# The h5diff’s current behaviors and shortcomings

## Purpose

List h5diff’s current behaviors, so sensible and insensible behaviors can be figured out. With the result, we can decide what to do with current h5diff’s issues and also prevent those issues from the new h5compre tool.

## Background

Many features got implemented into the current h5diff command tool over the past. However some of the feature development procedures failed to address related updates which must go along with the feature (ex: default output or exit code and so on). Also lack of fundamental definitions contributed other insensible behaviors in a narrow scope of development procedures from a big picture.

As a result, the current h5diff provides some inconsistent and insensible behaviors which confuse users to detect differences between HDF5 files.

In deciding developing a new comparing tool ‘h5ocompare’ for the future replacement, this document can be referred for comparing behaviors in sensible manner.

## Description

There are four main sections to show the current h5diff tool’s behaviors. Some section may contain overlapped contents from other section.

1. Default behaviors - main purposed behaviors
2. Optional behaviors - additional or exceptional behaviors
3. How to handle common, extra or not-comparable object and attribute with results - what user would look and feel in a big picture
4. Insensible behaviors - known issues (this can grow as identifying more issues)

All of these will contain example output and exit-code from h5diff.

**Highlights indicate either insensible or inconsistent results at current stage.**

## Example HDF5 files

All the HDF5 files that are used in each section can be found in “**Appendix A**”, so refer to specific files to verify the h5diff’s behaviors from examples in each section.

## Default behaviors

Assume all the objects are comparable object. (ex: same name, type, space)

|  |  |  |
| --- | --- | --- |
|  | **Compare *object* vs *object***(ex: $h5diff file1.h5 file2.h5 /dset1 /dset2) | **Compare *file* vs *file***(ex: $h5diff file1.h5 file2.h5) |
| Dataset : Number | $ **h5diff dset-num1.h5 dset-num2.h5 /dset /dset** dataset: </dset> and </dset>2 differences foundEXIT-CODE : 1 | $ **h5diff dset-num1.h5 dset-num2.h5**dataset: </dset> and </dset>2 differences foundEXIT-CODE : 1 |
| Dataset : String | $ **h5diff dset-str1.h5 dset-str2.h5 /string** dataset: </string> and </string>6 differences foundEXIT-CODE : 1 | $ **h5diff dset-str1.h5 dset-str2.h5**dataset: </string> and </string>6 differences foundEXIT-CODE : 1 |
| more … |  |  |

## Continue (default behaviors)

|  |  |  |
| --- | --- | --- |
|  | **Compare *object* vs *object*** | **Compare *file* vs *file*** |
| TODO* Dataset : array
* Dataset : obj ref
* Dataset : region Ref
* Dataset: compound
* Dataset: enum
* Dataset: opaque
* Dataset: Vlen
* Named Datatype
* Group
* Attribute: number
* Attribute: string
* Attribute: obj ref
* Attribute: region ref
* Soft link
* External link
 | TODO for each | TODO for each |

## Optional behaviors

|  |  |  |
| --- | --- | --- |
|  | **Compare *object* vs *object*** | **Compare *file* vs *file*** |
| -h / --help | Display help page. (Too long to put here) |
| -v / –verboseDataset : Number | $ **h5diff -v dset-num1.h5 dset-num2.h5 /dset**dataset: </dset> and </dset>size: [2] [2]position dset dset difference ------------------------------------------------------------[ 0 ] 1 0 1 [ 1 ] 2 0 2 2 differences foundEXIT-CODE : 1 | $ **h5diff -v dset-num1.h5 dset-num2.h5**file1 file2--------------------------------------- x x /  x x /dset group : </> and </>0 differences founddataset: </dset> and </dset>size: [2] [2]position dset dset difference ------------------------------------------------------------[ 0 ] 1 0 1 [ 1 ] 2 0 2 2 differences foundEXIT-CODE : 1 |
| -v / –verboseDataset : String | $ **h5diff -v dset-str1.h5 dset-str2.h5 /string /string**dataset: </string> and </string>size: [2] [2]position string string difference ------------------------------------------------------------[ 0 ] y c[ 0 ] y a[ 0 ] y t[ 1 ] z d[ 1 ] z o[ 1 ] z g6 differences foundEXIT-CODE : 1 | $ **h5diff -v dset-str1.h5 dset-str2.h5**file1 file2--------------------------------------- x x /  x x /string group : </> and </>0 differences founddataset: </string> and </string>size: [2] [2]position string string difference ------------------------------------------------------------[ 0 ] y c[ 0 ] y a[ 0 ] y t[ 1 ] z d[ 1 ] z o[ 1 ] z g6 differences foundEXIT-CODE : 1 |
| more … |  |  |

## Continue (optional behaviors)

|  |  |  |
| --- | --- | --- |
|  | **Compare *object* vs *object*** | **Compare *file* vs *file*** |
| TODO* -v or –verbose : others
* -v1 or –verbose=1
* -v2 or –verbose=2
* -r or –report
* -q
* --follow-symlinks
* --no-dangling-links
* -N or –nan
* -n count or –count=
* -d D or –delta D
* -p R or –relative=R
* --use-system-epsilon
* --exclude-path “path”
* -- missing RM ---
* -c or –compare
 | TODO for each | TODO for each |

## Handle common, extra or not-comparable object and attribute with results

|  |  |  |
| --- | --- | --- |
|  | **Compare object vs object** | **Compare file vs file** |
| **Common** object/attribute | If difference exist* exit code = 1
* Display output
	+ Default (without ‘-v’ option)
		- display number of differences with names of object or attribute
	+ ‘–v’ option
		- display details of the differences (data values)
		- display number of differences with name of object or attribute
 | If difference exist* exit code = 1
* Display output
	+ Default (without ‘–v’ option)
		- display number of differences with names of object or attribute
	+ ‘–v’ option
		- object status list
		- display details of the differences (data values)
		- display number of differences with name of object or attribute
 |
| **Extra** object | N/A because user specifies obj1 and obj2 directly; no extra object. | * exit code = 1
* Display output
	+ Default (without verbose option)
		- None
	+ Use ‘-v’ to verify
		- List of object status on top
 |
| **Extra** attribute | * exit code = 0
* Display output
	+ Default (without verbose option)
		- None
	+ Use ‘-v1’ or ‘-v2’ to verify
		- attribute status line and list
 | * exit code = 0
* Display output
	+ Default (without verbose option)
		- None
	+ Use ‘-v1’ or ‘-v2’ to verify
		- attribute status line and list
 |

**Words definitions**:

* ‘**Common**’ means all of name, type and space are same for comparing the two HDF5 files, or all of type and space are same for comparing the two HDF5 objects.
* ‘**Extra object**’ means un-matching named object(s) exist in any HDF5 file for comparing the two HDF5 files.
* ‘**Extra attribute**’ means un-matching named attribute(s) exist in any object for comparing the two HDF5 objects.
* ‘**Non-comparable**’ means any of type or space is different for comparing two HDF5 files or objects.

## Continue (handle common, extra or not-comparable object and attribute with results)

|  |  |  |
| --- | --- | --- |
|  | **Compare object vs object** | **Compare file vs file** |
| **Non-comparable** object | * exit code = 0
* Display output
	+ Default (without ‘–v’ or’ –c’)
		- display “Some objects are not comparable”
	+ With ‘–v’ or ‘-c’
		- Display reason of not-compatible; each type or space of given objects
 | Dataset vs dataset* exit code = 0
* Display output
	+ Default (without ‘-v’ or ‘-c’)
		- display “Some objects are not comparable”
	+ with ‘-v’ or ‘-c’
		- Display reason of not-comparable; each type or space of not-comparable dataset
 |
| Mixing different objects (dataset , group, type) as same name* Bug - [HDFFV-7644](http://jira.hdfgroup.uiuc.edu/browse/HDFFV-7644)
 |
| **Non-comparable** attribute | * exit code =0
* Display output
	+ Default (without ‘-v’ or ‘-c’)
		- display “Some objects are not comparable”
	+ Use ‘-v’ or ‘-c’
		- Display reason of not-comparable; each type or space of not-comparable attribute
	+ Use ‘-v1’ or ‘-v2’ to view details
		- attribute status line or list
 | * exit code =0
* Display output
	+ Default (without ‘-v’ or ‘-c’)
		- display “Some objects are not comparable”
	+ Use ‘-v’ or ‘-c’
		- Display reason of not-comparable; each type or space of not-comparable attribute
	+ Use ‘-v1’ or ‘-v2’ to view details
		- attribute status line or list
 |

## Insensible behaviors

|  |  |  |
| --- | --- | --- |
|  | **Compare *object* vs *object*** | **Compare *file* vs *file*** |
| Extra object | N/A | $ **h5diff dset-extra1.h5 dset-extra2.h5**<None>EXIT-CODE : 1 |
| Extra attribute | $ **h5diff extra-attr1.h5 extra-attr2.h5 /do /do**<None>EXIT-CODE : 0 | $ **h5diff extra-attr1.h5 extra-attr2.h5** <None>EXIT-CODE : 0 |
| Not-comparable object | $ **h5diff obj-nocomparable1.h5 obj-nocomparable2.h5 /obj1**--------------------------------Some objects are not comparable--------------------------------Use -c for a list of objects.EXIT-CODE : 0 | $ **h5diff obj-nocomparable1.h5 obj-nocomparable2.h5**--------------------------------Some objects are not comparable--------------------------------Use -c for a list of objects.EXIT-CODE : 0 |
| Not-comparableattribute |  |  |
| Empty dataset |  |  |
| Empty file |  |  |
| Display help page on error |  |  |
| … |  |  |

# Appendix A – List of example HDF5 files

|  |  |
| --- | --- |
| **Dset-num1.h5** | **dset-num2.h5** |
| HDF5 "dset-num1.h5" {GROUP "/" { DATASET "dset" { DATATYPE H5T\_STD\_I32LE DATASPACE SIMPLE { ( 2 ) / ( 2 ) } DATA { (0): 1, 2 } }}} | HDF5 "dset-num2.h5" {GROUP "/" { DATASET "dset" { DATATYPE H5T\_STD\_I32LE DATASPACE SIMPLE { ( 2 ) / ( 2 ) } DATA { (0): 0, 0 } }}} |

|  |  |
| --- | --- |
| **dset-str1.h5** | **dset-str2.h5** |
| HDF5 "dset-str1.h5" {GROUP "/" { DATASET "string" { DATATYPE H5T\_STRING { STRSIZE 3; STRPAD H5T\_STR\_NULLTERM; CSET H5T\_CSET\_ASCII; CTYPE H5T\_C\_S1; } DATASPACE SIMPLE { ( 2 ) / ( 2 ) } DATA { (0): "yyy", "zzz" } }}} | HDF5 "dset-str2.h5" {GROUP "/" { DATASET "string" { DATATYPE H5T\_STRING { STRSIZE 3; STRPAD H5T\_STR\_NULLTERM; CSET H5T\_CSET\_ASCII; CTYPE H5T\_C\_S1; } DATASPACE SIMPLE { ( 2 ) / ( 2 ) } DATA { (0): "cat", "dog" } }}} |

|  |  |
| --- | --- |
| **extra-obj1.h5** | **extra-obj2.h5** |
| HDF5 "extra-obj1.h5" {GROUP "/" { DATASET "do" { DATATYPE H5T\_STD\_I32LE DATASPACE SIMPLE { ( 2 ) / ( 2 ) } DATA { (0): 1, 1 } }}} | HDF5 "extra-obj2.h5" {GROUP "/" { DATASET "do" { DATATYPE H5T\_STD\_I32LE DATASPACE SIMPLE { ( 2 ) / ( 2 ) } DATA { (0): 1, 1 } } GROUP "mi" { } DATASET "re" { DATATYPE H5T\_STD\_I32LE DATASPACE SIMPLE { ( 1, 2 ) / ( 1, 2 ) } DATA { (0,0): 3, 3 } }}} |
|  | This file has extra object dataset “re” and group “mi”. |

|  |  |
| --- | --- |
| **extra-attr1.h5** | **extra-attr2.h5** |
| HDF5 "extra-attr1.h5" {GROUP "/" { DATASET "do" { DATATYPE H5T\_STD\_I32LE DATASPACE SIMPLE { ( 2 ) / ( 2 ) } DATA { (0): 1, 1 } ATTRIBUTE "attr1" { DATATYPE H5T\_STD\_I32LE DATASPACE SIMPLE { ( 1 ) / ( 1 ) } DATA { (0): 1 } } }}} | HDF5 "extra-attr2.h5" {GROUP "/" { DATASET "do" { DATATYPE H5T\_STD\_I32LE DATASPACE SIMPLE { ( 2 ) / ( 2 ) } DATA { (0): 1, 1 } ATTRIBUTE "attr1" { DATATYPE H5T\_STD\_I32LE DATASPACE SIMPLE { ( 1 ) / ( 1 ) } DATA { (0): 1 } } ATTRIBUTE "attr2" { DATATYPE H5T\_STD\_I32LE DATASPACE SIMPLE { ( 1 ) / ( 1 ) } DATA { (0): 2 } } }}} |
|  | The object “do” has extra attribute “attr2” |

|  |  |
| --- | --- |
| **obj-nocomparable1.h5** | **obj-nocomparable2.h5** |
| HDF5 "obj-nocomparable1.h5" {GROUP "/" { DATASET "obj1" { DATATYPE H5T\_STD\_I32LE DATASPACE SIMPLE { ( 2 ) / ( 2 ) } DATA { (0): 1, 2 } } DATASET "obj2" { DATATYPE H5T\_STD\_I32LE DATASPACE SIMPLE { ( 1, 1 ) / ( 1, 1 ) } DATA { (0,0): 10 } }}} | HDF5 "obj-nocomparable2.h5" {GROUP "/" { DATASET "obj1" { DATATYPE H5T\_STRING { STRSIZE 5; STRPAD H5T\_STR\_NULLPAD; CSET H5T\_CSET\_ASCII; CTYPE H5T\_C\_S1; } DATASPACE SIMPLE { ( 1, 1 ) / ( 1, 1 ) } DATA { (0,0): "abcde" } } DATASET "obj2" { DATATYPE H5T\_STD\_I32LE DATASPACE SIMPLE { ( 1, 1 ) / ( 1, 1 ) } DATA { (0,0): 10 } }}} |

# Acknowledgements

This work is supported by a commercial client of the HDF group.

# Revision History

|  |  |
| --- | --- |
| *August 11, 2011:* | Version 1 draft 1 circulated for directional comment within The HDF Group. |