The State of HDF5 – 2022-3

May 31, 2022

The HDF Group

Dana Robinson
The HDF Group

Overview

- Upcoming Features
- Release schedule for 2022
- Whither HDF5 ?

Upcoming Features

Upcoming Features

The HDF Group

- Onion VFD
- VFD SWMR
- Selection I/O
- Subfiling
- Multi-Dataset I/O
- VOL API compatibility flags

- Committed to getting these out in HDF5 1.14.0

Onion VFD

The HDF Group

- Versioning for HDF5 files

- Requires an external "onion" file for versions > 0

  - Version info could be stored in the original file
    - But would violate the "don't touch the original" principle
    - Plans for this, but not implemented

- Integrated with the command-line tools

- To be released in **HDF5 1.13.2**

VFD SWMR

The HDF Group

- "SWMR 2.0"
- Different scheme than "legacy SMWR" released in HDF5 1.10.0
  - Uses an external metadata snapshot file instead of flush dependencies
  - Deals with file metadata, not raw dataset data
  - Unlike "legacy SWMR" allows most operations (including object creation, etc.)
  - Legacy SWMR stays in place (for now)
  - Will eventually work with network file systems like NFS/SMB
- To be released in **HDF5 1.13.2**

Selection I/O

- Extends the VFD layer so I/O operations can be passed vectors of reads and writes
- Allows the creation of more complicated derived MPI types instead of breaking complex I/O into multiple read/write calls
- Mainly for parallel HDF5, but has applications in serial HDF5 (e.g., S3 VFD access)
- Already in develop branch
- To be released in **HDF5 1.13.2**

Subfiling

- I/O concentrators in parallel HDF5
- Middle ground between file-per-process and single-shared-file
- Lets the user make the tradeoff between number of files and shared file lock contention
- Implemented at the VFD layer
- To be released in **HDF5 1.13.3**

Multi-Dataset I/O

- Spreads I/O among multiple datasets in parallel HDF5
- Minimizes the number of I/O calls when writing to multiple open datasets
- Limited to datasets stored in a single file
- To be released in **HDF5 1.13.3**

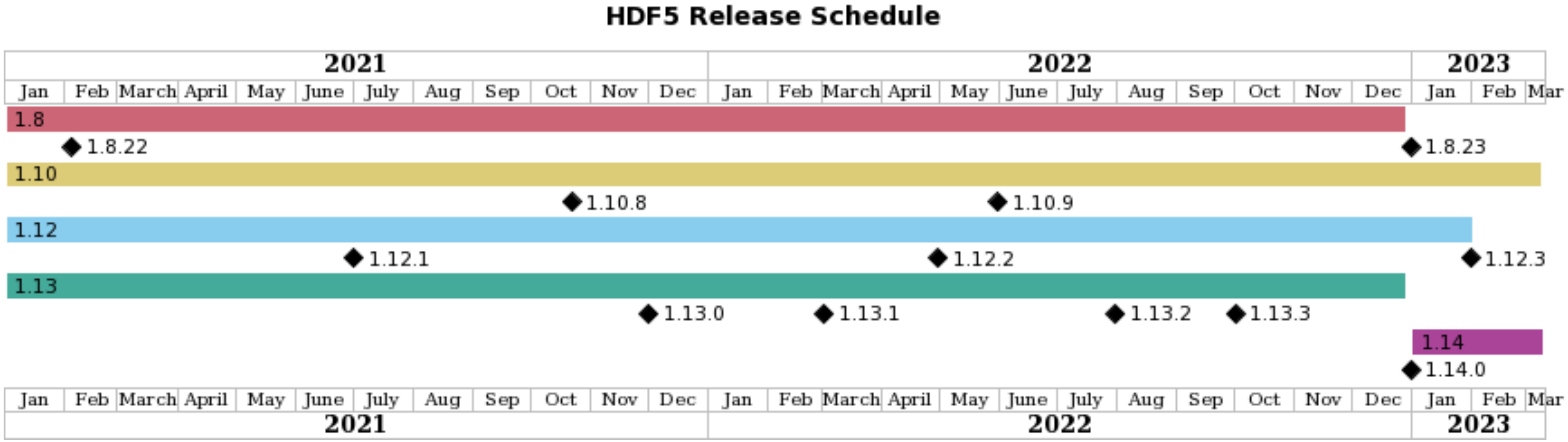
VOL API Compatibility Flags

- Will allow a virtual object layer (VOL) connector to specify which aspects of the HDF5 API it supports
- Will allow applications to determine if a VOL connector is suitable for its HDF5 utilization
- Will work with the vol-test repository to provide better testing
- No set release date right now
- Email me if you want to be a part of this (derobins@hdfgroup.org)

Releases

Release Schedule

- On GitHub (https://github.com/HDFGroup/hdf5), in the main README.md file
- Will be kept current



HDF5 Release Schedule

Experimental vs Maintenance branches

- Even number minor releases are maintenance releases (e.g., 1.12.x)
    - Usual HDF5 maintenance branches you know and love
    - Binary compatibility
    - Stable file format

- Odd number minor releases are experimental releases (e.g., 1.13.x)
    - No binary compatibility guarantees
    - API calls can change
    - File format can change
    - Unready features may be dropped
    - Used to try out new features as we prepare for the next maintenance release

- See the blog posts on this for more clarity
    - https://www.hdfgroup.org/2021/12/hdf5-1-13-0-introducing-experimental-releases

HDF5 1.8

- HDF5 1.8.23 will be released at the end of the year
- Last version of HDF5 1.8
- Main thing keeping 1.8 alive is performance issues, which should be addressed this summer

HDF5 1.10

- HDF5 1.10.9 just released
- Depending on the summer's performance gains, we may have a fall release of HDF5 1.10
- Will almost certainly live on into 2023
- Plan to retire at the end of 2023 or early 2024

HDF5 1.12

- HDF5 1.12.2 released in April
- HDF5 1.12.3 will be the last release of the HDF5 1.12 maintenance branch
- Incompatible VOL layer requires retiring this branch

HDF5 1.13 & 1.14

- HDF5 1.13.2 to be released in July
  - Selection I/O
  - Onion VFD
  - VFD SWMR

- HDF5 1.13.3 to be released in September
  - Multi-Dataset I/O
  - Subfiling

- HDF5 1.14.0 to be released in Nov/Dec

HDF5 1.16 & HDF5 2.0

- After 1.14.0 releases, develop will switch to 1.15, though it's not clear if the next major release will be 1.16.0 or 2.0.0

- Ideally HDF5 2.0 and allow larger API changes
  - Finally support semantic versioning properly
  - Drop deprecated API calls
  - Condense metadata categories & multi VFD → split VFD
  - Overhaul baroque and inconsistent error reporting scheme

- Address long-standing technical debt
  - Windows Unicode
  - Thread-safety
  - Drop Autotools support
  - Internal refactoring

Whither HDF5 ?

The next two years

- 2022

  - Mainly going to be about getting under-development features out the door
  - Will push to make performance improvements over the summer
  - Drop several maintenance branches
  - Community engagement and participation

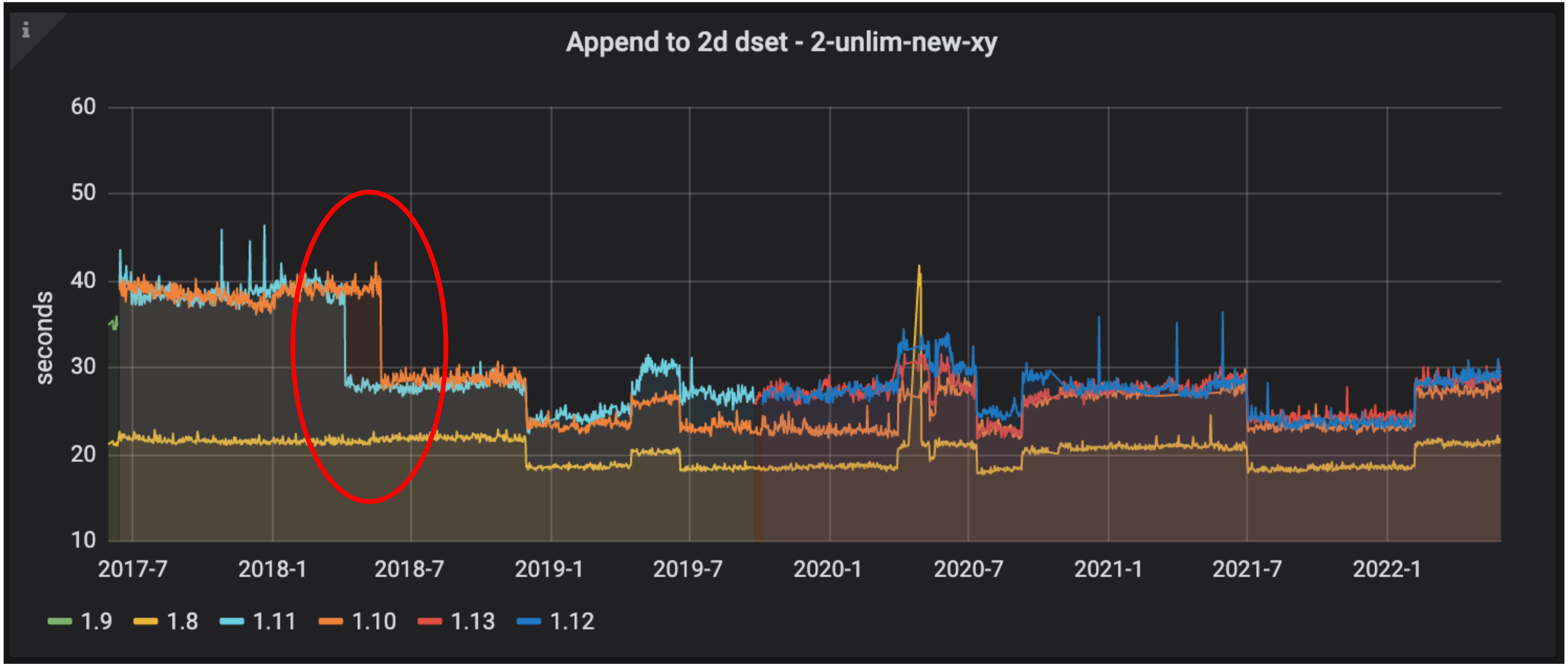
- 2023

  - Making HDF5 a maintainable piece of software
  - "HDF5 2.0"

Performance

- Significant performance regression between 1.8 and 1.10 release branches
- Several fixes have brought 1.10 closer to 1.8 levels of performance
- Remainder be addressed over the summer, as best we can
- Critical work so we can retire 1.8

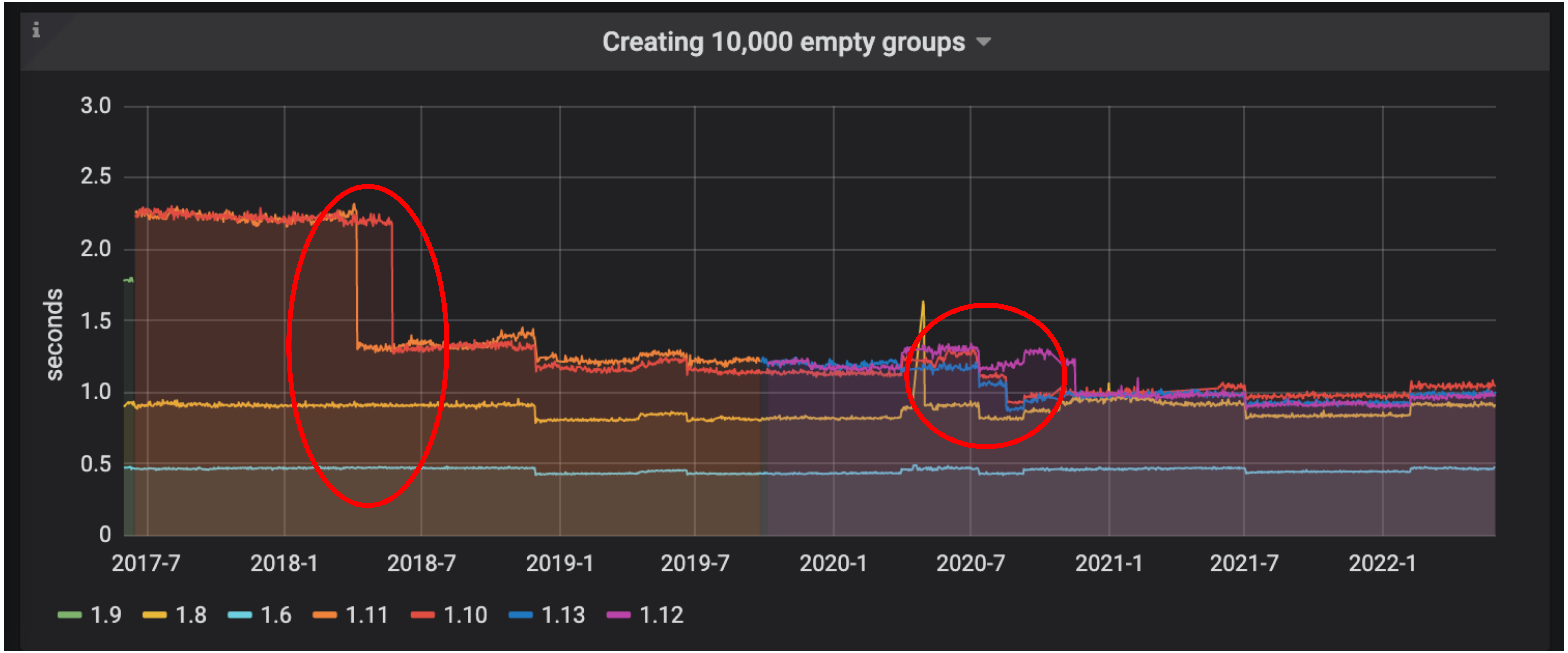
Performance (Bad)



https://grafana.hdfgroup.org/login (demo:demo)

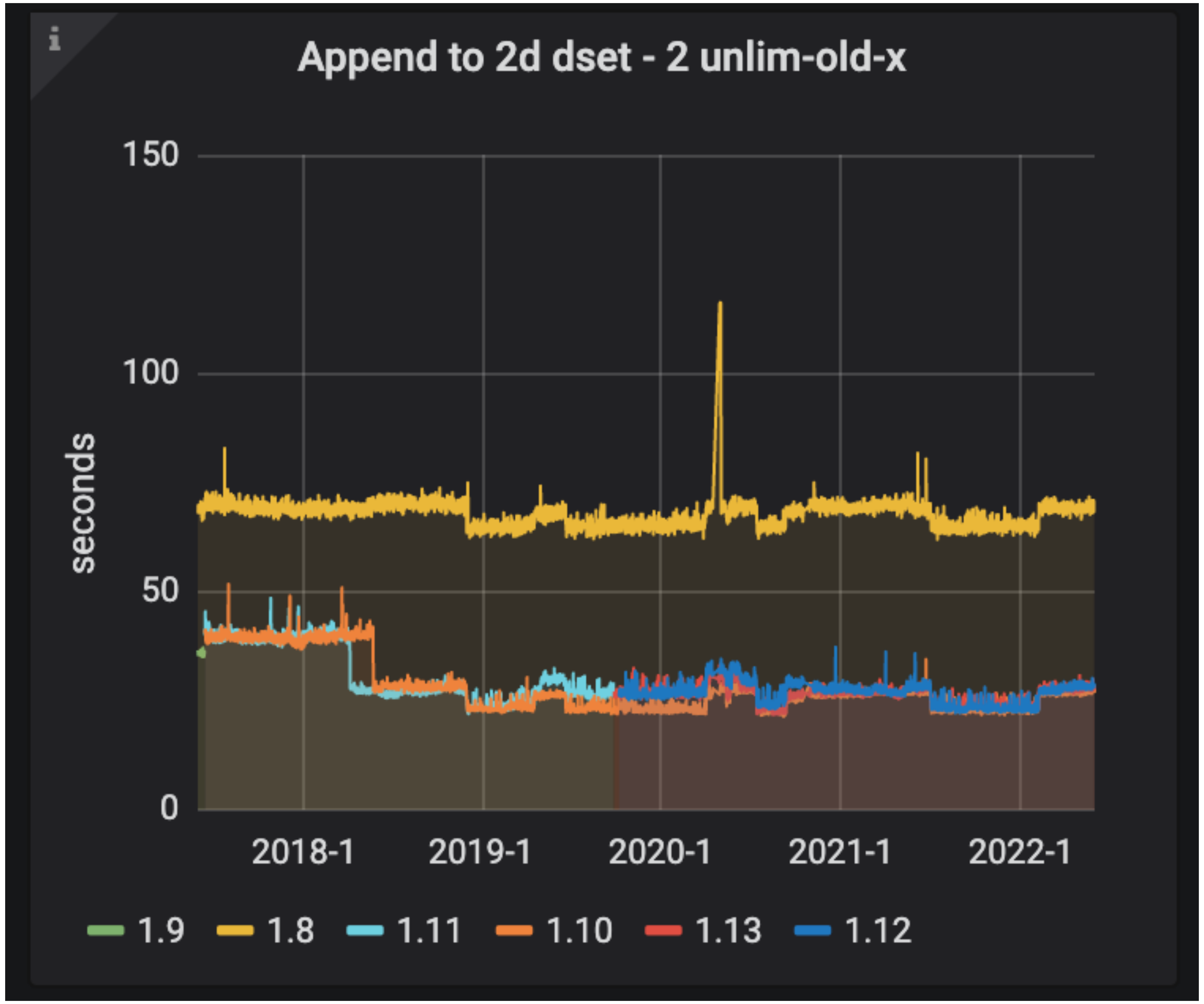
Performance (Good?)

The HDF Group



https://grafana.hdfgroup.org/login (demo:demo)

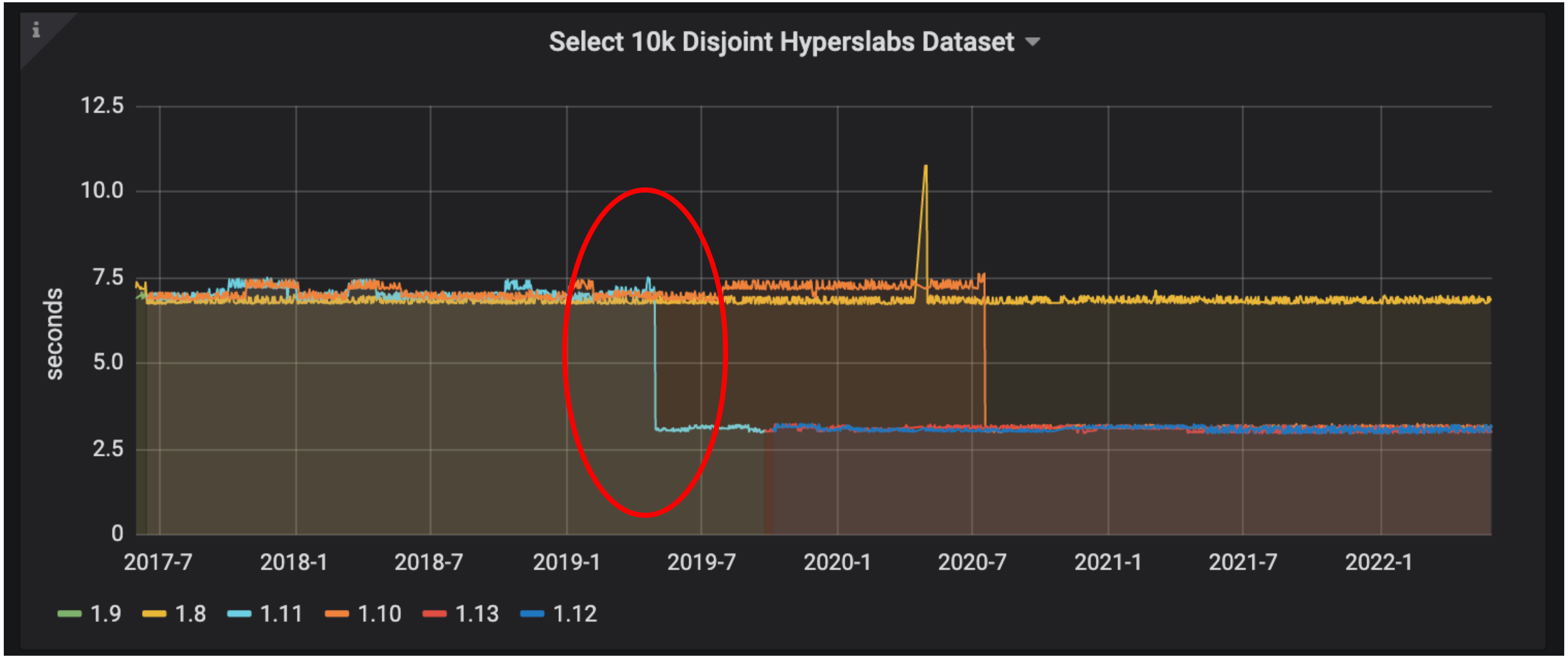
Performance (Good)



https://grafana.hdfgroup.org/login (demo:demo)

Performance (Good)

The HDF Group



https://grafana.hdfgroup.org/login (demo:demo)

Funding "Sustainable Engineering"

The HDF Group

- No pile of money for this
- Most funding pays for features, not maintenance
- We basically get enough money to do releases, testing, and staff the help desk
- Addressing maintainability issues has largely been a labor of love by a small number of developers
- This has been a challenge for The HDF Group for many years

Whither HDF5 ?

How do we make HDF5 a sustainable piece of software?

1. More maintainable code
   - Reduce
   - Refactor
   - Simplify
   - Document
   - Automate

2. More community participation
   - Tune in tomorrow!

THANK YOU!

Questions & Comments?