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 the solution.

Issues

h5diff currently compares attributes correctly only when two objects have the same number of attributes and attribute names are identical.

The current implementation creates two problems when the above conditions are not met:

- In some cases, h5diff displays lack of information about attribute differences.

- In other cases, h5diff returns incorrect exit code and lack of information about the attribute differences. This also results in incorrect or undesired output.

Case details

Case1: If obj1 and obj2 have a different number of attributes

- **Result**: h5diff returns exit code 1 (signifying a difference found) but does not report anything to stdout or stderr, even with -r, -c or -v.

- **Expected Fix**: the output should show more details for the attribute differences

Case2: If obj1 and obj2 have the same number of attributes, but they do not match up (by name),

- **Result**: h5diff will report the differences (or lack thereof) of the attributes in obj1 that match up to those in obj2, but only the attributes that come before the earliest created attribute in obj1 that is not in obj2. For example, if obj1 contains attributes A, B and C, and obj2 contains A, C, and D, h5diff will examine and report only attribute A. After doing this, h5diff will continue examining the rest of the file but will return 2 (indicating an error occurred). h5diff will iterate in the order the attributes were created in obj1.

- **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: if obj1 and obj2 contain the same attributes (same number and names)

- **Result**: h5diff will examine and report differences (or lack thereof) in a way similar to how it works for datasets. h5diff will iterate in the order the attributes were created in obj1.

- **Expected Fix**: work correctly. no fix need.

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.

Fix the incorrect exit code from '2' to '1' for the above 'case2'. (2 means error, 1 means different)

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>
Attributes status:  2 common, 2 only in obj1,  1 only in obj2
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

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
Attributes status:  2 common, 2 only in obj1,  1 only in obj2
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
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

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.
Feb 9,2011:	Version 6 incorporates feedback from internal review, especially in section '1.1 case details' updated for more details and also updated 2.2.
March 28,2011:	Version 7 Final updates for section 3. Output examples. Changed the location of simple attribute status line.