



PANGEO

Enabling Open, Reproducible, and scalable big data geoscience for everyone

Anne Fouilloux (Simula Research Laboratory, Norway),

Tina Odaka (IFREMER, France),

Jean-Marc Delouis (CNRS, France),

Alejandro Coca-Castro (The Alan Turing Institute, UK),

Pier Lorenzo Marasco (Provare LTD, UK)

What is Pangeo and what are we aiming for ?

We are a *global community* initiative for **Big Geoscience Data**

- Based on
 - Open Community
 - Open Platform
 - Open deployments
- Try it yourself!



We work on ensuring **more equitable access to data, software and hardware.**

A culture of collaboration

Lamont-Doherty Earth Observatory
COLUMBIA UNIVERSITY | EARTH INSTITUTE



Helmholtz Centre for Ocean Research Kiel



UNIVERSITY OF LEEDS



UiO University of Oslo



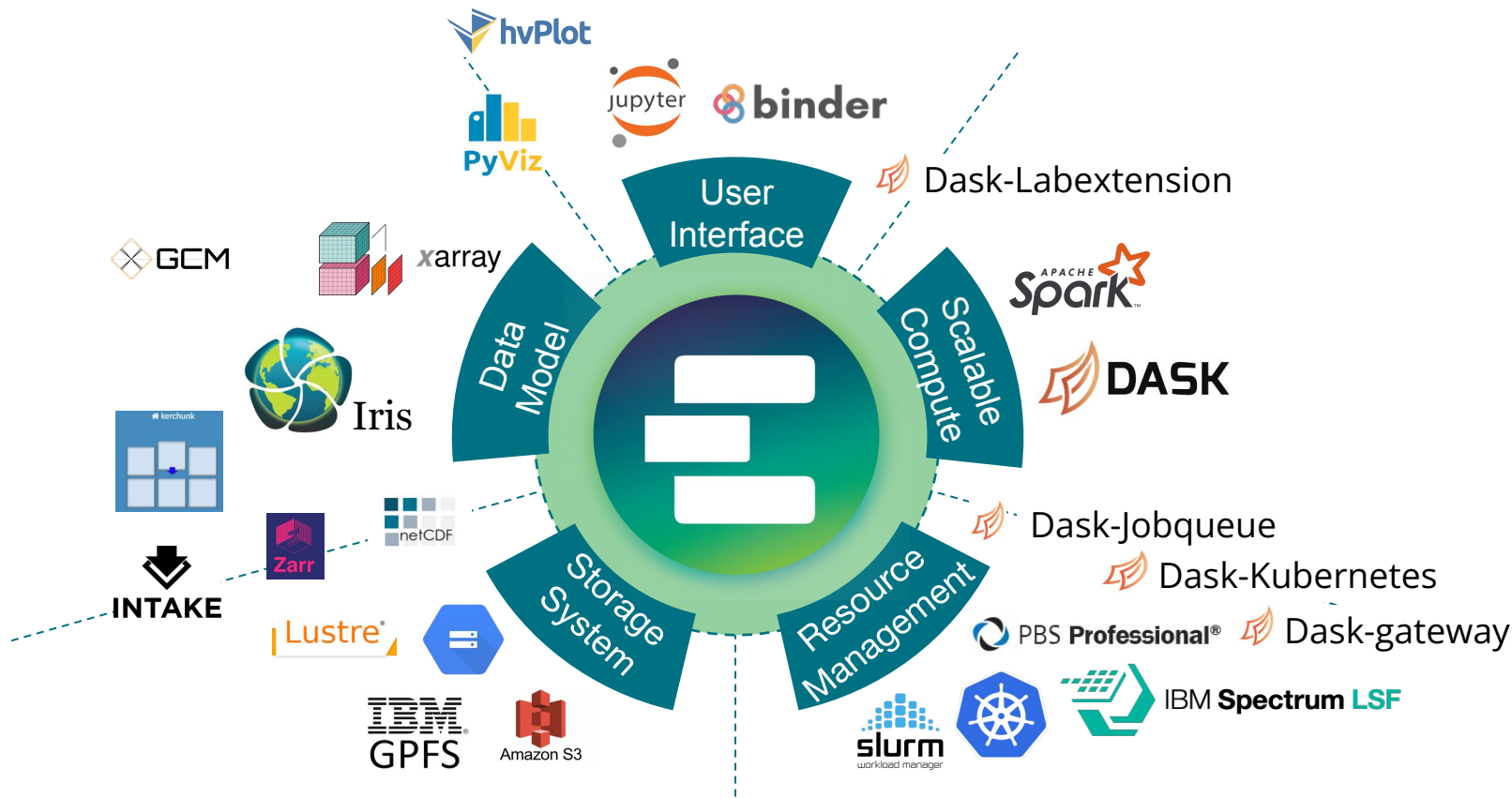
Massachusetts Institute of Technology®



Based on GitHub and papers affiliations

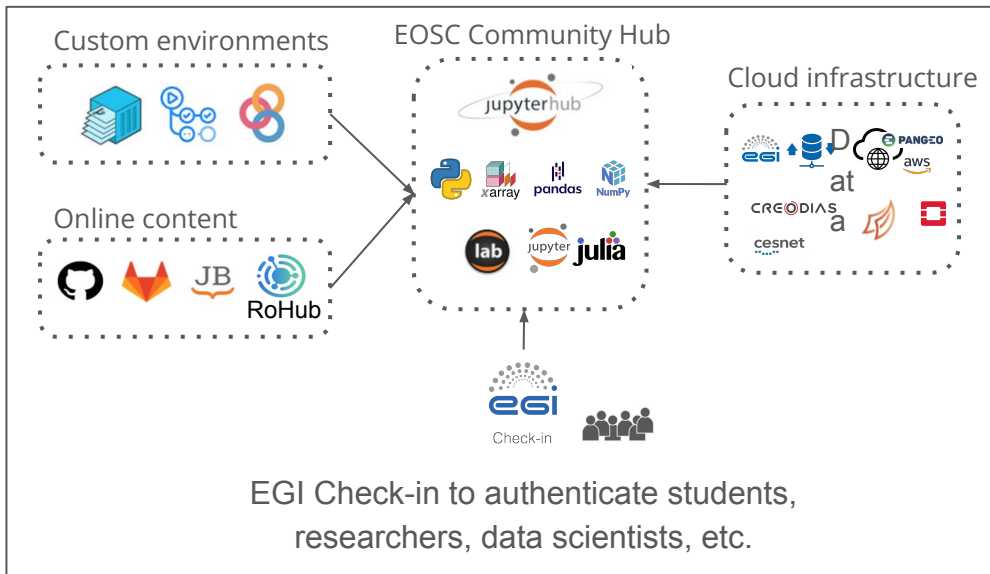


Modular, open and scalable platform



from laptop to cloud and HPC

Pangeo eossc

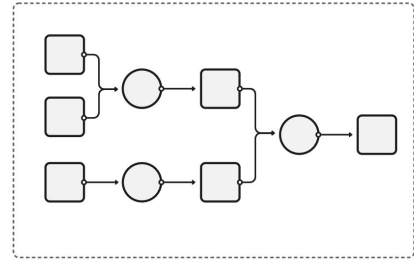


Collections
(create task graphs)

→ Task Graph


→ Schedulers
(execute task graphs)

- Dask Array
- Dask DataFrame
- Dask Bag
- Dask Delayed
- Futures



- Single-machine (threads, processes, synchronous)
- Distributed

High level collections are used to generate task graphs which can be executed by schedulers on a single machine or a cluster.

 **dask** Extends pandas and NumPy APIs to larger- than-memory data (chunks)

Reuse existing [2i2c Pangeo deployment configuration](#)



<https://pangeo-data.github.io/pangeo-eosc/>

Pangeo Training Infrastructure as a Service

FOSS4G workshop

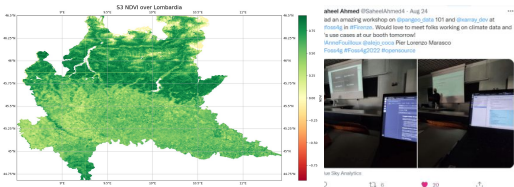
Overview

Questions

- Why do chunking matter?
- How can I read datasets by chunks to optimize memory usage?

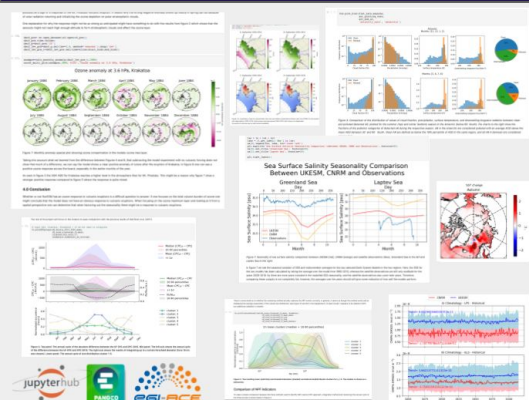
Objectives

- Learn about chunking
- Learn about zarr
- Use kerchunk to consolidate chunk metadata and prepare single ensemble datasets for parallel computing



Pangeo 101 workshop at the FOSS4G conference

eScience Course

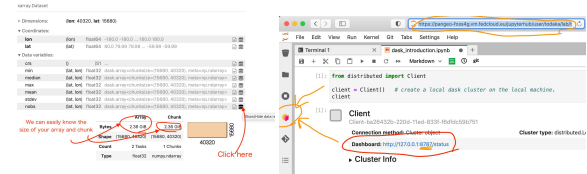


Integrate EOSC into master students' courses

CLIVAR Arctic Bootcamp



Supporting researchers in becoming competent practitioners in their scientific domain



Pangeo & OpenEO at BiDS



Request PTIaaS to train and teach!



Pangeo: where can I find information?

- More info & links here - <https://pangeo.io/meeting-notes.html>
- <http://pangeo.io/index.html#what-is-pangeo>
 - (<https://github.com/pangeo-data/pangeo>)
- Try some examples in the Pangeo Gallery
 - <http://gallery.pangeo.io>
- Subscribe in discourse, start discussions
 - <https://discourse.pangeo.io/>
- Follow pangeo-data from Github
- Discord channel
 - <https://discord.gg/ex5qqEyyTz>
- Interesting informations exist also in blogs
 - <https://medium.com/pangeo>
- Code of conduct.
 - https://github.com/pangeo-data/governance/blob/master/conduct/code_of_conduct.md



Thank you for your attention!

