Support HDF5 1.8 in the HDFView

Peter Cao
Nidhi Gupta

The HDF Group

HDF5 1.8 includes a number of new features that offer many users of HDF5 substantial performance improvements and expanded capabilities. Many of these features can only be accessed via revised API calls. To support these new APIs and new features in the HDF-Java products, we need make changes to the current the HDF-Java products. The changes may affect applications that use the HDF-Java products, especially the HDF5 Java wrapper.

The work for supporting HDF5 1.8 APIs in JHI5 and in the HDF-Java object layer has been completed. For details of the work, see RFC at https://www.hdfgroup.uiuc.edu/RFC/HDF5/hdf-java/.

The purpose of this document is to scope out the work needed to be done to support HDF5 1.8 in HDFView and set priorities of the work. Separate design documents will be needed based on the complexity of some of the features.

Table of Contents

1 Introduction

HDF5 1.8 represents a major update to the HDF5 Library, utilities, and file format. The changes introduced in HDF5 1.8 provide new capabilities and improve performance.

There are three distinct HDF5 Java Products:

- Java HDF5 Interface (JHI5): the Java Native Interface to the standard HDF5 library.
- Java HDF Object Package: a Java package that implements HDF data objects in an object-oriented form.
- HDFView: a visual tool for browsing and editing HDF4 and HDF5 files.

HDFView is built on the Java HDF Object Package, which in turn is built on the Java HDF5 Interface. Modifications will be needed in all three products to support HDF5 1.8 features. The necessary changes for JHI5 and the object layer have been completed in May 2010. This RFC mainly addresses the issue of supporting HDF5 1.8 in HDFView.

2 Current features

2.1 Testing current features

New features have been added to JHI5 and the HDF-Java object layer. Although the changes have been tested at JHI5 and the HDF-Java object layer, we want to make sure that the changes do not break any of the current features in HDFView.

HDFView will be tested according to the HDFView Test Checklist that was created and used for testing HDFView releases.

2.2 Adding new tests for copying objects

Since the method for copying HDF5 objects in the object layer has been replaced by H5Ocopy(), we need to test the change in HDFView. New tests should be added to the HDFView Test Checklist for HDF5.

- Copying a large dataset with contiguous layout, e.g. 3GB dataset. The old copy() method implemented in Java cannot copy large dataset because it loads all the entire dataset in memory from the source and writes it to the destination.
- Copying a chunked dataset where storage in file is not allocated. We want to make sure that the file storage of the copied dataset is not allocated.
- Copying a chunked dataset where file storage is partially allocated. The copied dataset should have the same file storage. The old copy() method allocates the entire dataset even if the source dataset is partially allocated because there is no way for the application to know which chunk is allocated and which is not.

3 New features

Since the HDFView is built on HDF-Java objects, the new features of HDFView will be limited to what is supported in the HDF5 Java object layer.

3.1 Adding new methods to the abstract layer of the object package

Since HDFView only uses the abstract layer of the HDF-Java objects, we have to add new methods that are needed to support the new features. For example, new methods will be needed to add to the current abstract objects to handle external links. The work for adding new methods should be small since the implementation has been already done in the HDF5 Java objects.

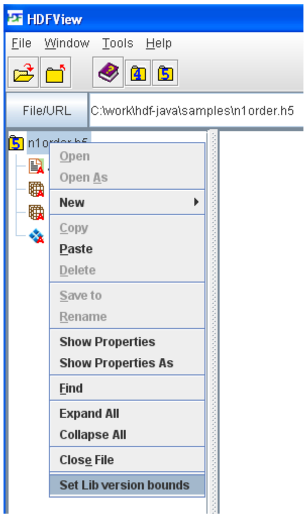
3.1.1 Library versions

HDF5 1.8 allows users to create object with a specified version. By default, objects will be created the earliest possible format that will handle the data being stored and accommodate the action being taken. Changes will be made to HDF-Java object to support this feature.

Changes:

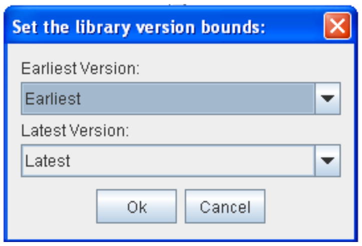
1. Add an option to set the bounds of library versions when opening a file. The file should not be opened as read-only.

   - If the file is HDF5 file, then right click on the file. The user can then set the library version bounds by selecting it from the popup menu.



File/URL C:\work\hdf-java\samples\n1order.h5

   - A new dialog box appears. The user has to select the earliest and latest library versions to be set.



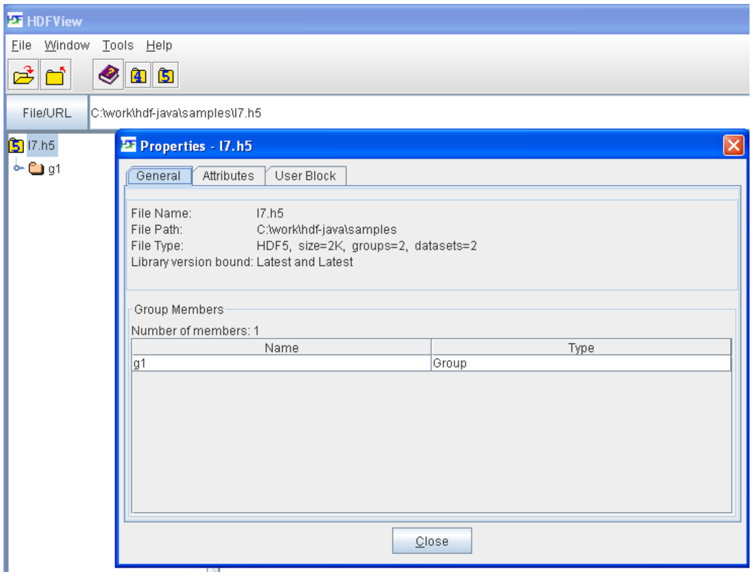
Set the library version bounds:

Earliest Version: Earliest

Latest Version: Latest

2. Allow users to see the bounds of library versions of an open file that was set above.

   - If the library version bounds are set above, then right click on the file and select 'show properties'. This will display the library versions that had been set.



File Name: l7.h5
File Path: C:\work\hdf-java\samples
File Type: HDF5, size=2K, groups=2, datasets=2
Library version bound: Latest and Latest

Group Members
Number of members: 1

Name	Type
g1	Group

3.1.2 Groups

Compact-or-indexed groups enable much-compressed link storage for groups with very few members and improved efficiency and performance for groups with very large numbers of members.

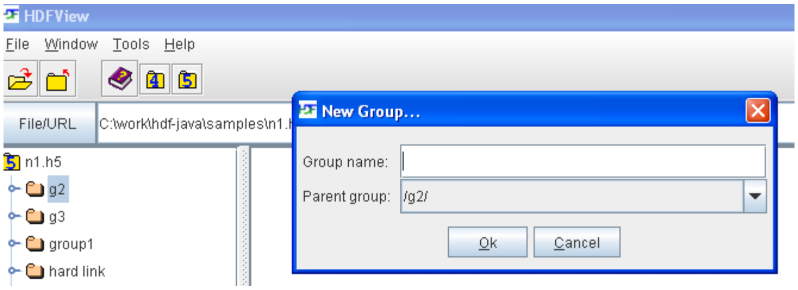
H5Pset_link_creation_order sets creation order tracking and indexing for links in a group.

We will add new options to allow users to set link storage and to use creation order other than alphabetical order.

Changes:

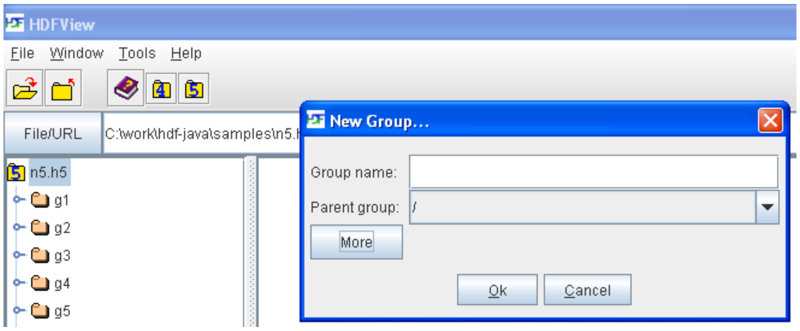
1. ***Allow users to set link storage and use creation order.***

   3. If the format is not HDF5, then the create new group GUI is same as original GUI:



Group name:
Parent group: /g2/

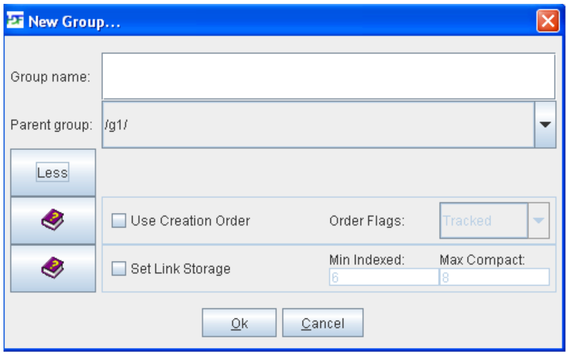
   4. If the format is HDF5, then the New Group GUI has an extra button, ‘More’, which would allow users to set the link storage and use creation order.



File/URL: C:\work\hdf-java\samples\n5.h

Parent group: /

5. After clicking the More button, the create New Group GUI would be like:



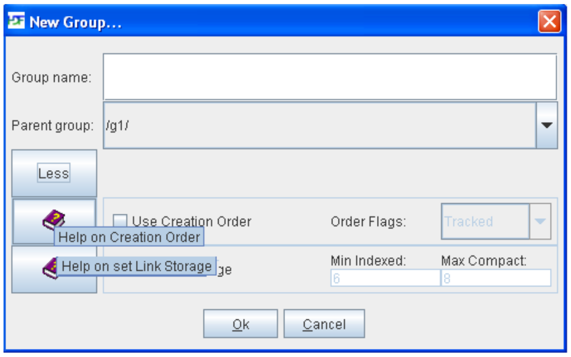
Parent group: /g1/

□ Use Creation Order Order Flags: Tracked

□ Set Link Storage Min Indexed: 6 Max Compact: 8

6. The More button changes to a less button. The Less button when clicked, removes the option for using creation order and link storage.

7. If the user wants more information about creation order and link storage, then clicking on the help buttons will provide detailed information.

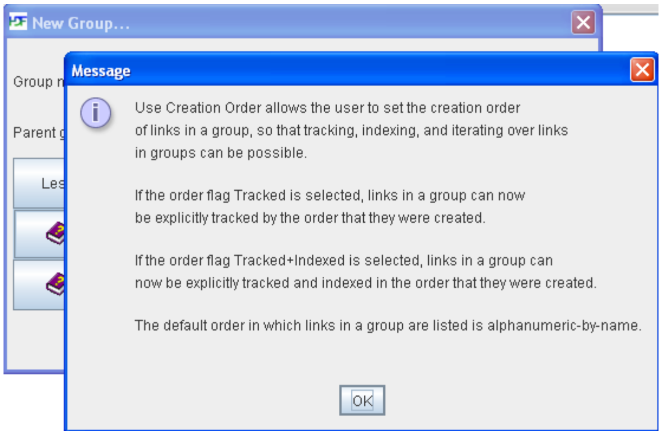


Parent group: /g1/

□ Use Creation Order Order Flags: Tracked

Min Indexed: 6 Max Compact: 8

8. By clicking on the Help on Creation Order help button, the following message is displayed:



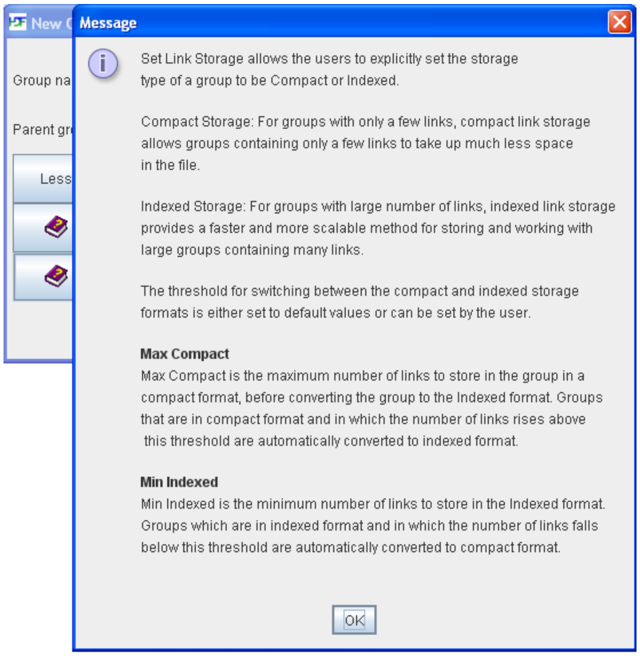
Use Creation Order allows the user to set the creation order of links in a group, so that tracking, indexing, and iterating over links in groups can be possible.

If the order flag Tracked is selected, links in a group can now be explicitly tracked by the order that they were created.

If the order flag Tracked+Indexed is selected, links in a group can now be explicitly tracked and indexed in the order that they were created.

The default order in which links in a group are listed is alphanumeric-by-name.

9. By clicking on the set link storage help button, the following message is displayed:



Set Link Storage allows the users to explicitly set the storage type of a group to be Compact or Indexed.

Compact Storage: For groups with only a few links, compact link storage allows groups containing only a few links to take up much less space in the file.

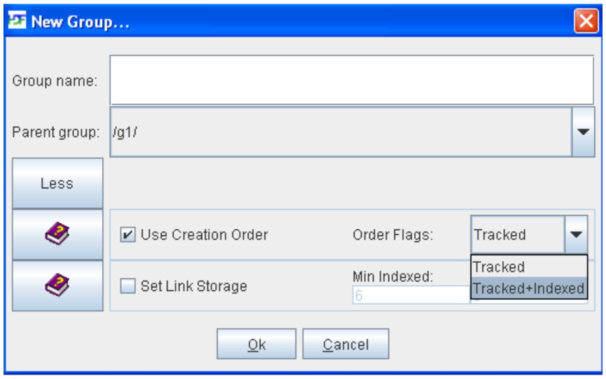
Indexed Storage: For groups with large number of links, indexed link storage provides a faster and more scalable method for storing and working with large groups containing many links.

The threshold for switching between the compact and indexed storage formats is either set to default values or can be set by the user.

Max Compact
Max Compact is the maximum number of links to store in the group in a compact format, before converting the group to the Indexed format. Groups that are in compact format and in which the number of links rises above this threshold are automatically converted to indexed format.

Min Indexed
Min Indexed is the minimum number of links to store in the Indexed format. Groups which are in indexed format and in which the number of links falls below this threshold are automatically converted to compact format.

10. By clicking on the check box, use creation order, the Order Flags are enabled and can be set.



Group name:

Parent group: /g1/

☑ Use Creation Order Order Flags: Tracked

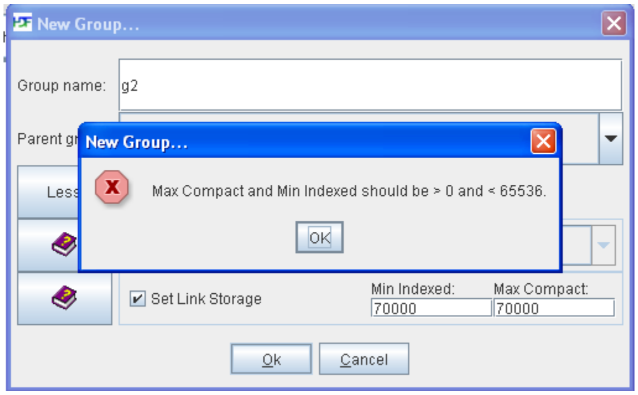
Tracked
Tracked+Indexed

□ Set Link Storage Min Indexed: 6

11. Users can set Link storage, by clicking on the check box, then the Min Indexed and Max Compact text fields are enabled with default values 6 and 8 respectively.

12. The max compact and min indexed can be set by the user. These text fields allow only numeric input and the length is set to 5 digits.

13. If max compact and min indexed are set to large values, then an error message is displayed.



New Group...

Group name: g2

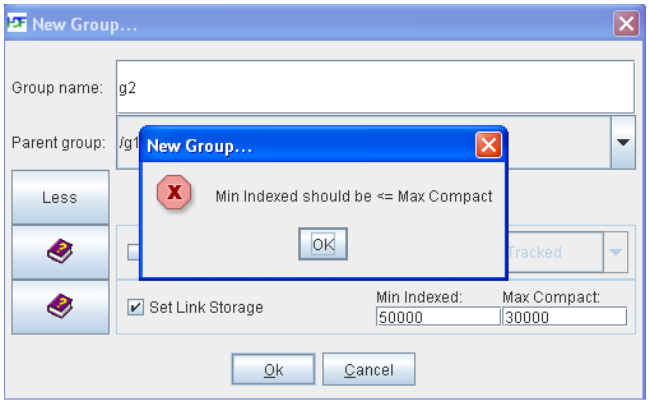
Max Compact and Min Indexed should be > 0 and < 65536.

☑ Set Link Storage

Min Indexed: 70000

Max Compact: 70000

14. An error message is displayed if the value of min indexed is greater than max compact.



New Group...

Group name: g2

Parent group: /g1

Min Indexed should be <= Max Compact

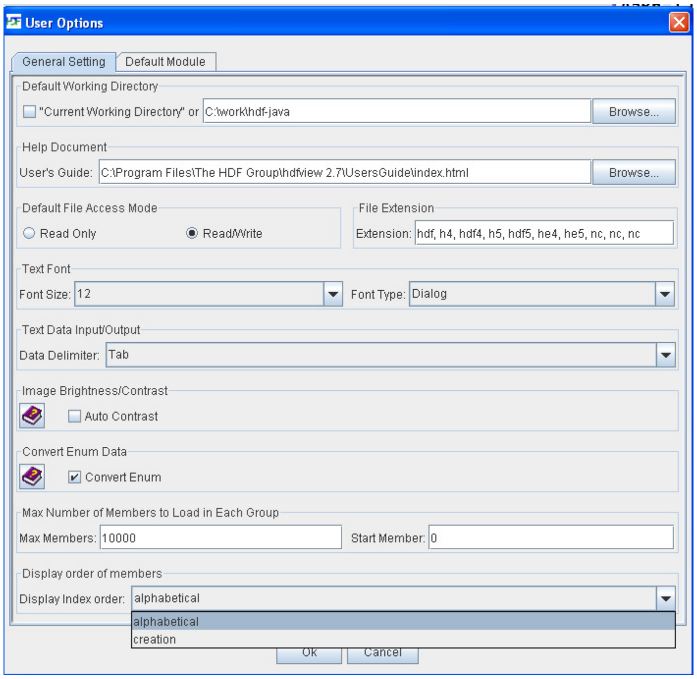
☑ Set Link Storage

Min Indexed: 50000

Max Compact: 30000

2. *Allow users to see objects in creation or alphabetical order.*

15. Added an option in User Options, to set the display order:



User Options

General Setting | Default Module

Default Working Directory
☐ "Current Working Directory" or C:\work\hdf-java Browse...

Help Document
User's Guide: C:\Program Files\The HDF Group\hdfview 2.7\UsersGuide\index.html Browse...

Default File Access Mode
○ Read Only ● Read/Write

File Extension
Extension: hdf, h4, hdf4, h5, hdf5, he4, he5, nc, nc, nc

Text Font
Font Size: 12 Font Type: Dialog

Text Data Input/Output
Data Delimiter: Tab

Image Brightness/Contrast
☐ Auto Contrast

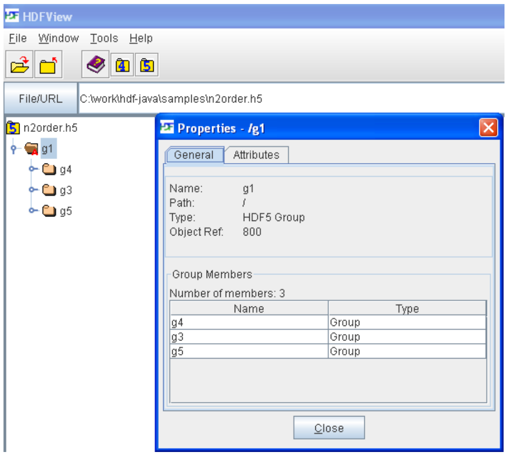
Convert Enum Data
☑ Convert Enum

Max Number of Members to Load in Each Group
Max Members: 10000 Start Member: 0

Display order of members
Display Index order: alphabetical
alphabetical
creation

Ok Cancel

16. Now objects in the group are listed in the order selected. For example, if it is creation order, the objects in group g1 are listed as:



Properties - /g1

General | Attributes

Name: g1
Path: /
Type: HDF5 Group
Object Ref: 800

Group Members
Number of members: 3

Name	Type
g4	Group
g3	Group
g5	Group

Close

3.1.3 External links and soft links

External links enable the insertion of a link into an HDF5 group in one file that points to an HDF5 object in a different HDF5 file.

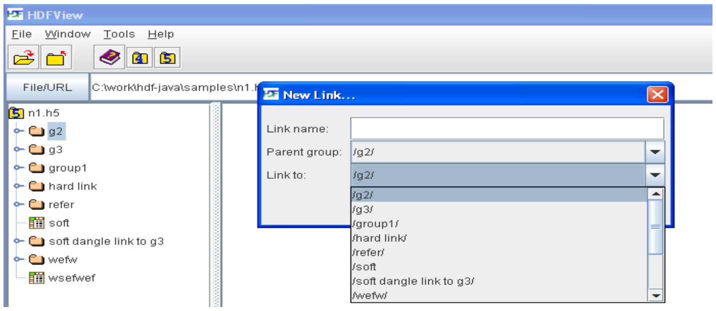
Soft links are symbolic links that use a name to point to an HDF5 object in the same file.

With the completion of the new HDF-Java JNI, reading/writing content of an object pointed by an external link will work with the HDF-Java object package. However, the current object package does not allow users to create an external and soft links and get link information.

1. ***Add an option or menu item to allow users to create external and soft links.***

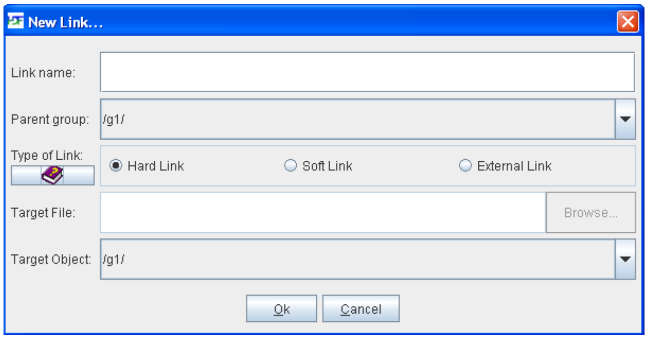
Originally:

17. We have the option to create hard links only.

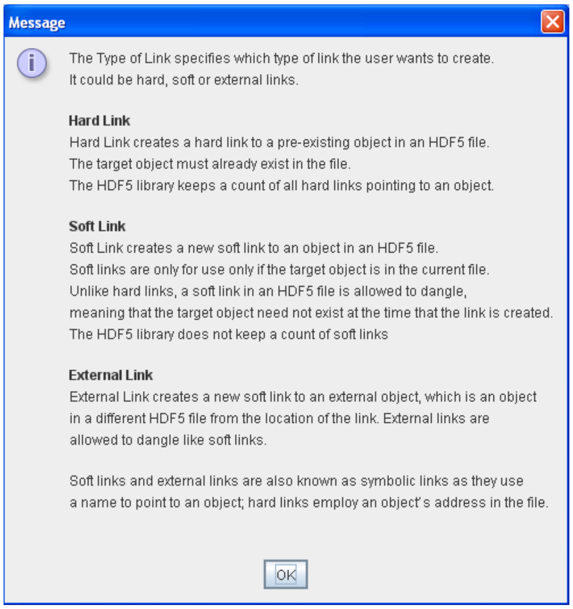


Changes:

18. Add a new option or menu item to allow users to create hard, soft or external links.



- If the user wants help on the type of link to create, then the help button can be clicked and the following message is displayed:



The Type of Link specifies which type of link the user wants to create.
It could be hard, soft or external links.

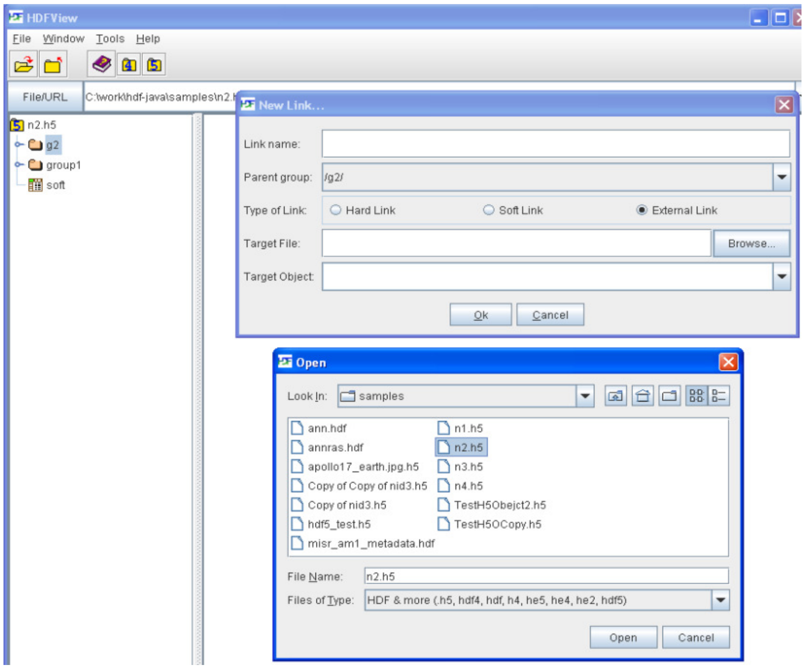
Hard Link
Hard Link creates a hard link to a pre-existing object in an HDF5 file.
The target object must already exist in the file.
The HDF5 library keeps a count of all hard links pointing to an object.

Soft Link
Soft Link creates a new soft link to an object in an HDF5 file.
Soft links are only for use only if the target object is in the current file.
Unlike hard links, a soft link in an HDF5 file is allowed to dangle,
meaning that the target object need not exist at the time that the link is created.
The HDF5 library does not keep a count of soft links

External Link
External Link creates a new soft link to an external object, which is an object
in a different HDF5 file from the location of the link. External links are
allowed to dangle like soft links.

Soft links and external links are also known as symbolic links as they use
a name to point to an object; hard links employ an object's address in the file.

- Add a Target File text field and a browse button , which is enabled only in the case of external links. The browse button allows user to select a target file. The Target File text field allows users to type in the target file.



File/URL: C:\work\hdf-java\samples\n2.h

n2.h5
g2
group1
soft

Link name:

Parent group: /g2/

Type of Link: ○ Hard Link ○ Soft Link ● External Link

Target File:

Target Object:

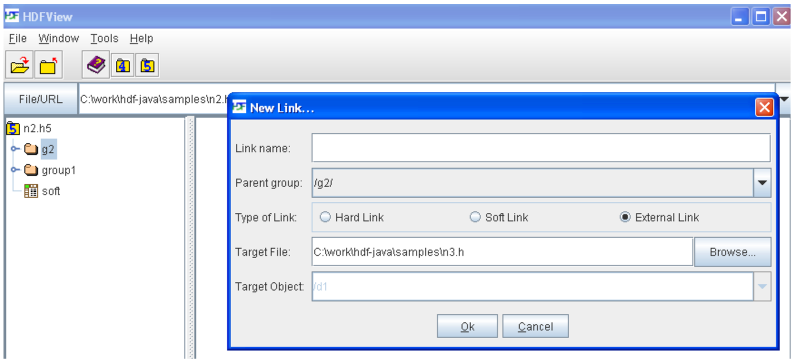
Look In: samples

ann.hdf
annras.hdf
apollo17_earth.jpg.h5
Copy of Copy of nid3.h5
Copy of nid3.h5
hdf5_test.h5
misr_am1_metadata.hdf
n1.h5
n2.h5
n3.h5
n4.h5
TestH5Obejct2.h5
TestH5OCopy.h5

File Name: n2.h5

Files of Type: HDF & more (.h5, hdf4, hdf, h4, he5, he4, he2, hdf5)

- The Target objects are automatically populated. For external links, target objects are automatically populated, both in case of user entering the target file manually or selecting a file by browsing. If user has entered a file that is invalid, then the target object option is disabled.



File/URL: C:\work\hdf-java\samples\n2.h

n2.h5
g2
group1
soft

Link name:

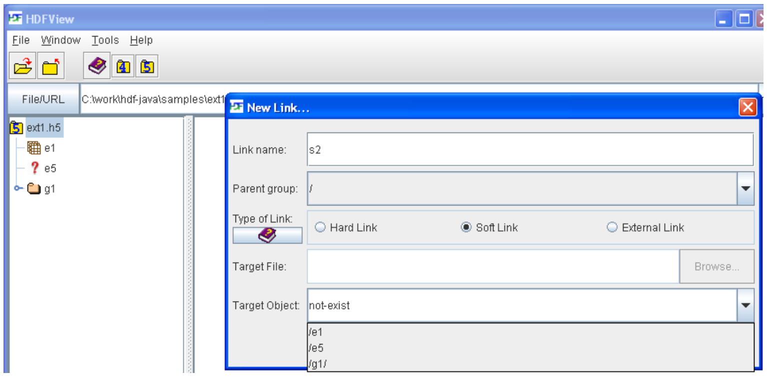
Parent group: /g2/

Type of Link: ○ Hard Link ○ Soft Link ● External Link

Target File: C:\work\hdf-java\samples\n3.h

Target Object: /d1

- Users can now create dangling links. That is, they can create soft or external links to objects that do not exist in the file, by typing in the target object name.



Link name: s2

Parent group: /

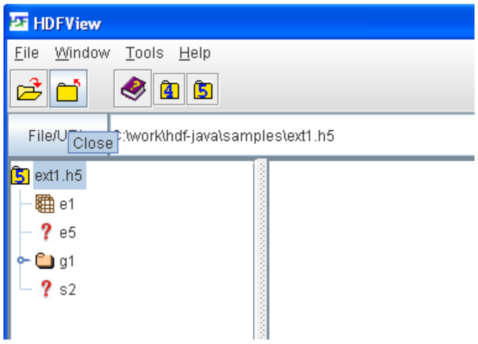
Type of Link: ○ Hard Link ● Soft Link ○ External Link

Target File:

Target Object: not-exist

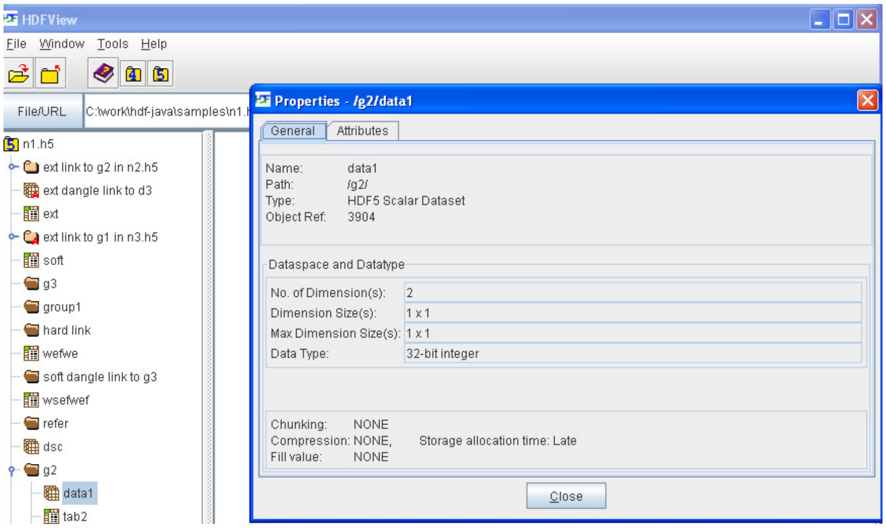
/e1
/e5
/g1/

- The new link object created e.g. S2, would have a “question” icon as the type is unknown.



2. *Show Link Information in HDFView.*

- Show object(that is not a soft or external link) information (same as the original)



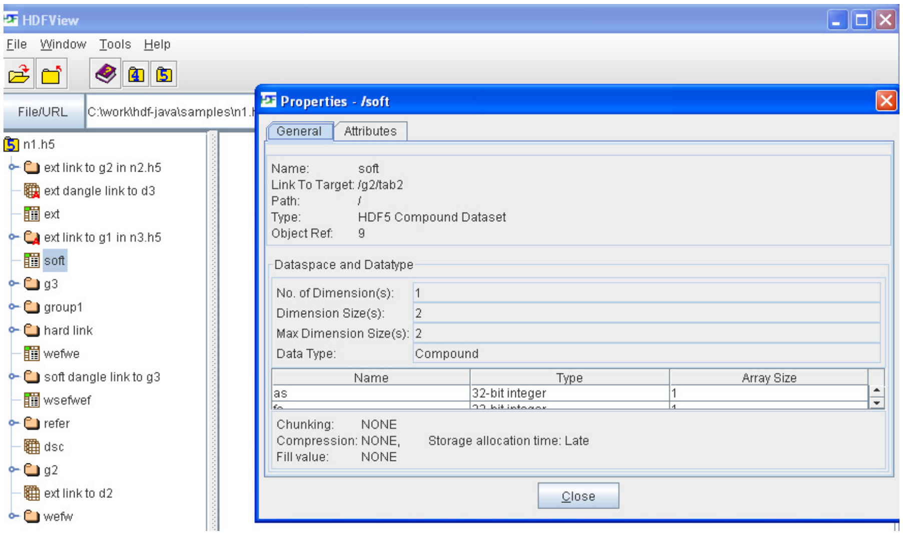
Name: data1
Path: /g2/
Type: HDF5 Scalar Dataset
Object Ref: 3904

Dataspace and Datatype

No. of Dimension(s): 2
Dimension Size(s): 1 x 1
Max Dimension Size(s): 1 x 1
Data Type: 32-bit integer

Chunking: NONE
Compression: NONE, Storage allocation time: Late
Fill value: NONE

- Show link information for soft links (path to target object)



Name: soft
Link To Target: /g2/tab2
Path: /
Type: HDF5 Compound Dataset
Object Ref: 9

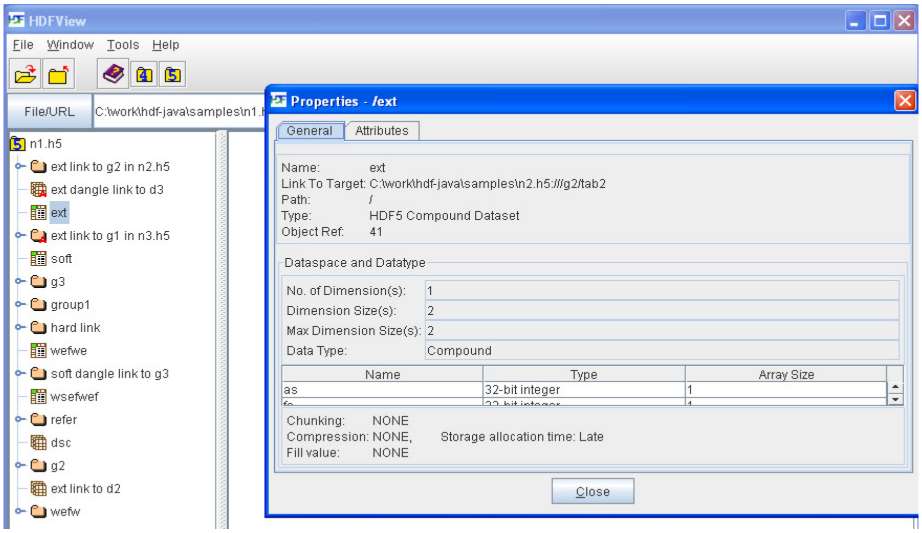
Dataspace and Datatype

No. of Dimension(s): 1
Dimension Size(s): 2
Max Dimension Size(s): 2
Data Type: Compound

Name	Type	Array Size
as	32-bit integer	1

Chunking: NONE
Compression: NONE, Storage allocation time: Late
Fill value: NONE

- Show information for external links (external file and path) in HDFView.



Name: ext
Link To Target: C:\work\hdf-java\samples\n2.h5:///g2/tab2
Path: /
Type: HDF5 Compound Dataset
Object Ref: 41

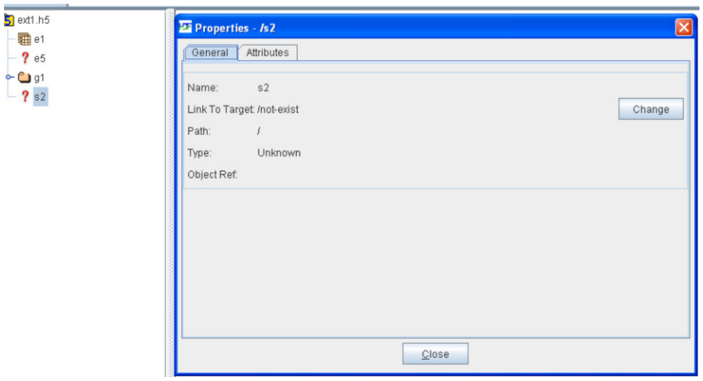
Dataspace and Datatype

No. of Dimension(s): 1
Dimension Size(s): 2
Max Dimension Size(s): 2
Data Type: Compound

Name	Type	Array Size
as	32-bit integer	1

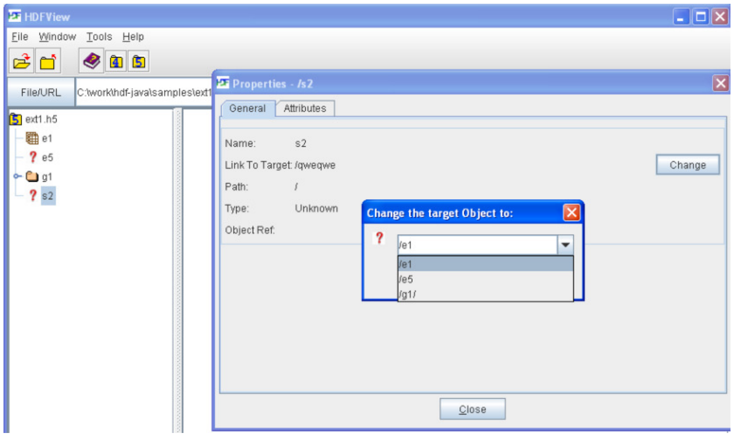
Chunking: NONE
Compression: NONE, Storage allocation time: Late
Fill value: NONE

- Show information for soft/external links that point to a non existing object



Name: s2
Link To Target: /not-exist
Path: /
Type: Unknown
Object Ref:

- The user can change the target object that the link is linking to, by clicking on the Change button.
- If the user changes the target object for a **soft link**, then a message box appears, listing the current objects in the file. The user can enter a target object that does not exist. By clicking on the OK button, the target object that the soft link is pointing to will change.

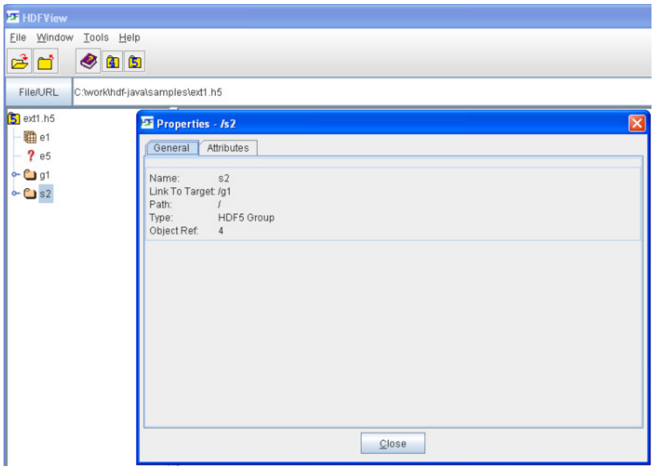


File/URL C:\work\hdf-java\samples\ext1

Name: s2
Link To Target: /qweqwe
Path: /
Type: Unknown
Object Ref:

Change the target Object to: /e1
/e1
/e5
/g1/

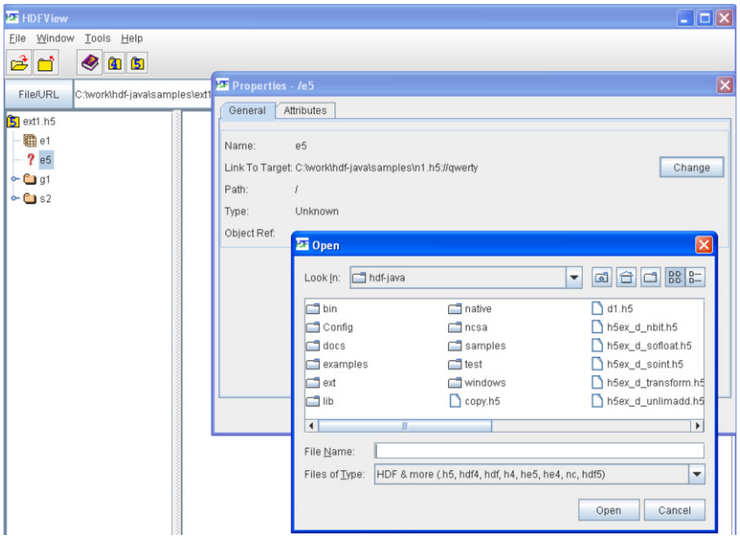
- For example, if the user chooses /g1/ as the target object then S2 will now link to g1 and be of type HDF5 group. This is shown below:



File/URL C:\work\hdf-java\samples\ext1.h5

Name: s2
Link To Target: /g1
Path: /
Type: HDF5 Group
Object Ref: 4

- If the user changes the target object for an **external link** (by clicking on Change button), then the user will get an option to choose the target file



File/URL C:\work\hdf-java\samples\ext1

Name: e5

Link To Target: C:\work\hdf-java\samples\n1.h5://qwerty

Path: /

Type: Unknown

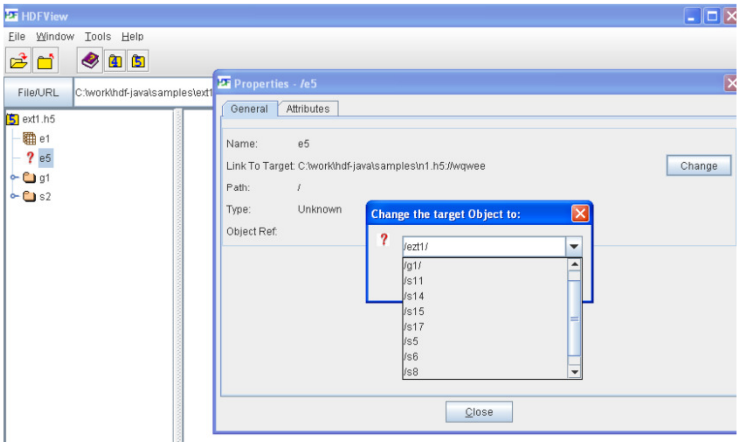
Object Ref:

Look In: hdf-java

File Name:

Files of Type: HDF & more (.h5, hdf4, hdf, h4, he5, he4, nc, hdf5)

- By selecting a file, the list of target objects in the target file will then be listed as:



File/URL C:\work\hdf-java\samples\ext1

Name: e5

Link To Target: C:\work\hdf-java\samples\n1.h5://wqwee

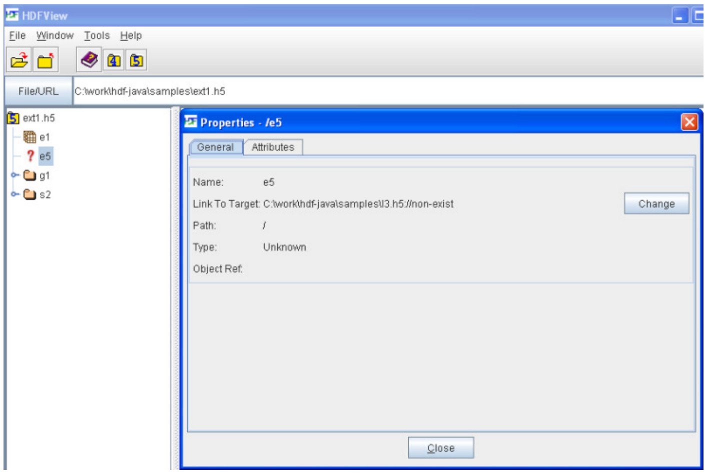
Path: /

Type: Unknown

Object Ref:

Change the target Object to: /ezt1/

- The user can select the target object from the list or enter a target object that does not exist.
- For example for external link e5, if the user wants to Change the target object and enters a non existing target object in a different file(l3.h5) and clicks on OK, then e5 will be:



File/URL C:\work\hdf-java\samples\ext1.h5

Name: e5

Link To Target: C:\work\hdf-java\samples\l3.h5://non-exist

Path: /

Type: Unknown

Object Ref:

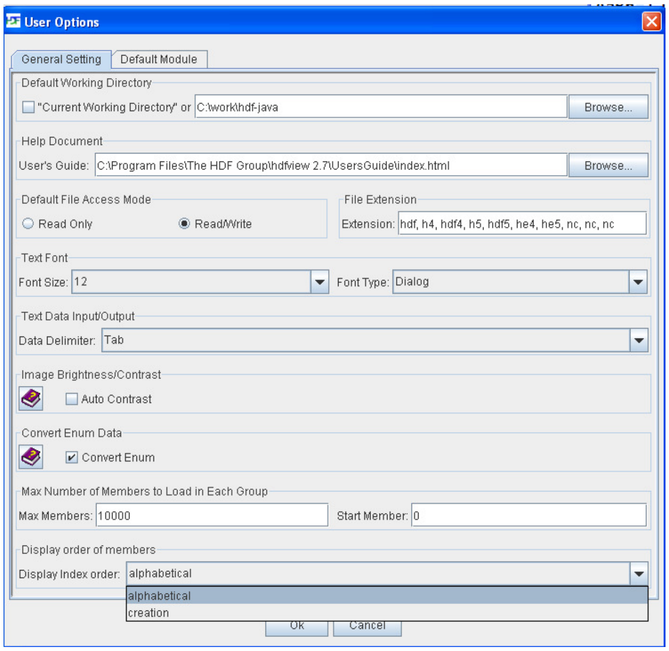
3.1.4 Attributes

Like groups, we will allow users to retrieve attributes in creation order. By default, attributes are listed in alphabetical order.

Changes:

1. Add an option to show attributes in creation order.

- Added an option in User Options, to set the display order:



Default Working Directory
□ "Current Working Directory" or C:\work\hdf-java Browse...

Help Document
User's Guide: C:\Program Files\The HDF Group\hdfview 2.7\UsersGuide\index.html Browse...

Default File Access Mode
○ Read Only ● Read/Write

File Extension
Extension: hdf, h4, hdf4, h5, hdf5, he4, he5, nc, nc, nc

Text Font
Font Size: 12 Font Type: Dialog

Text Data Input/Output
Data Delimiter: Tab

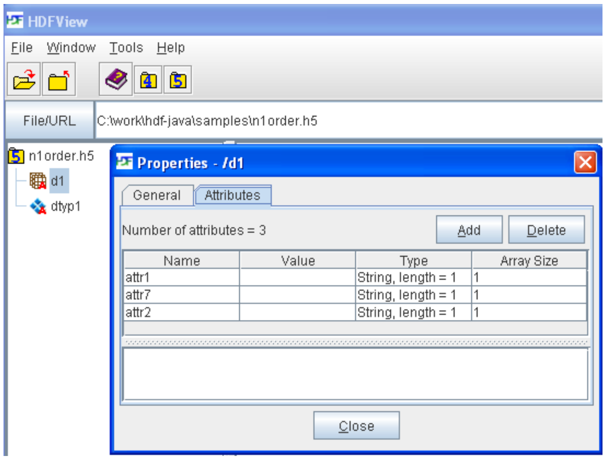
Image Brightness/Contrast
□ Auto Contrast

Convert Enum Data
☑ Convert Enum

Max Number of Members to Load in Each Group
Max Members: 10000 Start Member: 0

Display order of members
Display Index order: alphabetical
alphabetical
creation

- Now attributes are listed in the order selected. For example, if it is creation order, the attributes in d1 are listed as:



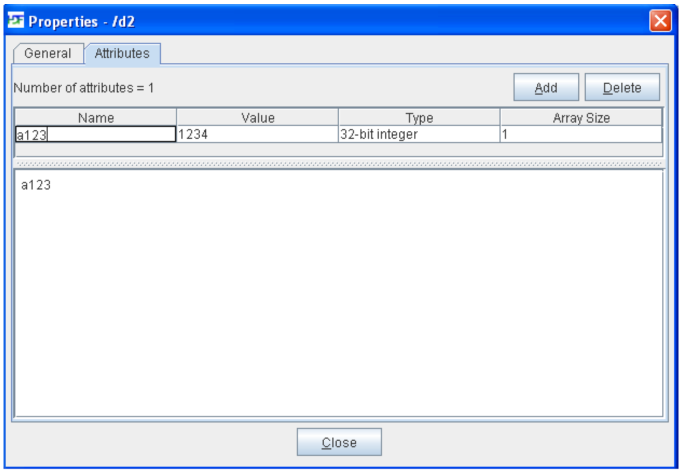
C:\work\hdf-java\samples\n1order.h5

Number of attributes = 3

Name	Value	Type	Array Size
attr1		String, length = 1	1
attr7		String, length = 1	1
attr2		String, length = 1	1

2. Rename Attributes

- If the user double clicks on the name cell, then the attribute name can be changed.

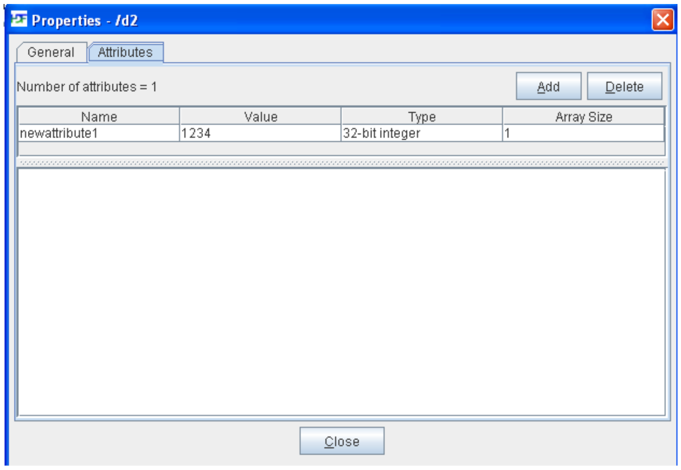


Number of attributes = 1

Name	Value	Type	Array Size
a123	1234	32-bit integer	1

a123

- For example *a123* is changed to *newattribute1*



Number of attributes = 1

Name	Value	Type	Array Size
newattribute1	1234	32-bit integer	1

4 Priorities and time estimation

The following is a list of tasks discussed above with priorities, 1 – high, 2 – medium, 3 – low. The work will include implementation, documentation, and testing. The estimation of number of hours is based on the fact that the staff will have good understanding of the hdf-java products.

Task	Priority	Work(hrs)	Notes
Test current features	1		
Test copying objects	1		
Add functions to the abstract layer	1		
Set the bounds of library versions	1		
Create external links	1		
Show link information	1		
Show and modify dangling links	1		
Use link creation order in groups	2		

Rename attributes	2		
Set link storage	2		
Use creation order in attributes	3		
Fix bugs and other unexpected			
Total			

Revision History

May 14, 2010	Version 1: initial RFC for circulating internally.
June 14, 2010	Version 2, added screenshots for the prototype design of the new features
July 8, 2010	Version 3, updated based on the feedback from the group
July 21, 2010	Version 4, added renaming attributes and setting library version bounds