
1 Nov 2022

CUAHSI CYBERSEMINAR



UA – CIROH 2i2c JUPYTERHUB INFRASTRUCTURE

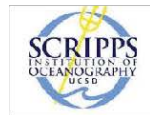
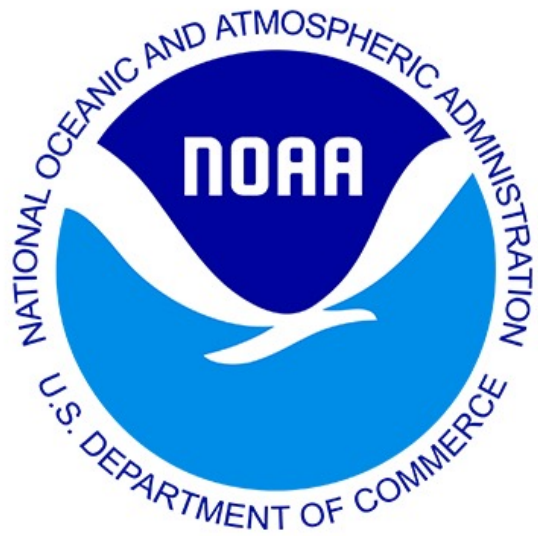


James Halgren
jshalgren@ua.edu

Arpita Patel

apatel54@ua.edu





NATIONAL WATER MODEL DATA CHALLENGE

- Size is formidable; 1TB + per day operational output, 100+ TB for complete archive of “retrospective” simulation
- Format is focused on output needs from models
- Queries tend to be orthogonal (and inefficient) to storage
- Big data providers host many of these datasets; more are coming
- More here: <https://github.com/2i2c-org/awi-ciroh-image/issues/16>

Current River Forecast Points (~3,600)



+

NWM Streamflow Output Points (~2.7 mil)



2i2c – SERVICES

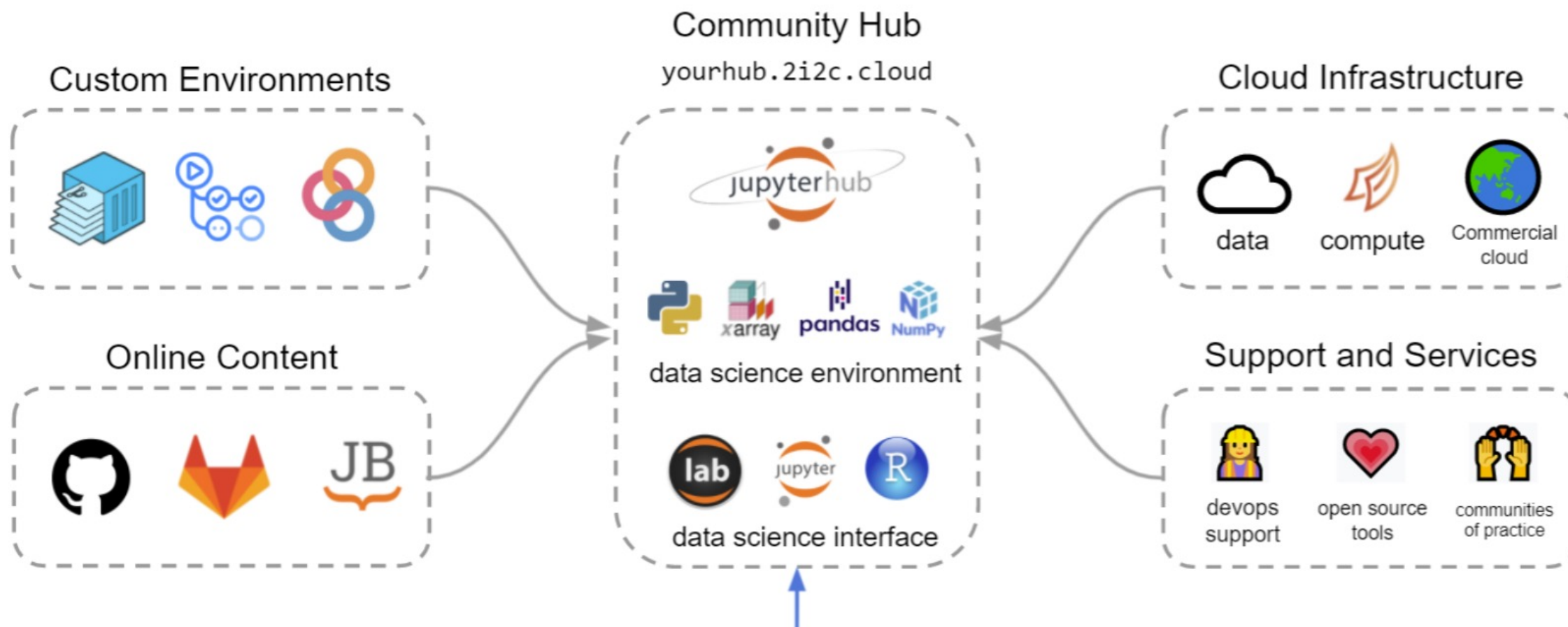
2i2c

<https://2i2c.org/>

- **JupyterHub** is an open-source project for providing and managing interactive computing sessions to multiple users, usually deployed on shared infrastructure.
 - **JupyterHub for Kubernetes** is a distribution of JupyterHub designed for use with the scalable Kubernetes platform.
 - **The Right to Replicate** is a guiding principle of 2i2c Hub infrastructure. It gives communities the right to replicate their infrastructure in its entirety elsewhere, with or without 2i2c.
 - **Expert** guidance and support
-

2I2C JUPYTER HUB

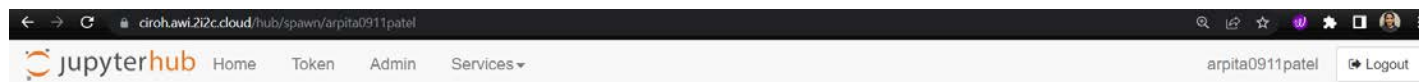
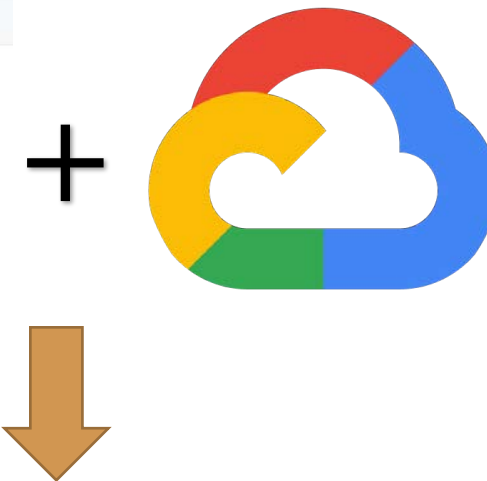
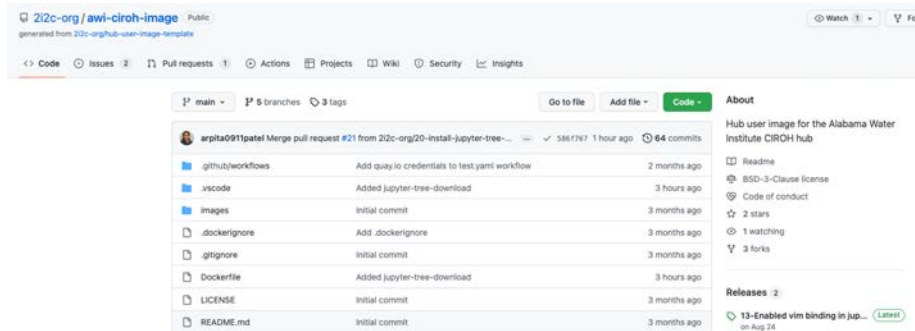
- Interactive computing and storage for the community
- Managed **JupyterHub** service
- Based on **JupyterHub** and **JupyterHub for Kubernetes** - open-source projects for providing and managing interactive computing sessions to multiple users, deployable on shared infrastructure designed for use with the scalable Kubernetes platform.
- Provided with **The Right to Replicate** is a guiding principle of 2i2c Hub infrastructure. It gives communities the right to replicate their infrastructure in its entirety elsewhere, with or without 2i2c.



CIROH PARTNERSHIP WITH 2I2C

- Provide cloud-based hub to exploit Google-cloud hosted NWM output
 - Provision compute collocated with NWM output data
 - Establish template open architecture “Right-to-Replicate”
 - Address the volume problem with collocation
 - Create a portal to work on the format challenges
-

CIROH-UA JUPYTERHUB



- **Production env:**

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

- **Staging env:**

<https://staging.awi.2i2c.cloud/>

- **Image repo:**

<https://github.com/2i2c-org/awi-ciroh-image>

Server Options

- ☒ **Small**
5GB RAM, 2 CPUs
- ☐ **Medium**
11GB RAM, 4 CPUs
- ☐ **Large**
24GB RAM, 8 CPUs
- ☐ **Huge**
52GB RAM, 16 CPUs

Start

HUB LANDING PAGE



ciroh.awi.2i2c.cloud/user/arpita0911patel/lab/tree/git/data_access_examples/parquet/example_remote_dl_parquet-mediumrange-multipledays.ipynb

File Edit View Run Kernel Git Tabs Settings Help

Filter files by name

/ git / data_access_examples /

Name	Last Modified
data	9 days ago
parquet	3 days ago
README.md	a month ago

```
[2]: %load_ext nb_black
import pandas as pd
from pyarrow.parquet import ParquetFile
import dask.dataframe as dd
import os
import xarray as xr
import ujson
import pprint
```

```
[3]: # These fs options don't work for http... beware!
so = dict(mode="rb", anon=True, default_fill_cache=False, default_cache_ty
```

```
[4]: def gen_json(u, fs, outf=None):
    with fs.open(u, **so) as infile:
        h5chunks = SingleHdf5ToZarr(infile, u, inline_threshold=300)
        p = u.split("/")
        date = p[3]
        fname = p[5]
        if outf:
            # outf = f'{json_dir}{date}.{fname}.json'
            with open(outf, "wb") as f:
                f.write(ujson.dumps(h5chunks.translate()).encode())
        else:
            return h5chunks.translate()
```

Cell Tags
Add Tag +

Slide Type
▼

Raw NBConvert Format
▼

Advanced Tools
▲

TERMINAL VIEW



#WOCinTech

ciroh.awi.2i2c.cloud/user/arpita0911patel/lab/tree/git/data_access_examples

File Edit View Run Kernel Git Tabs Settings Help

Filter files by name

/ git / data_access_examples /

Name	Last Modified
data	9 days ago
parquet	3 days ago
README.md	a month ago

example_remote_cX example_remote_cX Untitled.ipynb Terminal 2

```
(notebook) jovyan@jupyter-arpita0911patel:~$ cat /etc/os-release
PRETTY_NAME="Ubuntu 22.04 LTS"
NAME="Ubuntu"
VERSION_ID="22.04"
VERSION="22.04 LTS (Jammy Jellyfish)"
VERSION_CODENAME=jammy
ID=ubuntu
ID_LIKE=debian
HOME_URL="https://www.ubuntu.com/"
SUPPORT_URL="https://help.ubuntu.com/"
BUG_REPORT_URL="https://bugs.launchpad.net/ubuntu/"
PRIVACY_POLICY_URL="https://www.ubuntu.com/legal/terms-and-policies/privacy-policy"
UBUNTU_CODENAME=jammy
(notebook) jovyan@jupyter-arpita0911patel:~$
```

No properties to inspect.

ADMIN SUPPORT



← → ↻ ciroh.awi.2i2c.cloud/hub/admin#/ 🔍 ⌂ ☆ ⚙️ 🗑️ 👤

jupyterhub Home Token Admin Services ▾ arpita0911patel Logout

Search users [> Manage Groups](#)

User ↕	Admin ↕	Server ↕	Last Activity ↕	Running ↕	Actions
Add Users			Start All Stop All	Shutdown Hub	
▼ choldgraf	admin		Never	Start Server Spawn Page	Edit User
▼ sgibson91	admin		3 months ago	Start Server Spawn Page	Edit User
▼ colliand	admin		2 months ago	Start Server Spawn Page	Edit User
▼ georgianaelena	admin		Never	Start Server Spawn Page	Edit User
▼ yuvipanda	admin		8 days ago	Start Server Spawn Page	Edit User
▼ jameshalgren	admin		3 days ago	Start Server Spawn Page	Edit User
▼ consideratio	admin		Never	Start Server Spawn Page	Edit User
▼ jmunroe	admin		2 months ago	Start Server Spawn Page	Edit User
▼ damianavila	admin		Never	Start Server Spawn Page	Edit User
▼ whitelightning450			1 minutes ago	Stop Server Access Server	Edit User

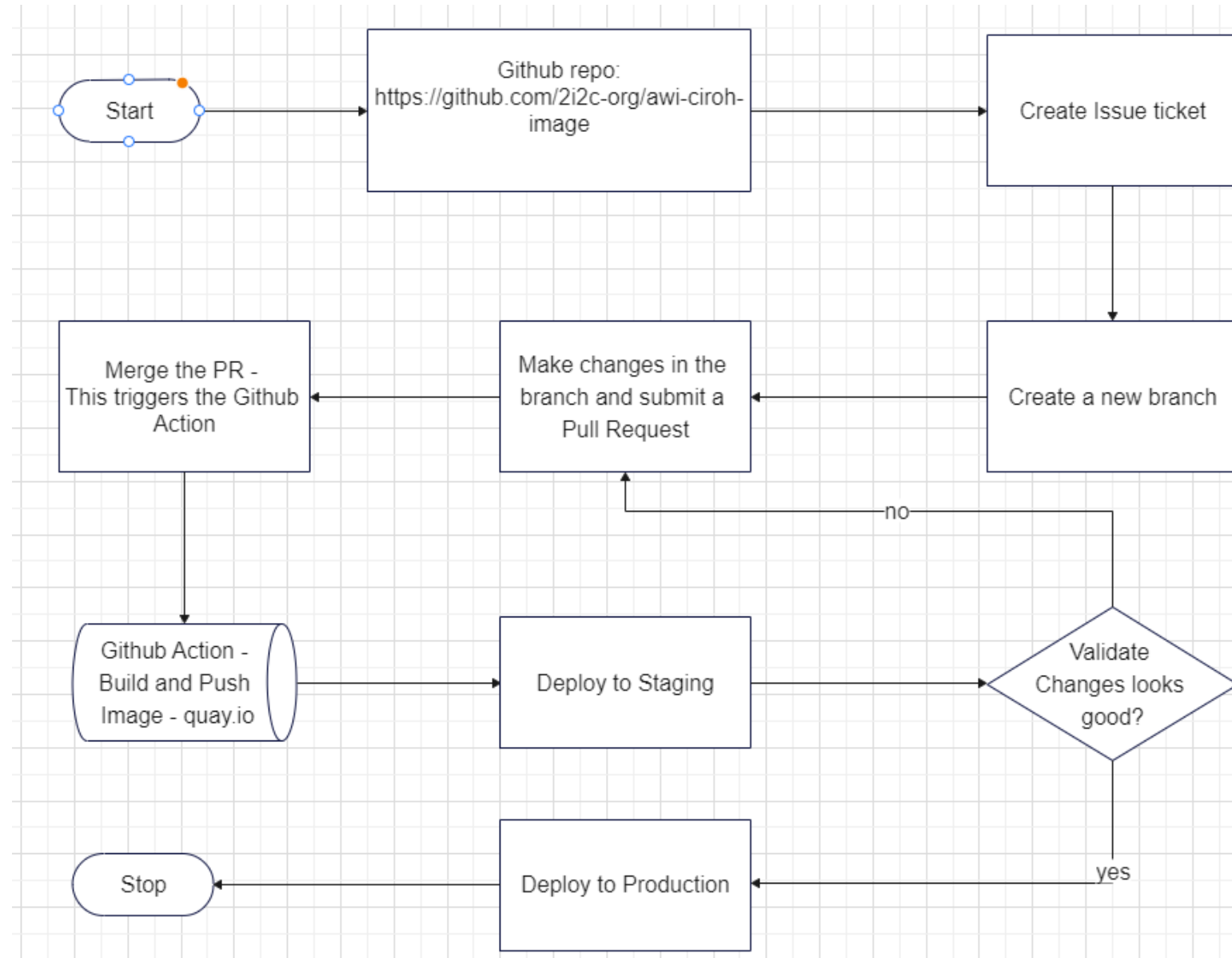
Fork

2I2C - TECHNOLOGY STACK

- **Docker** – to build customized image for the users
- **Kubernetes** - to manage resources on the cloud
- **Helm v3** - to configure and control the packaged JupyterHub installation
- **Terraform** – to build Kubernetes clusters on cloud providers.
- **Storage** – object storage, cloud storage (s3, gcsfs)
- **Cloud Providers** - Google Cloud, Microsoft Azure, Amazon EC2, IBM Cloud...
- **JupyterHub** – to give users access to a Jupyter computing environment
- **Domain registration** - make the hub available at <https://your-domain-name.com>



CI/CD STEPS



TEST DRIVE!





REFERENCES

- AWI-CIROH Hub Image: <https://github.com/2i2c-org/awi-ciroh-image>
 - 2i2c Hub Service Guide: <https://docs.2i2c.org/en/latest/>
 - 2i2c Infrastructure Guide: <https://infrastructure.2i2c.org/en/latest/>
 - Questions: support@ciroh.org
-