CAP
ASSIGNMENT 4, OPTIONAL PART
Fall 2018

Using CUDA with Data from Cyberinfrastructure

Goals

This (non-mandatory) assignment aims at exposing students to a cyber-infrastructure (CI) use case by leveraging their CUDA implementation.

It is proposed to use data from the Ocean Observatories Initiative, which delivers marine observation data through its cyber-infrastructure.

OOI data can be obtained using different mechanisms, including the data portal (https://ooinet.oceanobservatories.org) and a machine-to-machine interface (REST API) for programmatic data access.

You can find more information and video tutorials of the OOI data portal in the following link:

https://oceanobservatories.org/data-portal/

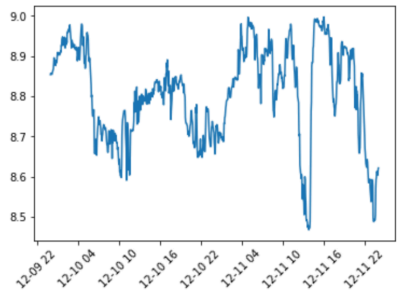
From the large number of sensors and data streams available in OOI it is provided sample data from CTD sensors (measure the conductivity, temperature, and pressure of seawater) at different depth points of the water column, and a BOTPT sensor (Bottom Pressure and Tilt Meter, e.g., used to detect earthquakes and predict tsunami and eruptions) at Axial Seamount, a submarine volcano located on the Juan de Fuca Ridge off the coast of Oregon.

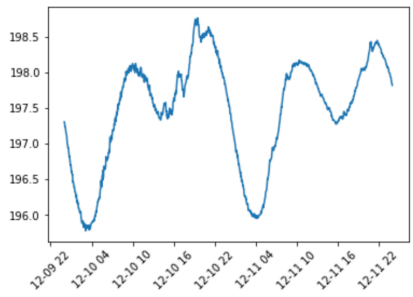
CTD sensor at 194m depth	Water temperature (file 194m_temp.txt) Water pressure (file 194m_pres.txt)
CTD sensor at 581m depth	Water temperature (file 581m_temp.txt) Water pressure (file 581m_pres.txt)
CTD sensor at 1542m depth	water temperature (file 1542m_temp.txt) Water pressure (file 1542m_pres.txt)
CTD sensor at 2903m depth	Water temperature (file 2903m_temp.txt) Water pressure (file 2903m_pres.txt)
BOTPT sensor at Axial Seamount	Water pressure (file botpt_pres.txt)

The python script provided (sample.py) is an example for obtaining the sample data. Please note that you will need to replace YOUR_API_USERNAME and YOUR_TOKEN with your user information in the OOI data portal (you can use a google account to log in).

The python script does also provide simple plotting capabilities as shown below.

```
#CTD 194m

ref_designator = 'CE04OSPS/PC01B/4A-CTDPFA109' #subsite/node/sensor
method = 'streamed'
stream = 'ctdpf_optode_sample'

req = requests.get(prefix_data_query+ref_designator+'/'+method+'/'+stream+'?beginDT='+begin+'&endDT='+end+'&limit='+lim
data = req.json()

df = pd.DataFrame(data)
ntp_diff = (datetime.datetime(1970, 1, 1) - datetime.datetime(1900, 1, 1)).total_seconds()
df['time'] = pd.to_datetime(df.time - ntp_diff, unit='s')
df = df.set_index(df.time)
plt.plot(df['time'], df['seawater_temperature'])
locs, labels = plt.xticks()
plt.setp(labels, rotation=45)
plt.show()

plt.plot(df['time'], df['seawater_pressure'])
locs, labels = plt.xticks()
plt.setp(labels, rotation=45)
plt.show()

#df['seawater_temperature'].to_csv('194m_temp.txt', index=None,)
#df['seawater_pressure'].to_csv('194m_pres.txt', index=None,)
```





The query requests 1,000 data points between representing the measurements between Dec 10 and Dec 12, 2018 (data is decimated or down-sampled).

You will need to use your environment to run the provided python code. For example, you can use the anaconda framework (multi-platform) - https://anaconda.org

Proposed Optional Exercises

1. Use your CUDA code using one specific data set and explore the effect of the moving average on the data when it is plotted. You can use different window sizes (e.g., 10, 15, 20 previous points for the moving average).

2. Use your CUDA code over the provided data (e.g., pressure) and describe the results. Please note that OOI provides thousands of data streams that could be potentially processed in parallel.

3. Explore the OOI data portal and OOI website (oceanobservatories.org) to obtain other data sets and analyze them with your CUDA code.

Survey

Please rate (1-5) the following questions in a table as shown below.

1. How much did you know about CI before this course?

2. How much do you know about CI after introducing this topic in this course?

3. Would you like to know more about CI?

4. Do you find useful to use CI in the context of this course?

5. Do you think adding more CI-related topics and activities would improve this course?

6. Would you like to have an assignment/assignments using HPC/CI-related problems (e.g., MPI using python)?

7. Do you have experience using python, data and scientific libraries (e.g., pandas, matplotlib, etc.), GitHub, Jupyter Notebooks, etc.?

8. How positively (5) or negatively (1) do you think including more CI-related topics and related tools (items listed in the previous question) would improve your skills and/or professional competences?

9. How useful do you think including CI-related topics in the textbook materials would be?

10. How useful do you think including formal or informal videos including demos/tutorials and use case scenarios would be in this course?

Question #	1	2	3	4	5	6	7	8	9	10
Rating (1-5)										

Format

Please provide a PDF document including all materials.

Submission Date

As part of the 4th Assignment due on December 27

Via direct email to the instructor no later than January 2