RFC: Adding ability to do hyperslab selection to hdp

Binh-Minh Ribler

DRAFT

This RFC is to get suggestions and decisions for the design of adding the ability to do hyperslab selection to hdp.

# Introduction

The HDF4 tool hdp currently reads and displays the entire data set. This RFC proposes an additional feature to hdp so that it will allow users to read data set in hyperslab selections. At this stage, this RFC is seeking for opinions on what the new feature should look like and what we are willing to provide, hence, there are decisions we need to make on restrictions.

# Tentative Design

We start with the new flag -***y*** (for h***y***perslab) to indicate that hyperslab selection is coming, then additional information needs to be provided for the start, stride, and length of each dimension. Section 2.1 lists a few options to provide those information and Section 2.2

## Option Selection

There are a few options that I can think of. They are not in any particular order, although the last option is my least favorite because it doesn’t start with -y. In the above examples, -i 0 means selecting data set with index 0.

Option 1: -y then s for starts, t for strides, or e for edges

 *Example:* hdp dumpsds -i 0 -ys(0,0,0) -yt(2,2,2) -ye(5,5,5) filename.hdf

Option 2: -y then start(…), stride(…), or edge(…)

 *Example:* hdp dumpsds -i 0 -ystart(0,0,0) -ystride(2,2,2) -yedge(5,5,5) filename.hdf

Option 3: -y, colon (:), then start(…), stride(…), edge(…), separated by commas

 *Example:* hdp dumpsds -i 0 -y: start(0,0,0),stride(2,2,2),edge(5,5,5) filename.hdf

Option 4: is an exception; specify starts/strides/edges without -y, like this: -start(…) -stride(…) -edge(…)

 *Example:* hdp dumpsds -i 0 -start(0,0,0) -stride(2,2,2) -edge(5,5,5) filename.hdf

## Notes and Decisions Needed

There are several questions that we need to weigh in while considering the above options and any additional ones that anyone may suggest.

1. When hyperslab selection is requested,
	1. Do the starts, strides, edges apply to all selected data sets?
	2. Or, should we restrict the hyperslab selection to one data set per run?
	3. Or, should we allow specifying hyperslab selection to individual data sets in one run. For example, the following command would specify two separate sets of starts/strides/edges for two data sets, index 0 and index 2, assuming we choose to implement option 1 in the previous section.

hdp dumpsds **-i** 0 **-ys(**1,1,1**) -yt(**2,2,2**) -ye(**5,5,5**) -i** 2 **-ys(**10,10**) -yt(**2,2**) -ye(**5,5**)** filename.hdf

 Obviously, this is a more complicated option than simply allowing only one data set per run or one set of starts/strides/edges for all the selected data sets, in which case the selected data sets must have the same number of dimensions. Which scenario is the most practical in user applications?

* 1. Do we allow mixed selection? For example, the following command would select two data sets, index 0 and index 2, but one of them is read by hyperslab, the other is not.

hdp dumpsds **-i** 0 **-ys(**0,0,0**) -yt(**2,2,2**) -ye(**5,5,5**) -i** 2 filename.hdf

1. Should we allow specifying hyperslab selection across multiple files?

For example, the following command would specify that all the data sets in the 3 files are read with the specified hyperslab selection.

hdp dumpsds **-ys(**10,10**) -yt(**2,2**) -ye(**5,5**)** file1.hdf file2.hdf file3.hdf

1. Intensive error checking will need to be done on hyperslab selection for each selected data set. If error is detected, do we want to continue on to the rest of the selected data sets or exit at that point? Does it make sense to continue?
2. Please add any other points that are missed here.