Interactive distributed computing with Jupyter, Python and Dask on Jetstream2 for Science Gateways

Andrea Zonca, Le Mai Weakley and Julian Pistorious
Deploy Kubernetes and JupyterHub

- On Jetstream 2 Openstack deployment
- Kubernetes for orchestration/networking/logging/resiliency
- JupyterHub runs a pod for each user across the cluster

https://docs.jetstream-cloud.org/general/k8skubespray/
https://zonca.dev/2022/03/jetstream2-jupyterhub.html
Laptop user 1

Jetstream VM Main

Kubernetes

Notebook+Kernels

user 1

Laptop user 2

Jetstream VM Worker

Notebook+Kernels

user 2
Hands-on: Login to JupyterHub

Point browser to:

bit.ly/gw23jhub

Authenticate with any Github account

Open terminal:

git clone https://github.com/zonca/dask-jetstream-tutorial

Execute `01_environment.ipynb`
Hands-on: Dask tutorial with local threads

Point browser to:

bit.ly/gw23jhub

Dask - single machine - local threads scheduler

- `00_install_dot.ipynb`
- `02_dataframe.ipynb`
- `03_array.ipynb`
- `04_delayed.ipynb`
Deploy Dask Gateway

- Give more computational resources to users
- Use dask for high-level distributed computing

https://www.zonca.dev/posts/2023-09-28-dask-gateway-jupyterhub
Hands-on: Dask tutorial with Dask gateway workers

Point browser to:

bit.ly/gw23jhub

Dask - 2 workers - distributed scheduler

- `05_dask_gateway.ipynb`
- `06_array_distributed.ipynb`
Parallel data storage with Object Store and Zarr

- Read/write data in parallel to Object Store with dask in Zarr format
- Zarr is a cloud-native file format for chunked/compressed/multi-dimensional arrays

https://www.zonca.dev/posts/2022-04-04-zarr_jetstream2
Hands-on: Write data to Object Store in Zarr format

Point browser to:

bit.ly/gw23jhub

Dask - write to object store with Zarr

- `07_zarr_object_store.ipynb`
The end

Links, slides, tutorials: zonca.dev/posts/gw23

Email Andrea Zonca (zonca@ucsd.edu), or let's meet in Pittsburgh!

Acknowledgments: thanks to Indiana University and the Jetstream team for funding this work.
Backup slides
Laptop user 1

- Chrome
- Chrome
- Chrome

Jetstream VM Main

Kubernetes

Jetstream VM Worker

Dask worker

Dask worker

Dask worker

Notebook+Kernels

user 1

JupyterHub

Dask gateway API

Dask gateway Proxy

Object Store

Zarr

Gateway

Dask scheduler

Flask