

HPC-HDF5 - Feature # 619: Create new 'single chunk' indexing method for chunked datasets

Status:	New	Priority:	Normal
Author:	Quincey Koziol	Category:	
Created:	02/17/2011	Assigned to:	
Updated:	02/18/2011	Due date:	
Impact:	3 - Medium		
Expected Use:	4 - Common		
OS:	All		
Support Group:	Any		
Subject:	Create new 'single chunk' indexing method for chunked datasets		
Description:	When a dataset's dimensions are fixed, and the same as the chunk dimensions, the library "knows" that the dataset will only have one chunk. So, the index can be eliminated entirely, and the information about the chunk can be moved directly into the object header message, eliminating an I/O access (to access the index) and saving some space in the file.		

History

02/18/2011 11:48 am - Mark Miller

Based on description, I could see caller's interface to invoking this behavior being implemented a few ways. The sort of `_passive_` way is just to have HDF5 lib `_watch_` dataset creation and chunk property settings, if any, and when dataset is fixed size and chunk is same size as dataset, then this `_single chunk_` simple indexing mechanism magically happens. That's an ok approach I guess but it's also prone to error as caller needs to make two calls with equivalent parameters to invoke the single chunk mechanism.

So, I'd suggest an explicit property, something like a new layout property "SINGLE_CHUNK" or maybe a new `H5Pset_single_chunk()` method. The latter gives you an opportunity then also to document why this approach might be useful as well as the `_partial readback_` pitfalls it may introduce.