Jetstream2: Accelerating cloud computing via Jetstream

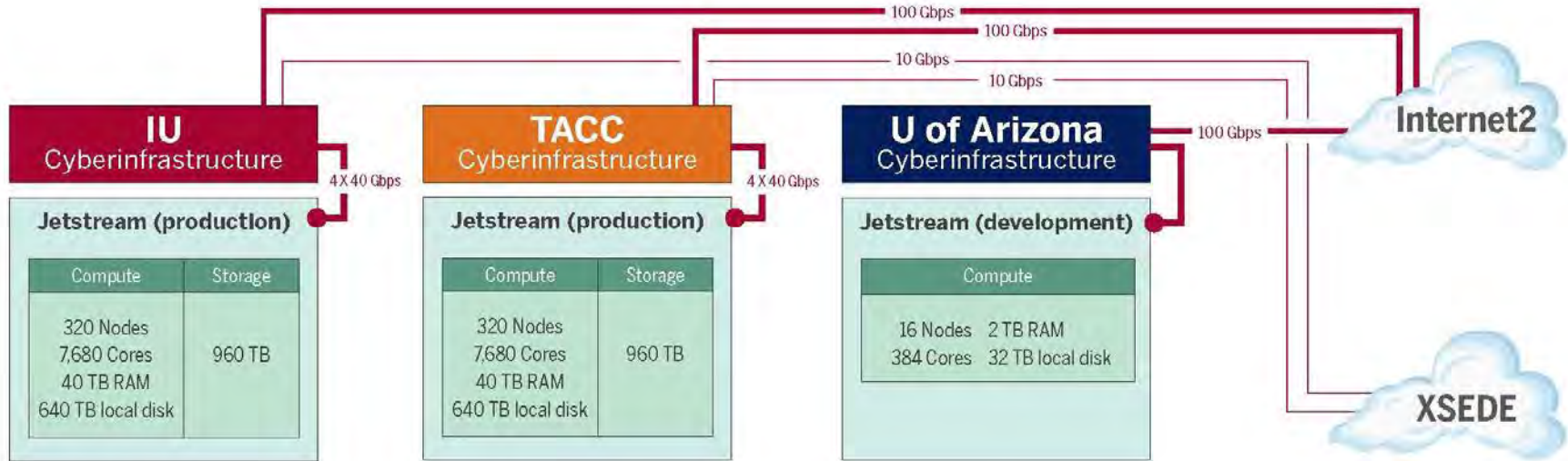**Jeremy Fischer – Indiana University**

Manager, Jetstream Cloud

South Big Data Hub All Hands Meeting – July 28, 2021

Jetstream2

Jetstream1 System Overview

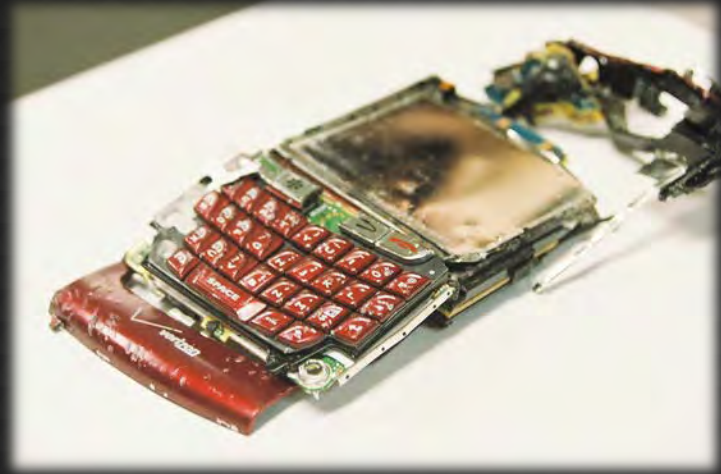http://wiki.jetstream-cloud.org/Network+configuration+and+policies

Quick Jetstream1 Facts

- vCPU ranges from 1 core to 44 cores

- Ram on flavors ranges from 2gb to 120gb

- Atmosphere gives 100gb of block storage per user by default (may request more)

- API side allows 1TB per allocation by default (shared between all allocation users – may request more)

- GPUs (NVIDIA V100) available on a limited basis as ¼ vGPU, ½ vGPU, or 1 vGPU

Jetstream2

Flickr user MattHurst – Broken Blackberry

What worked?

- Allowing API access and full control (root privileges)

- "Indefinite workflows" – allowing instances to run continuously – providing PIs renew their allocations

- Development of trial allocations

What didn't work?

- Forcing small allocations into the research allocation process

- Lack of multi-year allocations

- Lack of shared data set storage

Jetstream2

Lessons learned

Challenges -> Inspired changes

- Storage capacity -> Larger HDD pool and new flash storage

- Homogeneous hardware -> Inclusion of NVIDIA GPUs (w/MIG) and memory diversity

- Separate OpenStack domains -> Unification of "Atmosphere" domain

- Virtual networking architecture/maintenance -> Increase offload capabilities via Cumulus Networks software and Mellanox hardware (NAT & simulation)

- Acceptance & integration into national CI ecosystem -> Changes to our metrics/KPIs & accounting processes

- Deployment diversity -> Leverage single technology for config management

D.Y. Hancock – Castello di Nipozzano 2017

Jetstream2

Jetstream2 Capabilities

Enhancing IaaS model of Jetstream:

- Improved orchestration support

- Elastic virtual clusters

- Federated JupyterHubs

- Ease storage sharing (CephFS w/Manilla)

Commitment to >99% uptime

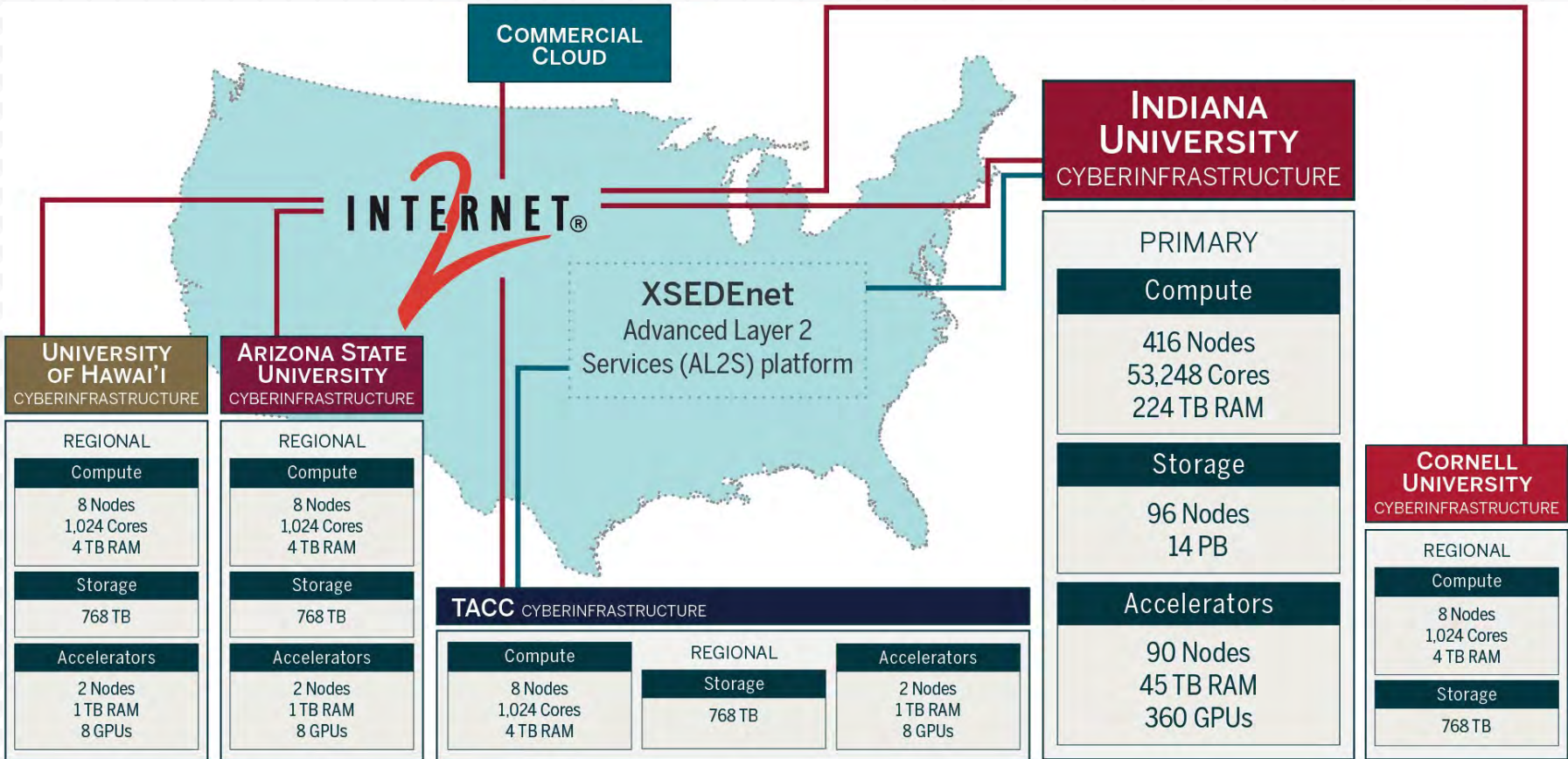- Critical for science gateway hosting

- Hybrid-cloud support

Revamped User Interface

- Unified instance management

- Multi-instance launch

Feb 12, 2019 – Jet stream region called "Jet N6"
NASA/JPL-Caltech/SwRI/MSSS/Kevin M. Gill

- >57K cores of next-gen AMD EPYC processors

- >360 NVIDIA A100 GPUs will provide vGPUs via NVIDIA's MIG feature

- >17PB of storage (NVMe and disk hybrid)

- 100GbE Mellanox network

Jetstream2

COMMERCIAL CLOUD

INDIANA UNIVERSITY
CYBERINFRASTRUCTURE

INTERNET2®

XSEDEnet
Advanced Layer 2
Services (AL2S) platform

PRIMARY

Compute

416 Nodes
53,248 Cores
224 TB RAM

Storage

96 Nodes
14 PB

Accelerators

90 Nodes
45 TB RAM
360 GPUs

UNIVERSITY OF HAWAI'I
CYBERINFRASTRUCTURE

REGIONAL

Compute

8 Nodes
1,024 Cores
4 TB RAM

Storage

768 TB

Accelerators

2 Nodes
1 TB RAM
8 GPUs

ARIZONA STATE UNIVERSITY
CYBERINFRASTRUCTURE

REGIONAL

Compute

8 Nodes
1,024 Cores
4 TB RAM

Storage

768 TB

Accelerators

2 Nodes
1 TB RAM
8 GPUs

TACC CYBERINFRASTRUCTURE

Compute

8 Nodes
1,024 Cores
4 TB RAM

REGIONAL

Storage

768 TB

Accelerators

2 Nodes
1 TB RAM
8 GPUs

CORNELL UNIVERSITY
CYBERINFRASTRUCTURE

REGIONAL

Compute

8 Nodes
1,024 Cores
4 TB RAM

Storage

768 TB

Jetstream2

Timeline

- Jetstream now in 5th year of operations

- Jetstream extension granted by the NSF through November 2021

- Extension through March 2022 in process

- Jetstream2
  - Early operations planned for December 2021
  - Production operations by January 2022

Flickr user Oiluj Samall Zeid - Lejos de Yulín

Jetstream2

Acknowledgements

Jetstream2 partners

PERVASIVE
TECHNOLOGY INSTITUTE

RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES