Using Python and H5Py with Freeze/Thaw HDF5 Granules

Download H5Py

Download location: http://code.google.com/p/h5py/downloads/list

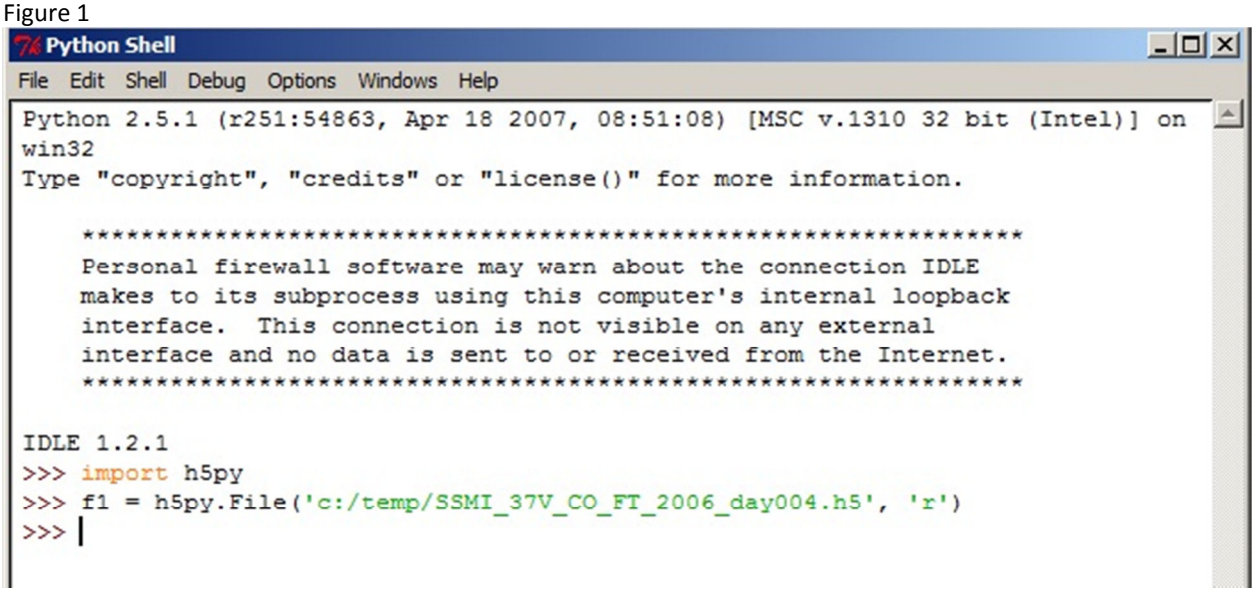
Windows Users:
Download the appropriate Windows Installer Package

Linux Users:
Use the appropriate UNIX Package

Opening and using Freeze/Thaw HDF5 Granules with H5Py

1) Open your preferred Python IDE (Integrated Development Environment) and load the F/T Granule. An example is displayed in Figure 1.

Figure 1



2) After the F/T Granule is open, you can look at the all of the groups within the F/T granule using the keys() function. Groups are the containers of the HDF5 file that hold associated datasets. The F/T Granule contains two groups: DATA and METADATA. Once all of the HDF5 groups are listed using the keys() function you can now view all of the datasets within each group. The F/T dataset and associated datasets are located within the DATA group. You can list all of the datasets within the group using the same keys() function mentioned above. Each F/T Granule contains three datasets: FT_SSMI, cell_lat and cell_long. The FT_SSMI dataset is the F/T data while the cell_lat and cell_long datasets are the geographical location of each pixel.

If you want to view the global F/T dataset you can do this by opening the FT_SSMI dataset. Use can use the print command (*print dataset* in Figure #2) to see the number of rows and columns with the F/T dataset and well as the data type. You can also view the contents of the dataset by printing an array dataset[0:,0:]. With this knowledge one can now begin to query data using array functionality. The most recommended Python Array Package is NumPy (http://new.scipy.org/download.html). A step by step example is displayed in Figure 2.

Figure 2

```
IDLE 1.2.1
>>> import h5py
>>> f1 = h5py.File('c:/temp/SSMI_37V_CO_FT_2006_day004.h5', 'r')
>>> f1.keys()
['DATA', 'METADATA']
>>> group = f1['DATA']
>>> group.keys()
['FT_SSMI', 'cell_lat', 'cell_lon']
>>> dataset = group['FT_SSMI']
>>> print dataset
<HDF5 dataset "FT_SSMI": shape (586, 1383), type "|u1">
>>> print dataset[0:,0:]
[[255 255 255 ..., 255 255 255]
 [255 255 255 ..., 255 255 255]
 [255 255 255 ..., 255 255 255]
 ...,
 [255 255 255 ..., 255 255 255]
 [255 255 255 ..., 255 255 255]
 [255 255 255 ..., 255 255 255]]
>>>
```

Viewing Freeze/Thaw HDF5 Metadata using H5Py

1) Each F/T Granule contains collection level metadata and the dataset level metadata. The metadata within the F/T Granule uses the NetCDF Climate and Forecast (CF) Metadata Conventions. You can view all of the F/T Granule metadata using H5Py. Once the HDF5 file is open you can select the METADATA group instead of the DATA group. Once the METADATA group is selected you can select the metadata using the attrs() function. If you want just view the metadata fields you can use the keys() function. If you want to view the metadata fields and value associated with that field you can use the items() function. The example in Figure 3 allows you to view the collection level metadata. If you want to view the dataset metadata follow the same steps except select the appropriate dataset. A step by step example is displayed in Figure 3.

Figure 3

```
IDLE 1.2.1
>>> import h5py
>>> f1 = h5py.File('c:/temp/SSMI_37V_CO_FT_2006_day004.h5', 'r')
>>> group = f1['METADATA']
>>> metadata = group.attrs
>>> listcollectionlevelmetadata = metadata.items()
>>> print listcollectionlevelmetadata
[('PRODUCT_ESDR_VERSION', 'v1.0 2006.004.00:00:01'), ('title', 'An ESDR fo
r Land Surface Freeze/Thaw State (NASA Measures)'), ('history', 'F/T ESDR
created at University of Montana 2009-2010'), ('institution', 'University
of Montana, Flathead Lake Biological Station -- F/T team'), ('source', 'Un
iversity of Montana, under NASA Contract No. NNX08AQ63A'), ('references',
'Developing a Global Data Record of Daily Landscape Freeze/Thaw Status usi
ng Satellite Microwave Remote sensing. IEEE Transactions on Geoscience and
Remote Sensing. Kim,Y.,J.S. Kimball, K.C. MacDonald, and J.Glassy.(pending
)2010'), ('comment', 'This NASA ESDR F/T Stage I release from SSMI data is
formatted in HDF5 v1.8.x'), ('Producer', 'NTSG University of Montana SMAP
Team'), ('PointOfContact', 'NSIDC'), ('Units', 'F/T status classification
code (dimensionless index)'), ('TimeResolution', 'daily'), ('PixelResoluti
on', '25km'), ('GranuleName', 'SSMI_37V_CO_FT_2006_day004'), ('Overpass co
de AM', 'Morning overpass   '), ('Overpass code PM', 'Afternoon overpass '
), ('Overpass code CO', 'Combined overpasses'), ('ProjectID', 'NASA Measur
es Freeze/Thaw ESDR Project'), ('Instrument', 'SSM/I'), ('Channel', '37'),
('Polarization', 'V'), ('Overpass_Code', 'CO'), ('Rank', 2L), ('Y_DIM', 58
6L), ('X_DIM', 1383L), ('Total_Pixels', 810438L), ('N_Channels', 1L), ('Ma
pProjection', 'EASE-GRID Polar-Cylindrical 25km'), ('Year', '2006'), ('Yea
rday', '004'), ('Sponsor', 'NASA Measures'), ('GEOGRAPHIC_LATLON_NW', arra
```

Python Packages mentioned in this Tutorial

H5Py – http://code.google.com/p/h5py/
Python interface to the Hierarchical Data Format Library Version 5

NumPy – http://numpy.scipy.org/
Python scientific computing package

Useful Documentation

H5Py – http://h5py.alfven.org/docs/guide/quick.html
Quick Start Guide

NumPy – http://docs.scipy.org/doc/numpy/reference/
Users Guide

Contact

joe.glassy@ntsg.umt.edu