HDF5 File Space Management

1. Introduction

The space taken up by an HDF5 file is called its *file space*. When a user first creates an HDF5 file, the HDF5 library immediately allocates space to store information called file metadata. *File metadata* is information the library uses to describe the HDF5 file and to identify its associated objects. When a user subsequently writes data to HDF5 dataset objects, the HDF5 library allocates space to store the data values, as well as the necessary additional file metadata. When a user removes HDF5 objects from an HDF5 file, the space associated with those objects becomes *free space*. The HDF5 library manages this free space.

The HDF5 library *file space management* activities encompass the allocation of space and the management of free space. The HDF5 library implements several *file space management strategies*, and the strategy used for a given HDF5 file is set when the file is created. Depending on the file’s usage patterns, one strategy may be better than the others; an inappropriate strategy can lead to storage and access performance issues. HDF5 files that will have objects added or deleted in later sessions, or that will never have objects deleted, may benefit from the use of a non-default strategy.

This document describes how the file space management strategies affect storage space and access time for various HDF5 file usage patterns. It also presents the HDF5 utilities and HDF5 library public routines that help users select appropriate file space management strategies for their specific needs.