https://ntrs.nasa.gov/search.jsp?R=20180003369 2019-08-31T15:47:12+00:00Z

GES DISC Data Recipes in Jupyter Notebooks

ED23D-0324

NASA/Goddard EARTH SCIENCES DATA and INFORMATION SERVICES CENTER (GES DISC)

Ben Banavige[1,2], Karthik Garimella[1,3], Justin Rice[1], Angela Li[1], Suhung Shen[1,4], Zhong Liu[1,4]

Abstract

The Earth Science Data and Information System (ESDIS) Project manages twelve Distributed Active Archive Centers (DAACs) which are geographically dispersed across the United States. The DAACs are responsible for ingesting, processing, archiving, and distributing Earth science data produced from various sources (satellites, aircraft, field measurements, etc.). In response to projections of an exponential increase in data production, there has been a recent effort to prototype various DAAC activities in the cloud computing environment. This, in turn, led to the creation of an initiative, called the Cloud Analysis Toolkit to Enable Earth Science (CATEES), to develop a Python software package in order to transition Earth science data processing to the cloud. This project, in particular, supports CATEES and has two primary goals. One, to transition data recipes created by the Goddard Earth Science Data and Information Service Center (GES DISC) into an interactive and educational environment using Jupyter Notebooks. Two, to acclimate Earth scientists to cloud computing.

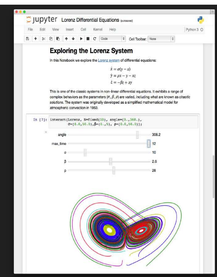
To accomplish these goals, we create Jupyter Notebooks to compartmentalize the different steps of data analysis and help users obtain and parse data from the command line. We also develop a Docker container, comprised of Jupyter Notebooks, Python dependencies, and command line tools, and configure it into an easy-to-deploy package. The end result is an end-to-end product that simulates the use case of end users working in the cloud computing environment.

GES DISC Data Recipes

- Static web tutorials to instruct users:
  - Request data from DAACs
  - Read data
  - Plot data

Jupyter Notebook

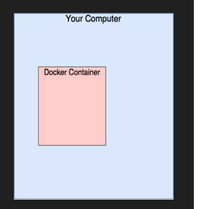
- Run Python Scripts (2.x or 3.x)
- Format Markdown Text
- Visual Python Environment



python jupyter

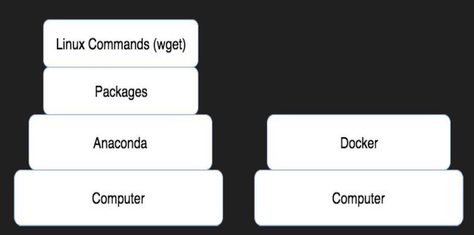
Docker

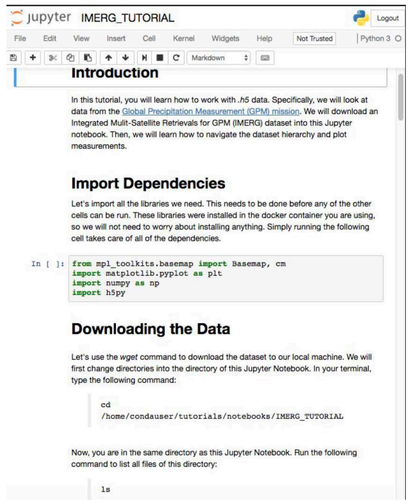
- Lightweight Virtual Machine
- (Almost) No Cross-Platform Compatibility Issues
- Simple to Install/Run
- No Dependency Problems



docker

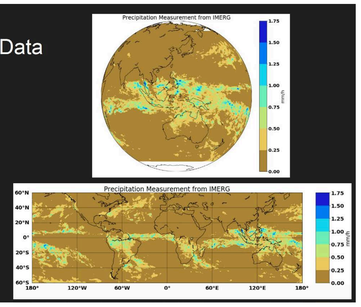
Without Docker / With Docker





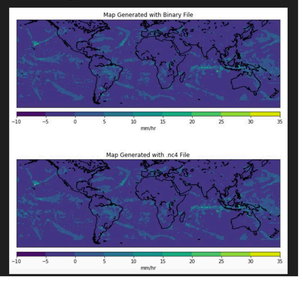
Reading IMERG Data

- Objectives
  - wget
  - h5py
  - HDF5 Files
  - basemap projections



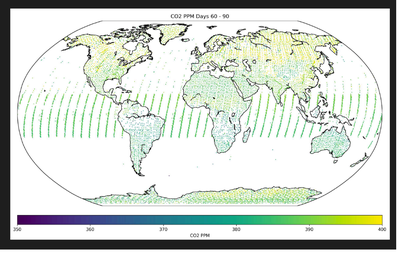
Reading Level-3 TRMM Binary Data

- Objectives
  - wget
  - basemap
  - Parsing binary data
  - Comparing netCDF4 and Binary



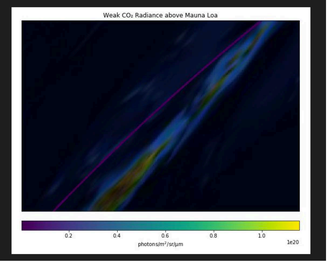
ACOS Data Handling

- Objectives
  - wget
  - netCDF4
  - Basemap



Spatial and Variable Subsets of L1B OCO-2 Data

- Objectives
  - wget
  - OPeNDAP
  - pydap server
  - Indexing Data
  - Indexing URL



Combining Jupyter and Docker

- Allows Distribution of Jupyter Notebooks
- Solves Dependency Issues
- Allows for Interactive Learning/Experimentation
- Provides an effective way to bring analysis closer to data when all DAACs have their data in the cloud
- Github: https://github.com/karthenjamin/data_recipes

Authors

[1] NASA Goddard Space Flight Center
[2] Harvard University, College of Arts and Sciences, Dept. of Statistics
[3] ORNL DAAC
[4] George Mason University