

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
Ministère de L'Enseignement Supérieur et de la Recherche Scientifique
Université Saad Dahlab blida 1
Faculté Des Sciences
Département d'Informatique



MÉMOIRE

Présentée pour l'obtention du **diplôme** de **MASTER**

En : Informatique

Spécialité : Systemes Informatiques et Réseaux

Par : GUELMOUNA Zine Eddine et HAIF Abdelhamid

Sujet

**Etude et mise en place d'une plate-forme Entreprise
d'hébergement et de partage de fichiers**

Soutenue, le 25/06/2023, devant le jury composé de :

M.	NOM Prénom	Professeur	à l'USDB	Président
Mme.	NOM Prénom	Maître de Conférences A	à l'USDB	Examineur
Mme.	BEY Fella	Maître de Conférences B	à USDB	Promotrice
M.	NEFFAH Mohamed	Directeur du Centre d'Innovation	à SONATRACH DG	Encadreur

Dédicaces

À ma famille bien-aimée et aux membres du club CSCC,

Je souhaite vous dédier ce mémoire avec tout mon amour et ma gratitude infinis. Votre soutien indéfectible, votre encouragement constant et votre compréhension ont été essentiels dans la réalisation de ce travail.

À ma famille, vous êtes ma force et ma motivation. Votre amour inconditionnel et votre soutien indéfectible ont été les piliers qui m'ont permis de mener à bien ce travail.

Je souhaite également exprimer ma reconnaissance envers tous les membres du club CSCC , en particuliers REMMIDE Mohamed Abdelkarim , BOUZID Manel , DJEMOUAI Sarah , MAAZOUZ Ilhem , BELAISSI Wissam et HAOUCHET Yanis pour leur soutien, leur collaboration et leur motivation tout au long de ce projet. Votre engagement et votre esprit d'équipe ont contribué à rendre cette expérience encore plus enrichissante.

Enfin, je tiens à remercier toutes les personnes qui ont contribué de près ou de loin à ce travail, qu'elles soient mentionnées ou non. Votre contribution a apporté une valeur ajoutée à ce mémoire et a enrichi mon parcours académique.

Cette dédicace vous est dédiée, car vous êtes mes partenaires, mes guides et mes sources d'inspiration. Votre présence et votre soutien ont été inestimables tout au long de ce voyage.

Avec tout mon amour et ma reconnaissance,

GUELMOUNA Zine Eddine et HAIF Abdelhamid

Liste des mots clé

NAS :Network Attached Storage

NFS :Network File System

LDAP : Lightweit Directory Access Protocol

Remerciement

Nous exprimons notre profonde gratitude envers Allah, le Tout-Puissant, pour Sa guidance, Sa protection et Ses innombrables bienfaits. C'est grâce à Sa grâce et à Sa bienveillance que nous avons pu accomplir ce projet. Que notre reconnaissance envers Allah soit constante et que nous Le remercions pour toutes les bénédictions qu'Il nous accorde dans notre vie. Qu'Il accepte nos efforts et nous guide sur le droit chemin.

Je tiens à exprimer mes sincères remerciements à tous ceux qui ont contribué à la réalisation de ce mémoire. Tout d'abord, je souhaite exprimer ma profonde gratitude à ma promotrice, Madame BEY FELLA, pour sa guidance précieuse, son soutien constant et ses conseils avisés tout au long de ce projet. Votre expertise et votre dévouement ont été d'une valeur inestimable, et j'ai eu la chance de bénéficier de votre encadrement attentif.

Je tiens également à remercier chaleureusement mon encadreur, Monsieur NEFFAH Mohamed, pour son implication, ses orientations pertinentes et son accompagnement tout au long de cette recherche. Votre expertise et votre disponibilité ont été essentielles pour mener à bien ce travail.

Je suis également reconnaissant envers les membres du jury pour leur temps, leur attention et leurs précieux commentaires lors de l'évaluation de ce mémoire.

Un grand merci à ma famille et à mes amis pour leur soutien inconditionnel, leur encouragement constant et leur compréhension tout au long de ce parcours. Votre présence et vos encouragements ont été une source de motivation et de réconfort pour moi.

Enfin, je tiens à exprimer ma reconnaissance envers toutes les personnes qui ont contribué de près ou de loin à ce travail, qu'ils soient cités ou non. Votre contribution a enrichi ce mémoire et a contribué à sa qualité.

Mes remerciements vont à chacun d'entre vous, car sans votre soutien et votre contribution, ce mémoire n'aurait pas pu voir le jour. Votre apport a été inestimable et je suis reconnaissant de vous avoir à mes côtés tout au long de ce voyage académique.

Merci du fond du cœur.

Cordialement, GUELMOUNA Zine Eddine et HAIF Abdelhamid

ملخص

تركز هذه المذكرة على تطوير منصة مشاركة ملفات قابلة للتطوير والتكرار لمواجهة تحديات إدارة ومشاركة الملفات في العصر الرقمي. تهدف المنصة إلى توفير وظائف موثوقة وفعالة لمشاركة الملفات، وتلبية المتطلبات المتزايدة لسوناتراخ يستكشف البحث أفضل الممارسات والمعايير والبروتوكولات لتنفيذ المنصة، مع التركيز على التكرار وقابلية التوسع لزيادة الموثوقية والأداء. ويأخذ التنفيذ في الاعتبار هيكل الشبكة، وتقنيات تكرار البيانات وموازنة الأحمال. يهدف المشروع إلى توفير حل كامل يضمن التعاون السلس ويلبي الاحتياجات المتطورة للمنظمات في عالم يعتمد على البيانات.

Abstract

This dissertation focuses on the development of a scalable and redundant file sharing platform to meet the challenges of managing and sharing files in the digital age. The platform aims to provide reliable and efficient file sharing functionality, meeting the growing requirements of organization Sonatrach. The research explores best practices, standards and protocols for implementing the platform, focusing on redundancy and scalability for increased reliability and performance. The implementation takes into account network architecture, data replication and load balancing techniques. The project aims to provide a complete solution that guarantees smooth collaboration and meets the evolving needs of organizations in a data-driven world.

Résumé

Ce mémoire porte sur le développement d'une plateforme de partage de fichiers scalable et redondante pour répondre aux défis de la gestion et du partage de fichiers à l'ère numérique. La plateforme vise à fournir une fonctionnalité de partage de fichiers fiable et efficace, répondant aux exigences croissantes d'organisation Sonatrach. La recherche explore les meilleures pratiques, les normes et les protocoles pour la mise en œuvre de la plateforme, en se concentrant sur la redondance et la scalabilité pour une fiabilité et des performances accrues. La mise en œuvre tient compte de l'architecture du réseau, de la réplication des données et des techniques d'équilibrage de la charge. Le projet vise à fournir une solution complète qui garantisse une collaboration harmonieuse et réponde aux besoins évolutifs des organisations dans un monde axé sur les données.

Table des matières

Table des figures	i
Liste des tableaux	iii
Introduction Générale	1
1 Etat de l'art sur les plateformes de partage de fichiers	3
1.1 Introduction	3
1.2 Le partage de fichiers	3
1.3 L'importance de plateformes de partage de fichiers scalable et redondante	4
1.4 Classer les plateformes de partage de fichiers en fonction de leur conception, de leur architecture et de leur mise en œuvre	7
1.5 Examiner les solutions existantes en matière de performance et de scalabilité et de redondance pour les plates-formes de partage de fichiers	8
1.6 Taxonomie des plateformes de partage de fichiers existantes	10
1.7 Les problèmes dans le partage de fichiers	14
1.7.1 Analyse des problèmes de scalabilité	14
1.7.2 Analyse des problèmes de performances	15
1.7.3 Analyse des problèmes de redondance	16
1.8 Les avantages et des inconvénients de chaque type de plateforme de partage de fichiers	17
1.8.1 Plateformes de partage de fichiers basées sur le cloud	17
1.8.2 Plateformes de partage de fichiers peer-to-peer (P2P)	18
1.8.3 File transfer protocols	18
1.9 Conclusion	19
2 Conception	20
2.1 Introduction	20
2.2 Analyse des besoins	20

2.2.1	Accessibilité et disponibilité	20
2.2.2	Sauvegarde et sécurité du stockage	21
2.2.3	Partage facile et collaboration	21
2.2.4	Adaptabilité et flexibilité	21
2.2.5	Mises à jour automatiques	21
2.3	Architecture global	21
2.4	Load balancing	24
2.4.1	Equilibrage de charge statique	24
2.4.2	Equilibrage de charge dynamique	25
2.5	Network File System	27
2.6	Clustering	28
2.7	Cache	28
2.8	LDAP	29
2.9	Conclusion	30
3	Mise en œuvre de la solution	31
3.1	Introduction	31
3.2	Environnement de travail	31
3.2.1	Partie matérielle	31
3.2.2	Partie Logiciels	32
3.3	déploiement	36
3.3.1	Première étape	36
3.3.2	Deuxieme étape	37
3.3.3	Troisieme étape	41
3.3.4	Phase2 (Configuration de TrueNAS)	43
3.3.5	phase3 (configurer galera cluster et haproxy de la base de données) :	45
3.4	Conclusion	48
	Conclusion Générale	49
	Bibliographie	50

Table des figures

1.1	logo de Hidrive share	10
1.2	logo de google drive	11
1.3	logo de file.io	11
1.4	logo de swiss transfer	12
1.5	logo de WeTransfer	12
1.6	logo de Filemail	12
1.7	logo de TransferNow	13
1.8	logo de Dropbox transfer	13
1.9	logo de Nextcloud	14
2.1	Architecture global de notre solution	23
2.2	Fonctionnement de l'Algorithme Roundrobin	25
3.1	logo de Proxmox	32
3.2	logo de HaProxy	33
3.3	logo de Php	33
3.4	logo de Apache	33
3.5	logo de MariaDB	34
3.6	logo de Galera Cluster	34
3.7	logo de TrueNAS	35
3.8	logo de Redis	35
3.9	créer un nouveau user	39
3.10	Interface d'installation	39
3.11	Installation finale	40
3.12	interface Proxmox après la création des machines	41
3.13	Fichier de configuration haproxy	42
3.14	Ajouter address ip de haproxy dans config.php sur web1 et web2	43
3.15	Créer un pool de stockage	44
3.16	Autoriser le service NFS	44

3.17 Autoriser les address ip a stocker dans le pool	45
3.18 Fichier de configuration galera1	46
3.19 Fichier de configuration galera2	46
3.20 Vérification de cluster	47
3.21 configuration haproxydb	47

Liste des tableaux

3.1	Caractéristique du Serveur principal	31
3.2	Caractéristique des VMs	32
3.3	Liste des technologies utilisées avec sa version	36

Introduction Générale

Le partage de fichiers est une pratique qui englobe le partage ou l'accès à des informations et des ressources numériques, comprenant des documents, du multimédia (audio/vidéo), des graphiques, des programmes informatiques, des images et des livres électroniques. Il s'agit d'une distribution de données ou de ressources, qu'elle soit privée ou publique, au sein d'un réseau, avec divers niveaux de privilèges de partage. Cette pratique permet d'échanger des fichiers de manière efficace, facilitant la collaboration, la diffusion d'informations et l'accès à des contenus pertinents. Que ce soit pour le partage de documents professionnels, le téléchargement de médias ou la diffusion de connaissances, le partage de fichiers joue un rôle central dans la société moderne et les échanges numériques.

Dans un monde de plus en plus numérique, la gestion et le partage efficaces des informations sont essentiels au bon fonctionnement de presque toutes les organisations, qu'il s'agisse d'une entreprise, d'un établissement d'enseignement ou d'un gouvernement. Cependant, avec l'augmentation rapide de la quantité d'informations numériques, la gestion de ces informations est devenue un défi majeur.

C'est dans ce contexte que ce mémoire porte sur la recherche et la mise en place d'une plateforme de partage de fichiers redondante et extensible. Les systèmes de partage de fichiers sont au cœur de nombreuses applications, de la collaboration en ligne au stockage de données en passant par le partage de médias. Cependant, à mesure que les besoins de stockage de données et de partage de fichiers augmentent, ces systèmes doivent être à la fois fiables et capables d'évoluer à mesure que la demande augmente.

En informatique, la redondance fait référence à la duplication des composants ou des fonctions dans un système pour augmenter sa fiabilité, généralement sous forme de sauvegarde ou de sécurité. L'évolutivité, quant à elle, est la capacité d'un système, d'un réseau ou d'un processus à gérer une charge croissante en ajoutant des ressources, généralement en termes de puissance de traitement ou de stockage de données. Dans ce mémoire, nous allons explorer différentes astuces et techniques pour obtenir la redondance et l'évolutivité

dans les systèmes de partage de fichiers.

Chapitre 1

Etat de l'art sur les plateformes de partage de fichiers

1.1 Introduction

Les plateformes de partage de fichiers sont des services essentiels qui offrent aux utilisateurs la possibilité de partager des fichiers numériques avec d'autres personnes. Ce chapitre se concentre sur la présentation des plateformes de partage de fichiers, en mettant en avant leurs avantages et inconvénients respectifs. De plus, nous allons les classer en fonction de leur conception, de leur architecture et de leur mise en œuvre, afin de mieux comprendre les différences et les implications pratiques de chaque type de plateforme. Cette analyse approfondie nous permettra de fournir un aperçu détaillé des différentes plateformes de partage de fichiers, en évaluant leurs caractéristiques clés et en mettant en évidence les aspects importants à prendre en compte lors du choix d'une plateforme adaptée à des besoins spécifiques.

1.2 Le partage de fichiers

Les plateformes de partage de fichiers sont des services en ligne qui permettent aux utilisateurs de stocker, partager et transférer des fichiers numériques avec d'autres. Ces fichiers peuvent inclure des documents, des images, des vidéos, des fichiers audio, etc. Les plateformes de partage de fichiers fournissent généralement une capacité de stockage en ligne où les utilisateurs peuvent télécharger des fichiers, puis générer des liens de partage auxquels d'autres peuvent accéder. [1]

Les plateformes de partage de fichiers offrent généralement les fonctionnalités suivantes :

Stockage en ligne : Ils fournissent un espace de stockage virtuel où les utilisateurs peuvent télécharger des fichiers. Cela libère de l'espace sur l'appareil local et permet d'accéder aux fichiers à partir de n'importe quel appareil en réseau. [2]

Partage de fichiers : les utilisateurs peuvent générer des liens de partage pour des fichiers spécifiques ou des dossiers entiers. Ces liens peuvent être envoyés aux destinataires, qui peuvent ensuite télécharger ou visualiser le fichier partagé. [2]

Collaboration : certaines plateformes de partage de fichiers offrent des fonctionnalités de collaboration en temps réel. Plusieurs utilisateurs peuvent travailler sur le même fichier en même temps, apporter des modifications et voir les mises à jour en temps réel.

Contrôle d'accès : les plateformes de partage de fichiers permettent aux utilisateurs de définir des autorisations d'accès pour les fichiers partagés. Cela inclut des options telles que la possibilité d'afficher, de modifier ou de partager uniquement des fichiers. [2]

Sécurité : les plateformes de partage de fichiers ont mis en place des mesures de sécurité pour protéger les fichiers contre les accès non autorisés ou les atteintes à la vie privée. Cela peut inclure des options de cryptage des données, des contrôles d'authentification et des mesures de protection contre les virus et les logiciels malveillants. [3]

Sauvegarde : à l'aide d'une plateforme de partage de fichiers, les utilisateurs peuvent créer des copies de sauvegarde de fichiers importants en les téléchargeant sur le service. Cela permet d'éviter la perte de données en cas de problème avec le périphérique local. [3]

Les plateformes de partage de fichiers offrent des solutions pratiques et efficaces pour partager des fichiers numériques, collaborer avec d'autres et accéder à des fichiers à distance.

1.3 L'importance de plateformes de partage de fichiers scalable et redondante

Les plateformes de partage jouent un rôle vital dans notre société moderne. Ils permettent aux individus et aux organisations de partager des informations, des ressources

et des idées de manière efficace et rapide. Voici quelques raisons qui soulignent la nécessité de disposer de telles plateformes :

Collaboration : les plateformes partagées facilitent la collaboration entre les individus et les groupes, quelle que soit leur distance géographique. Ils fournissent des outils pour partager des documents, des fichiers, des calendriers et d'autres ressources, facilitant la collaboration et la coordination de projets. [4]

Gestion de la croissance : les entreprises et les organisations génèrent et échangent constamment de plus en plus de données et de fichiers. Une plate-forme de partage de fichiers scalable peut s'adapter à cette augmentation de la capacité de stockage, du traitement et du transfert de données. Il gère efficacement les augmentations de volume de fichiers partagés sans compromettre les performances. [2]

Accès simultané : Dans un environnement de travail collaboratif, il est courant que plusieurs utilisateurs aient besoin d'accéder aux mêmes fichiers en même temps. La plateforme scalable de partage de fichiers permet à plusieurs utilisateurs d'accéder et de modifier des fichiers simultanément, facilitant ainsi la collaboration en temps réel. [4]

Tolérance aux pannes : les plates-formes de partage de fichiers redondantes sont conçues pour résister aux pannes matérielles ou aux interruptions de service. En déployant des systèmes de secours et redondants, les fichiers partagés restent accessibles même en cas de panne du serveur ou de l'infrastructure. Cela garantit une disponibilité continue des fichiers et réduit les temps d'arrêt. [5]

Sauvegarde et récupération des données : les plateformes de partage de fichiers redondantes incluent fréquemment des fonctions de sauvegarde et de récupération automatiques des données. Cela protège les dossiers des pertes accidentelles, des erreurs humaines ou des dommages matériels. Il est possible de restaurer les fichiers des sauvegardes si nécessaire. [6]

Scalabilité : En fonction des besoins de leur entreprise, les entreprises peuvent surveiller les fluctuations du trafic et de la demande. Une plateforme de partage de fichiers scalable peut gérer ces changements en augmentant ou en réduisant automatiquement ses ressources en fonction des besoins. Même lors des pics d'utilisation, cela garantit une expérience utilisateur cohérente.

Performance et vitesse : Les plateformes de partage de fichiers scalables utilisent fréquemment des technologies de mise en cache, de distribution de contenu et d'optimisation du réseau pour offrir une performance optimale et une vitesse de transfert élevée des fichiers. Cela permet aux utilisateurs d'accéder rapidement et de partager efficacement des fichiers, indépendamment de leur taille. [7]

Analyse : Certaines plateformes de partage de fichiers offrent des fonctionnalités d'analyse avancées, telles que des rapports d'utilisation, des statistiques de téléchargement et de partage, ainsi que des informations sur les modèles d'utilisation des fichiers. Cela peut aider les entreprises à mieux comprendre comment leurs fichiers sont utilisés, à améliorer leurs processus et à faire des choix intelligents. [7]

Conformité aux réglementations : certaines industries et zones ont des normes strictes pour la conformité et la protection des données. Les plateformes de partage de fichiers scalables et redondantes peuvent être configurées pour respecter le RGPD (Règlement général sur la protection des données), garantissant ainsi la conformité aux lois. [8]

Suivi et historique des modifications : les plateformes de partage de fichiers peuvent enregistrer les modifications des fichiers et fournir un historique des versions. Cela permet de surveiller les changements, de récupérer des versions antérieures si nécessaire et de conserver une trace de l'évolution des fichiers au fil du temps. [8]

Mobilité : Les utilisateurs peuvent accéder à leurs fichiers depuis n'importe quel appareil connecté à Internet grâce à des plateformes de partage de fichiers scalables et redondantes. Cela améliore la productivité car les utilisateurs peuvent accéder, modifier et partager des fichiers où qu'ils soient, ce qui facilite la mobilité et le travail à distance.

Sécurité des données : Les plateformes de partage de fichiers redondantes peuvent offrir des fonctionnalités avancées de sécurité des données telles que le chiffrement des fichiers, la gestion des droits d'accès et les contrôles d'authentification. Cela empêche les accès non autorisés et les violations de la confidentialité des fichiers sensibles. [8]

En somme, disposer de plateformes de partage de fichiers scalable et redondante offre des avantages significatifs en termes de sécurité, de mobilité, de collaboration, de conformité réglementaire, de suivi des modifications et d'analyse des données. Ces fonctionnalités contribuent à une gestion efficace des fichiers et à une meilleure productivité pour les utilisateurs et les organisations.

1.4 Classer les plateformes de partage de fichiers en fonction de leur conception, de leur architecture et de leur mise en œuvre

Les plates-formes de partage de fichiers peuvent être classées en fonction de leur conception, de leur architecture et de leur implémentation comme suit :

Plates-formes de stockage basé sur le cloud : ces plates-formes sont basées sur une architecture de stockage basé sur le cloud dans laquelle les fichiers sont téléchargés sur des serveurs distants et accessibles via Internet. Ils utilisent généralement un protocole de transfert de fichiers tel que HTTP ou FTP. Les exemples courants incluent Dropbox, Google Drive et Microsoft OneDrive. [9]

Réseaux peer-to-peer (P2P) : les plates-formes P2P permettent aux utilisateurs de partager des fichiers directement entre eux sans passer par un serveur central. Les fichiers sont éparpillés et distribués sur les ordinateurs des utilisateurs du réseau. Des exemples populaires de plates-formes P2P sont BitTorrent et eMule. [10]

Serveur FTP (File Transfer Protocol) : les serveurs FTP sont conçus pour les transferts de fichiers. Ils utilisent le protocole FTP pour permettre aux utilisateurs de télécharger et de télécharger des fichiers via des clients FTP. Les plates-formes de partage de fichiers basées sur FTP sont souvent utilisées pour les transferts de fichiers volumineux ou dans les environnements professionnels. [11]

Les plates-formes de partage de fichiers peuvent être classées selon différents critères tels que leur nature, leur accessibilité, leur modèle de déploiement et d'autres fonctionnalités. Concernant la nature de la plate-forme, elle est divisée en plate-forme commerciale et plate-forme grand public. Les plates-formes commerciales sont proposées par les entreprises et offrent des fonctionnalités avancées, un support client et une sécurité renforcée, tandis que les plates-formes grand public sont destinées aux utilisateurs individuels ou aux petits groupes et offrent des options gratuites ou abordables. [9]

En termes d'accessibilité et de méthodes de partage, nous trouvons des plates-formes Web qui permettent l'accès via un navigateur Web sans installer d'applications, ainsi que des plates-formes avec des applications de bureau qui fournissent des applications pour synchroniser et gérer des fichiers directement depuis votre ordinateur. [9]

Sur la base du modèle de déploiement, les plateformes de partage de fichiers peuvent être classées comme des plateformes de cloud public, hébergées sur une infrastructure de cloud public par des fournisseurs tiers, ou des plateformes de cloud privé, déployées sur des serveurs privés ou une infrastructure de cloud privé gérée par l'organisation elle-même. [9]

Enfin, en ce qui concerne les fonctionnalités supplémentaires, certaines plates-formes de partage de fichiers se concentrent sur la collaboration, offrant des fonctionnalités permettant aux utilisateurs de travailler sur le même fichier en même temps, de publier des commentaires, de suivre les modifications, etc. D'autres plates-formes se concentrent davantage sur le stockage et le partage de fichiers en toute sécurité et moins sur les fonctionnalités de collaboration. [9]

Il est important de noter que certaines plates-formes peuvent chevaucher ou combiner plusieurs de ces conceptions et architectures, en fonction de leur fonctionnalité et de leur objectif spécifiques. Le choix de la plate-forme de partage de fichiers dépendra des besoins individuels et des préférences en matière de stockage, de sécurité, de disponibilité et de scalabilité .

1.5 Examiner les solutions existantes en matière de performance et de scalabilité et de redondance pour les plates-formes de partage de fichiers

Les performances, la scalabilité et la redondance sont des considérations essentielles pour une plate-forme de partage de fichiers, en particulier lorsqu'il s'agit de gérer efficacement un grand nombre d'utilisateurs et de fichiers. Voici quelques solutions existantes dans ces domaines :

Infrastructure de serveur puissante : une plate-forme de partage de fichiers doit être hébergée sur une infrastructure de serveur solide et fiable. Cela peut impliquer l'utilisation de serveurs puissants avec une grande capacité de stockage, une bande passante élevée et une architecture en cluster pour répartir la charge. [12]

Équilibrage de charge : les plates-formes de partage de fichiers peuvent utiliser des techniques d'équilibrage de charge pour répartir uniformément la charge de travail entre plusieurs serveurs. Cela permet de gérer efficacement un grand nombre d'utilisateurs et de réduire les goulots d'étranglement. [12]

Mise en cache des données : les plates-formes de partage de fichiers peuvent utiliser des mécanismes de mise en cache pour mettre en cache les fichiers les plus demandés. Cela accélère l'accès aux fichiers fréquemment utilisés et réduit la charge sur les serveurs de stockage. [12]

CDN (Content Delivery Network) : Un CDN est un réseau géographiquement distribué de serveurs qui fournissent des fichiers aux utilisateurs à partir de serveurs proches de leur emplacement géographique. Cela réduit la latence et améliore les performances en garantissant une livraison rapide des fichiers. [12]

Réplication et sauvegarde des données : pour assurer la redondance et la résilience, les plateformes de partage de fichiers peuvent répliquer les données sur plusieurs serveurs. Ainsi, si un serveur tombe en panne, les fichiers sont toujours accessibles depuis d'autres serveurs. Des sauvegardes régulières peuvent également être effectuées pour éviter la perte de données. [13]

Scalabilité horizontale : Afin de faire face à l'augmentation du nombre d'utilisateurs et de fichiers, la plateforme de partage de fichiers doit être conçue avec une scalabilité horizontale à l'esprit. Cela signifie qu'ils doivent pouvoir ajouter de nouveaux serveurs au fur et à mesure de leur croissance afin de répartir la charge et de maintenir de bonnes performances. [14]

Surveillance et optimisation des performances : les plate-formes de partage de fichiers doivent mettre en œuvre des outils de surveillance pour surveiller les performances du système et identifier les problèmes potentiels. Des optimisations peuvent être apportées pour améliorer le temps de réponse, l'utilisation des ressources et la gestion du trafic. [10]

Les fournisseurs de plates-formes de partage de fichiers doivent mettre en œuvre ces solutions et adopter une approche proactive pour garantir des performances optimales, une scalabilité suffisante et une redondance fiable pour répondre aux demandes croissantes des utilisateurs et garantir une expérience utilisateur fluide. [10]

En mettant en œuvre ces solutions, une plate-forme de partage de fichiers peut fournir des performances élevées, une scalabilité flexible et une redondance fiable. Cela permet aux utilisateurs de partager et de télécharger des fichiers rapidement et efficacement tout en garantissant la disponibilité et l'intégrité des données.

1.6 Taxonomie des plateformes de partage de fichiers existantes

Il existe plusieurs types de plateformes de partage de fichiers disponibles. Les plateformes de partage de fichiers basées sur le cloud sont un type qui offre une sécurité de partage de fichiers dans le cloud . Les plateformes de partage de fichiers peer-to-peer (P2P) sont un autre type qui permet aux utilisateurs d'un réseau de partage de fichiers P2P de partager des ressources et de rechercher des données avec d'autres réseaux de partage de fichiers P2P [15]. AnonChain est une plateforme de partage de fichiers qui intègre le partage de fichiers Inter Planetary File Sharing (IPFS) avec la technologie blockchain pour fournir un partage de fichiers sécurisé et anonyme [16]. Des applications de partage de fichiers peer-to-peer open-source, telles que Ares, ont également été utilisées pour le partage de fichiers [17]. Les sites de réseaux sociaux peuvent également inclure des outils de partage de médias, des outils de partage de fichiers et des plateformes de réseautage [18]. De plus, il existe des services de partage de fichiers conçus à des fins spécifiques, tels que BioTorrents, qui est un service de partage de fichiers pour les données scientifiques [19]. Des plateformes de partage de fichiers autonomes sont également disponibles, où l'hétérogénéité des appareils et des services fournissant le partage de fichiers est masquée pour l'utilisateur final [20].

Nous vous présentons ci-dessous certaines des plateformes de partage de fichiers gratuites les plus adaptées au partage de fichiers qui ne peuvent pas être envoyés par e-mail en tant que pièce jointes.

HiDrive Share : Avec HiDrive Share, vous pouvez importer jusqu'à 2 Go de fichiers gratuitement, en toute sécurité et sans inscription. Vous envoyez le lien de téléchargement créé automatiquement par mail ou message privé aux utilisateurs de votre choix. Ces derniers peuvent télécharger les fichiers aussi souvent qu'ils le souhaitent et vous pouvez également envoyer le lien de téléchargement à un nombre illimité de personnes. Une fois la période de validité de sept jours écoulée, le lien expire et toutes les données importées sont supprimées en toute sécurité et de manière irrévocable. HiDrive Share offre un degré élevé de sécurité Cloud grâce à ses serveurs allemands sécurisés et conformes au RGPD.



Figure 1.1: logo de Hidrive share

Google Drive : Google Drive est un service de stockage en nuage offert par Google. Il permet aux utilisateurs de stocker et de partager des fichiers, y compris des documents, des photos et des vidéos. Google Drive est intégré à d'autres services Google, tels que Google Docs et Google Sheets, permettant la collaboration en temps réel sur des documents. [21]



Figure 1.2: logo de google drive

File.io : Avec File.io, vous pouvez importer des fichiers dont le volume s'élève à jusqu'à 2 Go sans déboursier un centime ni devoir vous inscrire. Vous partagez un lien de téléchargement avec les utilisateurs de votre choix par mail ou message privé. Ils peuvent ensuite télécharger les fichiers une fois avant qu'ils ne s'autodétruisent. Pour 25 par mois, les utilisateurs peuvent partager des fichiers d'un volume allant jusqu'à 10 Go avec téléchargements illimités des fichiers et autodestruction optionnelle. File.io fait également office de stockage Cloud avec un volume de stockage de 2 To. Si vous avez besoin d'un volume de stockage encore plus élevé, vous pouvez passer au forfait premium et partager des fichiers d'une taille supérieure à 100 Go. [22]



Figure 1.3: logo de file.io

Swiss Transfer : Swiss Transfer est un service fourni par l'entreprise d'hébergement Web suisse Infomaniak. Les utilisateurs peuvent partager jusqu'à 50 Go de fichiers gratuitement. Swiss Transfer adhère aux normes de protection des données suisses. Les fichiers peuvent être partagés à l'aide d'un lien de téléchargement valide pendant maximum 30 jours ou via e-mail. Les téléchargements sont protégés à l'aide d'un mot de passe.

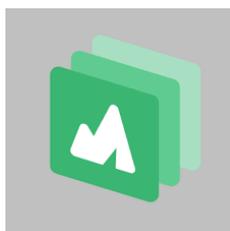


Figure 1.4: logo de swiss transfer

WeTransfer : WeTransfer compte parmi les services de partage de fichiers ou de Cloud les plus connus et peut être utilisé sans installation ni inscription. Cet outil en ligne permet de partager jusqu'à 2 Go de fichiers gratuitement. En optant pour un compte Pro, le volume d'import grimpe à 1 To. Le processus est simple : vous n'avez qu'à entrer votre adresse e mail et celle de votre destinataire. Une fois l'import effectué, ce dernier reçoit un lien de téléchargement. La protection par mot de passe et la personnalisation de la durée de validité d'un lien ne sont disponibles qu'avec la souscription Pro.



Figure 1.5: logo de WeTransfer

Filemail : Filemail est né en Norvège et offre l'option appréciable d'envoyer des fichiers volumineux par e-mail gratuitement. Il s'agit d'un service basé sur le Cloud qui peut être utilisé sur un navigateur ou via une application mobile. Avec la version gratuite, le destinataire reçoit un lien de téléchargement qui reste accessible durant 1 ou 7 jours, selon la décision de l'expéditeur. Si vous vous inscrivez, vous obtiendrez l'accès à du stockage gratuit ainsi qu'à un volume de téléchargement de 5 Go et profiterez de scans antivirus et du chiffrement des données. En version Pro, la durée de disponibilité des téléchargements peut être prolongée.



Figure 1.6: logo de Filemail

TransferNow : TransferNow est une plateforme de partage de fichiers pratique et sans inscription permettant d'envoyer des fichiers dont la taille peut aller jusqu'à 5 Go. Pour un volume d'upload plus élevé, vous pouvez étendre celle-ci jusqu'à 100 Go en souscrivant au service pour 8,33 par mois. La version premium offre 1 To de stockage, ce qui transforme TransferNow en stockage Cloud. Chaque transfert de fichier bénéficie d'une protection HTTPS.



Figure 1.7: logo de TransferNow

Dropbox Transfer : Dropbox Transfer bénéficie du fait que de nombreuses personnes sont déjà habituées à utiliser Dropbox. Ce service d'hébergement de fichiers compte parmi les leaders du marché du stockage Cloud, au coude-à-coude avec Google Drive et OneDrive. Mais les alternatives à Dropbox constituent également une concurrence de taille. Avec DropBox Transfer, vous pouvez envoyer jusqu'à 100 MB gratuitement à l'aide d'un lien de téléchargement. Pour ce faire, inscrivez-vous gratuitement ou connectez-vous à votre compte Dropbox existant.



Figure 1.8: logo de Dropbox transfer

Nextcloud : nextcloud est une solution cloud open source qui peut être considérée comme une alternative aux services commerciaux tels que DropBox, Google Cloud ou aux services cloud de Microsoft (par exemple OneDrive) et d'Apple (iCloud). Le système de cloud a été mis sur le marché en 2016 par Frank Karlitschek, qui avait auparavant quitté le projet précédent Owncloud avec un certain nombre d'autres collaborateurs. Les développeurs de Nextcloud attachent une importance particulière à ce que le logiciel soit conçu pour une utilisation la plus large possible et ce gratuitement. Owncloud, le concurrent de Nextcloud, en revanche, est actuellement principalement spécialisé dans les solutions commerciales payantes.

Les plateformes de partage de fichiers diffèrent en ce qui suit : D'autres se concentrent sur la facilité d'utilisation, tandis que certaines offrent des fonctionnalités avancées telles que la synchronisation automatique et la gestion des versions. La capacité de stockage varie de quelques gigaoctets à plusieurs téraoctets, et les mesures de sécurité peuvent inclure le chiffrement des données et l'authentification à deux facteurs. La collaboration peut prendre la forme de discussions intégrées, de modifications en temps réel et de commentaires sur les fichiers. Certaines plateformes offrent des plans gratuits avec des limitations, tandis que d'autres proposent des abonnements payants avec plus de fonctionnalités et de stockage.



Figure 1.9: logo de Nextcloud

1.7 Les problèmes dans le partage de fichiers

Les plateformes de partage de fichiers soulèvent différents problèmes, notamment en ce qui concerne la scalabilité, la redondance et les performances. Dans les sections suivantes, nous examinerons chacun de ces problèmes en détail :

1.7.1 Analyse des problèmes de scalabilité

Les problèmes de scalabilité avec les plates-formes de partage de fichiers peuvent se manifester de différentes manières. Voici quelques problèmes de scalabilité courant :

Performances lents : à mesure que le nombre d'utilisateurs et de demandes de transfert de fichiers augmente, la plate-forme peut commencer à connaître des ralentissements et des temps de réponse plus lents. Cela conduit à une mauvaise expérience utilisateur et à une faible satisfaction des utilisateurs. [23]

Goulots d'étranglement : certains composants de l'architecture de la plate-forme peuvent devenir des points de congestion lorsqu'ils sont fortement chargés. Par exemple, une base de données centrale pour l'indexation des fichiers peut devenir un goulot d'étranglement si elle ne peut pas gérer efficacement les demandes concurrentes. [24]

Incapacité à gérer l'augmentation du volume de fichiers : si la capacité de stockage de la plate-forme n'est pas suffisamment scalable, elle peut atteindre sa limite lorsque le volume de fichiers augmente. Cela peut entraîner des problèmes tels que le rejet de nouvelles soumissions de fichiers ou un ralentissement des performances lors de l'accès aux fichiers existants. [25]

Défis liés à la gestion des métadonnées : les plates-formes de partage de fichiers doivent souvent gérer de grandes quantités de métadonnées liées aux fichiers, telles que les informations d'indexation, les autorisations d'accès et les métadonnées utilisateur. La gestion efficace de ces métadonnées peut devenir compliquée à mesure que la plate-forme se développe, ce qui peut entraîner des problèmes de scalabilité . [24]

Coûts élevés : la mise à l'échelle d'une plate-forme de partage de fichiers pour répondre à la demande croissante peut entraîner des coûts énormes en termes d'infrastructure, de stockage, de bande passante et de ressources requises. La gestion de ces coûts tout en maintenant une scalabilité suffisante peut constituer un défi pour les fournisseurs de services. [2]

1.7.2 Analyse des problèmes de performances

Les problèmes de performances de la plate-forme de partage de fichiers peuvent entraîner des retards, des ralentissements et une mauvaise expérience utilisateur. Voici quelques problèmes de performances courants :

Latence élevée : la latence fait référence au délai entre une action de l'utilisateur (téléchargement, partage, accès au fichier) et la réception du résultat. Une latence élevée peut ralentir le processus et frustrer les utilisateurs. [26]

Vitesses de transfert lentes : les utilisateurs s'attendent à des vitesses de transfert rapides lors du téléchargement ou du partage de fichiers. Des vitesses de transfert lentes peuvent augmenter considérablement les temps de transfert, en particulier pour les fichiers volumineux. [27]

Temps de chargement de la page : les plates-formes de partage de fichiers ont généralement une interface Web où les utilisateurs peuvent interagir avec les fichiers. Si une page prend trop de temps à se charger, cela peut affecter l'expérience utilisateur et entraîner de la frustration. [27]

Problèmes de disponibilité : de mauvaises performances peuvent également entraîner une disponibilité réduite des fichiers. Par exemple, si les utilisateurs rencontrent des erreurs de connexion fréquentes ou des temps d'arrêt, cela peut affecter leur capacité à accéder ou à partager des fichiers. [2]

Traitement inefficace des requêtes concurrentes : lorsque de nombreux utilisateurs accèdent à la plate-forme en même temps pour télécharger ou partager des fichiers, des problèmes de performances peuvent résulter d'un traitement inapproprié des requêtes concurrentes. [27]

1.7.3 Analyse des problèmes de redondance

Les problèmes de redondance dans les plates-formes de partage de fichiers concernent principalement la disponibilité et la sécurité des fichiers stockés. Voici quelques problèmes de redondance courants :

Perte de données : si la plate-forme de partage de fichiers ne dispose pas de mécanismes de redondance et de sauvegarde adéquats, il existe un risque de perte de données en raison d'une panne matérielle, d'une corruption de fichiers ou d'autres problèmes techniques. La perte de données peut être catastrophique pour les utilisateurs et entraîner la perte irréversible de fichiers importants. [28]

Interruption prolongée : en cas de panne du système ou de maintenance planifiée, les utilisateurs peuvent perdre l'accès à leurs fichiers. Les temps d'arrêt prolongés peuvent entraîner des interruptions d'activité, une perte de productivité et des utilisateurs mécontents. [28]

Faible disponibilité : si la plate-forme de partage de fichiers ne dispose pas d'un mécanisme de redondance approprié, les fichiers peuvent ne pas être disponibles dans certaines circonstances, telles qu'une panne matérielle, un bogue logiciel ou des problèmes de réseau. Cela peut être frustrant pour les utilisateurs qui dépendent d'une disponibilité constante des fichiers. [29]

Risque de sécurité : la redondance des fichiers est également importante pour assurer la sécurité des données. La redondance permet de restaurer les fichiers à partir de copies de sauvegarde si le serveur est perdu ou endommagé. Sans redondance, les fichiers pourraient être vulnérables aux attaques, à la perte de confidentialité ou à des modifications non autorisées. [29]

Temps de récupération allongé : En cas de panne ou de perte de données, le temps nécessaire à la restauration des fichiers peut être important si la plateforme ne dispose pas de mécanismes de redondance efficaces. Des sauvegardes régulières et une stratégie de récupération appropriée sont essentielles pour minimiser le temps de récupération des fichiers et réduire les perturbations pour les utilisateurs. [29]

1.8 Les avantages et des inconvénients de chaque type de plateforme de partage de fichiers

Chacune des plateformes de partage de fichiers existantes a ses propres avantages et inconvénients. Dans cette section, nous discuterons des avantages et des inconvénients de chaque type de plateforme de partage de fichiers afin que nous puissions les comparer entre eux.

1.8.1 Plateformes de partage de fichiers basées sur le cloud

Les plateformes de partage de fichiers basées sur le cloud ont les avantages suivants :

Accessibilité : les fichiers sont facilement accessibles depuis n'importe quel appareil connecté à Internet. [30]

Partage facile : partagez facilement des fichiers avec d'autres utilisateurs en leur envoyant un lien de téléchargement. [30]

Synchronisation automatique : les modifications apportées aux fichiers sur un appareil sont automatiquement synchronisées avec tous les autres appareils connectés. [30]

Sauvegarde de fichiers : les fichiers sont sauvegardés sur un serveur distant, offrant une certaine sécurité en cas de perte ou de dommage de l'appareil. [30]

Cependant, il convient de noter qu'ils présentent les inconvénients suivants :

Dépendance à Internet : pour accéder aux fichiers, une connexion Internet est requise. [30]

Limites de stockage : les plateformes de stockage cloud gratuites ont généralement des limites de stockage, et un espace de stockage supplémentaire peut nécessiter un abonnement payant. [30]

Confidentialité et sécurité : les utilisateurs doivent faire confiance aux fournisseurs de services de stockage basé sur le cloud pour protéger leurs données, ce qui soulève des problèmes de confidentialité. [30]

1.8.2 Plateformes de partage de fichiers peer-to-peer (P2P)

Les plateformes de partage de fichiers peer-to-peer (P2P) ont les avantages suivants :

Vitesse de téléchargement élevée : les fichiers sont téléchargés simultanément à partir de plusieurs sources, ce qui permet des vitesses de téléchargement rapides. [10]

Résilience du réseau : en utilisant un réseau décentralisé, les fichiers sont disponibles tant que des utilisateurs partagent le même fichier. [10]

Partage illimité : les utilisateurs peuvent partager des fichiers sans être limités par des quotas de stockage ou des politiques de partage. [10]

Cependant, il convient de noter qu'ils présentent les inconvénients suivants :

Contenu illégal : les réseaux P2P peuvent faciliter le partage de contenu piraté

Risque de sécurité : les fichiers téléchargés via les réseaux P2P peuvent contenir des logiciels malveillants ou des virus. [31]

Manque de contrôle : les utilisateurs comptent sur la disponibilité des autres utilisateurs pour télécharger des fichiers, ce qui peut entraîner des téléchargements lents ou des fichiers indisponibles. [31]

1.8.3 File transfer protocols

File transfer protocols ont les avantages suivants :

Contrôle total : les utilisateurs ont un contrôle total sur leur serveur FTP, ce qui leur permet de gérer les autorisations d'accès et la sécurité des fichiers. [11]

Transferts rapides : Le protocole FTP est conçu pour des transferts de fichiers rapides et efficaces. [11]

Cependant, il convient de noter qu'ils présentent les inconvénients suivants :

Exigences techniques : la configuration d'un serveur FTP nécessite des connaissances techniques et peut être compliquée pour les utilisateurs non qualifiés. [32]

Restrictions d'accès à distance : L'accès aux fichiers sur les serveurs FTP peut être restreint lorsqu'il n'est pas possible de se connecter à distance au serveur. [32]

Coûts d'infrastructure : la configuration et la maintenance d'un serveur FTP peuvent entraîner des coûts d'infrastructure et de bande passante supplémentaires. [32]

1.9 Conclusion

Dans ce chapitre, nous avons examiné l'importance des plateformes de partage de fichiers ainsi que les problèmes liés à l'évolutivité et à la redondance. De plus, nous avons réalisé une comparaison approfondie des solutions existantes. En analysant ces aspects clés, nous avons pu mettre en évidence les défis auxquels sont confrontées les plateformes de partage de fichiers et identifier les stratégies et les approches utilisées pour relever ces défis. Cette comparaison nous permet de mieux comprendre les avantages et les inconvénients de chaque solution et d'orienter notre choix vers la plateforme de partage de fichiers la plus adaptée à nos besoins spécifiques.

Chapitre 2

Conception

2.1 Introduction

Ce chapitre se concentre sur la conception d'une plateforme de partage de fichiers redondante et scalable, en réponse aux défis identifiés dans le premier chapitre. Nous abordons ici la sélection des technologies appropriées, la définition de l'architecture du système, la planification de la capacité et la mise en œuvre des mesures de sécurité.

Notre objectif est de concevoir une plateforme capable de répondre aux exigences actuelles et de s'adapter aux besoins futurs. Ce chapitre détaillera un plan de conception qui pourra servir de modèle pour le développement de plates-formes de partage de fichiers similaires à l'avenir.

2.2 Analyse des besoins

Plusieurs facteurs sont à l'origine de l'évolution des technologies de partage de fichiers, notamment le besoin croissant de partage de fichiers, les progrès techniques, l'augmentation de la mobilité et de l'accès à distance. Ces éléments sont essentiels au progrès continu des technologies de partage de fichiers. Parmi les différentes méthodes de partage de fichiers modernes, le service cloud se démarque en tant que méthode la plus couramment utilisée. Les services cloud offrent de nombreux avantages par rapport aux autres méthodes de partage de fichiers. Voici quelques-uns de ces avantages :

2.2.1 Accessibilité et disponibilité

Les fichiers stockés dans le cloud sont accessibles de n'importe où, à n'importe quel moment, à condition de disposer d'une connexion internet. Cela signifie une plus grande

flexibilité pour travailler à distance, collaborer avec d'autres utilisateurs et accéder aux fichiers à partir de différents appareils. [4]

2.2.2 Sauvegarde et sécurité du stockage

Les services cloud offrent de vastes possibilités de stockage en toute sécurité. Les fichiers sont stockés sur des serveurs éloignés, ce qui protège les utilisateurs contre la perte de données en cas de défaillance matérielle ou de cyberattaque. En outre, les fournisseurs de services cloud ont mis en place des mécanismes de sécurité étendus pour empêcher tout accès indésirable. [4]

2.2.3 Partage facile et collaboration

Les utilisateurs peuvent échanger facilement des données avec d'autres personnes. Cela encourage le travail d'équipe et la collaboration, que ce soit pour des tâches professionnelles ou personnelles. [4]

2.2.4 Adaptabilité et flexibilité

Les services en nuage étant très adaptables, facile à augmenter ou à réduire l'espace de stockage en fonction des besoins de l'utilisateur. Il est possible également de choisir parmi une large gamme de plans et de services proposés par les fournisseurs de services en nuage, ce qui permet de trouver la solution qui correspond le mieux aux besoins individuels. [4].

2.2.5 Mises à jour automatiques

Ces services sont régulièrement mis à jour avec de nouvelles fonctionnalités et améliorations. Ces mises à jour sont généralement effectuées de manière transparente, sans nécessiter d'action de la part des utilisateurs. Ceux-ci peuvent également bénéficier des dernières fonctionnalités et des améliorations en cours sans effort supplémentaire. [4]

2.3 Architecture global

Notre architecture Nextcloud est construite en tenant compte de la haute disponibilité, de la redondance et de la flexibilité. Elle est composée des éléments suivants :

Load balancer : Il est essentiel pour gérer les demandes entrantes et les distribuer à d'autres instances Nextcloud. L'équilibrage de charge peut transférer automatiquement le trafic vers un autre serveur en cas de défaillance d'un serveur.

Nextcloud : C'est notre application principale, déployée sur plusieurs serveurs pour une distribution équilibrée de la charge et une redondance.

La base de données : MariaDB est utilisé pour stocker les données des utilisateurs de Nextcloud. Galera Cluster prend en charge la synchronisation multi-maître de MariaDB, ce qui signifie que nous pouvons écrire sur n'importe quel nœud du cluster et que les données seront immédiatement disponibles sur tous les autres nœuds. Cela garantit la cohérence et la disponibilité de nos données.

Stockage NFS : Tous nos fichiers Nextcloud sont stockés sur notre système de stockage en réseau (NAS). Nous montons le système de fichiers NAS sur nos serveurs Nextcloud en utilisant NFS (Network File System), ce qui permet à Nextcloud de voir les données comme si elles étaient locales.

Cache : Un système de cache, comme Redis ou Memcached, est utilisé pour améliorer les performances de Nextcloud en stockant temporairement les données fréquemment utilisées en mémoire.

LDAP : Nous utilisons un service LDAP pour la gestion centralisée des utilisateurs et des groupes. Nextcloud est configuré pour utiliser ce service LDAP pour l'authentification et l'autorisation, ce qui permet une gestion cohérente et centralisée des accès utilisateurs.

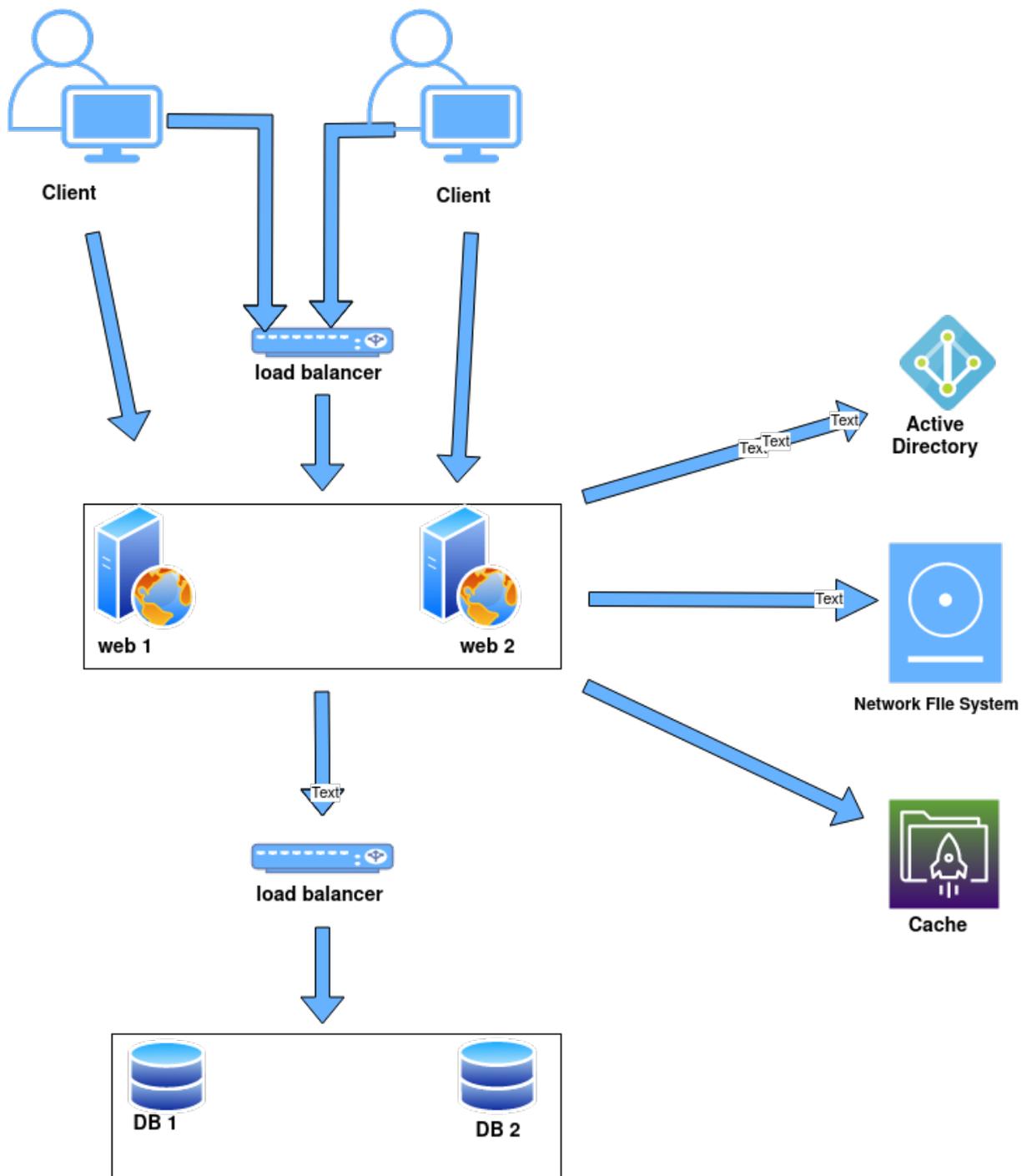


Figure 2.1: Architecture global de notre solution

Dans cette figure, les requêtes des utilisateurs sont gérées par l'équilibrage de charge, qui les redirige vers l'une de nos instances Nextcloud. Ces instances stockent leurs fichiers sur le NAS via NFS et les données utilisateur dans MariaDB, tout en utilisant un système de cache pour améliorer les performances. De plus, Nextcloud est configuré pour authentifier les utilisateurs via le service LDAP, garantissant une gestion centralisée et cohérente des

accès utilisateurs.

2.4 Load balancing

L'équilibrage de charge est une technique utilisée pour répartir la charge totale du réseau entre plusieurs ressources ou appareils. Cette approche améliore l'efficacité de l'utilisation des ressources et réduit le temps de réponse pour chaque tâche. De plus, cela évite les situations où certains centres sont sous-chargés et d'autres sont surchargés. La fiabilité et l'accessibilité des données peuvent être améliorées grâce à la redondance en utilisant différents composants pour l'équilibrage de charge. Le système mesure la charge en termes de charge réseau, d'utilisation de la mémoire et de charge CPU . [33]

2.4.1 Equilibrage de charge statique

L'équilibrage de charge statique est réalisé grâce à des informations préalables fournies dans le système. Les performances du processus sont gérées au moment de l'exécution. Les informations de charge sont calculées sur un nœud et soumises à un nœud distant pour exécution. Sur la base des calculs de charge de travail distribuée, la charge est distribuée et l'exécution démarre quelle que soit la charge précédente. La nature non préemptive est obtenue grâce à l'équilibrage de charge statique. La charge du système ne dépend pas de l'état actuel, mais elle nécessite des ressources et des poids système. Dans l'équilibrage statique de la charge de travail, l'algorithme de répartition de la charge est pré-exécuté. Plusieurs algorithmes d'équilibrage de charge statique sont disponibles pour une bonne répartition de la charge.. [33]

Algorithme de roundrobin : L'une des approches les plus simples est l'équilibrage de charge "round robin", qui répartit les informations des clients entre un certain nombre de serveurs. Cette stratégie est utilisée dans un petit réseau avec un nombre restreint de serveurs. Dans cet algorithme, nous avons de nombreux clients et serveurs. Lorsqu'un client lance une requête et la transmet à un serveur, la requête est attribuée au premier serveur. Lorsque la deuxième demande est lancée, elle est acheminée vers le deuxième serveur, tandis que la troisième demande est envoyée au troisième serveur. Toutefois, si le serveur reçoit une quatrième demande, celle-ci est attribuée au premier serveur et une allocation de type anneau est créée. La limite du serveur est le nombre de demandes qu'il peut traiter. [33]

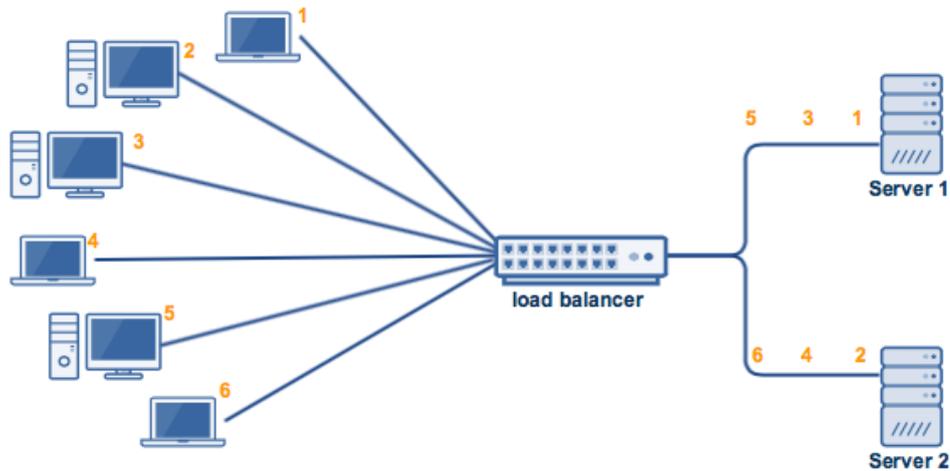


Figure 2.2: Fonctionnement de l'Algorithme Roundrobin

Algorithme aléatoire : Nous choisissons les nœuds au hasard et non selon certains critères spécifiques. Le serveur envoie la demande à un nœud avec une valeur de charge inférieure, et la demande peut alors être traitée par ce nœud. Cette action spécifique est effectuée par l'équilibreur de charge. Avec cet algorithme, nous avons un problème de surcharge, où tout nœud avec une charge élevée peut transmettre la charge à un nœud déjà surchargé. [33]

Algorithme du gestionnaire central : Dans l'algorithme du gestionnaire central d'équilibrage de charge, nous sélectionnons un nœud central parmi plusieurs nœuds en fonction du nombre de demandes qu'il peut recevoir. Une fois le nœud central sélectionné, nous gérons les demandes à travers celui-ci. Le nœud central transfère la charge du nœud surchargé au nœud peu chargé et crée le chemin approprié pour définir le travail principal. Cette méthode n'opère pas de manière distribuée. Nous devons gérer toutes les fonctions à travers un nœud central. Cela est inefficace pour un réseau étendu et dispersé. [33]

2.4.2 Equilibrage de charge dynamique

Dans cette technique, plusieurs nœuds sont chargés de prendre les bonnes décisions d'équilibrage de charge. De manière dynamique, le nœud est incertain et la méthode de requête n'est pas simple. Parfois, la charge est faible et parfois élevée. Par conséquent, nous avons utilisé un algorithme d'équilibrage dynamique qui donne des résultats corrects en termes de temps d'exécution. Chaque nœud contient des informations sur d'autres nœuds ou des nœuds voisins. Dans un scénario distribué, la défaillance d'un nœud ne

sera pas ignorée, nous avons donc un nouveau concept pour rendre le système tolérant aux pannes. Par conséquent, dans l'approche actuelle, nous effectuons un équilibrage de charge dynamique au lieu d'un équilibrage de charge statique. [33] Plusieurs algorithmes dynamiques sont discutés ci-dessous :

Algorithme du plus proche voisin (APV) : Chaque processeur de l'algorithme du plus proche voisin exécute l'algorithme d'équilibrage de la charge. Nous recherchons un nœud adjacent dans cette procédure, et si nous localisons un nœud surchargé parmi les voisins, nous déplaçons la demande vers ce nœud. Dans le cas contraire, ces voisins transmettront leur charge à l'un de leurs voisins. Les informations relatives à la charge sont transmises localement, ce qui signifie que nous travaillons dans un contexte contraint. Dans cette approche, tous les processus utilisent des processeurs proches et la répartition de la charge entre tous les nœuds est égale. Un équilibreur de charge global est utilisé pour garantir la réussite de la répartition locale de la charge. [33]

Algorithme RAND : Si un nœud spécifie une charge de travail qui dépasse le seuil limite, il est transféré à un nœud sélectionné de manière aléatoire dans la méthode d'équilibrage de la charge basée sur la sélection aléatoire des nœuds . Cette méthode n'examine pas les informations relatives au nœud pour déterminer s'il est surchargé ou sous-chargé. Il ne suit pas non plus la charge locale et ne communique pas d'informations sur la charge aux autres processeurs. En outre, il est simple à créer et à appliquer. Toutefois, en raison de la sélection aléatoire d'un nœud fortement chargé, il génère une surcharge de communication. Par conséquent, le partage des informations est géré localement et chaque nouveau nœud est choisi au hasard. Le fait que la charge soit répartie de manière asynchrone est un aspect essentiel de cette approche. [33]

Dans notre infrastructure, nous avons choisi d'implémenter HAProxy en tant que répartiteur de charge, afin de distribuer efficacement le trafic entre nos serveurs web et nos bases de données. Pour ce faire, nous avons opté pour l'algorithme Round Robin, une stratégie d'équilibrage de charge qui attribue les requêtes de manière circulaire et équitable à chaque serveur. Cette approche permet de maximiser l'utilisation des ressources et d'améliorer la performance globale de notre système, tout en évitant les goulots d'étranglement et les surcharges d'un serveur unique. Cette stratégie garantit non seulement une répartition équitable du trafic, mais contribue également à la fiabilité et à la haute disponibilité de notre infrastructure.

2.5 Network File System

Network File System (NFS) est un système de fichiers qui permet d'accéder aux fichiers à partir de plusieurs hôtes partageant un réseau informatique. Cela permet de partager des fichiers et des ressources de stockage entre plusieurs utilisateurs sur plusieurs machines. NFS offre un accès transparent et distant aux systèmes de fichiers. Il est important de choisir NFS car il peut être indépendant de la machine et du système d'exploitation, propose des mécanismes de récupération après une panne avancés et est hautement évolutif. [34]

Il existe plusieurs façons de créer un stockage NFS (Network File System). Voici quelques méthodes couramment utilisées :

Configuration manuelle sur le serveur : C'est une méthode qui consiste à configurer manuellement le serveur NFS en installant et en configurant le logiciel NFS approprié sur le serveur. Cette approche implique généralement la configuration des partages NFS, des autorisations d'accès, des options de sécurité et des exportations de fichiers. [34]

distribution Linux : De nombreuses distributions Linux telles que Ubuntu, CentOS et Fedora ont des fonctionnalités intégrées pour configurer un serveur NFS. L'outil de configuration fourni par la distribution peut être utilisé pour créer et gérer des partages NFS. [34]

Solutions de stockage en réseau (NAS) : Les périphériques NAS sont des périphériques de stockage autonomes qui offrent une solution complète de stockage en réseau avec une capacité NFS native. Ils sont conçus pour être simples à configurer et à utiliser, avec une interface intuitive qui permet aux administrateurs de créer et de gérer des partages NFS. Les fonctions avancées des périphériques NAS comprennent le contrôle des autorisations d'accès, la sécurité des données et la surveillance des performances. [35]

Dans le cadre de notre stratégie de gestion du stockage, nous avons adopté TrueNAS comme solution de stockage en réseau (Network File System, NFS). TrueNAS, grâce à sa robustesse et sa fiabilité, nous a permis de stocker et de partager efficacement nos données sur le réseau. Cette plateforme offre non seulement une grande flexibilité et une performance élevée, mais assure aussi la sécurité et l'intégrité des données, des aspects essentiels pour notre infrastructure.

2.6 Clustering

Le clustering est une technologie utilisée pour combiner plusieurs serveurs ou nœuds afin qu'ils fonctionnent ensemble comme un système distribué. Un cluster présente les caractéristiques importantes suivantes : [36]

Évolutivité horizontale : Le clustering facilite l'ajout de serveurs supplémentaires au cluster, augmentant ainsi la capacité et permettant de gérer les charges de travail croissantes.

Équilibrage de charge : Le cluster répartit la charge de travail entre les serveurs afin d'assurer une utilisation équilibrée des ressources et de minimiser les goulots d'étranglement. [36]

Haute disponibilité : Les méthodes de redondance peuvent être mises en œuvre à l'aide du clustering pour garantir la disponibilité continue des services même en cas de défaillance d'un ou de plusieurs serveurs. [36]

Tolérance aux pannes : si un nœud du cluster tombe en panne, d'autres nœuds peuvent prendre le relais et assurer la continuité du service, minimisant ainsi les interruptions. [36]

Réplication synchrone ou asynchrone : le cluster vous permet de choisir entre la réplication synchrone (les données sont écrites sur plusieurs nœuds en même temps) ou la réplication asynchrone (les données sont répliquées périodiquement). [37]

nous avons choisi d'implémenter Galera Cluster pour gérer nos bases de données. Cette solution assure une réplication multi-maître synchrone de nos bases de données, ce qui contribue à améliorer la disponibilité et la fiabilité de nos données. Galera Cluster offre une haute performance, une tolérance aux pannes et garantit que toutes les instances de la base de données restent en synchronisation, ce qui est essentiel pour maintenir l'intégrité de nos données et assurer un fonctionnement ininterrompu de nos services. [37]

2.7 Cache

La mise en cache est une technique qui accélère l'accès aux données en conservant une copie des données fréquemment utilisées dans un espace de stockage à accès rapide. La

mise en cache présente plusieurs avantages, notamment une vitesse améliorée, une diminution de la charge du serveur, une réduction du trafic réseau, une expérience utilisateur améliorée et des économies de coûts. Voici quelques exemples de systèmes de gestion de cache populaires.

Redis : Redis est un magasin de données en mémoire qui peut être utilisé pour gérer des caches. Redis prend en charge les chaînes, les listes, les ensembles, les ensembles ordonnés et d'autres structures de données. [38]

Memcached : Memcached est une solution de gestion de cache bien connue. Elle est destinée à être utilisée dans des situations où d'importants volumes de données doivent être mis en cache. Memcached enregistre les informations sous forme de paires clé-valeur. [38]

Varnish : Varnish est un séparateur de charge HTTP ou un accélérateur d'applications web. Il est largement utilisé pour mettre en cache du matériel en ligne afin d'accélérer la diffusion des pages web. [38]

CDN (Content Delivery Network) : Les CDN sont des réseaux de serveurs distribués à l'échelle mondiale qui fournissent du matériel internet aux utilisateurs aussi rapidement que possible en fonction de leur localisation. Pour augmenter la vitesse de diffusion du contenu, le réseau de diffusion de contenu (CDN) met en cache le contenu en ligne sur un réseau de serveurs. [38]

Redis a été choisi comme notre système de mise en cache. Cette technologie éprouvée a été intégrée afin d'optimiser les performances en stockant les données fréquemment demandées dans une mémoire rapide et accessible, réduisant ainsi les requêtes coûteuses vers la base de données. Cela a permis d'améliorer significativement la réactivité de notre plateforme de partage de fichiers, offrant une expérience utilisateur plus fluide et rapide.

2.8 LDAP

Le protocole réseau LDAP (Lightweight Directory Access Protocol) permet d'accéder aux services d'annuaire et de les gérer. Ce protocole est utilisé dans de nombreuses applications, telles que l'authentification des utilisateurs, la configuration du système et les services de carnet d'adresses. LDAP utilise une structure de base de données hiérarchique pour gérer les données, qui contiennent des éléments tels que des utilisateurs, des groupes et des ressources, ainsi que leurs caractéristiques connexes. Il comprend un schéma qui

décrit les nombreux types d'objets et d'entrées qui peuvent être produits dans une base de données LDAP, ainsi que leurs propriétés nécessaires et optionnelles. LDAP est couramment utilisé pour l'authentification et l'autorisation des utilisateurs, et est généralement utilisé en combinaison avec d'autres systèmes tels que Kerberos ou SAML. Pour la tolérance aux pannes et la distribution des frais, les annuaires LDAP peuvent être dupliqués sur de nombreux serveurs. Enfin, LDAP est un protocole d'application ouvert, indépendant du fournisseur et largement utilisé dans les entreprises. Il est souvent utilisé dans les entreprises pour la gestion centralisée des utilisateurs et comme colonne vertébrale pour les systèmes d'authentification unique (SSO), ainsi que pour le stockage des données de configuration du réseau pour des ressources telles que les imprimantes et les serveurs. OpenLDAP est un serveur LDAP open source de premier plan, et Active Directory de Microsoft utilise LDAP comme l'un de ses principaux protocoles d'accès. [39]

2.9 Conclusion

Dans ce chapitre, nous avons effectué une analyse approfondie des besoins spécifiques de Sonatrach, puis nous avons développé une architecture pour une plateforme de partage de fichiers redondante et scalable en réponse à ces besoins. Nous avons décrit en détail tous les composants de cette architecture, les technologies employées, et fourni une justification pour chaque décision de conception. Notre objectif était de nous assurer que chaque choix effectué répondait à une exigence précise et de fournir une compréhension complète du processus de conception afin de garantir une mise en œuvre réussie.

Chapitre 3

Mise en oeuvre de la solution

3.1 Introduction

Dans ce chapitre, nous examinons la mise en œuvre de notre solution. Tout d'abord, nous détaillons les environnements de travail que nous avons utilisés, tant sur le plan matériel que logiciel. Ensuite, nous passons à la partie implémentation, où nous décrivons en détail les étapes de notre travail.

3.2 Environnement de travail

Dans cette section, nous allons examiner les environnements de travail que nous avons utilisés.

3.2.1 Partie matérielle

Serveur principal : Voici les caractéristiques de notre serveur principal :

Pc	Processeur	Ram	Disque
Dell R730	Intel Xeon E5 2600 v4	256 Go	2*300 Go

Tableau 3.1: Caractéristique du Serveur principal

Les machines Virtuelle : Voici les caractéristiques de nos machines virtuelle :

Nom	Id	RAM	Stockage	SE	Coeur
Web1	101	16 Go	32 Go	Linux	4 Coeurs
Web2	105	16 Go	32 Go	Linux	4 Coeurs
Galera1	104	16 Go	32 Go	Linux	4 Coeurs
Galera2	108	16 Go	32 Go	Linux	4 Coeurs
TrueNas	102	8 Go	64 Go	FreeBSD	4 Coeurs
Haproxy	100	16 Go	32 Go	Linux	4 Coeurs
HaproxyDB	109	16 Go	32 Go	Linux	4 Coeurs

Tableau 3.2: Caractéristique des VMs

3.2.2 Partie Logiciels

Dans cette partie, nous décrivons les logiciels utilisés dans notre travail. Nous avons sélectionné des outils performants et adaptés à nos besoins afin d’accomplir nos tâches de manière efficace et efficiente.

3.2.2.1 Proxmox

Proxmox est une plateforme de virtualisation open-source basée sur Debian. Elle combine la virtualisation de serveurs et la gestion de conteneurs en une seule solution. Ses principales caractéristiques sont la virtualisation basée sur KVM, la conteneurisation basée sur LXC, une interface web conviviale, des capacités de haute disponibilité, une gestion centralisée, une sauvegarde et une récupération intégrées, et une communauté active. Proxmox est largement utilisé pour la virtualisation de serveurs et la gestion de conteneurs dans les environnements d’entreprise.



Figure 3.1: logo de Proxmox

3.2.2.2 HaProxy

HAProxy est un logiciel open source utilisé pour l’équilibrage de charge et le proxy inverse. Il équilibre le trafic entre les serveurs backend, améliorant ainsi les performances

et la disponibilité des applications web. HAProxy offre des fonctionnalités avancées telles que la haute disponibilité, l'équilibrage de charge avancé et la sécurité. C'est une solution populaire pour optimiser les performances des applications web.



Figure 3.2: logo de HaProxy

3.2.2.3 Php

PHP est le langage de programmation utilisé dans le développement de Nextcloud. Il gère les demandes des utilisateurs, l'authentification et l'autorisation, et permet d'étendre les fonctionnalités par le biais d'extensions et de plugins. PHP facilite également l'intégration avec d'autres composants et services. En bref, PHP joue un rôle central dans le fonctionnement et la fonctionnalité de Nextcloud.



Figure 3.3: logo de Php

3.2.2.4 Apache

Apache est un serveur web couramment utilisé avec Nextcloud. Il héberge les fichiers et les ressources nécessaires à l'application. Apache offre sécurité, performance et intégration avec d'autres modules. Il constitue une base solide pour faire fonctionner Nextcloud de manière fiable et sécurisée.



Figure 3.4: logo de Apache

3.2.2.5 MariaDB

MariaDB est une base de données open source, compatible avec MySQL, offrant des performances accrues, une haute disponibilité, une sécurité avancée et une large communauté de support. Il s'agit d'une alternative populaire pour les applications nécessitant une gestion de base de données robuste et évolutive.



Figure 3.5: logo de MariaDB

3.2.2.6 Galera Cluster

Galera Cluster est une solution de clustering multi-maîtres pour bases de données, utilisée notamment avec MariaDB. Elle offre une réplication synchrone, une haute disponibilité et une capacité d'écriture distribuée. Galera Cluster permet la réplication en temps réel des transactions sur tous les nœuds du cluster, garantissant la cohérence des données et la résilience en cas de défaillance. Il s'agit d'une solution performante et robuste pour les applications nécessitant une haute disponibilité des bases de données. [40]



Figure 3.6: logo de Galera Cluster

3.2.2.7 TrueNas

TrueNAS est une plateforme de stockage complète basée sur ZFS. Elle offre des fonctionnalités avancées pour la gestion du stockage, la sauvegarde, la réplication et la virtualisation. TrueNAS garantit la haute disponibilité et la redondance des données, et prend en charge plusieurs protocoles de stockage. Il est doté d'une interface conviviale pour une administration aisée et est utilisé dans divers environnements pour répondre aux besoins de stockage et de gestion des données. [35]



Figure 3.7: logo de TrueNAS

3.2.2.8 Redis

Redis est une base de données clé-valeur ce qui signifie que Chaque donnée est associée à une clé unique qui permet d’y accéder rapidement. Elle est souvent utilisée comme cache de données pour améliorer les performances, mais peut également être utilisée comme système de stockage primaire. Redis offre de hautes performances, une persistance optionnelle, une large gamme de structures de données et des fonctionnalités avancées telles que la réplication et le partitionnement. Il est largement utilisé dans les applications nécessitant une grande vitesse et une grande flexibilité. [38]



Figure 3.8: logo de Redis

Le tableau ci desous vas resumer toute les technologies utilisé avec sa version :

Technologie	Version
Linux	Ubuntu 22.04.2 LTS
Proxmox	7.3-3
Nextcloud	24.0.9.2
Php	8.1.2
Apache	2.4.52
Haproxy	2.4.18
TrueNAS	13.0-U4
MariaDB	10.6.12

Tableau 3.3: Liste des technologies utilisées avec sa version

3.3 déploiement

Dans cette partie, nous allons fournir une description détaillée de l'implémentation de notre solution. Nous allons explorer chaque aspect de notre approche, en mettant l'accent sur les étapes clés et les choix techniques que nous avons faits tout au long du processus.

3.3.1 Première étape

La première étape consiste à configurer la plate-forme Proxmox avec toutes les machines virtuelles requises. Pour cette configuration particulière, nous aurons besoin des VM suivantes.

HAproxy pour les serveurs web : Cette machine sera chargée de répartir les charges sur les serveurs web Nextcloud. Elle acheminera le trafic des utilisateurs vers les serveurs disponibles afin de garantir une disponibilité et une performance optimales.

Deux machines pour les serveurs web Nextcloud : Ces machines hébergeront les instances de Nextcloud, offrant une interface web conviviale pour les utilisateurs.

HAproxy pour Galera cluster : Cette machine achemine le trafic de la base de données vers les serveurs de base de données du cluster Galera, ce qui garantit une distribution uniforme des requêtes et une bonne disponibilité de la base de données.

deux machines pour galera cluster : Ces VM formeront une cluster de bases de données synchronisées en temps réel, assurant la redondance et la résilience en cas de défaillance d'un serveur. Ils assureront l'intégrité et la disponibilité des données de Nextcloud.

Une machine pour TrueNAS : Cete machine sera utilisé comme une plateforme de stockage à part entière basée sur NFS. Elle fournira des fonctions avancées de gestion du stockage, de sauvegarde, de réplication et de virtualisation pour Nextcloud.

3.3.2 Deuxieme étape

Cette étape est consacrée à l'installation des logiciels essentiels sur les machines virtuelles (VMs) .Il est nécessaire de mettre à jour le systeme ubuntu avant le debut de l'installation avec la commande suivante :

```
$ sudo apt update
```

HAproxy pour les serveurs web : Dans cette machine, nous avons besoin d'installer le logiciel HAProxy. La commande suivante permet le téléchargement et l'installation :

```
$ sudo apt-get install haproxy
```

Il est nécessaire d'activer et de démarrer le service HAProxy en utilisant les commandes suivantes :

```
$ systemctl start haproxy  
$ systemctl enable haproxy
```

les deux serveurs Nextcloud : Les étapes suivantes sont les mêmes pour les deux serveurs :

Premièrement, nous avons installé, activé et démarré le serveur web Apache :

```
$ sudo apt install apache2  
$ sudo systemctl enable apache2  
$ sudo systemctl start apache2
```

Deuxiemment, nous avons installé PHP avec ses extensions :

```
$ sudo apt-get install php8.1 php8.1-cli php8.1-common php8.1-  
imap php8.1-redis php8.1-snmp php8.1-xml php8.1-zip php8.1-m  
bstring php8.1-curl php8.1-gd php8.1-mysql
```

Ensuite, nous avons procédé à l'installation de MariaDB en local, dans le but de le migrer ultérieurement vers une machine indépendante dédiée :

```
$ sudo apt install mariadb-server  
$ sudo systemctl start mariadb  
$ sudo systemctl enable mariadb
```

Après l'installation du serveur de base de données, nous avons créé une base de données appelée "nextcloud" et un utilisateur à qui nous avons attribué tous les privilèges pour accéder à la base de données.

```
$ mysql -V  
$ CREATE DATABASE nextcloud ;  
$ GRANT ALL PRIVILEGES ON nextcloud.* TO 'user_'@'localhost '  
IDENTIFIED BY 'Password' ;  
$ FLUSH PRIVILEGES ;  
$ exit ;
```

Dans cette étape, nous avons téléchargé le fichier "nextcloud.zip" sur notre serveur. Une fois le téléchargement terminé, nous avons extrait ce fichier et modifié les autorisations correspondantes.

```
$ cd /var/www/html  
$ wget https://download.nextcloud.com/server/releases/nextcloud-24.0.1.  
$ unzip nextcloud-24.0.1.zip  
$ chown -R www-data:www-data /var/www/html/nextcloud
```

Enfin, Nous avons finalisé l'installation en accédant à l'interface de Nextcloud via notre navigateur en saisissant l'adresse IP en boucle locale. :

http://127.0.0.1

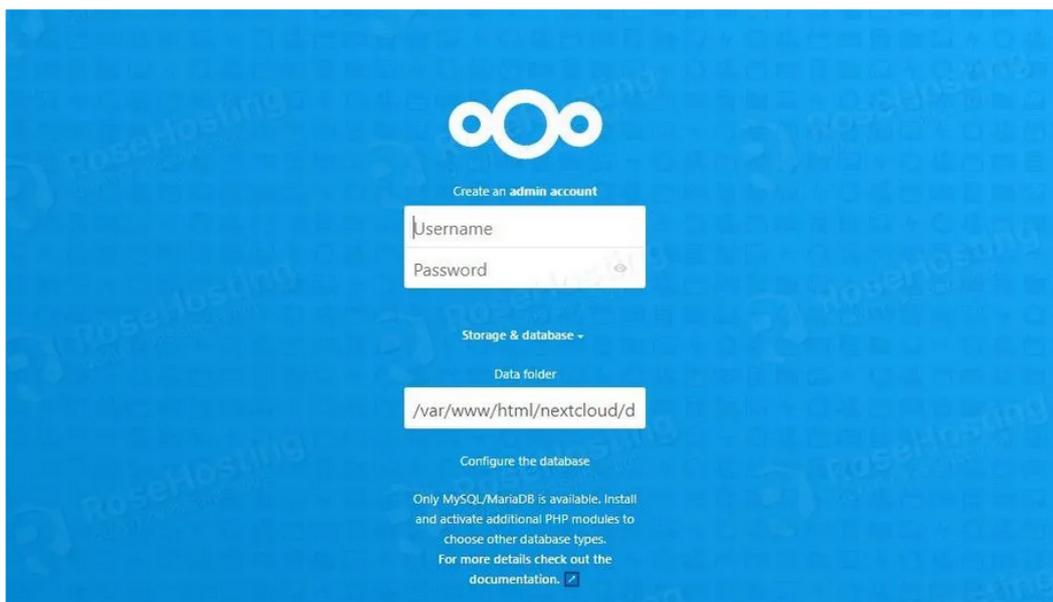


Figure 3.9: créer un nouveau user

Il nous reste à faire défiler vers le bas de la page et remplir les champs de création d'un nouvel utilisateur, ainsi que le nom de la base de données que nous avons précédemment créée.

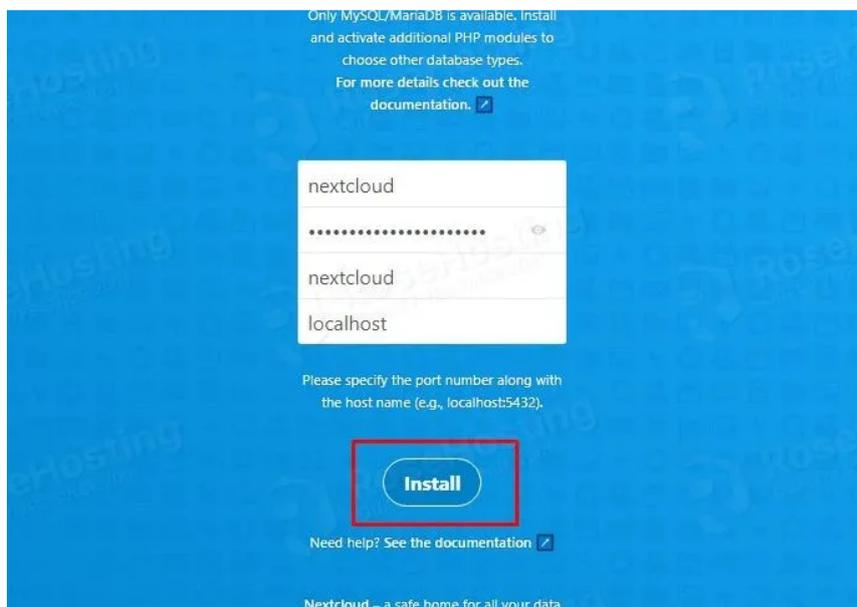


Figure 3.10: Interface d'installation

Dans la fenêtre suivante, nous avons installé les applications recommandées :

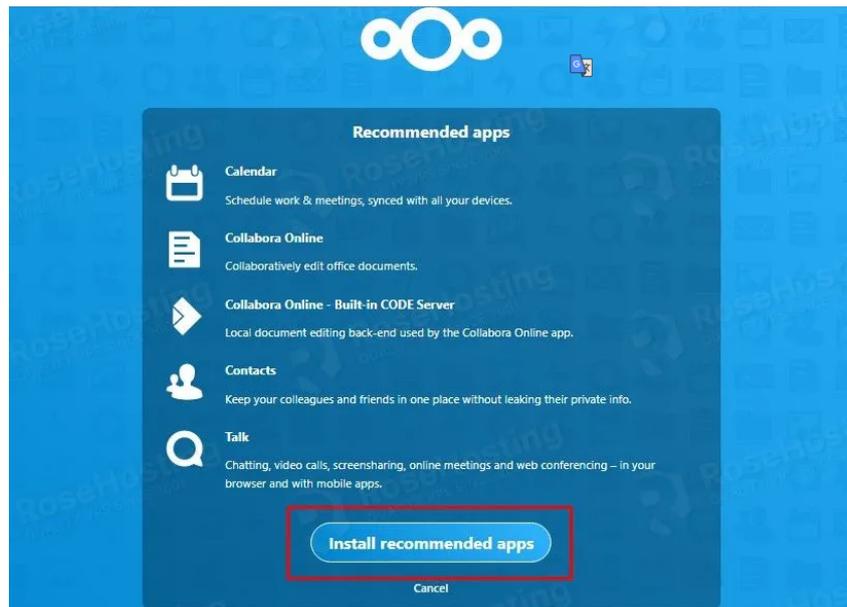


Figure 3.11: Installation finale

HAproxy pour Galera cluster : Les mêmes procédures ont été suivies dans la première machine HAproxy.

deux machines pour galera cluster : Dans cette partie, nous allons installer MariaDB dans les deux machines :

```
$ sudo apt install mariadb-server  
$ sudo systemctl start mariadb  
$ sudo systemctl enable mariadb
```

Voici l'interface de Proxmox après la création des Machines Virtuelles :

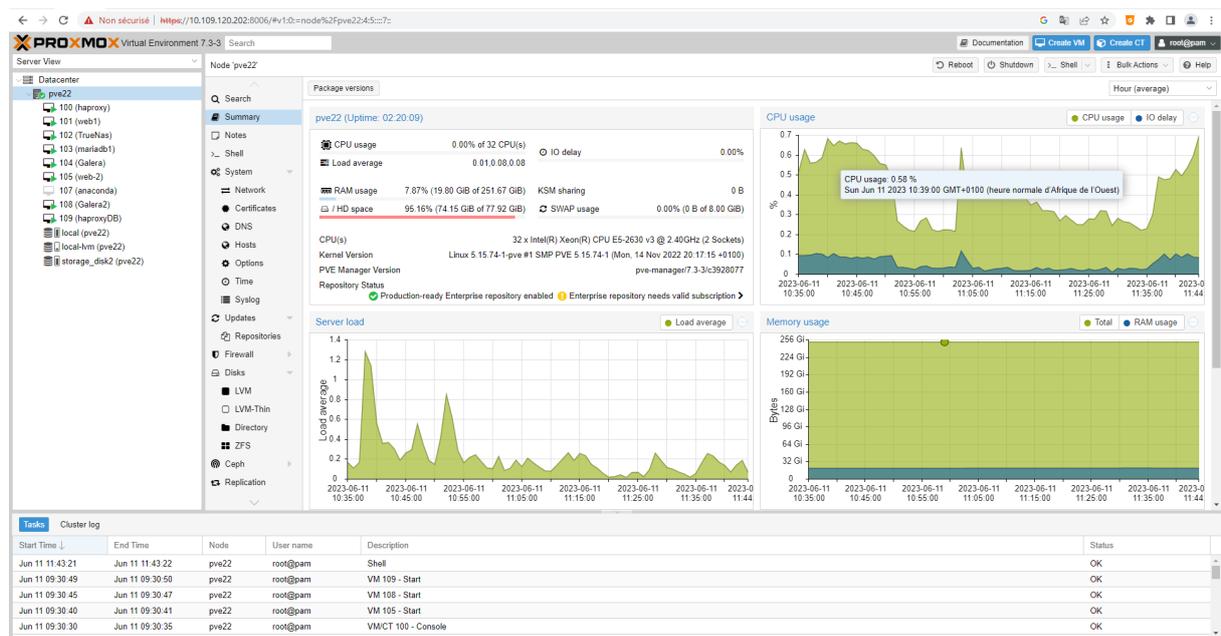


Figure 3.12: interface Proxmox après la création des machines

3.3.3 Troisième étape

La troisième étape est composée de trois phases. la première phase consiste à configurer et mettre en marche l'équilibrage de charge, Ensuite, dans la deuxième phase, le système NFS est migré vers TrueNAS, Enfin, dans la dernière phase, le focus est mis sur la configuration du cluster.

Phase1(configuration de load balancing) : Pour configurer HAProxy, nous avons modifié le fichier haproxy.cfg qui se trouve dans le répertoire /etc/haproxy en ajoutant les ligne suivantes :

```

root@HAproxy:/home/web1
root@HAproxy:/home/web1# cd /etc/haproxy/haproxy.cfg
bash: cd: /etc/haproxy/haproxy.cfg: Not a directory
root@HAproxy:/home/web1# cat /etc/haproxy/haproxy.cfg
global
    log /dev/log      local0
    log /dev/log      local1 notice
    chroot /var/lib/haproxy
    stats socket /run/haproxy/admin.sock mode 660 level admin expose-fd list

defaults
    stats timeout 30s
    user haproxy
    group haproxy
    daemon

    # Default SSL material locations
    ca-base /etc/ssl/certs
    crt-base /etc/ssl/private

    # See: https://ssl-config.mozilla.org/#server=haproxy&server-version=2.0
    #.3&config=intermediate
    ssl-default-bind-ciphers ECDHE-ECDSA-AES128-GCM-SHA256:ECDHE-RSA-AES128-
    GCM-SHA256:ECDHE-ECDSA-AES256-GCM-SHA384:ECDHE-RSA-AES256-GCM-SHA384:ECDHE-ECDSA
    -CHACHA20-POLY1305:ECDHE-RSA-CHACHA20-POLY1305:DHE-RSA-AES128-GCM-SHA256:DHE-RSA
    -AES256-GCM-SHA384
    ssl-default-bind-ciphersuites TLS_AES_128_GCM_SHA256:TLS_AES_256_GCM_SHA
    384:TLS_CHACHA20_POLY1305_SHA256
    ssl-default-bind-options ssl-min-ver TLSv1.2 no-tls-tickets

defaults
    log          global
    mode         http
    option       httplog
    option       dontlognull
    timeout     connect 5000
    timeout     client 50000
    timeout     server 50000
    errorfile   400 /etc/haproxy/errors/400.http
    errorfile   403 /etc/haproxy/errors/403.http
    errorfile   408 /etc/haproxy/errors/408.http
    errorfile   500 /etc/haproxy/errors/500.http
    errorfile   502 /etc/haproxy/errors/502.http
    errorfile   503 /etc/haproxy/errors/503.http
    errorfile   504 /etc/haproxy/errors/504.http

frontend myfrontend
    bind *:80
    default_backend myservers

backend myservers
    balance roundrobin
    mode http
    cookie SERVER_USED insert indirect nocache
    server web1 10.109.120.5:80 check cookie web1
    server web2 10.109.120.54:80 check cookie web2
root@HAproxy:/home/web1#

```

Figure 3.13: Fichier de configuration haproxy

commande `bind * :80` permet de lier toutes les adresses IP disponibles sur le port 80 pour recevoir les requêtes.

`default` back-end pour rediriger ces adresses ip vers le backend-end `myservers` qui porte les adresses ip de nos serveurs web. `balance roundrobin` pour choisir l'algorithme d'équilibrage

de charge .

La commande cookie pour conserver la session lors de la redirection vers l'un des deux serveurs.

les commandes server web1 10.109.120.5 cookie web1 , server web2 10.109.120.54 cookie web2 pour définir nos serveurs web.

on sauvegarde le fichier et on redémarre le service haproxy :

```
$ sudo systemctl restart haproxy
```

Maintenant, nous allons effectuer une modification dans le fichier config.php situé dans le répertoire /var/www/html/nextcloud sur les deux serveurs Nextcloud. Nous ajouterons l'adresse IP de la machine HAProxy .

```

GNU nano 6.2                                config.php
?php
$CONFIG = array (
  'instanceid' => 'oc0vdruo99oo',
  'passwordsalt' => 'Dwf2wyRxDT5DC+4dugg3oiYkvR/CEp',
  'secret' => 'Lb1UW/UQVlnEiC4OabpfScS/W0r/87CJcmlld1HgznYHeS/6k',
  'trusted_domains' =>
    array (
      0 => '10.109.120.249',
      1 => '10.109.120.5',
      2 => '10.109.120.54',
    ),
  'datadirectory' => '/mnt/data',
  'dbtype' => 'mysql',
  'version' => '24.0.9.2',
  'overwrite.cli.url' => 'http://10.109.120.5/nextcloud',
  'dbname' => 'nextcloud',
  'dbhost' => '10.109.120.205:3307',
  'dbport' => '',
  'dbtableprefix' => 'oc_',
  'mysql.utf8mb4' => true,
  'dbuser' => 'nextcloud',
  'dbpassword' => 'nextcloud',
  'installed' => true,
);
    
```

Figure 3.14: Ajouter address ip de haproxy dans config.php sur web1 et web2

3.3.4 Phase2 (Configuration de TrueNAS)

Cette étape consiste à créer un pool de stockage et à autoriser le stockage NFS pour les deux serveurs web. Ensuite, nous allons monter les données stockées localement vers ce

pool de stockage. Création de pool de stockage : Dans le répertoire Storage/pool, nous allons créer un pool nommé "nextclouddata". Ensuite, nous cliquerons sur le bouton "Add" pour effectuer cette action.

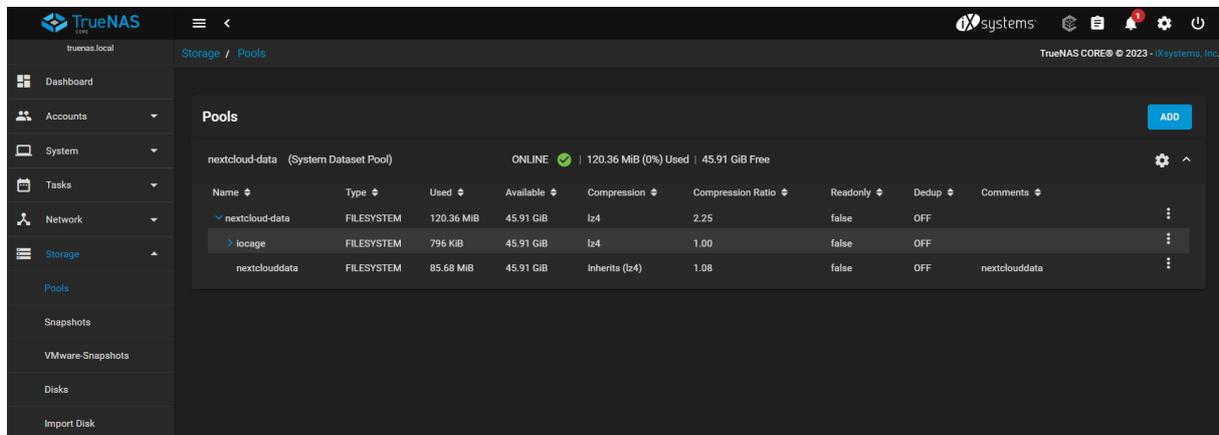


Figure 3.15: Créer un pool de stockage

Autoriser le service NFS : Pour activer le service NFS, nous avons accédé à la section "Services" puis nous avons activé la fonctionnalité NFS.

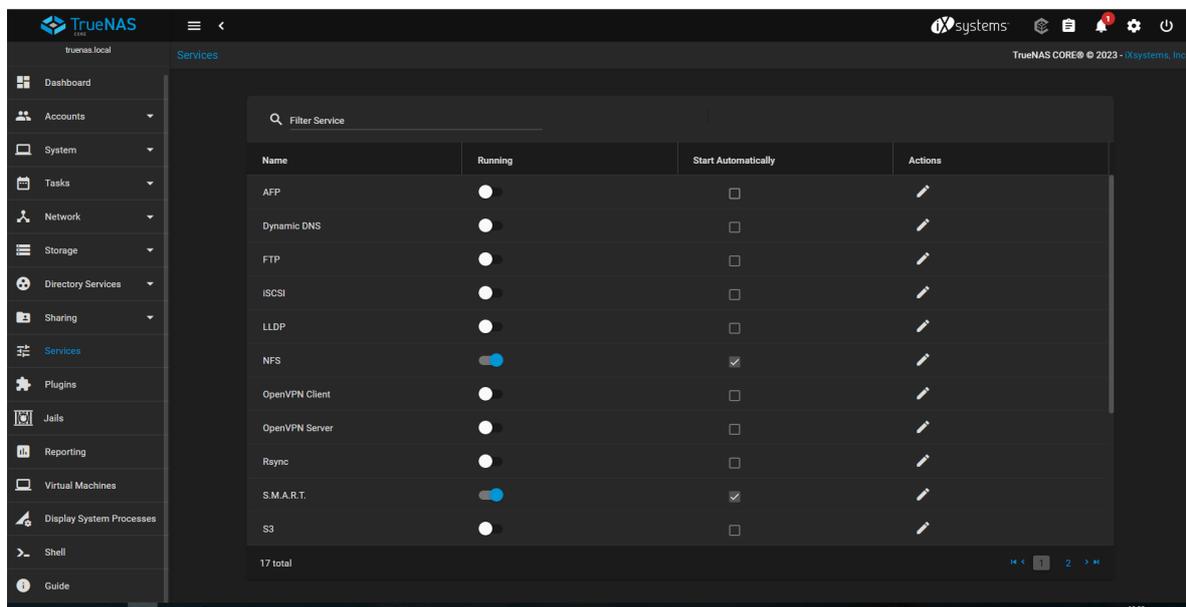


Figure 3.16: Autoriser le service NFS

Autoriser les address ip a stocker dans le pool Pour permettre à notre serveur Nextcloud d'accéder au stockage dans le TrueNAS, nous naviguons vers la section "Stockage" puis "Unix shares (NFS)". Ensuite, nous avons ajouté les adresses IP des deux serveurs.

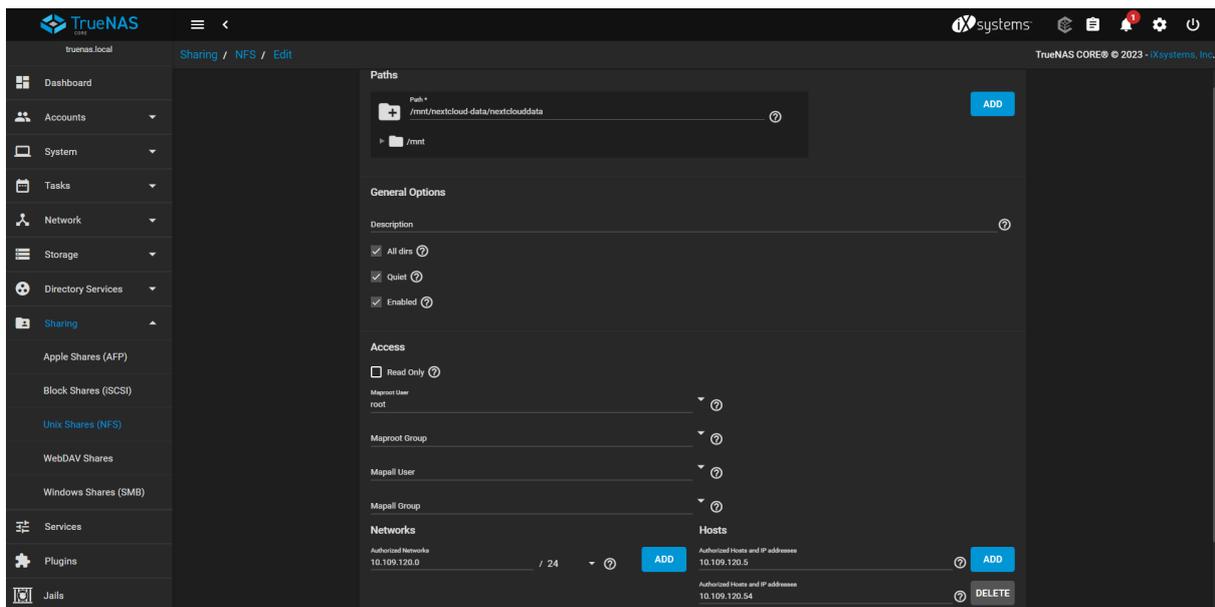


Figure 3.17: Autoriser les address ip a stocker dans le pool

monter les donnes vers TrueNAS : la commande mount permet de monter le stockage vers TrueNAS :

```
$ mount -t nfs 10.109.120.83:/mnt/nextclouddata /mnt/data
```

3.3.5 phase3 (configurer galera cluster et haproxy de la base de données) :

Dans cette phase, nous allons créer un cluster Galera entre les deux serveurs de base de données et mettre en place un équilibrage de charge avec HAProxy pour les interconnecter.

Création de cluster Pour le premier noeud, nous allons le stopper en utilisant la commande `systemctl stop mariadb`. Ensuite, nous allons modifier dans le fichier de configuration `/etc/mysql/mariadb.conf.d/50-server.cnf`. Nous allons commenter la ligne "bind-address" Ensuite, nous allons ajouter les lignes suivantes dans le fichier :

```

root@galera: /etc/mysql/mariadb.conf.d
GNU nano 6.2      50-server.cnf
[mariadb]

# This group is only read by MariaDB-10.6 servers.
# If you use the same .cnf file for MariaDB of different versions,
# use this group for options that older servers don't understand
[mariadb-10.6]
[galera]
wsrep_on                = ON
wsrep_cluster_name      = "Mariadb_Cluster"
wsrep_provider          = /usr/lib/galera/libgalera_smm.so
wsrep_cluster_address   = "gcomm://10.109.120.77,10.109.120.85"
binlog_format           = row
default_storage_engine  = InnoDB
innodb_autoinc_lock_mode = 2
bind-address            = 0.0.0.0
wsrep_node_address      = "10.109.120.77"

^G Help      ^O Write Out  ^W Where Is   ^K Cut        ^T Execute
^X Exit      ^R Read File  ^\ Replace    ^U Paste      ^J Justify
    
```

Figure 3.18: Fichier de configuration galera1

Ensuite, on va créer notre cluster avec la commande `galera_new_cluster`

Pour le deuxième nœud, nous allons commenter la ligne "bind-address" et ajouter les lignes suivantes : `systemctl restart mariadb`

```

root@galera2: /etc/mysql/mariadb.conf.d
GNU nano 6.2      50-server.cnf
[mariadb-10.6]
[galera]
wsrep_on                = ON
wsrep_provider          = /usr/lib/galera/libgalera_smm.so
wsrep_cluster_address   = "gcomm://10.109.120.77,10.109.120.85"
binlog_format           = row
default_storage_engine  = InnoDB
innodb_autoinc_lock_mode = 2
bind-address            = 0.0.0.0
wsrep_cluster_name      = "Mariadb_Cluster"
wsrep_node_address      = "10.109.120.85"

^G Help      ^O Write Out  ^W Where Is   ^K Cut        ^T Execute  ^C Location  M-U Undo
^X Exit      ^R Read File  ^\ Replace    ^U Paste      ^J Justify  ^/_ Go To Line M-E Redo
    
```

Figure 3.19: Fichier de configuration galera2

Enfin, nous allons redémarrer le service MariaDB en utilisant la commande

```
$ systemctl restart mariadb
```

on peut vérifier que le cluster a été bien créé en accédant à la base de données avec la commande

```
$ mysql -v"
—MariaDB[none]>"SHOW STATUS LIKE 'wsrep_cluster_size'
```

```

root@galera: /home/galera
Last login: Sun Jun 11 10:03:36 2023 from 10.109.120.202
galera@galera:~$ sudo -s
[sudo] password for galera:
root@galera: /home/galera# mysql -u root -p
Enter password:
Welcome to the MariaDB monitor.  Commands end with ; or \g.
Your MariaDB connection id is 37
Server version: 10.6.12-MariaDB-0ubuntu0.22.04.1-log Ubuntu 22.04

Copyright (c) 2000, 2018, Oracle, MariaDB Corporation Ab and others.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

MariaDB [(none)]> show status like 'wsrep_cluster_size';
+-----+-----+
| Variable_name | Value |
+-----+-----+
| wsrep_cluster_size | 2 |
+-----+-----+
1 row in set (0,002 sec)

MariaDB [(none)]>
    
```

Figure 3.20: Vérification de cluster

mettre le load balancer : La prochaine étape de cette phase consiste à mettre en place un équilibrage de charge entre les deux nœuds en ajoutant les lignes suivantes dans le fichier haproxy.cfg :

```

GNU nano 6.2 haproxy.cfg
frontend myfrontends
  bind *:3307
  mode tcp
  default_backend myservers

backend myservers
  mode tcp
  balance roundrobin
  server galera1 10.109.120.77:3306 check
  server galera2 10.109.120.85:3306 check
    
```

Figure 3.21: configuration haproxydb

migrer la base de données des serveurs nextcloud vers le cluster : pour migrer la base de données qui a été précédemment dans les serveurs nextcloud il faut créer une base de données et un nouvel utilisateur :

```

$ mysql -u root -p root
$ CREATE DATABASE nextcloud ; GRANT ALL PRIVILEGES ON nextcloud.* TO
  @localhost
  IDENTIFIED BY Password ;
    
```

```
$ FLUSH PRIVILEGES ;  
$ exit ;
```

apres on va faire un backup de la base de données nextcloud :

```
$ mysql -V  
$ mysqldump -u root -p root nextcloud > backup.sql
```

ensuite on va transférer backup.sql vers le cluster via ssh :

```
$ scp backup.sql root@10.109.120.77:/home
```

on va executer le fichier "backup.sql" avec la commande suivante :

```
$ mysql -u root -p root  
$ source /home/backup.sql
```

il nous reste que changer le fichier "config.php" dans les deux serveurs nextcloud en changeant l'address ip de la base de données local avec la l'address ip de load balancer de base de données.

3.4 Conclusion

Dans ce chapitre, nous avons présenté les étapes de mise en place de notre architecture de partage de fichiers scalable et redondante, tout en décrivant brièvement l'environnement de travail associé.

Conclusion Générale

Afin de répondre aux besoins croissants de collaboration et de gestion efficace des données au sein de Sonatrach, un projet a été initié pour mettre en place une plateforme de partage de fichiers redondante et scalable. Dans ce cadre, nous avons proposé une architecture basée sur Nextcloud, ainsi que faire une division des responsabilités entre le serveur web et la base de données. De plus, nous avons opté pour l'utilisation de la technologie de load balancing afin d'assurer une répartition équilibrée des charges et la technologie de stockage NFS pour stocker les fichiers sur un serveur distant.

Cette approche permet d'assurer une disponibilité élevée, une performance optimale, une redondance de l'information, open source et une capacité de stockage scalable pour répondre aux exigences de Sonatrach en matière de partage de fichiers.

À l'avenir, nous prévoyons d'améliorer l'architecture pour renforcer la sécurité et la résilience en ajoutant un service de sauvegarde. En outre, nous prévoyons d'introduire le déploiement automatique de machines virtuelles sur des machines physiques. Ces développements assureront une meilleure protection des données, une tolérance accrue aux pannes et une plus grande efficacité dans la gestion des ressources informatiques. Notre objectif est de garantir une infrastructure robuste, sécurisée et hautement disponible pour répondre aux besoins en constante évolution de Sonatrach.

Bibliographie

- [1] J Duffy and M Muchmore. The best cloud storage and file-sharing services for 2020. *online*, June, 2019.
- [2] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D Joseph, Randy Katz, Andy Konwinski, Gunho Lee, David Patterson, Ariel Rabkin, Ion Stoica, et al. A view of cloud computing. *Communications of the ACM*, 53(4) :50–58, 2010.
- [3] Shucheng Yu, Cong Wang, Kui Ren, and Wenjing Lou. Achieving secure, scalable, and fine-grained data access control in cloud computing. In *2010 Proceedings IEEE INFOCOM*, pages 1–9. Ieee, 2010.
- [4] John W Rittinghouse and James F Ransome. *Cloud computing : implementation, management, and security*. CRC press, 2016.
- [5] Mahesh Balakrishnan, Asim Kadav, Vijayan Prabhakaran, and Dahlia Malkhi. Differential raid : Rethinking raid for ssd reliability. *ACM Transactions on Storage (TOS)*, 6(2) :1–22, 2010.
- [6] Wayne Jansen, Tim Grance, et al. Guidelines on security and privacy in public cloud computing. 2011.
- [7] Peter Mell, Tim Grance, et al. The nist definition of cloud computing. 2011.
- [8] Graham Greenleaf. The influence of european data privacy standards outside europe : implications for globalization of convention 108. *International Data Privacy Law*, 2(2) :68–92, 2012.
- [9] Jerry FitzGerald and Alan Dennis. *Business data communications and networking*. John Wiley & Sons, 2007.
- [10] Quang Hieu Vu, Mihai Lupu, and Beng Chin Ooi. *Peer-to-peer computing : Principles and applications*. Springer, 2010.
- [11] Ping Lai, Hari Subramoni, Sundeep Narravula, Amit Mamidala, and Dhabaleswar K Panda. Designing efficient ftp mechanisms for high performance data-transfer over infiniband. In *2009 International Conference on Parallel Processing*, pages 156–163. IEEE, 2009.

- [12] Rajkumar Buyya. Introduction to the iee transactions on cloud computing. *IEEE Transactions on cloud computing*, 1(1) :3–21, 2013.
- [13] Sushant Goel and Rajkumar Buyya. Data replication strategies in wide-area distributed systems. In *Enterprise service computing : from concept to deployment*, pages 211–241. IGI Global, 2007.
- [14] Fabricio AB da Silva and Hermes Senger. Scalability limits of bag-of-tasks applications running on hierarchical platforms. *Journal of Parallel and Distributed Computing*, 71(6) :788–801, 2011.
- [15] Chung-Ming Huang and Tz-Heng Hsu. A resource exchange architecture for peer-to-peer file sharing applications. *The Computer Journal*, 48(1) :53–71, 2005.
- [16] Shamama Anwar, Ruchika Tulsyan, Souvik Saha, and Sudip Kumar Sahana. Anon-chain : A secure file sharing framework using ipfs integrated blockchain. *International Journal of Mathematical, Engineering and Management Sciences*, 7(6) :844, 2022.
- [17] Jasmine V Eggestein and Kenneth J Knapp. Fighting child pornography : A review of legal and technological developments. *Journal of Digital Forensics, Security and Law*, 9(4) :3, 2014.
- [18] Sooyoung Kim and Sun Joo Yoo. Age and gender differences in social networking : effects on south korean students in higher education. *Social networking and education : Global perspectives*, pages 69–82, 2016.
- [19] Morgan GI Langille and Jonathan A Eisen. Biotorrents : a file sharing service for scientific data. *PLoS One*, 5(4) :e10071, 2010.
- [20] Jussi Kiljander, Matti Eteläperä, Janne Takalo-Mattila, Juha-Pekka Soininen, and Kari Keinänen. Autonomous file sharing for smart environments. In *International Conference on Pervasive and Embedded Computing and Communication Systems*, volume 2, pages 191–196. SCITEPRESS, 2011.
- [21] Ian Lamont. *Google Drive & Docs in 30 minutes*. The Digital Media Machine, 2012.
- [22] Stack Overflow. Stack overflow - où les développeurs apprennent, partagent, & construisent des carrières, 2008.
- [23] Ashish Gupta, Inderpal Singh Mumick, et al. Maintenance of materialized views : Problems, techniques, and applications. *IEEE Data Eng. Bull.*, 18(2) :3–18, 1995.
- [24] Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung. The google file system. In *Proceedings of the nineteenth ACM symposium on Operating systems principles*, pages 29–43, 2003.
- [25] Apache Hadoop. Welcome to apache hadoop. *Hämtat från http ://hadoop. apache.org*, 2014.

- [26] Jürgen Cito, Philipp Leitner, Thomas Fritz, and Harald C Gall. The making of cloud applications : An empirical study on software development for the cloud. In *Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering*, pages 393–403, 2015.
- [27] Kevin Lai and Mary Baker. Nettetimer : A tool for measuring bottleneck link bandwidth. In *USITS*, volume 1, pages 11–11, 2001.
- [28] Jon Toigo. *Disaster recovery planning : Preparing for the unthinkable*. Prentice Hall Professional Technical Reference, 2002.
- [29] Jason Cannon. *High Availability for the LAMP Stack : Eliminate Single Points of Failure and Increase Uptime for Your Linux, Apache, MySQL, and PHP Based Web Applications*. CreateSpace Independent Publishing Platform, 2014.
- [30] Kereshmeh Afsari, Charles M Eastman, and Dennis R Shelden. Cloud-based bim data transmission : current status and challenges. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, volume 33, page 1. IAARC Publications, 2016.
- [31] William Stallings. *Computer security principles and practice*. 2015.
- [32] Jon Postel and Joyce Reynolds. File transfer protocol. Technical report, 1985.
- [33] Shantanu Shukla and Raghuraj Singh. Survey on load balancing techniques. 02 2019.
- [34] Susan Snedaker. *Business continuity and disaster recovery planning for IT professionals*. Newnes, 2013.
- [35] Taojiannan Yang, Linjie Yang, Xiaojie Jin, and Chen Chen. Revisiting training-free nas metrics : An efficient training-based method. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 4751–4760, 2023.
- [36] Guojun Gan, Chaoqun Ma, and Jianhong Wu. *Data clustering : theory, algorithms, and applications*. SIAM, 2020.
- [37] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.
- [38] Josiah Carlson. *Redis in action*. Simon and Schuster, 2013.
- [39] Tim Howes, Mark Smith, and Gordon S Good. *Understanding and deploying LDAP directory services*. Addison-Wesley Professional, 2003.
- [40] Bagus Aditya and Tutun Juhana. A high availability (ha) mariadb galera cluster across data center with optimized wrs scheduling algorithm of lvs-tun. In *2015 9th International Conference on Telecommunication Systems Services and Applications (TSSA)*, pages 1–5. IEEE, 2015.