HDF Assignment part 1

You will turn in all of the steps that begin with a number. There are 7 of these.

We'll start off by getting hands-on experience with HDF.

- download HDFView and run through the tutorial.
    - Click here for the installation page: http://www.hdfgroup.org/hdf-java-html/hdfview/
    - Work through the tutorial at http://www.hdfgroup.org/HDF5/Tutor/hdfview.html.
    - You can skip over the parts of the tutorial concerning "h5dump."
    - **(1)** Send me the file that you created in the tutorial.
- Download Weather Data from NASA.
    - Go to the MERRA MDISC Data Subset, http://disc.sci.gsfc.nasa.gov/daac-bin/FTPSubset.pl
    - Select the following data subset
        - IAU 3d assimilated state on pressure (inst3_3d_asm_Cp)
        - 1979, Jan 1, one day only
        - Select only Surface Pressure and Air Temperature
        - Leave the "Time Subset" setting to default
        - Select "1 (1000 hPA)" for the Verticle Level Subset
        - Output to HDF
    - **(2)**What kind of data formats are available to output to?
    - Open the downloaded file in HDFView
        - **(3)** What is the surface pressure at hour 12, latitude 48.125, longitude 158.125 ?
- Download an example dataset "test.h5" available on the class unit website
    - download "test.h5" provided on the website
    - **(4)** Define a "hyperslab" in HDF. You may want to look up the definition.
    - **(5)** Note that the database "Utility" is a 6-dimensional array. Ask yourself how HDFView represents such arrays. Your job is to find the correct hyperslab that will allow you to see a specific database entry. What is the value of utility stored at the location (0,1,2,3,4,5) of the database? Hint: Open the dataset "utility" by right-clicking and selecting "open as." Here you can specify the hyperslab you wish to view. Select the dimensions you wish to vie and press "dims..." to select the location of the slice. Note that you will likely see a table of zeros when you first open the file. As you browse among the databases dimensions, you will find non-zero values close to the point you are searching for.

Now, we will learn about the importance of HDF and similar technologies.

- skim this explanation, "Why HDF?" http://www.hdfgroup.org/why_hdf/
    - **(6)** What is the difference between HDF (or related technologies) an relational databases? List two that are mentioned in this article.
- Read a little about what users are saying about HDF.

- Read the discussion posted online:
  http://stackoverflow.com/questions/4871670/experience-with-using-h5py-to-do-analytical-work-on-big-data-in-python
- **(7)** What problem did this user run into that led to a need for HDF?

Additional Note:

"Traditional methods for handling scientific data include the use of flat sequential files of machine- specific binary data or portable ASCII data. For applications such as visualization, such methods are inefficient in storage, access, and ease-of-use. Relational database systems fail to accommodate the multidimensional or hierarchical structures often found in scientific data sets. In addition, relational systems do not provide adequate performance for the size, complexity, and type of access required for many scientific data sets."

p 308   COMPUTERS IN PHYSICS, VOL 7, NO. 3, MAY/JUN 1993, http://www-d0.fnal.gov/software/languages/cip_port.txt