PERVASIVE
TECHNOLOGY INSTITUTE

RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES

Jetstream2: Accelerating cloud computing via Jetstream

Jeremy Fischer – Indiana University

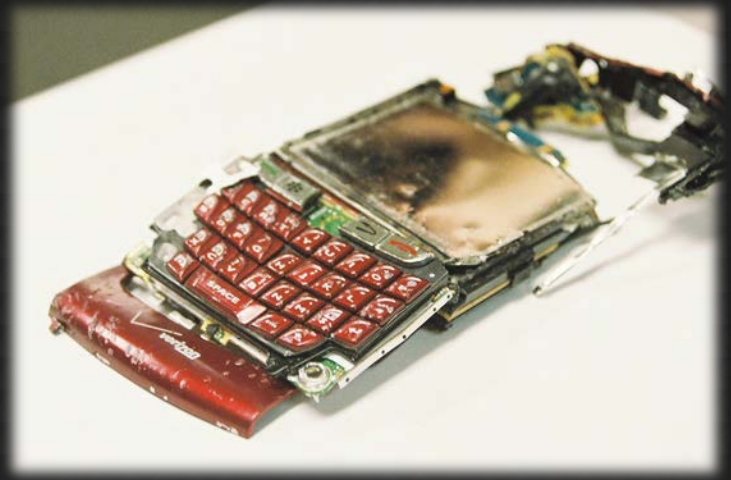
Manager, Jetstream Cloud

XSEDE Campus Champions Meeting – January 18, 2022

Jetstream2

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



Flickr user MattHurst – Broken Blackberry

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

D.Y. Hancock – Castello di Nipozzano 2017

- 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

Jetstream2

What is “the” Jetstream?

- Fast moving air currents
- Hot/Cold air boundaries
- An NSF-funded cloud environment
- A project re-defining state-of-the-ART

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

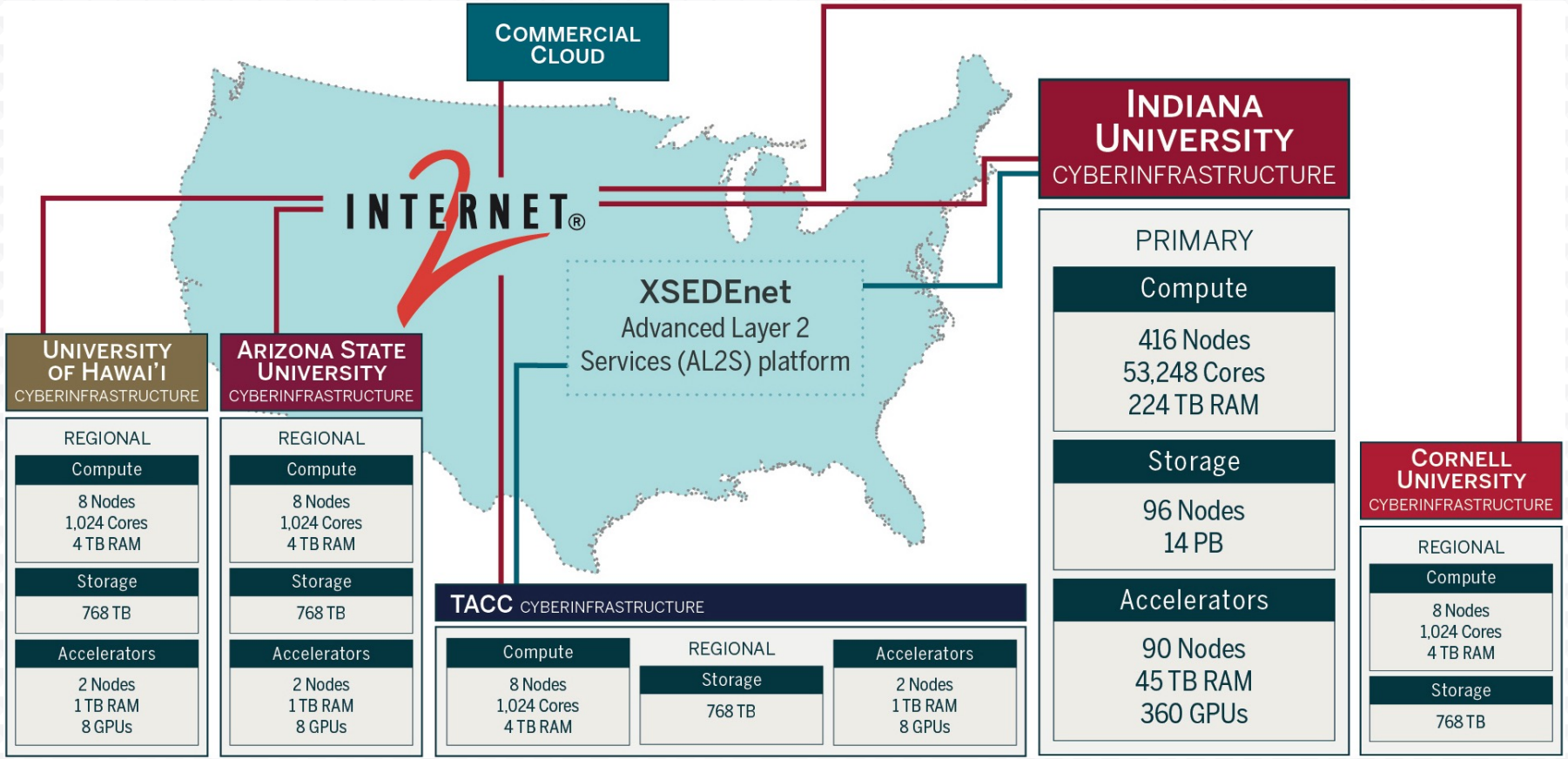
Revamped User Interface

- Unified instance management
- Multi-instance launch

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

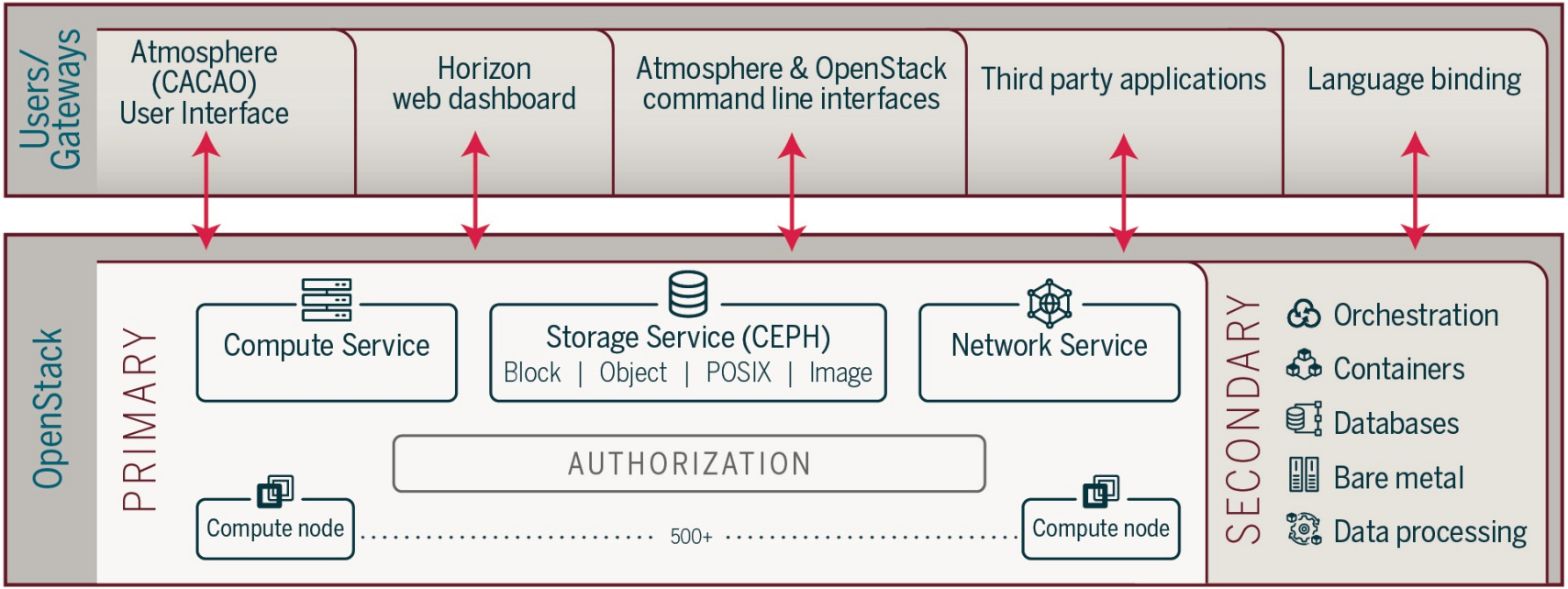
- >57K cores of next-gen AMD EPYC processors
- >360 NVIDIA A100 GPUs will provide vGPUs via NVIDIA's MIG feature
- >17PB of storage (NVMe and disk hybrid)
- 100GbE Mellanox network

Jetstream2


COMMERCIAL CLOUD
INTERNET2®
XSEDEnet
Advanced Layer 2
Services (AL2S) platform
INDIANA UNIVERSITY
CYBERINFRASTRUCTURE
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

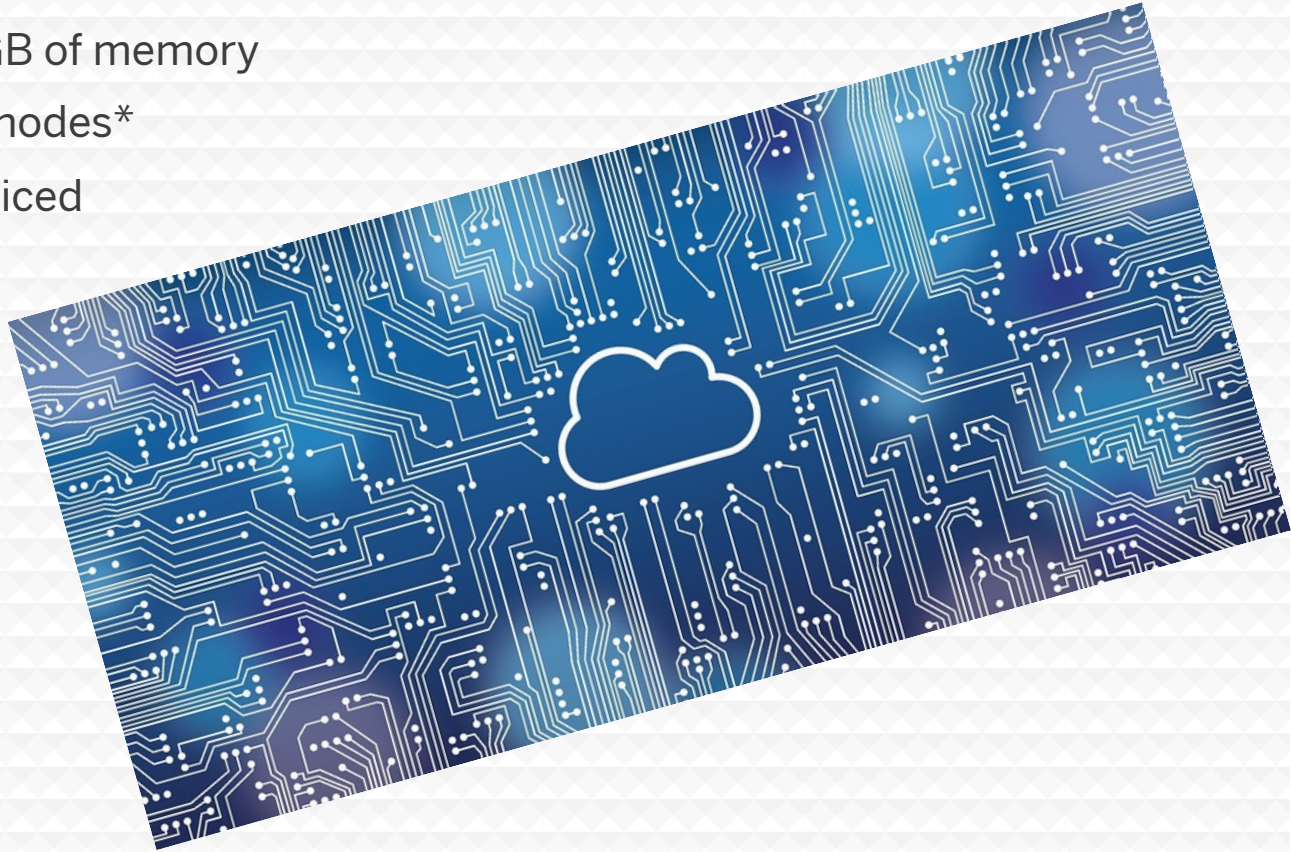
Jetstream2 Architecture



Jetstream2

Big Memory, Larger Instances, GPUs

- Smallest node has 512GB of memory
- 32 Larger 1TB memory nodes*
- A100 GPUs sliced and diced
- 128 Core nodes



Jetstream2

Allocatable Resources

- Primary cloud (IU) only
  - Jetstream (CPU Only)
    - 1 core hour = 1 SU
  - Jetstream LM (1TB Large Memory nodes)
    - 1 core hour = 2 SUs, 128 SU/hr minimum based on proposed flavors
  - Jetstream GPU (NVIDIA A100 GPU nodes)
    - 1 core hour = 4 SU, 16 SU/hr minimum based on proposed flavors
  - Jetstream Storage
    - Only with CPU/GPU/LM allocation
    - NVMe space available on limited basis and must be requested via Jetstream Help with proper justification
- Regional cloud access by invitation or request from provider

Jetstream2

Startup and Champion Defaults

- Primary cloud (IU) only
  - Jetstream (CPU Only) – 200,000 SU
  - Jetstream LM (1TB Large Memory nodes) – 400,000 SU
  - Jetstream GPU (NVIDIA A100 GPU nodes) – 600,000 SU
  - Jetstream Storage (requires one of the compute resources) – 1TB
- Regional clouds have no default startup values as you must have a primary cloud account first – quotas are determined by cloud provider if access is granted and may be significantly smaller than primary cloud

Jetstream2

Proposed VM flavors

Table 1. VM CPU Instance Configurations

Instance Type	vCPUs (128 total)	RAM (500GiB available)	Ephemeral Storage (in GB)	Instances/Node
m3.tiny	1	3	20	128
m3.small	2	6	20	64
m3.quad	4	15	20	32
m3.medium	8	30	60	16
m3.large	16	60	60	8
m3.xl	32	125	60	4
m3.2xl	64	250	60	2
m3.3xl	128	500	60	1

Table 2. VM GPU Instance Configurations

Instance Type	vCPUs (128 total)	vGPUs (7 slices)* + GPU RAM	RAM (500GiB available)	Ephemeral Storage (in GB)	Instances/Node
g3.small	4	1 / 5gb	15	120	28**
g3.medium	8	2 / 10gb	30	120	16
g3.large	16	3 / 20gb	60	120	8
g3.xl	32	7 / 40gb	125	120	4

*7 GPU slices = 1 NVIDIA 40GB Ampere A100 GPU

** https://docs.nvidia.com/datacenter/tesla/mig-user-guide/#a100-profiles - 7 slices max

Table 3. Large Memory Instance Configurations

Instance Type	vCPUs (128 total)	RAM (1000GB available)	Ephemeral Storage (in GB)	Instances/Node
r3.large	64	500GB	60	2
r3.xl	128	1000GB	60	1

Jetstream2

XRAC Limits

No established limits for any resource except storage. Awards will be based on merit as determined by XRAC and availability.

Jetstream2 Storage is limited to 40TB unless explicitly approved by JS2 allocations staff

Jetstream2

Scaling and Code Performance on JS2

- Continuing the trend with JS1
  - Traditional code gets measured in the normal ways, paying attention to cores and memory usage
    - As scaling generally doesn't span nodes, it's attention to single node (or less) performance and choosing the best VM flavor and length of running VM
    - For startups and education, internally we often allow for VM management time (10 to 20% as VMs don't automatically shut down when the runs complete)
  - Infrastructure allocations like gateways, Kubernetes deployments, or other support VMs are measured differently
    - General equation for an "always on" service VM:
      - # of cores * 24 hours/day * 365 days = VM cost in SUs
      - Will use the appropriate multiplier for LM or GPU
    - Providing a table of VMs, purpose, and SUs along with the detailed description is desirable for infrastructure allocations

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 January 2022
  - Production operations by February 2022

Flickr user Oiluj Samall Zeid - Lejos de Yulín

Jetstream2

PERVASIVE TECHNOLOGY INSTITUTE

RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES

Acknowledgements

NSF Awards 1053575 & 1548562 (XSEDE), 1445604 (Jetstream) and 2005506 (Jetstream2)

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 & Jetstream2 partners

- PI David Y. Hancock, J. Michael Lowe, Therese Miller, Maria Morris, Winona Snapp-Childs, and George Turner

PERVASIVE TECHNOLOGY INSTITUTE

RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES

Jetstream2 partners

Jetstream2

http://jetstream-cloud.org/
National Science Foundation
Award #ACI-2005506


PERVASIVE
TECHNOLOGY INSTITUTE
RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES