

# Quelle solution de stockage choisir dans la prochaine décennie ?

## **Sébastien Finkbeiner**

Direction du Numérique  
Université de Strasbourg  
16, rue René Descartes

## **Alain Heinrich**

Direction du Numérique  
Université de Strasbourg  
14, rue René Descartes

## **Laurence Moindrot**

Direction du Numérique  
Université de Strasbourg  
16, rue René Descartes

## **Simon Piquard**

Direction du Numérique  
Université de Strasbourg  
14, rue René Descartes

## **Xavier Pierre**

Direction du Numérique  
Université de Strasbourg  
16, rue René Descartes

## **Résumé**

*Les projets de mise en place de nouvelles solutions de stockage sont actuellement au coeur de nos métiers. On nous demande des solutions de plus en plus intégrées, offrant de plus en plus de services et avec de plus en plus d'espace, de performance, de fonctionnalités. Aujourd'hui, dans nos établissements, nous avons tous un historique de déploiement de diverses solutions qui répondent chacune à des besoins différents, qui sont très rarement compatibles et qui rendent le service obscur pour nos utilisateurs. Quel outil choisir ?*

## **Mots-clefs**

*Cloud, Stockage, Solution Collaborative, ...*

# 1 Introduction

En 2014 l'Université de Strasbourg a fait le choix de déployer la solution de stockage cloud "*Seafile*" pour l'ensemble de ses personnels et étudiants pour compléter les autres solutions de stockage déjà proposées à nos utilisateurs via l' *Active Directory*. Si cette solution donne entière satisfaction concernant le simple stockage et le partage de documents individuels, nous nous sommes aperçus qu'elle avait cependant plusieurs limites. Nous avons ainsi initié une nouvelle étude pour faire évoluer notre service de stockage et tenter de trouver une solution répondant aux nouveaux besoins. Plusieurs possibilités s'offraient à nous :

- continuer à investir dans *Seafile* ;
- changer de solution ;
- attendre les premiers résultats de la solution proposée par Renater ;
- lancer un projet communautaire avec d'autres établissements de l'ESR.

Cet article a pour but d'échanger avec la communauté sur l'aboutissement de notre étude.

## 2 Les solutions de stockage déployées actuellement dans nos établissements

On peut constater que nous retrouvons dans la plupart de nos établissements les solutions de stockage suivantes : un *Active Directory (AD)*, une solution de cloud (*NextCloud, OwnCloud, Seafile, Pydio, ...*), une solution de Ged (*Nuxeo, Alfresco*) et encore des cours en ligne (*Moodle*). Un projet de Ged pédagogique est actuellement à l'étude dans notre établissement. Nous avons également déployé un grand nombre d'applications métier qui possèdent elles-mêmes leur propre solution de stockage (local, NFS, base de données).

### 2.1 Active Directory

Le déploiement de l'*AD* a démarré en 2011 à l'Université. L'*AD* est une solution historique fortement couplée au poste de travail qui permet, au travers des profils itinérants et des dossiers partagés, de sauvegarder l'ensemble des documents de l'utilisateur de manière relativement transparente sur les postes fixes. L'intégration des postes et des utilisateurs s'est faite sur une période de 3 ans. Ce mécanisme répondait également ponctuellement aux demandes de partage de documents par la création d'espaces communs qui permettent aux personnes d'un même service de s'échanger leurs documents de travail. Mais très vite cela n'a plus suffi car certains services avaient des activités nécessitant de nombreux échanges. Des espaces ont donc été mis en place pour ces services. Cependant la création de ces espaces requiert systématiquement une intervention des équipes informatiques pour la création et la gestion des droits. Les partages ainsi mis en place sont souvent réservés aux services centraux de l'Université car l'ensemble des personnes visées font partie de l'établissement. Cette solution n'est donc absolument pas adaptée aux enseignants, aux chercheurs et à toute autre personne ayant besoin d'échanger avec l'extérieur. C'est ainsi que nous avons vu se propager l'usage de solutions de cloud externes (tels que *dropbox, google drive ...*) avec tous les problèmes engendrés. (Confidentialité des données, fuites de données à caractère personnel). Une autre pratique courante et largement employée a été l'usage du mail, qui pose d'autres problèmes. Quelle est la dernière version du document ? Qui la possède ? Qui l'a envoyée ?

## 2.2 Stockage Cloud

Pour pallier au manque de souplesse des partages via l'AD, tout en maîtrisant un lieu de stockage commun (contrairement au mail), les établissements de l'ESR ont commencé à mettre en place leur propre solution de cloud hébergée en interne. De notre côté, après un recueil des besoins de nos utilisateurs nous avons retenu la solution *Seafile*. Cette solution a été déployée pour l'ensemble des personnels de l'Université dans un premier temps, puis élargie aux étudiants afin de faciliter les échanges entre les enseignants et les étudiants. Nous avons ainsi pu répondre aux nouveaux besoins. Nos utilisateurs ont pu s'approprier la solution et sont parfaitement autonomes pour les partages, la gestion des droits. Cette solution reste en constante évolution pour s'adapter aux nouveaux besoins de nos utilisateurs. Parmi ceux-ci, nous pouvons citer la création de comptes invités pour les échanges avec l'extérieur, l'augmentation des quotas alloués, des bibliothèques de service liées à un groupe et non plus à une personne pour pallier le problème de mobilité (retraite, mutation,...).

## 2.3 GED

En parallèle à cette solution de cloud, s'est montée une solution de Ged administrative ayant pour objectif de stocker et publier tous les documents administratifs finaux de référence. On y retrouve donc les documents de l'intranet provenant des services centraux (Ressources humaines, direction des finances, ...). On retrouve également l'ensemble des documents provenant des différents conseils (CA, CEVU, ...). Là encore le public de ces documents est très différent. Pour certains documents, l'ensemble de l'établissement doit pouvoir les consulter, mais d'autres documents beaucoup plus sensibles, sont à destination d'un public très limité et parfaitement identifié.

À ce stade, il est évident que nous ne pourrions trouver un outil qui pourra parfaitement répondre à toutes les demandes, d'où cette multiplication des outils de stockage.

Parmi tous ces outils déjà mentionnés, il ne faut pas oublier nos applications métier qui pour des besoins de dématérialisation embarquent également des solutions de stockage plus ou moins propriétaires.

## 2.4 Stockage collaboratif

Depuis peu, une nouvelle solution a émergé à l'Unistra, qui est notre nouvel Environnement Numérique et Social de Travail (ENST - "*Ernest*").

Ce dernier possède également un stockage autonome. Ce choix pouvant sembler redondant, vient du coût excessif de l'intégration de notre solution *Seafile* dans cet outil. On constate donc que certains choix sont en fait pris par défaut et ne sont pas forcément issus de stratégies à moyen et longs termes.

## 3 Problématiques et évolution des besoins

Cette multiplicité d'outils, utilisés de différentes manières, par un public hétérogène, entraîne des incompréhensions et des problèmes de fonctionnement au sein de l'université. Au sein d'un même service, certains utilisateurs nomades préféreront utiliser *Seafile* quand d'autres garderont leurs habitudes sur *AD*. Ils doivent donc multiplier les outils pour retrouver l'ensemble de leurs fichiers ou pire, ils retrouveront les mêmes fichiers sur plusieurs solutions de stockage. Une meilleure intégration des outils semble donc nécessaire.

De plus, de nouveaux usages se démocratisent (ex : édition simultanée). Nos outils ne couvrent pas suffisamment ces besoins, ce qui entraîne l'apparition d'autres outils de stockage externes (ex : *office365*, *google drive*).

L'évolution des technologies, avec des débits plus rapides, incluant des fichiers plus volumineux, engendre de manière logique une utilisation plus importante des systèmes de stockage *clouds*. Pour palier

à cela, des solutions doivent être mises en place tels que des systèmes de stockage plus performant, avec, de notre côté un passage d'un système de fichier *NFS* à *CEPH*. De plus, l'augmentation du nombre d'utilisateurs vient accroître cette tendance à la hausse de l'utilisation du stockage.

À l'heure actuelle, il devient complexe pour les utilisateurs de voir des outils autonomes sans interopérabilités. La multiplication des besoins ne devrait donc pas engendrer une multiplication des outils. Ainsi nos solutions de stockage devraient s'intégrer pleinement dans les autres outils que peuvent être le *webmail* (pour envoyer des pièces jointes volumineuses), *Moodle* (afin de stocker les cours) sans même que l'utilisateur final soit contraint de s'interroger sur le stockage à utiliser.

Il devient donc urgent de rationaliser nos outils et de mieux les intégrer. Ces actions seront à la fois bénéfiques pour nos utilisateurs mais également pour nous en terme d'exploitation et de maintenance.

## 4 Évolution des infrastructures de stockage

Le premier déploiement de *Seafile* s'est fait sur la version communautaire. Cette version ne supportant pas de stockage objet, nous avons stocké les données de *Seafile* sur un partage *NFS* hébergé par du matériel *Netapp*. Cette solution n'était pas adaptée pour un déploiement à l'échelle d'une université, elle posait des problèmes évidents de mise à l'échelle.

Avec le passage à la version pro de *Seafile*, nous avons migré les données vers un cluster *Ceph*. Cette migration a nécessité la copie des 7 To de données hébergées à ce moment et a été effectuée grâce à un script fourni par *Seafile*. Elle n'a nécessité que quelques heures de coupures grâce aux mécanismes de copie incrémentale du script.

Le cluster *Ceph* utilisé pour *Seafile* est mutualisé avec notre infrastructure de virtualisation, les données utilisateurs sont hébergées sur des disques *SATA* de 8To et les métadonnées sur des disques *SSD*, le tout en triple réplication. Les données sont sauvegardées vers un autre site avec le mécanisme interne de réplication applicative de *Seafile*.

L'utilisation de *Ceph* nous permet de répondre efficacement à l'augmentation de la volumétrie, tout en garantissant un taux de disponibilité et un niveau de performance excellent.

Nous étudions la possibilité de faire évoluer cette architecture vers une utilisation de la passerelle *S3* de *Ceph* en remplacement de la *Librados*, ainsi que de l'utilisation de pools en *erasure coding*.

## 5 Les solutions de stockage Cloud actuelles

### 5.1 État de l'art des solutions

Lors de notre première étude, nous avons pris en compte les performances, les fonctionnalités, la facilité de prise en main et la complexité de l'architecture à mettre en place pour comparer les solutions du moment. Suite à cette étude, le choix s'est porté sur *Seafile*.

De ce point de vue, la solution nous convient toujours.

Si nous devons refaire une étude aujourd'hui, il ne serait pas pertinent de baser le choix de solution sur les fonctionnalités car les solutions sont maintenant très proches les unes des autres. Les indicateurs utilisés dans les premières études ne sont donc plus valides.

Ce nouveau choix serait sans aucun doute guidé par nos nouvelles problématiques que sont les chantiers d'intégration et de mutualisation des développements. C'est pour cette raison que nous nous intéressons aujourd'hui à l'interopérabilité des outils et des choix faits dans les différents établissements de l'ESR.

## 5.2 Quels sont les choix que vous avez effectués dans vos établissements ?



QR code pointant vers un petit sondage à destination des collègues des établissements de l'ESR :

1. Nom de l'établissement \* ;
2. Solution retenue \* ;
3. Quotas proposés (personnels / étudiants) ;
4. Type d'hébergement de la solution ;
5. Taille de l'établissement (nombre d'étudiants / personnels) ;
6. Email de contact pour échanges futurs.

## 5.3 Le Futur

À l'avenir, nous imaginons que l'utilisateur n'aura plus à se demander quel outil il doit utiliser. Il se connectera à son ENST et aura accès à sa messagerie, à ses espaces de travail collaboratifs. Il ne se souciera plus de savoir où sont stockés ses documents, il déposera dans son espace sans se soucier de savoir où son document est effectivement stocké. La seule question qu'il devrait avoir à se poser c'est : "À qui ai-je besoin de partager ce document ? Est-ce un document de travail ? Est-ce une version finale devant être publiée ?" En fonction des réponses à ces questions nos systèmes devront s'adapter pour faire le nécessaire de manière transparente pour l'utilisateur.

Car de son point de vue peu lui importe qu'il utilise *Nuxeo*, *Seafile*, *nextCloud*, *Moodle*, seul le résultat final compte.

## 6 Conclusion

Si par le passé, nous avons déployé de nombreuses solutions pour répondre aux besoins de nos utilisateurs, nous devons maintenant entamer un chantier de rationalisation et faciliter les interactions entre ces outils. La communauté de l'ESR a ainsi son rôle à jouer car ce chantier est conséquent et il sera difficile de le porter et de le mener seul au sein de nos établissements. On peut donc se réinterroger sur les choix passés de nos outils. *Seafile* a été un choix que nous assumons parfaitement et qui répondait à nos besoins jusque là. Mais faut-il persister dans ce choix si la majorité des établissements de l'ESR s'est orienté vers des solutions *Next/OwnCloud* ?