3. HDF5 File Space Allocation and Tuning

Audience:

An HDF5 application developer who has knowledge of the HDF5 library API.

The HDF5 library performs file space management activities related to tracking free space and allocating space to store file metadata and *raw data*, the data values in HDF5 dataset objects. Every HDF5 file has an associated file space management strategy that determines how the HDF5 library carries out these activities for the file.

3.1 File Space Allocation

The HDF5 library includes three different mechanisms for allocating space to store file metadata and raw data:

* Free-Space Managers

The HDF5 library’s free-space managers track sections in the HDF5 file that are not being used to store file metadata or raw data. These sections will be of various sizes. When the library needs to allocate space, the free-space managers search the tracked free space for a section of the appropriate size to fulfill the request. If a suitable section is found, the allocation can be made from the file’s existing free space.

* Aggregators

The HDF5 library has two aggregators. Each aggregator manages a block of contiguous bytes in the file that have not been allocated previously. One aggregator allocates space for file metadata from the block it manages; the other aggregator handles allocations for raw data. The maximum number of bytes in each aggregator’s block is tunable.

If the library’s allocation request exceeds the maximum number of bytes an aggregator’s block can contain, the aggregator cannot fulfill the request and the request falls through to the virtual file driver level. After space has been allocated from an aggregator’s block, that space it is no longer managed by the aggregator. Unallocated bytes continue to be managed by the aggregator.

When an aggregator cannot fulfill an allocation request from the remaining space in its block, it requests a new block of contiguous bytes and any unallocated blocks that remain in the existing block become free space.

* Virtual File Driver

The HDF5 library’s virtual file driver dispatches requests for additional space to the allocation routine of the file driver associated with an HDF5 file. For example, if the H5FD\_SEC2 file driver is being used, its allocation routine will increase the size of the single file on disk that stores the HDF5 file contents to accommodate the additional space that was requested.

File Space Management Strategies

The HDF5 library provides several file space management strategies that control how it tracks free space and uses the free-space managers, aggregators, and virtual file driver to allocate space for file metadata and raw data. The strategies are:

|  |  |
| --- | --- |
| Use all space allocation mechanisms. Track file free space across sessions. | H5F\_FILE\_SPACE\_ALL\_PERSIST (or ALL\_PERSIST) |
| Use all space allocation mechanisms. Track file free space only in current session. | H5F\_FILE\_SPACE\_ALL (or ALL) |
| Use only aggregator and VFD mechanisms.  Never track free space. | H5F\_FILE\_SPACE\_AGGR\_VFD (or AGGR\_VFD) |
| Use only VFD mechanism.  Never track free space. | H5F\_FILE\_SPACE\_VFD (or VFD) |

**Strategy 1: H5F\_FILE\_SPACE\_ALL\_PERSIST (*also called* ALL\_PERSIST)**

With this strategy, the HDF5 library’s free-space managers track the free space that results from the manipulation of HDF5 objects in an HDF5 file. The tracked free space information is saved when the HDF5 file is closed, and reloaded when the file is re-opened. The tracked free space information **persist**s across HDF5 file sessions, and the free space managers remain aware of free space sections that became available in any file session.

With this strategy, when space is needed for file metadata or raw data, the HDF5 library first requests space from the free-space managers. If the request is not satisfied, the library requests space from the aggregators. If the request is still not satisfied, the library requests space from the virtual file driver. That is, the library will use **all** of the mechanisms for allocating space.

The H5F\_FILE\_SPACE\_ALL\_PERSIST strategy offers every possible opportunity for reusing free space. The HDF5 file will contain extra file metadata information about tracked free space. The HDF5 library will perform additional “accounting” operations to track free space, and to search the free space sectors when allocating space for file metadata and raw data.

**Strategy 2: H5F\_FILE\_SPACE\_ALL (*also called* ALL)**

This strategy is the HDF5 library’s default file space management strategy. Prior to HDF5 Release 1.9.x, it was the only file space management strategy directly supported by the library.

With this strategy, the HDF5 library’s free-space managers track the free space that results from the manipulation of HDF5 objects in an HDF5 file. The free space managers are aware of free space sections that became available in the current file session, but the tracked free space information is not saved when the HDF5 file is closed. Free space that exists when the file is closed becomes unaccounted space in the HDF5 file. Unallocated space in the aggregators’ blocks may also become unaccounted space when the session ends.

As with the strategy ALL\_PERSIST, the library will try **all** of the mechanisms for allocating space with the ALL strategy. When space is needed for file metadata or raw data, the HDF5 library first requests space from the free-space managers. If the request is not satisfied, the library requests space from the aggregators. If the request is still not satisfied, the library requests space from the virtual file driver.

The H5F\_FILE\_SPACE\_ALL strategy allows free space incurred in the current session to be reused in the current session. There is no extra file metadata information about tracked free space in the HDF5 file. However, if free space exists when the file is closed the HDF5 file will contain unaccounted space that can never be reused. The HDF5 library will perform some additional “accounting” operations to track free space, but the amount of free space tracked and searched will usually be less than with the ALL\_PERSIST strategy, so the number of operations should be less.

**Strategy 3: H5F\_FILE\_SPACE\_AGGR\_VFD (*also called* AGGR\_VFD)**

With this strategy, the HDF5 library does not track the free space that results from the manipulation of HDF5 objects in an HDF5 file. All free space immediately becomes unaccounted space. Unallocated bytes in the aggregators’ blocks when the file is closed may also become unaccounted space.

With this strategy, when space is needed for file metadata or raw data, the HDF5 library first requests space from the aggregators. If the request is not satisfied, the library requests space from the virtual file driver. That is, the library will try the **aggr**egator and **v**irtual **f**ile **d**river mechanisms for allocating space.

The H5F\_FILE\_SPACE\_AGGR\_VFD strategy never reuses free space. Because small allocation requests can be satisfied from the aggregators’ blocks of contiguous bytes, this strategy will deliver better access performance for some file usage patterns. It may be appropriate when access performance is the highest priority and there are many small writes. Because there are different aggregators for file metadata and raw data, this strategy tends to co-locate file metadata more than some other strategies that can reuse free space scattered throughout the file.

**Strategy 4: H5F\_FILE\_SPACE\_VFD (*also called* VFD)**

With this strategy, the HDF5 library does not track the free space that results from the manipulation of HDF5 objects in an HDF5 file. All free space immediately becomes unaccounted space.

With this strategy, when space is needed for file metadata or raw data, the HDF5 library requests space from the **v**irtual **f**ile **d**river.

The H5F\_FILE\_SPACE\_VFD strategy never reuses free space. Because allocation requests go directly to the virtual file driver, this strategy is best suited for HDF5 files whose primary file usage pattern consists of writing large amounts of raw data to extend dataset object(s).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Summary of File Space Management Strategies** | | | | | | |
| Full Name  *used with H5Pset\_file\_space* | Short Name  *used with HDF5 repack* | Track Free Space | | Allocate Space Using | | |
| across multiple sessions | within single session | free-space managers | aggregators | virtual file driver |
| H5F\_FILE\_SPACE\_ALL\_PERSIST | ALL\_PERSIST | Y | Y | Y | Y | Y |
| H5F\_FILE\_SPACE\_ALL | ALL | N | Y | Y | Y | Y |
| H5F\_FILE\_SPACE\_AGGR\_VFD | AGGR\_VFD | N | N | N | Y | Y |
| H5F\_FILE\_SPACE\_VFD | SPACE\_VFD | N | N | N | N | Y |

Specifying a File Space Management Strategy

The strategy for a given HDF5 file is specified when the file is created; it cannot be changed thereafter.

The HDF5 library provides the *H5Pset\_file\_space* file creation property routine so that users can specify the file space management strategy that should be used when a new HDF5 file is created (see entry in HDF5 Reference Manual). The signature for the routine is:

*herr\_t H5Pset\_file\_space(hid\_t fcpl\_id, H5F\_file\_space\_t strategy, hsize\_t threshold)*

There first *H5Pset\_file\_space* parameter, *fcpl\_id,* is the file creation property list identifier that will be used when the HDF5 file is created. The second parameter, *strategy,* is one of the four strategies described above. The third parameter, *threshold,* is the free-space section threshold used by the library’s free-space managers. This parameter is mainly for performance tuning purposes, and is discussed in more detail elsewhere. Passing a value of zero for either *strategy* or *threshold* indicates that the file’s corresponding existing value should not be modified as a result of the call.

The library provides a companion routine that retrieves the file space management information for an HDF5 file (see entry in HDF5 Reference Manual):

*herr\_t H5Pget\_file\_space*(*hid\_t* *fcpl\_id*, *H5F\_file\_space\_t \* strategy*, hsize\_t \**threshold*)

The first parameter, *fcpl\_id,* is the file creation property list identifier associated with the HDF5 file. If the second parameter, *strategy,* is not NULL, the library will retrieve the existing file space management strategy in use for the file and store it in *strategy*. If the third parameter, *threshold,* is not NULL, the library will retrieve the existing free-space section threshold used by the library’s free-space manager and store it in *threshold*.

The following code sample shows how these two public routines are used to create an empty HDF5 file, *persist.h5,* with the file space management strategy ALL\_PERSIST:

/\* Create a file creation property list template \*/

fcpl\_id = H5Pcreate(H5P\_FILE\_CREATE);

/\* Set the file space management strategy \*/

/\* Don’t update the free-space section threshold \*/

H5Pset\_file\_space(fcpl\_id, H5P\_FILE\_SPACE\_ALL\_PERSIST, (hsize\_t)0);

/\* Create an HDF5 file with the file creation property list *fcpl\_id* \*/

fid = H5Fcreate(“persist.h5”, H5F\_ACC\_TRUNC, fcpl\_id, H5P\_DEFAULT);

/\* The strategy retrieved will be #1 H5F\_FILE\_SPACE\_ALL\_PERSIST \*/

/\* The threshold retrieved will be 1 which is the library default \*/

H5Pget\_file\_space(fcpl\_id, &strategy, &threshold);

/\* Close the file \*/

H5Fclose(fid);

The *h5dump* command line utility reports the file space management information for an HDF5 file. See the following *h5dump* *–B* output for the file *persist.h5*:

HDF5 "persist.h5" {

SUPER\_BLOCK {

SUPERBLOCK\_VERSION 2

:

:

:

FILE\_SPACE\_STRATEGY H5F\_FILE\_SPACE\_ALL\_PERSIST

FREE\_SPACE\_THRESHOLD 1

}

:

:

The output indicates that the HDF5 library will use the file space management strategy H5F\_FILE\_SPACE\_ALL\_PERSIST and a free-space section threshold of 1 (which is the default) when performing free space management activities on the HDF5 file *persist.h5.*

## 3.2 Tuning File Space Management

Each of the four file space management strategies has benefits and drawbacks. The appropriate strategy depends on the HDF5 file’s usage pattern. In this section we cover the pros and cons of the various strategies in more detail, and use additional scenarios to demonstrate their effect on file size.

Recall the two HDF5 files used in 1\_Primer, *no\_persist\_A.h5* and *persist\_B.h5*. By using the default file creation property identifier (H5P\_DEFAULT) when creating *no\_persist\_A.h5*, the HDF5 librarywill automatically use the file space management strategy H5F\_FILE\_SPACE\_ALL for the file. The code sample in the previous section demonstrated how to create the file *persist.h5* that would be managed using the H5F\_FILE\_SPACE\_ALL\_PERSIST strategy *.*

The prior sections have shown that strategy ALL\_PERSIST has the benefit of reusing the tracked free space in the file across multiple file sessions, while strategy ALL has the drawback of accumulating unaccounted space in the file over multiple sessions. The key factor contributing to the benefit of strategy ALL\_PERSIST is the usage pattern of manipulating (adding/deleting) HDF5 objects across multiple sessions. The fragmentation and unaccounted space with strategy ALL increases with the manipulation of HDF5 objects across sessions.

The H5F\_FILE\_SPACE\_AGGR\_VFD and H5F\_FILE\_SPACE\_VFD strategies never use the HDF5 library’s free-space manager to track released file space. Therefore, any fragments of unused space that result from the manipulation of HDF5 objects will be unaccounted space that can never be reused. For the AGGR\_VFD and VFD strategies, the number of sessions in which manipulations occur has negligible (AGGR\_VFD) or no (VFD) effect on the file size.

Discussions of the scenarios presented earlier in the Primer are expanded below, and HDF5 files with AGGR\_VFD and VFD management policies are also shown. The datasets in this section are identical to those used in the previous scenarios.

Scenario C: ALL\_PERSIST Strategy in Single Session

### Session 1: Create file, manipulate objects

In the only session of this scenario, a user creates an HDF5 file named *persist\_C.h5* using the H5F\_FILE\_SPACE\_ALL\_PERSIST strategy. The user then adds four datasets (*dset1, dset2, dset3,* and *dset4*), deletes *dset2*, and adds *dset5* before closing the file.

The file management strategy is the same strategy that was used in Scenario B. The HDF5 objects are manipulated in the same order as they were in Sessions 1-3 of Scenario B.

*h5stat –S* for *persist\_C.h5* shows the following:

Filename: ./persist\_C.h5

Summary of file space information:

File metadata: 2409 bytes

Raw data: 4640 bytes

Amount/Percent of tracked free space: 117854 bytes/94.4%

Unaccounted space: 0 bytes

Total space: 124903 bytes

The file size for *persist\_C.h5* is about 4000 bytes smaller than the file size for *persist\_B.h5* after Session 3 of Scenario B. This is because there are some space savings, in both free space and file metadata (fewer free space sections to track), when the HDF5 object manipulations occur in a single session.

Scenario D: ALL Strategy in Single Session

### Session 1: Create file, manipulate objects

In the only session of this scenario, a user creates an HDF5 file named *no\_persist\_D.h5* using the H5F\_FILE\_SPACE\_ALL strategy. The user then adds four datasets (*dset1, dset2, dset3,* and *dset4*), deletes *dset2*, and adds *dset5* before closing the file.

The file management strategy is the same strategy that was used in Scenario A. The HDF5 objects are manipulated in the same order as they were in Sessions 1-3 of Scenario A.

*h5stat –S* for *no\_persist\_D.h5* shows the following:

Filename: ./no\_persist\_D.h5

Summary of file space information:

File metadata: 2216 bytes

Raw data: 4640 bytes

Amount/Percent of tracked free space: 0 bytes/0.0%

Unaccounted space: 117976 bytes

Total space: 124832 bytes

The file size for *no\_persist\_D.h5* is about 6000 bytes smaller than the file size for *no\_persist\_A.h5* after Session 3 of Scenario A. This is because the HDF5 library was able to reuse some of the free space it was tracking when all of the object manipulations took place in a single session. *no\_persist\_D.h5*, created in Scenario D, still has a substantial amount of unaccounted space (117976 bytes) – almost 95% of the total file space.

Comparing file space information for *persist\_C.h5* (Scenario C)and *no\_persist\_D.h5* (Scenario D)*, no\_persist\_D.h5* is a bit smaller. For both files, the library’s free-space manager tracks the free space resulting from the deletion of *dset2,* and reuses the free space for the addition of *dset5*. Looking at the size of the file metadata for the two files, the greater amount of file metadata in *persist\_C.h5* is due to the extra metadata needed to keep free space information persistent when the file is closed. This demonstrates that using strategy ALL, as was done for *no\_persist\_D.h5,* has some saving in file space compared to strategy ALL\_PERSIST when the HDF5 object manipulation occurs in a single session. The exact amount of space savings will depend on the number and size of HDF5 objects that are added and deleted, as well as on the value of the free-space section threshold and other advanced tuning parameters.

Scenario E: AGGR\_VFD Strategy in Single Session

### Session 1: Create file, manipulate objects

In the only session of this scenario, a user creates an HDF5 file named aggrvfd\_E*.h5* using the H5F\_FILE\_SPACE\_AGGR\_VFD strategy. The user then adds four datasets (*dset1, dset2, dset3,* and *dset4*), deletes *dset2*, and adds *dset5* before closing the file.

*h5stat –S* output shows:

Filename: ./aggrvfd\_E.h5

Summary of file space information:

File metadata: 2208 bytes

Raw data: 4640 bytes

Amount/Percent of tracked free space: 0 bytes/0.0%

Unaccounted space: 121936 bytes

Total space: 128784 bytes

Scenario F: VFD Strategy in Single Session

### Session 1: Create file, manipulate objects

In the only session of this scenario, a user creates an HDF5 file named vfd\_F*.h5* using the VFD strategy. The user then adds four datasets (*dset1, dset2, dset3,* and *dset4*), deletes *dset2*, and adds *dset5* before closing the file.

*h5stat –S* output shows:

Filename: ./vfd\_F.h5

Summary of file space information:

File metadata: 2208 bytes

Raw data: 4640 bytes

Amount/Percent of tracked free space: 0 bytes/0.0%

Unaccounted space: 120272 bytes

Total space: 127120 bytes

Comparison of HDF5 Files from Scenarios A-F after HDF5 Object Manipulation

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Scenario /  # Sessions | Strategy | File Name | File Size  (bytes) | File Metadata  (bytes) | Raw Data (bytes) | Tracked Free Space (bytes) | Unaccounted Space (bytes) |
| A / 3 | ALL | no\_persist\_A.h5 | 130880 | 2216 | 4640 | 0 | 124024 |
| B / 3 | ALL\_PERSIST | persist\_B.h5 | 128921 | 2427 | 4640 | 121854 | 0 |
| C / 1 | ALL\_PERSIST | persist\_C.h5 | 124903 | 2409 | 4640 | 117854 | 0 |
| D / 1 | ALL | no\_persist\_D.h5 | 124832 | 2216 | 4640 | 0 | 117976 |
| E / 1 | AGGR\_VFD | aggrvfd\_E.h5 | 128784 | 2208 | 4640 | 0 | 121936 |
| F / 1 | VFD | vfd\_F.h5 | 127120 | 2208 | 4640 | 0 | 120272 |

Files *no\_persist\_A.h5* and *persist\_B.h5,* which were written over three sessions, have the largest file sizes. Since the unused space in *persist\_B.h5* is tracked free space, it may be reused in later sessions if more HDF5 objects are added to the file, or if new data values are added to existing dataset objects.

The file sizes of *aggrvfd\_E.h5* and *vfd\_F.h5* are larger than *persist\_C.h5* and *no\_persist\_D.h5* – files that were also created in a single session. This is because strategies AGGR\_VFD and VFD do not track free space, even within a single session, and therefore do not reuse any space that is released as HDF5 objects are manipulated. *aggrvfd\_E.h5* is larger than *vfd\_F.h5* because bytes in the aggregators’ blocks have become unaccounted in the process of managing space. The VFD strategy does not use the aggregators, but allocates space directly from the file driver.

The final Scenarios G and H illustrate that the strategies AGGR\_VFD and VFD have the benefit of saving file space when the usage pattern is adding HDF5 objects without deletion. They may also be faster, because no time is spent tracking free space in the file.

Scenario G: AGGR\_VFD Strategy in Single Session, no Objects Deleted

### Session 1: Create file, add objects

In the only session of this scenario, a user creates an HDF5 file named aggrvfd\_G*.h5* using the H5F\_FILE\_SPACE\_AGGR\_VFD strategy. The user then adds four datasets (*dset1, dset2, dset3,* and *dset4*) and closes the file.

*h5stat –S* shows:

Filename: ./aggrvfd\_G.h5

Summary of file space information:

File metadata: 2208 bytes

Raw data: 120640 bytes

Amount/Percent of tracked free space: 0 bytes/0.0%

Unaccounted space: 1936 bytes

Total space: 124784 bytes

Scenario H: VFD Strategy in Single Session, no Objects Deleted

### Session 1: Create file, add objects

In the only session of this scenario, a user creates an HDF5 file named vfd\_H*.h5* using the H5F\_FILE\_SPACE\_VFD strategy. The user then adds four datasets (*dset1, dset2, dset3,* and *dset4*) and closes the file.

*h5stat –S* shows:

Filename: ./vfd\_H.h5

Summary of file space information:

File metadata: 2208 bytes

Raw data: 120640 bytes

Amount/Percent of tracked free space: 0 bytes/0.0%

Unaccounted space: 0 bytes

Total space: 122848 bytes

Comparison of HDF5 Files from Scenarios A, B, G, H after HDF5 Objects Added

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Scenario /  # Sessions | Strategy | File Name | File Size  (bytes) | File Metadata  (bytes) | Raw Data (bytes) | Tracked Free Space (bytes) | Unaccounted Space (bytes) |
| A / 2 | ALL | no\_persist\_A.h5 | 124832 | 2216 | 120640 | 0 | 1976 |
| B / 2 | ALL\_PERSIST | persist\_B.h5 | 124885 | 2319 | 120640 | 1854 | 0 |
| G / 1 | AGGR\_VFD | aggrvfd\_G.h5 | 124784 | 1936 | 120640 | 0 | 1936 |
| H / 1 | VFD | vfd\_H.h5 | 122848 | 2208 | 120640 | 0 | 0 |

The table above shows the file space information for HDF5 files after four datasets have been added. This corresponds to the state of the files after Session 2 in Scenarios A and B.

The *aggrvfd\_G.h5* and *vfd\_H.h5* files are smaller than *no\_persist\_A.h5* and *persist\_B.h5*. The HDF file *vfd\_H.h5,* managed with the VFD strategy, has the smallest size with no tracked free space or unaccounted space. Even though the file *aggrvfd\_G.h5* has less saving in file space than *vfd\_H.h5*, it will have the benefit of better I/O performance due to the use of aggregators for servicing space allocation requests. Metadata in *aggrvfd\_G.h5* will also tend to be more concentrated in contiguous blocks than in *vfd\_H.h5*.

The section *Performance Report for File Space Management* provides more information about selecting file space management strategies to optimize access performance.

Particularly aggravated when all within one setting all\_one\_setting () (~4k, 2k) or

add\_close\_adddelete() (this one is worser?hm…not really diff 2k but I have the bug in this one for VFD) : for #3, then #4

Bug in add\_close\_adddelete() for VFD