Exporting to HDF5 Files

Overview

Hierarchical Data Format, Version 5, (HDF5) is a general-purpose, machine-independent standard for storing scientific data in files, developed by the National Center for Supercomputing Applications (NCSA). HDF5 is used by a wide range of engineering and scientific fields that want a standard way to store data so that it can be shared. For more information about the HDF5 file format, read the HDF5 documentation available at the HDF Web site (`http://www.hdfgroup.org`).

MATLAB® provides two methods to export data to an HDF5 file:

- High-level functions that simplify the process of exporting data, when working with numeric datasets
- Low-level functions that provide a MATLAB interface to routines in the HDF5 C library

> **Note** For information about exporting to HDF4 files, which have a separate and incompatible format, see Export to HDF4 Files.

Using the MATLAB High-Level HDF5 Functions to Export Data

The easiest way to write data or metadata from the MATLAB workspace to an HDF5 file is to use these MATLAB high-level functions.

> **Note:** You can use the high-level functions only with numeric data. To write nonnumeric data, you must use the low-level interface.

- `h5create` — Create an HDF5 dataset
- `h5write` — Write data to an HDF5 dataset
- `h5writeatt` — Write data to an HDF5 attribute

For details about how to use these functions, see their reference pages, which include examples. The following sections illustrate some common usage scenarios.

Writing a Numeric Array to an HDF5 Dataset

This example creates an array and then writes the array to an HDF5 file.

1. Create a MATLAB variable in the workspace. This example creates a 5-by-5 array of `uint8` values.

   ```
   testdata = uint8(magic(5))
   ```

2. Create the HDF5 file and the dataset, using `h5create`.

   ```
   h5create('my_example_file.h5', '/dataset1', size(testdata))
   ```

3. Write the data to the HDF5 file.

   ```
   h5write('my_example_file.h5', '/dataset1', testdata)
   ```

Using the MATLAB Low-Level HDF5 Functions to Export Data

MATLAB provides direct access to dozens of functions in the HDF5 library with *low-level* functions that correspond to the functions in the HDF5 library. In this way, you can access the features of the HDF5 library from MATLAB, such as reading and writing complex data types and using the HDF5 subsetting capabilities.

The HDF5 library organizes the library functions into collections, called *interfaces*. For example, all the routines related to working with files, such as opening and closing, are in the H5F interface, where *F* stands for file. MATLAB organizes the low-level HDF5 functions into classes that correspond to each HDF5 interface. For example, the MATLAB functions that correspond to the HDF5 file interface (H5F) are in the @H5F class folder.

The following sections provide more detail about how to use the MATLAB HDF5 low-level functions.

- Map HDF5 Function Syntax to MATLAB Function Syntax
- Map Between HDF5 Data Types and MATLAB Data Types

- Report Data Set Dimensions
- Write Data to HDF5 Data Set Using MATLAB Low-Level Functions
- Write a Large Data Set
- Preserve Correct Layout of Your Data

Note: This section does not describe all features of the HDF5 library or explain basic HDF5 programming concepts. To use the MATLAB HDF5 low-level functions effectively, refer to the official HDF5 documentation available at `http://www.hdfgroup.org`.

Map HDF5 Function Syntax to MATLAB Function Syntax

In most cases, the syntax of the MATLAB low-level HDF5 functions matches the syntax of the corresponding HDF5 library functions. For example, the following is the function signature of the `H5Fopen` function in the HDF5 library. In the HDF5 function signatures, `hid_t` and `herr_t` are HDF5 types that return numeric values that represent object identifiers or error status values.

```
hid_t H5Fopen(const char *name, unsigned flags, hid_t access_id) /* C syntax */
```

In MATLAB, each function in an HDF5 interface is a method of a MATLAB class. The following shows the signature of the corresponding MATLAB function. First note that, because it's a method of a class, you must use the dot notation to call the MATLAB function: `H5F.open`. This MATLAB function accepts the same three arguments as the HDF5 function: a text string for the name, an HDF5-defined constant for the flags argument, and an HDF5 property list ID. You use property lists to specify characteristics of many different HDF5 objects. In this case, it's a file access property list. Refer to the HDF5 documentation to see which constants can be used with a particular function and note that, in MATLAB, constants are passed as text strings.

```
file_id = H5F.open(name, flags, plist_id)
```

There are, however, some functions where the MATLAB function signature is different than the corresponding HDF5 library function. The following describes some general differences that you should keep in mind when using the MATLAB low-level HDF5 functions.

- **HDF5 output parameters become MATLAB return values** — Some HDF5 library functions use function parameters to return data. Because MATLAB functions can return multiple values, these output parameters become return values. To illustrate, the HDF5 `H5Dread` function returns data in the `buf` parameter.

  ```
  herr_t H5Dread(hid_t dataset_id,
                                          hid_t mem_type_id,
                                          hid_t mem_space_id,
                                          hid_t file_space_id,
                                          hid_t xfer_plist_id,
                                          void * buf )   /* C syntax */
  ```

  The corresponding MATLAB function changes the output parameter `buf` into a return value. Also, in the MATLAB function, the nonzero or negative value `herr_t` return values become MATLAB errors. Use MATLAB `try-catch` statements to handle errors.

  ```
  buf = H5D.read(dataset_id,
                                           mem_type_id,
                                                mem_space_id,
                                           file_space_id,
                                                plist_id)
  ```

- **String length parameters are unnecessary** — The length parameter, used by some HDF5 library functions to specify the length of a string parameter, is not necessary in the corresponding MATLAB function. For example, the `H5Aget_name` function in the HDF5 library includes a buffer as an output parameter and the size of the buffer as an input parameter.

  ```
  ssize_t H5Aget_name(hid_t attr_id,
                      size_t buf_size,
                      char *buf ) /* C syntax */
  ```

The corresponding MATLAB function changes the output parameter `buf` into a return value and drops the `buf_size` parameter.

```
buf = H5A.get_name(attr_id)
```

- **Use an empty array to specify NULL** — Wherever HDF5 library functions accept the value `NULL`, the corresponding MATLAB function uses empty arrays (`[]`). For example, the `H5Dfill` function in the HDF5 library accepts the value `NULL` in place of a specified fill value.

```
herr_t H5Dfill(const void *fill,
               hid_t fill_type_id, void *buf,
               hid_t buf_type_id,
               hid_t space_id ) /* C syntax */
```

When using the corresponding MATLAB function, you can specify an empty array (`[]`) instead of `NULL`.

- **Use cell arrays to specify multiple constants** — Some functions in the HDF5 library require you to specify an array of constants. For example, in the `H5Screate_simple` function, to specify that a dimension in the data space can be unlimited, you use the constant `H5S_UNLIMITED` for the dimension in `maxdims`. In MATLAB, because you pass constants as text strings, you must use a cell array to achieve the same result. The following code fragment provides an example of using a cell array to specify this constant for each dimension of this data space.

```
ds_id = H5S.create_simple(2,[3 4],{'H5S_UNLIMITED' 'H5S_UNLIMITED'});
```

Map Between HDF5 Data Types and MATLAB Data Types

When the HDF5 low-level functions read data from an HDF5 file or write data to an HDF5 file, the functions map HDF5 data types to MATLAB data types automatically.

For *atomic* data types, such as commonly used binary formats for numbers (integers and floating point) and characters (ASCII), the mapping is typically straightforward because MATLAB supports similar types. See the table Mapping Between HDF5 Atomic Data Types and MATLAB Data Types for a list of these mappings.

Mapping Between HDF5 Atomic Data Types and MATLAB Data Types

HDF5 Atomic Data Type	MATLAB Data Type
Bit-field	Array of packed 8-bit integers
Float	MATLAB single and double types, provided that they occupy 64 bits or fewer
Integer types, signed and unsigned	Equivalent MATLAB integer types, signed and unsigned
Opaque	Array of `uint8` values
Reference	Array of `uint8` values
String	MATLAB character arrays

For *composite* data types, such as aggregations of one or more atomic data types into structures, multidimensional arrays, and variable-length data types (one-dimensional arrays), the mapping is sometimes ambiguous with reference to the HDF5 data type. In HDF5, a 5-by-5 data set containing a single `uint8` value in each element is distinct from a 1-by-1 data set containing a 5-by-5 array of `uint8` values. In the first case, the data set contains 25 observations of a single value. In the second case, the data set contains a single observation with 25 values. In MATLAB both of these data sets are represented by a 5-by-5 matrix.

If your data is a complex data set, you might need to create HDF5 data types directly to make sure that you have the mapping you intend. See the table Mapping Between HDF5 Composite Data Types and MATLAB Data Types for a list of the default mappings. You can specify the data type when you write data to the file using the `H5Dwrite` function. See the HDF5 data type interface documentation for more information.

Mapping Between HDF5 Composite Data Types and MATLAB Data Types

HDF5 Composite Data Type	MATLAB Data Type

Array	Extends the dimensionality of the data type which it contains. For example, an array of an array of integers in HDF5 would map onto a two dimensional array of integers in MATLAB.
Compound	MATLAB structure. Note: All structures representing HDF5 data in MATLAB are scalar.
Enumeration	Array of integers which each have an associated name
Variable Length	MATLAB 1-D cell arrays

Report Data Set Dimensions

The MATLAB low-level HDF5 functions report data set dimensions and the shape of data sets differently than the MATLAB high-level functions. For ease of use, the MATLAB high-level functions report data set dimensions consistent with MATLAB column-major indexing. To be consistent with the HDF5 library, and to support the possibility of nested data sets and complicated data types, the MATLAB low-level functions report array dimensions using the C row-major orientation.

Write Data to HDF5 Data Set Using MATLAB Low-Level Functions

This example shows how to use the MATLAB® HDF5 low-level functions to write a data set to an HDF5 file and then read the data set from the file.

Open This Example

Create a 2-by-3 array of data to write to an HDF5 file.

```
testdata = [1 3 5; 2 4 6];
```

Create a new HDF5 file named `my_file.h5` in the system temp folder. Use the MATLAB `H5F.create` function to create a file. This MATLAB function corresponds to the HDF5 function, `H5Fcreate`. As arguments, specify the name you want to assign to the file, the type of access you want to the file (`'H5F_ACC_TRUNC'` in this case), and optional additional characteristics specified by a file creation property list and a file access property list. In this case, use default values for these property lists (`'H5P_DEFAULT'`). Pass C constants to the MATLAB function as strings.

```
filename = fullfile(tempdir,'my_file.h5');
fileID = H5F.create(filename,'H5F_ACC_TRUNC','H5P_DEFAULT','H5P_DEFAULT');
```

`H5F.create` returns a file identifier corresponding to the HDF5 file.

Create the data set in the file to hold the MATLAB variable. In the HDF5 programming model, you must define the data type and dimensionality (data space) of the data set as separate entities. First, use the `H5T.copy` function to specify the data type used by the data set, in this case, `double`. This MATLAB function corresponds to the HDF5 function, `H5Tcopy`.

```
datatypeID = H5T.copy('H5T_NATIVE_DOUBLE');
```

`H5T.copy` returns a data type identifier.

Create a data space using `H5S.create_simple`, which corresponds to the HDF5 function, `H5Screate_simple`. The first input, `2`, is the rank of the data space. The second input is an array specifying the size of each dimension of the dataset. Because HDF5 stores data in row-major order and the MATLAB array is organized in column-major order, you should reverse the ordering of the dimension extents before using `H5Screate_simple` to preserve the layout of the data. You can use `fliplr` for this purpose.

```
dims = size(testdata);
dataspaceID = H5S.create_simple(2, fliplr(dims), []);
```

`H5S.create_simple` returns a data space identifier, `dataspaceID`. Note that other software programs that use row-major ordering (such as `H5DUMP` from the HDF Group) might report the size of the dataset to be 3-by-2 instead of 2-by-3.

Create the data set using `H5D.create`, which corresponds to the HDF5 function, `H5Dcreate`. Specify the file identifier, the name you want to assign to the data set, the data type identifier, the data space identifier, and a data set creation property list identifer as arguments. `'H5P_DEFAULT'` specifies the default property list settings.

```
dsetname = 'my_dataset';
datasetID = H5D.create(fileID,dsetname,datatypeID,dataspaceID,'H5P_DEFAULT');
```

`H5D.create` returns a data set identifier, `datasetID`.

Write the data to the data set using `H5D.write`, which corresponds to the HDF5 function, `H5Dwrite`. The input arguments are the data set identifier, the memory data type identifier, the memory space identifier, the data space identifier, the transfer property list identifier and the name of the MATLAB variable to write to the data set. The constant, `'H5ML_DEFAULT'`, specifies automatic mapping to HDF5 data types. The constant, `'H5S_ALL'`, tells `H5D.write` to write all the data to the file.

```
H5D.write(datasetID,'H5ML_DEFAULT','H5S_ALL','H5S_ALL',...
    'H5P_DEFAULT',testdata);
```

Close the data set, data space, data type, and file objects. If used inside a MATLAB function, these identifiers are closed automatically when they go out of scope.

```
H5D.close(datasetID);
H5S.close(dataspaceID);
H5T.close(datatypeID);
H5F.close(fileID);
```

Open the HDF5 file in order to read the data set you wrote. Use `H5F.open` to open the file for read-only access. This MATLAB function corresponds to the HDF5 function, `H5Fopen`.

```
fileID = H5F.open(filename,'H5F_ACC_RDONLY','H5P_DEFAULT');
```

Open the data set to read using `H5D.open`, which corresponds to the HDF5 function, `H5Dopen`. Specify as arguments the file identifier and the name of the data set, defined earlier in the example.

```
datasetID = H5D.open(fileID,dsetname);
```

Read the data into the MATLAB workspace using `H5D.read`, which corresponds to the HDF5 function, `H5Dread`. The input arguments are the data set identifier, the memory data type identifier, the memory space identifier, the data space identifier, and the transfer property list identifier.

```
returned_data = H5D.read(datasetID,'H5ML_DEFAULT',...
    'H5S_ALL','H5S_ALL','H5P_DEFAULT');
```

Compare the original MATLAB variable, `testdata`, with the variable just created, `returned_data`.

```
isequal(testdata,returned_data)
```

```
ans =

     1
```

The two variables are the same.

Write a Large Data Set

To write a large data set, you must use the chunking capability of the HDF5 library. To do this, create a property list and use the `H5P.set_chunk` function to set the chunk size in the property list. Suppose the dimensions of your data set are `[2^16 2^16]` and the chunk size is 1024-by-1024. You then pass the property list as the last argument to the data set creation function, `H5D.create`, instead of using the `H5P_DEFAULT` value.

```
dims = [2^16 2^16];
plistID = H5P.create('H5P_DATASET_CREATE'); % create property list

chunk_size = min([1024 1024], dims); % define chunk size
H5P.set_chunk(plistID, fliplr(chunk_size)); % set chunk size in property list

datasetID = H5D.create(fileID, dsetname, datatypeID, dataspaceID, plistID);
```

Preserve Correct Layout of Your Data

When you use any of the following functions that deal with dataspaces, you should flip dimension extents to preserve the correct layout of the data.

- `H5D.set_extent`
- `H5P.get_chunk`
- `H5P.set_chunk`
- `H5S.create_simple`
- `H5S.get_simple_extent_dims`
- `H5S.select_hyperslab`
- `H5T.array_create`
- `H5T.get_array_dims`