The State of HDF
2024 HDF User Group

Chicago, IL
5 August, 2024

Dana Robinson
Director of Engineering
The HDF Group

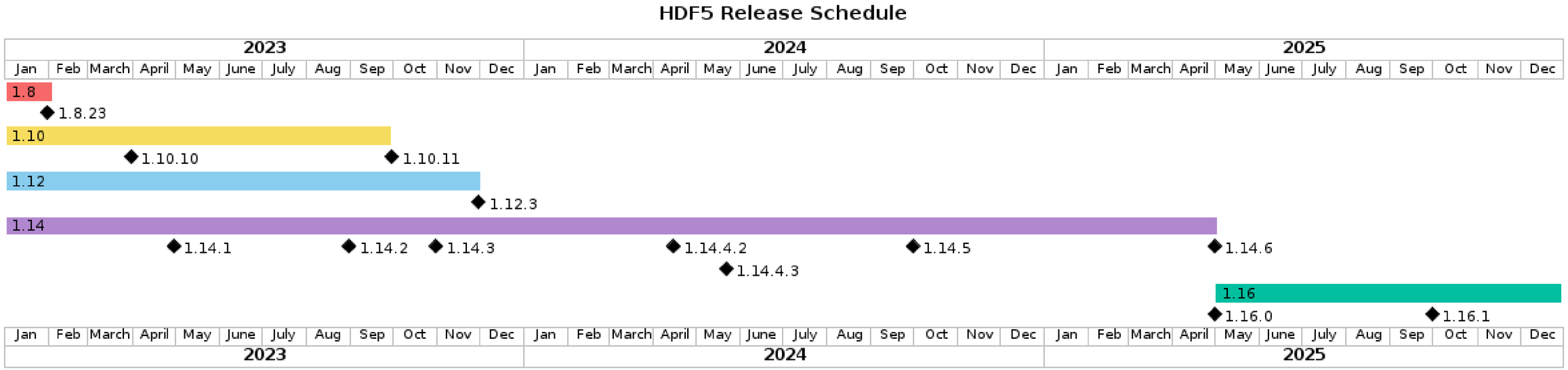
Contents

The HDF Group

- HDF Product Status
- Community Engagement
- Future Plans

HDF Product Status

HDF5 Release Schedule

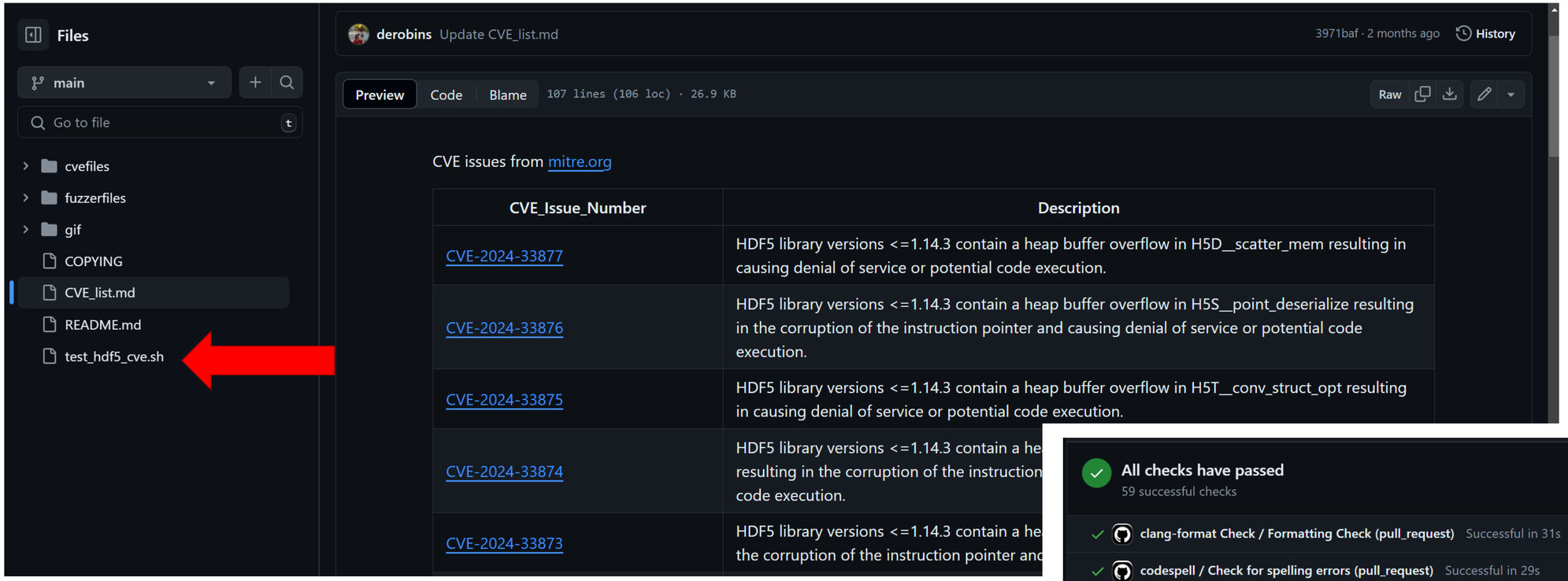


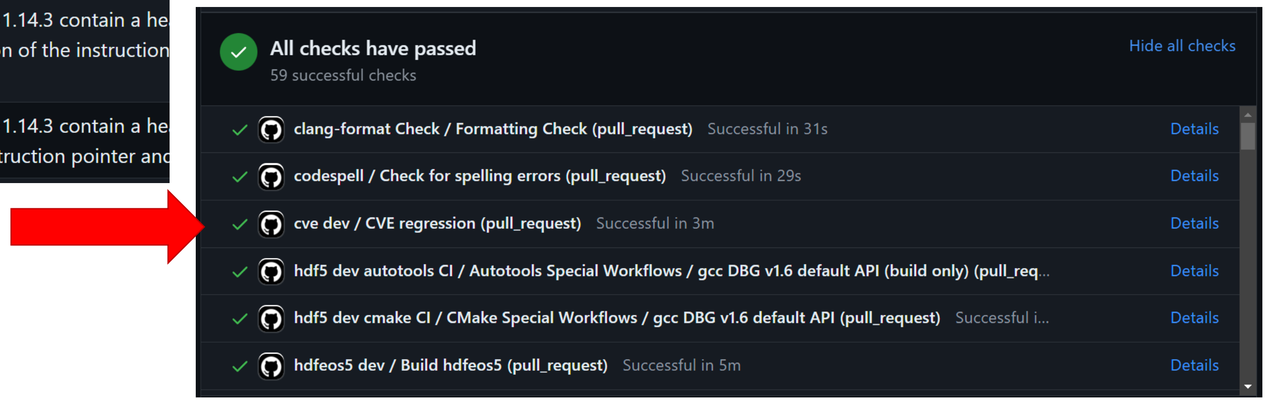
- Latest release: 1.14.4 (April 2024)
  - _Float16 support
  - Cloud-optimized HDF5 tweaks (relaxed page buffer checks, temp security credentials)
  - Chunk offsets returned from API calls now always take the user block into account
  - Improved datatype conversion performance
  - Many CVE fixes (35 - https://github.com/HDFGroup/cve_hdf5/blob/main/CVE_list.md)

HDF5 CVE repository

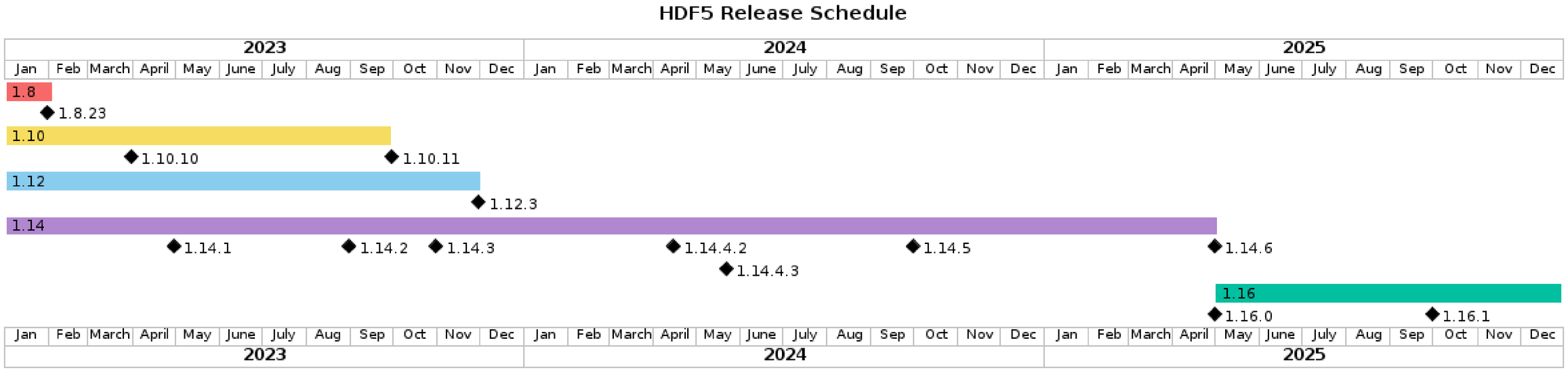
The HDF Group

https://github.com/HDFGroup/cve_hdf5



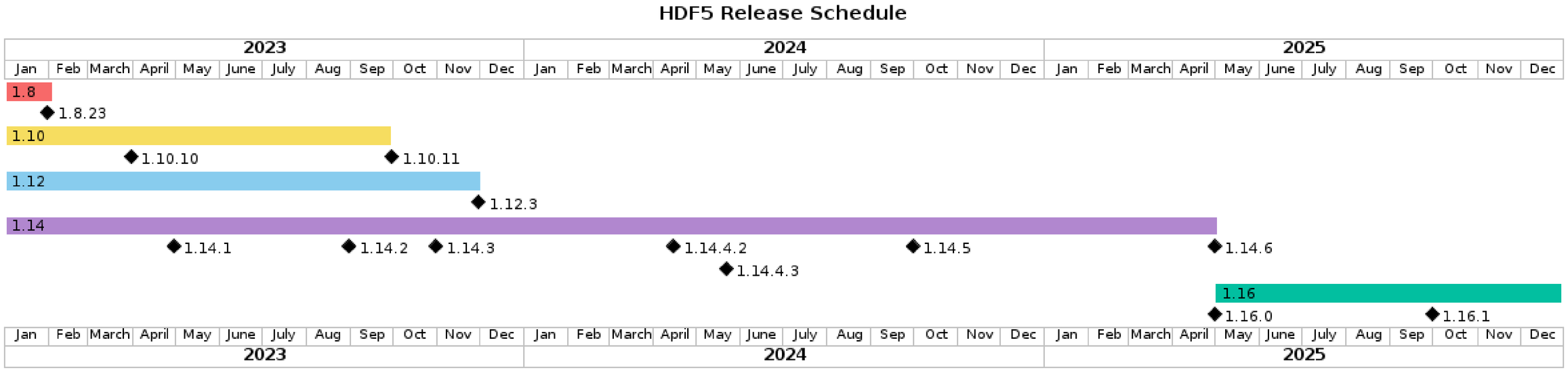


HDF5 Release Schedule



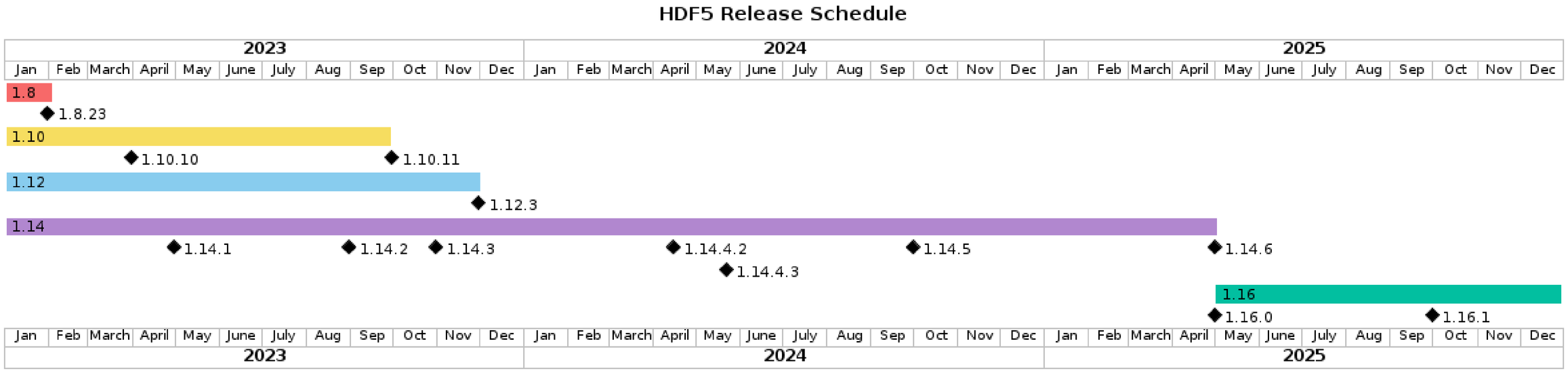
- Latest release: 1.14.4 (April 2024)
  - Vector I/O + subfiling VFD
  - Selection I/O + MPI-I/O VFD (on by default)
  - Many other minor tweaks and bugfixes
  - **NOTE: 1.14.4.3 is a patch release that fixes an ID leak when converting compound datatypes**

HDF5 Release Schedule



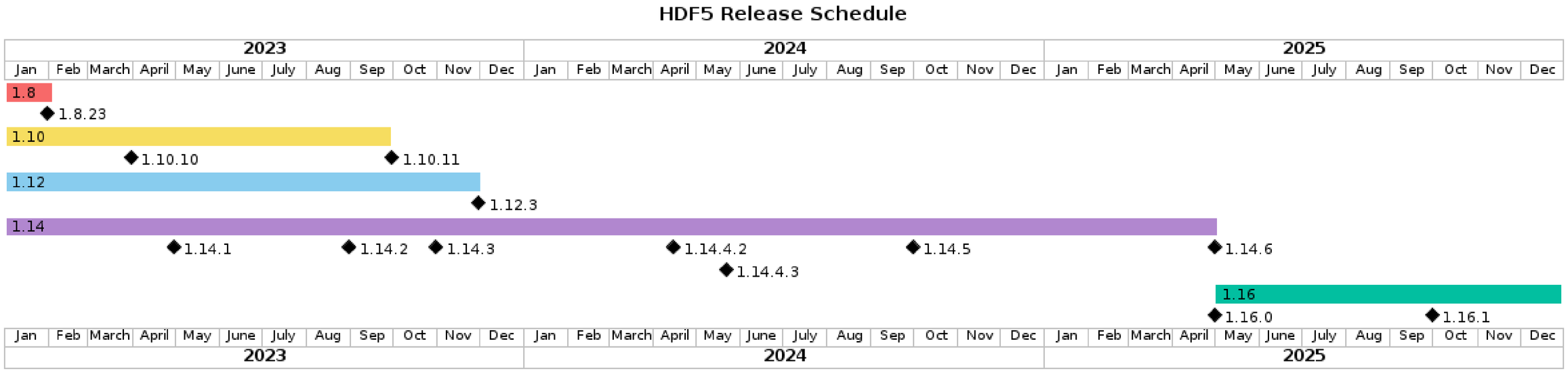
- n-1 release: 1.14.3 (November 2023)
  - Improved cross-compiling support (no more h5detect or h5make_libinfo)
  - Cloud-optimized HDF5 tweaks (16 MiB ros3 VFD cache)
  - Better support for Intel oneAPI, AMD aocc, Nvidia nvhpc
  - Async Fortran API calls
  - Parallel HDF5 tweaks and improvements

HDF5 Release Schedule



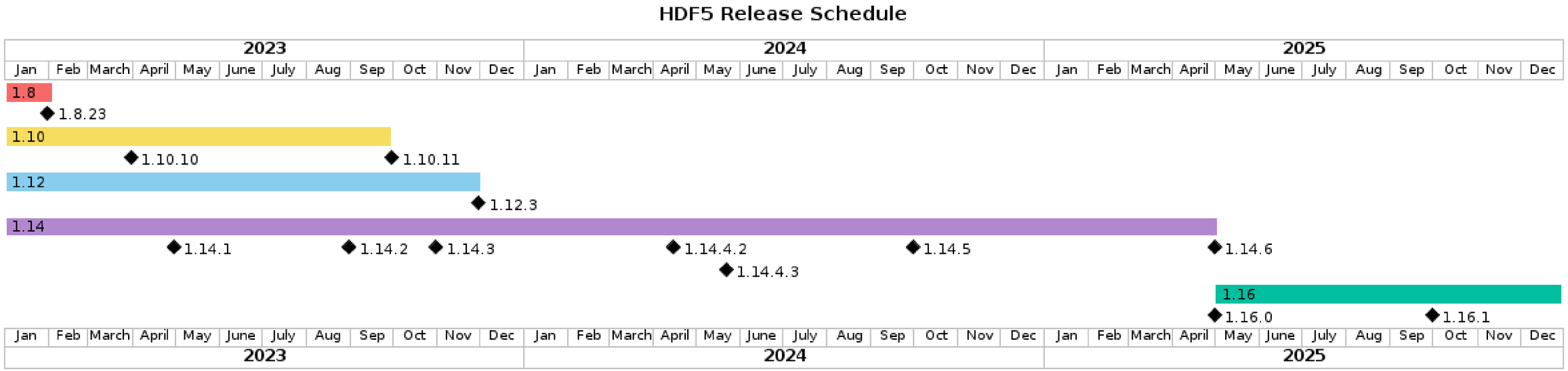
- 1.14.5 will release in late September
  - Possible last release of 1.14.x
  - oss-fuzz fixes
  - Cloud-optimized HDF5 tweaks (S3 object URL support, AWS_* env var support, tool page buffer options)
  - zlib-ng support (https://github.com/zlib-ng/zlib-ng)

HDF5 Release Schedule



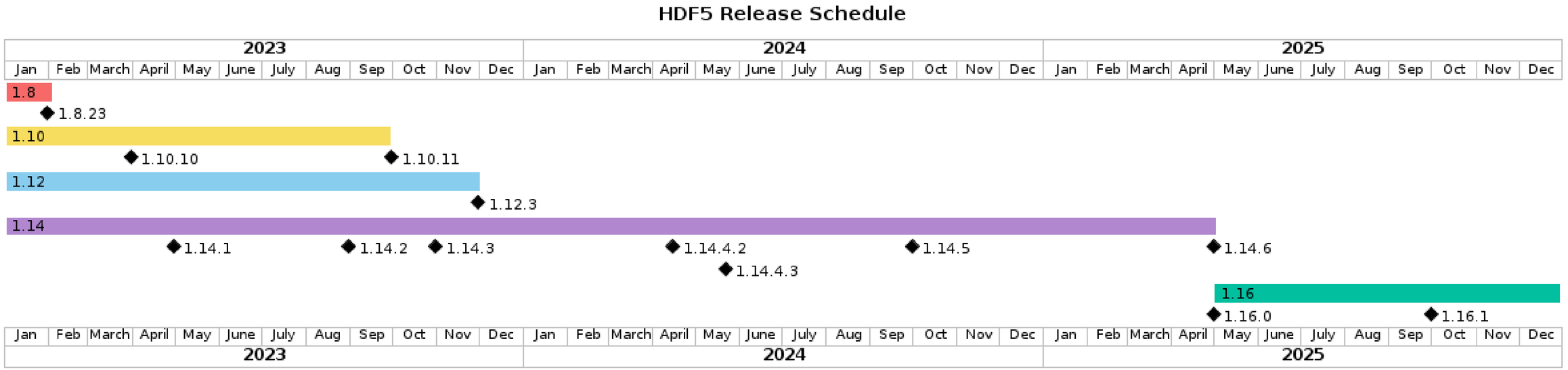
- 1.14.6
  - May not happen, depending on timing of 1.16.0 (see next slide)
  - If it happens, it'll probably just be a bugfix release

HDF5 Release Schedule



- HDF5 1.16.0
  - Complex number support
  - Changes to library defaults (cache sizes, etc.)
  - Other small API changes that require a major version number bump (off_t --> HDoff_t)
  - Intended to be an easy migration from 1.14.x
  - **Will no longer support 1.14 after 1.16.0 releases**

HDF5 Release Schedule



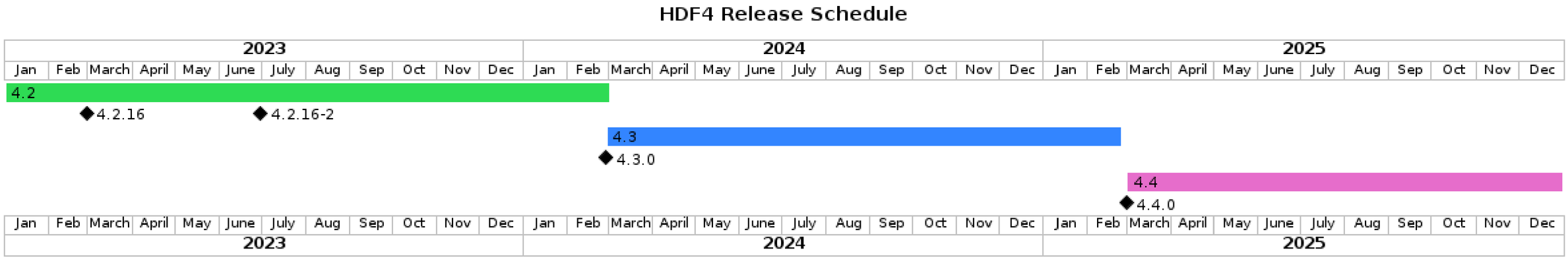
- When?
  - Official calendar says April
  - I’m hoping it’s earlier, either mid-November or sometime in January

HDF4

The HDF Group

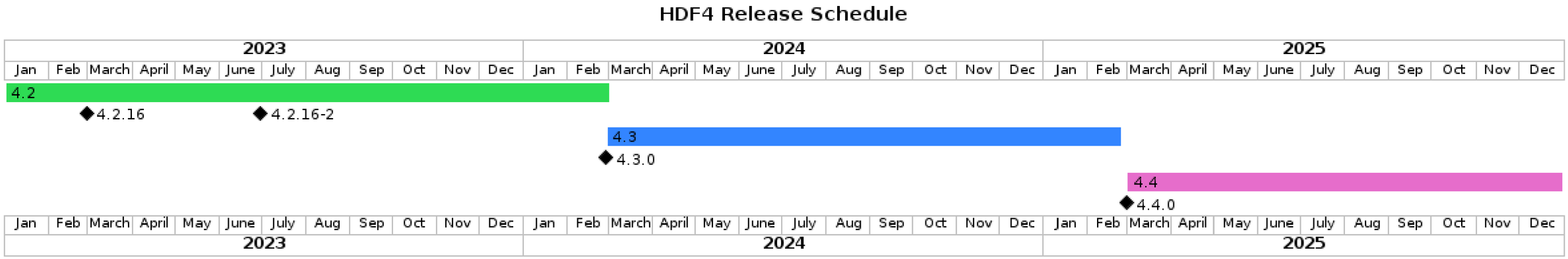
- Overall strategy:
  - Maintain the code and continue to release as long as we are able
  - Fix a few long-standing issues
  - Minimize long-term cost

HDF4 Release Schedule



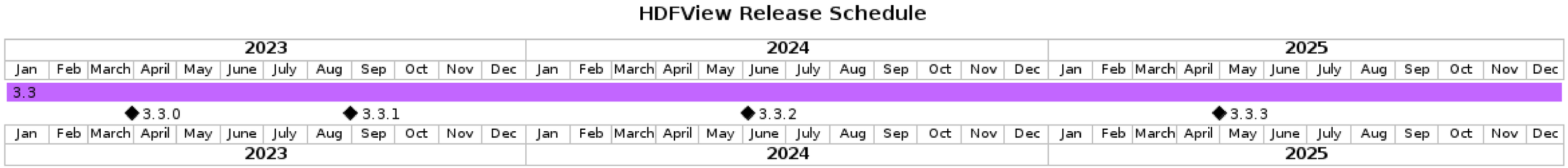
- Latest release: 4.3.0 (Feb 2024)
  - Defined public API (no longer release internal header files / API calls)
  - Removed XDR dependency
  - Bug fixes, etc.

HDF4 Release Schedule



- Next release: 4.4.0 (Feb 2025)
  - Fix library naming inconsistency between CMake and Autotools (libhdf vs ~~libdf~~)
  - Fortran will be in a separate library in the Autotools
  - Build system will look like HDF5, making it easier to maintain

HDFView Release Schedule



- Current version: 3.3.2 (May 2024)
  - HDF5 1.14.4 and HDF 4.3.0
- Next version: 3.3.3 (April 2025)
  - HDF5 1.16.0 and HDF 4.4.0
  - Hope to enable paging so large files can be loaded/navigated
- No longer maintain versions based on older versions of the HDF5 library

About release versions…

The HDF Group

- A product's release schedule is posted via README.md on GitHub
- We are trying to minimize the number of versions we support
  - 1 maintenance branch per product
- Releases are often driven by features and key projects, but we'll also try to settle in on a schedule, to make planning more predictable
  - HDF4 - 1x per year, in the spring
  - HDF5 - 2x per year, spring and fall
  - HDFView - 1x per year, after HDF4 releases
- We're working on automating as much of the process as possible

Community Engagement

HDF5 Working Group Meeting

The HDF Group

- Purpose is to discuss HDF5 library development
  - RFCs
  - Go over recent PRs / issues / discussions in GitHub
  - Other technical development discussion
  - Not for support (use the forum for that)
- Thursday at 10 am central time
- Open to anyone
  - https://us06web.zoom.us/j/89601195963
  - https://github.com/HDFGroup/hdf5/wiki/HDF5-Working-Group

HDF5 Plugin Working Group Meeting

The HDF Group

- Purpose is to discuss HDF5 plugin maintenance and deployment
  - https://github.com/HDFGroup/hdf5_plugins
- Every 2nd Thursday of the month at 10 am central time
  - Replaces the usual HDF5 Working Group meeting
  - August meeting will be the 3rd Thursday - August 15th
- Open to anyone
  - https://us06web.zoom.us/j/89601195963
  - https://github.com/HDFGroup/hdf5_plugins/wiki/HDF5-Plugin-Working-Group

HDF5 Enhancement Proposals (HEPs)

The HDF Group

- Borrowing Python's PEP infrastructure
- https://github.com/HDFGroup/heps (currently empty)
- HEP-1 will be like PEP-1 and describe the process
- Look for this in September

Technical Advisory Board

The HDF Group

I don’t have a lot of info here, but I do want to state that I’d like to restart this in the fall/winter.

Contact me if you are interested in being a part of this

Future Plans

Future Plans

The HDF Group

- Continued work on Cloud Optimized HDF5
- Machine Learning Optimized HDF5
- Documentation improvements
- Work towards HDF5 2.0.0

Cloud Optimized HDF5

The HDF Group

- Make the ros3 VFD a first-class citizen
- Performance tuning, possible selection I/O support
- Expand to other cloud object stores (multi-cloud: Azure, GCP)
- Continued tweaks and enhancements
- Add documentation to the user guide

Machine Learning Optimized HDF5

The HDF Group

- Similar what we've done with cloud optimized HDF5
- Analyze I/O patterns
- Documentation, demos, best practices

Documentation Improvements

The HDF Group

- User guide needs attention
- Cloud-optimized HDF5
- Migration guide for people who are updating versions
- Best practices guide(s)

HDF5 2.0.0

The HDF Group

Tentative features (no release date yet):

- Multi-threaded HDF5 (via Lifeboat, LLC and Nvidia)
- Sparse data storage and improved variable-length data storage (via Lifeboat, LLC)
- Crashproofing
- Encryption and digital signatures
- Full SWMR (VFD SWMR)

See https://github.com/HDFGroup/hdf5 for a full list

NOBUGS 2024 Meeting

The HDF Group

- September 24-26 in Grenoble, France
- Organized by the European Synchrotron Radiation Facility (ESRF) and the Institut Laue-Langevin (ILL)
- Can attend remotely (only €50!)
- Meeting website: https://indico.esrf.fr/event/114/
- HDFG is giving a talk on crashproofing (Tuesday, Sept 24 at 5pm)

Support Our Non-Profit Mission

The HDF Group

To ensure efficient and equitable access to science and engineering data across platforms and environments, now and forever.

THESE DON'T COST YOU A DIME	HELP US TO KEEP IT THAT WAY
Help Desk Support	Become a Code Owner
Sustaining Engineering	Consult with Us
HDF Clinic, Working Group	Purchase Custom Development
Webinars, User Events	Get HDF Software Priority Support
HDF User Forum	Donate or be a Sponsor
Community Outreach	Collaborate with Us on a Proposal
Assured Longevity of HDF Technologies	Become an HDF Advocate

Contact: info@hdfgroup.org

https://www.hdfgroup.org/donate

THANK YOU!

Questions & Comments?