

My CoRe, partage de fichiers et nomadisme

David Rousse

CNRS/Direction des Systèmes d'Information
358 rue Pierre Gilles de Gennes
Bâtiment Tour GAIA - Quartier Bouysset
31 670 Labège

Résumé

Depuis fin 2012, le CNRS propose aux laboratoires une offre de services spécifiques (<http://ods.cnrs.fr/>), adaptée aux métiers de la Recherche et à leurs contraintes de sécurité : cloud sécurisé (serveurs virtuels, hébergement sécurisé de sites web, stockage brut de données), messagerie intégrée et CoRe (portail collaboratif multitutelle).

Un nouveau service, My CoRe (<http://ods.cnrs.fr/mycore.html>), basé sur le logiciel libre ownCloud (<http://www.owncloud.org/>), vient enrichir cette offre (le service My Com de messagerie instantanée, de visioconférence et de partage d'écran va également bientôt être disponible). Via un espace personnel "gratuit" de 20 Go, ce service permet aux agents travaillant dans des unités CNRS :

- d'échanger des fichiers avec d'autres personnes, partenaires privés ou personnels du monde de la recherche, de façon souple,
- de partager et de synchroniser des fichiers dans un espace accessible de n'importe quel lieu disposant d'un accès internet (nomadisme), avec tout type de plate-forme,
- d'assurer une sauvegarde "simple" et "sécurisée" de ses fichiers professionnels.

Mots clés

ownCloud, partage de fichiers, synchronisation de fichiers, nomadisme

1 Introduction

L'objet de cet article est de proposer un retour d'expérience sur le déroulement du projet de mise en œuvre du service My CoRe. En brossant le tour d'horizon des différentes étapes majeures du déroulement du projet, les points de satisfaction ainsi que les difficultés majeures rencontrées seront abordées.

2 Étude préalable

La première étape qui a précédé au lancement du projet s'est déroulée de janvier à septembre 2013 : le but était de recueillir au mieux les besoins des utilisateurs et d'appréhender, en termes fonctionnel et technique, le domaine du partage et de la synchronisation de fichiers.

2.1 Étude des besoins

La première étape a été de réaliser diverses enquêtes "terrain" afin de cerner au mieux les besoins autour des outils de partage et de synchronisation de fichiers, auxquels les offres publiques, de type Dropbox¹, répondent. Deux enquêtes auprès d'utilisateurs, dans deux délégations régionales différentes du CNRS (Côte d'Azur et Paris B), ont été menées :

- Enquête autour de l'usage d'outils de type Dropbox par les personnels.
- Enquête autour de la sauvegarde des postes de travail (volumétrie, type de documents à sauvegarder et existence du besoin).

Les usages principaux suivants ont émergé :

- Échange facile de fichiers, envoi ponctuel à des utilisateurs distants de fichiers ne passant pas par email.
- Partage d'une base de documents pour une communauté de personnes, avec les notions de facilité et de pérennité des partages.
- Synchronisation de documents locaux sur un espace accessible en ligne.
- Sauvegarde de données².

L'état de l'art des solutions de partage et de synchronisation de fichiers a en parallèle été étudié, ceci afin de bien cerner les outils disponibles sur le "marché". Quelques-uns ont fait l'objet de maquettes³ afin de mettre les besoins métier en regard des modalités techniques d'implémentation. Au final, un tour d'horizon des outils liés aux quatre besoins précités a été fait⁴.

2.2 Choix d'ownCloud

Pourquoi ownCloud a-t-il été finalement choisi ? Suite à l'étude des besoins issus du terrain et au tour d'horizon des différentes solutions disponibles à l'époque, les fonctions de partage et de synchronisation sont apparues comme les plus importantes. Tous les laboratoires remontent un besoin de ce type. L'aspect multi-plateformes et la simplicité d'utilisation d'un outil de type Dropbox sont les aspects les plus demandés.

L'outil répondant à la plupart des usages, et donc étant le meilleur compromis, s'est avéré être ownCloud. Déjà utilisé dans le monde de l'enseignement supérieur et de la recherche, facilement implémentable, bénéficiant d'un bon "bouche à oreille" dans la communauté, intégrant l'authentification via une fédération d'identité de type shibboleth, ownCloud répond aux principaux besoins : de manière native pour les trois premiers besoins d'échange, de partage et de synchronisation, de façon "basique" pour la sauvegarde (en considérant que le fait de synchroniser certains répertoires locaux des terminaux des utilisateurs permet de sauvegarder les données qui y sont stockées).

3 Cadrage fonctionnel et technique

De octobre 2013 à mai 2014, le cadrage fonctionnel et technique a été réalisé. Les principaux travaux de la période ont été de dérouler des tests de charge et d'en tirer des projections d'évolutions d'architecture selon le nombre d'utilisateurs à servir, ainsi que des projections financières. La liste détaillée des besoins fonctionnels, et la manière d'y répondre via les fonctions existantes d'ownCloud, ou via des développements complémentaires, a également été faite.

1. Offres "grand public" de partage et de synchronisation de fichiers, comme notamment Dropbox et Google Drive

2. Besoin de sauvegarde cité dans une moindre mesure par rapport aux trois autres

3. ownCloud, Horizon Data de VMWare et Duplicati.

4. Ce tableau est consultable à <https://mycore.core-cloud.net/public.php?service=files&t=37bc2f2b014dca70488e0859b7894ff2> (outils sélectionnés et qui étaient "disponibles" en 2013)

3.1 Cadrage fonctionnel

Le cadrage fonctionnel du projet ayant été formellement posé, les différents usages nécessaires ont été mis au regard des fonctions disponibles sur ownCloud⁵. Le tableau suivant les résume⁶ :

Synthèse			
	Titre	Priorité	Conformité
IN1	Accès simple au service	Haute	100
BP1	Partage de fichiers	Haute	100
BP2-BP3	Création de groupes et d'utilisateurs supplémentaires	Moyenne	25
BP4	Auto-complétion	Moyenne	100
SY1	Synchronisation multi-plateforme	Haute	100
SY2	Synchronisation agenda	Faible	0
SY3	Synchronisation contacts	Faible	0
SY4	Traçage des mises à jour d'un document	Moyenne	100
C1	Copie distante de fichier	Moyenne	100
TR0	Visualisation sur le client léger de fichiers MS Office	Moyenne	0
TR1	Protection contre la suppression/mise à jour par un non propriétaire	Moyenne	100
TR2	Edition collaborative en ligne de fichiers	Très faible	25
TR3	Gestion de la corbeille	Haute	50
TR4	Purge manuelle de la corbeille	Moyenne	100
TR5	Gestion des versions	Moyenne	50
TR6	Tableau de bord d'activité par document et pour l'utilisateur	Moyenne	80
I1	Interface CoRe vers ownCoRe	Moyenne	50
I2	Interface LABINTEL vers ownCoRe	Moyenne	25
I3	Interface agenda MS Exchange	Faible	0
SE1	Recommandation aux utilisateurs sur la sécurité	Haute	N/A
SE2	Chiffrement des documents	Haute	N/A
SE3	Blocage en cas de vol	Haute	N/A
A2	CGU	Haute	N/A
A3	Accès à l'ODS depuis divers clients légers	Haute	80
AV	Antivirus	Haute	80
TE1	Mise à disposition « gratuite » du client mobile	Faible	N/A
TE2	Page d'administration des utilisateurs organisée par DR / unité	Haute	0
TE3	Outil de filtrage d'accès au service	Haute	0
TE4	Tableau de bord d'indicateurs de suivi global	Moyenne	0
TE5	Version packagée CNRS du client lourd de synchronisation	Moyenne	N/A

Figure 1: Besoins métier initiaux

Au final, ownCloud en version "entreprise" répondait à la majeure partie des besoins métier identifiés en importance haute. Certains besoins annexes nécessaires à un déploiement généralisé de l'outil ownCloud n'étaient cependant pas pris en compte dans la version étudiée, mais l'outil représentait un bon compromis entre les besoins exprimés, la couverture fonctionnelle et l'acceptation de la solution par les utilisateurs finaux, modulo le prix de la licence "entreprise" et indépendamment des aspects techniques. En version communautaire, un effort supplémentaire apparaissait nécessaire afin d'avoir la même couverture fonctionnelle que la version "entreprise". En particulier, des développements spécifiques étaient nécessaires afin que la solution rende un service industriel. Le choix de la DSI s'est au final porté sur la version communautaire d'ownCloud, les efforts budgétaires qui auraient dû être faits pour acquérir la version "entreprise" ont été reportés vers des développements d'apps⁷, avec tout l'intérêt de reverser les développements effectués dans la communauté.

3.2 Cadrage technique

L'objet du cadrage technique a été de choisir les briques principales pour fournir le futur service, et de convenir des modalités d'hébergement et d'exploitation de ce dernier.

5. Les fonctions de synchronisation d'agendas et de contacts sont mises à 0% de conformité dans la figure 1 car leur implémentation dans la version d'ownCloud évaluée (version 5 "entreprise") n'était pas satisfaisante

6. La colonne de droite présente une complétude du besoin exprimée en pourcentage

7. Un app peut être vue comme un module complémentaire, un "plugin", ajouté à ownCloud

Les briques principales, solution de virtualisation, solution de base de données et solution de stockage, ont été choisies sur des critères de coût, de facilité d'installation et d'exploitation, et de capacité à être évolutif. La solution de virtualisation qui s'est imposée est VMWare⁸, car déjà maîtrisée par les équipes internes et par les prestataires.

La solution de base de données relationnelles a été choisie en comparant divers critères de quatre solutions : Galera/MariaDB, MySQL, Oracle et SQL Server. Les critères de choix étaient : le coût d'achat, le coût de maintenance, le nombre de nœuds supportés, la charge théorique supportée par nœud (afin de minimiser le nombre de serveurs) et la simplicité d'installation et d'exploitation. Au final, le choix s'est porté sur une solution Galera/MariaDB, à laquelle l'outil ClusterControl a été adossé. Cette solution a les avantages suivants : open source⁹, gestion jusqu'à 32 cores en termes de nœuds, connue par des équipes internes DSI donc en cohérence avec la stratégie globale de celle-ci, facile d'installation et d'exploitation grâce à l'outil ClusterControl.

Le stockage a fait l'objet d'une étude particulière, car étant donné les volumes à gérer à la cible, si le service a du succès, une solution résiliente au meilleur coût était nécessaire. Parmi les diverses solutions du marché, trois ont fait l'objet d'une attention poussée, car répondant à divers critères internes. Ces trois solutions ont au final été jugées ainsi :

Critère	Scality	Dell Compellent	EMC ² Isilon
Architecture	Solution distribuée adaptée aux stratégies cloud	Solution industrielle centralisée	Solution industrielle distribuée
Maturité	Solution récente mais disposant de bons ReX	Excellente	Excellente
Administrabilité	Exploitation complexe mais outillage suffisant	Très industrialisée	Solution la plus simple à exploiter
Coûts	Plus cher à la cible (dans le strict contexte projet)	Moins cher à la cible	Coût à la cible important
Evolutivité	Forte (multi projets et multi sites)	Faible	Bonne (mais en dessous de Scality)
Qualité de service attendue	Pas de différenciation nette	Pas de différenciation nette	Pas de différenciation nette
Conclusion	Solution la plus souhaitable mais la plus onéreuse	Solution qui répond aux besoins projet de stockage, sans permettre d'évolution	Solution facile d'utilisation mais trop liée à un matériel donné (par opposition à Scality)

Tableau 1: Comparatif des solutions de stockage

La solution de stockage distribuée Scality¹⁰, à base de matériels Dell, est la solution qui a été choisie, car la plus adaptée au projet. Elle est la plus évolutive et est sensée rendre le meilleur service.

L'hébergement est réalisé par le Centre de Calcul de l'IN2P3 : le débit réseau disponible sur le site, avec un point d'accès à un nœud réseau Renater disponible, la capacité de disposer d'une solution de sauvegarde éprouvée pour de forts volumes de données ainsi que le fait d'héberger le service dans des locaux CNRS (et donc d'avoir une véritable "étanchéité" des données) ont fait de ce choix une évidence.

Les modalités d'exploitation ont été quant à elles choisies afin de pouvoir assurer la meilleure qualité de service possible en termes de plages d'ouvertures et donc de contraintes opérationnelles d'exploitation liées. A ce titre cette fonction est assurée par un prestataire de la DSI.

8. Des solutions équivalentes, de type Proxmox par exemple, auraient pu convenir pour les besoins de virtualisation sur le projet

9. Briques de BDD open source, à l'exception de l'outil ClusterControl (voir <http://www.severalnines.com/>)

10. On peut voir Scality comme un équivalent propriétaire à Ceph, avec un accès de type "file system" stable

3.3 Tests de charge

Le principe a été de découper la réflexion en deux phases : d’une part une étude théorique du comportement d’ownCloud, et d’en extrapoler des hypothèses de charges et donc de dimensionnement d’infrastructure, puis d’autre part de réaliser des tests de charges pour confirmer ou infirmer l’étude théorique.

Le point de départ de ces deux études a été de prendre des hypothèses d’usage d’un service de type Dropbox, capitales pour dimensionner l’infrastructure technique :

- Service potentiellement ouvert à une population de 100000 utilisateurs (agents travaillant dans des unités du CNRS).
- 50% des utilisateurs vont au maximum réellement utiliser le service.
- Le quota par défaut sera de 10 Go par utilisateur.
- Environ 1000 fichiers par utilisateurs.
- Chaque fichier fait 5 Mo.
- 50 mises à jour de fichiers par jour pour un utilisateur.
- Chaque fichier est répliqué vers différentes cibles :
 - 3 terminaux par utilisateur.
 - 15% des fichiers d’un utilisateur est partagé vers 5 autres utilisateurs.

En termes d’hypothèses techniques, ces valeurs ont été retenues :

- Hypothèses de tenue de charge sur les serveurs web de type Apache (8 coeurs, 16 Go de RAM) : 530 requêtes simultanées.
- Hypothèses de tenue de charge sur MariaDB (8 coeurs, 16 Go de RAM) :
 - SELECTs : maximum 3857 par seconde.
 - INSERTs : maximum 22000 par seconde.
 - UPDATEs : maximum 3857 par seconde.

A partir des hypothèses précitées, une étude détaillée du comportement d’ownCloud a été faite afin de connaître les charges et dimensionnements potentiels associés. Cette étude théorique a ensuite été mise en regard de réels tests de charges, réalisés sur une infrastructure matérielle composée de deux serveurs hébergeant deux hyperviseurs et d’une baie de stockage adossée à ces derniers¹¹. Les briques logicielles en place étaient les suivantes sur cette infrastructure :

- 2 Virtual Machines (VM) Piranha en actif/passif.
- 2 VM Apache/ownCloud¹² en actif /actif.
- 1 VM MariaDB.
- 1 VM exposant en NFS le stockage présent sur une baie de stockage.
- 1 VM Shinken pour le monitoring.
- 5 VM pour les injecteurs Tsung.

Plusieurs campagnes de tirs ont été faites pour affiner progressivement les chiffres : campagne initiale sur une installation brute, campagne après diverses optimisations système, campagne avec reverse-proxies et antivirus. Ces campagnes ont été automatisées via l’emploi de l’outil Tsung¹³. Au final, le dimensionnement de l’infrastructure, issu de l’étude théorique, a été confirmé par les tests de charges¹⁴. Cela est résumé ci-après.

11. Infrastructure (sur du matériel Dell) pour les tests de charge My CoRe disponible sur https://github.com/CNRS-DSI-Dev/mycore_press/blob/master/architecture/schema_tests_charges_2014.png

12. ownCloud en version 6 communautaire

13. Le détail de cette étude n’a pas été rendu publique du fait d’un manque de temps, voir le principe global sur https://github.com/CNRS-DSI-Dev/mycore_press/blob/master/architecture/schema_tests_charges_tsung_2014.png

14. Le détail de ces tests est disponible à https://github.com/CNRS-DSI-Dev/mycore_press/blob/master/CNRS-LINAGORA-ownCloud-load-estimate.ods

— dimensionnement de la base de données :

SQL servers		Number of SQL nodes (~ VM) for the estimated SQL load (8 cores/16GB RAM per node)						
		Number of users (N)						
% of active users		1	1000	5000	30000	50000	70000	100000
5,00%		1	1	1	3	5	7	10
10,00%		1	1	1	6	9	12	18
15,00%		1	1	2	8	13	18	25
20,00%		1	1	2	11	18	24	35
30,00%		1	1	3	16	27	38	54
50,00%		1	1	3	18	30	41	59

Figure 2: Dimensionnement base de données pour ownCloud

— dimensionnement des nœuds web :

Web servers		Number of web nodes for the estimated load (8 cores/16GB RAM per node)						
		Number of users (N)						
% of active users		1	1000	5000	30000	50000	70000	100000
5,00%		1 serv	1 serv	1 serv	3 serv	5 serv	7 serv	10 serv
10,00%		1 serv	1 serv	1 serv	6 serv	10 serv	14 serv	19 serv
15,00%		1 serv	1 serv	2 serv	9 serv	15 serv	20 serv	29 serv
20,00%		1 serv	1 serv	2 serv	12 serv	19 serv	27 serv	38 serv
30,00%		1 serv	1 serv	3 serv	17 serv	29 serv	40 serv	57 serv
50,00%		1 serv	1 serv	5 serv	29 serv	48 serv	67 serv	95 serv

Figure 3: Dimensionnement web pour ownCloud

— dimensionnement du réseau :

Total DL		Network bandwidth simulation for download [Sync own+Sync share] (global architecture)						
		Number of users (N)						
% of active users		1	1000	5000	30000	50000	70000	100000
5,00%		0 mb/s	13 mb/s	64 mb/s	382 mb/s	637 mb/s	891 mb/s	1 273 mb/s
10,00%		0 mb/s	25 mb/s	127 mb/s	764 mb/s	1 273 mb/s	1 782 mb/s	2 546 mb/s
15,00%		0 mb/s	38 mb/s	191 mb/s	1 146 mb/s	1 910 mb/s	2 674 mb/s	3 819 mb/s
20,00%		0 mb/s	51 mb/s	255 mb/s	1 528 mb/s	2 546 mb/s	3 565 mb/s	5 093 mb/s
30,00%		0 mb/s	76 mb/s	382 mb/s	2 292 mb/s	3 819 mb/s	5 347 mb/s	7 639 mb/s
50,00%		0 mb/s	127 mb/s	637 mb/s	3 819 mb/s	6 366 mb/s	8 912 mb/s	12 731 mb/s
Total UL		Network bandwidth simulation for upload [Sync own+Sync share] (global architecture)						
		Number of users (N)						
% of active users		1	1000	5000	30000	50000	70000	100000
5,00%		0 mb/s	5 mb/s	23 mb/s	139 mb/s	231 mb/s	324 mb/s	463 mb/s
10,00%		0 mb/s	9 mb/s	46 mb/s	278 mb/s	463 mb/s	648 mb/s	926 mb/s
15,00%		0 mb/s	14 mb/s	69 mb/s	417 mb/s	694 mb/s	972 mb/s	1 389 mb/s
20,00%		0 mb/s	19 mb/s	93 mb/s	556 mb/s	926 mb/s	1 296 mb/s	1 852 mb/s
30,00%		0 mb/s	28 mb/s	139 mb/s	833 mb/s	1 389 mb/s	1 944 mb/s	2 778 mb/s
50,00%		0 mb/s	46 mb/s	231 mb/s	1 389 mb/s	2 315 mb/s	3 241 mb/s	4 630 mb/s

Figure 4: Dimensionnement réseau pour ownCloud

L'analyse de la solution ownCloud menée d'un point de vue théorique a montré la faisabilité technique du projet. Toutefois, le dimensionnement du nombre de serveurs nécessaires indique que le partage de fichiers entre utilisateurs et synchronisés sur différents terminaux est très consommateur en ressources, nécessitant l'acquisition de nombreux serveurs. Le critère de coût est donc un élément primordial dans le choix de mise en œuvre d'un tel projet. Par ailleurs, le grand nombre de serveurs nécessaires implique qu'une optimisation d'architecture ou de configuration peut amener à des gains très importants en terme de nombre de serveurs et donc de budget.

A noter qu'une mini-étude à un choix alternatif à une mise en œuvre en "interne" d'ownCloud, par rapport à une solution équivalente en mode SaaS ¹⁵, a montré des coûts à l'utilisateur sensiblement inférieurs pour la solution gérée en interne CNRS.

4 Conception et mise en œuvre

La mise en œuvre a été faite ensuite sur la période de juin à décembre 2014. Cette phase projet a consisté à réaliser les spécifications détaillées, à installer et à paramétrer la solution, tout en rédigeant les diverses procédures techniques, et enfin à recetter et à préparer la conduite du changement. Le plus intéressant à noter sur cette phase projet a été les divers développements réalisés autour du cœur applicatif ownCloud. L'architecture telle que déployée au final pour rendre le service est également présentée.

4.1 Développements spécifiques

Afin d'adapter le comportement d'ownCloud aux besoins métier, divers développements ont été réalisés. La liste à ce jour est la suivante :

- App de tableau de bord = <https://github.com/CNRS-DSI-Dev/dashboard>
- App pour faciliter la gestion des groupes pour les administrateurs = <https://github.com/CNRS-DSI-Dev/lotsofgroups>
- App de gestion personnelle de groupes = https://github.com/CNRS-DSI-Dev/group_custom
- App pour renforcer la dureté des mots de passe locaux = https://github.com/CNRS-DSI-Dev/password_policy
- App de gestion des CGU = <https://github.com/CNRS-DSI-Dev/gtu>
- App de filtrage applicatif = <https://github.com/CNRS-DSI-Dev/gatekeeper>
- App d'intégration à shibboleth pour l'authentification et le provisionning des comptes via Janus ¹⁶ = https://github.com/CNRS-DSI-Dev/user_servervars2
- App de gestion des partages = <https://github.com/LydSC/ShareWatcher>
- App d'actions lors de la création/suppression d'utilisateurs = https://github.com/CNRS-DSI-Dev/user_account_actions
- App de migration de fichiers entre deux comptes d'un même utilisateur = https://github.com/CNRS-DSI-Dev/user_files_migrate
- App et scripts pour la sauvegarde et la restauration de fichiers = https://github.com/CNRS-DSI-Dev/mycore_backup_restore_user_files et https://github.com/CNRS-DSI-Dev/user_files_restore
- Script d'intégration à des listes SYMPA = https://github.com/CNRS-DSI-Dev/mycore_sympa
- Thème spécifique = <https://github.com/CNRS-DSI-Dev/mycore>
- Script de build = https://github.com/CNRS-DSI-Dev/mycore_build

Ces développements concernent la partie serveur ownCloud : aucun développement à ce jour n'a été fait sur les divers clients ownCloud.

A noter que le code source des développements précités a été rendu public via Github. Ceci est principalement dû au fait que l'équipe en place lors des développements était habituée aux workflows qui y sont disponibles. En tout rigueur, il serait préférable à terme de migrer ce code sur un service plus proche du milieu de l'Enseignement Supérieur et de la Recherche (ESR), comme SourceSup ¹⁷.

15. Solution SaaS comparée, Yaziba Drive (<http://www.yazibadrive.net/>)

16. Janus est la solution de gestion des identités du CNRS, voir <https://janus.dsi.cnrs.fr/Documentation/>. Dans My CoRe, une personne disposant d'un compte dans l'annuaire du CNRS (<https://www.dsi.cnrs.fr/labintel/>) peut donc s'authentifier dans Janus et accéder à My CoRe

17. SourceSup est la plateforme d'hébergement de projets informatiques de l'ESR, disponible à <https://services.renater.fr/sourcesup/index>

4.2 Architecture

L'architecture technique en place à ce jour est synthétisée dans ce schéma :

*Production architecture for 2015
Located on the CNRS IN2P3 Computing Center*

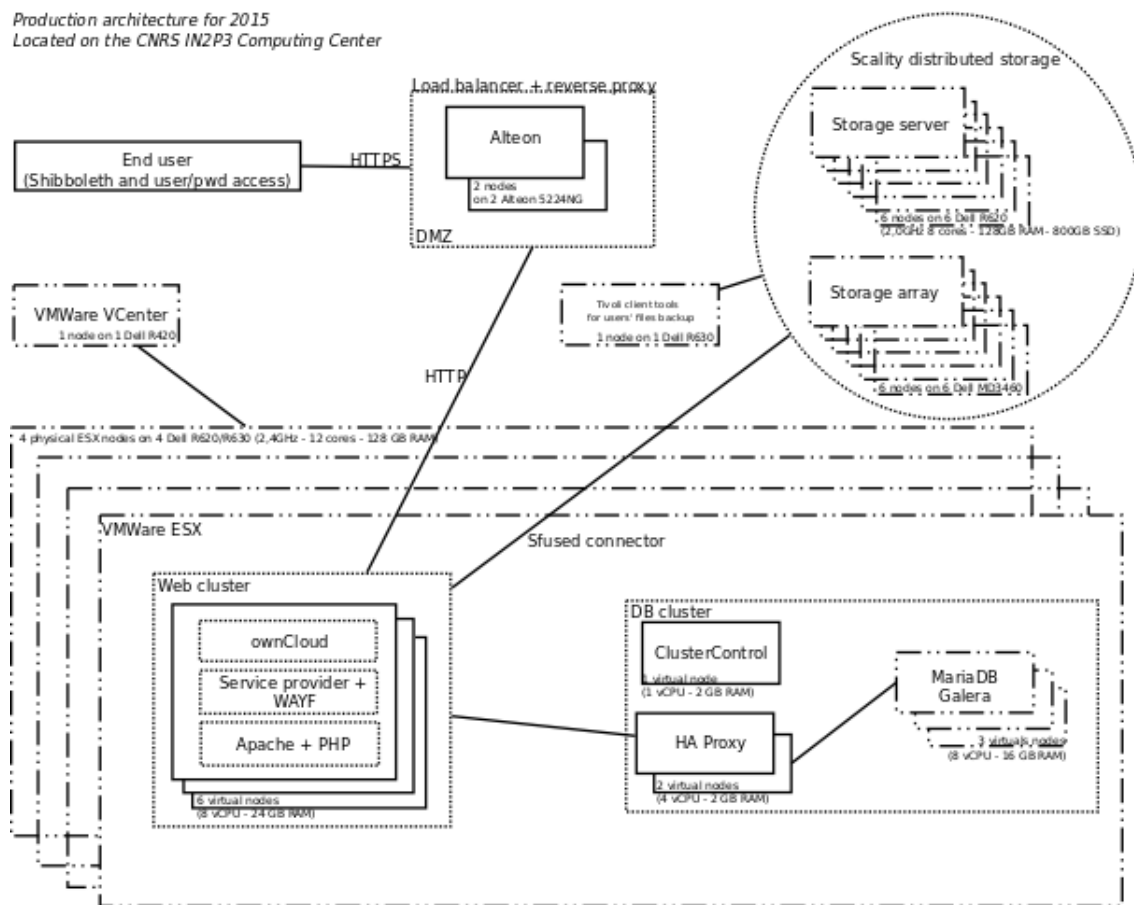


Figure 5: Architecture My CoRe pour 2015

5 Déploiement pilote

Le service a été ouvert en pilote à partir de début 2015, de janvier à septembre, sur une partie des unités de trois délégations régionales du CNRS (Alsace, Côte d'Azur et Paris B).

La phase pilote a eu comme intérêt, au niveau fonctionnel, de faire valider par les utilisateurs que la solution choisie, et que les modules complémentaires développés, répondent à leurs besoins. Elle a aussi permis de mettre à jour les développements complémentaires à prévoir¹⁸, en particulier deux besoins majeurs : la mise à disposition de comptes "invités" (afin de faciliter les collaborations avec des personnes n'ayant pas la possibilité de demander un compte My CoRe) et la sauvegarde/restauration des fichiers utilisateurs.

Par ailleurs, au niveau technique, la comparaison des chiffres théoriques sur la charge avec les mesures réelles n'a pas été satisfaisante. En effet, la plate-forme n'a pas été aussi sollicitée qu'espéré¹⁹. Cela peut s'expliquer par le fait que les utilisateurs restent réticents à utiliser un service étiqueté en mode "pilote". Il faudra réitérer la comparaison ci-dessus après quelques mois d'exploitation en production.

18. Voir https://github.com/CNRS-DSI-Dev/mycore_press/blob/master/besoins/besoins_mycore_apres_pilote.png

19. Voir à https://github.com/CNRS-DSI-Dev/mycore_press/raw/master/myCore_comparison_estimate-load_real-load_on_ownCloud.pdf la comparaison des métriques système entre théorie et phase pilote

6 État actuel

Depuis octobre 2015, le service est ouvert à tout le CNRS. Il est prévu d'ajouter dans les mois à venir diverses fonctions à la version actuelle :

- Migration en ownCloud 8 ²⁰.
- App de gestion des partages (<https://github.com/LydSC/sharewatcher>), app d'aide à la valorisation du mot de passe local My CoRe lors du login initial et app de "posts" sur la page de login.
- Comptes "invités" afin de donner des accès temporaires à des utilisateurs hors CNRS.
- Gestion des versions de fichiers activée pour tout le monde.
- Page d'administration des utilisateurs optimisée.
- Changement d'identifiant dû à Réséda ²¹.
- Chiffrement des données côté serveur.
- Logiciels clients PC et mobiles personnalisés (en particulier permettant un accès unique via Janus).

Au niveau technique, selon comment se comporte la plate-forme, il y a deux scénarii possibles à ce jour. Le premier est un scénario basique ²² d'ajout de nœuds web et/ou base de données, selon les besoins. Le deuxième scénario est un passage de la base de données sur des nœuds physiques ²³. On pourrait aussi faire une évolution plus profonde en changeant le mode d'accès aux données, en passant d'un mode d'accès de type "file system" à un mode objet. Pousser la fonction "Server to server sharing" d'ownCloud, ceci afin de permettre à des instances ownCloud de communiquer entre elles, est aussi envisageable.

7 Conclusion

Quels retours d'expérience tirer du déroulement du projet My CoRe ? Tout d'abord les points positifs principaux sont :

- L'utilité du projet : besoin identifié, utilité avérée. Les diverses études de terrain ont permis de conforter l'équipe projet sur la voie à suivre.
- Le bon dialogue MOA-MOE : réunion de suivi régulière, échange de visu, communication ouverte entre les deux équipes ont permis d'avoir un bon mode de fonctionnement.
- Les échanges avec d'autres organismes en France ²⁴ et dans les pays voisins ²⁵, ont été bénéfiques, à la fois pour la maturation du besoin fonctionnel mais aussi pour comparer/valider les orientations techniques.
- L'espace projet collaboratif sous CoRe ²⁶ : il permet de centraliser la réflexion et donne une vision à jour à tous les intervenants projet.
- La qualité de la réflexion autour du stockage, de la haute disponibilité du SGBD et du dimensionnement de la solution ont permis d'aboutir à un cadrage technique avant mise en œuvre bien structuré.
- Le bon déroulement des développements, fait par une équipe de personnes CNRS et hors CNRS.
- La bonne collaboration avec le centre de calcul de l'IN2P3 sur les solutions de stockage distribué, grâce à leur expérience sur le sujet, ainsi que pour la sauvegarde d'importants volumes de données.

20. ownCloud en version 7 communautaire est actuellement utilisé sur le service

21. Plus d'informations sur Réséda à <http://intranet.cnrs.fr/intranet/actus/150205-reseda.html>

22. Schéma d'évolution d'architecture My CoRe pour 2016 à https://github.com/CNRS-DSI-Dev/mycore_press/raw/master/architecture/schema_2016.pdf

23. Schéma d'option d'évolution d'architecture My CoRe pour 2016 à https://github.com/CNRS-DSI-Dev/mycore_press/raw/master/architecture/schema_2016_option.pdf

24. Voir <https://groupes.renater.fr/sympa/info/owncloud-fr>

25. Divers échanges ont eu lieu avec le CERN et l'Université de Vienne en particulier

26. Voir <http://ods.cnrs.fr/core.html>

Par opposition, avec du recul, certains points auraient pu être mieux faits, ou faits autrement :

- Le difficulté de concilier le besoin d'un outil de partage et de synchronisation avec un besoin de sauvegarde pur : cette dualité a généré diverses discussions et arbitrages projet.
- La durée très longue du projet, qui peut s'expliquer en particulier par diverses difficultés rencontrées lors de la mise en œuvre (notamment la complexité de prise en main de certaines briques, la difficulté inhérente à la mise en place à distance d'une infrastructure par des exploitants non situés physiquement près du site d'hébergement et le travail préalable nécessaire pour identifier précisément les coûts induits par un tel projet).
- Le choix de faire les tests de charges sur une nouvelle architecture dédiée n'était pas adéquat : plus de temps a été passé à l'installer qu'à configurer et tester ownCloud au final.
- Le choix d'une méthode projet "classique", à cycle long, n'était peut être pas adaptée au projet. Une mise en place de cycles courts, avec tout de suite notamment des pilotes (dès le début du projet), aurait donné des usages métier et facilité le dimensionnement.
- Le manque d'outils de tests de régression pour valider chaque montée de version ou modification est problématique. L'investissement initial est important mais il fiabilise le service sur le long terme.
- Une équipe MOA trop réduite, car affectée pour une toute petite partie de son temps sur le projet, a tendance à ne pas pouvoir tenir une charge de travail trop longtemps sur le projet, notamment pour les tests fonctionnels de chaque version.
- Un outil de chat aurait été utile, car certains échanges informels, utiles à l'animation et au dialogue de l'équipe, n'ont pas de valeur ajoutée à se faire par email.
- La mise en place d'une infrastructure dédiée à My CoRe sur les couches techniques basses (stockage, réseau, BDD, virtualisation), sans mutualisation ou réutilisation avec d'autres services, aboutie à des coûts d'entrée importants. "Monter" des couches IaaS et PaaS communes à divers services métier, sur lesquelles construire des solutions de type SaaS, aurait pu être un préalable au projet, et aurait peut être été plus conforme à l'état de l'art.

8 Remerciements

Les informations données dans ce document sont issues du travail de diverses personnes ayant participé au déploiement de My CoRe, l'équipe projet d'une part : Jonathan Bouchiquet (support), Marc Dexet (développeur), Philippe Dubrulle (activité transverse), Gilian Gambini (responsable technique), Eric Gervasoni (co-responsable MOA), Jérôme Jacques (responsable technique), Nadine Marouze (responsable du service dédié à l'offre de services de la DSI du CNRS), Paulo Mora de Freitas (co-responsable MOA), Olivier Lenormand (activité transverse), Jean-Yves Lopez (activité technique transverse), Patrick Paysant (développeur), Lyderic Saint-Criq (développeur) et Alexandre Salvat (gestion de projet). L'équipe du département infrastructure de la DSI du CNRS ainsi que son graphiste, les équipes du centre de calcul de l'IN2P3 et les équipes des prestataires de la DSI impliquées sur le projet sont d'autre part à remercier.