Turning CephFS into a collaborative space with CERNBox

Giuseppe Lo Presti[1,*], *Sebastian* Bukowiec[1], *Diogo* Castro[1], and *Hugo* G. Labrador[1]

[1]CERN - European Organization for Nuclear Research, 1211 Geneva, CH

Abstract. CERNBox is an innovative scientific collaboration platform, built using solely open-source components to meet the unique requirements of scientific workflows. Used at CERN for the last decade, the service satisfies the 35K users at CERN and seamlessly integrates with batch farms and Jupyter-based services. Powered by `Reva`, an open-source HTTP and gRPC server written in Go, CERNBox has demonstrated the provision of sync&share capabilities on top of multiple storage systems such as EOS and CephFS, as well as enabling federated sharing with other institutions.

In this contribution, we present the evolution of CERNBox towards supporting the low-latency Windows applications use-cases at CERN. As we are migrating out of DFS, the legacy Windows storage provided by Microsoft, and commissioning Windows Workspaces powered by CephFS, we show how CERNBox provides a flexible software stack to seamlessly integrate the Windows-based community, which includes the Engineering sector of the Organization.

We conclude by emphasizing the multiple synergies enabled by this approach. On one hand, Windows-based data-centric workflows can leverage the multi-protocol accesses (sync, web, SMB) provided by CERNBox. On the other hand, the widespread adoption of CephFS within the scientific community positions CERNBox as an out-of-the-box solution for implementing a scalable collaborative cloud storage service.

1 Introduction and Motivation

At CERN, the offer of storage for the Windows[8]-based user community has historically relied upon the Microsoft Distributed File System (DFS) [9], which has been in production since the late 1990s to support several use cases and workflows. However, DFS does not meet the requirements of a modern storage system with multiple access patterns and sync&share capabilities. Thus, CERN IT has launched a project to consolidate DFS into CERNBox, which offers cloud storage to the whole organization.

The personal storage areas hosted on DFS have been migrated in the past, and we reported about that migration in a previous work [2]. In this contribution, we report on the experience migrating all other workspaces and use-cases, which span from general teamwork areas to specific Windows-based workflows. Figure 1 shows the distribution of files hosted in DFS for the most common file types, with a grand total of about half a billion files.

By analyzing the usage pattern of these applications, we defined a strategy to determine the optimal storage solution for each use case. While most workloads could be accommodated within the existing CERNBox EOS-based offer [2], some use cases had more stringent

*e-mail: giuseppe.lopresti@cern.ch

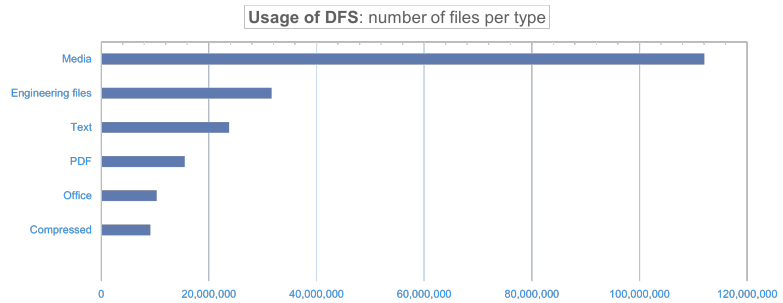

Figure 1. DFS Usage at CERN. While the volume is not large by today's standards, the large variety of file types and use cases supported by the system for multiple decades is significant.

requirements in terms of latency. Therefore, we designed and prototyped a complementary solution, and invited early adopters and critical stakeholders to identify potential issues in the early stage of deployment.

To implement the new solution, multiple alternatives were analyzed, including NetApp, CephFS with HDDs and SSDs, and we compared them to the existing EOS-based CERNBox storage system. We developed dedicated tests to compare DFS with these solutions, in particular when it comes to metadata I/O throughput and latency, as these turned out to be the most critical performance indicators for several Windows applications used across the Laboratory, such as the ones used by the engineering sector.

The tests concluded that EOS had significantly higher latency than DFS, whereas CephFS performed comparably. NetApp performed slightly better than all other systems (Figure 2). Additionally, we note that the performance differences between Ceph with HDDs, Ceph with SSDs, and NetApp were not significant, and all of them met the low latency required to support Windows applications. As CERN's IT Storage Group already operates a multi-cluster Ceph service that could be leveraged, and since Ceph offers an open source API to facilitate its integration, we chose CephFS as the back end to develop the new solution.

The new space offer has been nicknamed ***WinSpaces***, as the most prominent feature is the support of always-online, low-latency Windows applications (see Figure 3). we will describe the architecture of the system and how we can cover the entire spectrum of use cases.

However, CephFS is not limited to serving Windows applications alone. Over the past few years, Ceph has emerged as a de-facto standard for storage provisioning, widely adopted across multiple sites and data centres, due to its scalability, flexibility, and robust performance. This widespread adoption has made Ceph a cornerstone of modern storage infrastructure. By integrating CephFS into CERNBox, new opportunities arise to leverage CERNBox as a collaborative sync&share platform seamlessly built on top of existing, on-premise Ceph-based storage systems. This integration not only enhances the versatility of CERNBox but also allows organizations to maximize the value of their current Ceph deployments, enabling efficient data sharing and collaboration while maintaining control over their storage infrastructure.

2 The CERNBox Ecosystem

Based on EOS [1] storage, CERNBox [5] is the CERN cloud storage hub that has become the reference platform for collaboration across the Organization.

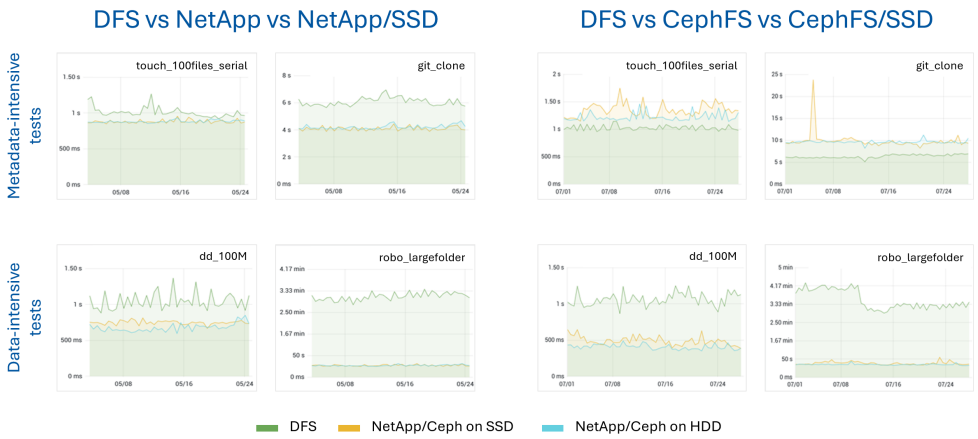

Figure 2. Subset of test results comparing DFS with Ceph and NetApp on different backends. Metadata-intensive tests show that CephFS is on par with DFS whereas NetApp performs better. Data-intensive tests show that Both CephFS and NetApp perform significantly better than DFS.

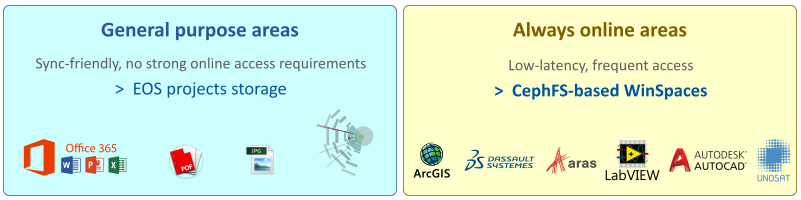

Figure 3. The criteria adopted in order to choose the most appropriate target for each existing area and workspace in DFS.

The CERNBox ecosystem today consists of several components, in order to support the different access protocols and options offered to users (see Figure 4). In particular:

- A powerful web subsystem that serves desktop and mobile sync clients via the WebDAV protocol and a feature-rich web user interface, which enables users to manage access permissions, project spaces, and restore functionalities such as a trash bin and file versioning. The back end component that enables web and sync functionalities is code-named `Reva`, and it has been made available as open source to the broader community [6];
- Access to storage by native Linux clients via FUSE mounts and native Windows clients through the Microsoft Server Message Block (SMB) file-sharing protocol, also known as the Common Internet File System (CIFS) protocol;

The system is completed by a backup component, `cback` [4], which is based on *Restic* and ensures a coherent backup strategy for all EOS- and CephFS-based storage.

3 CephFS in CERNBox

We now focus on the implementation of CephFS-based spaces for the Windows users in the Organization.

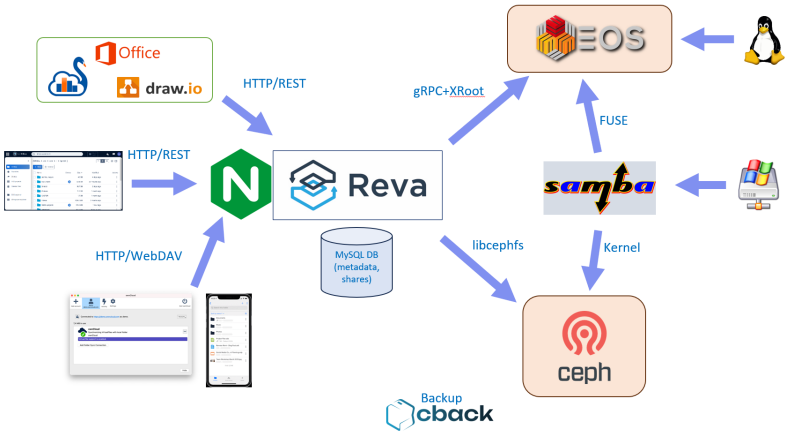

Figure 4. The CERNBox ecosystem

At CERN, we provide critical support to diverse communities that heavily rely on Windows-based applications to manage and process vast amounts of data. These communities span a wide range of scientific, engineering, administrative and humanitarian endeavours, each with unique requirements for data storage, accessibility, and collaboration. One notable example is the collaboration with UNOSAT (United Nations Institute for Training and Research Operational Satellite Applications Programme), a technology-driven initiative that harnesses satellite imagery and geospatial data to address global challenges. UNOSAT's work in humanitarian aid, disaster response, sustainable development, and peacekeeping relies on advanced data analysis and visualization tools, many of which are Windows-based, to deliver actionable insights to stakeholders worldwide.

Within CERN itself, the Product Lifecycle Management (PLM) group leverages Windows storage solutions as part of the Aras PLM platform to manage the intricate lifecycle of CERN's cutting-edge scientific equipment and infrastructure. The Aras PLM system plays a pivotal role in overseeing the design, development, deployment, and maintenance of highly specialized systems, such as those used in the Large Hadron Collider (LHC) and other experimental setups.

To support all such activities, which require SMB/CIFS network access to the storage, we extended our Highly-Available Samba [7] clusters, leveraging our experience in running a similar production service for EOS [3]. The Samba cluster supporting EOS-backed project spaces has been extended to include a kernel mount of the CephFS-based WinSpaces, and a new cluster has been deployed to serve the UNOSAT and Aras PLM communities with dedicated kernel-mounted CephFS volumes. In both cases, Samba has been configured to include the `acl_xattr` Samba VFS plugin, to enable support for Windows ACLs on CephFS. This ensures our clients the most seamless transition from the former DFS storage.

The whole CephFS capacity is provided by a dedicated Ceph cluster, which currently offers 600 TB of usable space with a 3-replica file layout. Figure 5 shows how the number of connections has been growing over the course of 2024 for the WinSpaces storage, as well as for the UNOSAT and ARAS storages.

Zooming in on one of the most relevant performance metrics, which is the data and metadata I/O operations per second (IOps), in Figure 6 we show the actual rates during about a month of operations. If aggregated data throughput is yet relatively modest with peaks at

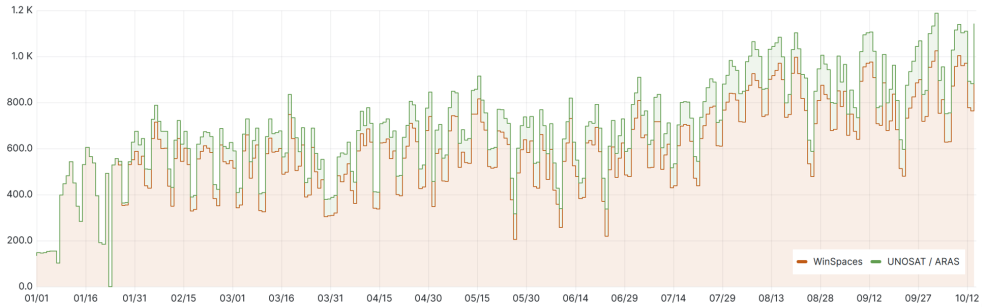

Figure 5. Number of connections to the Samba clusters hosting the WinSpaces storage (in red) and the UNOSAT / Aras storage (in green) over the year 2024.

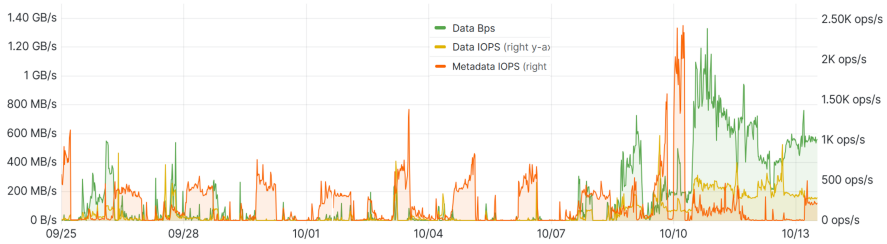

Figure 6. IOps rates over a month of operations. IOps are one of the most critical metrics for success.

about 1 GB/s, the data and metadata I/O operations rates already exceed the 1 kHz mark, even when serving such amount of data: this can be explained by the typical behaviour of Windows applications, which are extremely greedy when it comes to metadata operations, resulting in a remarkably high number of I/O operations per actual chunk of data served. As we expect the load to keep increasing with the ongoing migration, we recognize the need to scale out the capacity of the system to meet this peculiar usage.

Finally, Figure 7 shows the growth in terms of volume usage during 2024. In addition to the active migration of several DFS workspaces, UNOSAT has been migrating a significant fraction of their data from DFS, resulting in a large growth. We anticipate a similar growth in 2025, as we will perform the full migration of the remaining DFS workspaces, whereas in the future, we expect a growth comparable to the CERNBox EOS-based storage.

4 Conclusions and Outlook

In this contribution, we have described the evolution of CERNBox and how CephFS has been integrated into the ecosystem to address the Windows-based workflows at CERN. We have shown how the SMB/CIFS access is run in production, pairing a similar service for the EOS-based storage.

To complete the offer for Windows-based activities, the web access requires further development prior to being opened to the entire user community in production. During the course of 2025, we will implement a number of features, such as full support for ACLs and the ability to restore data from a trash bin, in order to provide a feature set similar to the EOS-based offer.

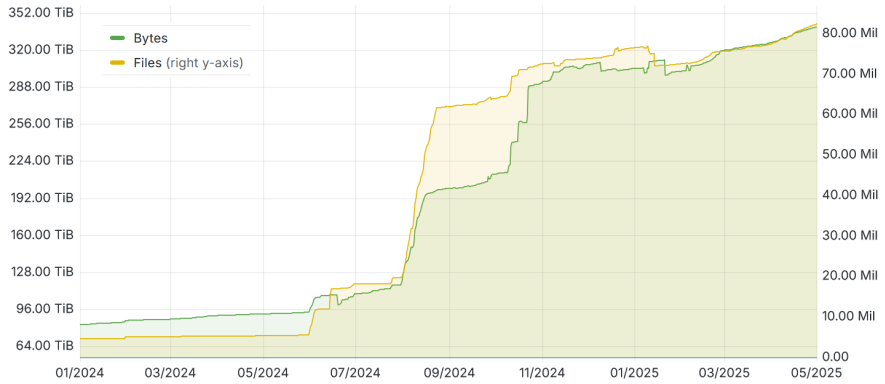

Figure 7. Volume usage rapidly grew in 2024, both in terms of data and number of files, as more and more use cases were migrated.

In this context, it is worth noting that we have received a number of expressions of interest from sites such as the University of Paris-Saclay, the University of Vienna, and other research institutes, who started to deploy proof-of-concept setups based on CERNBox.

We are confident that those developments, paired with the widespread adoption of CephFS within the scientific community, will make CERNBox an attractive out-of-the-box solution for implementing a scalable collaborative cloud storage service, able to serve all diverse workflows of the scientific community.

References

[1] A.J. Peters, E.A. Sindrilaru, G. Adde, *EOS*, Journal of Physics: Conference Series **664**, 062037 (2015)

[2] G. Lo Presti, S. Bukowiec, L. Mascetti, H. Gonzalez Labrador, V. Bippus, A. Smyrnakis, M. Kwiatek, J. Moscicki, *CERNBox as the hyper-converged storage space at CERN: integrating DFS use-cases and home directories*, poster at CHEP (2019), https://indico.cern.ch/event/773049/contributions/3474470

[3] G. Lo Presti, A. Brosa Iartza, S. Bukowiec, *Samba and CERNBox: Providing online access to Windows-based users at CERN*, EPJ Web of Conferences **251**, 02024 (2021)

[4] R. Valverde Cameselle, H. Gonzalez Labrador, *Addressing a billion-entries multi-petabyte distributed file system backup problem with cback: from files to objects*, EPJ Web of Conferences **251**, 02071 (2021)

[5] G. Lo Presti, *User and Collaboration Storage: CERNBox*, Tech Week Storage 2024 (Presentation), CERN, Geneva, https://indico.cern.ch/event/1353101/contributions/5805537

[6] Reva: the Interoperability Platform, https://github.com/cs3org/reva

[7] Samba, https://www.samba.org

[8] Microsoft Windows, https://www.microsoft.com/windows, used with permission from Microsoft

[9] Microsoft Distributed File System, https://learn.microsoft.com/en-us/windows-server/storage/dfs-namespaces/dfs-overview (access time: 03/01/2025)