



Getting Started on Jetstream2

Danny Havert – Indiana University

Systems Programmer/Analyst

Research Cloud Services

djhavert@iu.edu

Zach Graber – Indiana University

Systems Programmer/Analyst

Research Cloud Services

zegraber@iu.edu

Jenn Taylor – Indiana University

IT Project Manager

Research Cloud Services

jlrobiso@iu.edu

Julian Pistorius – Indiana University

Research Software Developer

Research Cloud Services

jupist@iu.edu



Agenda

Introduction to Jetstream2 (45 min)

- ACCESS account and training allocation
- What is Jetstream2? Whom is it for?
- How to apply for an allocation - ACCESS

Getting access to Jestream2 & Break (15 min)

- Logging in to Exosphere

Using Jetstream2 - Basics (45 min)

- Creating and accessing your instance
- Using the software collection
- Volume/Storage management
- Transferring Data to and from your instance
- Saving and Sharing Images
- Shelving Instance

Using Jetstream2 - Advanced (45 min)

- Other Interfaces: Horizon and CACAO
- GPU instances and 3D accelerated Desktop
- Additional Requests by Attendees



Accessing Jetstream2

Before you can use Jetstream2 you must first have an ACCESS ID and be on an active Jetstream2 allocation! Fill out this form to participate in the tutorial.

<https://forms.gle/JSbMkfXURPLMcaYA9>





AI For Everyone

What is Jetstream2?

- A user-friendly cloud computing environment for researchers and educators.
- NSF funded. Available to any US based researcher or educator.
- CPU, GPU, and large memory nodes.
- On-demand interactive computing and data analysis resources.
- Infrastructure for gateways and other “always on” services.
- Available for Education - use it for courses!
- Virtual clustering, programmable cyberinfrastructure



What's Special About Cloud Computing?

- Create your own virtual machine (VM or 'instance')
- Full sudo (admin) access. Use your instance how you want.
 - Install the software you need.
- On-demand resources. No sharing or queue times.
 - Keep your jobs running as long as you want.
- Interactive computing environment
 - With graphical desktop
- Full internet access with Persistent IPs
 - Great for web hosting

Jetstream2 Instance Sizes

- Three different resources:

- CPU

- Up to 64 cores, 250 GB RAM

- Large Memory

- Up to 128 cores, 1000 GB RAM

- GPU

- Up to 32 CPU cores, 125 GB RAM, and a full NVIDIA 40GB Ampere A100 GPU

- Storage

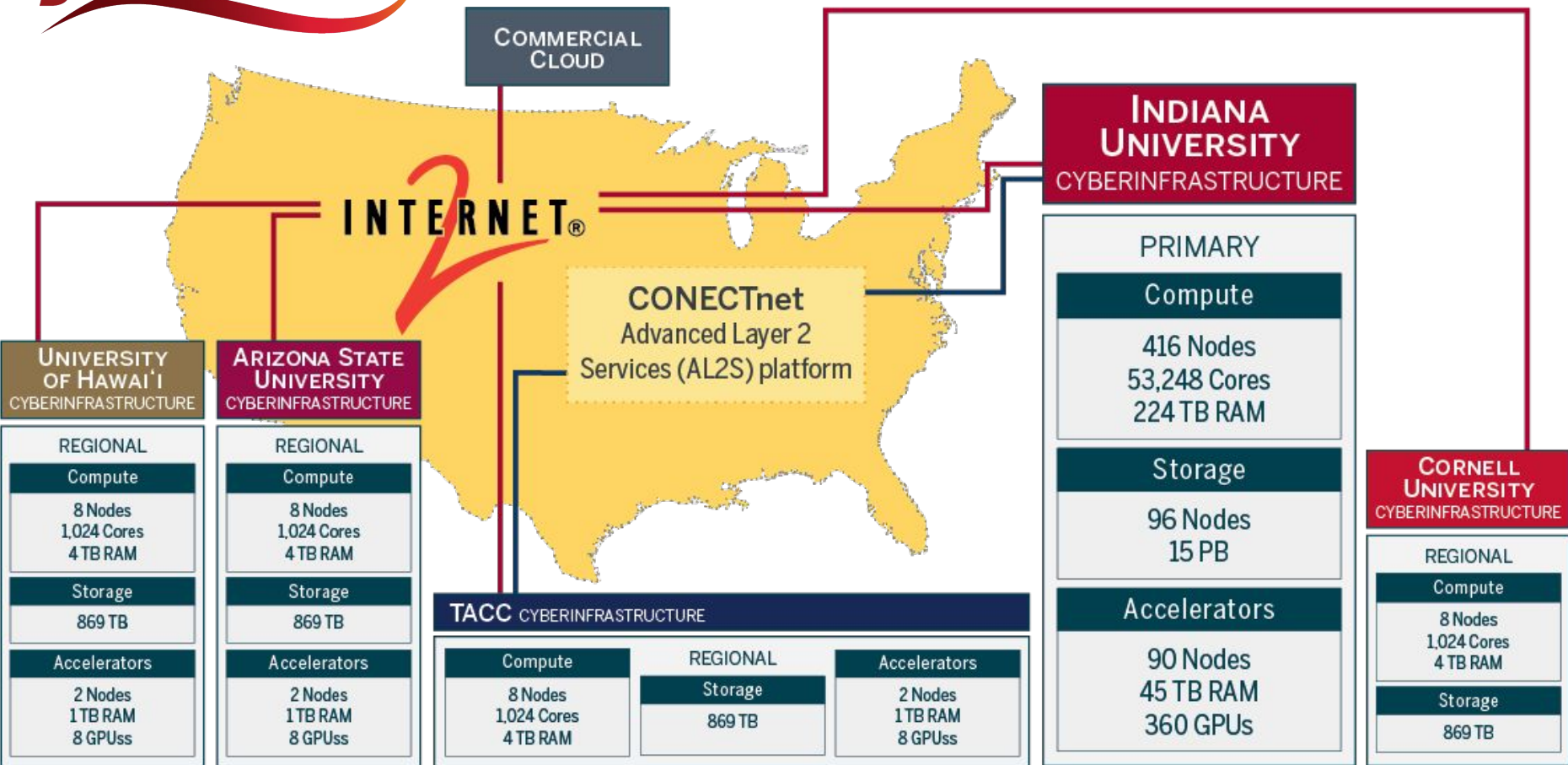
- Default: 1 TB
- More on Request

View the full list:

<https://docs.jetstream-cloud.org/general/vmsizes/>



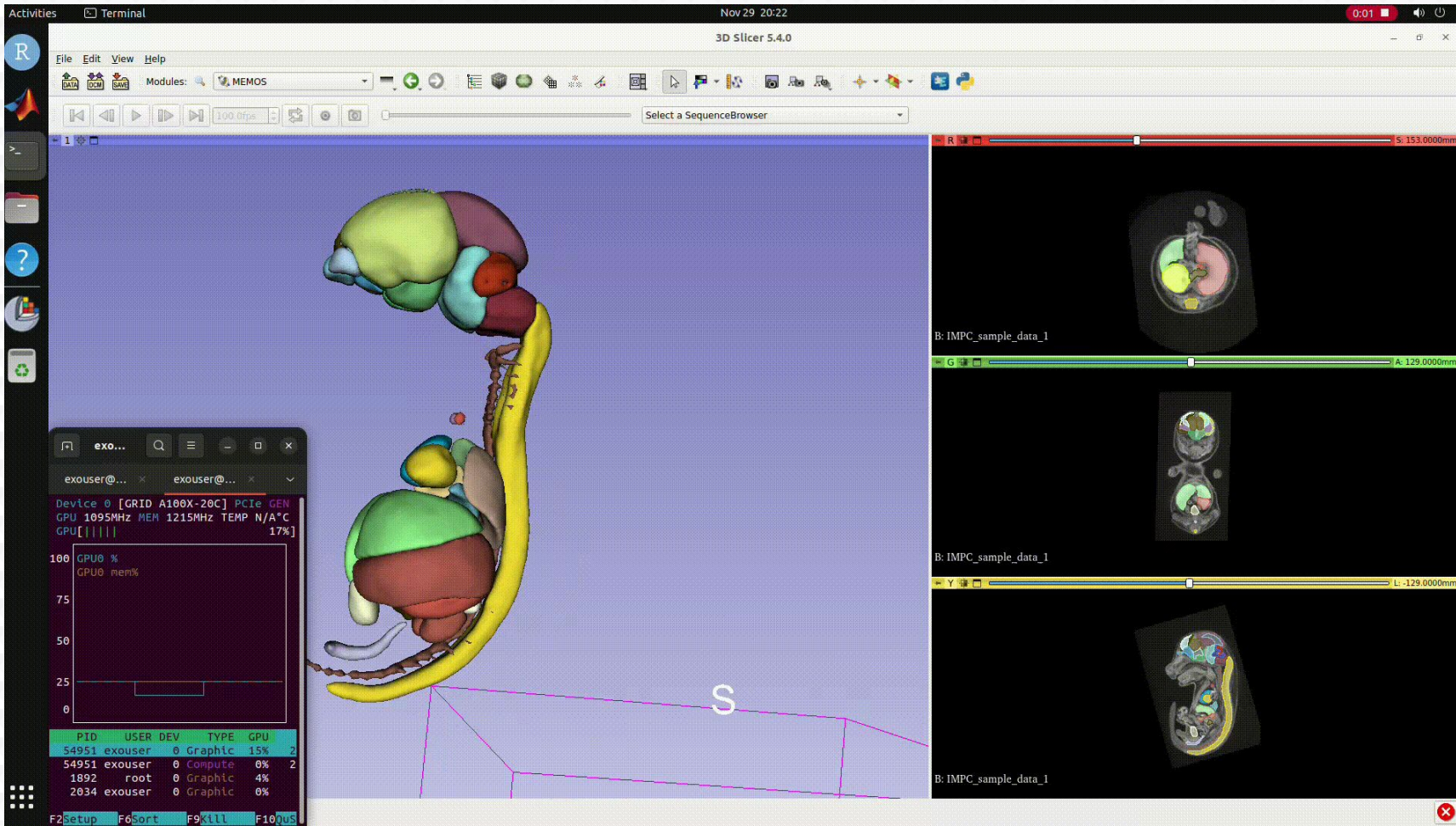
Jetstream2



Use Cases

- On-demand access to powerful computer for high performance analysis.
 - Useful for teaching a course in which students have this need.
- Teach a course where each student gets their own VM to work on.
 - Network security, reproducible development environment, etc.
- Write, debug, and execute code with interactive GUI applications
 - Use a GPU if needed.
- Host a file server or database. Receive, store, and serve data to colleagues.
- Host a website or web-app to share with students or colleagues.

Use Case: ML & 3D Biomedical Visualization



The image displays the 3D Slicer 5.4.0 software interface. The main window shows a 3D model of a biological specimen, likely a mouse embryo, rendered in various colors (yellow, green, red, blue, purple). The model is positioned on a white grid. The interface includes a menu bar (File, Edit, View, Help), a toolbar with various icons, and a status bar at the bottom. A terminal window is open in the bottom-left corner, displaying system metrics and a table of process information.

Terminal Output:

```
Device 0 [GRID A100X-20C] PCIe GEN  
GPU 1095MHz MEM 1215MHz TEMP N/A°C  
GPU[||||| 17%]  
  
100 GPU0 %  
GPU0 mem%  
75  
50  
25  
0  
  
PID USER DEV TYPE GPU  
54951 exouser 0 Graphic 15% 2  
54951 exouser 0 Compute 0% 2  
1892 root 0 Graphic 4%  
2034 exouser 0 Graphic 0%
```

Three orthogonal viewports are visible on the right side of the interface, showing different views of the 3D model:

- Top viewport: Axial view (labeled 'B: IMPC_sample_data_1') showing a cross-section of the specimen.
- Middle viewport: Coronal view (labeled 'B: IMPC_sample_data_1') showing a front-to-back cross-section.
- Bottom viewport: Sagittal view (labeled 'B: IMPC_sample_data_1') showing a side view of the specimen.

The viewports include sliders for zooming and panning. The top viewport has a red slider bar, the middle has a green slider bar, and the bottom has a yellow slider bar.

Use Case - Science Gateways and Portals

Galaxy:
“enables scientists to share, analyze and visualize their own data via a browser.”



The screenshot displays the Galaxy web interface. The top navigation bar includes "Workflow", "Visualize", "Shared Data", "Help", "User", and a notification bell. The main content area shows a workflow titled "Draw nucleotides distribution chart (Galaxy Version 1.0.1)". The workflow includes a "Statistics text file" tool that takes "6: FastQC on data 1: Raw Data" as input. Below the tool name, there are options for "Email notification" (set to "No") and an "Execute" button. A "What it does" section explains that the tool creates a stacked-histogram graph for nucleotide distribution in Solexa libraries. A "TIP" suggests using the "FASTQ Statistics" tool to generate a report file. An "Output Examples" section shows a barcode sequence: "GATCT". Below this is a stacked histogram chart with a legend on the right. The chart's x-axis is labeled "Nucleotide" and the y-axis is "Number of reads". The legend indicates colors for A (blue), C (red), G (green), and T (yellow). The histogram shows a clear peak for 'G' at the 5th position. Below the chart, a note states: "In the following chart, one can almost 'read' the most abundant sequence by looking at the dominant values: GATCT". The right sidebar shows a "History" panel with a search box and a list of recent jobs, including "6: FastQC on data 1: Raw Data", "5: FastQC on data 1: Web page", "4: FastQC on data 2: Raw Data", "3: FastQC on data 2: Web page", "2: EBI SRA: ERR3988762", and "1: EBI SRA: ERR3988762". The bottom of the browser window shows the URL: https://usegalaxy.org/tool_runner?tool_id=toolshed.g2.bx.psu.edu%2Frepos%2Fdevteam%2Ffastx_nucleotides_distribution%2Fcschl_fastx_nucleotides_distribution%2F1.0.1

Use Case - Science Gateways and Portals

Brainlife:
“providing an online, community-based platform where users can publish code (Apps) and Data while integrating HPC and cloud-computing resources to run Apps.”

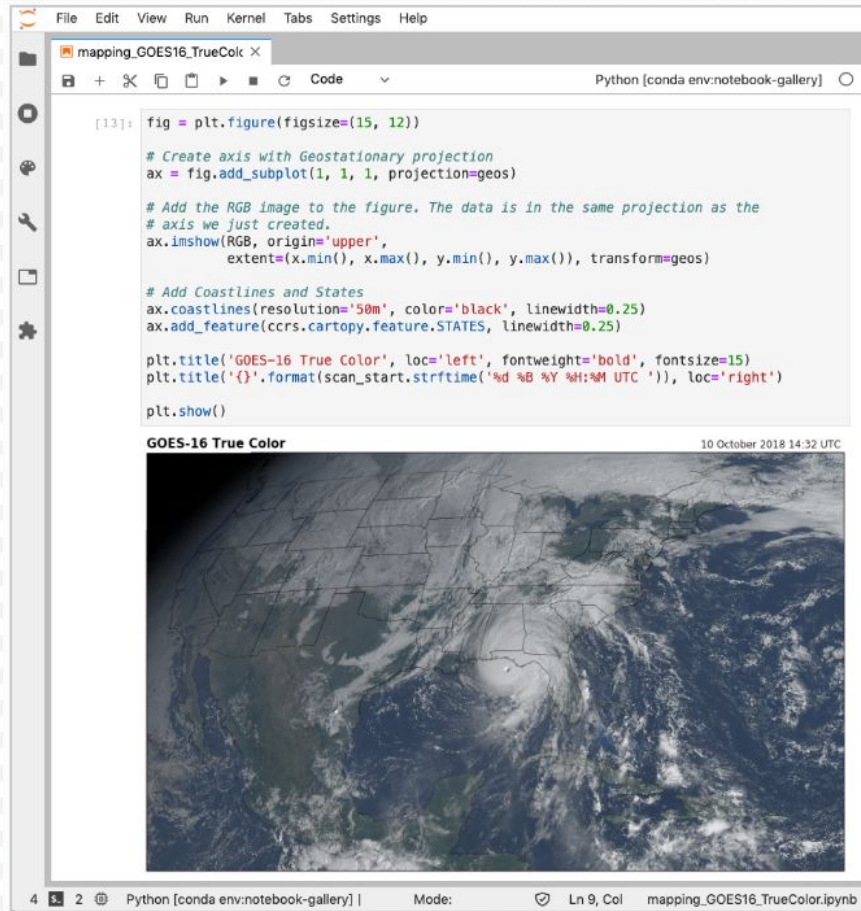
The screenshot displays the Brainlife app marketplace interface. On the left is a navigation sidebar with categories: DATASETS, PROJECTS, APPS (selected), PUBLICATIONS, DATATYPES, RESOURCES, and SETTINGS. Below the sidebar is the user profile for Danny Havert, including SLACK/CHAT, REPORT ISSUES, and DOCUMENTATION links. The main area is titled 'Search Apps' and features a grid of application cards. Each card includes the app name, a brief description, input/output information, and performance metrics (downloads, likes, time, and success rate). The apps shown include:

- Network Neuro Aggregator**: Aggregates outputs from networkneuro. (30 downloads, 8 likes, 5-15 min, 79.2% success)
- White Matter Anatomy Segmentation**: Automated segmentation of major white matter tracts. (43,403 downloads, 81 likes, 2.3-3.1 hour, 66.9% success)
- Tract Analysis Profiles**: Create plots of diffusion metrics. (23,012 downloads, 63 likes, 0-2.5 hour, 75.0% success)
- Tract Analysis Profiles - Dipy**: Create plots of diffusion metrics using Dipy. (23,012 downloads, 63 likes, 0-2.5 hour, 75.0% success)
- Tract Profile Aggregator**: Aggregates outputs from tractprofile-deprecated. (14 downloads, 4 likes, 7-53 min, 64.3% success)
- AFQ Tract Classification with LIFE**: Identifies major tract segments and quantifies tissue properties. (3,477 downloads, 35 likes, 25-84 min, 50.7% success)
- Tract Analysis Profiles (NODDI)**: Create plots of diffusion metrics for NODDI. (23,012 downloads, 63 likes, 0-2.5 hour, 75.0% success)
- Tract Analysis Profiles - Myelin mapping**: Create plots of diffusion metrics for myelin. (23,012 downloads, 63 likes, 0-2.5 hour, 75.0% success)
- LIFE (beta)**: Predicts measured diffusion signal using fascicle orientation. (13,134 downloads, 43 likes, 1.4-17.3 hour, 86.9% success)
- AFQ Tract Classification**: Automated fiber quantification and fe structure output. (3,477 downloads, 35 likes, 25-84 min, 50.7% success)
- C-PAC**: Configurable Pipeline for the Analysis of Connectomes. (3,907 downloads, 36 likes, 1.1-3.6 hour, 70.1% success)
- FreeSurfer 7.3.2**: Segments the tlw anatomical data into functionally different parts. (442,880 downloads, 530 likes, 6-7.6 hour, 47.8% success)
- TractSeg**: Tool for fast and accurate white matter bundle segmentation. (24,975 downloads, 82 likes, 0.8-2.7 hour, 67.6% success)
- Segment thalamic nuclei**: Segments the thalamus into its multiple components. (7,455 downloads, 23 likes, 12-36 min, 83.0% success)
- FreeSurfer 7.1.1**: Segments the tlw anatomical data into functionally different parts. (442,880 downloads, 530 likes, 6-7.6 hour, 47.8% success)
- TractSeg from peaks**: Tool for fast and accurate white matter bundle segmentation. (24,975 downloads, 82 likes, 0.8-2.7 hour, 67.6% success)

Use Case - Live Data Processing and Distribution

Unidata:

“sharing geoscience data and the tools to access and visualize that data.”



How to Use Jetstream2

- We have three Dashboards
 - Exosphere - The Primary dashboard. Most user friendly. Start here!
 - Stay for the interactive tutorial to see this in action.
 - Horizon - The default Openstack dashboard. For advanced users. More tools but not as user friendly.
 - CACAO - Still in active development but available to use. For advanced users. Useful for building clusters.
- Use Dashboard to...
 - Create, Shelve, and Delete Instances
 - Create and Attach/Remove Volumes
 - Access your instance via Web Shell or Desktop (Exosphere only)
- Command Line Interface - for programmatic interaction



Data Storage

- Jetstream2 is NOT primarily a Storage Device
 - Treat storage as a temporary location to stage data for analysis.
 - Default of 1 TB storage.
- Storage Options
 - Volumes - The default storage type
 - Attach to an instance to read and write your data from your programs.
 - Manilla - Shared file systems
 - accessible by multiple instances
 - Object Store - 'bucket' to store large amounts of data without a filesystem
 - S3 compatible

*Data that must be protected by Federal security or privacy laws (e.g., HIPAA, FERPA, ITAR, classified information, export control, etc.) is NOT permitted to be stored on Jetstream2 unless such storage and usage is specifically authorized by the responsible University administrator and complies with any processes for management of access to such information.



How to Access Jetstream2

Allocations are managed by ACCESS



What is Access?



Advancing
Innovation

Provides Access to National Advanced Computing Resources

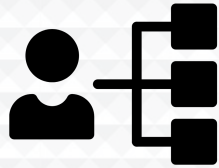
<https://access-ci.org/>

The NSF's ACCESS (Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support) program builds upon the successes of the 11-year XSEDE project, while also expanding the ecosystem with capabilities for new modes of research and further democratizing participation.

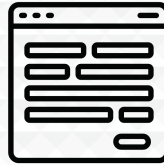
Allocation Process



CREATE
ACCOUNT



SELECT
OPPORTUNITY



REQUEST
ALLOCATION



RECEIVE
CREDITS



EXCHANGE
CREDITS

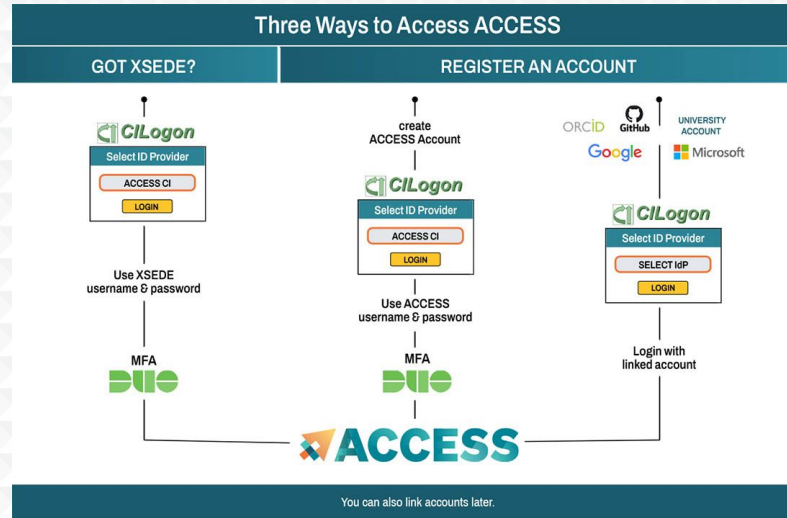
Getting an ACCESS ID



CREATE
ACCOUNT

In order to Apply for an allocation OR be able to login to any resources an Access ID is required.

Go To <https://identity.access-ci.org/new-user>



Getting Your Own Allocation



SELECT
OPPORTUNITY



<https://access-ci.org/>

Applying for an allocation is like applying for a small grant that provides computational resources.

The first level of application “Explore ACCESS” is a single form and only takes a few minutes. Allocations are usually approved in a few days.
~\$40,000 worth of cloud resources for a few minutes of effort.



SELECT
OPPORTUNITY

Allocation Tiers

Explore	Discover	Accelerate	Maximize
400,000 credits	1,500,000 credits	3,000,000 credits	N/A
CV and Brief Summary	CV & 1-page proposal	3-page proposal & panel review	10 page proposal & panel review
Grad student projects	Modest resource needs	Mid-scale resource needs	Large-scale projects in need of massive amounts of computational resources
Small labs	Mid-sized projects	Multi-grant collaborative projects	
Benchmarking	Campus Champions	Science Gateways with growing communities	
Small classes	Large classes		



REQUEST
ALLOCATION

Allocation Details

- Allocation must be submitted by the PI of the project
- Who can be a PI?
 - Any educator or research at a U.S. academic, non-profit research, or educational institution
- What about grad students?
 - Yes! Can receive Explore tier, just need a letter of collaboration from your PI.
- Anyone with an ACCESS ID can be added to a project
 - Getting an ACCESS ID is free and easy - useful for adding students
- Allocations must be renewed annually
- Projects can be “upgraded” by applying for the next-level allocation tier

Allocations Are Awarded Credits



- Credits are exchanged for computational time or storage on supercomputers
- For Jetstream2, 1 ACCESS credit = 1 Service Unit (SU) = 1 CPU hour
- What are some common usage scenarios?
 - Running the equivalent of a high-end laptop (8 cores, 32 GB RAM) for 40 hours a week for 2 years
 - 35,000 SUs
 - A server that can handle moderate traffic for 5 years uninterrupted
 - 175,000 SUs
 - Using an entire A100 GPU on average 40 hours a week for 2 years
 - 530,000 SUs
 - Note: We have smaller GPU instances that aren't as costly

What Resources are Available?



EXCHANGE
CREDITS

- Indiana Jetstream2
- Indiana Jetstream2 GPU
- Indiana Jetstream2 Large Memory
- Indiana Jetstream2 Storage



Indiana University (IU)
Jetstream2



Johns Hopkins University
(JHU)
Rockfish



National Center for
Supercomputing
Applications (NCSA)
Delta



Open Science Grid (OSG)



Open Storage Network
(OSN)



Pittsburgh Supercomputing
Center (PSC)
Bridges-2



Purdue
Anvil



San Diego Supercomputer
Center (SDSC)
Expanse



Stony Brook University
Ookami



Texas Advanced
Computing Center (TACC)
Stampede2
Ranch



Texas A&M University
(TAMU)
FASTER



University of Delaware
DARWIN



University of Kentucky
KyRIC



Accessing Jetstream2

 Go To <https://jetstream2.exosphere.app/>

Before you can use Jetstream2 you must first have an ACCESS ID and be on an active Jetstream2 allocation!



Documentation and Support

Jetstream2 Website

<https://jetstream-cloud.org/index.html>

Reach out to Jetstream2 support at

help@jetstream-cloud.org

Documentation:

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

Or directly to Research Cloud Services:

rcsadm@iu.edu

Jetstream2 Login (Exosphere)

<https://jetstream2.exosphere.app/>

My email:

djhavert@iu.edu



Demo

Acknowledgements

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

This material is based upon work supported by 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 Jetstream2 contributors

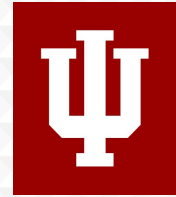
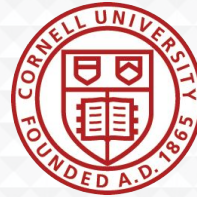
- PI David Y. Hancock, J. Michael Lowe, Malinda Husk, Winona Snapp-Childs, George Turner, and Jeremy Fischer

And the rest of the Jetstream2 Support team

- Zach Graber, Julian Pistorius, Jenn Taylor and Le Mai Weakley.



Jetstream2 partners



JOHNS HOPKINS
UNIVERSITY



UCAR



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