RFC: Supporting soft-link and external-link for h5diff

Jonathan Kim (jkm@hdfgroup.org)

This RFC is for discussing about supporting soft-link and external-link for h5diff as our Chicago customer requested for the feature in h5diff command line tool.

# Introduction

Currently the h5diff command like tool doesn’t have the ability to compare groups and datasets that are created through soft-link or external-link. The current h5diff’s behavior will be shown with the result output from test for both soft-link and external-link and will talk about what should be expected for correct behavior. Also discussion on user interface could be helpful.

# Motivation and Goal

* Add soft-link and external-link feature in h5diff command line tool as the Chicago customer requested.

# Current Behavior

## Test cases were created to demonstrate current behavior. There are 3 HDF5 data files were used. One file is for soft-link tests and the other two files are for external-link tests. The output from h5dump for each file is shown below.

* Soft-link test file contents:

HDF5 "soft\_link.h5" {

GROUP "/" {

SOFTLINK "softlink\_dset1\_1" {

LINKTARGET "target\_dset1"

}

SOFTLINK "softlink\_dset1\_2" {

LINKTARGET "target\_dset1"

}

SOFTLINK "softlink\_dset2" {

LINKTARGET "target\_dset2"

}

SOFTLINK "softlink\_group1" {

LINKTARGET "target\_group"

}

SOFTLINK "softlink\_group2" {

LINKTARGET "target\_group"

}

SOFTLINK "softlink\_noexist" {

LINKTARGET "no\_obj"

}

DATASET "target\_dset1" {

DATATYPE H5T\_STD\_I32LE

DATASPACE SIMPLE { ( 2, 4 ) / ( 2, 4 ) }

DATA {

(0,0): 0, 1, 2, 3,

(1,0): 1, 2, 3, 4

}

}

DATASET "target\_dset2" {

DATATYPE H5T\_STD\_I32LE

DATASPACE SIMPLE { ( 2, 4 ) / ( 2, 4 ) }

DATA {

(0,0): 0, 0, 0, 0,

(1,0): 0, 0, 0, 0

}

}

GROUP "target\_group" {

DATASET "dset" {

DATATYPE H5T\_STD\_I32LE

DATASPACE SIMPLE { ( 2, 4 ) / ( 2, 4 ) }

DATA {

(0,0): 0, 1, 2, 3,

(1,0): 1, 2, 3, 4

}

}

}

}

}

* External-link source file contents:

HDF5 "extlink\_source.h5" {

GROUP "/" {

EXTERNAL\_LINK "ext\_link\_dset1" {

TARGETFILE "extlink\_target.h5"

TARGETPATH "/target\_group/x\_dset"

DATASET "/target\_group/x\_dset" {

DATATYPE H5T\_STD\_I32LE

DATASPACE SIMPLE { ( 2, 4 ) / ( 2, 4 ) }

DATA {

(0,0): 0, 1, 2, 3,

(1,0): 1, 2, 3, 4

}

}

}

EXTERNAL\_LINK "ext\_link\_dset2" {

TARGETFILE "extlink\_target.h5"

TARGETPATH "/target\_group2/x\_dset"

DATASET "/target\_group2/x\_dset" {

DATATYPE H5T\_STD\_I32LE

DATASPACE SIMPLE { ( 2, 4 ) / ( 2, 4 ) }

DATA {

(0,0): 0, 0, 0, 0,

(1,0): 0, 0, 0, 0

}

}

}

EXTERNAL\_LINK "ext\_link\_grp1" {

TARGETFILE "extlink\_target.h5"

TARGETPATH "target\_group"

GROUP "target\_group" {

DATASET "x\_dset" {

HARDLINK "/target\_group/x\_dset"

}

}

}

EXTERNAL\_LINK "ext\_link\_grp2" {

TARGETFILE "extlink\_target.h5"

TARGETPATH "target\_group2"

GROUP "target\_group2" {

DATASET "x\_dset" {

HARDLINK "/target\_group2/x\_dset"

}

}

}

EXTERNAL\_LINK "ext\_link\_noexist1" {

TARGETFILE "extlink\_target.h5"

TARGETPATH "no\_obj"

}

EXTERNAL\_LINK "ext\_link\_noexist2" {

TARGETFILE "no\_file.h5"

TARGETPATH "no\_obj"

}

}

}

* External-link target file contents:

HDF5 "extlink\_target.h5" {

GROUP "/" {

DATASET "target\_dset1" {

DATATYPE H5T\_STD\_I32LE

DATASPACE SIMPLE { ( 2, 4 ) / ( 2, 4 ) }

DATA {

(0,0): 0, 1, 2, 3,

(1,0): 1, 2, 3, 4

}

}

GROUP "target\_group" {

DATASET "x\_dset" {

DATATYPE H5T\_STD\_I32LE

DATASPACE SIMPLE { ( 2, 4 ) / ( 2, 4 ) }

DATA {

(0,0): 0, 1, 2, 3,

(1,0): 1, 2, 3, 4

}

}

}

GROUP "target\_group2" {

DATASET "x\_dset" {

DATATYPE H5T\_STD\_I32LE

DATASPACE SIMPLE { ( 2, 4 ) / ( 2, 4 ) }

DATA {

(0,0): 0, 0, 0, 0,

(1,0): 0, 0, 0, 0

}

}

}

}

}

## Test results

* Soft-link test results:

===============================================================

# Case: group (soft linked) vs group (target)

CMD> h5diff -v soft\_link.h5 soft\_link.h5 /softlink\_group1 /target\_group

</softlink\_group1> is of type H5G\_LINK and </target\_group> is of type H5G\_GROUP

--------------------------------

Some objects are not comparable

--------------------------------

Use -c for a list of objects.

===============================================================

# Case: dset (soft linked) vs dset (target)

CMD> h5diff -v soft\_link.h5 soft\_link.h5 /softlink\_dset1\_1 /target\_dset1

</softlink\_dset1\_1> is of type H5G\_LINK and </target\_dset1> is of type H5G\_DATASET

--------------------------------

Some objects are not comparable

--------------------------------

Use -c for a list of objects.

===============================================================

# Case: dset (soft linked) vs non-exist object

CMD> h5diff -v soft\_link.h5 soft\_link.h5 /softlink\_dset1\_1 /softlink\_noexist

link : </softlink\_dset1\_1> and </softlink\_noexist>

1 differences found

===============================================================

# Case: group (soft linked and exist) vs group (soft linked and exist)

CMD> h5diff -v soft\_link.h5 soft\_link.h5 /softlink\_group1 /softlink\_group2

link : </softlink\_group1> and </softlink\_group2>

0 differences found

===============================================================

# Case: dset (soft linked and exist) vs dset (soft linked and exist) - same values

CMD> h5diff -v soft\_link.h5 soft\_link.h5 /softlink\_dset1\_1 /softlink\_dset1\_2

link : </softlink\_dset1\_1> and </softlink\_dset1\_2>

0 differences found

===============================================================

# Case: dset (soft linked and exist) vs dset (soft linked and exist) - diff values

CMD> h5diff -v soft\_link.h5 soft\_link.h5 /softlink\_dset1\_1 /softlink\_dset2

link : </softlink\_dset1\_1> and </softlink\_dset2>

1 differences found

* External-link test results

===============================================================

# Case group (ext linked in src file) vs group (in target file)

CMD> h5diff -v extlink\_source.h5 extlink\_target.h5 /ext\_link\_grp1 /target\_group

</ext\_link\_grp1> is of type H5G\_UDLINK and </target\_group> is of type H5G\_GROUP

--------------------------------

Some objects are not comparable

--------------------------------

Use -c for a list of objects.

===============================================================

# Case dset (ext linked in src file) vs dset (in target file)

CMD> h5diff -v extlink\_source.h5 extlink\_target.h5 /ext\_link\_dset1 /target\_group/x\_dset

</ext\_link\_dset1> is of type H5G\_UDLINK and </target\_group/x\_dset> is of type H5G\_DATASET

--------------------------------

Some objects are not comparable

--------------------------------

Use -c for a list of objects.

===============================================================

# Case dset (ext linked) vs object (non-exist / exsiting file)

CMD> h5diff -v extlink\_source.h5 extlink\_target.h5 /ext\_link\_dset1 /ext\_link\_noexist1

Object </ext\_link\_noexist1> could not be found in <extlink\_target.h5>

===============================================================

# Case dset (ext linked) vs object (non-exist / no file)

CMD> h5diff -v extlink\_source.h5 extlink\_target.h5 /ext\_link\_dset1 /ext\_link\_noexist2

Object </ext\_link\_noexist2> could not be found in <extlink\_target.h5>

===============================================================

# Case group (ext linked and exist) vs group (ext linked and exist)

CMD> h5diff -v extlink\_source.h5 extlink\_source.h5 /ext\_link\_grp1 /ext\_link\_grp2

external link: </ext\_link\_grp1> and </ext\_link\_grp2>

1 differences found

===============================================================

# Case dset (ext linked and exist) vs dset (ext linked and exist)

CMD> h5diff -v extlink\_source.h5 extlink\_source.h5 /ext\_link\_dset1 /ext\_link\_dset2

external link: </ext\_link\_dset1> and </ext\_link\_dset2>

1 differences found

# User Interface

Propose to the use of ‘-f’ option flowing by ‘soft’ for the soft-link and ‘ext’ for the external-link. The h5copy command line tool also uses the same user interface to handle soft-link and external-link. This will give the user more a intuitive way to handle the feature among tools. (‘-f’ could mean ‘follow’ in this case.)

Soft-link example:

* h5diff **–f soft** <h5\_file> <h5\_file> /softlink\_to\_group1 /softlink\_to\_group2

External-link example:

* h5diff **–f ext** <h5\_source\_file> <h5\_target\_file> /externlink\_to\_group1 /group1

Without these options, the behavior would stay same as now.

# Acknowledgements

The Chicago customer requested for the feature in h5diff command line tool. Thanks Peter for guideline.

# Revision History

|  |  |
| --- | --- |
| *December 18, 2009:* | Version 1 circulated for comment within The HDF Group. |