HDF5 Hands-On

Type description

Gerd Heber, The HDF Group
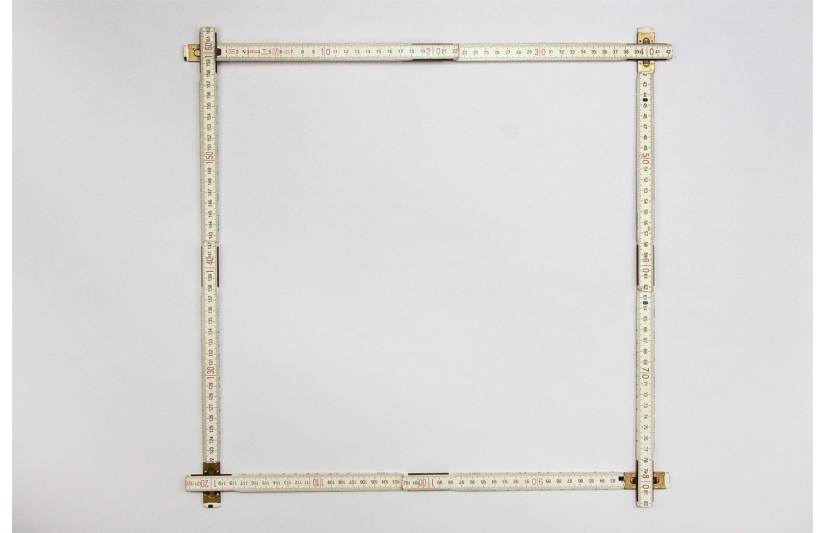
Outline

- Time is special
- Portable performance (?)
- Parallel HDF5
- **h5repack**
- Visualization
- Chunking
- Diagnosing and fixing performance problems

Time is special

- How to reach high-performance when reading large arrays whose first dimension is the time?
  - Reading the full array
  - Reading a time-slice
- Under what circumstances would one split the data into separate datasets (different chunk sizes) or even separate files?
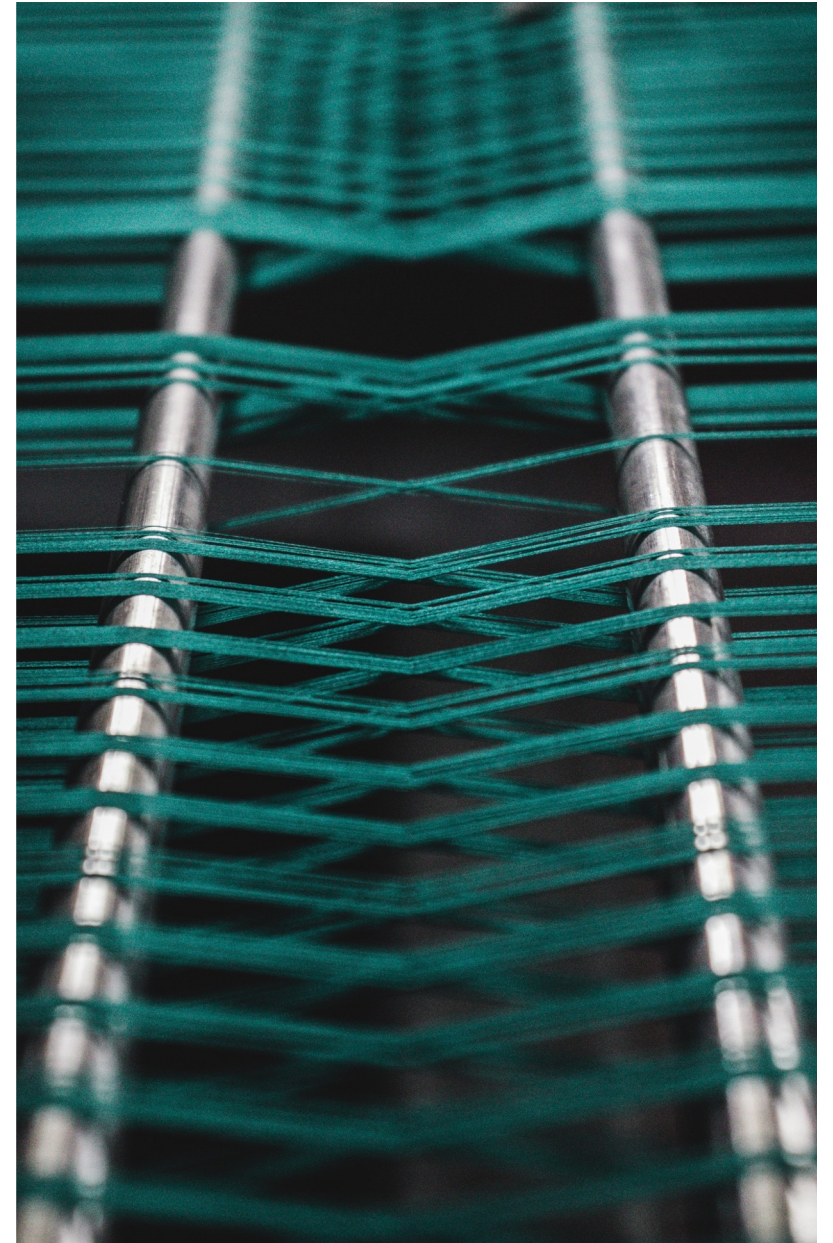- What are others doing? What's the sweet spot?

Portable Performance (?)

How to ensure high read (not necessarily write-) performance when sharing HDF5 files between machines/institutes?

Parallel HDF5

- What is parallel HDF5?
- Are parallel HDF5 files different from regular HDF5 files?

h5repack

What does it do? When and why would I use it?

Visualizing HDF5 Data

- Visualizing the structure of HDF5 files
- Visualizing the "meaning" of data stored in HDF5 files

Chunking

- Why and when to use a chunked layout?
- Is there a guide/algorithm to determine an optimal chunk size?
- How does chunk size depend on datatype and access patterns?

Diagnostics

- How to diagnose I/O performance problems?
- How to fix problems for good?