CIROH CyberInfrastructure And Case Studies

CIROH

Developers CONFERENCE

May 29 - June 1, 2024 | Salt Lake City, Utah

CIROH CyberInfrastructure and case studies

Agenda:

- Introduction
- CIROH Survey Results
- CIROH UA CyberInfrastructure Overview

Case Studies

- Tethys Platform and Google BigQuery (**BYU**)
- NextGen In A Box (NGIAB) and ngen-dataStream **(University of Alabama, Lynker, BYU, Aquaveo)**
- Integrated Evaluation System Prototype for Testing Research and Operational Advancements (**RTI**)
- Hydrology AI framework **(University of Iowa)**
- Automated System for River Ice Monitoring using Remote Sensing and Deep Learning Techniques Across the United States **(Stevens Institute of Technology)**
- Community Streamflow Evaluation System (CSES) **(University of Alabama)**
- Modernized Standards and Tools for Sharing and Integrating Real-time Hydrologic Observations Data **(Utah State University)**
- Using CIROH HPC and Cloud Resources for ML workflow **(Pennsylvania State University)**

Q&A

CIROH CyberInfrastructure And Case Studies

Presenter: James Halgren

Assistant Director of Science and Research Operations

CIROH CyberInfrastructure And Case Studies

- Multi/Hybrid Cloud and HPC
- Scalable Resources
- Pre-configured Environments
- Pre-installed Software
- Secure and Rapid Deployment

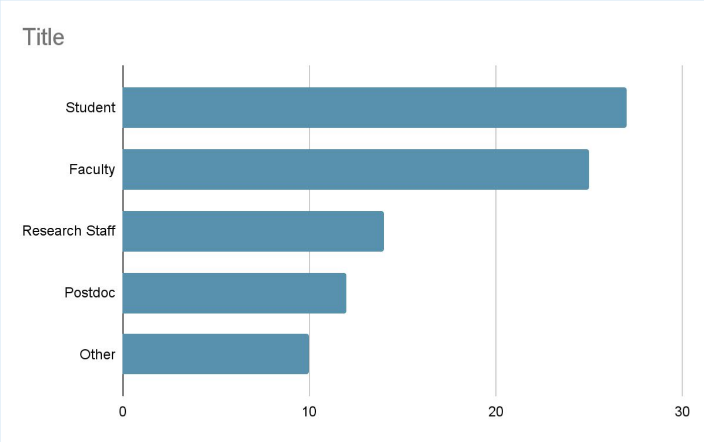
CIROH Survey Results

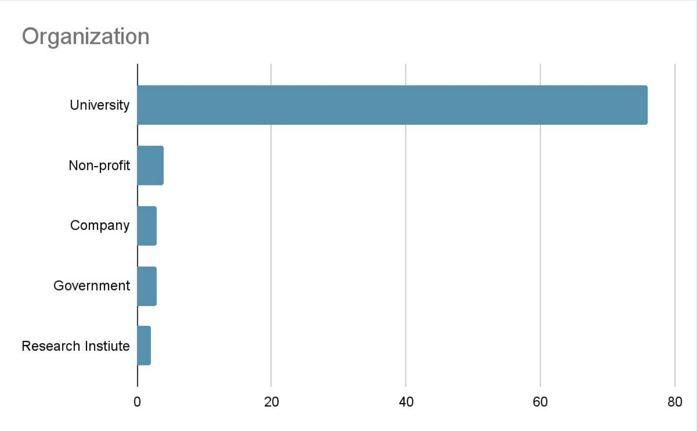
Presenter: **Dr. Jeffrey Carver**
James R. Cudworth Professor and Graduate Program Director
University of Alabama

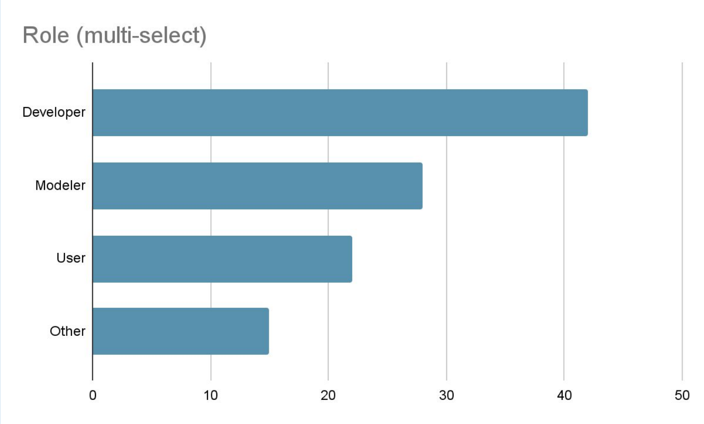
CIROH Survey Objectives

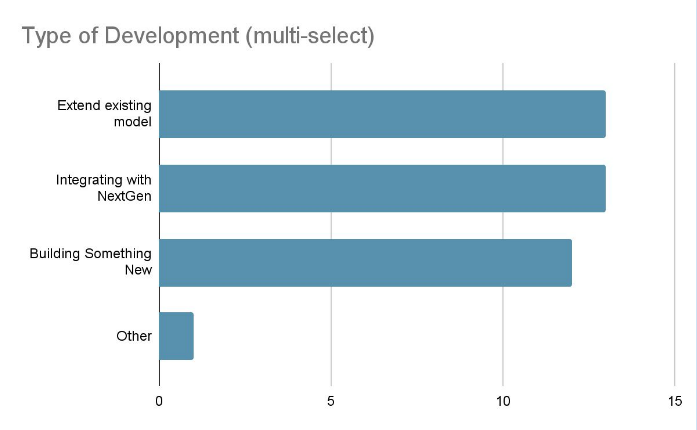
- Characterize the development environment and current practices and identify areas for needed improvement
- Understand community needs
- Identify training needs
- Build community of people with common interests

CIROH Survey: Demographics

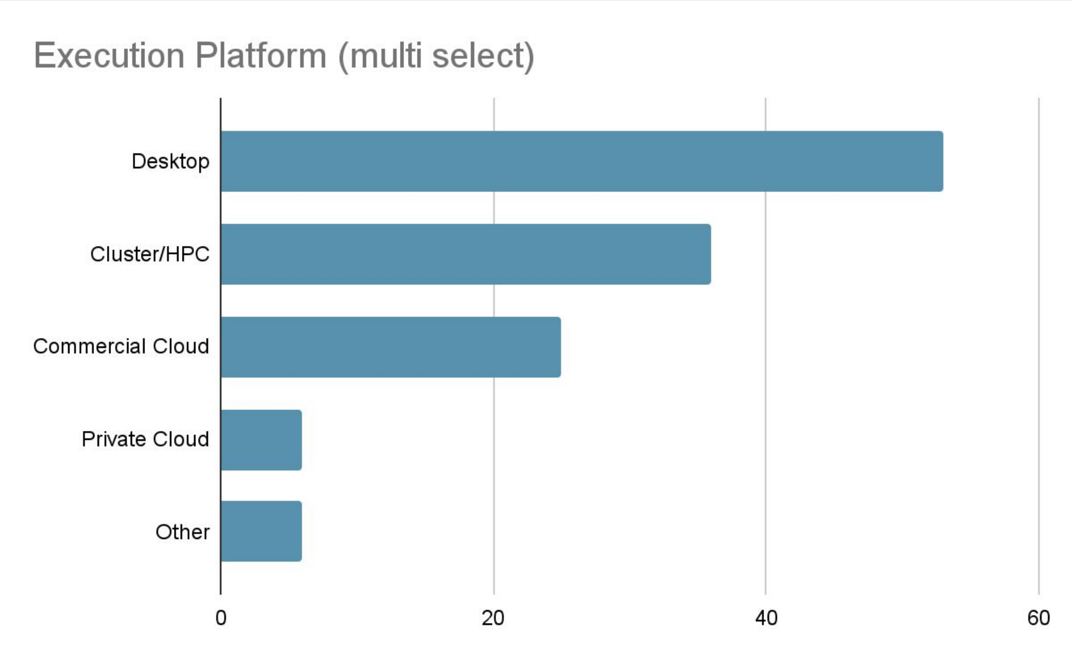








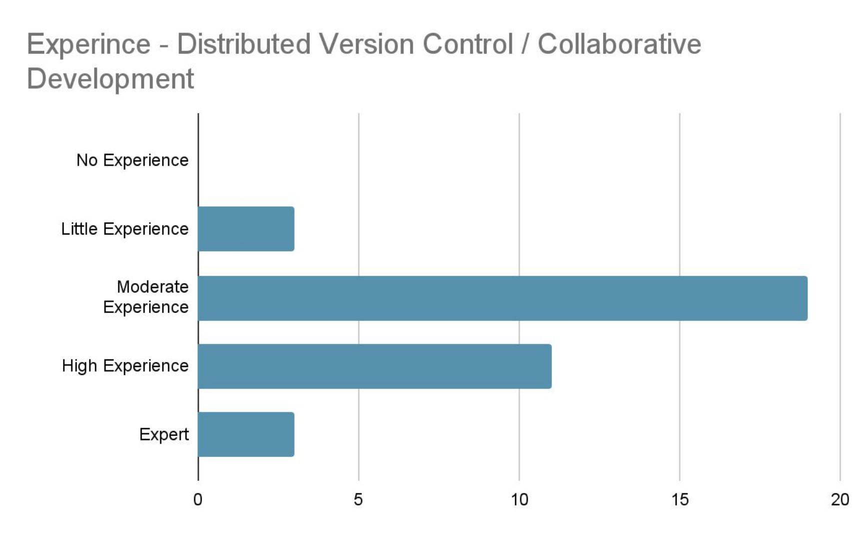
CIROH Survey: Computational Infrastructure

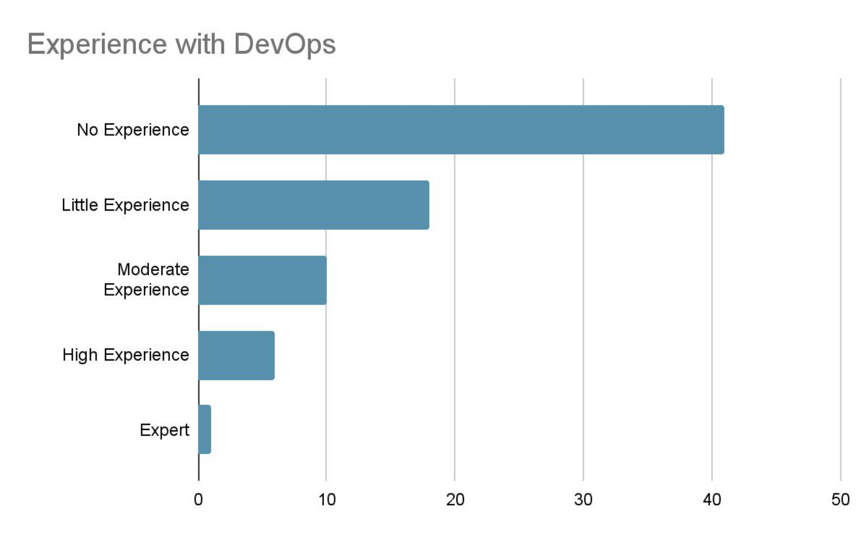


Training Needs

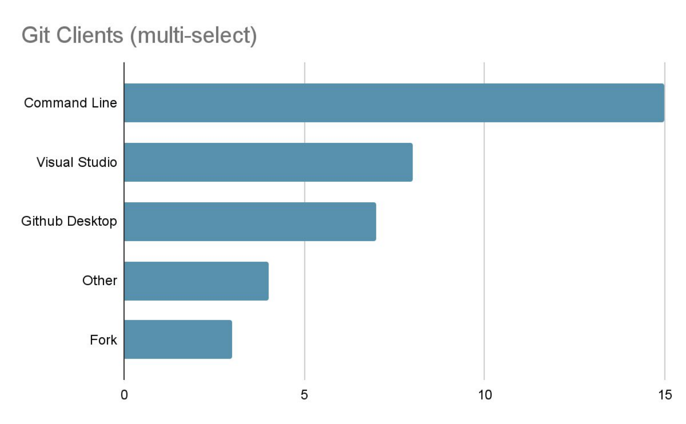
- Cloud (5)
- HPC for Hydro (4)
- NextGen (3)
- Distributed Computing (3)
- SE (3)
- Hydrological Modeling (2)
- Python (2)
- ML (2)

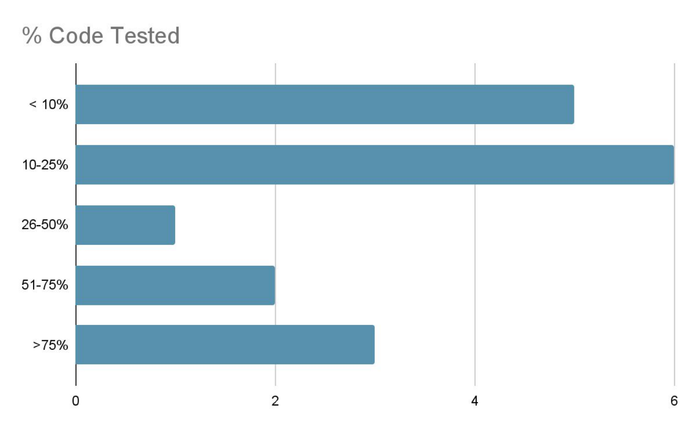
CIROH Survey: Software Development Process





CIROH Survey: Software Development Process

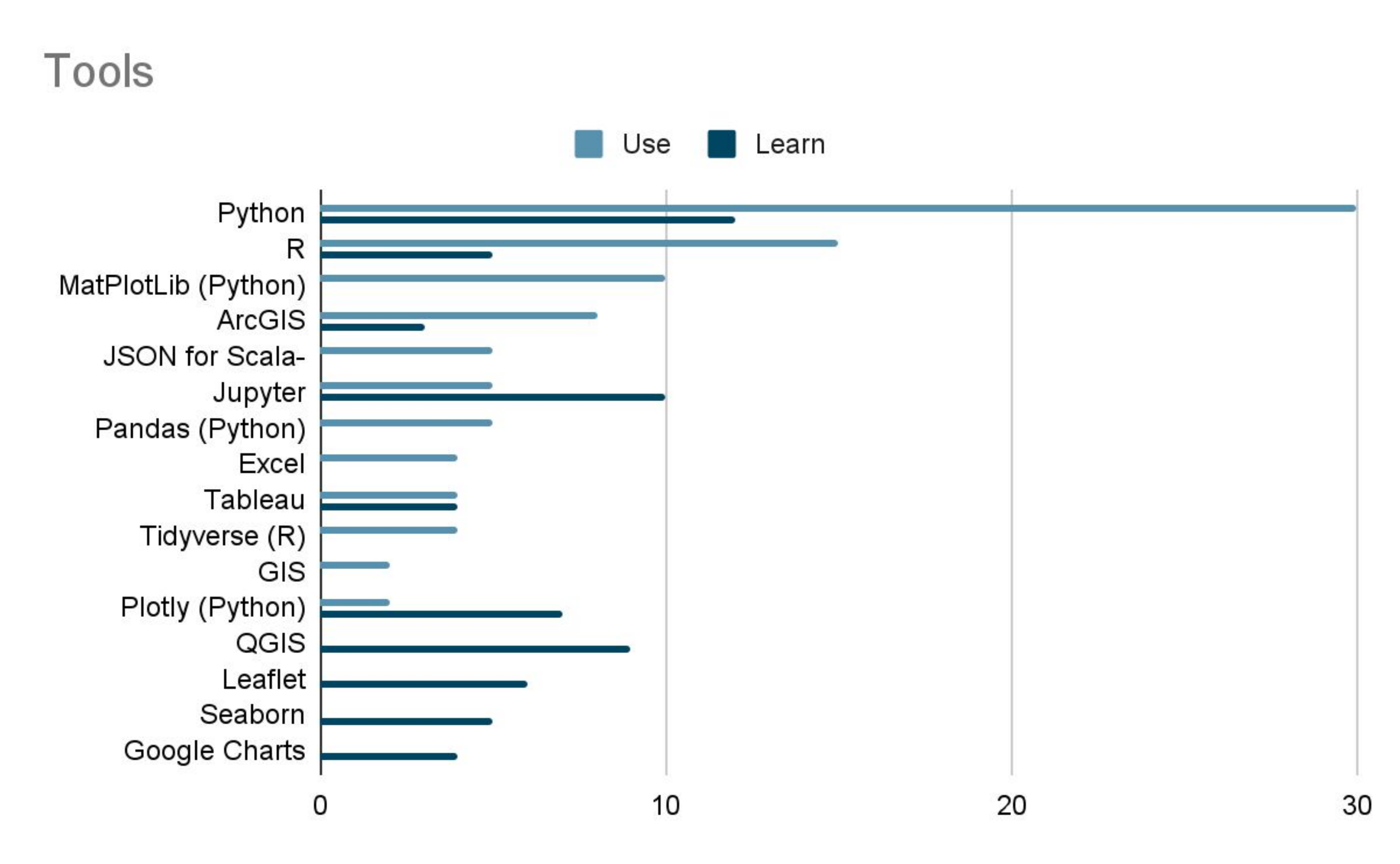




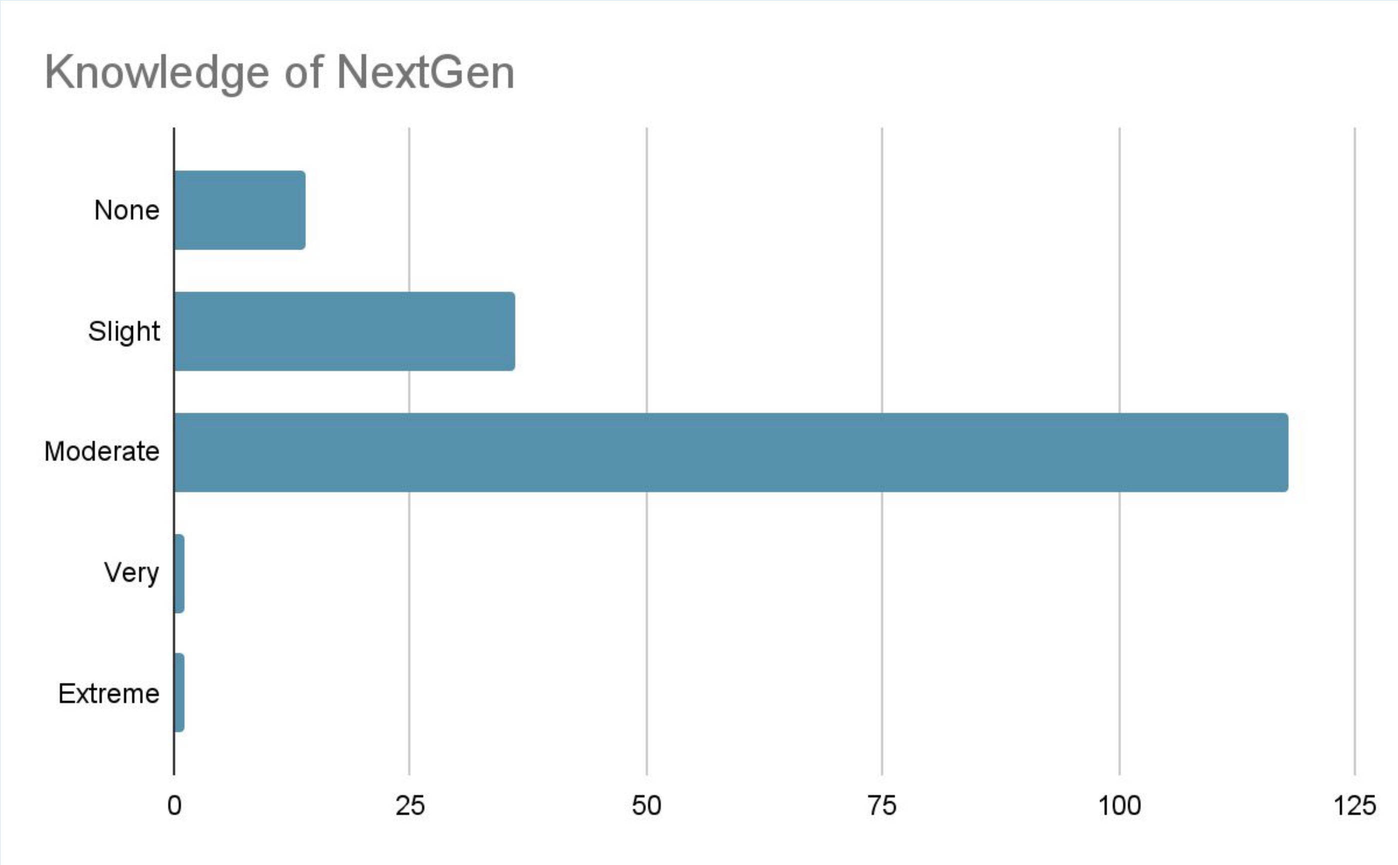
Training Needs

- CI/CD (7)
- SE (5)
- GitHub (5)
- DevOps (2)
- Project Standards (2)

CIROH Survey: Data Analysis/Visualization



CIROH Survey: Water Model



CIROH Survey: Water Model

Model	Use	Use in NextGen	Contributing
Not Applicable	26	43	36
SWAT	9	2	
HEC-HMS	7		
NWM	6	4	3
SUMMA	5		3
RHESSYS	4		
HEC-RAS	3		
WRF-HYDRO	3		2
GSSHA	2		
Noah-MP	2		

Others Used in NextGen

- CFE (2)
- HAND (2)
- SWEML (1)
- Neural Hydrology (1)

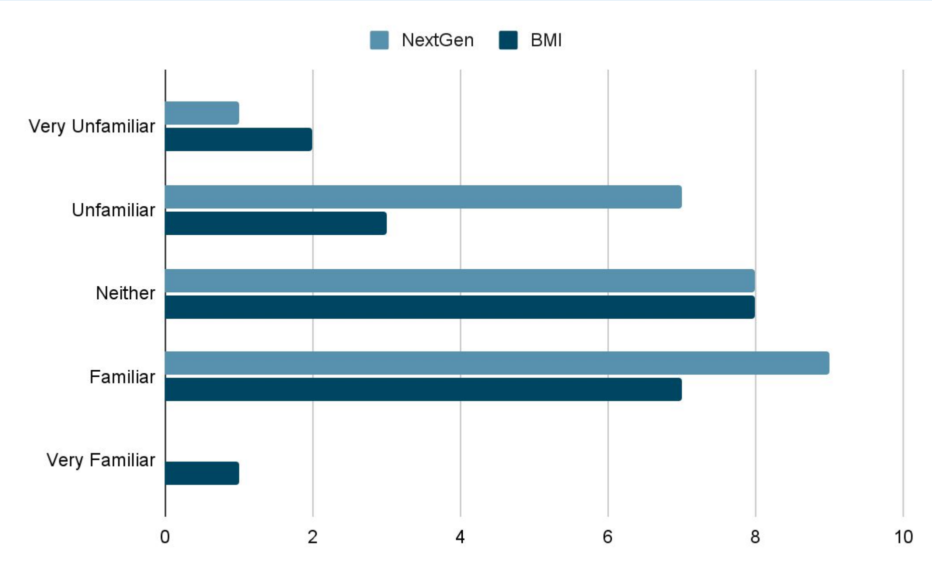
Others Contributing To

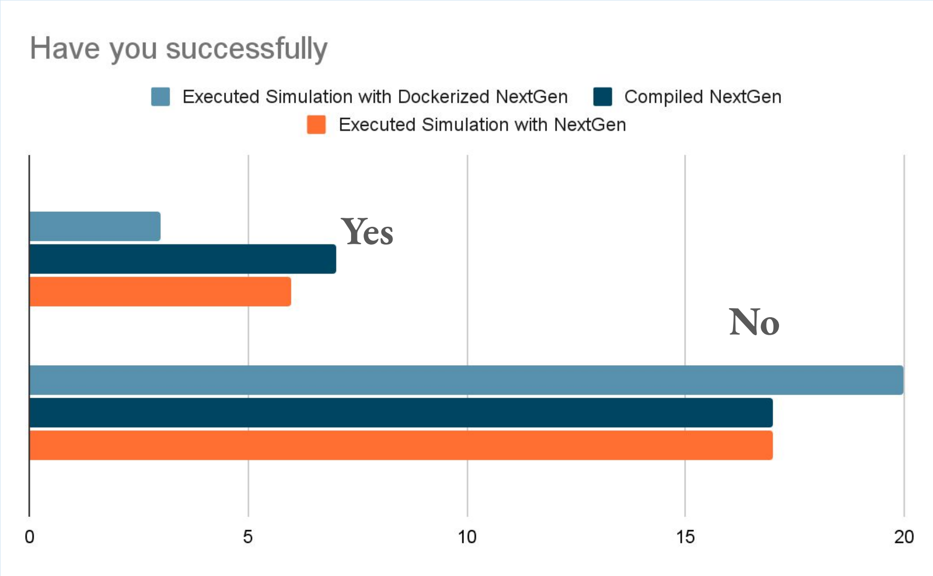
- HydroBID (1)
- Muskingum Cunge (1)

Biggest Training Needs

- Development/Integration with NextGen (11)
- BMI (4)

CIROH Survey: Water Resources Modeling





CIROH CyberInfrastructure

Presenter: Arpita Patel
DevOps Manager and Enterprise Architect
University of Alabama

CIROH CyberInfrastructure

"**CyberInfrastructure**" = IT for data-intensive research

At CIROH, we provide flexible computational platform to support collaborative hydrology research surrounding the NextGen framework.

CIROH R2OHC GOAL (CIROH Research to Operations Hybrid Cloud)

01

Promote **reproducible hydrologic computing experiments** with the NextGen Water Resource Modeling Framework;

02

Provide support for, and reduce the barrier to entry for performing NextGen-related **experiments at** various **scales**; and

03

Accelerate the interconnection and **integration of research products** and hydroinformatics innovations from the various ongoing CIROH experiments.

CyberInfrastructure Team @ CIROH

Dr. Steve Burian

Dr. Purushotham Bangalore

Dr. Jeffrey Carver

James Halgren

Arpita Patel

Benjamin Lee

Sepehr Karimi

Trupesh Patel

Manjila Singh

CIROH DocuHub: Community editable technical documentation hub!

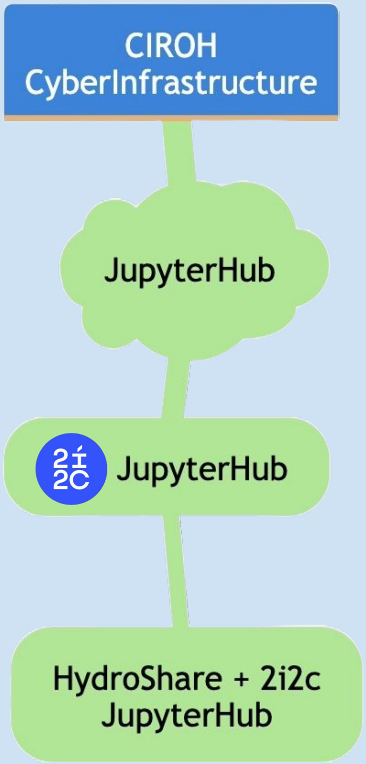
- CIROH's DocuHub is a open-source central repository for CIROH's projects, services **technical documentation**.

- **DocuHub Features:**
  - Products Documentation
  - Services Documentation
  - Training and Education
  - CIROH News
  - Blog

docs.ciroh.org

2i2c JupyterHub @ CIROH

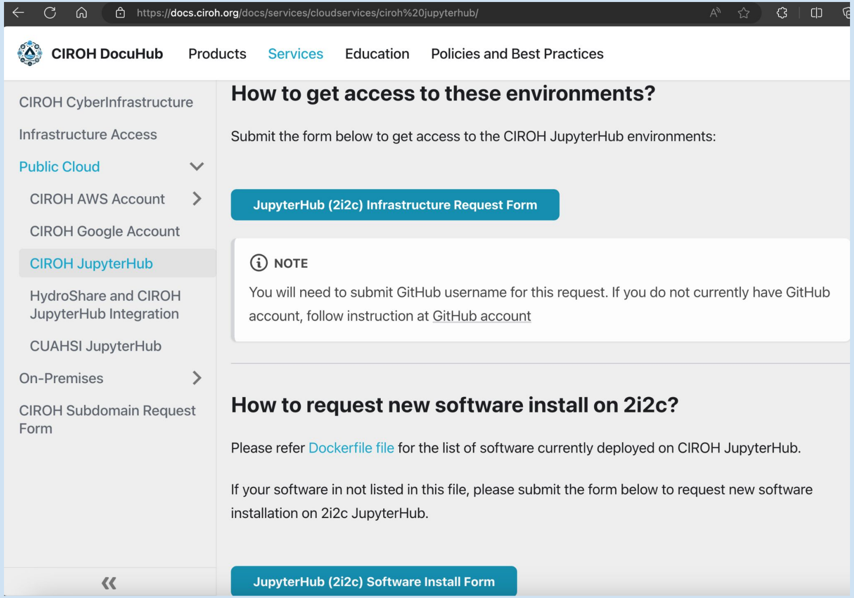
- CIROH provides dedicated JupyterHub Environment on **Google Cloud** via 2i2c.
- Runs the **pangeo-notebook** docker image.
- Ability to create custom CIROH images.
- Provides **scalable dask gateway** cluster that allows users to leverage scalable computing to analyze data more quickly.

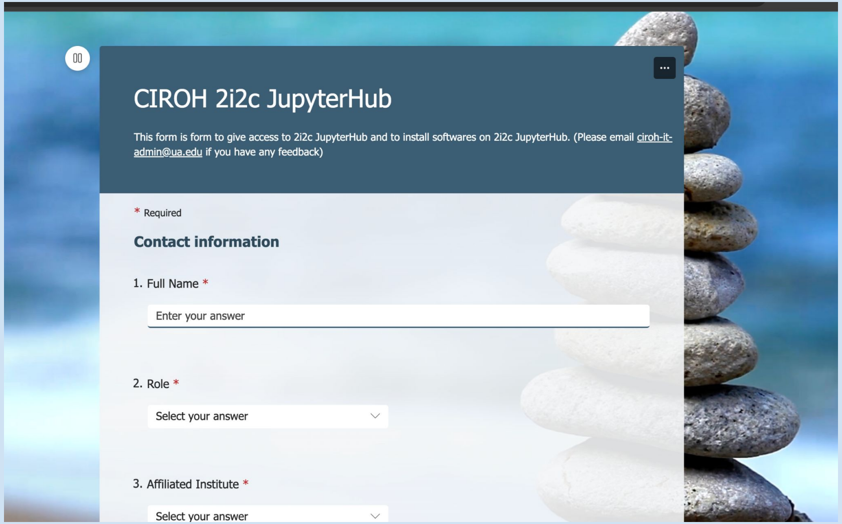


-

2i2c JupyterHub Access

- Access is handled by CIROH DevOps Staff.
- Submit form to get access
- Provide GitHub username





2i2c JupyterHub Overview

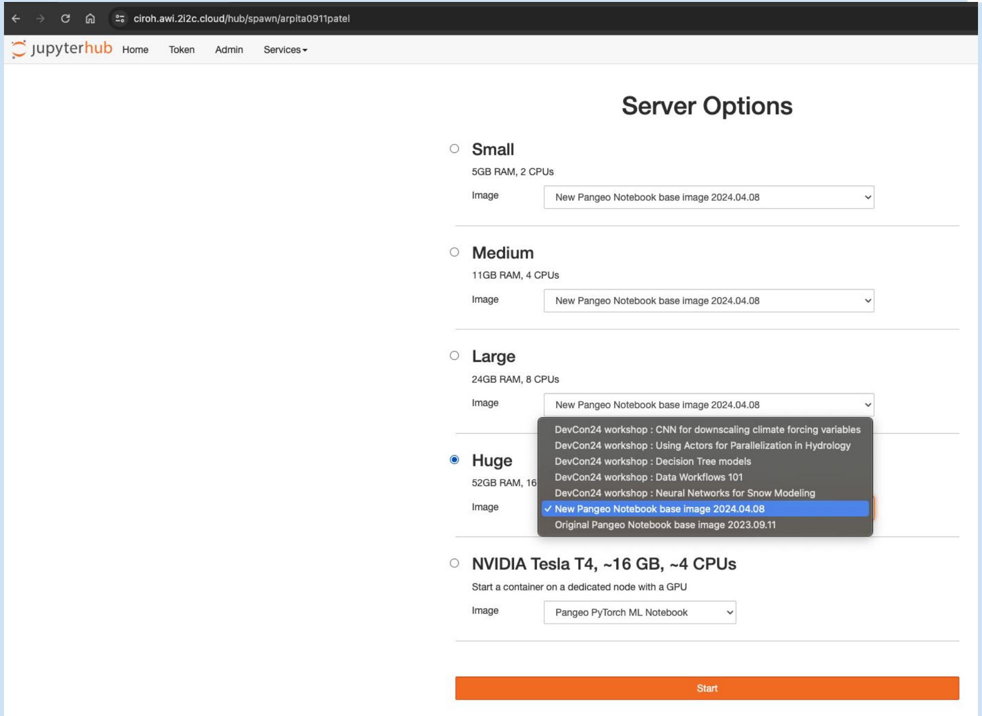
<u>Production Environment:</u>

https://ciroh.awi.2i2c.cloud/

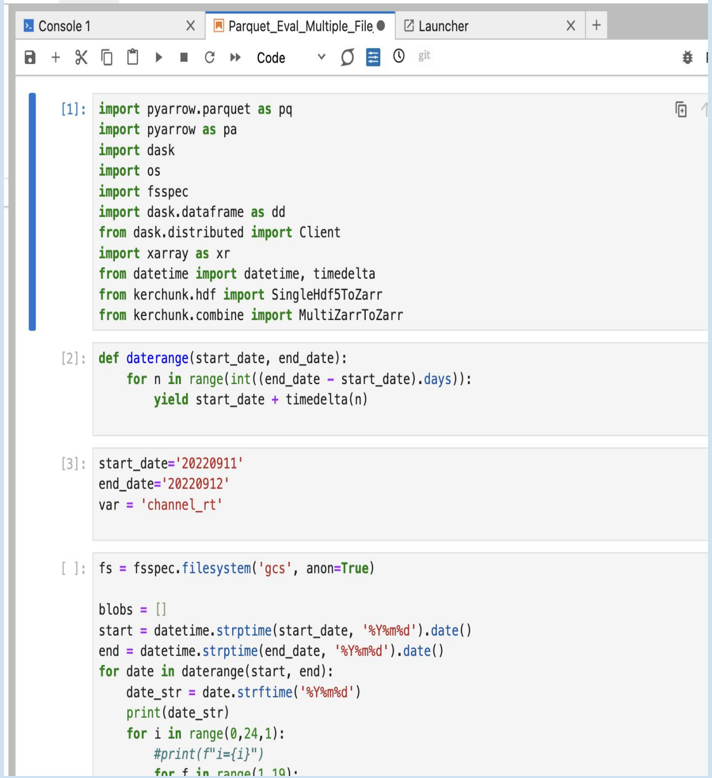
<u>Staging Environment:</u>

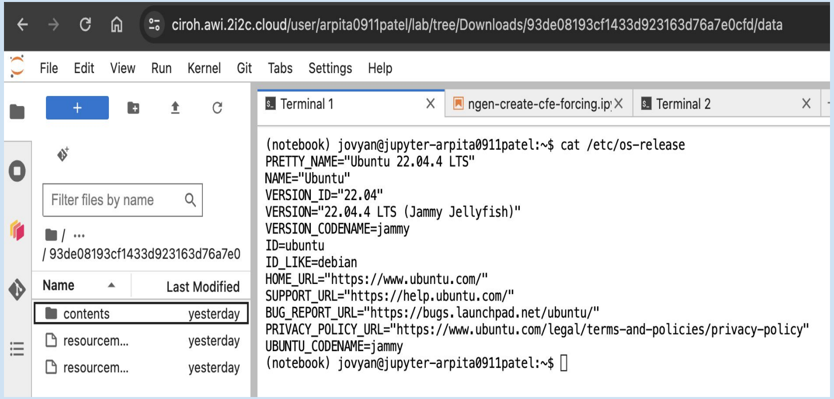
https://staging.ciroh.awi.2i2c.cloud/

- Provides different compute options
- Login using GitHub username
- Provided access to approx. 200+ researchers.
- More than 10 DevCon24 workshops to use 2i2c JupyterHub.
- Created custom image for the workshops.



2i2c JupyterHub Interface





2i2c JupyterHub + HydroShare

https://docs.ciroh.org/docs/services/cloudservices/HydroShare/

HYDROSHARE HOME MY RESOURCES DISCOVER COLLABORATE APPS HELP

CIROH Production JupyterHub

Authors: Anthony Castronova

Owners: Anthony M. Castronova | Arpita Patel

Type: App Connector

Storage: The size of this app connector is 0 bytes

Created: Oct 06, 2023 at 4:09 p.m.

Last updated: Apr 03, 2024 at 3:54 p.m.

Citation: See how to cite this resource

Abstract

CIROH 2i2c JupyterHub - Production Server

CIROH 2i2c Production environment - https://ciroh.awi.2i2c.cloud/

To gain access to this environment please reach out to ciroh-it-admin@ua.edu

https://docs.ciroh.org/docs/services/cloudservices/HydroShare/

CIROH DocuHub Products Services Education Policies and Best Practices

- CIROH CyberInfrastructure
- Infrastructure Access
- Public Cloud
  - CIROH AWS Account
  - CIROH Google Account
  - CIROH JupyterHub
  - HydroShare and CIROH JupyterHub Integration
  - CUAHSI JupyterHub
- On-Premises
- CIROH Subdomain Request Form

CIROH JupyterHub via HydroShare

HydroShare is a repository, website, and hydrologic information system for sharing hydrologic data and models aimed at giving users the cyberinfrastructure needed to innovate and collaborate in research to solve water problems.

INFO

HydroShare website

HydroShare and CIROH JupyterHub Integration

Users now have the capability to directly launch and execute computational notebooks from HydroShare resources into the CIROH Jupyterhub environments. Wondering how to do it?

Here are the steps for you to follow:

HYDROSHARE

AWS @ CIROH: Unlocking Water Research Potential

CIROH leverages the power and flexibility of Amazon Web Services (AWS) to advance critical water research initiatives.

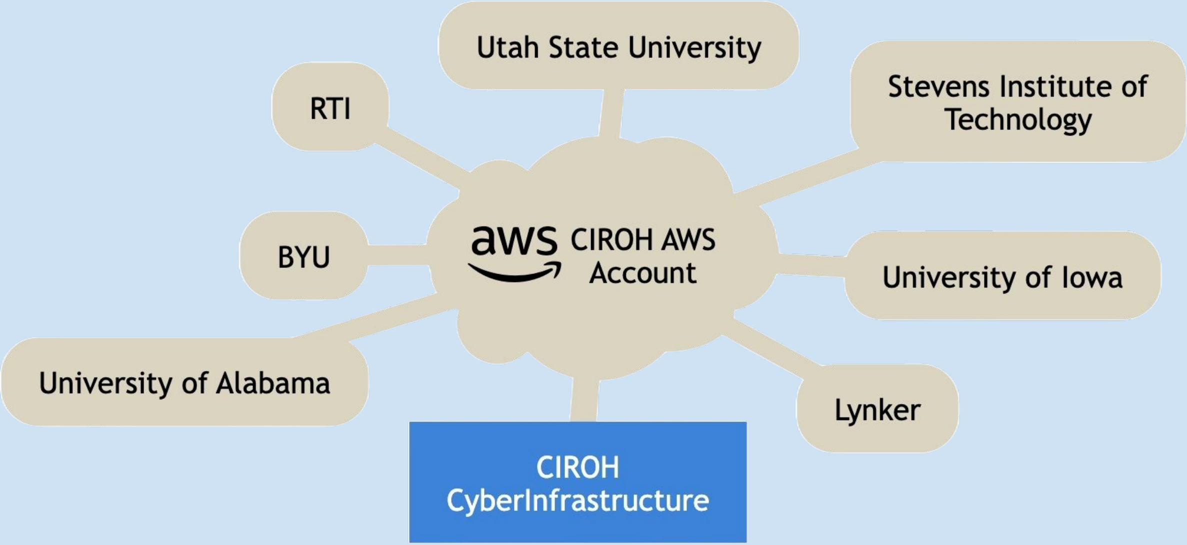
<u>**Goal:**</u>

- Bringing **compute closer to data** for faster analysis.
- Facilitating **collaborative research** through seamless data sharing.

Here's how AWS supports CIROH's mission:

- **Data Management Analysis:**
  - **Storing vast datasets:** We utilize **Amazon S3** to store massive volumes.
  - **Processing & analysis:** Amazon **EC2** provide the computational power.
- **Collaboration:**
  - **Sharing research data: Amazon S3** and open data platforms facilitate easy sharing of research data.
- **Scalability & Innovation:**
  - **Adapting to changing needs:** The **scalability** of AWS allows to quickly adjust resources to accommodate new projects and growing data volumes.
  - **Exploring cutting-edge technologies:** We leverage AWS machine learning services, like **Amazon SageMaker.**

Current CIROH AWS Users

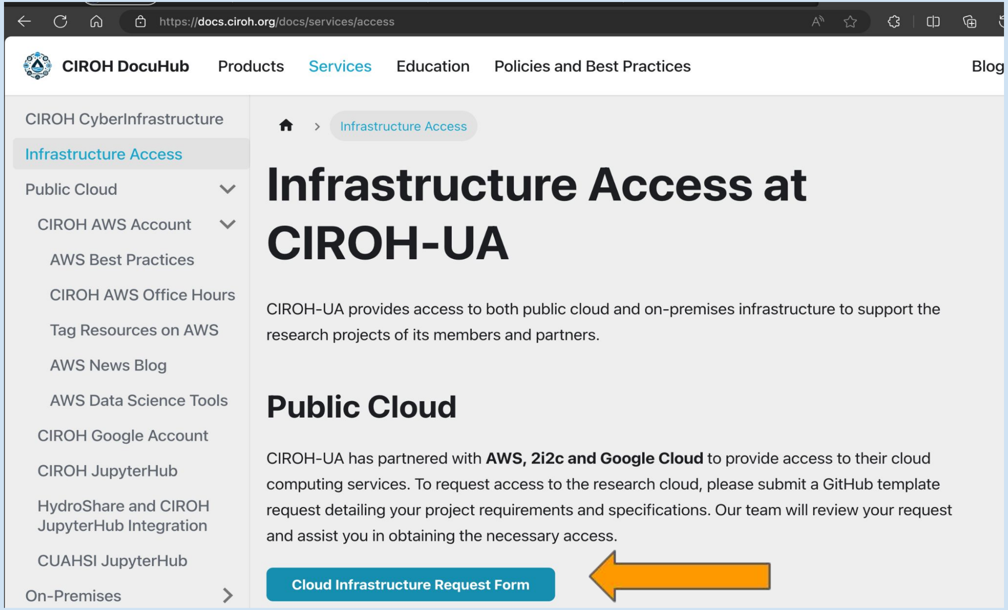


Key AWS Services utilized by CIROH:

- Control Tower
- IAM
- EC2 (Elastic Compute Cloud)
- S3 (Simple Storage Service)
- EFS (Elastic File Store)
- EKS (Elastic Kubernetes Services)
- Lambda
- SageMaker

CIROH AWS Access

How to get **Access** to AWS Account for CIROH projects?



Monthly AWS Office Hours

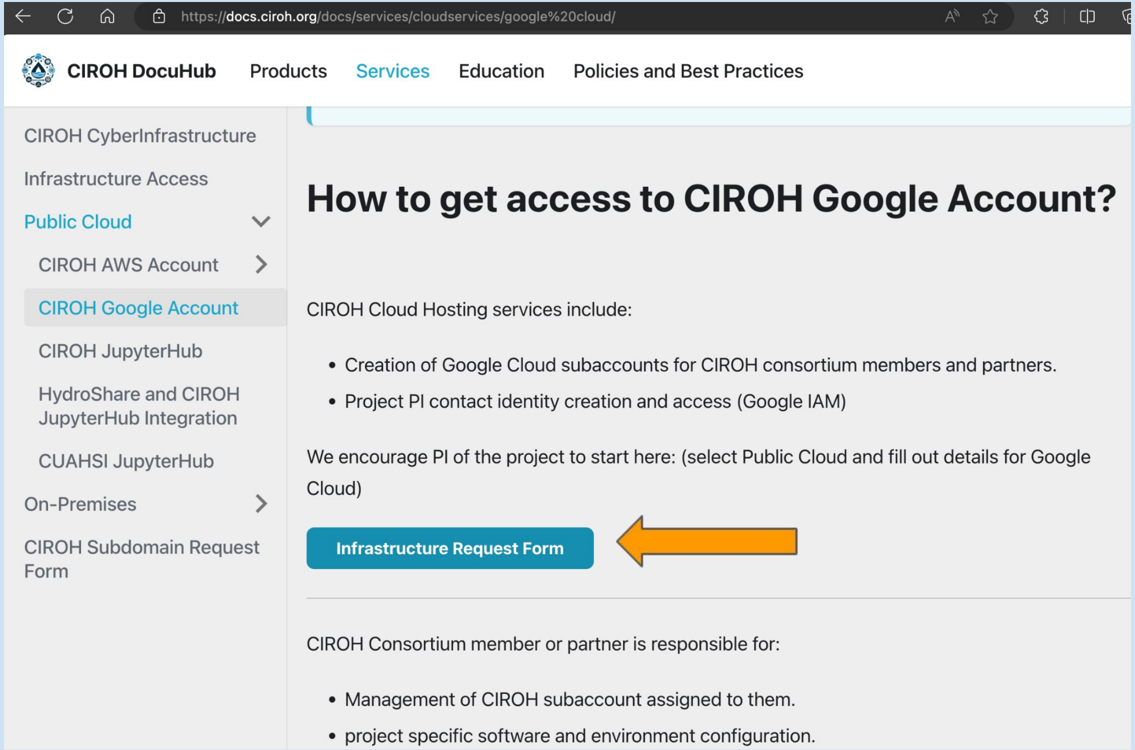
Submit your request or questions at:

CIROH Slack Channel **#aws-ciroh-support** or email us at **ciroh-it-support@ua.edu**

Google Cloud @ CIROH

AWS Google Cloud Access

How to get Access to Google Cloud services for CIROH projects?



https://docs.ciroh.org/docs/services/cloudservices/google%20cloud/

CIROH DocuHub Products Services Education Policies and Best Practices

- CIROH CyberInfrastructure
- Infrastructure Access
- Public Cloud
  - CIROH AWS Account
  - CIROH Google Account
  - CIROH JupyterHub
  - HydroShare and CIROH JupyterHub Integration
  - CUAHSI JupyterHub
- On-Premises
- CIROH Subdomain Request Form

How to get access to CIROH Google Account?

CIROH Cloud Hosting services include:

- Creation of Google Cloud subaccounts for CIROH consortium members and partners.
- Project PI contact identity creation and access (Google IAM)

We encourage PI of the project to start here: (select Public Cloud and fill out details for Google Cloud)

Infrastructure Request Form

CIROH Consortium member or partner is responsible for:

- Management of CIROH subaccount assigned to them.
- project specific software and environment configuration.

On-premise HPC Cluster @ CIROH "Pantarhei"

Please scan the QR code to visit CIROH's DocuHub for more information.

The System Includes

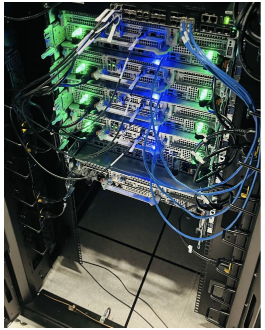
- **6 CPU** nodes (***40 Core*** each, ***1***x Intel Arria 10 GX FPGA)
- **1 GPU** node (***4***x **V100**)
- **384 GB** DDR4-2666 RAM (each)
- **4 TB** of NVMe memory
- **76 TB** RAID Storage system
- **Rocky Linux 9.2** Operating System
- **Slurm Workload Manager** for job scheduling.
- **Singularity** container platform
- The **environment modules system** provides access to diverse software modules (e.g., Python, MPI, CUDA, netCDF, BLAS, Visual Studio Code, etc.)
- Accessible through **remote connections** for convenient usage.

Benefits

- **Expedites research** with computationally intensive tasks.
- Enhances **reproducibility** through **containerization** and software environments.
- Provides **scalable**, **flexible** solutions for diverse research needs.

Current Users:

- University of Alabama
- Pennsylvania State University
- Michigan State University
- University of Saskatchewan
- University of Calgary
- Lynker



On-premise AI/ML Cluster @ CIROH “Wukong”

The System Includes

- **2 Intel Xeon Platinum** 8470 **processor**
- **104 Cores** (2 Thread(s) per core)
- **1 TB** DDR5-4800 RAM
- **8 NVIDIA HGX A100 80GB** 500W SXM4 **GPUs**, fully interconnected with NVLink
- **56 TB** RAID NVMe storage system
- **Rocky Linux 9.2** Operating System
- **Singularity** container platform
- The **environment modules system** provides access to diverse software modules (e.g., Python, MPI, CUDA, netCDF, BLAS, Visual Studio Code, etc.)
- It is **optimized for AI/ML workloads** and comes with pre-installed **deep learning frameworks** and libraries (e.g., TensorFlow, PyTorch, etc.)
- Accessible through **remote connections** for convenient usage.

Please scan the **QR code** to visit CIROH’s DocuHub for more information.

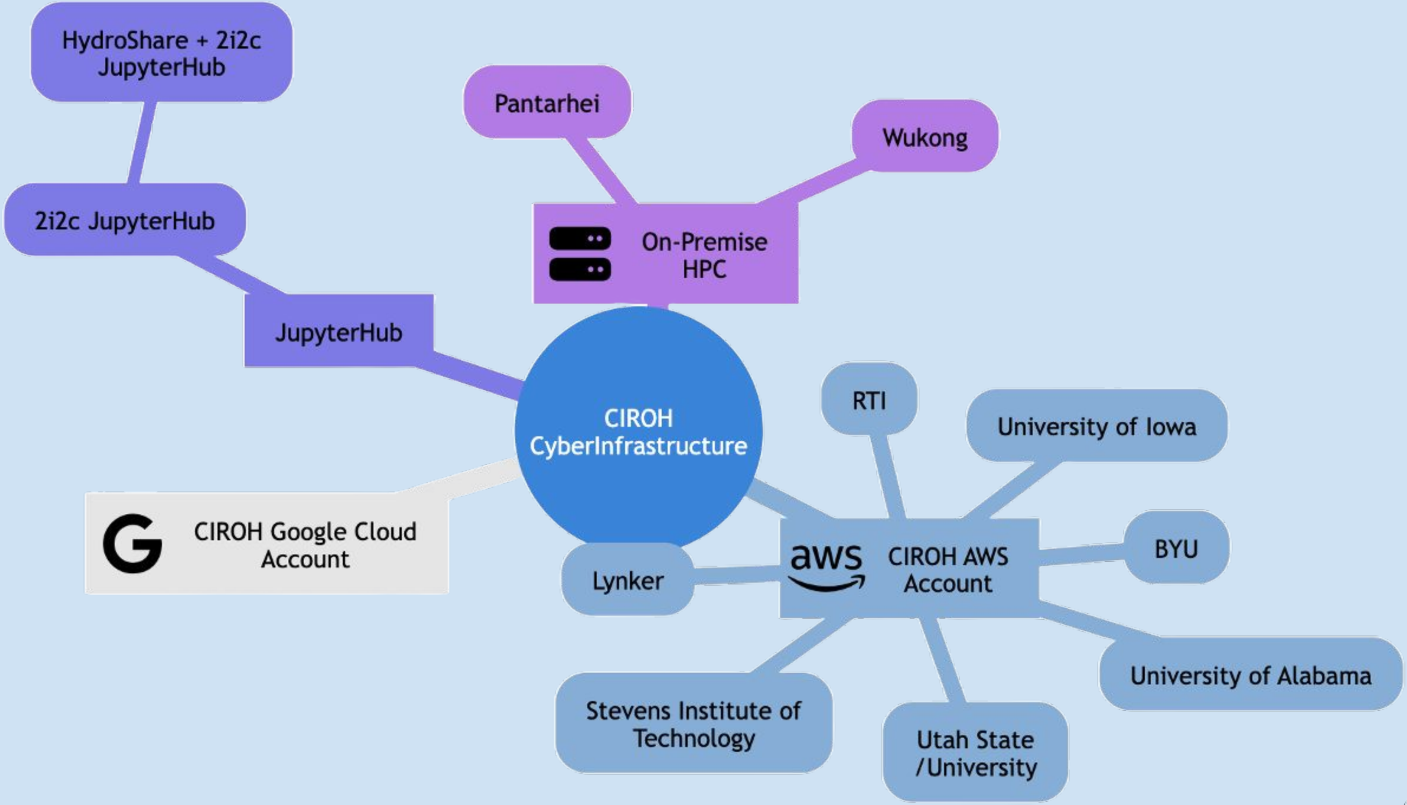
Benefits

- Enables researchers to **develop** and **deploy AI/ML models.**
- Accelerates **model training** and **inference times,** boosting research productivity
- Enhances **reproducibility** through **containerization** and software environments.
- Provides **scalable**, **flexible** solutions for diverse research needs.

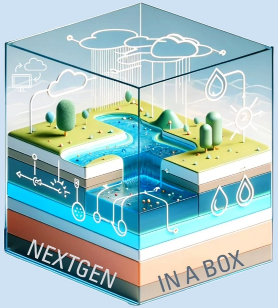
Current Users:

- Pennsylvania State University
- University of Alabama

CIROH Services



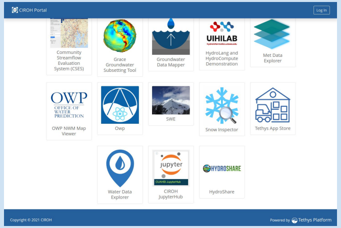
CIROH Products: Advancing Water Science Research and Applications



CIROH develops and maintains a diverse range of software products that empower researchers, practitioners, and stakeholders to address complex water-related challenges.

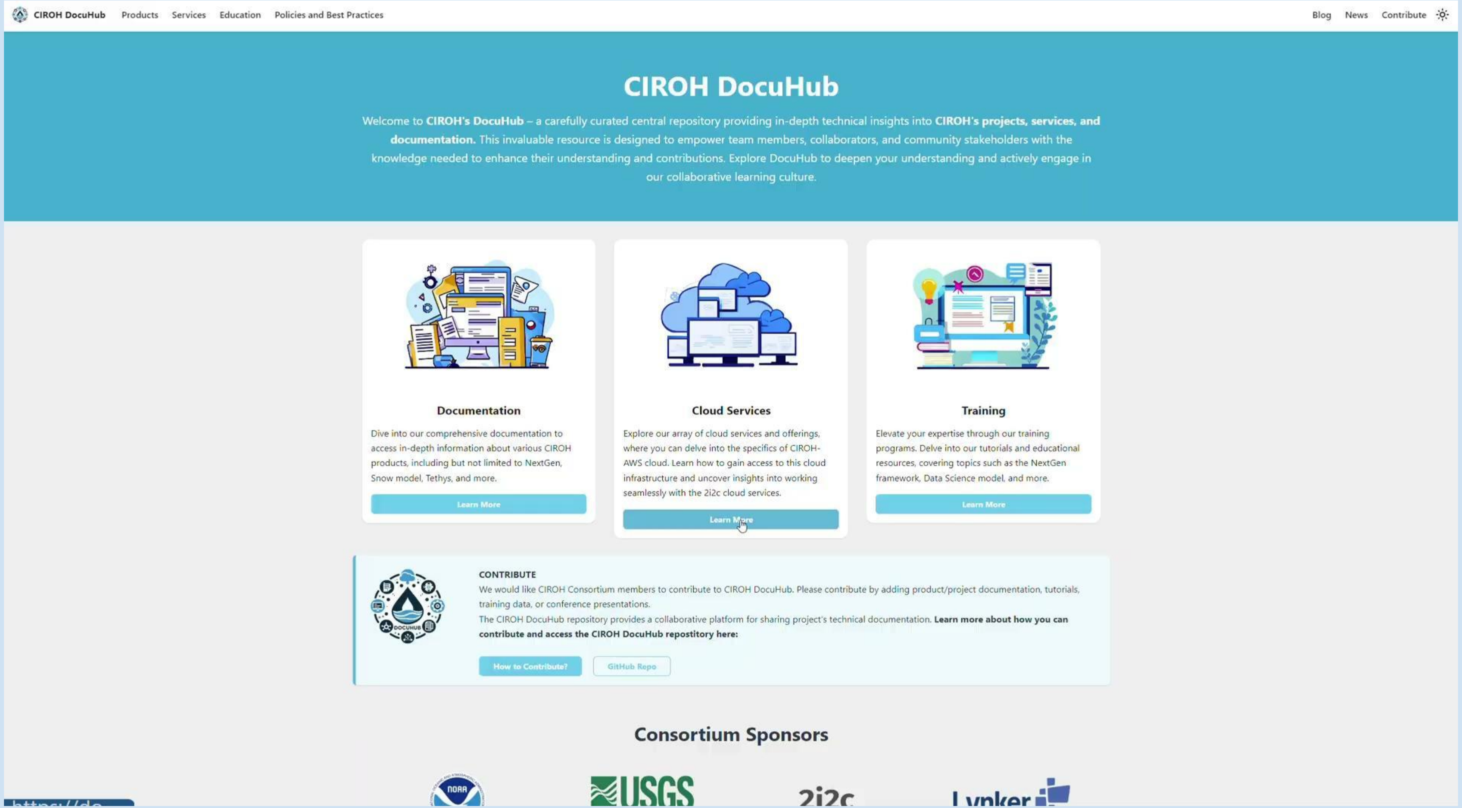
Product Categories:

- **Hydrologic Modeling Frameworks:**
  - **NGIAB (NextGen In A Box):** A user-friendly platform that simplifies the setup and execution of NextGen models.
  - **Ngen-dataStream**
- **Data Management and Access Tools:**
  - **HydroShare:** A collaborative platform for sharing, discovering, and publishing hydrologic data and models.
  - **AWS S3**
- **Visualization and Analysis Tools:**
  - **Tethys Platform:** A web-based platform for building and deploying interactive water resources applications.
- **Evaluation Tools:**
  - **TEEHR**: A set of tools for hydrologic model/forecast evaluation that are scalable and flexible and enable highly exploratory evaluation and foster open community development.
  - **Community Streamflow Evaluation System (CSES)**: A python-based tool, user friendly, fast and model agnostic streamflow evaluation tool.



More products: https://docs.ciroh.org/docs/products/intro

CIROH DocuHub



CIROH DocuHub - Promote your CIROH research

How and What to Contribute to CIROH DocuHub?

https://docs.ciroh.org/contribute

CIROH project/products PIs/team members : Share your open-source GitHub URL to add to products page via email at ciroh-it-admin@ua.edu

Thank You!

Case Study: Tethys Platform and Google BigQuery - (BYU)

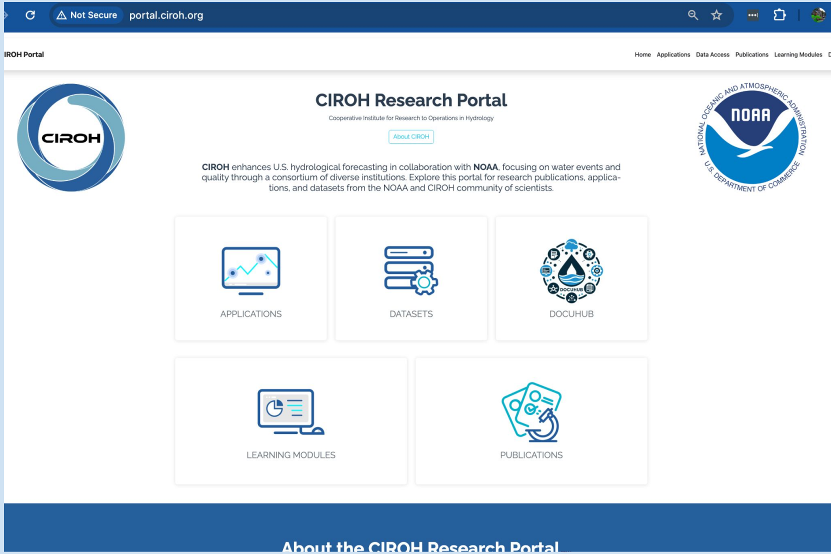
Presenter: Dan Ames
Professor at Brigham Young University

Project Overview

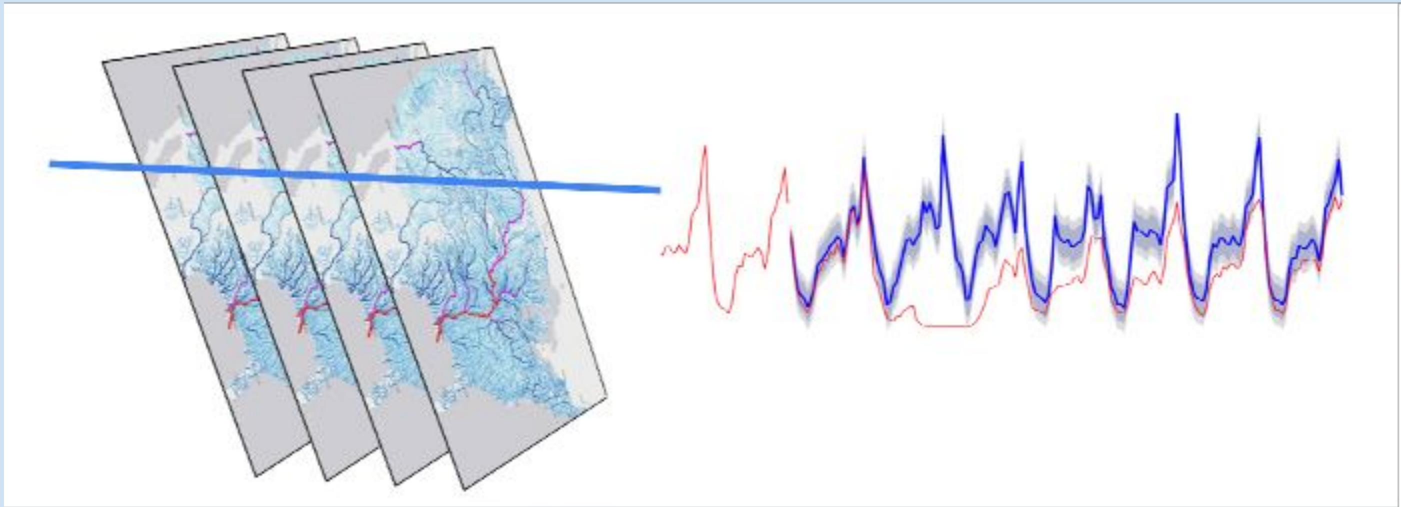
Phase one project: Design and Implementation of a Big Query Dataset and Application Programmer Interface (API) for the U.S. National Water Model

Phase two project: CIROH Research Portal: A one-stop shop portal for NOAA CIROH Data, Research, and Applications

portal.ciroh.org



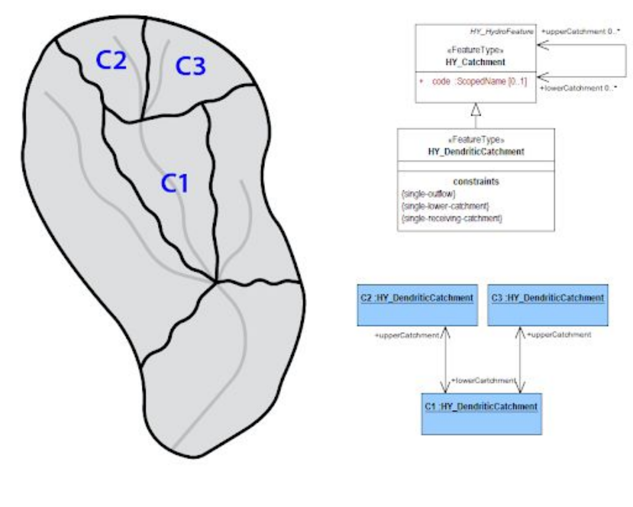
Contribution to CIROH

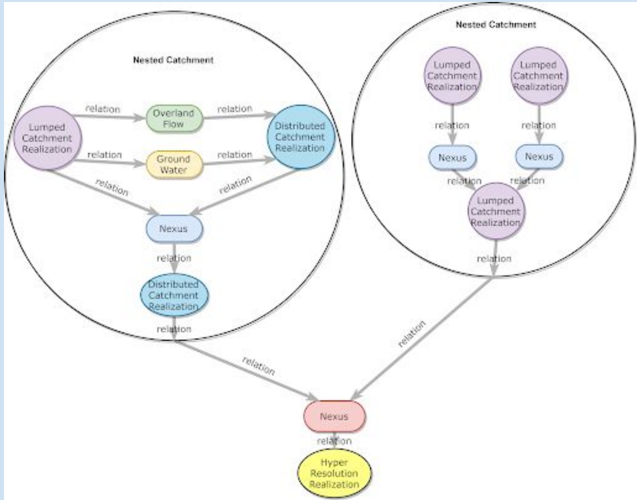


The CIROH BigQuery API will be deployed on CIROH Cloud, supporting research workflows by CIROH members

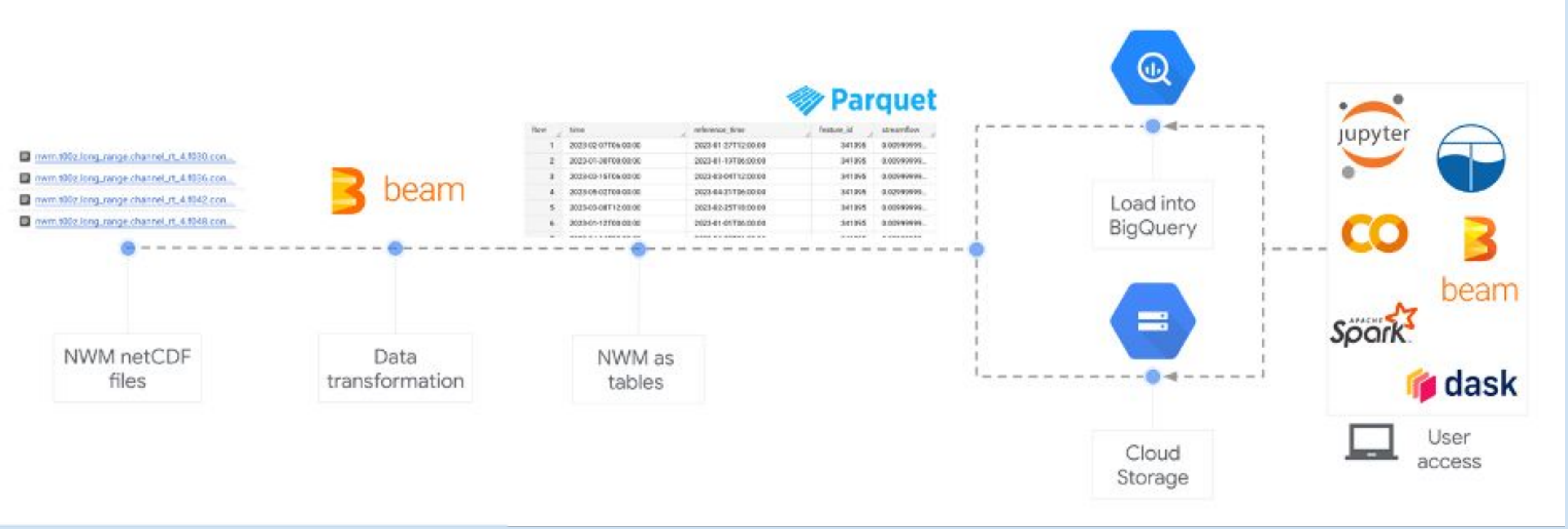
Relevance to NextGen

Ready access to current and historical forecasts will enable and simplify research including comparison of model results with observations, assessment of NWM 3.0 as well as NextGen based model, and localization and regional testing and improving of the NWM.

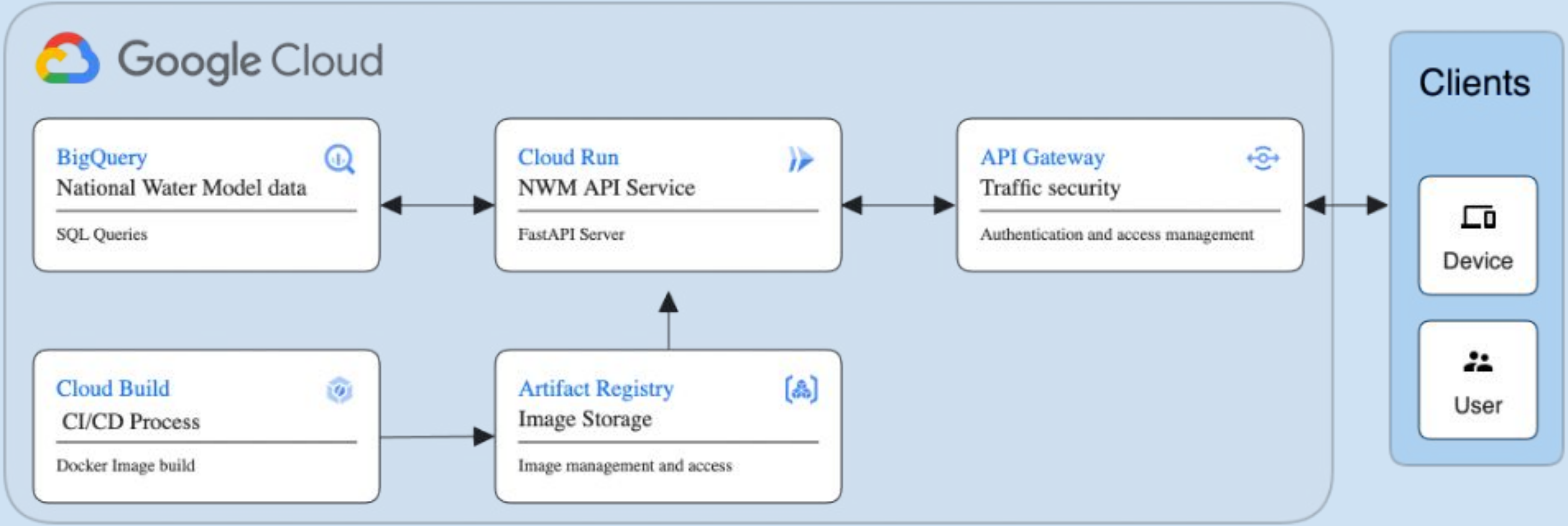




Infrastructure Utilized



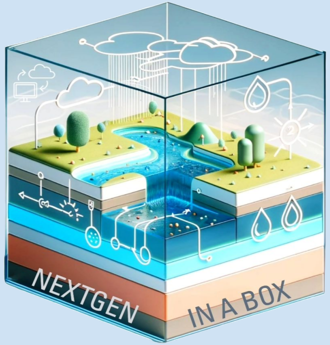
Additional Infrastructure requirements

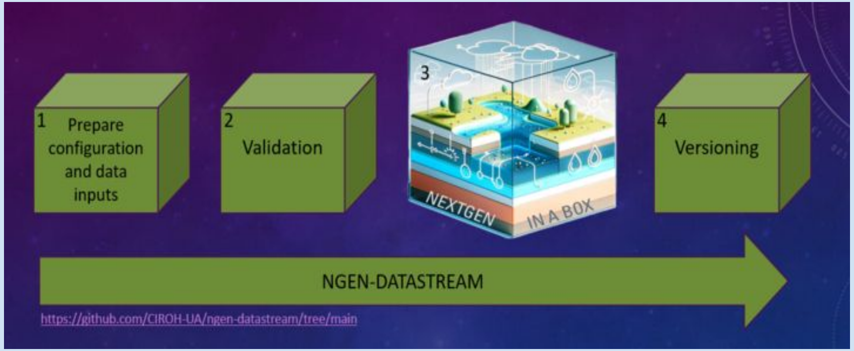


Case Study: NextGen In A Box and ngen-dataStream (UA, Lynker, BYU, Aquaveo)

Presenter: Zach Wills

NextGen In A Box and ngen-datastream






Nels Frazier


Zach Wills


Jordan Laser


James Halgren


Arpita Patel


Shahab Alam


Benjamin Lee


Hari Teja Jajula


Trupesh Patel


Josh Cunningham


Chad Perry


Manjila Singh

Contribution to CIROH

Turning the ideal workflow, into reliable community software.

<u>Massive Barriers to Community Contribution:</u>

- Software Compilation Expertise
- No public test data
- No local install documentation
- Software Licensing Costs
- Inefficient scaling to larger domains
- Incompatible input/output data
- Unreproducible outputs or results

All of this severely limits the community's ability to cooperate and collaborate. It forsters mistrust and a high barrier to entry.

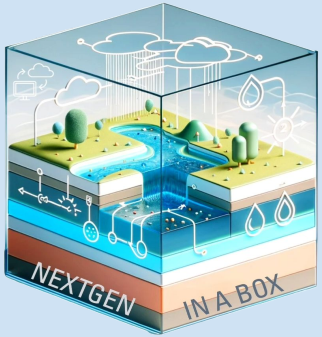
<u>Community Benefits of NextGen In a Box:</u>

- Pre-compiled, and tested, in public
- Multiple public test data sets
- Open Local and Cloud deployment code
- No Cost
- Scales from backyard to CONUS
- Data Standard allows for interoperation
- Fully reproducible, we even publish hashes!

This is all for the community, right now, for free.

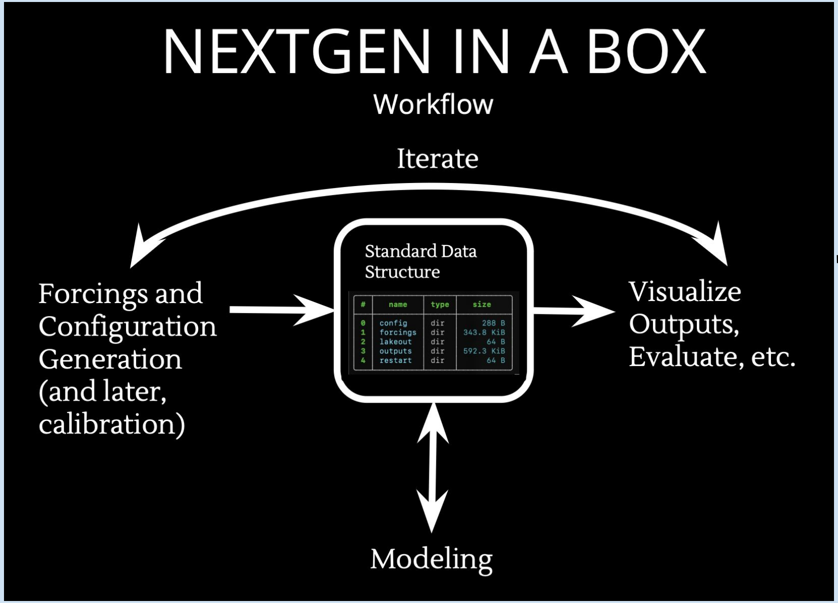
Relevance to NextGen

Thanks Josh S. for the logo!

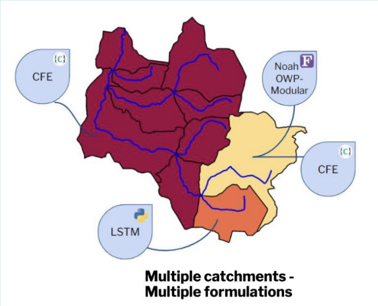


Go from zero to running 8 NextGen BMI-Compatible models, locally or in the cloud, in under an hour.

No HPC or Software Licenses necessary.




```
bash-5.1# ../dmod/bin/ngen-parallel --info
NGen version: 0.1.0
Build configuration summary:
  Generator: Unix Makefiles
  Build type:
  System: Linux
  C Compiler: /usr/bin/cc
  C Flags:
  CXX Compiler: /usr/bin/c++
  CXX Flags:
  Flags:
    NGEN_WITH_MPI: ON
    NGEN_WITH_NETCDF: ON
    NGEN_WITH_SQLITE: ON
    NGEN_WITH_UDUNITS: ON
    NGEN_WITH_BMI_FORTRAN: ON
    NGEN_WITH_BMI_C: ON
    NGEN_WITH_PYTHON: ON
    NGEN_WITH_ROUTING: ON
    NGEN_WITH_TESTS: ON
    NGEN_WITH_COVERAGE: OFF
    NGEN_QUIET: OFF
  Extern Models:
    NGEN_WITH_EXTERN_ALL: ON
    NGEN_WITH_EXTERN_SLOTH: ON
    NGEN_WITH_EXTERN_TOPMODEL: ON
    NGEN_WITH_EXTERN_CFE: ON
    NGEN_WITH_EXTERN_PET: ON
    NGEN_WITH_EXTERN_NOAH_OWP_MODULAR: ON
Environment summary:
  Boost:
    Version: 1.79.0
    Include: /ngen/boost
  MPI (C):
    Version: 3.1
    Library: /usr/local/lib/libmpi.so
    Include: /usr/local/include
  MPI (CXX):
    Version: 3.1
    Library: /usr/local/lib/libmpicxx.so,/usr/local/lib/libmpi.so
    Include: /usr/local/include
  NetCDF:
    Version: 4.8.1
    Library: /usr/lib/libnetcdf.so
    Library (CXX): /usr/local/lib64/libnetcdf-cxx4.so
    Include: /usr/include
    Include (CXX): /usr/local/include
    Parallel: TRUE
  SQLite:
    Version: 3.34.1
    Library: /usr/lib64/libsqlite3.so
    Include: /usr/include
  UDUNITS2:
    Library: /usr/lib64/libudunits2.so
    Include: /usr/include/udunits2
  Fortran:
    BMI_FORTRAN_ISO_C_LIB_PATH:
    BMI_FORTRAN_ISO_C_LIB_NAME: OFF
    BMI_FORTRAN_ISO_C_LIB_DIR: OFF
  Python:
    Version: 3.9.18
    Virtual Env: <none>
```



All NextGen
No Compromises

Infrastructure Utilized

CIROH Community Resources

From Container, to CONUS.

Amazon S3

AWI-CIROH AWS
Using Datastream

Local Development

#	name	type	size
0	config	dir	288 B
1	forcings	dir	343.8 KiB
2	lakeout	dir	64 B
3	outputs	dir	592.3 KiB
4	restart	dir	64 B

Local Tests and Runs with docker

docker

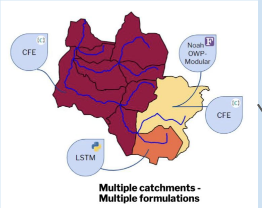
Scale to CONUS on EC2 using your exact configuration in the Datastream

Cost and Time efficient
No HPC Queue
ARM & x86 Compatible

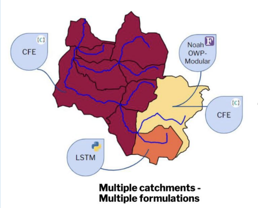
Hashed outputs for full reproducibility

Additional Infrastructure requirements

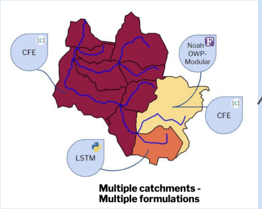
CIROH Project #1



CIROH Project #2



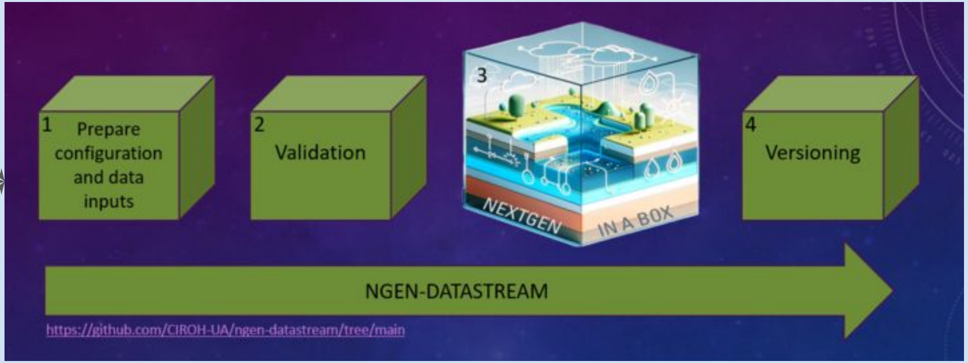
CIROH Project #3



Integration of Improvements

<u>Community Datastream</u>

“The Daily State of the Science”



Future Compute:
GPU?
FPGA?
Unified Memory?

<u>All of the Above!</u>

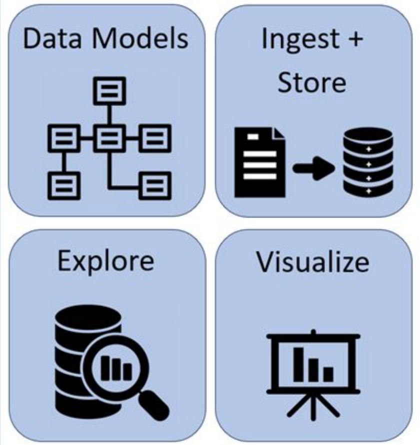
Case Study: Integrated Evaluation System Prototype for Testing Research and Operational Advancements (RTI)

Presenter: Matt Denno
Senior Environmental Applications Developer
RTI

Project Overview

Integrated Evaluation System Prototype for Testing Research and Operational Advancements

Create a set of tools (and system) for hydrologic model and forecast evaluation that are **scalable and flexible** for the wide range of use cases and users, that enable **highly exploratory evaluation**, and that foster open community development



Katie van Werkhoven (PI)
Matt Denno (co-PI, Development Lead)
Sam Lamont (Developer)

Contribution to CIROH

- **TEEHR-HUB** – A **JupyterHub** environment running the **TEEHR** image + **EFS** and **S3** for data storage.
- Using CIROH CI as a platform for testing scalability and performing research.
- Using TEEHR-HUB in AWS:
  - Processed the AORC (v3.0 retrospective) gridded precipitation to the MERIT basins -> TEEHR data model
  - CONUS 40-yr Retrospective (v3.0 and USGS) -> TEEHR data model
  - OWP post flood-event forecast evaluations
- Testbed integration

Relevance to NextGen

- A standard way to evaluate research and innovations to allow comparison of different methods is an essential part of NextGen (R2O).
- A system capable of supporting large CONUS+ scale evaluations while allowing exploration of the data is essential to understanding and selecting innovations for NextGen.
- Using the CIROH Cyberinfrastructure for research and development is helping us to meet that goal.

Infrastructure Utilized

Elastic Kubernetes Service (EKS) (including supporting AWS services) – Scalable computing resources to host JupyterHub, Dask and Spark

Elastic File System (EFS) – Shared data drive for cached data and shared documents (notebooks, etc.)

Simple Storage Service (S3) – Bucket storage for large public and private datasets

Additional Infrastructure requirements

- We don't anticipate utilizing any new Cyberinfrastructure services. We will continue to use the ones we are – EKS, EFS and S3.
- Developing the evaluation system around Kubernetes (EKS) cluster architecture makes it deployable anywhere and extendable without additional cloud services (beyond storage and compute).
- Supporting larger and larger volumes of data (which is a good thing) will require more storage and compute.

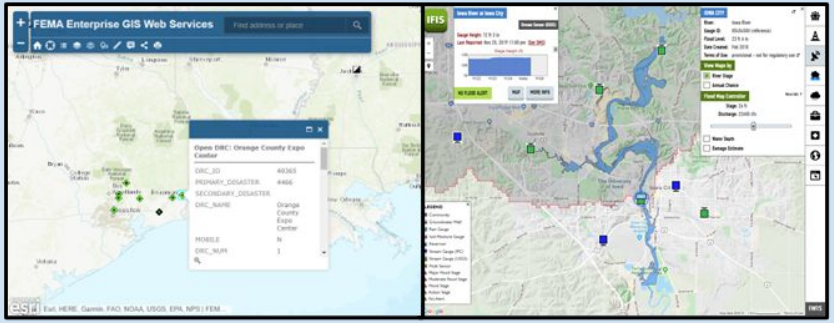
Case Study: Hydrology AI Framework (University of Iowa)

Yusuf Sermet
Asst. Research Engineer at UIOWA

Ibrahim Demir
Assoc. Professor at UIOWA

Presenter: Yusuf Sermet
University of Iowa

Project Overview



Challenges

- Complex Data Integration
- Predictive Modeling Limitations
- Accessibility and Inclusivity
- Collaborative Barriers

Solution

- Hydrology AI Framework
- Advanced AI Integration
- Democratization of Tools
- Community-Driven Development

Contribution to CIROH

Data Resources
- Hydrological & Geospatial Data
- Socioeconomic Data
- Academic Literature
- Sensor/IoT Data

Model Development
- LLMs
- MLLMs
- Mini Models
- Hydro AI Models

Cyber Infrastructure
- Interactive Web Portal
- Educational Workshops
- Exhibitions & Tours
- Webinars and Hackathons

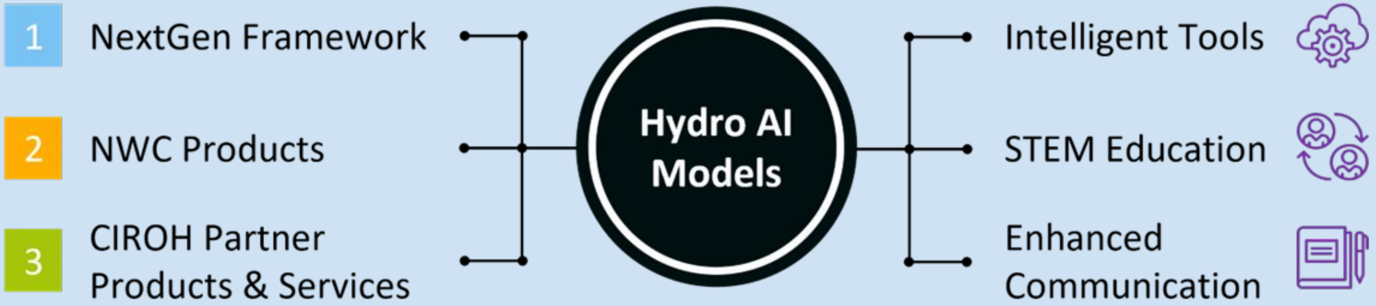
AI Hub Web Platform
- SaaS Application and UI/UX
- Model Playground
- Educational Resources
- Community Forum

Community Engagement
- Workshops and Webinars
- Open-Source Collaboration
- Hackathons and Challenges
- Academic Integration

Potential Use Cases and Application Areas
- AI Automated Reviewers
- Raster & Map Generation
- Textbook Generation
- Manuscript & Proposal Preparation
- Hydrological Process Visualizations
- Watershed Management Support
- Flood Prediction & Management
- Policy & Decision-Making
- Public Engagement & Education
- Infrastructure Planning & Dev.
- WQ Monitoring

Relevance to NextGen

- **Intelligent Tools**: Development of tools like FPM, Hydro Assistant, and Virtual Forecaster accessible to both forecasters and operational teams.
- **Enhanced Communication**: Utilizing the NextGen framework for improved public communication and collaboration across various sectors.
- **STEM Education Enhancement**: Integrating AI tools into STEM curricula to educate the next generation of hydrologists and environmental scientists.



Infrastructure Utilized

- **AWS Elastic Compute Cloud (EC2)**
  - Utilization of P3 and P4 instance variants tailored for intensive GPU needs.
  - Supports the training of complex, large-scale AI models with enhanced computational power.

- **AWS Simple Storage Service (S3)**
  - Employed for efficient data management and access for large datasets integral to hydrological modeling.

- **AWS SageMaker**
  - Leveraged for convenient model building, training, and deployment.

Additional Infrastructure

- **Access to CIROH Datasets and Services**
  - Requirement for CIROH datasets and associated services to be hosted and easily accessible via a well-documented API.
  - Ensures seamless integration and utilization of critical data resources within our AI models and applications.

- **Stronger Computational Infrastructure**
  - Need for enhanced computational infrastructure to handle extensive data analytics tasks and the operational demands of large-scale datasets and models.
  - Upgrades to include more powerful servers, increased storage capabilities, and advanced GPU resources to support intensive modeling and data processing tasks.

Case Study: Automated System for River Ice Monitoring using Remote Sensing and Deep Learning Techniques Across the United States (Stevens Institute of Technology)

Dr. Marouane Temimi

Mohamed Abdelkader

Moheb Henein

Presenter: Mohamed Abdelkader

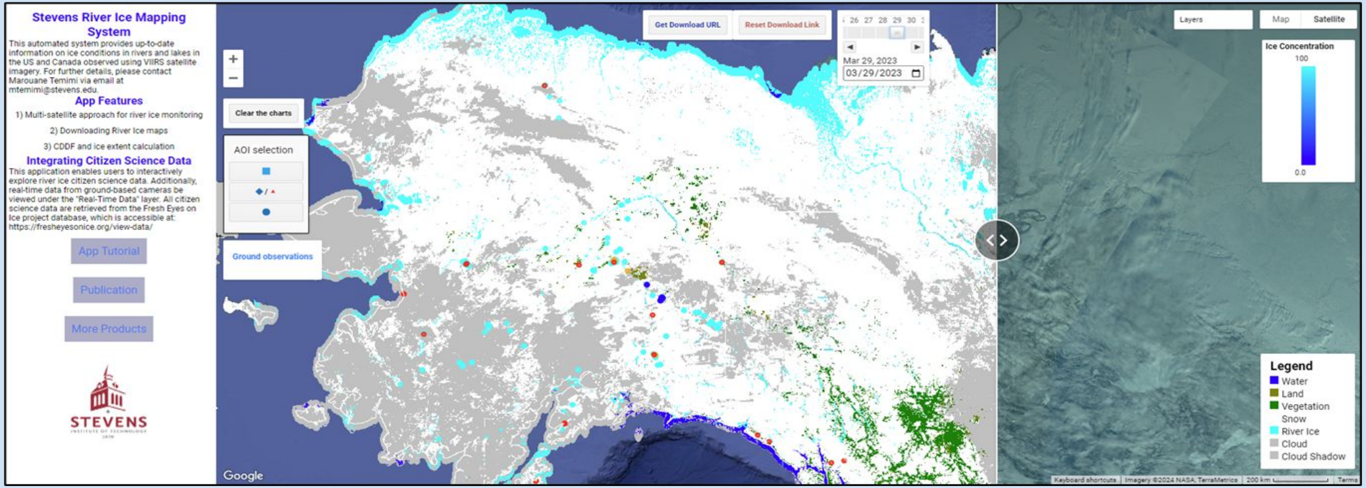
PhD Candidate

Project Overview

The goal of this project is to enhance the modeling and mapping of ice-induced flood inundation in northern watersheds in the Unites States.

The main objectives of the project are:

- Develop an enhanced operational system for river and lake ice mapping and monitoring across the United States.
- Develop a system to assess the risk of ice-induced flood inundation using NWM streamflow and other explanatory variables.
- Introduce the modeling of the thermodynamic ice formation in the NWM and its impact on streamflow and flood inundation.



Relevance to NextGen

Tailored Regional Formulations

- Testing region-specific models for ice-affected streams in northern watersheds.
- Generates detailed, region-specific ice condition data to enhance streamflow forecasts in cold regions.

Integration of Advanced Modeling Techniques

- Introducing thermodynamic ice formation modeling within the NWM.
- Producing remote sensing products designed for assimilation into NEXTGEN framework via the Basic Model Interface (BMI).

Facilitation of Model Evaluation

- Generates precise river ice maps for real-time and retrospective model evaluations.
- Providing tools and protocols for assessing ice-related flood risks.

Contribution to CIROH

1. Improvement of Water Resources Prediction Systems
   - ✓ Developing an enhanced automated system for river and lake ice mapping across the nation.

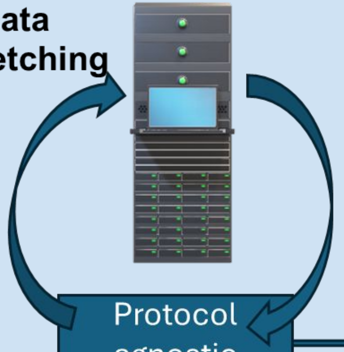
2. Advancement and Acceleration of Community Water Resources Modeling
   - ✓ Integrating artificial intelligence with conventional hydrological processes, accelerates the transition of new modeling techniques into operational practice

3. Advancement and Augmentation of Hydroinformatics
   - ✓ Developing advanced user interfaces for the automated river ice monitoring system.
   - ✓ Play a role in the CIROH Developers Conference Hydroinformatics trainings.

Infrastructure Utilized

Identity and Access Management (IAM)

Data fetching



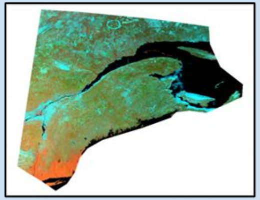
Protocol agnostic

Amazon CloudWatch

.yml

Lambda Function

Data projection



Data preprocessing

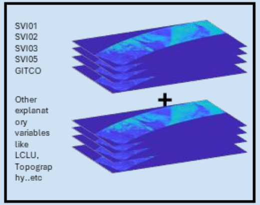
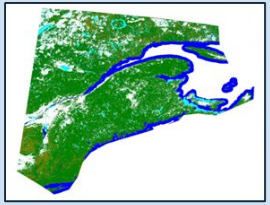


Image segmentation



Data postprocessing

Amazon EC2

- Instance type : *g4ad.4xlarge* (Ubuntu)
- The instance is set for automated start (Instance Scheduler).
- Delete temporary files (>> 3 GiB/ day)
- Once images are generated, the script is sending a stop signal to the instance

vCPUs	16
Memory (GiB)	64.0
Storage(GiB)	600
Pricing	$0.8670/hr On Demand

Additional Infrastructure requirements

Request Google Cloud resources to establish a seamlessly integrated system, utilizing Google Compute Engine for integrated processing and enhanced data management capabilities.

g4dn instances equipped with single and multiple GPUs to significantly accelerate the inference process for segmentation tasks compared to CPU-based processing.

Case Study: Community Streamflow Evaluation System (CSES) (University of Alabama)

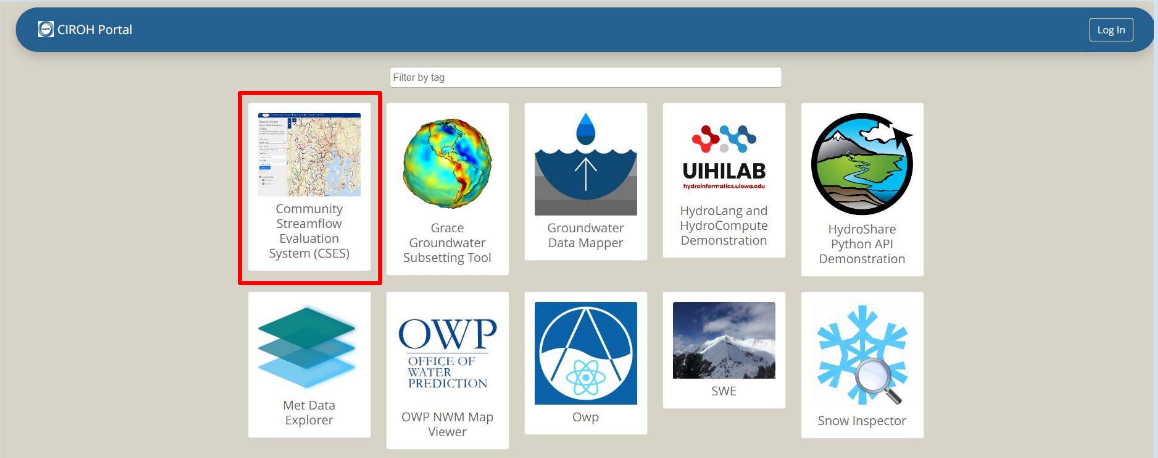
Presenter: Ryan Johnson
AI Research Scientist

Project Overview: Community Streamflow Evaluation System (CSES)

CSES is a model evaluation platform designed to lower the barrier for researchers and end-users in assessing hydrological models.

CSES provides a:

- Model agnostic evaluation framework
- Research-oriented version for developers
- Web-based version for end-users
- Interactive model evaluation and analysis



Contribution to CIROH

Centralized location for all CIROH model evaluation tools

- Reach-based Eval
- State-based Eval
- HUC-based Eval

Framework for sharing CIROH Streamflow prediction advancements with the greater hydrological community

Tethys- CSES GitHub

Tethys- CSES App

Relevance to NextGen

CSES-Python supports developers assess the model skill from different modeling formulations - regionally dominate hydrology.

CSES Web-App provides a platform to share modeling advancements with the hydrological community - researchers to end users.

CSES Python GitHub

Infrastructure Utilized

CIROH 2i2c **JupyterHub**

Amazon Web Services S3

Amazon Web Services EC2

UA Pantarhei HPC

Pantarhei

Additional Infrastructure

None!

Next Steps:

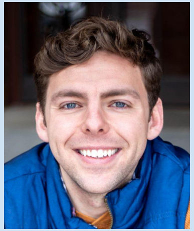
- Streamflow Extreme Event Dataset (SEED)
- HydroFabric integration
- Increase community participation
- RTI Tools for Exploratory Evaluation in Hydrologic Research (TEEHR)
- UA-MINES Snow-Water-Equivalent Evaluation Tool (SWEET)

Case Study: Modernized Standards and Tools for Sharing and Integrating Real-time Hydrologic Observations Data (Utah State University)

Presenter: Jeff Horsburgh

Associate Professor at Utah Water Research Laboratory

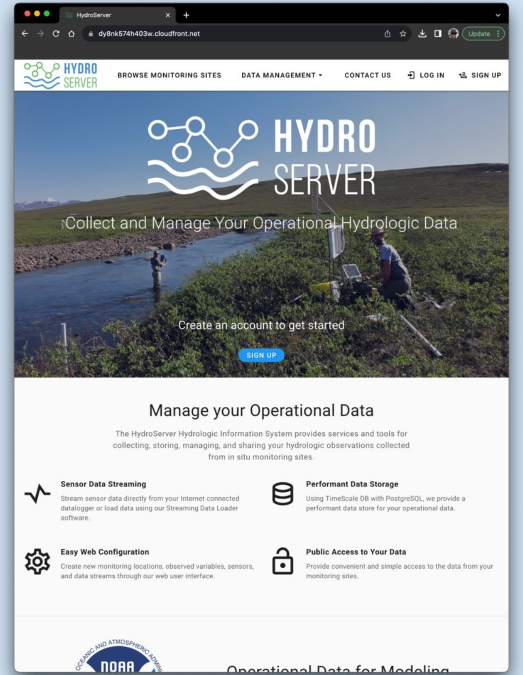
Project Overview



HydroServer is a modern hydrologic information system (HIS) software designed to streamline operational collection, management, and use of real-time hydrologic sensor data

Broadly, it supports both users and producers of hydrologic data by addressing the following core operational needs:

- Efficient Operational Data Storage
- Automated Data Ingestion from Sensors
- Data Management and Quality Control Tools
- Standardized Data Dissemination and Archival Services



Utah Water Research Laboratory
UtahStateUniversity

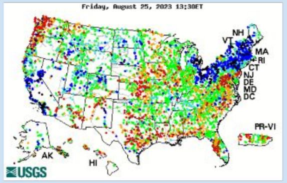
Contribution to CIROH

HydroServer addresses the three following CIROH Research Themes:

1. *<u>Improvement of Water Resources Prediction Systems</u>:* HydroServer enables the integration of a wider array of data sources into modeling and prediction systems - expanding beyond USGS NWIS.
2. *<u>Advancement and Acceleration of Community Water Resources Modeling</u>:* HydroServer supports CIROH's contribution to NextGen by allowing a broader range of organizations to share their hydrologic data for use in the National Water Model.
3. *<u>Advancement and Augmentation of Hydroinformatics</u>:* HydroServer is based on previous work developing a hydrologic information system while enhancing the concept with new and modern technologies.

Relevance to NextGen

- NextGen modeling will enable/rely on more localized model formulations
- Data assimilation and model testing can be enhanced by local-scale data
- HydroServer provides cyberinfrastructure for building a cooperative network of streamflow monitoring sites (and eventually water quality) to support NextGen modeling



USGS NWIS Gage Network

+

HYDRO SERVER

New HydroServer COOP Gage Network

=

Impact

- **Enhanced, National-Scale Stream Gage Network**
- **More data available for operational modeling**

Infrastructure Utilized

- AWS ElasticBeanstalk
  - Platform-as-a-Service (PaaS) for automatically deploying, scaling, and managing the Python/Django backend application
- AWS CloudFront
  - Content Delivery Network (CDN) for serving the HydroServer web app and other data files.
  - Used as a reverse proxy to route traffic to appropriate S3 and ElasticBeanstalk services
- AWS S3
  - Object storage service for HydroServer deployments.
  - Used to store live HydroServer code as well as multimedia data such as photos

Case Study: Using CIROH HPC and Cloud Resources for ML workflows

Chaopeng Shen
Professor at Pennsylvania State University

Additional Infrastructure Requirements

- Timescale Cloud
  - Relational database service used to host operational observations data
  - Built on PostgreSQL with the Timescale extension optimized for storing time-series data
- GitHub (https://github.com/hydroserver2)
  - GitHub Actions and Environments alongside Terraform infrastructure as code (IaC)
  - Provide a framework for automated continuous integration and deployment (CI/CD) workflows
  - Help streamline the deployment and management of HydroServer instances

Project Overview

Multiscale Hydrologic Processes and Intelligence (MHPI) group

Differentiable models are a genre of physics-informed machine learning model that mix process-based equations (called priors) and neural networks (NNs) with the following advantages

- **Interpretable**: they represent the full physical processes and output intermediate physical variables not used in training
- **Generalizability**: because of the use of priors, differentiable models often generalize better under data-sparse regions like in global hydrologic simulations.
- **Knowledge Discovery**: they can use NNs as question marks in the combined system to learn unknown/uncertain relationships from data.
- Respect physical laws and sensitivities

Contribution to CIROH

Differentiable models mainly contributes to the following CIROH Research Themes:

1. *Improving streamflow and routing simulation by integrating machine learning and physics-based modeling*
   - **Build a physics-informed and learnable flow routing method for global scale river networks**

2. *Developing and benchmarking data assimilation (DA) methods on a standardized testbed*
   - **Develop DA method for both ML models and physics-informed, differentiable models and expand DA to update flows throughout a river network using gauged sites.**

Relevance to NextGen

- We will enable backpropagation in the Nextgen framework to support the training and assimilation tasks of **pure AI models and differentiable models.**

- The differentiable model framework will be used to identify the best model structures among the **NextGen candidates** (differentiable SAC-SMA, HBV, PRMS, CFE, etc.) with a learnable regional parameterization module.

- Large scale differentiable models with the state-of-the-art performance can be used as NextGen models

Infrastructure Utilized

- **Wukong**:
  - Enable the training and execution of AI and differentiable models on a global and continental scale using multi-GPUs with substantial capacity.
  - Accelerate large-scale training works with powerful NVIDIA GPUs (e.g., NVIDIA A100/H100)
  - Provide high-speed storage for efficient data access.

- **AWS S3**
  - Data storage for global and continental scale modeling.
  - Enable the data transfer among local machines and Wukong

Q&A

CIROH Cyberinfrastructure Support Team

- Join us on CIROH Slack
  #ciroh-ua-it-admin

- Email admins at : ciroh-it-admin@ua.edu
- Support email : ciroh-it-support@ua.edu