

RFC: Integrating attribute feature with h5import

Jonathan Kim

This RFC proposes to integrate a feature generating attribute(s) with h5import tool.

1 Introduction

A feature to generate attribute(s) via a tool was requested by NPOESS project.

This feature can be integrated into h5import tool and this RFC is proposing some user interface details.

The 'h5import' tool currently can create or add dataset(s) to a HDF5 file with limited data types. The supported data types are 'Integer', 'Floating point' and 'String'. Other new data types are expected to be supported as necessary. (ex: reference type)

Since a dataset and an attribute in HDF5 are very similar to each other, the 'h5import' tool can be used to generate attribute(s) to object(s) in the similar manner. So users who are already familiar with h5import tool will be able to easily adopt to generate attribute(s) by using the same h5import tool.

For the implementation, certain code in the h5import can be reused along with the current framework. Since dataset and attribute is similar to each other, future additional code for other data types can be shared between dataset and attribute.

2 Suggestion details

2.1 Configuration directives

The h5import tool uses configuration file as input from command line to figure out generating a dataset. Portion of the current directives can be reused for generating an attribute in the same manner.

2.1.1 Current directives for creating 'Dataset' from h5import

- PATH
- INPUT-CLASS
- INPUT-SIZE
- RANK
- DIMENSION-SIZES
- OUTPUT-CLASS
- OUTPUT-SIZE
- OUTPUT-ARCHITECTURE
- OUTPUT-BYTE-ORDER

- CHUNKED-DIMENSION-SIZES
- MAXIMUM-DIMENSIONS
- COMPRESSION-TYPE
- COMPRESSION-PARAM
- EXTERNAL-STORAGE

2.1.2 Directives which can be reused for creating 'Attribute'

- PATH
 - A path to an attribute. This path should contain valid path to an object.
 - Ex: /group1/dset100/attr_test
 - /group1/dset100 must be an existing object in a given HDF5 file.
- INPUT-CLASS
- INPUT-SIZE
- RANK
- DIMENSION-SIZES
- OUTPUT-CLASS
- OUTPUT-SIZE
- OUTPUT-ARCHITECTURE
- OUTPUT-BYTE-ORDER

Note: directives not supported for attribute

- CHUNKED-DIMENSION-SIZES
- MAXIMUM-DIMENSIONS
- COMPRESSION-TYPE
- COMPRESSION-PARAM
- EXTERNAL-STORAGE

2.2 Distinguish between dataset and attribute

There are two suggestions here.

2.2.1 Suggestion1 – By a new directive 'TOPIC'

'TOPIC' can be specified in 'configuration file' of h5import tool.

Examples:

- TOPIC DATASET
 - Generating dataset
- TOPIC ATTRIBUTE
 - Generating attribute

2.2.1.1 Reasoning

- Creating dataset and adding attribute can be done combinational manner at once.
- A user can easily distinguish for the purpose of a configuration file.
- Command line interface can stay same as now.

2.2.1.2 Note

It would be better to enforce using this new directive for the previous customer. However if it cause too much issue, default action can be generating dataset without the 'TOPIC' directive.

2.2.2 Suggestion2 – By command line arguments

Examples:

- --datasets
 - Generating dataset only
- --attributes
 - Generating attribute only

Command line examples:

- h5import **--attributes** test.txt -c test.conf -o dsets.h5
 - Adding an attribute as specifying with test.txt and test.conf to an object existing in 'dsets.h5' file.

2.2.2.1 Requirement

- Generating dataset and attribute must be run separately due to issue for distinguishing the config files when specified together at the same time.
- For this case, It will need to support comment feature for handling configuration file, so each config file can keep a comment line describing what it's for.

2.2.2.2 Note

It would be better to enforce using this new argument for the previous customer. However if it cause too much issue, default action can be generating dataset without the '--datasets'.

Revision History

April 22, 2010: Version 1 draft for initial review.