HDF5 Roadmap
A Community Discussion

The HDF Group

October 13, 2021

Outline

- Update on Hierarchical Data Format version 5 (HDF5) software
- What is on your HDF5 wish list?

HDF5

- Release information, source code and Linux, macOS, Windows binaries, and compression plugins are available from [The HDF Group (THG) Portal]()

- We support HDF5 1.8.*, 1.10.* and 1.12.* maintenance releases
  - The latest releases are 1.8.22, 1.10.7, 1.12.1
  - Upcoming releases are 1.10.8 (imminent) and 1.8.23 in Fall 2021
    - Starting with HDF5 1.10.7 [szip]() compression was replaced with [aec]() (Open Source, BSD license)

- New series of "unstable" HDF5 1.13.* releases

HDF5 1.13.* Releases

- "Unstable" release – our work toward new major release
  - APIs subject to change
  - File format may change
  - No binary compatibility guarantees

- "Unstable" doesn't mean "buggy"
  - We deliver new features and would like you to try them ASAP
  - Please send us feedback if something could be improved!

- In the develop (default) branch in the hdf5 repository on GitHub
  - There will not be an hdf5_1_13 branch
  - All 1.13 releases will split off of the develop branch

- HDF5 1.14.0 (stable) targeted for late 2022

HDF5 1.13.* Releases (cont'd)

- Assumes modern compilers and systems
  - C99
  - C++11 (if building C++ wrappers)
  - Targeted at current POSIX platforms
  - Visual Studio 2015+ (due to C99 needs)
  - Fewer checks for obsolete things than in 1.12/1.10

- Memory sanity checks OFF by default
  - Heap canaries cause problems with buffer (re)allocation/free

HDF5 1.13.0 Features

- Release candidate can be found in
  https://gamma.hdfgroup.org/ftp/pub/outgoing/hdf5/snapshots/v113/
  HDF5-1.13.0-rc6.tar

- Asynchronous HDF5
  - New async versions of many API calls
  - New event set (H5ES) API calls
  - Requires an async-capable VOL connector (NOT in the native connector)

- Updated VOL connectors APIs
  - Webinar on 9/24/2021 covered both updated
  https://www.hdfgroup.org/2021/09/webinar-followup-new-features-in-the-hdf5-1-13-0-release/

- Pluggable VFDs

HDF5 1.13.0 Features in Works

The HDF Group

- Updates to HDF5 in support of HDF5 VOLs (DAOS, Async, and Cache VOLs)
  - Improvements to HDF5 to support Cache VOL LBNL's HDF5
  - Dimension Scale APIs with using new references

- h5dwalk tool
  - Invoking parallel instances of a serial tool, the dh5walk application can accept directories as input arguments.

  ```
  mpiexec -n 4 ./dh5walk -o show-h5dump-h5files.log -T ./h5dump <dir_name>
  ```
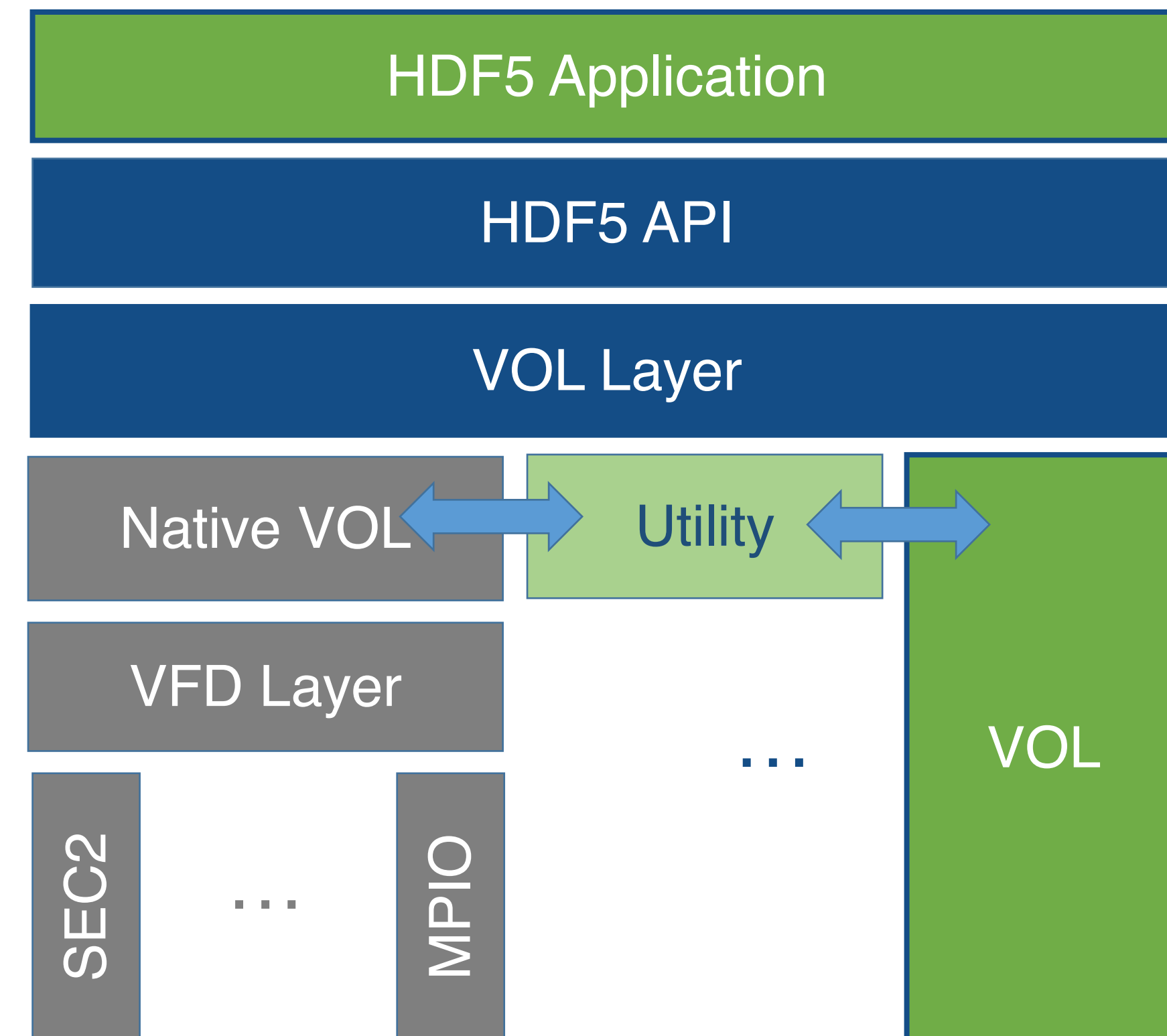
HDF5 1.13.*

- VFD SWMR, provenance (aka onion) VFD
- Selection I/O
- Subfiling VFD
- Performance improvements to close performance gap between 1.8.* and the later releases
- Parallel compression improvements

- This development will complete the work for HDF5 1.14.0 (Fall 2022)
- What is next?

2023 – 2024 Wish List

- Assess the state of HDF5 software
- Look into restructuring of VOL layer and Native VOL
  - Refactor "Utility" packages (ID handling, datatypes, selections, error handling) to be used by all VOL connectors
  - Create **real** native HDF5 VOL
- Remove redundant code (e.g.,SWMR)
- Rework the code to prevent library's bad behavior when reading corrupted Data
- Address MD performance problems

HDF5 Features in Pipeline

- HDF5 file robustness
  - H5nurse/h5doctor tools
- Support for sparse data
- Multi-threaded HDF5
- Full UTF-8 support
- Misc. drivers:
  - mmap driver
  - RW S3
- Data streaming
- Efficient handling of variable-length datatypes
- Performant sub-setting

Contribute

- HDF software is on GitHub
- HDF5 documentation is in Doxygen starting with HDF5 1.12.1 release
  - http://docs.hdf5.info/hdf5/v1_12/index.html
  - http://docs.hdf5.info/hdf5/develop/index.html
  - Searchable, indexed by Google
  - Help us to improve HDF5 documentation!

THANK YOU!

Questions & Comments?