

Overview and test of on-premise open-source and online proprietary collaboration systems for establishing IT business environment in small and medium-sized companies.

Matej Rabzelj, Aljaž Martinčič, Tjaša Jereb, Andrej Kos

University of Ljubljana, Faculty of Electrical Engineering

E-mail: matej@rabzelj.si, aljaz.martincic@lufe.org, tjasajereb@svet.fe.uni-lj.si, andrej.kos@fe.uni-lj.si

Abstract

Every company faces the problem of establishing a functional IT business environment in order to enable productive work. Selection of the most optimal digital collaboration solution requires balancing different factors, such as user privacy, service scalability, and its cost. This paper briefly describes different aspects of some of the most widely-used self-hosted open-source online collaboration software and its commercial cloud-hosted competitors. Furthermore, it describes the setup of a promising open-source project in a simulated small-sized IT business environment.

1 Introduction

Today, hundreds of different online collaboration services are available, ranging from essential file storage utilities to complete in-browser office solutions. The vast majority of their market share belongs to some of the most prominent online players, such as Google, Dropbox, Atlassian, and several others. These services are usually provided to professional customers for a certain monthly subscription, or in some cases entirely for free, but with severe usability restrictions or in the form of an ad-supported product. Nevertheless, nearly all of the available options come with some form of constraints regarding either the storage size and bandwidth capacity or privacy policy concerns and extensibility restrictions. More demanding customers or companies with numerous users may have to pay very high fees or even develop custom solutions in order to get a flexible service that truly fits their business model and needs. The combined costs of volume software licenses and multiple third-party service subscriptions may not always lie within the company's financial structure or justify their use. Attention to this problem has recently been brought to light by the several of the world's leading research institutions, including CERN's migration to open source software to avoid vendor and data lock-in [1] and NASA's switch to Linux on board of the International Space Station [2]. Following these examples, we selected the Nextcloud project from the vast array of available open-source collaboration services, compared it with two commercial third-party solutions, and set up an on-premise IT collaboration system.

2 Licensing, cost, privacy and security

License costs and subscription fees are only some of the aspects when debating about the usage of open source solutions in a business environment. Intellectual property and EULA (End User License Agreement) licenses must fit the company's policy before a decision of purchase is made. Some organizations, for example, may not be comfortable using a proprietary licensed software due to possible privacy concerns, while others may intentionally avoid AGPL, GPL or even LGPL license, as it obligates them to release the source code of distributed binaries or modified libraries [3].

Other issues, particular to cloud-based solutions include the questions of data ownership, its confidentiality, and the customer's privacy. When customers upload their data to a file sharing or online processing service, that data is treated regarding the service provider's ToS (Terms of Service) and its Privacy Policy. Sensitive data may cross borders and can be inspected by different computer algorithms or even forwarded and processed by subcontractors or third parties. These service providers and their privacy policies are not necessarily under the jurisdiction of the same regulatory framework as in the customer's state. Individual users may, therefore, be profiled and find themselves subject to a targeted advertisement.

On the other side, companies attempting to implement their own version of open source projects, which by most licenses comes "as is" and "without warranty or guarantee of any kind" [3] need to understand the risks involved and assume complete responsibility for their system's operation. Particularly noteworthy is adequate IT and hardware security expertise to aid in achieving compliance with local regulations and standards. Furthermore, a small to a medium-sized organization may never be able to achieve the same data security standards in terms of geolocation and physical redundancy as its most substantial cloud-providing counterparts.

3 Overview of typical online collaboration system's features

The essential components of online collaboration are file sharing and permission management. Such a system should support secure file storage, either in a centralized on-premise location or in a decentralized manner (a server cluster, geographically distributed servers). It should also provide user-friendly file access via the web or desktop and mobile client application. Advanced file-sharing solutions typically also offer file versioning and remote directory mounting via various file transfer protocols such as WebDAV, SMB and FTPS. Additionally, office collaboration requires some form of internal task management and team communication. Some systems implement file tagging, per-person task management, real-time document co-authoring, and private chat platforms. Rare solutions also include a self-hosted e-mail server, as most only offer an online interface for e-mail integration. While an integrated e-mail client does streamline the business workflow, it does not guarantee communication security and user privacy as data still resides on a third-party server. Lastly, per-employee PIM (personal information management) suite should include a system for calendar and contact information sharing to aid meetings and complete the digital transformation of business.

While it is desired that collaboration software includes the functionalities mentioned above, it should not sacrifice information security to do so. It is therefore expected that software follows the latest security guidelines, such as TLS communication encryption of communication, CSP (Content Security Policy) 3.0 for web interfaces, two-factor authentication support, support for per-file or full disk encryption, and revokable methods of file sharing. When comparing different solutions, it is also worth noting how scalable those systems are. Self-hosted solutions for small to medium businesses typically prevent vendor lock-in in the long run but can pose a problem as the company and its demands grow. The comparison should, therefore, include support for third-party extensions (distributed via a trusted source) or store data in a way that allows either API access or direct filesystem manipulation on POSIX systems. The latter two options enable companies to easily extend the open source software and further tailor the system to their needs.

4 Comparison of popular online collaboration solutions and their features

Despite evaluating several open-source and proprietary collaboration solutions centered around file-sharing, such as Nextcloud, Seafile, Pydio, Kolab Groupware, GSuite, Dropbox, and Box, their detailed comparison is far beyond the scope of this article. Instead, the following section provides a brief overview of some of the most popular solutions available on the internet and presents their most prominent features as well as potential disadvantages in the following table (Table 1).

Table 1: Comparisons of popular open-source and proprietary online collaboration suites [4]. Comparison is based on criteria with an emphasis on security, file manipulation features, document co-authoring, PIM, user communication options, and system extensibility.

Feature	GSuite	Nextcloud	Dropbox
On premises	No	Yes	No
Source model	Closed	Open	Closed
Setup complexity	Easy	Difficult	Easy
Server-side encryption	Yes	Yes	Yes
Client-side encryption	No	Yes	No
Brute-force prevention	Yes	Yes	Yes
ENV variable authentication	No	Yes	No
LDAP support	Yes	Yes	Yes
Two factor authentication	Yes	Yes	Yes
NFS support	No	Yes	No
CIFS support	No	Yes	No
File versioning	Yes	Yes	1 year
Access Control Lists	Yes	Yes	Yes
Online file preview	Yes	Yes	Yes
PIM features	Yes	Yes	No
Audio, Video, Text Chat	Y/Y/Y	Y/Y/Y	No
Integrated e-mail server	Yes	No	No
Inter-server sharing	No	Yes	No
Lan synchronization	No	No	Yes
Online office	Yes	Extension	Third-party
Desktop client	Yes	Yes	Yes
Mobile client	Yes	Yes	Yes

Compared features represent several different system aspects and can be roughly categorized into the following sections:

- licensing model and system setup complexity
- available security and authentication mechanisms
- file storage and file manipulation mechanisms
- PIM, live document co-authoring and team communication options

Feature selection criteria is based on approximation of real-world use cases in various business environments. The first category usually dictates the service setup and operational costs. While commercial cloud-based software tends to have simple installation and low implementation price, its costs tend to significantly grow with the increasing number of system users (company employees). Inversely an organization may freely obtain community releases of open source solutions without any costs but has to possess the expertise and network architecture to install and maintain such a service.

The second section provides an overview of common authentication mechanisms, which may be one of the deciding factors for integration with existing digital infrastructure. Adequacy of file storage mechanisms largely depends on the nature of work and the consequential file volume and size. The typical uplink bandwidth limitations may also play an important role and require a fast locally attached storage as an alternative. Lastly, a vast diversity of PIM, co-authoring, and team intra-communication functionalities justifies the title of online collaboration solutions.

5 Test environment

Research and testing showed that the Nextcloud system was one of the easiest and most polished self-hosted solutions to set up among comparable open-source projects. Its base installation provided most of the expected functionality. Nextcloud was also easily extended through an integrated extension store with open-source applications. The following functionality and performance test were therefore performed on a Nextcloud instance.

Tests were conducted on networking equipment that simulates a small business environment. Hardware configuration is listed below.

- Dell R710 Xeon X5670 6 core server with 12 TB RAID10 SSD + HDD storage
- RouterBOARD 3011UiAS-RM router
- Mikrotik CSS326-24G-2S + RM switch
- RIPE Atlas probe (uptime and link-state statistics)

The host system includes a modified Debian Linux distribution with Proxmox kernel and KVM virtualization utilities. It is installed on RAID + LVM stack providing redundancy and partitioning flexibility (Figure 1). All disks are LUKS (Linux Unified Key System) encrypted. The system can be remotely operated and provides a secure key-based SSH environment for boot-time disk decryption.

The platform hosts several virtualized servers with bridged paravirtualized (virtio) NICs and dedicated LVM partitions for near-native performance. Servers include an Nginx reverse proxy instance for web access and a dedicated Nextcloud installation on a LAMP (Apache 2.4, MariaDB and PHP 7) stack [5]. Nextcloud stores data directly on EXT4 formatted volumes, locally exposed via NFS. Nextcloud core installation is further extended with a two-factor authentication module, secure online password management interface (KeePass), complete PIM (contacts, calendars, notes, tasks, polls, and e-mail) suite with native support for iOS and Android synchronization via CardDAV and CalDAV protocol. Furthermore, the Collabora package provides a shared real-time cloud document editing. Individual files can be tagged and shared securely (configurable password protection, expiration date, user permissions) within local groups, or to third party users. The e-mail server, however, is configured as VUMS (Virtual User Mail System), relying on Postfix and Dovecot running on a separate virtual machine.

```

[root@pve ~ # lsblk
NAME
loop0
loop1
sda
├─sda1
├─sda2
├─sda3
└─sda4
   └─sda4_crypt
      ├─pve-swap
      ├─pve-root
      └─pve-virt_tmeta
         └─pve-virt-tpool
            ├─pve-virt
            │   ├─pve-vm--105--disk--1
            │   ├─pve-vm--105--disk--2
            │   ├─pve-vm--104--disk--1
            │   ├─pve-vm--106--disk--0
            │   └─pve-vm--106--disk--1
            └─pve-virt_tdata
               └─pve-virt-tpool

```

Figure 1: Visualization of the virtual storage configuration of the first SSD drive. Note the encrypted LUKS container and thin-provisioned LVM structure. The unencrypted boot partition, which includes initramfs with integrated networking and SSH environment is also visible.

6 Functionality and performance results

The system was test loaded with 1.7 TB of real user-generated data (Figure 2). Files ranged from small text documents to multi-gigabyte raw disk image backups of virtual servers, totaling in 617842 inodes.

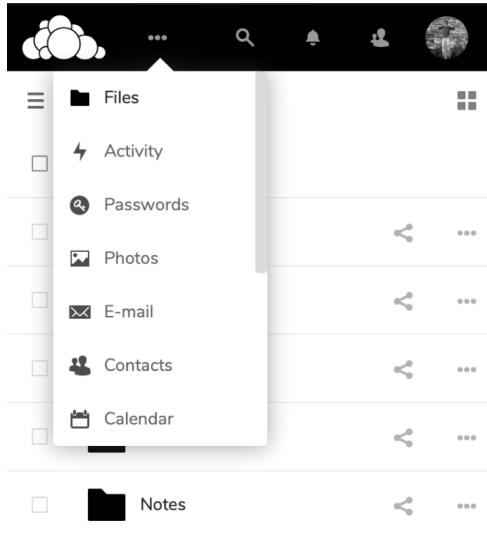


Figure 2: Nextcloud's web interface on a mobile device. File management is the system's central feature and a default application on the main screen. Additional functionalities are presented in the form of modular extensions, accessible via a drop-down menu.

Configured Nextcloud instance has been successfully operating for over 8 months, serving 9 active users and participating in 3 inter-server federated shares. The service was regularly updated and remained stable ever since. The virtual machine's CPU load mostly varied with user activity, file indexing, and scheduled cron operations. Memory consumption remained astonishingly low, and even noticeably dropped after one of the updates (Figure 3).

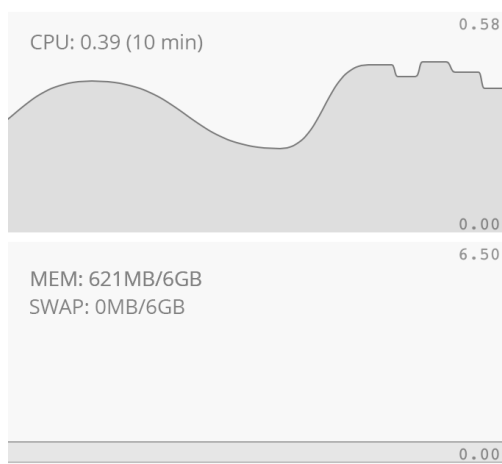


Figure 3: Virtual machine's CPU utilization and memory consumption in the idle state.

7 Conclusion

Online collaboration systems enable digital business transformation by providing a centralized environment for user communication and project management. The combined services of file sharing, task and resource management, event scheduling, live document co-authoring, and employee communication streamline a company's typical workflow and reduce its organizational overhead. Moreover, self-hosting open-source online collaboration systems allow organizations to cut down on costs of similar third-party services, provide faster content access in the local network environment, as well as brand, tailor and extend the software to their needs. Modern solutions of such systems prove to be secure and scalable, which allows businesses to keep data under their control and retain their privacy policy. On-premise hosting of such systems, therefore, provides a viable alternative for organizations of various size.

However, the major drawbacks of this solution include uneven cost distribution due to an initial hardware investment, a considerable set-up complexity, and the requirement of an organization to assume complete responsibility for the system's operation. The latter requires companies to possess technical know-how and makes this method unsuitable for businesses without a dedicated IT department. The costs of system maintenance, potential data loss or security breach due to an improper service configuration may by far exceed a license fee of a comparable commercial offering.

Finally, we conclude that the most notable delimiting factors of choice between a third-party, commercial collaboration suite and an open-source on-premise deployment seem to be requirements of privacy policy retention and specific use-cases of system configurability. A real-world example of the latter could be illustrated on a creative agency in need of fast, large drive arrays of locally exposed network storage for real-time, high bitrate content editing, and remote access for content deployment, live online script co-authoring and project timeline management functionality. On the other hand, a small startup company with uncertain future and limited resources could opt for an entirely cloud-based solution, which usually bears a low price tag for a small number of users and eliminates the burden of maintenance and configuration.

References

- [1] CERN, <https://home.cern/news/news/computing/migrating-open-source-technologies>
- [2] NASA, <https://training.linuxfoundation.org/solutions/corporate-solutions/success-stories/linux-foundation-training-prepares-the-international-space-station-for-linux-migration/>
- [3] GNU GPL v3.0, <https://www.gnu.org/licenses/gpl-3.0.html>
- [4] Nextcloud, <https://nextcloud.com/compare/>
- [5] Odprtokodni oblak, <https://stromar.si/wp-content/uploads/zapiski/2017/01/Odprtokodni-oblak-2017-2.pdf>