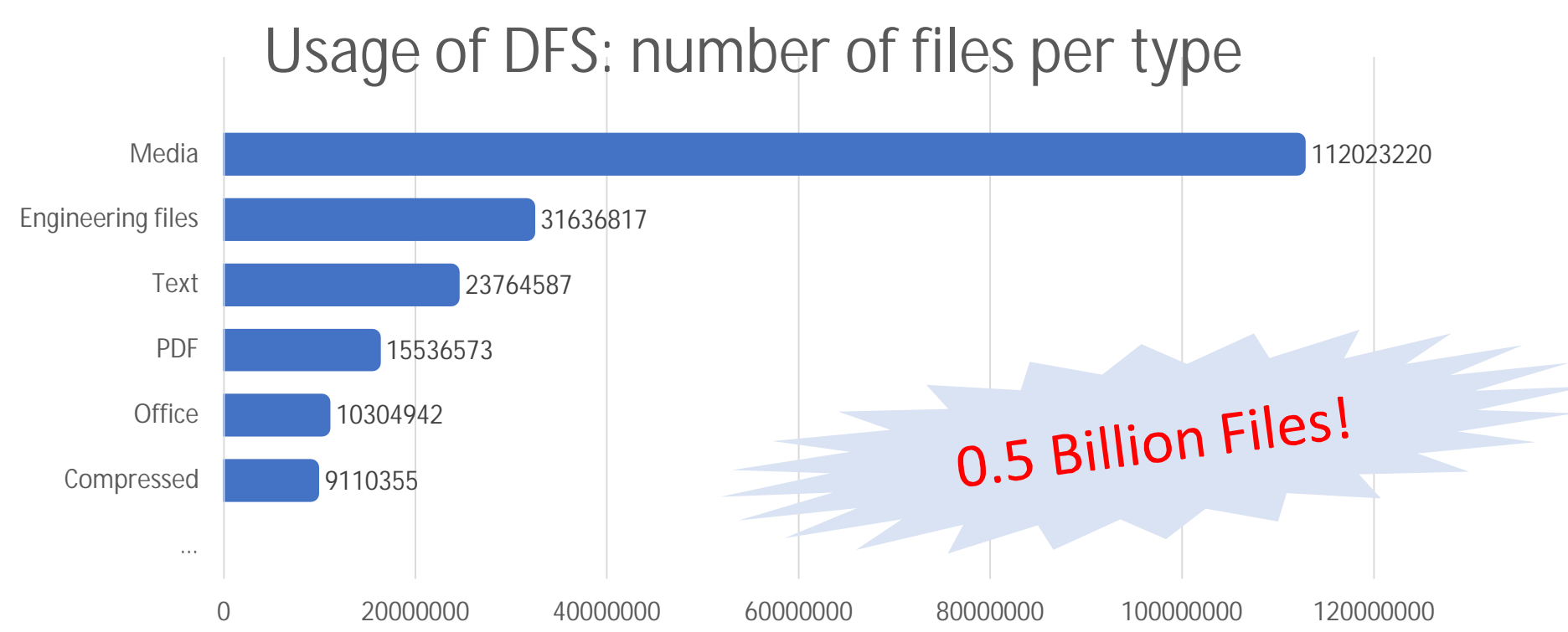# TURNING CEPHFS INTO A COLLABORATIVE SPACE WITH CERNBox

G. Lo Presti, S. Bukowiec, D. Castro, H. Labrador

CERN

CHEP 2024

Motivation and Context

CERN IT is consolidating DFS into CERNBox, the cloud storage offered to the whole Organisation.

Usage of DFS: number of files per type

Media	112023220
Engineering files	31636817
Text	23764587
PDF	15536573
Office	10304942
Compressed	9110355

0.5 Billion Files!

DFS has been used since the late 1990s as a general storage system for Windows, and it supports many use cases and workflows.

Multiple alternatives were analysed, including NetApp and CephFS with HDDs and SSDs. Dedicated tests were developed to compare DFS with other solutions.

An Analysis was made to establish the optimal target location for each application and use case.

Most can be accommodated in the existing CERNBox EOS-based offer.

Others have more stringent requirements in terms of latency → Designed and prototyped new solution, invited early adopters to identify potential issues.
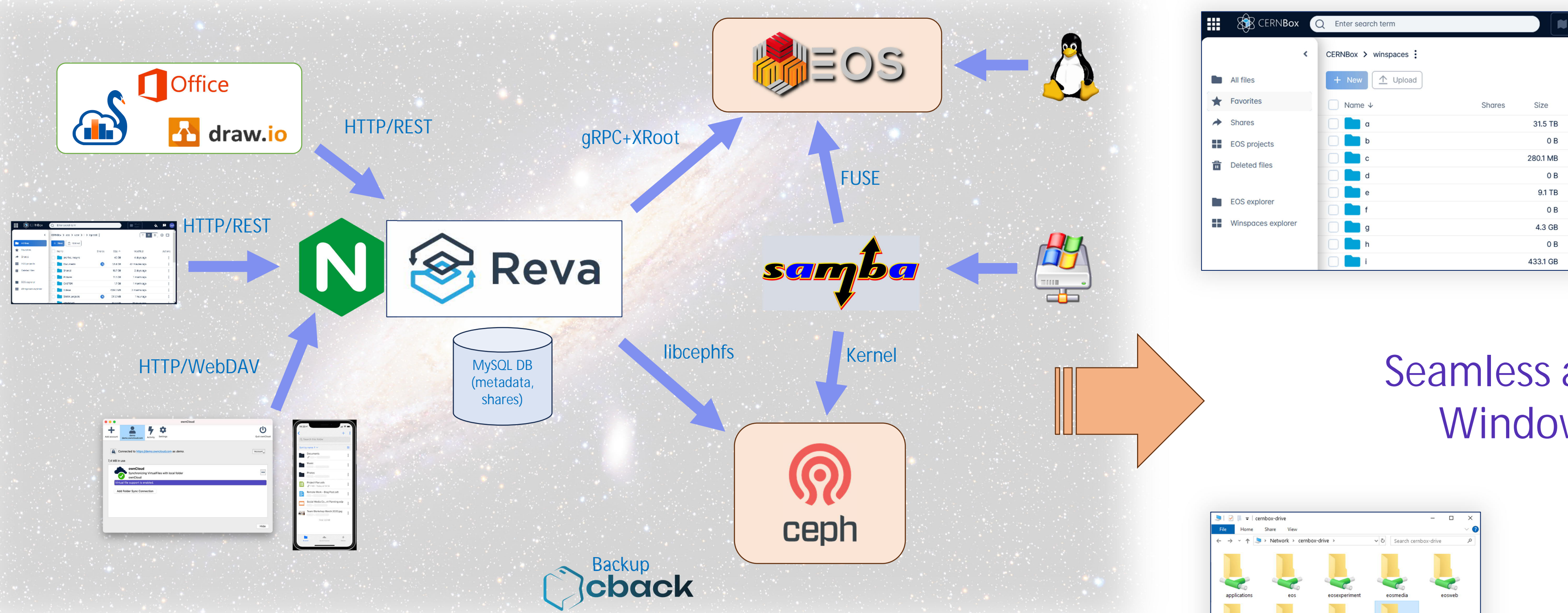
General purpose areas
Sync-friendly, no strong online access requirements
> EOS projects storage

Office 365

Always online areas
Low-latency, frequent access
> CephFS-based WinSpaces

ArcGIS   DASSAULT SYSTEMES   aras   LabVIEW   AUTODESK AUTOCAD   UNOSAT

CephFS is the best choice thanks to its raw IOps performance and feature set

CephFS in CERNBox

The «CERNBox Galaxy»: a complete deployment of CERNBox at CERN, including CephFS storage.

Office   draw.io
HTTP/REST
HTTP/REST
HTTP/WebDAV
N   Reva
MySQL DB (metadata, shares)
gRPC+XRoot   EOS   FUSE
samba   libcephfs   Kernel
ceph
Backup cback
image: Wikimedia

Web + WebDAV access powered by Reva, CERNBox's open-source backend based on gRPC. (github.com/cs3org/reva)

Seamless access via Web UI and Windows Network Shares

Windows Network Shares are served via High-Available SAMBA clusters, leveraging multi-year experience with production service for EOS, with ever-growing usage by clients.

CephFS support in CERNBox is readily available for other Institutions to try, please get in touch if interested!

Being deployed at multiple sites, e.g. Uni Saclay, Paris

doyle: CephFS Pool IO (last month)
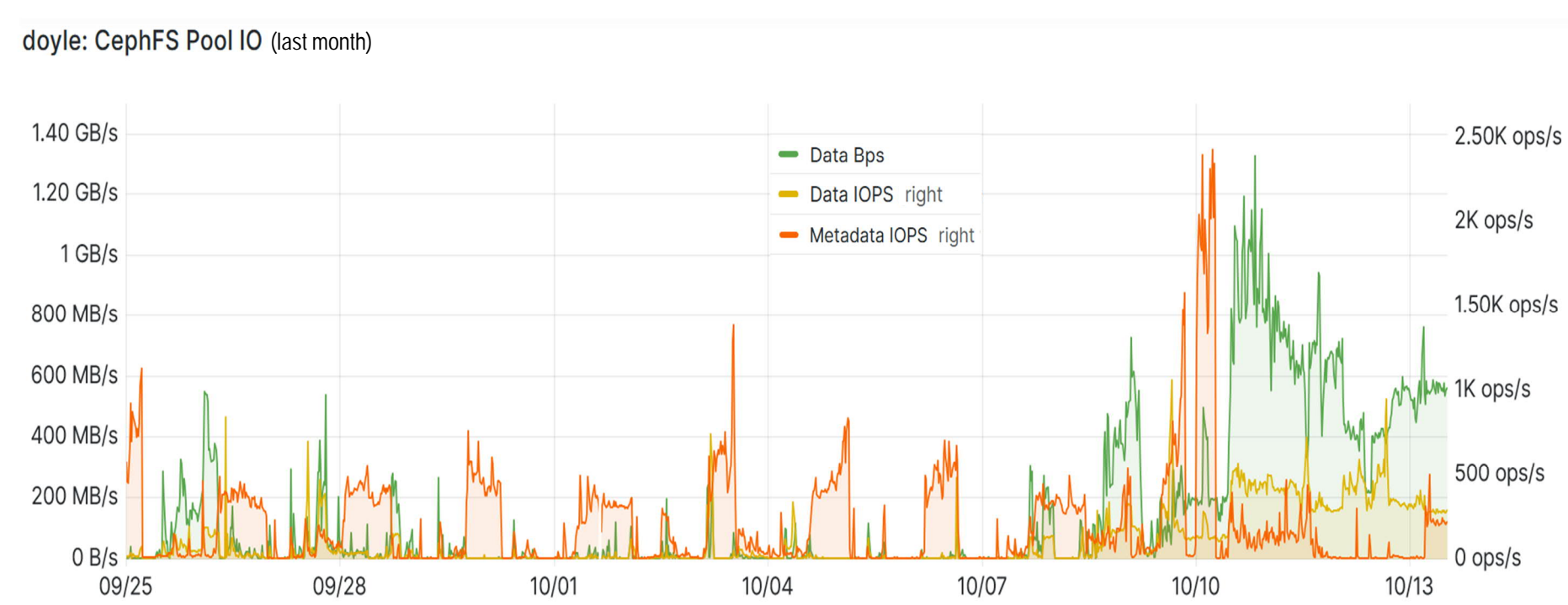
IOps Performance: the most critical metric for success.

doyle: Total Space and Namespace (year to date)

Volume usage rapidly growing as more and more use cases are migrated.
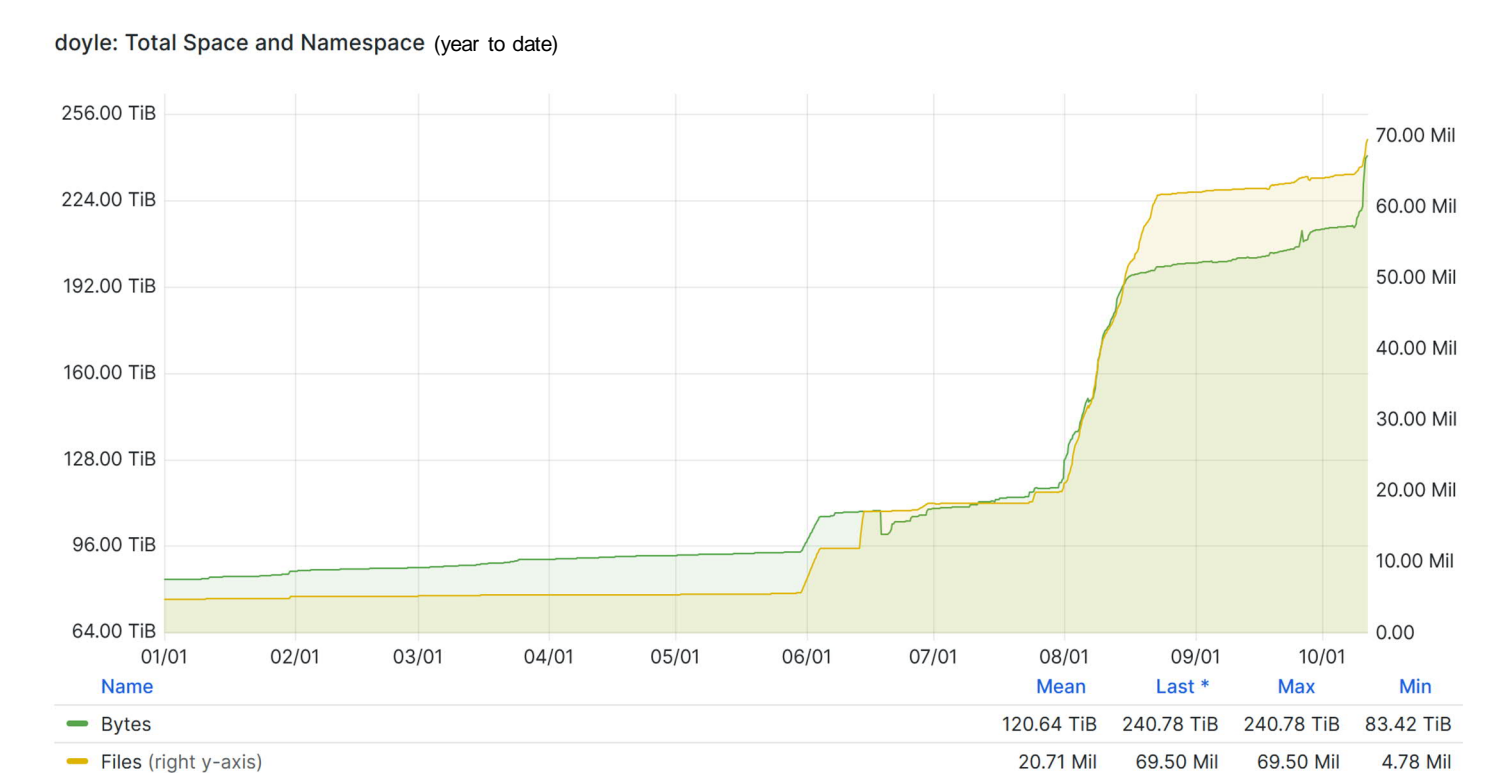
Outlook:
CERNBox as the hub to seamlessly integrate multiple heterogeneous storages, for all diverse workflows of the scientific community.

cern.ch/cernbox

Contact
Giuseppe.LoPresti@cern.ch