Compressed Datasets

Creating a Compressed Dataset

HDF5 requires you to use chunking to create a compressed dataset. *(To use chunking efficiently, be sure to see the advanced topic, Chunking in HDF5.)*

The following operations are required in order to create a compressed dataset:

- Create a dataset creation property list.
- Modify the dataset creation property list instance to enable chunking and to enable compression.
- Create the dataset.
- Close the dataset creation property list and dataset.

For more information on compression, see the FAQ question on Using Compression in HDF5.

Programming Example

Description

See HDF5 Introductory Examples for the examples used in the Learning the Basics tutorial.

The example creates a chunked and ZLIB compressed dataset. It also includes comments for what needs to be done to create an SZIP compressed dataset. The example then reopens the dataset, prints the filter information, and reads the dataset:

For details on compiling an HDF5 application: Compiling HDF5 Applications

Remarks

- The H5P_SET_CHUNK call modifies a Dataset Creation Property List instance to store a chunked layout dataset and sets the size of the chunks used.
- The H5P_SET_DEFLATE call modifies the Dataset Creation Property List instance to use ZLIB or DEFLATE compression. The H5Pset_szip call modifies it to use SZIP compression. There are different compression parameters required for each compression method.
- SZIP Limitations:
  - SZIP compression can only be used with atomic datatypes that are integer, float, or char. It cannot be applied to compound, array, variable-length, enumerations, or other user-defined datatypes. The call to H5D_CREATE **will fail** if attempting to create an SZIP compressed dataset with a non-allowed datatype. The conflict can only be detected when the property list is used.
  - There are restrictions for use of SZIP by commercial users. See the documents at SZIP Compression in HDF5 for further information on SZIP, including the *SZIP copyright notice*.