

RFC: Reserved Characters for HDF5 Applications

Peter Cao

This document recommends a set of reserved characters for use by HDF5 applications. The HDF5 library does not have any restrictions on characters that can be used in link or attribute names, except that the forward slash, '/', is reserved for use as a separator between link names in paths. This flexibility creates challenges for HDF5 applications, especially HDF5 command-line utilities, where different components of HDF5 files often need to be separated from each other on the command line or in output. In this document, we propose a set of characters that would be reserved for use by HDF5 applications.

1 Introduction

The HDF5 format does not impose any restrictions on characters that can be used in link names or attribute names, except that the forward slash, '/', is reserved for use as a separator between link names in paths. This policy gives users great flexibility in the names they use in HDF5 files. At the same time, it also creates problems for applications.

The issue of appropriate separators has been discussed numerous times for various HDF5 utilities, such as h5watch and h5diff, and other products built on HDF5, such as the hdf-java products. Reserving a set of characters that will be used consistently by HDF5 applications seems to be the best possible solution.

2 Reserved Characters for HDF5 Applications and products built on HDF5

Table 1 shows the set of characters we are recommending be reserved for use by the HDF5 utilities and other products built on HDF5. The reserved characters will be applied to both the inputs for command-line options and the output from the HDF5 utilities or other applications if applicable.

Table 1 -- Characters reserved for HDF5 Applications

Character	Description
'\'	Escape character
'*'	Wildcard character
','	Separate dimensions

'/'	Separates link names to objects (i.e the "path to an object")
'#'	Separates a file name from the beginning of the link name path
'@'	Indicates the beginning of an attribute name
'.'	Indicates a field in a compound datatype
"" and ""	Double quotes, anything enclosed by a pair of double quotes is treated as a single entry
'[and]'	Begin/end dataspace description (e.g., [3,4])
'{ and }'	Begin/end datatype description (e.g., {int})
'< and >'	Indicates array datatype dimensions
'(and)'	Indicates a variable-length datatype

The reserved characters will be only restricted in their use by HDF5 applications, and not by the HDF5 format or library. In other words, reserving the characters will not change the current behavior of the library. It will, however, affect applications working with HDF5 files where reserved characters are used in link or attribute names. For example, '@' is reserved to separate a link and an attribute. Using '@' in a link name or attribute name will not cause the library calls to fail or the HDF5 file to be invalid. However, some HDF5 utility features, such as the "--exclude-object-attribute" option for h5diff, may not work for the links and attributes with '@' in their name. We strongly recommend users not to use the reserved characters in link names and attribute names.

3 Use Cases

3.1 Passing file name and link name as a single parameter in hdf-java

Java static methods can be invoked without creating an instance of the class. Using a static method, such as `H5File.getObject(String fullPath)` in the `hdf-java` object package, users can retrieve a specific object without having an instance of `H5File` and the whole file structure in memory. The ability to pass a single parameter would be convenient for users. For example,

```
Dataset dset = (Dataset)H5File.getObject("/tmp/test.h5#/dset");
```

Where `"/tmp/test.h5"` is the file name and `"/dset"` is the path to the dataset.

The separator is also very useful for client/server applications. The file name and link name can be passed to the server in a single string, and is similar to using `"#"` to access a specific part of the page content in an URL.

3.2 Identifying a specific attribute in h5diff

The `--exclude-object-attribute` option is used to exclude certain attributes for comparing objects in `h5diff`. The proposed approach is to use a `<object_name, attribute_name>` pair of strings for the option. For details see RFC at https://www.hdfgroup.uiuc.edu/RFC/HDF5/tools/h5diff/ignore_obj/.

Some users proposed to use a separator between the link name and the attribute name so that the “**-exclude-object-attribute**” option takes a single string instead of a pair of strings. Using a reserved separator, the h5diff command-line interface or configuration file for excluding attributes will be consistent and much simpler. For example,

```
h5diff --exclude-object-attribute “/sensor_data@time_stamp” File5.h5 File6.h5
```

Where “/sensor_data” is the dataset and “time_stamp” is the attribute to be excluded from the dataset for comparison.

3.3 Separating nested compound fields in h5watch

The h5watch tool allows users to output new elements appended to a dataset as it grows. For details, see the RFC at <https://thehdfgroup-import.pbworks.com/h5watch>. One of the features in h5watch is able to show values of a specific field in a nested compound datatype. For this purpose, three reserved characters are used:

- A comma, ‘,’, is used to separate field names in a compound data type.
- A period, ‘.’, is used as the separator for nested fields.
- An escape character, ‘\’, is used for any “.” that is part of a field name.
- A slash character, ‘/’, is used to separate file name and object path (we may need to change it to use ‘#’ as we proposed in table 1).

The following is an example of a compound datatype for dataset, dset, in an HDF5, compd.h5.

```
DATATYPE H5T_COMPOUND {
  H5T_STD_I32LE "lon";
  H5T_STD_I32LE "lat";
  H5T_COMPOUND {
    H5T_STD_I32LE "high.day";
    H5T_STD_I32LE "low.day";
    H5T_STD_I32LE "high.wk";
    H5T_STD_I32LE "low.wk";
  } "temp";
}
```

The following is an example of h5watch command to show the values of the fields of “long”, “lat”, and “high.day” in the compound dataset, dset, in the file, compd.h5.

```
h5watch --fields=lon,lat,temp.high\day compd.h5/dset
```

3.4 Specifying a region of dataset in h5dump

Subsetting is available in h5dump by using the following options for a dataset:

- s START, --start=START Offset of start of subsetting selection
- S STRIDE, --stride=STRIDE Hyperslab stride
- c COUNT, --count=COUNT Number of blocks to include in selection
- k BLOCK, --block=BLOCK Size of block in hyperslab

',' is used to separate values of dimension sizes. For example,

```
h5dump -d /foo -s "0,1" -S "1,1" -c "2,3" -k "2,2" quux.h5
```

4 Recommendation

Our recommendation is to publish a standard set of characters reserved for HDF5 application. Although using reserved characters in link names and attribute names will not cause the library calls to fail but it may cause some applications to fail if the reserved characters are used in those applications for their intended purposes. Therefore, we strongly recommend users not to use the reserved characters in link names and attribute names.

Revision History

July 27, 2010: Version 1, first draft, requested comment from Ruth.

October 5, 2010: Version 2, revised based on Ruth's comments, circulated for comment within The HDF Group.