

The State of HDF5 HUG23

August 16, 2023



Dana Robinson
Director of Software Engineering
The HDF Group

Contents



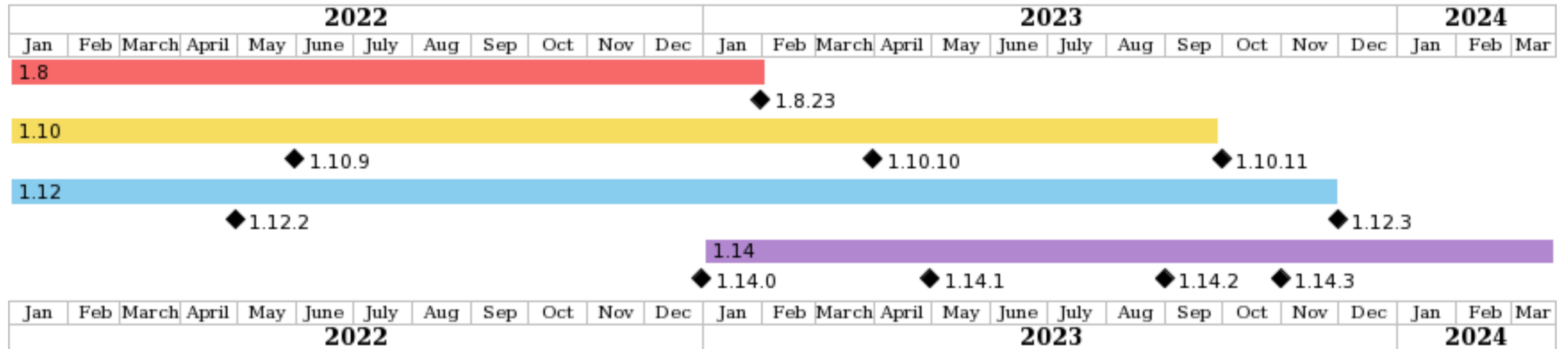
- HDF Product Status
- Community Engagement
- Future Plans

HDF Product Status

HDF5 Release Schedule



HDF5 Release Schedule

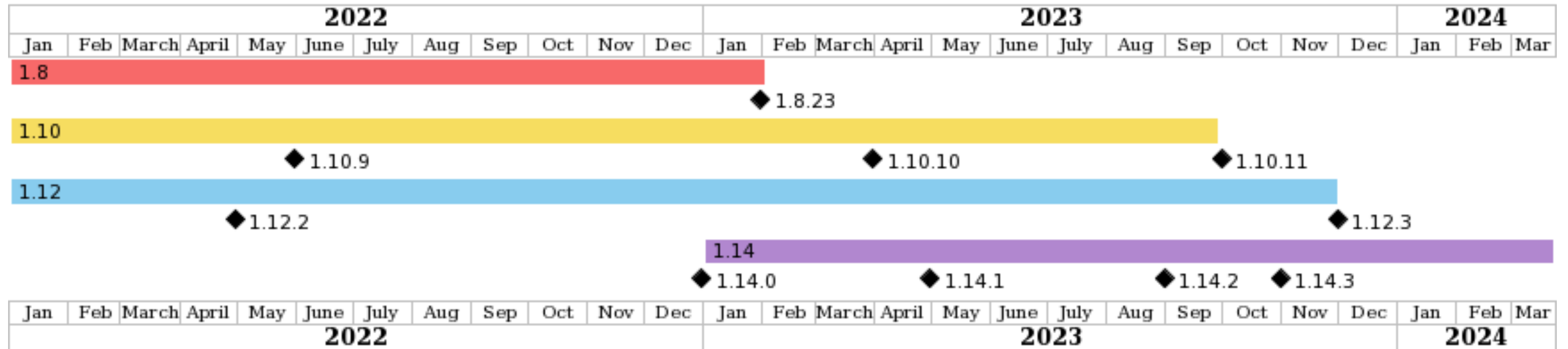


- Latest release: 1.14.2 (releasing this week)
 - In-place type conversion
 - Selection I/O on w/ MPI-I/O VFD
 - ros3 VFD supports temporary security tokens
 - Bug and CVE fixes (Only one known CVE issue remaining!)

HDF5 Release Schedule



HDF5 Release Schedule

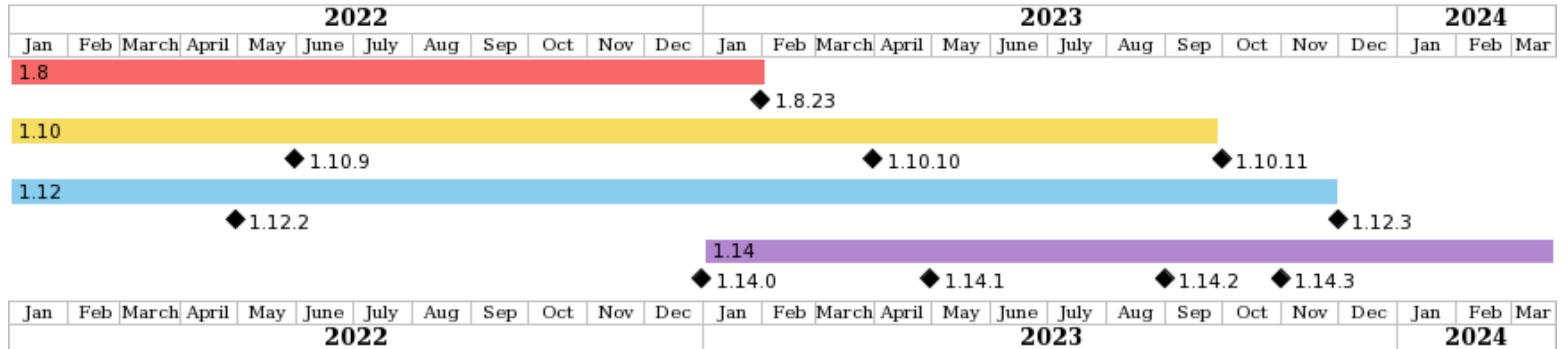


- 1.14.3 will release in October
 - Incorporated into the last E4S release for the Exascale Computing Project
 - Should be CVE-free
 - Focus will be on addressing any remaining ECP-related issues for E4S
- 1.14.4 release not planned yet (assume H1 2024)

HDF5 Release Schedule



HDF5 Release Schedule



- 1.8 line ended in January with 1.8.23
- 1.10.11 (September) - end of 1.10 line
- 1.12.3 (November) - end of 1.12 line

HDF5 Release Schedule



- Major 1.14.0 features (reminder)
 - Subfiling
 - Multi-dataset I/O
 - Selection I/O
 - Onion VFD
- VFD SWMR is still TBD
 - Looking for resources to finish this up

HDF4 Release Schedule



HDF4 Release Schedule

2023												2024											
Jan	Feb	March	April	May	June	July	Aug	Sep	Oct	Nov	Dec	Jan	Feb	March	April	May	June	July	Aug	Sep	Oct	Nov	Dec
4.2																							
◆ 4.2.16				◆ 4.2.16-2				◆ 4.2.17															
Jan	Feb	March	April	May	June	July	Aug	Sep	Oct	Nov	Dec	Jan	Feb	March	April	May	June	July	Aug	Sep	Oct	Nov	Dec
2023												2024											

- Latest release: 4.2.16-2 (June 2023)
 - Patch release (fixes a critical Java issue that affected HDFView)
 - A few other minor bug fixes
 - Build system improvements

HDF4 Release Schedule



HDF4 Release Schedule

2023												2024											
Jan	Feb	March	April	May	June	July	Aug	Sep	Oct	Nov	Dec	Jan	Feb	March	April	May	June	July	Aug	Sep	Oct	Nov	Dec
4.2																							
◆ 4.2.16		◆ 4.2.16-2				◆ 4.2.17																	
Jan	Feb	March	April	May	June	July	Aug	Sep	Oct	Nov	Dec	Jan	Feb	March	April	May	June	July	Aug	Sep	Oct	Nov	Dec
2023												2024											

- 4.2.16 (Feb 2023)
 - Significant code cleanup
 - Build system overhaul (no longer dependent on hard-coded machine values)
 - Many bug fixes

HDF4 Release Schedule



HDF4 Release Schedule



HDF4 Release Schedule

2023												2024											
Jan	Feb	March	April	May	June	July	Aug	Sep	Oct	Nov	Dec	Jan	Feb	March	April	May	June	July	Aug	Sep	Oct	Nov	Dec
4.2																							
◆ 4.2.16				◆ 4.2.16-2								◆ 4.2.17											
Jan	Feb	March	April	May	June	July	Aug	Sep	Oct	Nov	Dec	Jan	Feb	March	April	May	June	July	Aug	Sep	Oct	Nov	Dec
2023												2024											

- Next release: 4.2.17 / 4.3.0 (February 2024)
 - Defined public API (no longer release internal header files / API calls)
 - Remove XDR dependency
 - Bug fixes, etc.

HDFView Release Schedule

- Sorry, no release plot for HDFView (yet)
- Current version: 3.3.0 (April 2023)
 - HDF5 1.14.0
 - HDF4 4.2.16
- Next version: 3.3.1 (August/September 2024)
 - HDF5 1.14.2
 - HDF4 4.2.16.-2
- No longer maintain versions based on older versions of the library

About release versions...



- 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

Engineer of the Week



- We have a dedicated engineer whose job it is to monitor key products on GitHub (HDF5, HDF4, HDFView, HSDS)
- Rotates weekly through all HDFG engineers
- Issues and pull requests get tagged, assigned, discussed and the engineer helps push things forward
- Will add HDF Group forum monitoring going forward

HDF5 Working Group Meeting



- Purpose is to discuss HDF5 library development
 - RFCs
 - Go over recent PRs / issues / discussions in GitHub
 - Other technical development discussion
 - Not for support
- Every Thursday at 10 am central time
- Open to anyone. Send email to derobins@hdfgroup.org for an invite to the Teams meeting.

Everything is moving to GitHub



- Releases
- Bug reports --> GitHub issues
 - Will move existing Jira issues to GitHub
 - Exception: Some issues with private data will stay private
- New feature requests --> GitHub Discussions
- Support questions --> help@hdfgroup.org or HDF Group forum
- Development documentation, HDF5 WG info, etc. --> GitHub wiki

Some things are not great on GitHub (e.g. code review, roadmaps), so this will be a work in progress while we look for good solutions.

Code Owners



- We're looking for people outside of The HDF Group to take a more active role in HDF product development
- Still working on what this looks like, from an organizational perspective
 - I don't want to make too many demands on volunteers
- Have already added some external users to the CODEOWNERS files in HDF5 and HDF4
- Contact me if you are interested!

Change Management



- Aside from binary, file format, API, etc. compatibility, there are gray areas
 - Build system options
 - Tool options
 - Semantics
- We're working on a policy for how we will handle things like this

Change Management



Before making a change that faces users, we will...

- Create a discussion post on GitHub
- Create a sticky post in the HDF Forum
- Make an announcement in the newsletter
- Bring the issue up at an HDF5 WG meeting (HDF5)

- Allow at least 30 days for comments

Change Management

- In place in HDF4, will move to other products soon

Categories

- View all discussions
- Announcements
- Breaking Change
- General
- Ideas
- Polls
- Proposed Change**
- Q&A
- Show and tell

- [PROPOSAL] Remove unsafe Fortran API calls**
derobins started 3 weeks ago in Proposed Change
- [PROPOSAL] Fix HDF4 library naming / organizational scheme**
derobins started 3 weeks ago in Proposed Change
- [PROPOSAL] Define a public interface and stop deploying internal headers**
derobins started 3 weeks ago in Proposed Change
- [PROPOSAL] Require Fortran 95**
derobins started 3 weeks ago in Proposed Change
- [PROPOSAL] Drop the pre-C99 types (at least from internal code)**
derobins started 3 weeks ago in Proposed Change
- [PROPOSAL] Remove XDR dependency**
derobins started 3 weeks ago in Proposed Change

Future Plans

What are we going to work on in H2 2023?



- Take security seriously
- Provide the best possible HDF5 library for the final E4S release
- Make the ros3 VFD a first-class citizen
- Discuss the next major features that we'll work on in 2024 (HDF5 2.0?)

Take Security Seriously



- Fix remaining CVE issues
 - Final releases of 1.10, 1.12 should also be CVE-free
- More improvement of the CVE test repo
 - https://github.com/HDFGroup/cve_hdf5
 - Still missing proof-of-vulnerability files from Talos issues
 - Need to specifically address GIF tool issues
 - Should run the test script via a GitHub action to prevent regression
- Better integration with OSS-fuzz, OpenSSF Scorecard, etc. and fix identified issues
- Restart Coverity scans

CVEs



- CVE = Common Vulnerability and Exposure
 - https://en.wikipedia.org/wiki/Common_Vulnerabilities_and_Exposures
- Tracked by mitre.org
 - <https://www.cve.org/>
- There are 60+ filed against HDF5 since 2016 (some are duplicates)
 - <https://cve.mitre.org/cgi-bin/cvekey.cgi?keyword=HDF5>
- We track these and now publish a table with each HDF5 release
 - https://github.com/HDFGroup/hdf5/blob/hdf5_1_14_2/CVE_list_1_14.md

hdf5 / CVE_list_1_14.md

derobins Add CVE matrix (#3370) ✓

Preview Code Blame 75 lines (74 loc) · 9.01 KB

CVE issue number	1.14.0	1.14.1	1.14.2
CVE-2022-26061	UNTESTED	UNTESTED	UNTESTED
CVE-2022-25972	UNTESTED	UNTESTED	UNTESTED
CVE-2022-25942	UNTESTED	UNTESTED	UNTESTED
CVE-2021-46244			
CVE-2021-46243			
CVE-2021-46242			
CVE-2021-45833			
CVE-2021-45832	UNTESTED	UNTESTED	UNTESTED
CVE-2021-45830			
CVE-2021-45829			
CVE-2021-37501	FAILED		

Provide the Best HDF5 Library for E4S



- More ECP VOL connector integration testing (async, etc.) via GitHub actions
- Fix as many parallel/HPC-related Jira/GitHub issues as possible
 - Move everything in Jira that can't be fixed to GitHub
- HPC CDash results linked via GitHub README.md
- Finish VOL test suite work
 - Work with VOL connector authors to address failing tests
- Better documentation for recent ECP features
- Internal MPI-related code cleanup

The ros3 VFD Becomes a First-Class Citizen



- Expand to Google, Azure
- Investigate performance
- Look into replacing cURL with the S3 C library
- Think about making cloud-optimized values the library defaults
- Update the HDF5 User's Guide
- Better testing (expanded test suite, GitHub actions)
- Update to use selection I/O
- Better tools support (especially HDFView)
- Look into improving h5repack performance

HDF5 2.0 / Future Major Work



- TBD!
- These 3:30 - 5:00 pm meetings are for us to get feedback from the community about what will go into HDF5 in the future
- I have other slides full of ideas and conversation starters...








European HUG



- Hamburg, Germany at DESY
- September 19-21, 2023
- <https://indico.desy.de/event/39343/>

Support Our Non-Profit Mission

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

THESE DON'T COST <u>YOU</u> A DIME
 Help Desk Support
 Sustaining Engineering
 HDF Clinic, Working Group
 Webinars, User Events
 HDF User Forum
 Community Outreach
 Assured Longevity of HDF Technologies

HELP US TO KEEP IT THAT WAY
Become a Code Owner
Consult with Us
Purchase Custom Development
Get HDF Software Priority Support
Donate or be a Sponsor
Collaborate with Us on a Proposal
Become an HDF Advocate

Contact: info@hdfgroup.org
<https://www.hdfgroup.org/donate>

THANK YOU!

Questions & Comments?