



PERVASIVE TECHNOLOGY INSTITUTE



UNIVERSITY INFORMATION TECHNOLOGY SERVICES
RESEARCH TECHNOLOGIES

Jetstream2

Research Cloud Infrastructure: A Cloudy Introspective

Jeremy Fischer – Indiana University
Research Cloud Infrastructure Manager

RT Lunch Series

Bloomington, IN – June 8, 2023

Research Cloud Infrastructure: An Origin Story



- Who we are:
 - Jeremy Fischer – Manager, former JS/JS2 EOT/support lead
 - Mike Lowe – Senior Cloud Architect, JS/JS2 architect and lead
 - Steve Bird – Cloud Administrator, JS/JS2 admin and ResCloud admin
 - Aaron Wells – Cloud Administrator, ResCloud admin
 - Sarah Williams – Infrastructure Specialist, Monitoring wizard

What RCI does

We subvert the dominant paradigm by engaging services that utilize technologies hosted by a different organization than the end user.

Translation: We run cloud services and some services related to RT operations.



No, really, what does RCI do?

- Jetstream2 – National Science Foundation funded research cloud
- IU Research Cloud (ResCloud) – a new cloud system built on the former Jetstream1 hardware
- Jump hosts
- Zabbix
- Vault
- Loghost
- Rundeck
- REDCap (transitioning to RCS)
- Other services to be named later

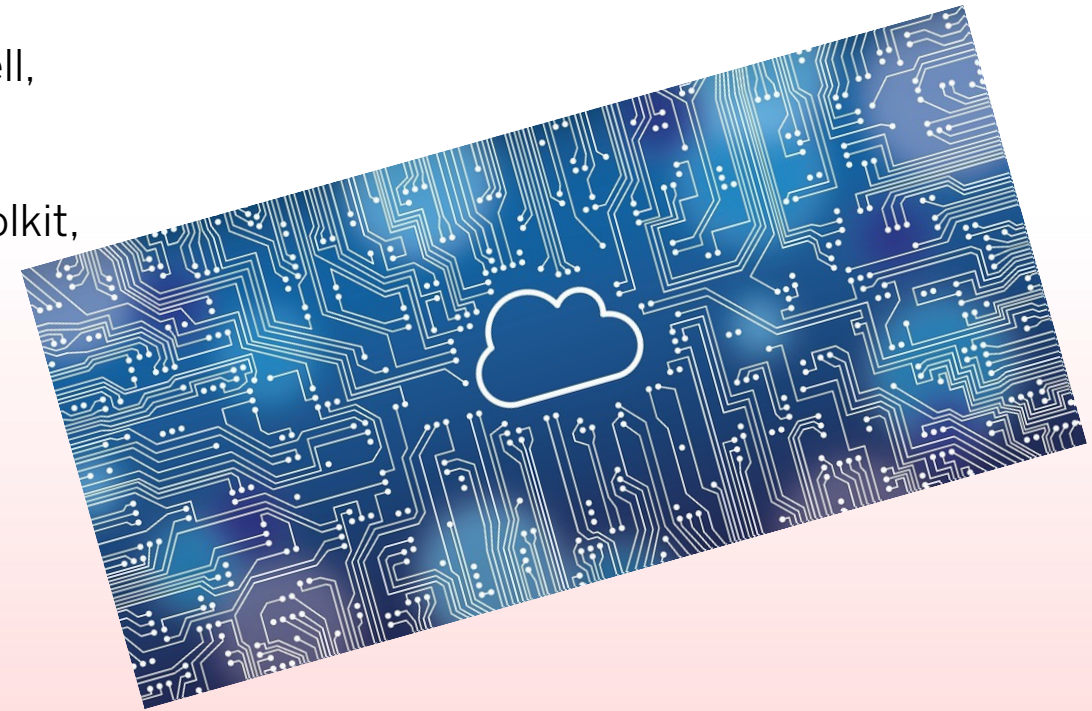
About Jetstream2

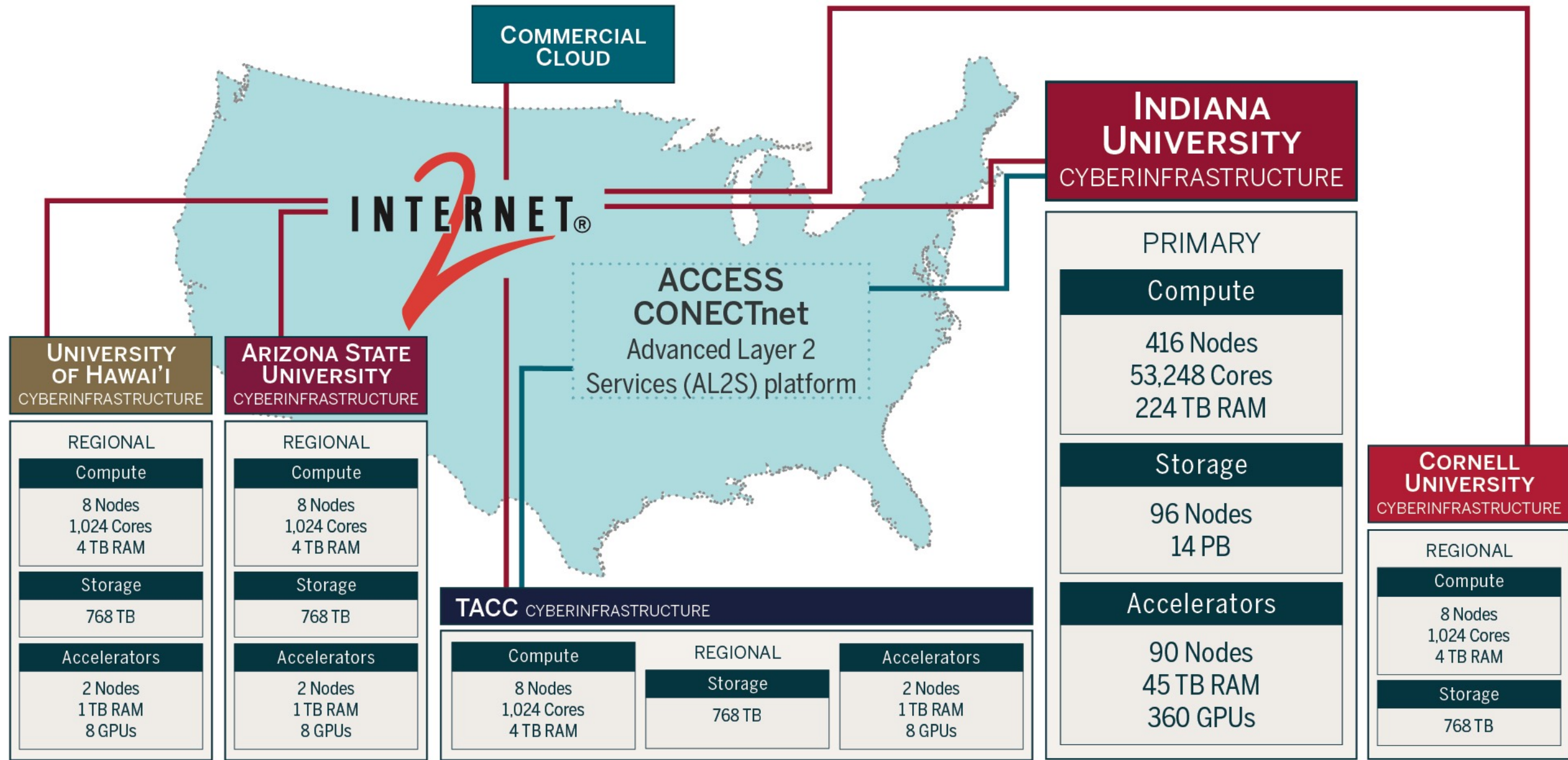
- Mostly the same as my two talks to RT in August 2022
- NSF-funded production cloud environment
- Ease-of-use focus, rapid on-ramp to ACCESS (allocated ONLY via ACCESS)
- **On-demand** interactive computing and persistent services for science gateways
- Enables configurable environments; *programmable cyberinfrastructure*
- Building on the success of Jetstream1
 - The 63 science gateways that utilized Jetstream indirectly supported over 183,197 people.
 - Six year of operations an overall availability of 98.54%, incl. planned and unplanned outages
 - An uptime of 99.9967% where the system was operating but at a reduced capacity



Jetstream2 Features

- Primary Cloud (IU)
 - 400 compute nodes -- AMD EPYC 3rd Generation Milan CPUs – 128 cores per node + 512gb RAM
 - 90 GPU nodes – 4 x NVIDIA A100 40gb per node
 - 32 Large Memory nodes with 1TB of RAM
- Regional Clouds available by invitation/request (Arizona State, Cornell, Hawaii, TACC)
- Shared application store with common applications (NVIDIA HPC Toolkit, multiple compilers, R/Rstudio, Matlab, Anaconda, etc)
- Federated JupyterHubs, Virtual Clusters, and orchestration are all available with features being added and refined
- Commitment to **>99%** uptime
 - **99.87% availability for 9-7-22 to 3-31-23** (last NSF reporting period)





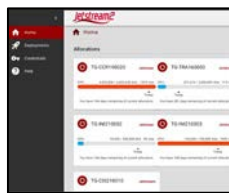
Jetstream2

Platform Overview



Horizon

OpenStack CLI



CACAO



Exosphere



Science Gateways

Authentication Service

ACCESS Accounting

OpenStack

Ceph

OpenStack



Jetstream2 Core System

Some sample use cases

- Science gateways
- Research-supporting infrastructure / Infrastructure as a service
- Education support – compute and desktops for courses, workshops, tutorials
- Domain science interactive compute
- Domain science long running compute
 - Small core counts, "pleasingly parallel", etc
- Jupyter notebooks and Hubs
- Research software development
- Machine learning – training and workflow development and data analysis
- [Your use case here]

A few key stats...

1,900

Users using Jetstream2
directly

170,000

Users using Jetstream2
via Science Gateway

67

Fields of science
represented on
Jetstream2

How do I access Jetstream2?

Jetstream2

Home > Project TG-TRA160003

Jetstream2 IU - TG-TRA160003 (logged in as jfischer@xsede.org)

Allocation usage 0 of 1,000,000 SUs

Remove Allocation [→] Create [v]

Instances used 10 of 100 total

Volumes used 9 of 50 total

Volume Name	Size
cmaaaaaaaaaart	10 GB
(Untitled volume)	20 GB
(Untitled volume)	20 GB

```
Openstack Admin - IU -- -bash -- 94x26
(openstack5) [JS2 IU Admin] [Entropy] jeremy ~-->openstack flavor list
+-----+-----+-----+-----+-----+-----+-----+
| ID | Name      | RAM   | Disk | Ephemeral | VCPUs | Is Public |
+-----+-----+-----+-----+-----+-----+-----+
| 1  | m3.tiny   | 3072  | 20   | 0          | 1     | True      |
| 13 | g3.xl     | 128000| 60   | 0          | 32    | False     |
| 2  | m3.small  | 6144  | 20   | 0          | 2     | True      |
| 3  | m3.quad   | 15360 | 20   | 0          | 4     | True      |
| 4  | m3.medium | 30720 | 60   | 0          | 8     | True      |
| 5  | m3.large  | 61440 | 60   | 0          | 16    | True      |
| 7  | m3.xl     | 128000| 60   | 0          | 32    | True      |
| 8  | m3.2xl   | 256000| 60   | 0          | 64    | True      |
+-----+-----+-----+-----+-----+-----+-----+
(openstack5) [JS2 IU Admin] [Entropy] jeremy ~-->
```

cacao

Jetstream2 Alpha Release

Thank you for participating in the Jetstream2 Alpha release. Please use the following links to view known issues and submit any additional feedback as you use the application.

VIEW KNOWN ISSUES | SUBMIT FEEDBACK

Allocations

Allocation ID	Project	CPU	Memory
TRA220028	Jetstream2 Affiliated Development Projects	48,279 / 1,000,000 SUs 5% Used	12,800 / 1,000,000 SUs 1% Used
TRA160003	Jetstream2 Staff Test Allocation	149,943 / 2,000,000 SUs 7% Used	45,303 / 2,000,000 SUs 3% Used
DS220045	Deep Learning Tutorial for Translational AI Center at Iowa State University	424,222 / 480,000 SUs 7% Used	-

Featured Learning

- Continuous Analysis 101
- Jetstream2 Basics
- Manage Resources

Project / Compute / Overview

Overview

Limit Summary

Compute

- Instances: Used 10 of 100
- VCPUs: Used 25 of 12,800
- RAM: Used 84GB of 48.8TB

Volume

- Volumes: Used 9 of 50
- Volume Snapshots: Used 0 of 10
- Volume Storage: Used 180GB of 1000GB

Network

- Floating IPs: Allocated 11 of 50
- Security Groups: Used 10 of 100
- Security Group Rules: Used 62 of 100
- Networks: Used 1 of 100
- Ports: Used 23 of 500
- Routers: Used 1 of 10

Usage Summary



<https://docs.jetstream-cloud.org/overview/overview-doc/>

Exosphere

The screenshot shows the Jetstream dashboard for Project TG-CCR190024. At the top, there's a navigation bar with the Jetstream logo, Messages, Settings, Get Support, About, and Logout. Below this, the breadcrumb trail reads 'Home > Project TG-CCR190024'. The main heading is 'iu.jetstream-cloud.org - TG-CCR190024', with 'Remove Allocation' and 'Create' buttons. The 'Instances' section shows usage bars: 11 of 25 total instances used, 26 of 132 total cores used, and 100 of 388 GB RAM used. A list of instances includes 'formally_trusty_urchin' (Ready), 'optionally_certain_longhorn with GUI' (Shelved), and 'wildly_united_mite' (Ready). A 'Show' button indicates 8 other instances are hidden. The 'Volumes' section shows 2 of 10 total volumes used and 279 of 1,100 GB storage used.

The screenshot shows the details for the instance 'formally_trusty_urchin'. The breadcrumb trail is 'Home > Project TG-CCR190024 > Instances > Instance formally_trusty_urchin'. The page title is 'iu.jetstream-cloud.org - TG-CCR190024', with 'Remove Allocation' and 'Create' buttons. The instance details include: 'Created 19 minutes ago / by user tg836338 / from image JS-API-Featured-CentOS8-Latest', 'Status: Ready', 'UUID: 2bc77f59-73bf-470f-95b6-51dc31d7577f', 'Flavor: m1.small', 'SSH Public Key Name: cmart', and 'Public IP Address: 149.165.157.3' with an 'Unassign' button. The 'Actions' section includes Lock, Suspend, Shelve, Image, Reboot, and Delete. The 'Action History' table shows a 'create' action 19 minutes ago. The 'System Resource Usage' section contains two line graphs: 'CPU Usage' (0-100% over time) and 'Memory Usage' (0-100% over time). The 'Interactions' section lists 'Web Shell', 'Web Desktop', 'Native SSH: exouser@149.165.157.3', and 'Console'. The 'Password' section provides a login instruction for 'exouser' and a 'Show password' button.



<https://exosphere.Jetstream-cloud.org> or try.exosphere.app

Using and preserving VMs

- You can install just about anything*
 - But generally limited to Linux**
- Snapshots are fairly simple and easily shared with your allocation
- One general practice is often to pull from Git(hub/lab) or pull a container

* Standard warnings about licensed software here.

** Here there be dragons.

Gateways use JS2 in several ways



Gateway web hosting



Datasets and
Database hosting



Gateway Security
Services

Integrated JupyterHub



Interactive
Computing



Elastic Virtual Clusters



Elastic Virtual Clusters

One Click OnDemand Cluster Augmenting the cloud capabilities

- Bundled lightweight HPC Stack, including SLURM.
- Users deploy scientific software with complete OS control.
- Dedicated and Responsive scheduler for rapid testing and development like workloads.
- Mounted persistent storage.

Jetstream2 Allocation/Usage Considerations

- No scheduled downtime for upgrades
 - Upgrades are done while the system stays live overall
 - 99.87% availability for 9-7-22 to 3-31-23
- Persistent IP addresses (for the life of an allocation if desired)
- No runtime limits – VMs can exist as long as there is an active allocation with SUs available
- No allocation limits for SUs – if you can justify it and we can provide it, we do
- Instance, core, and ram limits are flexible and extendable – if you can justify it and we can provide it, we do
- Storage allocations are reasonably generous – 1TB default up to 50TB in volume, shared, or object storage



The forecast calls for clouds...

- IU Research Cloud is in the process of being built
 - Repurposed Jetstream1 hardware with new storage and controllers
 - Condensed RAM from 320 computes with 128gb to 96 computes with 368gb
 - Additional RAM purchased for 1 chassis (16 nodes) of larger memory (768gb) nodes
 - 16 Intel Skylake nodes and 8 Intel Skylake nodes + NVIDIA V100s will also be integrated into ResCloud from Jetstream1 hardware

ResCloud Use Cases

- Primary short-term goal 1: provide system for Research Desktop services with room to expand
- Primary short-term goal 2: condense the RT services scattered over multiple virtualization systems onto one cloud
- Longer term goals: Provide a platform for new RT-run research services and for IU researcher-run VMs
- Other goals: Exploring accelerated research desktops with the V100s

Present state of ResCloud

- Several compute chassis online with more coming online
- Completing the network fault tolerance
- RED is up and running on ResCloud after a few cloud-induced hiccups
- Research Cloud Services has access to ResCloud now to start their testing and planning for service migrations to ResCloud



PERVASIVE TECHNOLOGY INSTITUTE



UNIVERSITY INFORMATION TECHNOLOGY SERVICES
RESEARCH TECHNOLOGIES

Jetstream2



National Science Foundation
Award #ACI-2005506

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 the Research Cloud Infrastructure team –
Mike, Steve, Aaron, and Sarah as well as the Jetstream2 PI
David Y. Hancock, Malinda Husk, Winona Snapp-Childs, and
George Turner (ret.)*



PERVASIVE TECHNOLOGY INSTITUTE



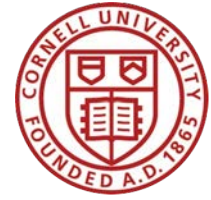
UNIVERSITY INFORMATION TECHNOLOGY SERVICES
RESEARCH TECHNOLOGIES

Jetstream2

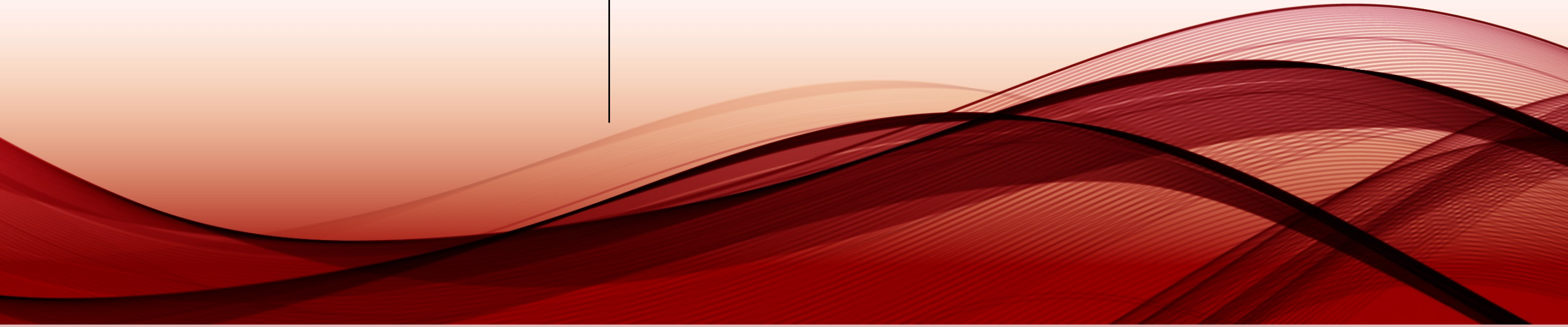


National Science Foundation
Award #ACI-2005506

Partners



JOHNS HOPKINS
UNIVERSITY





AI
For
EVERYONE

Jetstream2

