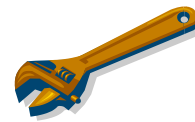




Introduction to HDF5 Command-line Tools



HDF5 Command-line Tools



Readers

- h5dump, h5diff, h5ls
- h5stat, h5check (*new in release 1.8*)



Writers

- h5import, h5repack, h5repart, h5jam/h5unjam
- h5copy, h5mkgrp (*new in release 1.8*)



Converters

- h4toh5, h5toh4, gif2h5, h52gif



h5dump

h5dump: exports (dumps) the contents of an HDF5 file

- Multiple output types
 - ASCII
 - binary
 - XML
- Complete or selected file content
 - Object header information (the structure)
 - Attributes (the metadata)
 - Datasets (the data)
 - All dataset values
 - Subsets of dataset values
 - Properties (filters, storage layout, fill value)
 - Specific objects: groups/ datasets/ attributes / named datatypes / soft links
- `h5dump -help`
 - Lists all option flags

September 9, 2008

SPEEDUP Workshop - HDF5 Tutorial

3

www.hdfgroup.org



Example: h5dump

No options: "All" contents to standard out

```
% h5dump Sample.h5
HDF5 "Sample.h5" {
GROUP "/" {
GROUP "Floats" {
DATASET "FloatArray" {
DATATYPE H5T_IEEE_F32LE
DATASPACE SIMPLE { ( 4, 3 ) / ( 4, 3 ) }
DATA {
(0,0): 0.01, 0.02, 0.03,
(1,0): 0.1, 0.2, 0.3,
(2,0): 1, 2, 3,
(3,0): 10, 20, 30
}
}
}
DATASET "IntArray" {
DATATYPE H5T_STD_I32LE
DATASPACE SIMPLE { ( 5, 6 ) / ( 5, 6 ) }
DATA {
(0,0): 0, 1, 2, 3, 4, 5,
(1,0): 10, 11, 12, 13, 14, 15,
(2,0): 20, 21, 22, 23, 24, 25,
(3,0): 30, 31, 32, 33, 34, 35,
(4,0): 40, 41, 42, 43, 44, 45
}
}
}
}
```

September 9, 2008

SPEEDUP Workshop - HDF5 Tutorial

4

www.hdfgroup.org



h5dump - object header information

-H option: Object header information

```
% h5dump -H Sample.h5
```

```
HDF5 "Sample.h5" {  
  GROUP "/" {  
    GROUP "Floats" {  
      DATASET "FloatArray" {  
        DATATYPE H5T_IEEE_F32LE  
        DATASPACE SIMPLE {(4, 3)/(4, 3)}  
      }  
    }  
    DATASET "IntArray" {  
      DATATYPE H5T_STD_I32LE  
      DATASPACE SIMPLE {(5, 6)/(5, 6)}  
    }  
  }  
}
```

September 9, 2008

SPEEDUP Workshop - HDF5 Tutorial

5

www.hdfgroup.org



h5dump – specific dataset

-d dataset option: Specific *dataset*

```
% h5dump -d /Floats/FloatArray Sample.h5
```

```
HDF5 "Sample.h5" {  
  DATASET "/Floats/FloatArray" {  
    DATATYPE H5T_IEEE_F32LE  
    DATASPACE SIMPLE {(4, 3)/(4, 3)}  
    DATA {  
      (0,0): 0.01, 0.02, 0.03,  
      (1,0): 0.1, 0.2, 0.3,  
      (2,0): 1, 2, 3,  
      (3,0): 10, 20, 30  
    }  
  }  
}
```

September 9, 2008

SPEEDUP Workshop - HDF5 Tutorial

6

www.hdfgroup.org



h5dump – dataset values to file

-o file option: Dataset values output to *file*

```
% h5dump -o Ofile -d /IntArray Sample.h5
```

```
HDF5 "Sample.h5" {  
  DATASET "/IntArray" {  
    DATATYPE H5T_STD_I32LE  
    DATASPACE SIMPLE { ( 5, 6 ) / ( 5, 6 ) }  
    DATA {  
    }  
  }  
}
```

```
% cat Ofile
```

```
(0,0): 0, 1, 2, 3, 4, 5,  
(1,0): 10, 11, 12, 13, 14, 15,  
(2,0): 20, 21, 22, 23, 24, 25,  
(3,0): 30, 31, 32, 33, 34, 35,  
(4,0): 40, 41, 42, 43, 44, 45
```

-y option: Do not output array indices with data values

September 9, 2008

SPEEDUP Workshop - HDF5 Tutorial

7

www.hdfgroup.org



h5dump – binary output

-b FORMAT option: Binary output, *FORMAT* can be:

MEMORY

- Data exported with datatypes matching memory on system where h5dump is run.

FILE

- Data exported with datatypes matching those in HDF5 file being dumped.

LE

- Data exported with pre-defined little endian datatype.

BE

- Data exported with pre-defined big endian datatype.

- Typically used with `-d dataset -o outputFile` options
 - Allows data values to be exported for use with other applications.
 - When `-b` and `-d` used together, array indices are not output.

September 9, 2008

SPEEDUP Workshop - HDF5 Tutorial

8

www.hdfgroup.org



h5dump – binary output

```
% h5dump -b BE -d /IntArray -o OBE Sample.h5
% od -b OBE | head -2
0000000 000 000 000 000 000 000 000 001 000 000 000 002 000 000 000 003
0000020 000 000 000 004 000 000 000 005 000 000 000 012 000 000 000 013

% h5dump -b LE -d /IntArray -o OLE Sample.h5
% od -b OLE | head -2
0000000 000 000 000 000 001 000 000 000 002 000 000 000 003 000 000 000
0000020 004 000 000 000 005 000 000 000 012 000 000 000 013 000 000 000

% h5dump -b MEMORY -d /IntArray -o OME Sample.h5
% od -b OME | head -2
0000000 000 000 000 000 001 000 000 000 002 000 000 000 003 000 000 000
0000020 004 000 000 000 005 000 000 000 012 000 000 000 013 000 000 000
```



h5dump – properties information

-p option: Print dataset filters, storage layout, fill value

```
% h5dump -p -H Sample.h5
```

```
HDF5 "Sample.h5" {
  GROUP "/" {
    GROUP "Floats" {
      DATASET "FloatArray" {
        DATATYPE H5T_IEEE_F32LE
        DATASPACE SIMPLE {(4, 3)/(4, 3)}
        STORAGE_LAYOUT {
          CONTIGUOUS
          SIZE 48
          OFFSET 3696
        }
        FILTERS {
          NONE
        }
        FILLVALUE {
          FILL_TIME H5D_FILL_TIME_IFSET
          VALUE 0
        }
        ALLOCATION_TIME {
          H5D_ALLOC_TIME_LATE
        }
      }
    }
  }
  ...
}
```



h5import

h5import: loads data into an existing or new HDF5 file

- Data loaded from ASCII or binary files
- Each file corresponds to data values for one dataset
- Integer (signed or unsigned) and float data can be loaded
- Per-dataset settable properties include:
 - datatype (int or float; size; architecture; byte order)
 - storage (compression, chunking, external file, maximum dimensions)
- Properties set via
 - command line

```
% h5import in in_opts [in2 in2_opts] -o out
```

- configuration file

```
% h5import in -c conf1 [in2 -c conf2] -o out
```

September 9, 2008

SPEEDUP Workshop - HDF5 Tutorial

11

www.hdfgroup.org



Example: h5import

Create *Sample2.h5* based on *Sample.h5*

```
% h5dump -d Floats/FloatArray -y Sample.h5
```

```
HDF5 "Sample.h5" {  
  DATASET "/Float/FloatArray" {  
    DATATYPE H5T_IEEE_F32LE  
    DATASPACE SIMPLE {( 4, 3 ) / ( 4, 3 )}  
    DATA {  
      0.01, 0.02, 0.03,  
      0.1, 0.2, 0.3,  
      1, 2, 3,  
      10, 20, 30  
    }  
  }  
}
```

```
% cat config.FloatArray
```

```
PATH /Floats/FloatArray  
INPUT-CLASS TEXTFP  
RANK 2  
DIMENSION-SIZES 4 3
```

```
% cat in.FloatArray
```

```
0.01 0.02 0.03  
0.1 0.2 0.3  
1 2 3  
10 20 30
```

September 9, 2008

SPEEDUP Workshop - HDF5 Tutorial

12

www.hdfgroup.org



Example: h5import

```
% cat config.IntArray
```

```
PATH /IntArray
INPUT-CLASS TEXTIN
RANK 2
DIMENSION-SIZES 5 6
```

```
% cat in.IntArray
```

```
0    1    2    3    4    5
10   11   12   13   14   15
20   21   22   23   24   25
30   31   32   38   34   35
40   41   42   43   44   45
```

➤ Input and configuration files ready; issue command

```
% h5import in.FloatArray -c config.FloatArray \
in.IntArray -c config.IntArray -o Sample2.h5
```

September 9, 2008

SPEEDUP Workshop - HDF5 Tutorial

13

www.hdfgroup.org



h5mkgp

h5mkgp: makes groups in an HDF5 file.

Usage: h5mkgp [OPTIONS] FILE GROUP...

OPTIONS

-h, --help Print a usage message and exit
-l, --latest Use latest version of file format to create groups
-p, --parents No error if existing, make parent groups as needed
-v, --verbose Print information about OBJECTS and OPTIONS
-V, --version Print version number and exit

Example:

```
% h5mkgp Sample2.h5 /EmptyGroup
```

Introduced in HDF5 release 1.8.0.

September 9, 2008

SPEEDUP Workshop - HDF5 Tutorial

14

www.hdfgroup.org



h5diff

h5diff: compares HDF5 files and reports differences

- compare two HDF5 files
 % `h5diff file1 file2`
- compare same object in two files
 % `h5diff file1 file2 object`
- compare different objects in two files
 % `h5diff file1 file2 object1 object2`

Option flags:

- none:** report number of differences found in objects and where they occurred
- r:** in addition, report the differences
- v:** in addition, print list of object(s) and warnings; typically used when comparing two files without specifying object(s)



Example: h5diff

```
% h5diff -v Sample.h5 Sample2.h5
```

```
file1 file2
-----
x x /
  x /EmptyGroup
x x /Floats
x x /Floats/FloatArray
x x /IntArray

group : </> and </>
0 differences found
group : </Floats> and </Floats>
0 differences found
dataset: </Floats/FloatArray> and </Floats/FloatArray>
0 differences found
dataset: </IntArray> and </IntArray>
size:      [5x6]      [5x6]
position   IntArray  IntArray  difference
-----
[ 3 3 ]    33       38       5
```



h5repack

h5repack: copies an HDF5 file to a new file with specified filter and storage layout

- Removes unused space introduced when...
 - Objects were deleted
 - Compressed datasets were updated and no longer fit in original space
 - Full space allocated for variable-length data not used
- Optionally applies filter to datasets
 - gzip, szip, shuffle, checksum
- Optionally applies storage layout to datasets
 - Continuous, chunking, compact



h5repack: filters

-f *FILTER* option: Apply filter, *FILTER* can be:

GZIP	to apply GZIP compression
SZIP	to apply SZIP compression
SHUF	to apply the HDF5 shuffle filter
FLET	to apply the HDF5 checksum filter
NBIT	to apply NBIT compression
SOFF	to apply the HDF5 Scale/Offset filter
NONE	to remove all filters

Compression will not be performed if data is smaller than 1K unless -m flag is used.



h5repack: storage layout

-f LAYOUT option: Apply layout, *LAYOUT* can be:

- CHUNK to apply chunking layout
- COMPA to apply compact layout
- CONTI to apply continuous layout



Example: h5repack (filter)

```
% h5repack -f SHUF -f GZIP=1 TES-Aura.he5 \
TES-rp.he5
```

Tropospheric Emission Spectrometer on Aura, the third of NASA's Earth Observing System's spacecrafts.

Makes global 3-d measurements of ozone and other chemical species involved in its formation and destruction.

```
% ls -sk TES-Aura.he5 TES-rp.he5
```

```
75608 TES-Aura.he5
56808 TES-rp.he5
```

➤ 33% reduction in file size



Example: h5repack (layout)

```
% h5repack -m 1 -l Floats/FloatArray:CHUNK=4x1 \  
Sample.h5 Sample-rp.h5
```

```
% h5dump -p -H Sample-rp.h5
```

```
HDF5 "Sample-rp.h5" {  
  GROUP "/" {  
    GROUP "Floats" {  
      DATASET "FloatArray" {  
        DATATYPE H5T_IEEE_F32LE  
        DATASPACE SIMPLE {(4, 3)/(4, 3)}  
        STORAGE_LAYOUT {  
          CHUNKED (4, 1)  
          SIZE 48  
        }  
        FILTERS {  
          NONE  
        }  
        FILLVALUE {  
          FILL_TIME H5D_FILL_TIME_IFSET  
          VALUE 0  
        }  
        ALLOCATION_TIME {  
          H5D_ALLOC_TIME_INCR  
        }  
      }  
    }  
  }  
  ...  
}
```

September 9, 2008

SPEEDUP Workshop - HDF5 Tutorial

21

www.hdfgroup.org



Performance Tuning & Troubleshooting

- HDF5 tools can assist with performance tuning and troubleshooting
 - Discover objects and their properties in HDF5 files
`h5dump -p`
 - Get file size overhead information
`h5stat`
 - Find locations of objects in a file
`h5ls`
 - Discover differences
`h5diff`, `h5ls`
 - Location of raw data
`h5ls -var`
 - Does file conform to HDF5 File Format Specification?
`h5check`

September 9, 2008

SPEEDUP Workshop - HDF5 Tutorial

22

www.hdfgroup.org



h5stat

h5stat: Prints statistics about HDF5 files

- Reports two types of statistics:
 - High-level information about objects:
 - Number of different objects (groups, datasets, datatypes)
 - Number of unique datatypes
 - Size of raw data
 - Information about object's structural metadata
 - Size of structural metadata (total/free)
 - Object header, local and global heaps
 - Size of B-trees
 - Object header fragmentation



h5stat

- Helps...
 - troubleshoot size overhead in HDF5 files
 - choose appropriate properties and storage strategies
- Usage:
 - `% h5stat -help`
 - `% h5stat file.h5`
- Full specification at :
 - <http://www.hdfgroup.uiuc.edu/RFC/HDF5/h5stat/>

Introduced in HDF5 release 1.8.0.



h5check

- Verifies that a file is encoded according to the *HDF5 File Format Specification*
 - <http://www.hdfgroup.org/HDF5/doc/H5.format.html>
- Does not use the HDF5 library
- Used to confirm that the files written by the HDF5 library are compliant with the specification
- Tool is not part of the HDF5 source code distribution
 - ftp://ftp.hdfgroup.org/HDF5/special_tools/h5check/



Questions?