/teaching allocations serving almost 1,000 students
- **Highest** user satisfaction in most recent XSEDE survey

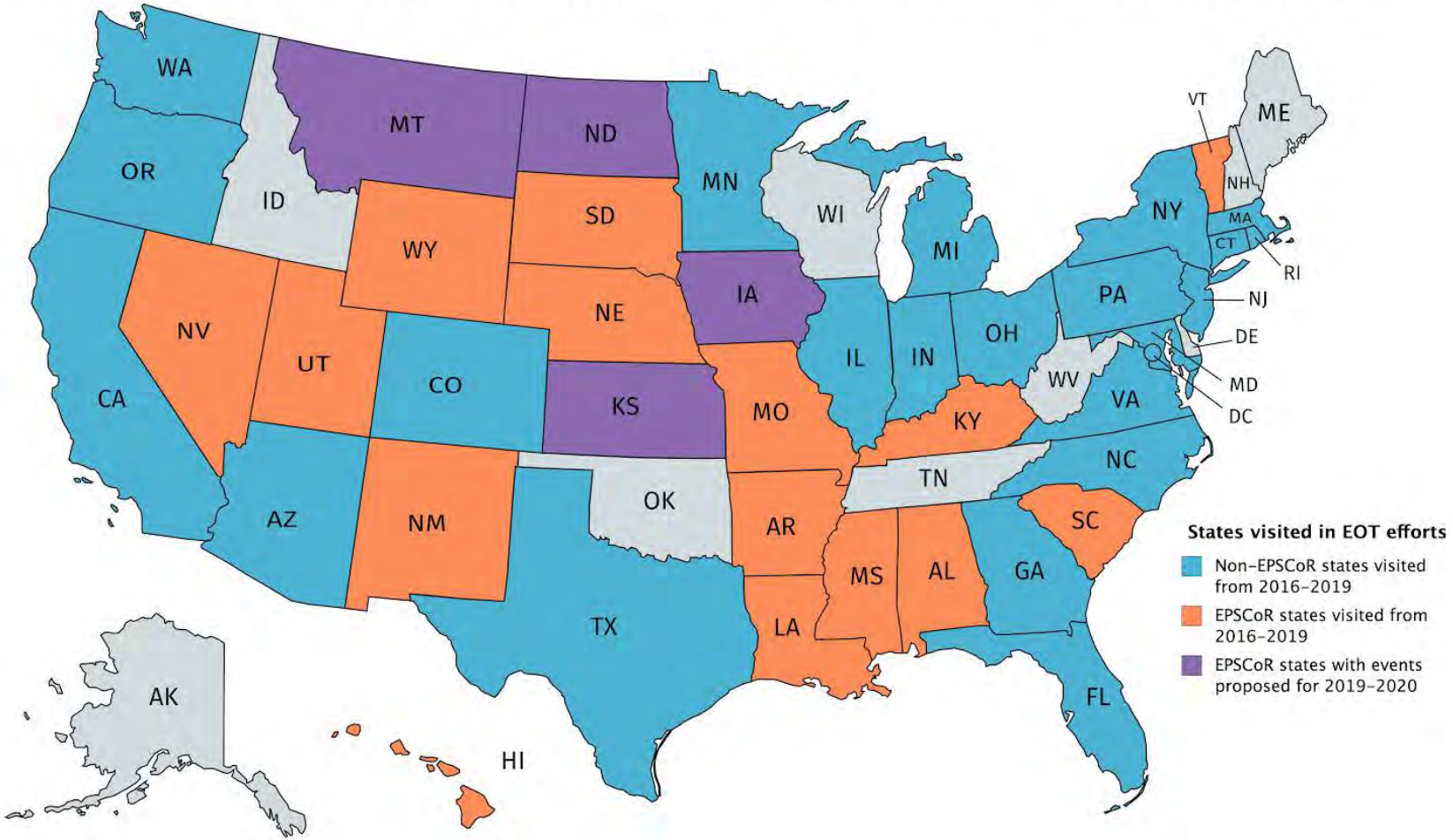
Workforce development & Training

Jetstream REU Program 2018 – courtesy UITS ITCO

- Research Experiences for Undergraduates
- Hosted 4 students in 2017, 6 in 2018 & 2019
- Work presented at PEARC

Attributed Joey deVilla – Sep 2008

States visited in Jetstream Outreach Efforts - 2016 to 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

Acknowledgements

NSF Awards 1053575 & 1548562 (XSEDE), 1341711 (Wrangler), and 1445604 (Jetstream).

This document was developed with support from the National Science Foundation. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

Special thanks to contributors & Jetstream partners:

- Craig Stewart, Therese Miller, Jeremy Fischer, J. Michael Lowe, Winona Snapp-Childs, George Turner, Brad Wheeler, Dan Calarco, and Maria Morris.
- TACC, University of Arizona, Johns Hopkins, University of Chicago, Cornell, UT San Antonio

A PARTNER
in new
POSSIBILITIES

RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES

RT.IU.EDU