Study Area and Programming Environment

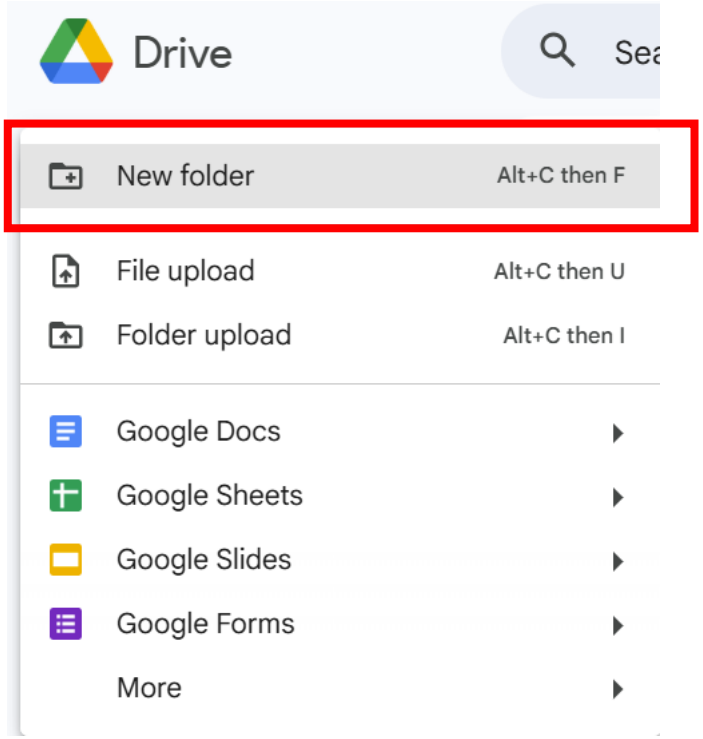
Vidya Samadi, Ph.D. , M.ASCE

Clemson University

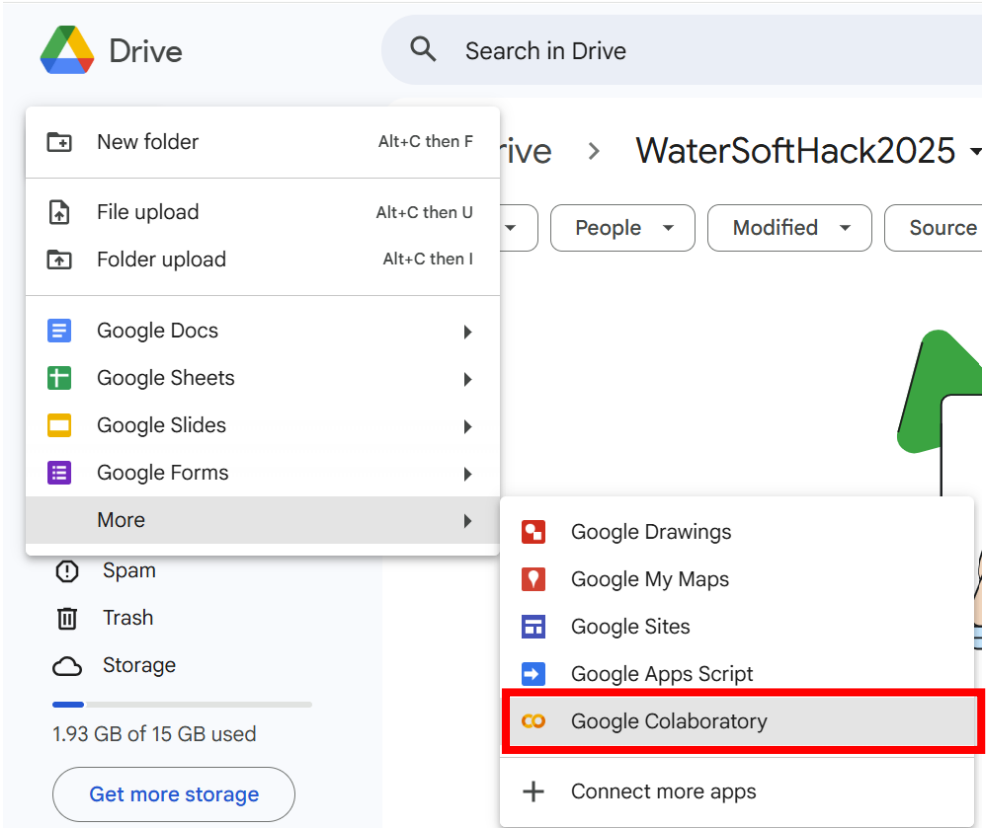
Google Colab

What is Colab?

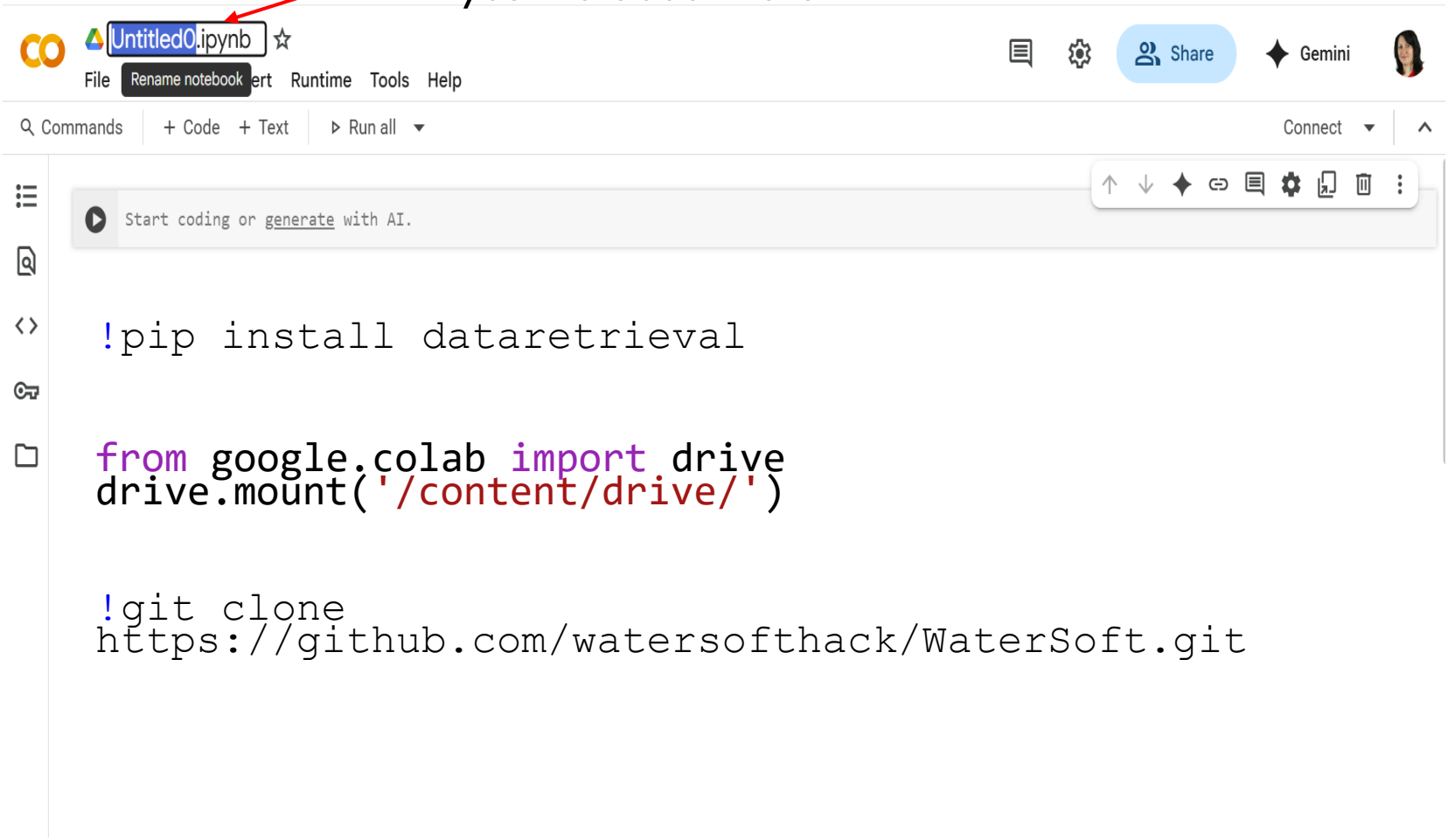
- Colaboratory is a Google research project created to help disseminate machine learning education and research.
- Colab a Jupyter notebook environment that requires no setup to use and runs entirely in the cloud
- Colab allows you to use free Tesla K80 GPU, which is a powerful hardware component useful for machine learning tasks.
- Colab provides 12GB of RAM, which is the memory available for your code and data.
- You can use it up to 12 hours in row (You need to restart the session after 12 hours).
- The session with the GPU and RAM is limited to 12 hours. After this period, you need to restart your Colab session to continue using the resources.


Drive
Sea
New folder Alt+C then F
File upload Alt+C then U
Folder upload Alt+C then I
Google Docs
Google Sheets
Google Slides
Google Forms
More

WaterSoftHack2025

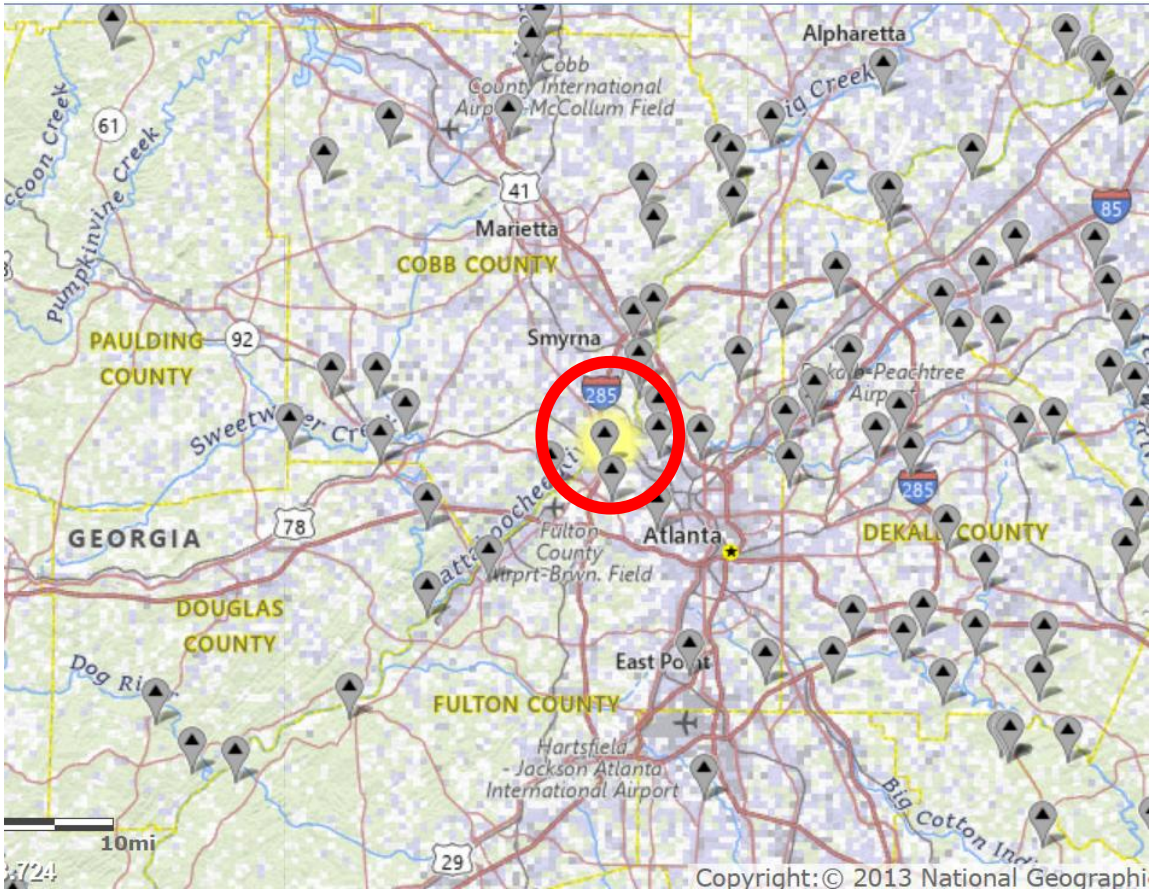

Drive
Search in Drive
New folder Alt+C then F
File upload Alt+C then U
Folder upload Alt+C then I
Google Docs
Google Sheets
Google Slides
Google Forms
More
Spam
Trash
Storage
1.93 GB of 15 GB used
Get more storage
rive
WaterSoftHack2025
People
Modified
Source
Google Drawings
Google My Maps
Google Sites
Google Apps Script
Google Colaboratory
Connect more apps

You can rename your notebook here



USGS Gauging Station

The Proctor Creek-ChattahoocheeRiver watershed (USGS 02336490), Fulton County, Georgia.



Application programming interface (API)

Vidya Samadi, PhD. , M.ASCE

Clemson University

How to Extract Data?

- "dataretrieval" Package
- Manually!
- Application Programming Interface (API)

An API is an interface between multiple applications. It allows two applications to communicate.

"dataretrieval" Python Package

Tools

dataretrieval-python

Keywords: data retrieval; hydrology; python; NWIS; water

Domain: Hydrology

Language: Python

Description:

dataretrieval-python was created to simplify the process of loading hydrologic data into the Python environment. Like the original R version dataRetrieval, it is designed to retrieve the major data types of U.S. Geological Survey (USGS) hydrology data that are available on the Web, as well as data from the Water Quality Portal (WQP), which currently houses water quality data from the Environmental Protection Agency (EPA), U.S. Department of Agriculture (USDA), and USGS. Direct USGS data is obtained from a service called the National Water Information System (NWIS). Note that the python version is not a direct port of the original: it attempts to reproduce the functionality of the R package, though its organization and interface often differ. If there's a hydrologic or environmental data portal that you'd like dataretrieval to work with, raise it as an issue.

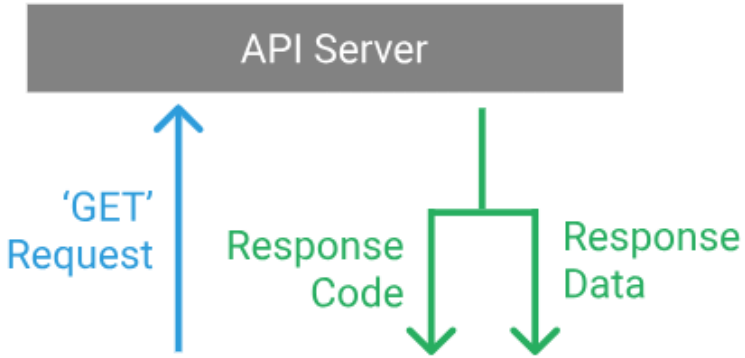
Citation: Hodson, T.O., Hariharan, J.A., Black, S., and Horsburgh, J.S., 2023, dataretrieval (Python): a Python package for discovering and retrieving water data available from U.S. federal hydrologic web services: U.S. Geological Survey software release, https://doi.org/10.5066/P94I5TX3 .

How APIs work!

An API follows the principle of question and answer, exactly as humans do.

Questions, and answer are the same as humans, but it's written in a technical way.

Request: "Give me USGS 02336490 data"
Answer: "Here's USGS 02336490 data"



How to Request an API

To communicate with APIs, we use four basic requests.

- **GET**- Get data (get all USGS 02336490 data)
- **PUT**- Update a piece of data (update USGS 02336490 stream data)
- **POST**-Create and save new data
- **DELETE**-Delete data (Delete USGS 02336490)

An API answers your request with a language called **JSON**.

JSON stands for "**JavaScript Object Notation**". It's a data representation easy to read by humans and easy to manipulate through code.

API Status Codes

Status codes indicate information about what happened with a request. Here are some codes that are relevant to *GET* requests:

200: Everything went okay, and the result has been returned (if any).

301: The server is redirecting you to a different endpoint. This can happen when a company switches domain names, or an endpoint name is changed.

400: The server thinks you made a bad request. This can happen when you don't send along the right data, among other things.

401: The server thinks you're not authenticated. Many APIs require login credentials, so this happens when you don't send the right credentials to access an API.

403: The resource you're trying to access is forbidden: you don't have the right per lessons to see it.

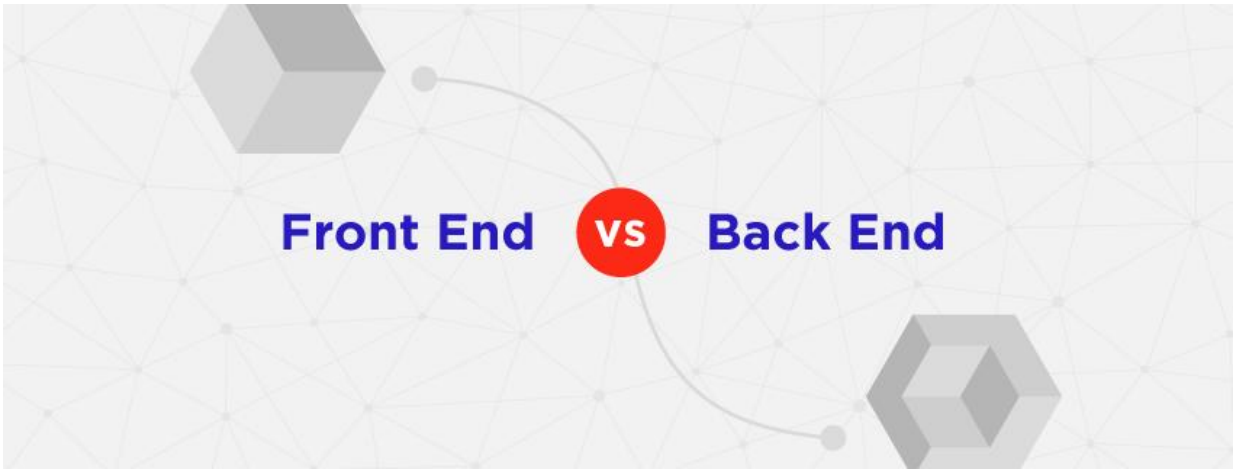
404: The resource you tried to access wasn't found on the server.

503: The server is not ready to handle the request.

If your status code starts with a '2' it was successful and if it starts with a '4' or '5' there was an error.

How an API Answers!

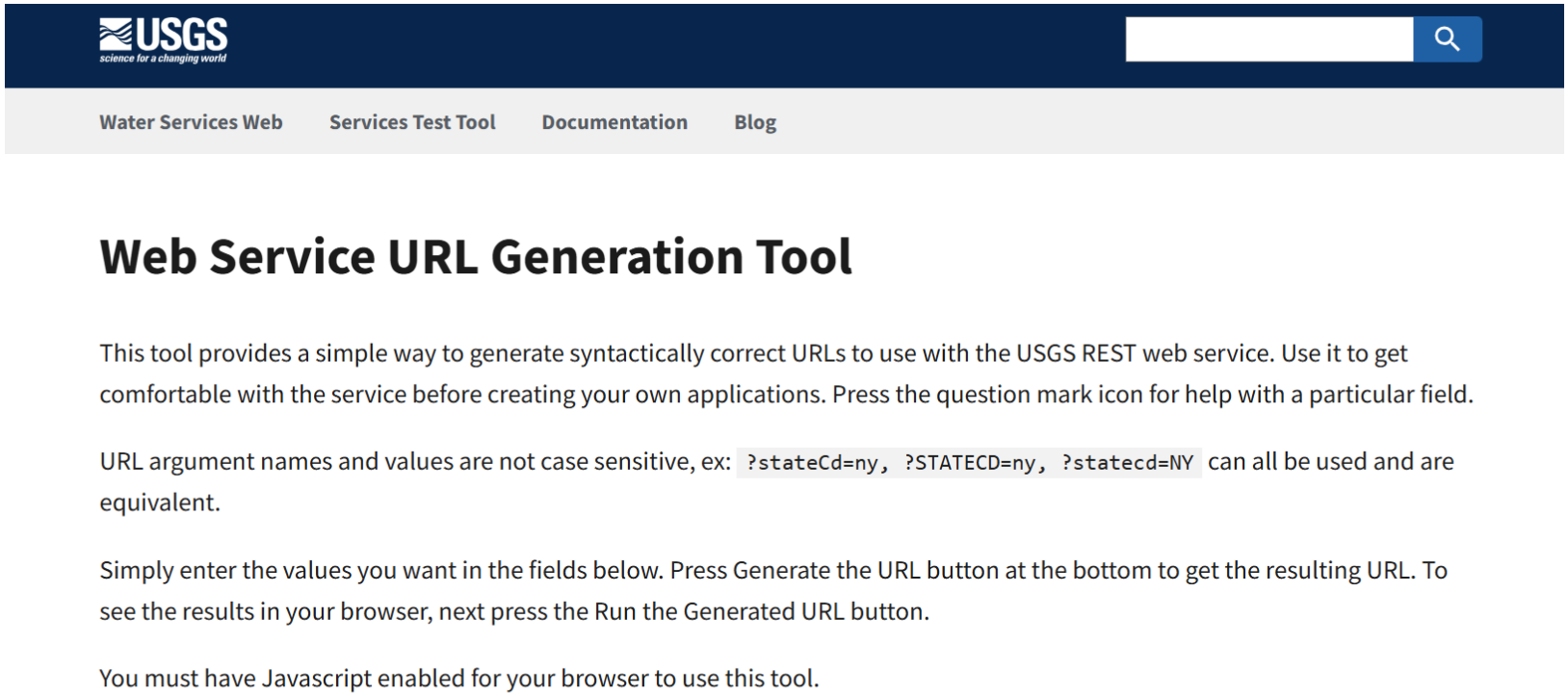
When we use an API on our project, we will divide the visual part (front-end) from the server (back-end). It means that our application will be at least divided into two projects.



Many federal agencies build their applications and give free access to their API at the same time.

Generate API URL for a USGS Gauging Station

API URL for any gauging stion in US: https://waterservices.usgs.gov/test-tools/?service=dv&siteType=&statTypeCd=all&major-filters=hydrologic-unit-codes&huc=02336490&format=json&date-type=type-none&statReportType=daily&statYearType=calendar&missingData=off&siteStatus=all&siteNameMatchOperator=start



Generate API URL for a USGS Gauging Station-cont.

Select Service to Test

Daily Values Service

1 Select a Major Filter

A **major filter selected with a value is required** in order for the generated URL to function properly.

Major Filter

Hydrologic Unit Codes

Hydrologic Unit Codes

02336490

Generated API URL

3 Current URL Generated

When a major filter and any optional filters have been selected, the generated URL below will give the data based upon the selections above. Be sure to use the best filters possible to reduce the records to only those that you need.

https://waterservices.usgs.gov/nwis/dv/?format=json&huc=02336490&siteStatus=all

Copy generated API URL | Run generated URL | dataRetrieval Code | Share

NAIRR Pilot National Artificial Intelligence Research Resource Pilot

NSF National Artificial Intelligence Research Resource (NAIRR) Pilot

Advancing US Innovation in Artificial Intelligence

Vidya Samadi, PhD. , M.ASCE

Clemson University

NSF NAIRR Resources

Computing Resources

NAIRR Pilot National Artificial Intelligence Research Resource Pilot

High-Performance Computing (HPC) Systems: Access to the DeltaAI at the National Center for Supercomputing Applications, Lonestar at the Texas Advanced Computing Center (TACC), and other systems.

Cloud Computing Platforms: Access to major cloud providers through dedicated programs, including AWS, Google Cloud, and Microsoft Azure, offering a wide range of services, compute instances (including GPUs and TPUs), and storage solutions.

Specialized AI Accelerators: Access to innovative computing hardware designed specifically for AI workloads, such as Cerebras Wafer-Scale Engine 2 (CS-2) AI Accelerator, SambaNova DataScale SN30 system, Graphcore Bow Pod 64, and Groq LPU Inference Engine, available through DOE's ALCF AI Testbed.

Access to Models, Software, and Tools: Access to pre-trained models, software tools, and privacy-enhancing technologies for various AI research needs.

NSF NAIRR Resources-cont.

Datasets

AI-ready Datasets: Access to a growing collection of curated datasets suitable for AI training and research.

Open Datasets: Availability of publicly accessible datasets from various domains through cloud provider registries (e.g., AWS Registry of Open Data).

Domain-specific Datasets: Specialized datasets for areas like agriculture, environmental research, etc.

NSF NAIRR Resources-cont.

Training & Education Resources

NAIRR Pilot National Artificial Intelligence Research Resource Pilot

Online Interactive Notebooks: Resources for educators and instructors to use in classrooms or for student projects.

Educational Platforms: Platforms like Vocareum offer access to notebooks and cloud resources to support hands-on AI education.

Training and Workshops: Opportunities to participate in workshops and training sessions to learn about and utilize AI technologies and platforms effectively.

Collaborations: Opportunities to collaborate with researchers and experts from various partner organizations, including EleutherAI.

Current NAIRR Opportunities

NAIRR Pilot Portal: A central website (nairrpilot.org) to discover and apply for access to these resources and view descriptions of awarded projects.

New Opportunity

START-UP PROJECTS RESOURCES

Request access to AI resources for start-up projects

Apply Now

Requires Application

RESEARCH RESOURCES

Access high-performance computing platforms tailored for AI research.

Apply Now

Requires Application

EDUCATIONAL RESOURCES

Request access to educational platforms (such as computational notebooks).

Apply Now

Freely Accessible

DATA, MODELS, AND MORE

Browse curated datasets, pre-trained models, and additional tools for training and testing your AI systems.

View Resources

NAIRR Resources Available for WaterSoftHack CyberTrainng

NRP NATIONAL RESEARCH PLATFORM

Please accept the Acceptable Use Policy

☐ Accept the NRP AUP

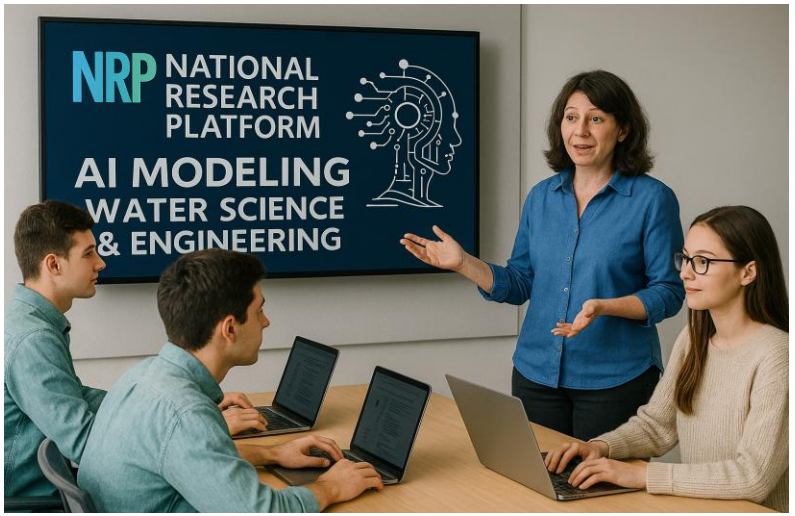
Required. By checking this box, I accept the Acceptable Use Policy and agree to follow the cluster policies.

The AUP can be found at https://nrp.ai/NRP-AUP.pdf.
Please review the cluster policies.

Continue

- ◉ NRP Deep Learning & Data Science Full, PyTorch
- ○ NRP Deep Learning & Data Science Full, TensorFlow
- ○ B-Data python scipy
- ○ B-Data Julia

.edu email address



Notebook Integration for Foundational AI Education (Assistant Professor/Postdoc)

vocareum

Home Manage Help Inbox samadi@clemson.edu

Clemson University (NAIRR) (Ends - Jan 01 2026)

Edit Assignments Enrollment Submissions Class Sessions Dashboard Settings

Assignment: Vocareum Notebook

Search: (Student View) adarsh...

My Student View

(Student View) adarshn@clemson.edu

Budget	Monthly	Total
Time (Min)	0	0
Spend	0	0

vocareum

CLEMSON UNIVERSITY IOWA CUAHSI allied for water science www.cuahsi.org Tulane University

Summary

- NAIRR pilot aims to democratize access to AI infrastructure and expertise, fostering innovation and training a diverse AI-savvy workforce.
- US-based researchers and educators affiliated with US-based institutions are eligible to apply.
- Note that individual resources may have specific eligibility requirements, and your institutional email must be used.
- Allocated resources to WaterSoftHack include CPU, GPU, and classroom notebook (Vocareum).

Email NAIRR_Pilot@nsf.gov with any questions!

THANK YOU

The support from the National Science Foundation (award # 2320979) is acknowledged.