



# CernVM Program of Work 2023

---

Jakob Blomer, Laura Promberger, Valentin Völkl



SFT Group Meeting

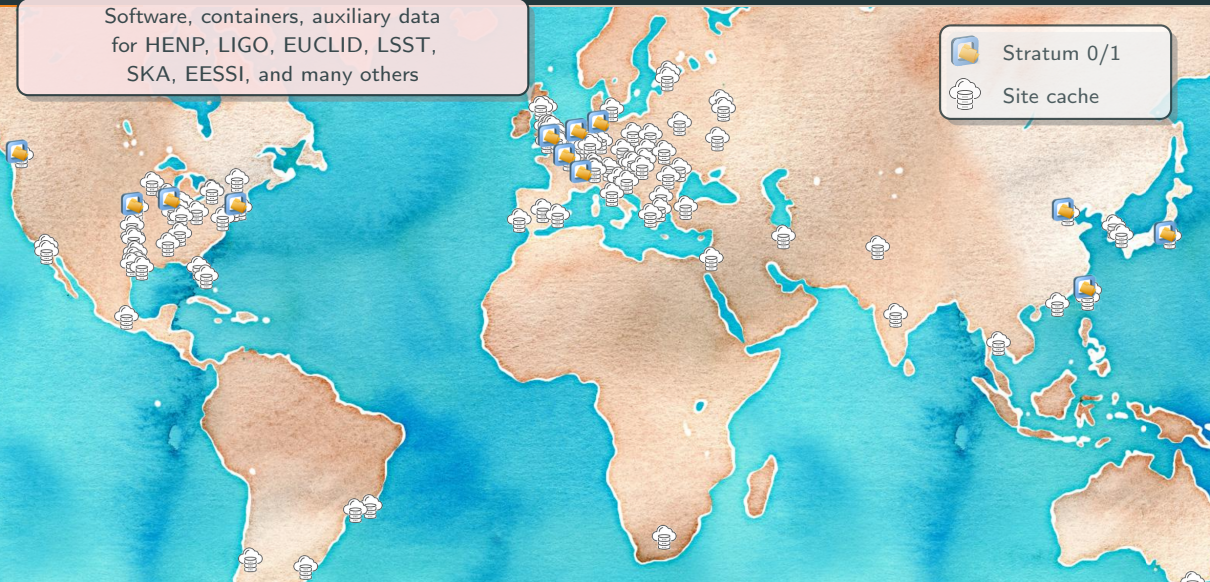
13 February 2023

# At a Glance: CernVM-FS Deployment (Grid)



Software, containers, auxiliary data  
for HENP, LIGO, EUCLID, LSST,  
SKA, EESSI, and many others

 Stratum 0/1  
 Site cache



# At a Glance: CernVM-FS Deployment (Grid)



Software, containers, auxiliary data  
for HENP, LIGO, EUCLID, LSST,  
SKA, EESSI, and many others



Stratum 0/1



Site cache

- ~ 15 Stratum 1s (Europe, North America, Asia)
- ~ 5.4 B files in the /cvmfs tree (+210 % in last year)
- ~ 2 PB of data accessible through /cvmfs  
out of which ~1 PB in *external* files  
proven to scale up to 100 PB
- ~ 3550 container images (+60 % in last year)
- ~ 260 repositories (+15 % in last year)



- Steady 50–100 commits per month
- 2022: ~13 000 LOC changed (-15 % wrt. 2021) by >10 contributors



cvmfs

⚙ Settings | 📄 Report Duplicate



High Activity

## Commits per Month

Zoom

1yr

3yr

5yr

10yr

All



## Review of 2022

---



## Highlights

- Substantial performance engineering on the client (hot cache)  
e.g. 30 % faster Athena builds on many-core machines  
full assessment in CHEP'23 contribution
- First version of proxy sharding (to be revised)
- Addressed long-standing issues around stale kernel caches
- Container tools integration with registry.cern.ch as an image proxy
- Moved from JIRA to GitHub issues
- Platform coverage: EL9 and OpenSSL 3, Ubuntu on ARM, SLES 15
- Construction of container-first CernVM 5  
resulted in successful bachelor thesis
- CernVM Workshop @ NIKHEF with >50 participants  
with speakers from Microsoft, Jump Trading, NTT
- Dissemination: [▶ DPHEP Report](#) (foreseen for EPJ-C) [▶ ACAT'22](#)



## Highlights

- Substantial performance engineering on the client (hot cache)  
e.g. 30 % faster Athena builds on many-core machines
- First version of proxy sharding (to be revised)
- ...

## Unfinished Tasks

- Refactoring of container conversion tools
- Release of containerd snapshotter
- macOS binary signatures
- In progress:
  - Client-side prefetching
  - Feature parity of local publisher and gateway publisher

CernVM Team

---



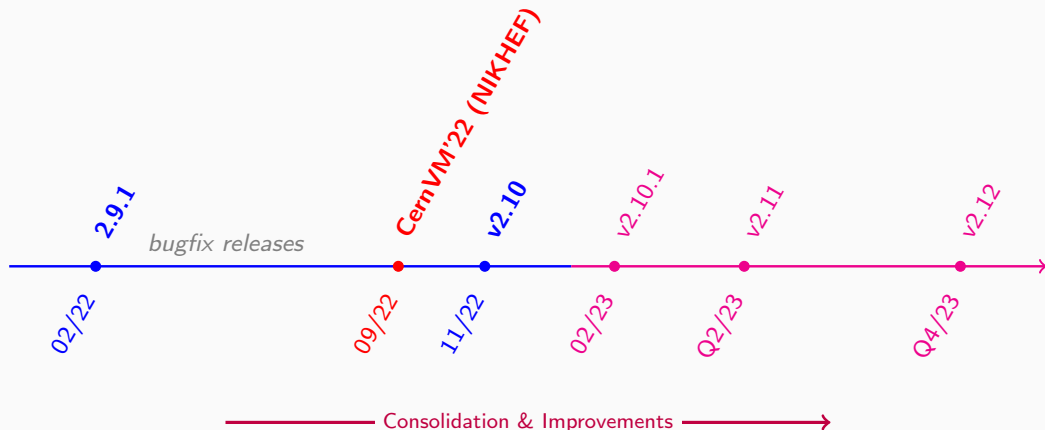


		2022	2023
Jakob Blomer	Staff	25 %	25 %
Laura Promberger	Fellow	40 %	100 %
Valentin Völkl	Staff	—	100 %
Radu Popescu	Staff	65 %	—
Jakob Eberhardt	Tech	75 %	—
<i>TBS</i>	Tech	—	25 %
FTE		~2.05	~2.5

Note: 1 fellow and 1 technical student externally funded by Jump Trading

Significant code contributors:

Matt Harvey (Jump Trading), Dave Dykstra (FNAL), Razvan Virtan (summer student)



Most of the client hot cache improvements will be part of the 2.11 release; some require recent kernel.

# CernVM Appliance 2023

---



1. Ready to use platform for HEP application stacks
2. Reference platform for **long-term data preservation**

As discussed at the CernVM workshop, CernVM Online has been decommissioned  
(offline by end of February 2023)

## 2023 Plan of Work

- Release EL 9 based CernVM 5 container **est 3 FTW [VV]**

First users

- LHCb apptainer base container
- Key4HEP / FCC software tutorials
- Support and advice for software preservation efforts

# CernVM File System 2023

---



## Extras:

- cvmfsexec
- cvmfs-servermon
- github-action-cvmfs
- cvmfs-x509-helper
- repository monitor
- ...

## Stand-alone utilities

Preloader

Shrinkwrap

## Services (Go)

containerd snapshotter  
(preproduction)

Container Publishing Tools

Gateway Services

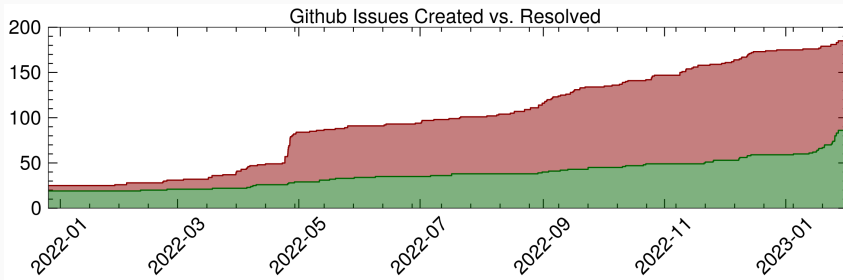
## Core Software

Client

Fuse module, libcvmfs,  
cache plugins

Server

publisher tools, libcvmfs\_server,  
Geo-API



## Support load in 2022:

- ~ 350 posts on mailing lists and forum
- Occasional tickets in SNOW, GGUS

## Tickets and PRs in 2022:

- 88 open issues, 66 closed
- 33 open pull requests
- 25 open bugs / 22 closed

## Growing backlog!

Bugfix sprint / hackathon planned for mid 2023



- Review of the technical documentation, removal of obsolete information **est 1 FTW**
- Integration tests ergonomics **est 3 FTW**
  - Regular stress tests [\[all\]](#)
  - Review of duplicated tests and flaky tests [\[all\]](#)
  - Stretch goal: continuous performance monitoring [\[summer student proposal\]](#)
  - Stretch goal: lightweight VM for each test [\[GSoC proposal\]](#)
- Refactored client-side file catalog updates **est 1 FTM** [\[LP\]](#)
- Continued Bash to C++ conversion (repository creation, destruction) **est 3 FTW**
- Investigate language upgrade to C++11 **est 3 FTW**
  - Main blocker: live upgrade of C++03 data structures during fuse module reload
- Feature life cycle management: deprecation procedure for rare / obsolete features (e.g., NFS HA mode, repository monitor and JavaScript client)





- De-duplication of open file descriptors in support of multiprocess frameworks **est 2 FTM [JB]**  
(Key User: ALICE)
- Cold cache performance engineering
  - Prototype of client-side object prefetching **est 3 FTM [VV]** (Key User: LHCb)
  - Refactoring of the HTTP client code **est 2 FTM [LP]**
    - Addresses many-core scalability limit identified in last year's summer student project
    - Facilitates site cache scalability, e.g. pluggable proxy health checks, I/O error tracing
- Full assessment of hot cache performance improvements (CHEP'23) **est 1 FTM [LP]**
- Investigation of ZSTD compression **est 2 FTM [LP]**



- Client resilience improvements for extreme conditions **est 4 FTW [LP, JB]**  
(several smaller improvements)
  - Out of memory recovery
  - Recovery of full disk conditions
  - Fix of rare races when root file catalogs are reloaded
- Optimized disk cache management for conditions data & grafted namespace **est 2 FTW [JB]**
- Improved macOS support **est 3 FTW [VV]**
  - Package notarization (no more special operations needed on install)
  - Native M1 builds
- Release of containerd snapshotter **est 2 FTW [VV]**
- Client-side kubernetes integration **est 3 FTM [TECH] (Key User: ATLAS)**
  - Assessment of use cases and available community approaches (daemon set, CSI driver)
  - Development of Helm charts, including web proxy deployment



- Feature-parity between gateway and single-publisher modes [VV]
  - Trigger garbage collection from remote publishers [est 3 FTW]
  - Use template transaction from remote publishers [est 1 FTW]
  - Improve fairness when multiple publishers are provisioned [est 2 FTW]
  - Full repository tagging support [est 3 FTW]
  - Stretch goal: rebase gateway receiver on libcvdfs\_server [est 3 FTW]
- Reduction of transaction overhead from <5 seconds to <1 second [est 2 FTW] [VV]
- New Bulk API for file grafting in libcvdfs\_server [est 1 FTM] [VV]
- REST API for container conversion service [est 1.5 FTM] [summer student proposal]
- Stretch goal: exploitation of modern overlayfs features [est 1 FTM] [LP]

# Community Interaction

---



- Developers and operators meet in a monthly coordination call (no changes for 2023)
- Monthly alignment with IT-ST (changed from weekly in 2022)
- NEW: establish contacts with experiment representatives in charge of operations
- Interaction with external users and industry (Jump Trading, Microsoft, EUCLID, LIGO, etc.)
- Conferences on the radar: CHEP, GDB, XRootD workshop, HEPiX
- Preparation of the CernVM Workshop 2024 (tba)

## Summary

---



## Main Priorities for 2023

1. Improved client hot cache and cold cache performance
2. Better support for kubernetes clusters (primarily client)
3. Continuous investment in robustness and scalability of the publishing
4. Fraction of effort continuously invested in code renovation



# CernVM Program of Work 2023

---

Jakob Blomer, Laura Promberger, Valentin Völkl

SFT Group Meeting

13 February 2023



## Backup Slides

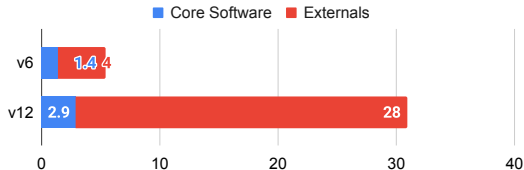
---

# On the Horizon: Software Management for HL-LHC

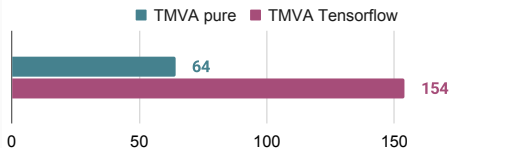
## Compared to run 1-2, we now find

- Multiple target architectures: x86\_64 micro-architectures (e. g. AVX512), AArch64, Power, GPUs
- A growing Python software ecosystem, in particular for machine learning tasks
- More agile software development: automated integration builds, nightly builds
- Generally we tend to add code and externals more often than removing components

CMSSW Single Version and Platform (Gigabytes)



Classification Tutorial: Number of File Lookups (in thousands)



My estimate: the software distribution problem for HL-LHC grows by a factor of 3-5 for most key metrics.