AN INTRODUCTION TO

Jetstream2

Zachary Graber
Systems Programmer/Analyst, Research Cloud Services
UITS Research Technologies, Indiana University

PERVASIVE
TECHNOLOGY INSTITUTE

RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES

Jetstream2 is a **flexible, user-friendly cloud computing environment** designed for everyone from researchers with minimal high-performance computing experience to software engineers looking for the latest in cloud-native approaches.

Jetstream2

PERVASIVE
TECHNOLOGY INSTITUTE

RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES
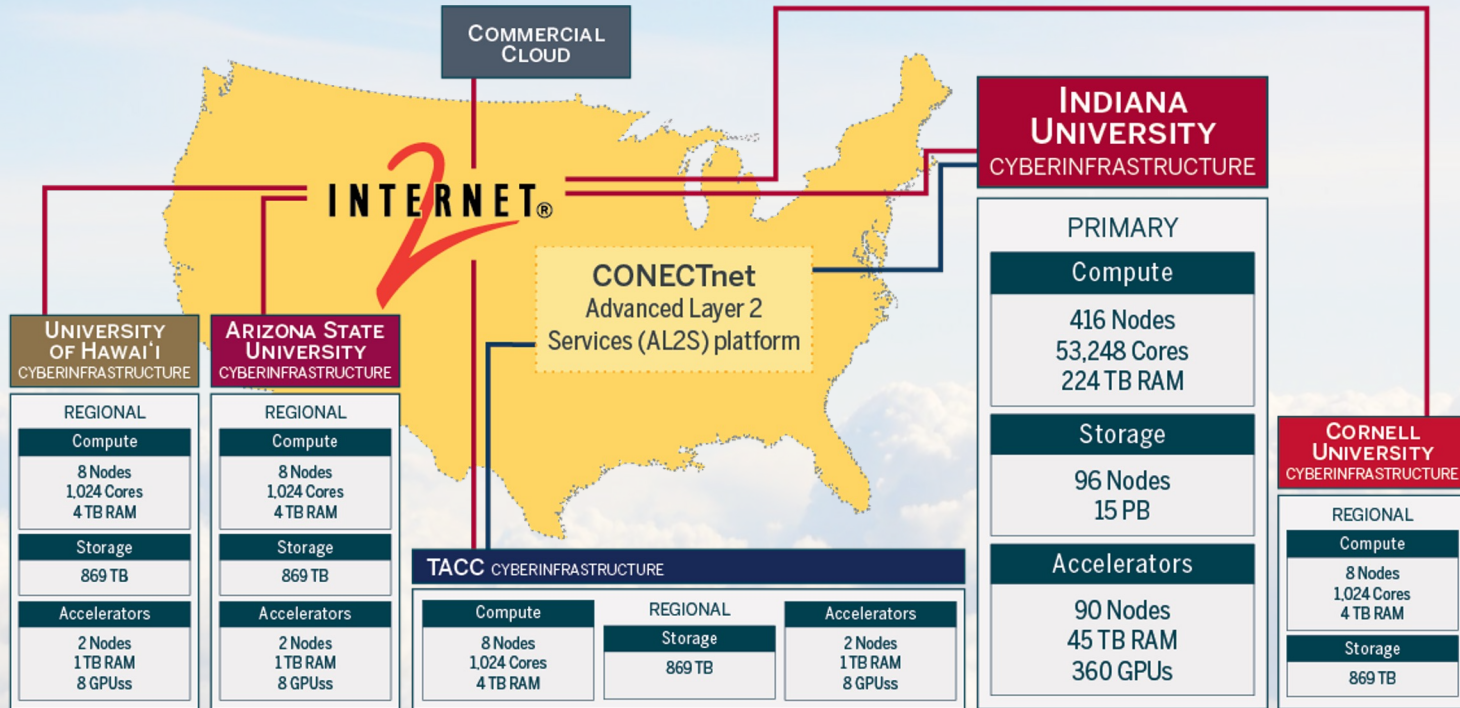
WHAT DOES JETSTREAM2 OFFER?

- **Scalable, customizable, on-demand computation**
  - NVIDIA A100 GPUs
  - Large Memory nodes (up to 1 TB RAM)

  - No waiting in queues or runtime limits

- **Scratch storage**
  - Default 1 TB per allocation
  - Network-attached shares

  - Amazon S3-compatible object store

- **Flexible networking**
  - Persistent public IP addresses
  - 100 Gb/s from compute hosts
  - Network-level firewall options

  - (Optional) user-defined routers & subnets

Jetstream2

PERVASIVE
TECHNOLOGY INSTITUTE

RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES

- **Various management interfaces**
  - **Exosphere** makes it easy to get started and access your resources
    - Browser-based "Web shell" and "Web desktop"
  - **CyVerse CACAO** aims to make complicated deployments simple
    - Fast templates for JupyterHub, Kubernetes, etc.
  - **Powerful API/CLI** for advanced users

- **Experimental features**
  - Virtual elastic Slurm clusters
  - One-click Binder deployments for interactive notebooks

- **Zero cost to the user**

WHAT CAN YOU DO WITH JETSTREAM2?

- **Receive and process data**

- **Write, debug, and execute code** with interactive GUI applications

- **Train, refine, and run machine learning models** with GPUs

- **Host a file server, database, or portal**

- **Share your applications** via static or dynamic website

- **Provide educational cyberinfrastructure** for workshops and courses

Jetstream2

PERVASIVE TECHNOLOGY INSTITUTE   RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES

Galaxy

Tools

search tools

⬆ Upload Data

Get Data
Send Data
Collection Operations

GENERAL TEXT TOOLS

Text Manipulation
Filter and Sort
Join, Subtract and Group
Datamash

GENOMIC FILE MANIPULATION

FASTA/FASTQ
FASTQ Quality Control
SAM/BAM
BED
VCF/BCF
Nanopore
Convert Formats
Lift-Over

COMMON GENOMICS TOOLS

Interactive tools
Operate on Genomic Intervals
Fetch Sequences/Alignments

GENOMICS ANALYSIS

Assembly
Annotation
Mapping

Workflow   Visualize   Shared Data   Help   Log in or Register   Using 0%

Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources. You can install your own Galaxy by following the tutorial and choose from thousands of tools from the Tool Shed.

Paper Alert!

Scalable, accessible, and reproducible reference genome assembly and evaluation in Galaxy

Larivière, D. et al. 2023

READ NOW!

Donate to the James P. Taylor Foundation for Open Science

Learn More

Want to learn the best practices for the analysis of SARS-CoV-2 data using Galaxy? Visit the Galaxy SARS-CoV-2 portal at https://galaxyproject.org/projects/covid19/

PennState

Galaxy
PROJECT

https://galaxyproject.org/

- Online data analysis platform
- Tackling transparency, collaboration, and reproducibility in biomedical research
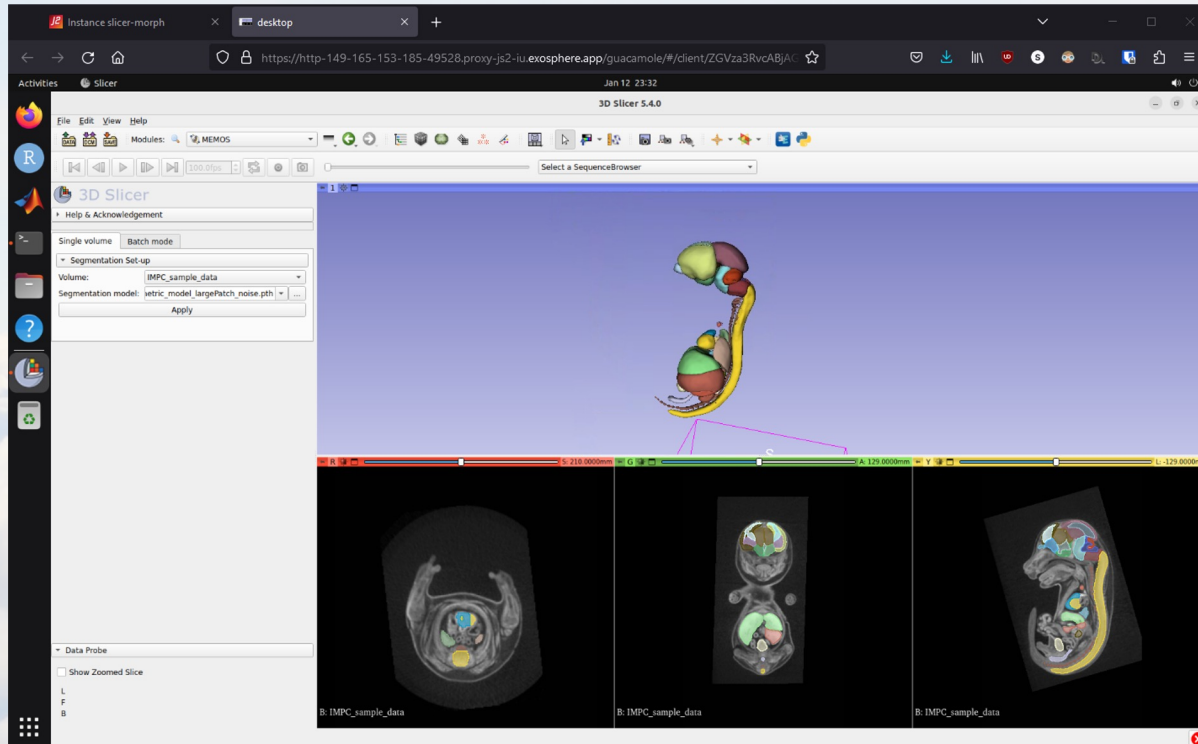- Using several ACCESS resources

"Advancing endangered species monitoring using bioacoustics and machine learning"

- Using Jetstream2 GPU

- Creating/training AI classification models

- Using those classifiers on tens of TB of passive

  acoustic monitoring data collected in Hawai'i

Photo by Kvnga on Unsplash

Example: 3D Slicer

JETSTREAM2 BENEFITS/IMPACT

- **Full sudo/admin access** to choose your operating system and install software you need

- **Broad range of resource "flavors"** including Large Memory and NVIDIA vGPU slicing
  - Easy and equitable access to GPUs helps spearhead AI/ML research nationally

- **Available on demand** without sharing or queues, and no runtime limit

- **Free to use** through our support from the National Science Foundation and ACCESS
  - "ACCESS allocations are available to any researcher or educator at a U.S. academic, non-profit research, or educational institution."
  - Available at any level, from community college to R1 university

ACCESS | Advancing Innovation

access-ci.org/about/get-started

1. **Create an account with ACCESS.**
   Required to apply for and log in to resources

2. **Choose your opportunity and submit your request.**
   Four tiers of allocations with their own application requirements

3. **Receive and spend your ACCESS credits.**
   1 ACCESS credit == 1 Jetstream2 SU == 1 GB; split however you'd like across Jetstream2's CPU, GPU, LM, and storage resources

Jetstream2

PERVASIVE TECHNOLOGY INSTITUTE   RESEARCH TECHNOLOGIES
UNIVERSITY INFORMATION TECHNOLOGY SERVICES

Jetstream2 Website

https://jetstream-cloud.org

Jetstream2 Support

help@jetstream-cloud.org

Documentation

https://docs.jetstream-cloud.org

Advanced Cyberinfrastructure Coordination
Ecosystem: Services & Support (ACCESS)

https://access-ci.org/