RFC: h5diff Attribute Comparisons

Jonathan Kim

Comparing attributes result in incorrect or undesired output in some cases. This RFC explains the issues and suggestions for solutions.

# Issues

h5diff currently compares attributes correctly only when two objects have the same number of attributes and attribute names are identical by creation order.  
  
The current implementation creates two problems when the above conditions are not met:

* In some cases, h5diff displays no information about attribute differences.
* In other cases, attributes are compared not by matching the names but by creation order.  This can yield incorrect or undesired output.

## Case details

### Case1: Same number of attribute & same names

* Result output: Showed the differences and exit code 1. Works correctly (OK).

### Case2: Same number of attribute & all different names

* Result output: Didn't show appropriate output
  + With no option:

Nothing

* + With -v or -r:

dataset: </d1> and </d2>

0 differences found

* Exit code: 2
* Expected Fix: Exit code should be 1 as no error occurred but some attributes are different. Also need to come up with displaying what's not compared.

### Case3: Different number of attributes & some same names

* Assumptions:
  + /dset1 has attributes ‘attr1’, ‘attr2’ and ‘aa’.
  + /dset2 has attributes attr1, attr2, bb and cc”.
  + The ‘attr1’ and ‘attr2’ has different contents.

* Result output:

dataset: </d1> and </d2>

1 differences found

* Exit code: 1
* Expected Fix: the output should show more details for the attribute differences

### Case4: Different number of attributes & all different names

* Result output:

dataset: </d1> and </d2>

1 differences found

* Exit code: 1
* Expected Fix: the output should show more details for the attribute differences

# Suggestions

## Add two new verbose options, -v1 (--verbose=1) and -v2 (--verbose=2), which will include detailed attribute information.  This framework will provide extensibility for future unknown output addition. See the "Output examples" section below.

## Compare attributes by matching names not by creation order, which is what h5diff does now when comparing datasets or objects in a group or file.

# Output examples

Assumption:

* object1 '/d1' has attributes named 'attr1', 'attr2', 'attr3' and 'attr5'.
* object2 '/d2' has attributes named 'attr1', 'attr2', and 'attr4'.

Example 1, with "**-v1**":  
  
$ h5diff -v1 attrs1.h5 attrs1.h5 /d1 /d2  
  
dataset: </d1> and </d2>  
attribute: <attr1 of </d1>> and <attr1 of </d2>>  
size:       [2]          [2]  
position       attr1 of </d1>  attr1 of </d2>  difference         
------------------------------------------------------------  
[ 1 ]       1              2                1                 
1 differences found  
attribute: <attr2 of </d1>> and <attr2 of </d2>>  
size:       [1]          [1]  
position       attr2 of </d1>  attr2 of </d2>  difference         
------------------------------------------------------------  
[ 0 ]       1              2  
[ 0 ]       c              d  
2 differences found  
**Attributes status:  2 common, 2 only in obj1,  1 only in obj2**  
3 differences found

Example 2, with "**-v2**" :  
  
$ h5diff -v2 attrs1.h5 attrs1.h5 /d1 /d2  
  
dataset: </d1> and </d2>  
**obj1   obj2  
---------------------------------------  
 x     x    attr1            
 x     x    attr2        
 x          attr3  
       x    attr4  
 x          attr5**  
  
attribute: <attr1 of </d1>> and <attr1 of </d2>>  
size:       [2]             [2]  
position       attr1 of </d1>  attr1 of </d2>  difference         
------------------------------------------------------------  
[ 1 ]       1              2                1                 
1 differences found  
attribute: <attr2 of </d1>> and <attr2 of </d2>>  
size:       [1]          [1]  
position       attr2 of </d1>  attr2 of </d2>  difference         
------------------------------------------------------------  
[ 0 ]        1            2  
[ 0 ]        c            d  
2 differences found  
**Attributes status:  2 common, 2 only in obj1,  1 only in obj2**  
3 differences found

A double 'x' means the attribute is common to both objects.  A single 'x' means the attribute exists only on one of the objects being compared.  (This matches current h5diff output when comparing objects in files or groups.)  
  
If no number is specified with '-v', h5diff will display the same output as is currently displayed to maintain backward compatibility of output.  (‘-v0’ will be same as ‘-v’.)  
  
It is not yet determined whether the attribute status information will appear at the top or bottom of the object section.

# Exit code

The exit code will stay same as showing on h5diff help page currently.  
  0 - if no differences  
  1 - if differences found  
  2 - if error

# Acknowledgements

This work was supported by a customer of The HDF Group.

# Revision History

|  |  |
| --- | --- |
| *Jan 18, 2011:* | Version 1 circulated for feedback within tool team. |
| Jan 20, 2011: | Version 2 incorporates feedback from customer, especially in –v interface. |
| Jan 24, 2011: | Version 3 incorporates feedback from customer, especially in ‘exit code’. |
| Feb 1, 2011: | Version 4 circulated for internal review. |
| Feb 8, 2011: | Version 5 incorporates feedback from internal review, especially in section ‘1.1 case details’ to give some more details and 2.2 commenting about h5diff compares objects by matching name, which is the expected behavior as reference. |
|  |  |
|  |  |