File Space Management

This page briefly describes the documentation available to those who use the file space management feature found in the HDF5 library.

The HDF5 library's file space management activities encompass both the allocation of file space and the management of free space. When an HDF5 object (group, dataset, etc.) is created and written, file space is allocated for storing its metadata and raw data. When an object is removed, the space associated with the object becomes free space.

The following are the sections on this page:

- File Space Management Strategies
- File Space Management User's Guide *(not yet available)*
- HDF5 Library APIs
- Tools
- Differences between HDF5-1.10 and HDF5-1.8
- How to Remove the Free Space in an Existing File

File Space Management Strategies

The HDF5 library uses three mechanisms to manage space in an HDF5 file. They are:

- Free-space managers that track free-space sections of various sizes in the file that are not currently allocated.
- Aggregators, which are contiguous blocks of free space in the file.
- Virtual file drivers, which use the virtual file driver interface to request additional space from the file driver associated with the file.

There are four file space-handling strategies available to users that use these mechanisms:

H5F_FSPACE_STRATEGY_FSM_AGGR

This strategy has always been available in HDF5 and is the default. The mechanisms used for this strategy are free-space managers, aggregators, and virtual file drivers.

H5F_FSPACE_STRATEGY_PAGE

The current HDF5 file space allocation accumulates small pieces of metadata and raw data in aggregator blocks. However, these blocks are not page aligned and vary widely in sizes. The paged aggregation feature provides efficient paged access of these small pieces of metadata and raw data. It accumulates metadata and raw data into well-aligned pages called file space pages. The library defines a default file space page size but a user can set the page size via a new public routine.

The mechanisms used for this strategy are free-space managers with embedded paged aggregation and virtual file drivers.

See the RFC on this feature for complete details.

H5F_FSPACE_STRATEGY_AGGR

With this strategy the library will request space from either the metadata or raw data aggregator depending on the file space type. If the request is not satisfied, the library will request space from the virtual file driver.

The mechanisms used for this strategy are aggregators and virtual file drivers. It does not use the free-space manager.

H5F_FSPACE_STRATEGY_NONE

This strategy will request space from the virtual file driver. The only mechanism used is the virtual file driver. It does not use the free-space manager.

File Space Management User's Guide

(This document is not yet available.)

HDF5 Library APIs

The APIs listed below from the *HDF5 Reference Manual* provide a means for users to directly manage the file space management feature.

H5F_GET_FREE_SECTIONS	Retrieves free-space section information for a file
H5F_GET_FREESPACE	Returns the amount of free space in a file
H5F_GET_INFO2	Returns global information for a file
H5P_GET_FILE_SPACE_STRATEGY	Retrieves the File Space Strategy for a file creation property list
H5P_SET_FILE_SPACE_STRATEGY	Sets the File Space Strategy for a file creation property list
H5P_GET_FILE_SPACE_PAGE_SIZE	Retrieves the file space page size for paged aggregation
H5P_SET_FILE_SPACE_PAGE_SIZE	Sets the file space page size for paged aggregation

Tools

The tools listed below have been modified to preserve or modify file freepace information appropriately when processing files employing this feature.

h5dump	When printing the file creation property information for the superblock via the -B option, `h5dump` includes the block size obtained via H5P_GET_FILE_SPACE_PAGE_SIZE.
h5stat	When printing the file space information via the `-S` option, `h5stat` includes the block size obtained via H5P_GET_FILE_SPACE_PAGE_SIZE.
h5repack	The following options were added to `h5repack`: `-G FS_PAGESIZE,--fs_pagesize=FS_PAGESIZE` enables the file space page size to be changed to `FS_PAGESIZE`. `-P FS_PERSIST,--fs_persist=FS_PERSIST` sets the persisting free space to persist (`1`) or to not persist (`0`). `-S FS_STRATEGY, --fs_strategy=FS_STRATEGY` sets the file space management strategy. `-T FS_THRESHOLD, --fs_threshold=FS_THRESHOLD` sets the free-space section threshold.

Differences between HDF5-1.10 vs HDF5-1.8

HDF5-1.10

File space management strategies were introduced via H5P_SET_FILE_SPACE_STRATEGY to manage the unused space in a file.

While a file is open, HDF5 tracks and re-uses the unused space in the file according to the strategy used. If using a strategy that uses the free space manager, then free space can be tracked across file opens by use of the "persist" flag and a minimum free space threshold can be specified. If not specifying a strategy that uses the free space manager, then when the file is closed, any free space is lost and will remain in the file.

HDF5-1.8

File space management only occurs between the HDF5 file open and close, and the free space is NOT tracked beyond file closed. In other words, when you delete a dataset, the space used by the dataset becomes free space that can be re-used as long as the file is open. Once the file is closed, the free space is lost and will remain in the file.

How to Remove the Free Space in an Existing File

The h5repack utility can be used to remove the unused space in a file, by writing the file to a new file. This utility comes with the HDF5 binary distribution.