While you wait

1. Connect to TAMU VPN and Login to Grace
   
   `ssh <username>@grace.tamu.edu`

2. Go to your scratch directory
   
   `cd $SCRATCH`

3. Clone the notebook repository from github
   
   `git clone https://github.com/abishekg7/python_geos.git`

   (OR)

   Copy notebooks from Grace scratch
   
   `cp -r /scratch/training/python_geos/notebooks .`
Python Tools for Geosciences

Fall 2021 HPRC Short Course
Nov 5, 2021

Abishek Gopal
Assistant Research Scientist
iHESP, Texas A&M Oceanography
Texas A&M High Performance Research Computing
Expectations for this course

• Get an overview of some recent Python libraries designed to support geoscientific analysis

• Learn about the data structures in xarray, how to load and visualize netCDF files, and some basic operations

• Explore other geoscience packages built on top of xarray

• Intended to be a starting point for switching your workflow to Python
Helpful HPRC resources

- Grace quick start guide
  - [https://hprc.tamu.edu/wiki/Grace:QuickStart](https://hprc.tamu.edu/wiki/Grace:QuickStart)

- Introduction to HPRC – Short course
  - [https://hprc.tamu.edu/training/intro_hprc.html](https://hprc.tamu.edu/training/intro_hprc.html)

- Submit tickets to [help@hprc.tamu.edu](mailto:help@hprc.tamu.edu)
Upcoming relevant HPRC short courses

- **Nov 12: Introduction to Fortran**
  - **Instructor:** Abishek Gopal
  - **Time:** Friday, Nov 12, 10:00AM - 12:30PM

- **Nov 19: Introduction to R**
  - **Instructor:** Ridham Patoliya
  - **Time:** Friday, Nov 19, 10:00AM - 12:30PM

- **Nov 19: Introduction to Julia**
  - **Instructor:** Jian Tao
  - **Time:** Friday, Nov 19, 1:30PM - 4:00PM

[https://hprc.tamu.edu/training/index.html](https://hprc.tamu.edu/training/index.html)
Acknowledgements

• Course materials adapted from detailed xarray, xgcm and Siphon tutorial notebooks
  – https://github.com/xarray-contrib/xarray-tutorial
  – https://gallery.pangeo.io/repos/xgcm/xgcm-examples/

• The HPRC team supporting the short course operations

• Sanjiv R., Steve Y., Fred C., Dapeng Li (iHESP)

• Kristen Thyng (previously: TAMU, now: Axiom Data Science)
Launching a JupyterLab notebook from Grace portal

1. Go to https://portal.hprc.tamu.edu/

2. Interactive Apps -> JupyterLab Geoscience
JupyterLab - Geoscience

This app will launch a JupyterLab server on the Grace cluster for the Python Tools for Geosciences short course.

Module

Anaconda3/5.3.0

Anaconda/3-x.x.x.x and Anaconda3 use Python3

Optional Environment to be activated

/scratch/training/python_geos/conda/envs/training

Enter the name of the environment to be activated.

Account

This field is optional.

Email

This field is optional.

☐ I would like to receive an email when the session starts

Press Launch

Check environment path

/scratch/training/python_geos/conda/envs/training

Hit Launch

* The JupyterLab - Geoscience session data for this session can be accessed under the data root directory.
Connect to JupyterLab session

```
cd $SCRATCH

git clone https://github.com/abishekg7/python_geos.git

cp -r /scratch/training/python_geos/notebooks .
```
Check if the virtualenv works correctly

1. Data structures in xarray
2. Reading and visualizing climate data
3. Computation with Xarray
4. Lazy loading and chunking with Dask arrays
5. Introduction to Dask

Exercise: Print Hello, world!

Each notebook will have exercises for you to solve. You'll be given a blank or partially completed cell, followed by a hidden cell with a solution. For example.

Print the text "Hello, world!".

```
[1]: print('Hello world')

Hello world

***
```

Run the cell below. Let's make sure the virtual environment is loaded correctly

```
[3]: import xarray
import cartopy

[ ]:
```
In case of session not starting or virtualenv issues

Email output.log to help@hprc.tamu.edu
Conda virtual environment

- cd $SCRATCH
  # Load Anaconda
  ml Anaconda3/2020.07

- # Create the virtual environment
  conda create python=3.7 -n training -c conda-forge

- # Activate the virtual environment
  source activate training

- # Install packages into the virtual environment
  conda install -c conda-forge cartopy matplotlib xarray
  conda install -c conda-forge xgcm dask esmpy

- # Deactivate the environment
  source deactivate
Course outline

- Intro to the Pangeo stack
- xarray data structures
- Reading and writing netCDF files
- Plotting with matplotlib and cartopy

- Spatial operations in xgcm
- Vertical interpolation in xgcm
- ESMPy regridding/remapping
- Data access using Siphon
Current/last generation of post-processing tools

- Mature tools/languages for working with moderate resolution datasets
- Often optimized to do specific tasks really well/fast.
- Not designed with high-resolution datasets in mind.
Python geo-scientific software stack

Credit: Ryan Abernathey. Inspired by Jake VanderPlas PyCon 2019
Pangeo

https://pangeo.io/architecture.html
## Build Your Own Pangeo

<table>
<thead>
<tr>
<th>Storage Formats</th>
<th>HDF</th>
<th>OPeNDAP</th>
<th>Cloud Optimized COG/Zarr/Parquet/etc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ND-Arrays</td>
<td>NumPy</td>
<td>DASK</td>
<td>More coming…</td>
</tr>
<tr>
<td>Data Models</td>
<td>xarray</td>
<td>Iris</td>
<td>pandas</td>
</tr>
<tr>
<td>Processing Mode</td>
<td>jupyter</td>
<td>Batch</td>
<td>Serverless</td>
</tr>
<tr>
<td>Compute Platform</td>
<td>HPC</td>
<td>Cloud</td>
<td>Local</td>
</tr>
</tbody>
</table>

Builds on NumPy by applying metadata such as dimensions, coordinates, data variables and attributes to raw NumPy arrays.

Inherits Pandas functionality

xarray.Dataset is an in-memory representation of the netCDF file format

xarray works seamlessly with the dask library to enable parallel computations more easily
Apply operations over named dimensions

Select values by label or logical conditions, instead of integer location

Easily use the `split-apply-combine` paradigm with `groupby`

Keep track of arbitrary metadata in the form of a Python dictionary
A comprehensive library for creating static, animated, and interactive visualizations in Python.

Cartopy adds understanding of map projections to matplotlib plots

https://matplotlib.org/gallery/

https://scitools.org.uk/cartopy/docs/latest/gallery/index.html
Short break!
(15 minutes)

We will resume at 11:30 CDT
Course outline

- Intro to the Pangeo stack
- `xarray` data structures
- Reading and writing netCDF files
- Plotting with `matplotlib` and `cartopy`

Environment Setup (15 mins)

Basics (60 mins)

15 min Break

Intermediate (60 mins)

- Spatial operations in `xgcm`
- Vertical interpolation in `xgcm`
- ESMPy regridding/remapping
- Data access using Siphon
• xarray doesn’t implicitly understand GCM grids

• xgcm wraps xarray to add an understanding of grid topology

• Implements spatial derivative operators

• Understands only C-grids for now, but other works are in progress

• Grid-aware vertical interpolation

https://xgcm.readthedocs.io/en/latest/grids.html
A collection of Python utilities for downloading data from remote data services

Temperature forecast (°F) for 14 August 2018 21:00Z

https://matplotlib.org/gallery/

https://scitools.org.uk/cartopy/docs/latest/gallery/index.html
Some great Python modules to go along with xarray and dask!
pop-tools  

Tools to support analysis of POP2-CESM model solutions with xarray

- Wraps xgcm to provide support for POP2 grids.
- Inherits spatial derivative operators from xgcm
- Support for POP2 region masks

xroms

- Wraps xgcm to provide ROMS-specific grid manipulations and functions of interest to oceanographers.
- Developed by Kristen Thyng, Rob Hetland, et al. at TAMU
- Wraps cf-xarray to generalize coordinate and dimension calling.
- Wraps xcmocean to automatically choose colormaps for plotting!

https://github.com/kthyng/xroms
GeoCAT

A collection of Python utilities related to NCL

https://matplotlib.org/gallery/

A collection of Python tools for reading, visualizing, and performing calculations with weather data.

https://unidata.github.io/python-training/gallery/gallery-home/
Key Takeaways

• The Pangeo framework rethinks how we analyze large datasets
  – Resusable software design can help avoid re-writing analysis scripts that has already been developed by community
  – In its developmental stages, and will take a few more years to reach the depth/breadth of existing geoscience tools
  – For newer analysis tools development, consider using Pangeo

• NCO, CDO, Ferret, etc are still extremely handy for specific tasks
Additional Python resources

• Previously offered HPRC short courses
  – Introduction to Python
    • https://hprc.tamu.edu/training/intro_python.html
  – Introduction to Scientific Python
    • https://hprc.tamu.edu/training/intro_scientific_python.html
  – Introduction to Python for MATLAB users
    • https://hprc.tamu.edu/training/python_matlab.html

• NumPy for MATLAB users (Quick reference)
  – http://mathesaurus.sourceforge.net/matlab-numpy.html
Additional resources

• Official Documentation
  – xarray docs
  – xgcm docs

• Ask for help:
  – Use the python-xarray on StackOverflow
  – GitHub Issues for bug reports and feature requests
  – Pangeo forums http://discourse.pangeo.io/
Questions?