IOPscience iopscience.iop.org

Home Search Collections Journals About Contact us My IOPscience

CERNBox + EOS: end-user storage for science

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2015 J. Phys.: Conf. Ser. 664 062037

(http://iopscience.iop.org/1742-6596/664/6/062037)

View the table of contents for this issue, or go to the journal homepage for more

Download details:

IP Address: 188.184.3.56
This content was downloaded on 08/03/2016 at 22:14

Please note that terms and conditions apply.

CERNBox + EOS: end-user storage for science

L Mascetti, H González Labrador, M Lamanna, JT Mościcki and AJ Peters

CERN European Organization for Nuclear Research, 1211 Geneva, CH

E-mail: `luca.mascetti@cern.ch`

Abstract. CERNBox is a cloud synchronisation service for end-users: it allows syncing and sharing files on all major mobile and desktop platforms (Linux, Windows, MacOSX, Android, iOS) aiming to provide offline availability to any data stored in the CERN EOS infrastructure. The successful beta phase of the service confirmed the high demand in the community for an easily accessible cloud storage solution such as CERNBox. Integration of the CERNBox service with the EOS storage back-end is the next step towards providing "sync and share" capabilities for scientific and engineering use-cases. In this report we will present lessons learnt in offering the CERNBox service, key technical aspects of CERNBox/EOS integration and new, emerging usage possibilities. The latter includes the ongoing integration of "sync and share" capabilities with the LHC data analysis tools and transfer services.

1. Introduction

The IT Data & Storage Services group (IT-DSS) at CERN is responsible for the main data-related services for High Energy Physics (HEP) community. With its data services it supports users of experiments collaborations based at CERN, notably the LHC experiments (ALICE, ATLAS, CMS, LHCb), the major experiments using the CERN accelerator complex and AMS, the spectrometer installed on the International Space Station.

In recent years the access to data for physics analysis has become easier due to improvements in the CERN central data services used to store experimental data (CASTOR and EOS[1]) and user data (AFS). However, the integration of these large repositories with end-user computers (such as laptops) is left to the users. Many users handle data locally on their machines: in the long run this might become complex since each user is left with ad-hoc procedures to back up the local data, share it across different devices and users, import-export data with our central facilities.

From the point of view of mobility, users expect to access their data via different types of devices (smartphones and tablets) which increases the difficulty to maintain a coherent view of the data. None of the IT-DSS central services support disconnected operations.

A solution adopted by some users was to use popular synchronisation and sharing services like Dropbox™: analysing the network traffic and following up on users support requests we observed that these were used not only for private use but also to share professional data (presentations, root files, code, etc...). However, externally hosted cloud storage providers, such as Dropbox, are not integrated with the CERN data services, and are not suited to keep certain types of data (for example due to size or confidentiality policies).

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

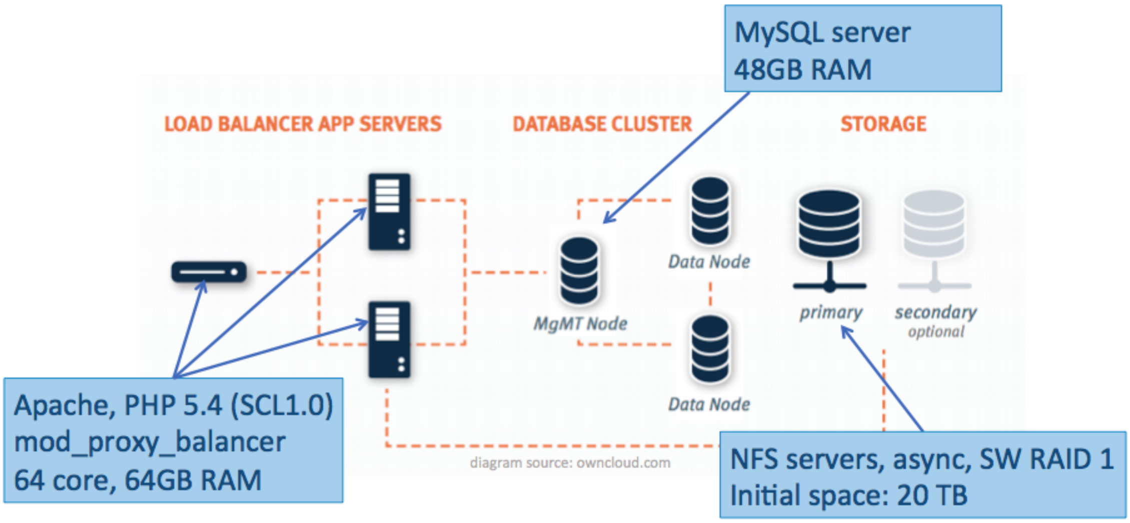

Figure 1. CERNBox pilot architecture with NFS storage as backend.

IT-DSS introduced CERNBox, a synchronisation and sharing service, to solve these problems and to promote a new integrated way of accessing data for scientific research[2].

The core of CERNBox is a synchronisation and sharing system based on components provided by the OwnCloud^{TM}open software stack[3]. CERNBox provides a service layer on top of storage systems already provided by IT-DSS. This allows coherent data access policies to be applied and for service levels compatible with CERN standards to be defined. The OwnCloud software source code is freely available which is beneficial for systems integration in novel ways and provides the desired level of control of the source code base. OwnCloud maintains clients for a variety of platforms: Windows, MacOSX, all mainstream Linux flavours, Android and iOS.

OwnCloud allows two main routes to access the data:

Direct web access: via any web browser the user can upload/download/remove files and share files or directories with other users. Some extensions (available as application plugins in the OwnCloud server) provide added functionality such as easing the access to certain types of files (e.g. image viewers). Apps for Android and iOS can make this type of access easier from smartphones and tablets (complementing direct web access via the device browsers).

Sync client: this is a component, integrated in the user desktop environment, which keeps one (or more) local folders in sync with the central storage server. Users can work on their laptop without connectivity and the sync client reconciles changes when the network connectivity is restored.

In 2014 we set up a pilot service (NFS storage as backend) which was operated with minimal operational effort, see Figure 1. This service demonstrated a growing demand for file sync and share, and reinforced our assumption on the potential popularity of such a service.

We have observed a constant increase of the number of users of the pilot service (several tens of new users weekly) both from the physics community and from other communities such as the general laboratory services and the accelerator departments. The type of usage was consequently broad, ranging from the synchronisation of data files (ROOT analysis files), arbitrary work files to office documents shared with colleagues.

2. Vision

We federated CERNBox with our disk storage system, EOS[4]. The current usable capacity of EOS is about 70 PB with proven capabilities in terms of performance (designed to support LHC analysis) and reliability (geographically distributed replicas)[5]. For comparison, the entire AFS hosts about 300 TB of user data.

CERNBox could become the main entry point for the entire CERN data store: users will selectively access or synchronise data with their devices while their entire data storage is accessible online by the batch jobs for massive data processing or for the creation of large data archives. The choice of EOS allows for the distribution of large user quotas (presently 1 TB per user).

A preliminary survey of use cases suggests that our users will benefit (on top of the standard synchronisation and sharing) from the seamless access of data at different levels (home directories, group spaces) in a shared way. The "disconnected operations" provided by the sync client is a welcome feature.

In the future other activities may benefit from a sync client becoming a one-way synchronisation agent between the data acquisition and the central services. For example small experiments need to develop ad-hoc scripts and procedures to upload data from their acquisition system to the CERN central data store: a one-way sync client will be a fundamental simplification. One should note that the current client sync behaviour does not yet support this mode (removal of files from the client will remove the same files in the central repository).

These are general use cases and are more widely applicable than in the High-Energy Physics workflows. CERNBox can satisfy them in a pretty straightforward way: images from surveys in the field (e.g. geo-sciences), readings of devices or upload of data from a remote machine (sensors, etc...).

3. Integration

The OwnCloud software provides the possibility of using any file system (and more recently some object store systems) for hosting the user data. Files are stored in a straightforward directory structure. Metadata (in particular, the one controlling the synchronisation) is stored in a relational database which ultimately controls the behaviour of the sync clients.

In its core implementation OwnCloud assumes that the backend file system is visible only via the OwnCloud servers: this is clearly incompatible with allowing users to directly access their data (via POSIX commands like cp, mv and rm). OwnCloud also allows an "External Storage Access", unfortunately this solution does not scale well: the whole directory tree on the external storage must be scanned periodically for updates and integration of independent storage systems in the way attempted by OwnCloud is operationally complex.

The need of a relational database is also incompatible with a scale-out system of the size that we need at CERN. For example EOS contains $\mathcal{O}(10^8)$ files and AFS $\mathcal{O}(10^9)$.

However, because of the state-based nature of OwnCloud synchronisation algorithm and straightforward filesystem organisation in the OwnCloud backend server it was possible to adapt EOS to conform to the OwnCloud synchronisation protocol.

3.1. Filesystem organisation

The solution to the database problem has been simply to remove it and delegate its functionality to EOS itself, since EOS is a shared filesystem with a hierarchical namespace. The key piece of information for the OwnCloud sync algorithm is a unique tag (called ETAG) to identify file updates and detect conflicting changes. This is now coherently handled and provided as an extended attribute by the EOS namespace (as a combination of inode number and the content checksum).

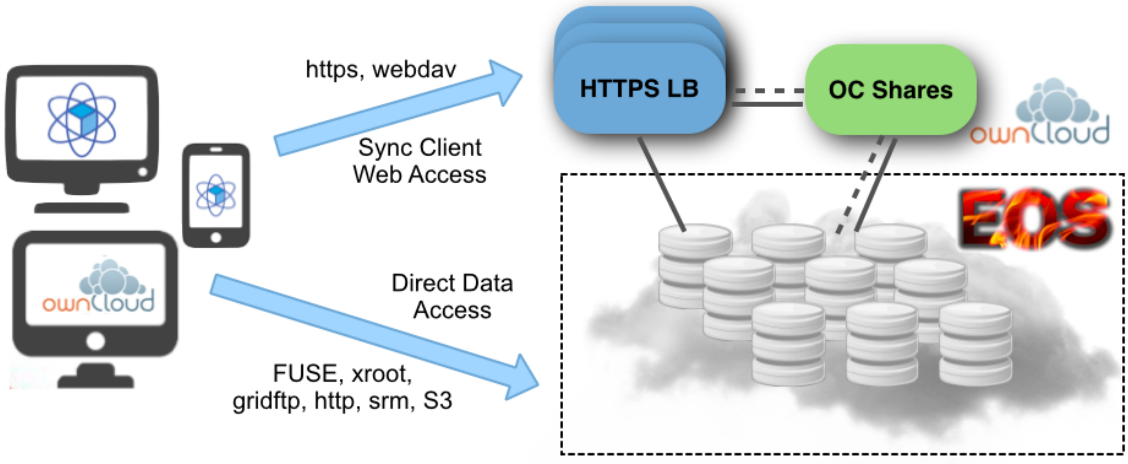

Figure 2. CERNBox architecture with EOS backend.

It is important to stress that the sync client is completely unchanged: we do not want to break the compatibility with standard clients developed by OwnCloud for several platforms. For the web server we have developed a set of plugins to integrate with EOS and we also worked on general improvements in the core architecture. We seek to integrate these changes with the OwnCloud upstream.

EOS presents a WebDAV interface (with OwnCloud protocol extensions) to the synchronisation clients. The data streams (file upload and download) are redirected to EOS storage nodes while the metadata operations (file removal, directory rename etc) are handled by the EOS head node. This "out-of-band" handling of traffic solves the potential data transfer bottleneck of having all transfers channeled through the OwnCloud server. The OwnCloud server is used to manage sharing and web access.

4. EOS backend for CERNBox

The backend storage for general use via CERNBox was setup as a dedicated EOS instance named *EOSUSER*.

Users have several ways of accessing their files directly on the *EOSUSER* instance in the context of scientific computing. Command line tools (notably the xrdcp command) allows the users to use sophisticated features such as 3rd-party copy. EOS can also be fuse-mounted and files can be accessed via standard unix command (e.g. ls, cp, mv). An alpha-tester programme is being set up to enable selected users to directly access *EOSUSER* and to get experience with this key functionality of the new system.

At present users have a default 1 TB quota. The accounts are created on the fly (using the CERN LDAP to authenticate users).

A prototype of full integration with CERN egroups (essential to easily share files with groups of colleagues) is ready and the version now in production will also handles nested groups.

A prototype of a full integration with the CERN SSO (Single Sign-On infrastructure) has been prepared as well.

5. Service evolution

To study the system during the acceptance and integration testing phases we developed a test-suite called smashbox: it is written in Python and is publicly available on github[6]. Complex tests where concurrent sync clients access the repository in a controlled way can easily be written.

We currently use smashbox to validate new releases of all CERNBox components and to assure smooth upgrades.

One of the most basic tests defined in smashbox is to have one sync client ("uploader") generate files with known checksums and a variable number of sync clients ("downloaders") to synchronise with the CERNBox repository and check data integrity. This is used as a simple functional test but also on a massive scale for stress-testing.

Smashbox receives contributions from other users of OwnCloud and the OwnCloud team complements their testing with smashbox tests.

File transfers in CERNBox are protected by checksums (as we do on CASTOR and EOS). This feature has been jointly developed by the CERN IT-DSS team and OwnCloud Inc. CERNBox is the first OwnCloud-based deployment in the world with the file checksumming in production. We presently collaborate with the OwnCloud development team to integrate other features to ease large-scale operations, such as a centrally managed re-upload functionality for the system admins.

We have also developed the necessary operational structure behind the service, most notably using the EOS archive functionality to provide disk-to-disk backup (disaster recovery: Meyrin/Wigner)[7]. The *undelete* functionality of EOS has been used to implement the trashbin for user's files.

6. Operational Experience

CERNBox has been providing a very stable and robust service with very little downtime for scheduled maintenance. In the future we expect to provide completely transparent operations. The performance of the system is on par with a rich set of features and has been appreciated by the end-users. This is often reflected in the direct feedback from our users.

The number of connected synchronisation clients and aggregated data transfer are the two main dimensions of the scalability of CERNBox. Sync clients poll for updates at regular time intervals and it is easy to measure the number of connected users in a unit of time (see Figure 3). There are currently 2100 user accounts in CERNBox. In the present configuration the *EOSUSER* namespace can handle at least one order of magnitude more users without visible performance impact. In terms of data transfer the system is horizontally scalable as fully independent front-end proxy servers may easily be added and the data traffic is split across several storage nodes on the EOS backend.

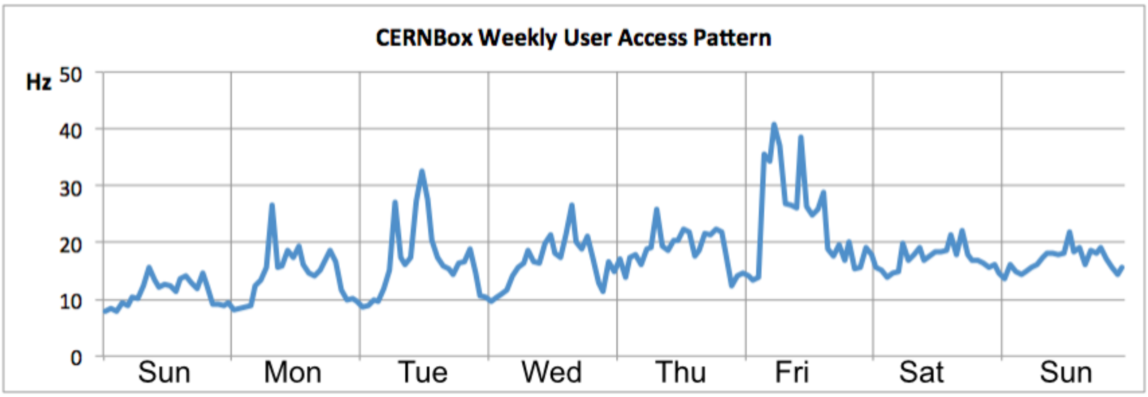

Figure 3. Access frequency of CERNBox clients during a week.

Journal of Physics: Conference Series **664** (2015) 062037 doi:10.1088/1742-6596/664/6/062037

7. Conclusion

The good experience in running CERNBox and the encouraging user feedback confirm our initial assumptions on the actual need of a cloud storage service for file sync and share. More importantly this service has the potential to become a central place for user data access since it addresses a number of issues, such as mobility and multi-platform access. It is an innovative service for sharing data between users and we expect this to become an environment that could fit physicists, engineers and general laboratory staff.

We believe that our system could provide new communities an entry point for "big data" activities as discussed in a recent cloud storage event organised at CERN[8].

References

[1] "EOS website" https://eos.cern.ch
[2] JT Mościcki and M Lamanna. "Prototyping a file sharing and synchronization service with Owncloud." *Journal of Physics: Conference Series*. Vol. **513**. No. 4. IOP Publishing, 2014.
[3] "OwnCloud website" https://www.owncloud.org
[4] AJ Peters and J Lukasz "Exabyte Scale Storage at CERN." *Journal of Physics: Conference Series*. Vol. **331**. No. 5. IOP Publishing, 2011.
[5] L Mascetti et al. "Disk storage at CERN." CHEP 2015 - these proceedings.
[6] "SmashBox repository" https://github.com/cernbox/smashbox
[7] EA Sindrilaru et al. "Archiving tools for EOS." CHEP 2015 - these proceedings.
[8] "Workshop on Cloud File Syncronisation and Sharing Services." CERN, Geneva (CH), Nov. 17/18 2014, https://indico.cern.ch/event/336753 .