

REPUBLIQUE DU SENEGAL



Un Peuple - Un But - Une Foi

Ministère de l'Enseignement Supérieur de la Recherche et de l'innovation

Direction Générale de l'Enseignement Supérieur Privé



**INSTITUT SUPERIEUR
D'INFORMATIQUE**

L'institut de référence dans les TIC

Km1, Avenue Cheikh Anta Diop BP : 28 110 Dakar Sénégal / Tel : 33 822 19 81 / Fax :

33 822 31 90

Site Web : www.isi.sn

**MEMOIRE DE FIN DE CYCLE EN VUE DE L'OBTENTION DU
DIPLOME DE LICENCE**

SPECIALITE : INFORMATIQUE

OPTION : RESEAU ET TÉLÉCOMMUNICATION

Etude et mise en place d'une architecture de stockage pour la
haute disponibilité

Présenté et soutenu par :

M. SAID AHMED Kaissane

Sous la direction de :

M. Racine Oumar Sall

Formateur en réseau et système

Année Académique : 2021 -2022

Dédicace

C'est avec gratitude que je dédie cet humble travail à ceux à qui je ne peux pas exprimer mon véritable amour, peu importe les moyens.

Merci à cet homme, mon chéri, qui me doit ma vie, ma réussite et tout mon respect :
mon cher père Abdallah MDAHOMA

À la femme qui souffre mais ne me laisse pas souffrir, ne demande jamais et ne ménage aucun effort pour me rendre heureux : ma charmante mère, Anziza AHMADA

A mon très cher grand frère que je le considère comme 2eme père SAID Kamardine et sa femme NAILA ; Rien au monde vaut leurs efforts qu'ils ont fournis pour mon éducation et mon bonheur. Ce travail est le résultat des sacrifices que vous faites pour moi, pour mon éducation et ma formation.

Dédié à mes grands-mères, oncles et tantes. Que Dieu leur accorde une vie longue et heureuse

A tous mes cousins, voisins et amis que j'ai connus jusqu'à présent. Merci pour leurs amours et leurs encouragements

Remerciements

En premier lieu, je remercie Dieu de ces faveurs, de sa gratitude et de m'avoir donné du courage, santé et patience pour élaborer ce magnifique travail.

Tout d'abord, ce travail n'aurait pas été aussi instructif et n'aurait pu être réalisé sans l'aide et la supervision de M. Racine Oumar Sall. Je le remercie pour son excellente qualité d'encadrement, sa patience, sa rigueur et sa disponibilité lors de la préparation de ce mémoire.

J'adresse mes remerciements les plus sincères au corps professoral et administratif de l'Institut Supérieur d'Informatique, pour la richesse et la qualité de leur enseignement et qui déploient de grands efforts pour assurer à leurs étudiants une formation actualisée.

Un grand merci à mon père Abdallah MDAHOMA, ma mère Anziza AHMADA et surtout mon grand frère SAID Kamardine, pour leurs amours, leurs conseils ainsi que leur soutien inconditionnel, à la fois moral et économique, qui m'a permis de réaliser les études que je voulais et par conséquent ce mémoire.

J'exprime aussi ma reconnaissance envers mes amis Attoisse MOHAMED, Mouan-ouia ALI, Iktissam HOUSSLOU pour leurs soutien spirituel et intellectuel tout au long du processus.

Enfin, mes remerciements les plus sincères à toutes personnes qui aurait contribué de près ou de loin à l'élaboration de ce travail ainsi qu'à la réussite de ces formidables années universitaires.

Avant-propos

L'Institut Supérieur d'Informatique (ISI) a été créé en 1988 par des étudiants de l'Université Cheikh Anta Diop de Dakar en collaboration avec des étudiants de l'Université de Laval, Québec, Canada, sous le nom de JET INFORMATIQUE. C'est ainsi qu'en 1994, il devient officiellement l'Institut Supérieur d'Informatique. L'ISI est une école spécialisée dans le domaine de l'informatique et se compose de trois départements : le département de réseautage, le département de génie logiciel et le département de gestion. Les diplômes délivrés sont DTS, BTS, Licence, Master...

Pour obtenir une licence pour un réseau de télécommunications, l'ISI demande aux étudiants de rédiger une thèse finale. C'est dans ce contexte que nous présentons ce document dont le sujet est : **Etude et mise en place d'une architecture de stockage pour la haute disponibilité.**

La perte de données sensibles peut avoir un impact énorme sur une entreprise et mettre en péril son existence même à court ou moyen terme. C'est pourquoi les entreprises de toutes tailles sont intéressées à la haute disponibilité de leurs données pour maintenir leurs activités. Dans cette rubrique, il s'agit de mettre en place une solution de haute disponibilité de données.

Ce document constitue notre premier effort de recherche académique, c'est pourquoi nous demandons au jury d'être très indulgent dans son appréciation.

Glossaire

CSV: comma separated values

DAS: Direct Attached Storage

FTP: File Transfer Protocol

HTTP: Hypertext Transfer Protocol

IP : Internet Protocol

JBOD (Just a Bunch Of Disks) : permet de regrouper plusieurs disques durs en une seule partition.

LUN : Logical Unit Number

MTBF : Mean Time Between Failure

NAS : Network Attached Storage est un serveur de stockage directement attaché au réseau IP fournissant un service de partage de fichiers aux clients /serveurs d'un environnement hétérogène (multi-OS).

RAID (Redundant Array of Independents Disks) : divise ou copie la tâche d'un parmi de nombreux disques durs (ou simplement de deux) afin d'améliorer les performances ou de générer la redondance des données en cas de panne du disque.

RAID 0 ou **Striping** (entrelacement ou agrégat par bande, parfois injustement appelé **striping**) consiste à stocker les données en les répartissant sur l'ensemble des disques de la grappe.

RAID 1 ou ***mirroring*** ou ***shadowing*** a pour but de dupliquer l'information à stocker sur plusieurs disques.

SAN : Storage Area Network est un réseau de stockage sur lequel transitent des blocs de données.

SCSI : est l'acronyme de **Small Computer System Interface**. SCSI est une norme qui permet de relier un ordinateur à un périphérique en mode bloc (disque, lecteur CDROM...). C'est un protocole client/serveur. Dans la norme SCSI, on parle d'initiateurs et de cibles.

SMTP: Simple Mail Transfer Protocol

TCP: Transmission Control Protocol

TFTP: Trivial File Transfer Protocol

UDP : User Datagram Protocol.

Liste des figures

Figure 1: Modèle TCP/IP	6
Figure 2: Modèle OSI.....	8
Figure 3: Environnement client/serveur	9
Figure 4: centre des données	10
Figure 5: Architectured'un DAS.....	11
Figure 6: architecture du NAS.....	13
Figure 7: architecture d'un SAN.....	15
Figure 8: technologies JBOD	21
Figure 9: technologies RAID0	22
Figure 10: technologie RAID1	23
Figure 11: technologie RAID4	24
Figure 12: technologie RAID5	25
Figure 13: Technologie RAID6.....	26
Figure 14: Comparaison RAID Standards.....	26
Figure 15: RAID 01.....	28
Figure 16: RAID10.....	29
Figure 17: RAID 05.....	30
Figure 18: RAID 15.....	30
Figure 19: RAID 51.....	31
Figure 20: Architecture du déploiement.....	33
Figure 21: photo du logiciel VMWARE	34
Figure 22: installation de rôle d'ISCSI	36
Figure 23: création des disques virtuels iscsi	36
Figure 24: initiation des disques virtuels.....	37
Figure 25: installation du rôle de clustering de basculement	38
Figure 26: ajout des serveurs pour le cluster	39
Figure 27: création d'un cluster	39
Figure 28: ajout les disques au cluster.....	40
Figure 29: les disques au cluster avec les configurations.....	40
Figure 30: installation du serveur de fichier.....	41
Figure 31: ajout le rôle au cluster.....	42

Figure 32: création du fichier de partage.....	43
Figure 33: image du partage	44
Figure 34: rôle du serveur de fichier	45
Figure 35: image de suspension d'un des serveurs	45
Figure 36: installation de 2 rôles pour la réplication.....	47
Figure 37: groupe de réplication	48
Figure 38: ajout des serveurs.....	49
Figure 39: choix du serveur principal.....	49
Figure 40: choix d'emplacement du fichier répliqué.....	50
Figure 41: creation du document dans le serveur principal.....	51
Figure 42: réplication des données	51

Liste des tableaux

Tableau 1: comparatif des solutions de stockage.....	17
Tableau 2: Comparatif de SAN et NAS	17
Tableau 3: Tableau comparatif des solutions des RAIDS	32

Sommaire

Introduction	1
I. Les Cadres théoriques et méthodologique	2
1.1. Cadre théorique	2
1.2. Cadre méthodologique	4
II. Cadre conceptuel	6
2.3. Rappel sur le réseau.....	6
2.5. Rappel sur la haute disponibilité	18
III. Mise en œuvre	33
3.6. Présentation de la solution.....	33
3.7. Implémentation.....	34
3.8 Test.....	41
3.9 Amélioration de la solution	47
Conclusion générale	53

Résumé

Toutes les entreprises sont conscientes de l'importance de leurs données et ils font leurs moyens pour qu'ils soient disponibles à tout moment. La haute disponibilité de stockage définit les moyens techniques mis en œuvre pour garantir peu ou pas de perte de données en cas d'incident. Ce niveau de disponibilité des données est assuré par la redondance des données sur plusieurs systèmes physiques différents, et en s'assurant que chaque écriture est effectivement effectuée sur plusieurs d'entre eux. Dans la réplication synchrone entre systèmes, les écritures sont suspendues jusqu'à ce qu'elles puissent être validées de manière fiable sur au moins deux systèmes.

Notre problème était donc de trouver une meilleure solution pour garder nos données hautement disponibles afin de couvrir les aspects critiques du processus de reprise après sinistre, en minimisant la perte de données et les interruptions d'activité.

L'objectif global de toute structure est d'assurer la continuité de ses activités dans de bonnes conditions. Cette continuité ne peut être assurée que par la mise en place d'une solution de stockage hautement disponible. Une solution de stockage haute disponibilité doit vous permettre de protéger vos données et documents numériques au niveau de l'entreprise.

Les principales fonctionnalités de notre solution incluent : le clustering la réplication, la redondance et le RAID5.

Abstract

All companies are aware of the importance of their data and they do their best to make it available at all times. Storage high availability defines the technical means implemented to guarantee little or no loss of data in the event of an incident. This level of data availability is provided by redundant data across multiple different physical systems, and ensuring that every write is actually performed across multiple of them. In synchronous replication between systems, writes are suspended until they can be reliably committed to at least two systems.

So, our problem was to find a better solution to keep our data highly available to cover the critical aspects of the disaster recovery process, minimizing data loss and business interruption.

The overall objective of any structure is to ensure the continuity of its activities under good conditions. This continuity can only be ensured by setting up a highly available storage solution. A high-availability storage solution should allow you to protect your digital data and documents at the enterprise level.

The main features of our solution include: clustering replication, redundancy and RAID5.

Introduction

Au cours des dix dernières années, le stockage des données est devenu l'une des ressources les plus importantes pour les grandes entreprises. Peu de domaines ont échappé à cette évolution au point que la cohésion de nos entreprises repose sur la disponibilité du stockage qui rythme notre activité de tous les jours : Pour éviter à ce problème de disponibilité, c'est là où les solutions des clusters et autres viennent se greffer à des nouvelles solutions pour assurer une haute disponibilité de stockage.

Dès lors, nous pouvons nous poser les questions de savoir : comment mettre en place une architecture de stockage de haute disponibilité ? quels sont les différents avantages de la disponibilité de ce service ? ces différentes interrogations nous apporteront les problèmes thématiques tout au long de notre sujet.

L'objectif général de notre recherche est de contrôler la disponibilité de stockage. Pour éviter de rencontrer des données inutilisables, Pour ce faire notre objectif spécifique repose sur le choix d'une solution de cluster.

Si nous avons choisi ce sujet c'est pour faire face aux problèmes que les entreprises rencontrent au quotidien suite à la non disponibilité des données.

Vue la nécessité pour une entreprise d'accéder à une donnée fiable en temps réel, nous supposons que le clustering et l'architecture de stockage SAN pourraient être une solution pour la disponibilité des données au sein d'une entreprise.

Afin de mener à bien nos objectifs, nous avons dû faire des recherches documentaires en lisant des livres, en lisant des mémoires et en cherchant sur Internet.

Pour effectuer cette tâche, notre recherche se divise en trois grandes parties. Dans un premier temps, nous nous intéresserons aux cadres théoriques et méthodologiques dans la première partie, puis nous aborderons les cadres conceptuels dans la deuxième partie. Enfin, la partie 3 implémente la solution de notre choix.

I. Les Cadres théoriques et méthodologique

Dans cette partie nous allons entamer le cadre théorique et méthodologique.

1.1. Cadre théorique

Dans ce chapitre nous traiterons du cadre théorique, c'est-à-dire nous parlerons de la problématique, des objectifs de recherche et les hypothèses de recherche.

1.1.1. Problématique

Aujourd'hui, la durabilité des entreprises repose principalement sur ses données informatiques. Les données sont les ressources les plus précieuses pour toute entreprise, et la perte des données peuvent avoir de graves conséquences. Ces résultats peuvent avoir un impact significatif (direct ou indirect) sur l'activité de l'entreprise. Par conséquent, les sauvegarde, vise à minimiser l'impact associé à la perte de données informatiques pour prévenir les risques naturels, les erreurs humaines, les virus ou les catastrophes. Mais,

- Comment mettre en place une telle architecture ?
- Quels sont les différents avantages permettant de mettre en place une telle architecture ?

De ce fait nous nous sommes fixés des objectifs pour apporter une solution au problème soulevé.

1.1.2. Objectifs

Dans le cadre de notre travail, par suite des questions émises dans la problématique sur notre thème, nous aurons pour objectif général la mise en place d'une architecture de stockage pour la haute disponibilité.

Nos objectifs spécifiques sont :

- ✓ Etudier les architectures de stockage
- ✓ Étudier les différentes solutions de haute disponibilité

- ✓ Proposer une solution répondant aux attentes
- ✓ Mettre en place la solution choisie

Cependant, plusieurs hypothèses ont été élaborées au travers de ces objectifs de recherche.

1.1.3. Hypothèse de recherche

Après avoir mentionné les problèmes et les objectifs de notre travail, j'ai senti que nous devions parler des hypothèses qui seraient les éléments de réponse à la question ci-dessus.

Les avantages de la haute disponibilité de stockage sont en autres : une meilleure disponibilité des données, une exploitation informatique simplifiée, des risques de perte de données réduits et des gains financiers importants.

C'est dans cette logique que nous allons passer au cadre méthodologique.

1.2. Cadre méthodologique

Cette partie qui constitue le cadre méthodologique de ce mémoire, elle nous permettra de faire une délimitation du sujet dans son contexte, les techniques de recherche et les difficultés dont nous avons eu à faire face.

1.2.1. Délimitation du champ de l'étude

Afin d'avoir une compréhension claire du sujet et de faciliter notre approche académique, tout travail se voulant académique doit être examiné à la fois temporellement et spatialement. Par conséquent, notre tâche est avant tout de faire une enquête détaillée sur la disponibilité du stockage, mais c'est aussi de proposer une solution.

Le champ d'étude ainsi présenté nous allons maintenant exposer les instruments qui ont servi à collecter des informations pour l'élaboration de notre document.

1.2.2. Les techniques de recherche

La réalisation et la rédaction d'un bon mémoire passe nécessairement par la collecte des informations acquises grâce à la bonne documentation sur notre sujet en parcourant les éléments suivants :

- ✓ Mémoires
- ✓ Recherche en ligne

A partir de ces nombreuses techniques de recherche nous allons pouvoir bien cerner notre sujet.

Ainsi, ces différentes techniques de recherche ont fait ressortir toutes les difficultés liées à ce sujet.

1.2.3. Les difficultés rencontrées

Autant de difficultés qui peuvent entraver l'élaboration de cet article, comme la difficulté à trouver des informations précises à travers notre projet, la traduction en français car la plupart des excellents documents sur notre sujet sont en anglais, et.

Toutes ces nombreuses difficultés ne nous ont pas empêchés de mener à bien notre travail.

C'est dans cette optique, que nous allons dans la deuxième partie traiter le cadre conceptuel.

II. Cadre conceptuel

Le cadre conceptuel est composé de deux chapitres à savoir un rappel sur le réseau et une généralité sur le stockage.

2.3. Rappel sur le réseau

Au niveau de ce chapitre nous parlerons du modèle TCP/IP, modèle OSI et de l'environnement client-serveur.

2.3.1 Modèle TCP/IP et OSI

Dans cette section nous vous expliquerons les mécanismes de communication du modèle TCP/IP et celui de l'OSI.

2.3.1.1 Modèle TCP/IP

TCP/IP désigne communément une architecture réseau, mais cet acronyme désigne en fait 2 protocoles étroitement liés : un protocole de transport, TCP (Transmission Control Protocol) qu'on utilise « par-dessus » un protocole réseau, IP (Internet Protocol). Ce qu'on entend par « modèle TCPIP ». C'est en fait une architecture réseau en 4 couches dans laquelle les protocoles TCP et IP jouent un rôle prédominant, car ils en constituent l'implémentation la plus courante.

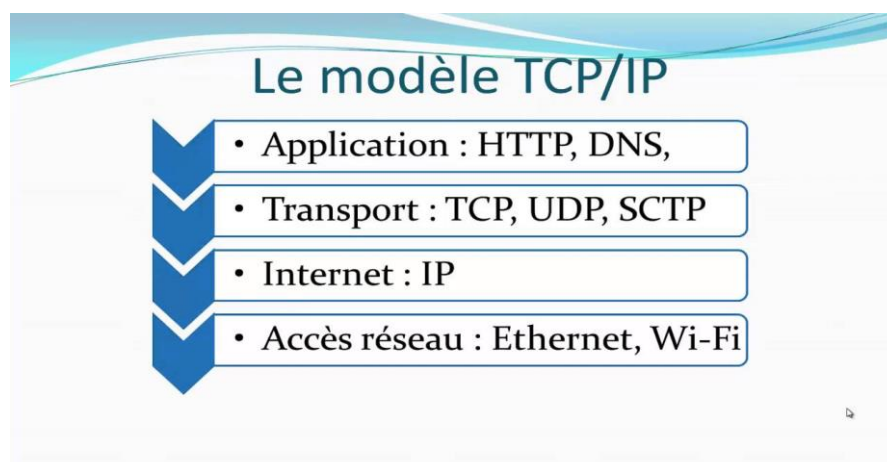


Figure 1: Modèle TCP/IP

Les couches du modèle TCP/IP

Il existe 4 couches dans ce model :

- ✓ Couche application : Elle est la couche de communication qui s'interface avec les utilisateurs. Cette couche contient tous les protocoles de haut niveau, comme par exemple Telnet, TFTP SMTP, HTTP. Le point important pour cette couche est le choix du protocole de transport à utiliser.
- ✓ Couche Transport : elle est responsable du dialogue entre les hôtes terminaux d'une communication. Cette couche n'a que deux implémentations : le protocole TCP (Transmission Control Protocol) et le protocole UDP (User Datagram Protocol).
- ✓ Couche Internet : Cette couche est la clé de voûte de l'architecture. Cette couche réalise l'interconnexion des réseaux distants sans connexion.

Son rôle est de permettre l'injection de paquets dans n'importe quel réseau et l'acheminement de ces paquets indépendamment les uns des autres jusqu'à destination. Comme aucune connexion n'est établie au préalable, les paquets peuvent arriver dans le désordre ; le contrôle de l'ordre de remise est éventuellement la tâche des couches supérieures.

- ✓ Couche accès réseau : elle organise le flux binaire et identifie physiquement les hôtes.

Au sens du modèle TCP/IP la couche Accès Réseau est vide, car la pile des protocoles Internet (TCP/IP) est censée "inter-opérer" avec les différentes technologies qui offrent un accès au réseau.

2.3.1.2 Modèle OSI

Le modèle OSI (Open Systems Interconnection Model) est un cadre conceptuel utilisé pour décrire les fonctions d'un système de mise en réseau.

Ce modèle définit une architecture hiérarchique qui répartit logiquement les fonctions nécessaires à la communication entre systèmes. Il y a au total sept couches qui sont responsables de tâches et de fonctionnalités spécifiques.

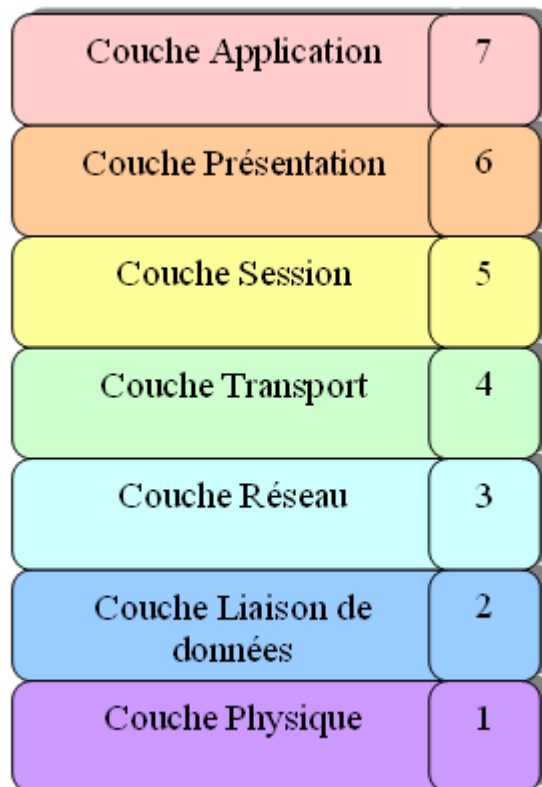


Figure 2:Modèle OSI

Les couches du modèle OSI

Comme on vient de l'énoncer au début. Le modèle OSI a 7 couches respectives :

- **Couche physique** : Elle est chargée de la transmission effective des signaux entre les interlocuteurs. Son service est limité à l'émission et la réception d'un bit ou d'un train de bits continu.
- **Couche liaison de données** : Elle gère les communications entre deux machines directement connectées entre elles, ou connectées à un équipement qui émule une connexion directe.
- **Couche réseau** : Elle gère les communications de proche en proche, généralement entre machines : routage et adressage des paquets.
- **Couche transport** : Elle gère les communications de bout en bout entre processus.

- **Couche session** : Elle gère la synchronisation des échanges et les transactions, permet l'ouverture et la fermeture de session.
- **Couche présentation** : Elle est chargée du codage des données applicatives, précisément de la conversion entre données manipulées au niveau applicatif et chaînes d'octets effectivement transmises.
- **Couche application** : Elle est le point d'accès aux services réseaux, elle n'a pas de service propre spécifique et entrant dans la portée de la norme.

2.3.2 L'environnement client/serveur

De nombreuses applications fonctionnent dans un environnement client/serveur. C'est-à-dire que les machines clientes (machines faisant partie du réseau) se connectent et servent généralement des serveurs dotés de capacités d'E / S très puissantes. Ces services sont des programmes qui stockent des données telles que l'heure, les fichiers et les connexions.

Les services sont exécutés par un programme s'exécutant sur la machine cliente, appelé programme client. Par conséquent, lors de la spécification d'un programme à exécuter sur une machine cliente, nous parlerons de clients capables de traiter les informations récupérées sur le serveur (client FTP, client de messagerie, etc.) (File pour le client FTP et Gmail pour le client de messagerie)

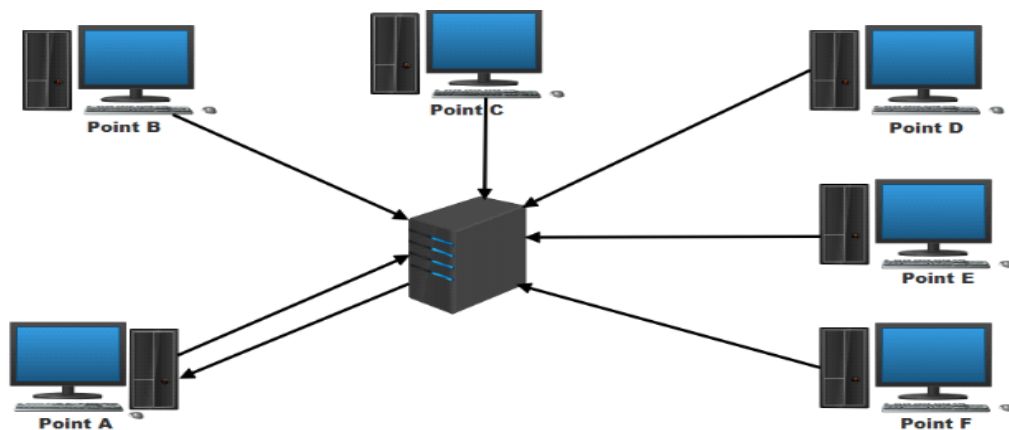


Figure 3:Environnement client/serveur

2.4 Généralité sur le stockage

2.4.1 Définition du stockage

Le stockage informatique est devenu au fil des années un enjeu crucial pour les entreprises.

Cette constatation est en grande partie due à la croissance exponentielle continue de la quantité de données informatiques que nous générons. Les experts estiment même que d'ici 2025, les humains auront besoin de stocker 163 zettaoctets de données... un nombre énorme ! (Source : Futura Sciences).

Les données, en revanche, ont une autre dimension pour les entreprises : avec la transformation numérique, elles sont considérées comme un actif commercial précieux et remboursable qui doit être protégé comme tout autre produit.

Par conclusion, le stockage est devenu un enjeu très important au niveau des entreprises.

Cependant des nombreuses architectures ont été élaborés pour avoir un stockage meilleur.



Figure 4: centre des données

2.4.2 Les différentes architectures de stockage

Il existe trois principales architectures de stockage :

2.4.2.1 Le DAS

DAS (Direct Attached Storage), également connu sous le nom de Direct Attached Disk, est une autre solution que les entreprises envisagent. Il s'agit d'un périphérique de stockage connecté directement à l'ordinateur. Les autres ordinateurs ne peuvent pas y accéder. Par exemple, le disque dur interne d'un ordinateur ou d'un serveur représente le stockage DAS.

Au sein d'une entreprise, les disques connectés aux serveurs sont considérés comme appartenant au domaine de stockage à connexion directe. Il est largement admis que le DAS parvient à fournir de meilleures performances que le NAS et le SAN car il n'est pas nécessaire de traverser le réseau pour effectuer le transfert de données. Par conséquent, certaines applications prennent en charge le stockage DAS.

Les DAS peuvent se connecter aux PC et aux serveurs via divers protocoles, notamment :

- ✓ FTP : (File Transfer Protocol) ;
- ✓ SMB : (Server Message Block) ;
- ✓ CIFS : (Common Internet File System) ;
- ✓ NFS : (Network File System) ;
- ✓ NCP : (Netware Core Protocol).

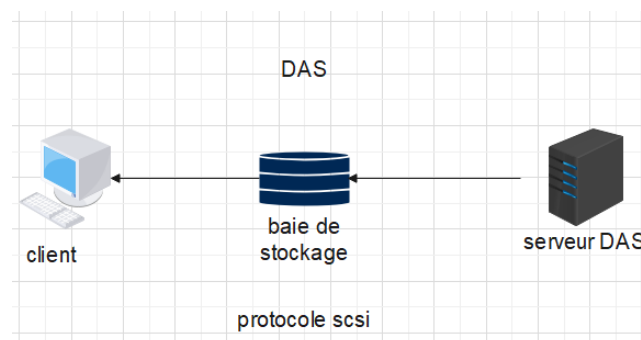


Figure 5:Architecture d'un DAS

Avantages

- ✓ Il peut fournir des performances meilleures que le stockage SAN ou NAS, car les données n'ont pas à être transmises via un réseau.
- ✓ Les supports amovibles peuvent être externalisés (il s'agit de mettre les sauvegardes à l'abri en dehors de l'entreprise). Si le lieu de production est très endommagé, les sauvegardes ne seront pas détruites. Cependant, le coût du lieu de stockage est à prendre en compte.
- ✓ Le coût de l'investissement est abordable à toutes PME, quelle que soit sa taille

Inconvénients

- ✓ Les périphériques de stockages sont gérés indépendamment les uns des autres, ce qui complique l'administration et la gestion du parc.
- ✓ Le partage de ressources, s'il est possible, impliquera une charge supplémentaire sur le réseau en place.
- ✓ Si les opérations de sauvegardes sont faites indépendamment sur chacun des serveurs cela alourdira encore l'administration de l'infrastructure.
- ✓ Distance entre système et périphériques
- ✓ Les applications telles que les bases de données doivent être fermées avant de lancer la sauvegarde. Pendant ce temps, il est impossible d'y effectuer toute modification. Cependant certains logiciels permettent de sauvegarder « à chaud » (c'est-à-dire en cours de fonctionnement) certaines bases de données.
- ✓ Les supports amovibles sont sensibles à l'environnement (électricité, température, humidité...). En particulier, les supports optiques sont facilement rayés et deviennent illisibles.
- ✓ La vitesse de transfert que proposent les lecteurs de bandes et de DVD est relativement faible. Il faut prévoir en moyenne entre deux à cinq heures pour effectuer

En raison de la tendance aux connexions de données et du besoin croissant d'un accès plus rapide aux données, deux infrastructures de stockage, NAS et SAN, ont été développées.

2.4.2.2 Le NAS

Le stockage en réseau (NAS) est une architecture de stockage basée sur des fichiers qui facilite l'accès des périphériques réseau aux données stockées.

Il est en quelque sorte un serveur spécialement configuré pour la gestion d'un espace de stockage à destination de clients hétérogènes (Multi-protocoles). Ce type de service vise à mettre à disposition de l'espace de stockage permettant de déposer des fichiers de tout type.

IL fournit un point d'accès de stockage unique au réseau qui combine la sécurité, la gestion et la tolérance aux pannes. Il peut être configuré comme une option de stockage compatible avec les conteneurs, dans laquelle le système de stockage est exposé à un ou plusieurs conteneurs. Les conteneurs désignent une technologie hautement flexible qui apporte un niveau d'évolutivité considérable à la distribution des applications et des ressources de stockage.

Comme les serveurs de fichiers, le réseau de stockage NAS est servi sur un réseau IP en utilisant un ou plusieurs des protocoles suivants :

- ✓ Common Internet File System (CIFS), également connu sous le nom de Server Message Block (SMB) ;
- ✓ Network File System (NFS) ;
- ✓ Apple Filing Protocol (AFP).

Parfois, les fichiers sont disponibles via le protocole de transfert de fichiers (FTP), WebDAV et les gestionnaires de fichiers réseau. Généralement, le NAS est configuré via l'interface Web.

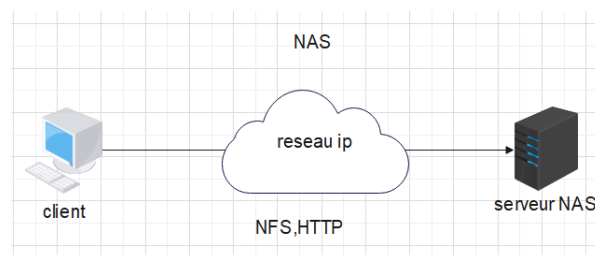


Figure 6:architecture du NAS

Avantages

- ✓ Évolutivité : Pour augmenter la capacité de stockage de votre NAS, ajoutez simplement des disques durs. Pas besoin de mettre à niveau ou de remplacer les serveurs, et encore moins de désactiver la mise en réseau.
- ✓ Performances : le NAS étant un système de fichiers, les autres périphériques du réseau ne sont pas responsables du partage de fichiers. De plus, le NAS est souvent configuré à des fins très spécifiques (comme le stockage de données volumineuses ou de contenu multimédia), ce qui permet d'obtenir de meilleures performances.
- ✓ Configuration simple : les architectures NAS sont souvent livrées avec des scripts simples ou un système d'exploitation simplifié préinstallé, ce qui réduit considérablement le temps de configuration et d'administration du système.
- ✓ Accessibilité : Tous les appareils en réseau peuvent accéder au NAS.
- ✓ Tolérance aux pannes : le NAS peut être formaté pour prendre en charge les disques répliqués, les systèmes RAID ou le codage d'effacement pour garantir l'intégrité des données.

Inconvénients

- ✓ La mise en place d'un système de stockage NAS domestique n'est pas bon marché. Un NAS solide qui peut gérer un grand nombre de fichiers multimédias différents vous coûtera au moins 500 \$, et selon la taille du disque dur que vous choisirez, vous pouvez encourir entre 50 \$ et 200 \$ supplémentaires.
- ✓ Cela rend le coût initial de configuration d'un lecteur NAS relativement élevé. La bonne nouvelle est que vous économiserez les coûts de votre fournisseur de sauvegarde à long terme. (Ces coûts sont tout aussi importants, sinon plus chers, à long terme.)

2.4.2.3 Le SAN

Un réseau de zone de stockage (SAN) est un réseau haut débit dédié et indépendant qui interconnecte plusieurs serveurs et leur offre des pools partagés de périphériques de stockage. Chaque serveur peut accéder à un stockage partagé comme s'il s'agissait d'un lecteur y étant directement rattaché. Il est l'architecture la plus couramment utilisée par les entreprises pour les applications critiques qui nécessitent un débit élevé et une faible latence.

Il a été conçu au départ avec un réseau d'accès spécifique au serveur de stockage, ceci afin de permettre l'accès à des vitesses proche de celui d'un DAS. Les SAN ont également profité des

évolutions technologiques des réseaux et peuvent aujourd'hui, sous certaines conditions, utiliser des réseaux non spécialisés.

Un SAN utilise généralement l'un des quatre protocoles suivants :

- ✓ FCP (Fibre Channel Protocol)
- ✓ iSCSI (Internet Small Computer System Interface)
- ✓ FCoE (Fibre Channel over Ethernet)
- ✓ FC-NVMe (Non-Volatile Memory Express over Fibre Channel)

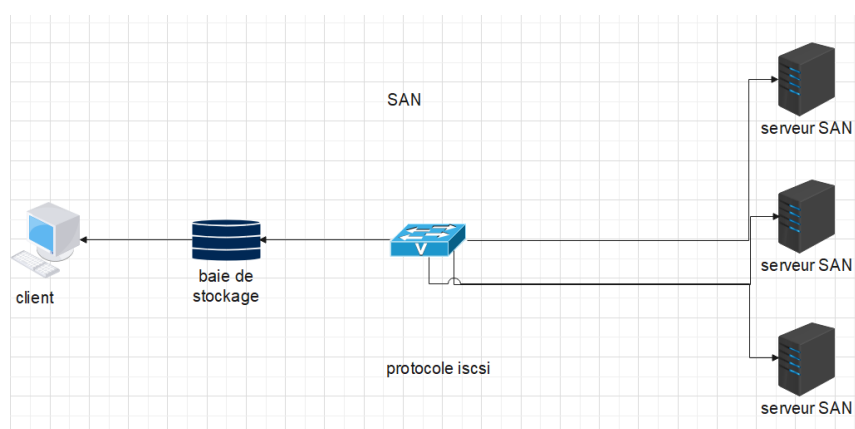


Figure 7:architecture d'un SAN

Avantages

Un réseau de stockage SAN peut offrir les avantages suivants :

- **Amélioration de la disponibilité des applications :** Le stockage existe indépendamment des applications, et il est accessible par différents canaux, ce qui améliore la fiabilité, la disponibilité et la facilité d'entretien.
- **Meilleures performances des applications :** Les réseaux de stockage SAN déchargent et déplacent le traitement du stockage des serveurs vers des réseaux distincts.
- **Centralisés et consolidés :** Les réseaux de stockage SAN permettent de simplifier la gestion, l'évolutivité, la flexibilité et la disponibilité.
- **Transfert et mise en lieu sûr des données sur un site distant :** Ils protègent les données contre les sinistres et les attaques malveillantes grâce à une copie à distance.
- **Gestion centralisée simple :** Les réseaux de stockage SAN simplifient la gestion en créant des images uniques des supports de stockage.

Inconvénients

- En contrepartie, le coût d'acquisition d'un SAN est beaucoup plus onéreux qu'un dispositif NAS dans la mesure où il s'agit d'une architecture complète (switch, baie, câbles, ...)
- **Performances variables** : La performance de l'accès aux disques par un hôte varie en fonction des sollicitations de la part des autres hôtes sur les ressources partagées, elles peuvent donc varier dans le temps.
- Un grand nombre de produits et d'environnements souvent hétérogènes, une vérification de la compatibilité sera souvent nécessaire

2.4.3 Comparaison

Voici un petit tableau comparatif de nos solutions :

	DAS	NAS	SAN
Accès	Mode bloc	Mode fichier	Mode bloc
Connexion	Série -SAS Parallèle -SCSI	Ethernet	Fiber Channel
Performances d'accès	Très bonnes	Moins bonnes	Très bonnes
Limite des Performance	Sous-système SCSI du noyau	Système de fichiers NFS/CIFS	Commutation Fiber Channel
Augmentation de Capacité	Arrêt du système obligatoire	Très facile	Complexe suivant l'architecture
Evolutivité et Continuité d'exploitation	Faible	Moyenne	Élevée

Tableau 1: comparatif des solutions de stockage

	NAS	SAN
Fonction principale	Serveur spécialité qui sert les fichiers et les données stockées aux postes clients et aux autres serveurs à travers le réseau	Le stockage est accessible à travers un réseau qui lui est spécialement dédié. Sa principale fonction est de fournir aux serveurs un stockage consolidé basé sur le Fiber Channel
Applications bien adaptées	Ideal pour servir les fichiers	Ideal pour les bases de données et le traitement des transactions en ligne
Transfert des données	A travers un LAN ou un WAN	A travers le SAN vers un serveur vers un LAN ou un WAN
Ressources de stockage et de sauvegardes	Les sauvegardes peuvent être attachées directement à des Appliance NAS intermédiaire ou être distribuées et attachées à un LAN ou un WAN	Les ressources de stockage et de sauvegarde peuvent être attachées directement au serveur ou à travers une structure Fiber Channel
Disponibilités	Des alimentations et des ventilateurs redondants sont couramment utilisés	Des composants matériels et logiciels redondant donnent au système une haute disponibilité. Le système peut être configuré sans le moindre point de panne
Scalabilité	Plusieurs serveur NAS peuvent être ajoutés au réseau et du stockage peut être ajouté aux serveurs NAS intermédiaire	Le stockage peut être étendu par l'ajout de switches Fiber Channel et des dispositifs de stockage

Tableau 2: Comparatif de SAN et NAS

2.4.4 Présentation de la solution

La perte de données sensibles peut avoir des conséquences dramatiques pour une entreprise et menacer sa survie à court ou moyen terme. Par conséquent, les entreprises de toutes tailles sont intéressées à rendre ces données disponibles à tout moment. Dans ce sujet il s'agit de mettre en place une architecture de stockage de haute disponibilité. Cette solution met en place le stockage NAS pour qu'on arrive à créer un cluster de basculement en cas de fail d'un serveur

2.5 Rappel sur la haute disponibilité

2.5.1 Présentation de la haute disponibilité

La disponibilité est aujourd'hui un enjeu important des infrastructures informatiques. Une étude de 2007 estime que la non-disponibilité des services informatiques peut avoir un coût de 440 000 euros de l'heure, ces coûts se chiffrant en milliards d'euros à l'échelle d'un pays. L'indisponibilité des services informatiques est particulièrement critique dans le domaine de l'industrie, notamment en cas d'arrêt d'une chaîne de production

Elle est un modèle de système architectural qui tente de protéger le système contre certaines conditions de défaillance. Cela signifie que même si certaines parties du système échouent, l'ensemble du système reste accessible et utilisable. Les solutions de haute disponibilité protègent souvent contre des conditions spécifiques telles que les pannes de serveur, les pannes de composant unique, les pannes de dépendance, les augmentations de charge variables et les divisions de réseau où les composants du système deviennent inaccessibles sur le réseau.

Deux moyens complémentaires sont utilisés pour améliorer la disponibilité :

La mise en place d'une infrastructure matérielle spécialisée, généralement en se basant sur de la redondance matérielle. Est alors créé un cluster de haute-disponibilité (par opposition à un cluster de calcul) : une grappe d'ordinateurs dont le but est d'assurer un service en évitant au maximum les indisponibilités ;

La mise en place de processus adaptés permettant de réduire les erreurs, et d'accélérer la reprise en cas d'erreur. ITIL contient de nombreux processus de ce type.

2.5.2 Les technologies des RAIDS

Le terme RAID (Redundant Array of Independent/Inexpensive Disks) désigne un groupe de disques indépendants qui est vu comme une seule unité logique. Cette technologie permet notamment de cumuler les performances de plusieurs disques en répartissant les données sur l'ensemble et en accédant simultanément à tous les disques, de manière synchronisée ou non, selon les cas. Elle permet aussi d'assurer la sécurisation des données, soit par copie, soit par calcul de parité ; cette dernière méthode nécessitant, pour être réellement efficace, l'utilisation d'un ou plusieurs contrôleurs RAID matériel dédiés.

La technologie RAID (Redundant Array of Inexpensive Disks, parfois Redundant Array of Independent Disks) permet de constituer une unité de stockage à partir de plusieurs disques durs. L'unité ainsi créée (appelée grappe) a donc une grande tolérance aux pannes (haute disponibilité), ou bien une plus grande capacité/vitesse d'écriture. La répartition des données

sur plusieurs disques durs permet donc d'en augmenter la sécurité et de fiabiliser les services associés.

Cette technologie a été mise au point en 1987 par trois chercheurs (Patterson, Gibson et Katz) à l'Université de Californie (Berkeley). Depuis 1992 c'est le RAID Advisory Board qui gère ces spécifications.

Elle consiste à constituer un disque de grosse capacité (donc coûteux) à l'aide de plus petits disques peu onéreux (c'est-à-dire dont le MTBF, Mean Time Between Failure, soit le temps moyen entre deux pannes, est faible). Chacun de ces niveaux constitue un mode d'utilisation de la grappe, en fonction : des performances ; du coût ; des accès disques. En jouant sur la mémoire tampon des disques, les contrôleurs RAID et les algorithmes, les constructeurs affinent les performances pour les adapter à différents types d'applications. Il existe plusieurs types de RAID, qui ont des avantages et leurs inconvénients bien précis. Dans les grandes lignes, on peut classer les différentes techniques de RAID dans deux catégories : les **RAID standards** et les **RAID combinés**.

2.5.2.1 Les technologies RAID standard

La technologie RAID dite standard a été inventée pour la première fois. Ce sont les technologies JBOD, RAID 0, RAID 1, RAID2, RAID3, RAID 4, RAID5 et RAID6.

Technologie JBOD

Permet de combiner plusieurs disques durs en une seule partition. JBOD signifie juste un tas de disques. Avec cette technique, on attend que le disque dur soit plein pour lancer le suivant. La capacité totale du disque dur s'additionne. Il est tout à fait possible d'utiliser des disques durs de tailles différentes.

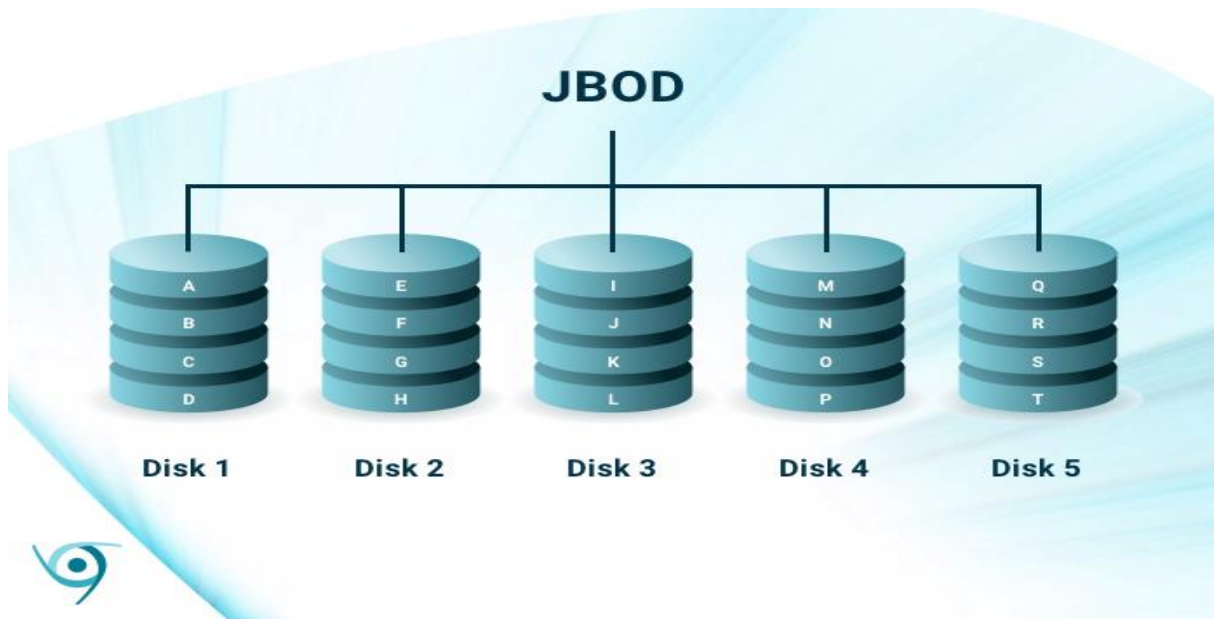


Figure 8: technologies JBOD

Technologie RAID0

Il est également connu sous le nom d' « entrelacement de disques » ou de « volume agrégé par bandes » (*striping* en anglais), est une configuration RAID permettant d'augmenter significativement les performances de la grappe. Les données sont simplement réparties sur plusieurs disques sans aucune protection des données. Plus il y a de disques, plus la capacité totale est importante et meilleures sont les performances, mais en cas de panne d'un seul disque, tout est perdu. De cette façon, il n'y a pas de redondance, on ne peut donc pas parler de tolérance aux pannes. En effet en cas de défaillance de l'un des disques, l'intégralité des données réparties sur les disques sera perdue. Toutefois, étant donné que chaque disque de la grappe a son propre contrôleur, cela constitue une solution offrant une vitesse de transfert élevée. Le RAID 0 consiste ainsi en la juxtaposition logique (agrégation) de plusieurs disques durs physiques. En mode RAID-0 les données sont écrites par "bandes" (en anglais stripes). Il est recommandé d'utiliser des disques de même taille pour faire du RAID-0 car dans le cas contraire le disque de plus grande capacité ne sera pas pleinement exploité

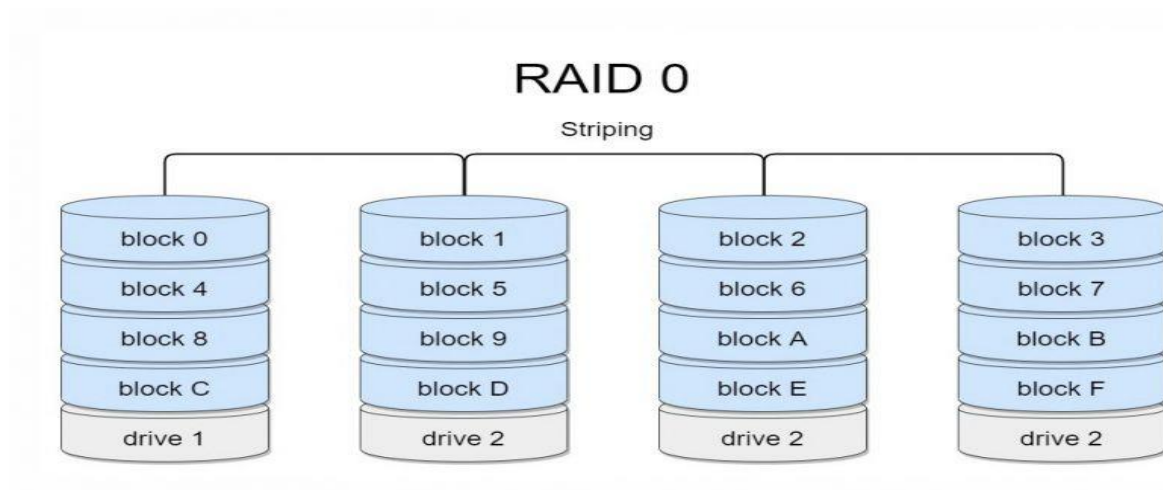


Figure 9: technologies RAID0

Technologie RAID1

Elle est Aussi appelée mode « miroir ». Les données ne sont pas réparties mais la protection est assurée par duplication des données sur deux disques distincts. Le stockage utile est donc divisé par deux. Les performances en lecture sont améliorées grâce à l'accès simultané aux disques en miroirs. Les performances en écriture sont identiques à un disque unique. En cas de panne d'un disque, son miroir est utilisé et les données restent sans protection jusqu'au remplacement et à la reconstruction complète du disque défectueux. On obtient ainsi une plus grande sécurité des données, car si l'un des disques tombe en panne, les données sont sauvegardées sur l'autre. D'autre part, la lecture peut être beaucoup plus rapide lorsque les deux disques sont en fonctionnement. Enfin, étant donné que chaque disque possède son propre contrôleur, le serveur peut continuer à fonctionner même lorsque l'un des disques tombe en panne, au même titre qu'un camion pourra continuer à rouler si un de ses pneus crève, car il en a plusieurs sur chaque essai. En contrepartie la technologie RAID1 est très onéreuse étant donné que seule la moitié de la capacité de stockage n'est effectivement utilisée

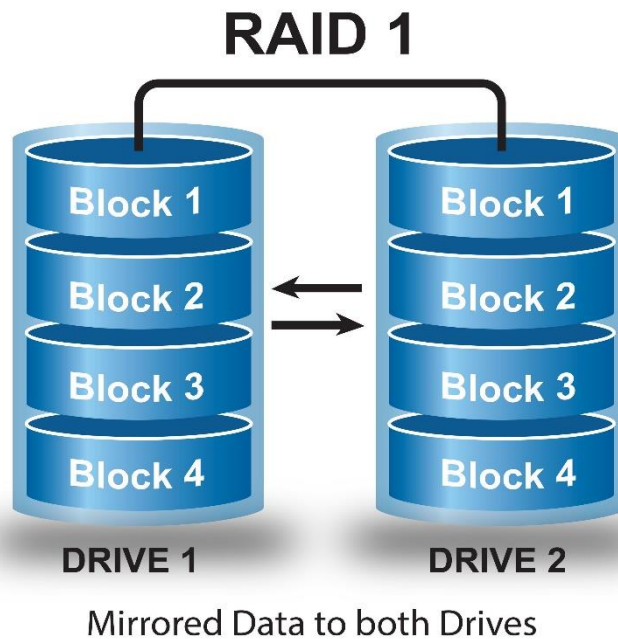


Figure 10:technologie RAID1

Les RAID 2, 3, 4, 5 et 6 utilisent des bits de parité. Cela leur permet d'obtenir une résilience aux pannes plus attrayante que RAID1. Ces technologies ont tendance à utiliser un seul disque dur pour stocker des données redondantes, nettement moins que RAID1. D'autre part, la récupération après une panne est plus lente car les données d'origine doivent être reconstruites par des calculs impliquant toutes les données du disque dur ainsi que les données de parité.

Technologie RAID2,3 et 4

Les données sont réparties sur n disques dont un est dédié aux calculs de parité. Le stockage utile est donc n-1 disques.

Le RAID 2 est aujourd'hui obsolète. Il combine la méthode du volume agrégé par bande (*striping* en anglais) à l'écriture d'un code de contrôle d'erreur par code de Hamming (code ECC) sur un disque dur distinct. Cette technologie offre un bon niveau de sécurité, mais de mauvaises performances.

En RAID 3, les calculs de distribution et de parité se font au niveau octet, ce qui est bien pour limiter les requêtes pour les gros fichiers, tandis qu'en RAID 4, le système gère au niveau bloc

(plusieurs secteurs), ce qui est bien pour les petits fichiers. Demandes de fichiers. En cas de panne d'un disque, il y a une légère baisse des performances même si le système est toujours opérationnel. Pour retrouver des performances et des niveaux de sécurité optimaux, les utilisateurs doivent attendre le remplacement et la reconstruction complète du disque défectueux, ce qui est plus long qu'une simple copie en raison des calculs de parité. Si le deuxième disque tombe en panne avant la fin du processus, toutes les données seront perdues.

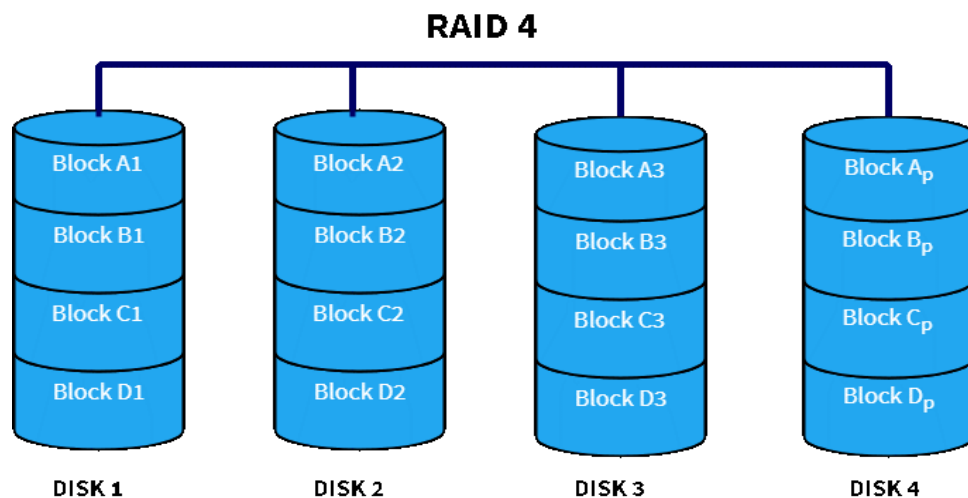


Figure 11:technologie RAID4

Technologie RAID5

Les données sont réparties sur n disques, de même que les blocs de parité, en changeant de disque à chaque bloc. Le stockage utile est de n-1 disques. Cette répartition est censée permettre d'améliorer encore les performances en fonctionnement normal, en autorisant notamment un accès désynchronisé aux disques, ce qui favorise les requêtes nombreuses, aussi bien en lecture qu'en écriture. Mais la défaillance d'un disque et la reconstruction qui s'en suit entraîne une dégradation sensible des performances pendant un long moment. Le mode RAID-5 permet d'obtenir des performances très proches de celles obtenues en RAID-0, tout en assurant une tolérance aux pannes élevée, c'est la raison pour laquelle c'est un des modes RAID les plus intéressants en termes de performance et de fiabilité

La capacité d'une batterie RAID 5 est la taille du plus petit disque multipliée par le nombre de disques dans la batterie moins un.

Ainsi, quatre disques durs de 120 Go chacun regroupés au sein d'une batterie RAID 5 seront considérés comme un disque dur de 360 Go par le système d'exploitation.

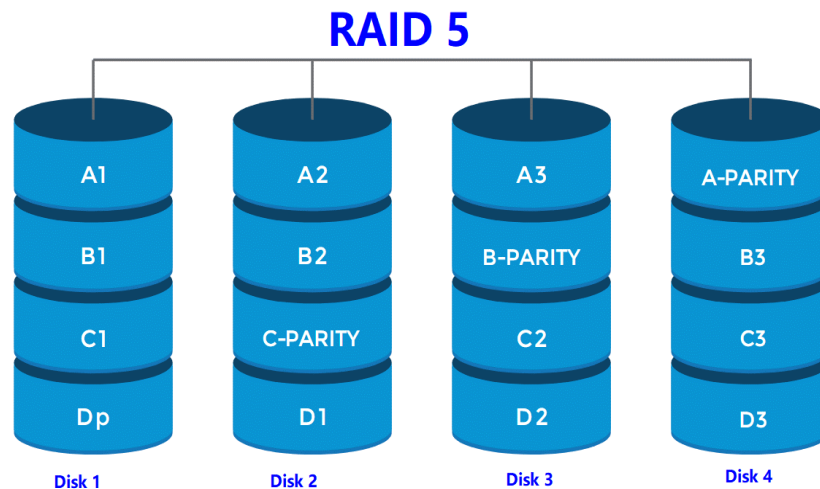


Figure 12:technologie RAID5

Technologie RAID6

C'est une amélioration du RAID 5 où les données de parité sont elles-mêmes dupliquées en plusieurs exemplaires. Cela permet de résister à la défaillance de plus d'un HDD (Hard Disk Drive). Le RAID 6 est une évolution du RAID 5 qui accroît la sécurité en utilisant « *n* » informations redondantes au lieu d'une. Il peut donc résister à la défaillance de « *n* » disques. Les défauts majeurs sont :

- ✓ Les temps d'écriture sont allongés à cause des calculs de redondance complexes.
- ✓ Le temps de reconstruction en cas de défaillance simultanée de 2 disques est nettement allongé.

Le RAID 6 était peu utilisé du fait de son surcoût. La récente envolée des capacités des disques ainsi que la vulgarisation de solution professionnelle à base de disque SATA a montré un intérêt nouveau dans l'utilisation du RAID 6, que ce soit par le biais de contrôleur Raid Hardware ou via du raid logiciel (Linux-2.6 intègre le RAID 6).

La capacité utile totale C_t , pour un système avec **k disques** dont **n** réservés pour la redondance est de $C_t = (k - n) \times c$ (c = capacité du plus petit des disques dur).

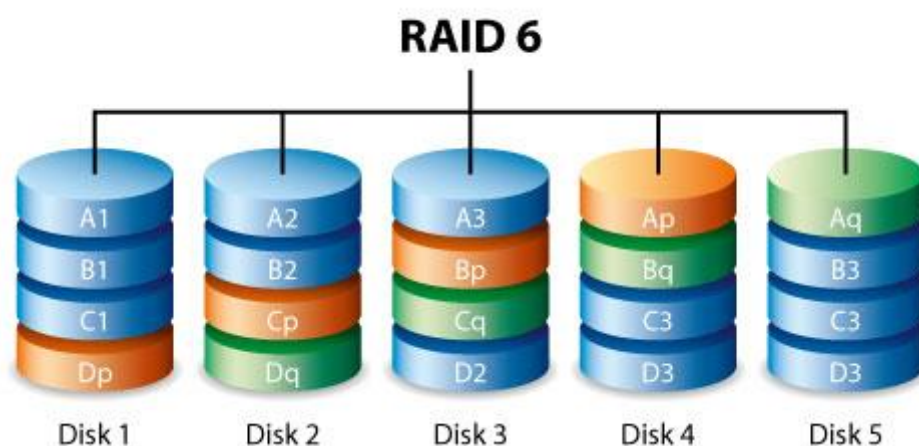


Figure 13:Technologie RAID6

Comparaisons des RAIDS Standards

Comparaison des niveaux RAID

Fonctions	RAID 0	RAID 1	RAID 5	RAID 6
N° mini de disques	2	2	3	4
Protection des données	Aucune	1 seul HD en panne	1 seul HD en panne	2 HD en panne
Performance de Lecture	Haute	Haute	Haute	Haute
Performance d'écriture	Haute	Moyenne	Faible	Faible
Capacité utile	100%	50%	67% - 94%	50% to 88%
Applications Types	Production Vidéo et Edition	OS, BDD transactionnelles, Serveur de fichiers, Serveur WEB	Entreposage de données, Serveur de Base de Données, NVR, DVR	Archivage de données, solutions à haute disponibilité

Figure 14:Comparaison RAID Standards

2.5.2.2 Les technologies RAID combiné

Fondamentalement, un niveau de RAID combiné est l'utilisation d'un concept de RAID classique sur des éléments constitutifs qui sont eux-mêmes le résultat d'un concept RAID classique. Le concept utilisé peut être le même ou différent.

La syntaxe est encore un peu floue mais on peut généralement considérer que le premier chiffre indique le niveau de raid des "grappes" et que le second indique le niveau de raid global. Dans l'absolu rien n'empêche d'imaginer des RAID combinés à 3 étages ou plus mais cela reste pour l'instant plus du domaine de la théorie et de l'expérimentation.

Les disques durs peuvent être combinés à l'aide de diverses techniques RAID. Par exemple, on peut utiliser RAID1 de disques en RAID 0 : on dit alors RAID 01. Il est également possible d'utiliser le RAID 0 des disques en RAID 1 : on dit alors RAID 10. De même, il est possible d'utiliser RAID 0 RAID 3, 4, 5 ou 6 : on parle respectivement de RAID 30, 40, 50, 60.

Le nombre important (et croissant) de permutations possibles signifie qu'il existe des tonnes de raids combinés, et nous ne les énumérerons pas tous. Cependant, nous pouvons couvrir les avantages et les inconvénients les plus courants. Pour les calculs suivants, on utilise les variables suivantes :

- G : nombre de disques par grappe ;
- N : nombre de grappes ;
- C : capacité d'un disque (tous les disques sont supposés identiques) ;
- V : vitesse d'un disque.

Le RAID 01 (ou RAID 0+1)

Il permet d'obtenir du *mirroring* rapide puisqu'il est basé sur des grappes en *striping* (RAID 0) combinées en miroir (RAID 1). Chaque grappe contenant au minimum deux éléments, et un minimum de deux grappes étant nécessaire, il faut au minimum quatre unités de stockage pour créer un volume RAID 0+1.

La fiabilité est moyenne, car un disque défectueux entraîne le défaut de toute la grappe qui le contient. Par ailleurs, cela allonge nettement le temps de reconstruction et dégrade les performances pendant la reconstruction, puisque tous les disques sont sollicités.

L'intérêt principal est que dans le cas d'un miroir à trois grappes ou plus, le retrait volontaire d'une grappe entière permet d'avoir une sauvegarde instantanée sans perdre la redondance.

Capacité totale : $C_t = G \times C$

Vitesse maximale : $V_m = G \times V$

Seuil de mise en défaut : G disques

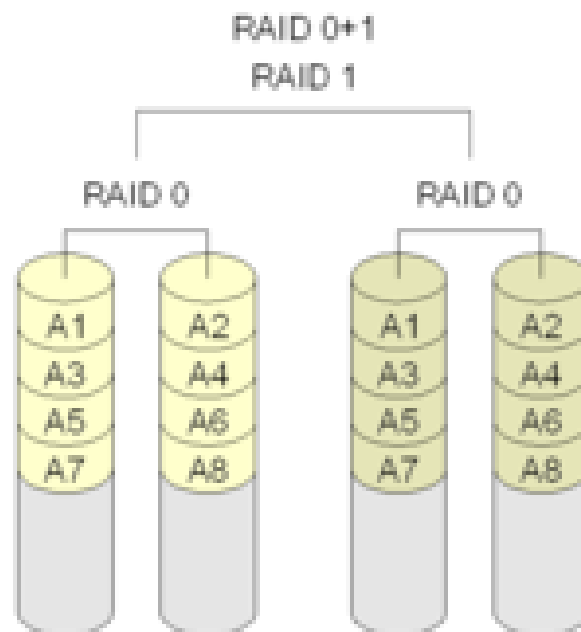


Figure 15: RAID 01

RAID 10 (ou RAID 1+0)

Il permet d'obtenir un volume agrégé par bande fiable (puisque'il est basé sur des grappes répliquées). Chaque grappe contenant au minimum 2 éléments et un minimum de 2 grappes étant nécessaire, il faut au minimum 4 unités de stockage pour créer un volume RAID10.

Sa fiabilité est assez grande puisqu'il faut que tous les éléments d'une grappe soient défectueux pour entraîner un défaut global. La reconstruction est assez performante puisqu'elle ne mobilise que les disques d'une seule grappe et non la totalité.

Capacité totale : $C_t = G \times C$

Vitesse maximale : $V_m = G \times V$

Seuil de mise en défaut : G disques

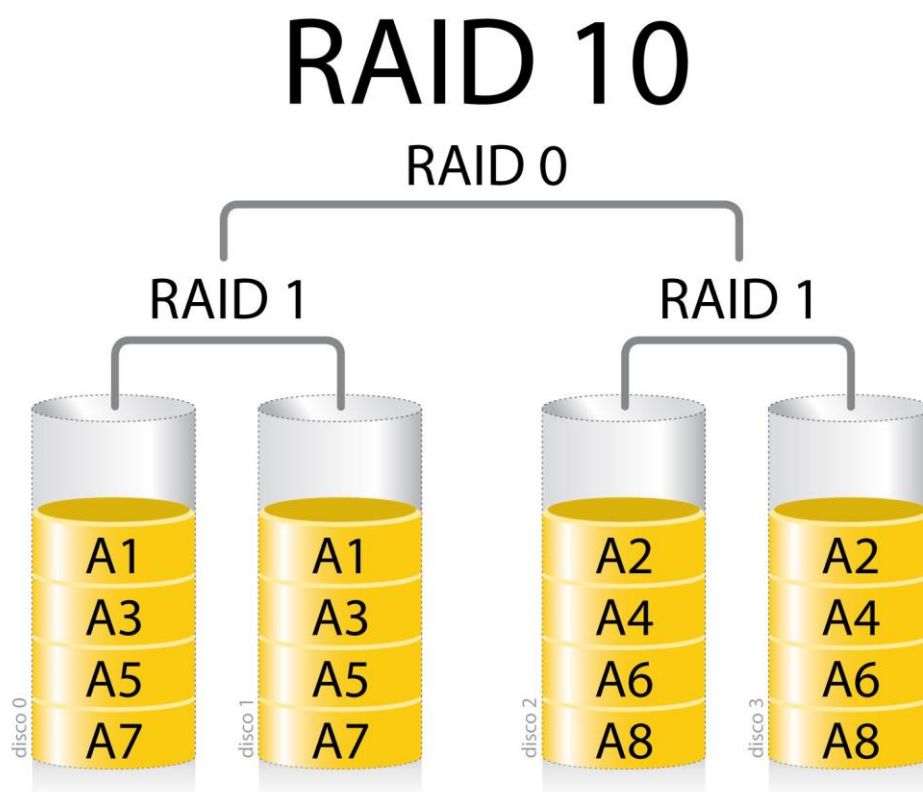


Figure 16: RAID10

RAID 05

Même principe que pour le raid 0+1 mais en employant du RAID5 pour la partie globale. Chaque grappe contenant au minimum 2 disques, et un minimum de 3 grappes étant nécessaire, il faut au minimum 6 unités de stockage pour créer un volume RAID05. Ce mode ne présente pas d'intérêt majeur par rapport à un RAID5 classique à $N * G$ disques. Il est donc très peu utilisé.

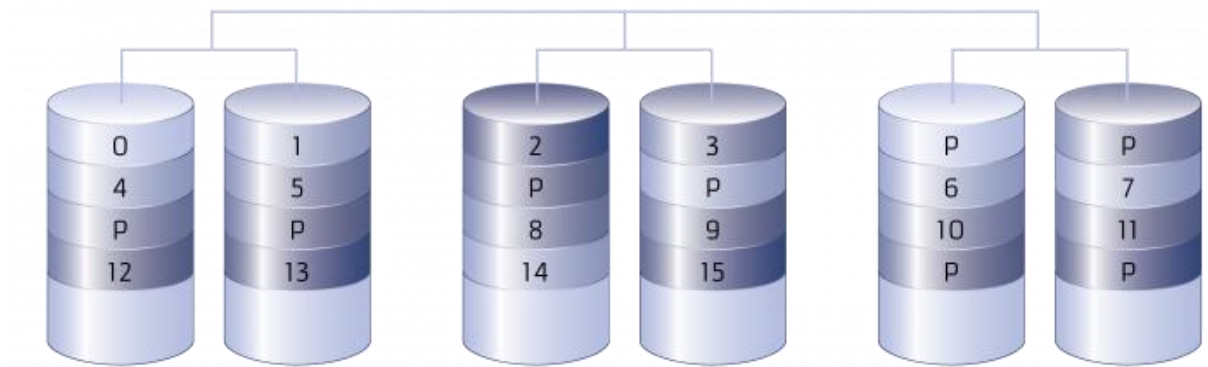


Figure 17: RAID 05

RAID 15

Il permet d'obtenir un volume agrégé par bandes avec redondance répartie très fiable (puisque'il est basé sur des grappes répliquées en miroir). Chaque grappe contenant au minimum 2 disques, et un minimum de 3 grappes étant nécessaire, il faut au minimum 6 unités de stockage pour créer un volume RAID15. Ce mode est très fiable puisque'il faut que tous les disques de 2 grappes différentes cessent de fonctionner pour le mettre en défaut. Ce mode est cependant coûteux par rapport à la capacité obtenue.

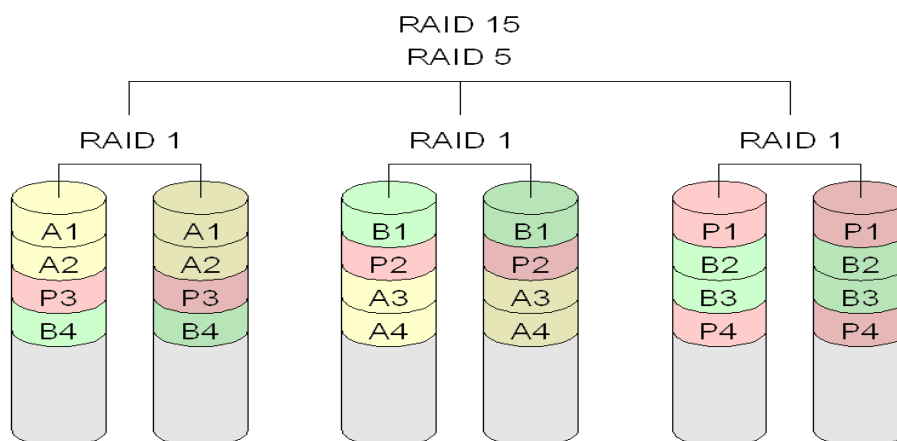


Figure 18: RAID 15

RAID 51

Il permet d'obtenir un volume répliqué basé sur des grappes en RAID5. Chaque grappe contenant au minimum 3 disques, et un minimum de 2 grappes étant nécessaire, il faut au

minimum 6 unités de stockage pour créer un volume RAID51. C'est un mode coûteux (faible capacité au regard du nombre de disques).

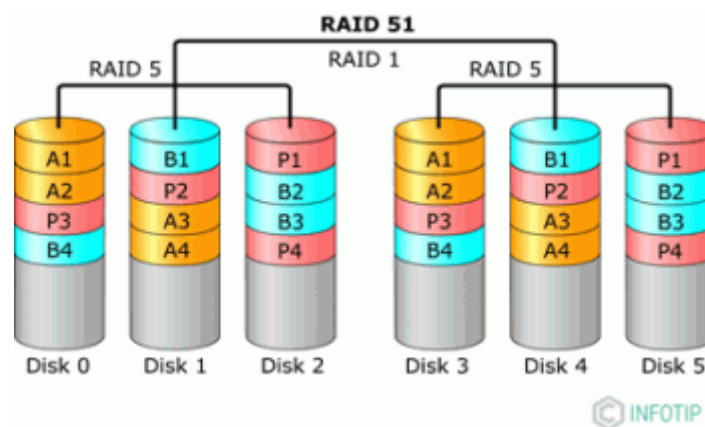


Figure 19: RAID 51

2.5.3 Comparaison

Les solutions RAID généralement retenues sont le RAID de niveau 1 et le RAID de niveau 5. Le choix d'une solution RAID est lié à trois critères : La sécurité : RAID 1 et 5 offrent tous les deux un niveau de sécurité élevé, toutefois la méthode de reconstruction des disques varie entre les deux solutions. En cas de panne du système, RAID 5 reconstruit le disque manquant à partir des informations stockées sur les autres disques, tandis que RAID 1 opère une copie disque à disque. Les performances : RAID 1 offre de meilleures performances que RAID 5 en lecture, mais souffre lors d'importantes opérations d'écriture. Le coût : le coût est directement lié à la capacité de stockage devant être mise en œuvre pour avoir une certaine capacité effective. La solution RAID 5 offre un volume utile représentant 80 à 90% du volume alloué (le reste servant évidemment au contrôle d'erreur). La solution RAID 1 n'offre par contre qu'un volume disponible représentant 50 % du volume total (étant donné que les informations sont dupliquées).

Fonctionnalités	RAID 0	RAID 1	RAID 1E	RAID 5	RAID 5EE	RAID 6	RAID 10	RAID 50	RAID 60
Nbre minimum de disques	2	2	3	3	4	4	4	6	8
Protection des données	Pas de protection	Panne d'un seul disque	Panne d'un seul disque	Panne d'un seul disque	Panne d'un seul disque	Panne de deux disques	Une panne de disque maxi dans chaque sous-pile	Une panne de disque maxi dans chaque sous-pile	Deux pannes de disque maxi dans chaque sous-pile
Performances en lecture	Elevées	Elevées	Elevées	Elevées	Elevées	Elevées	Elevées	Elevées	Elevées
Performances en écriture	Elevées	Moyennes	Moyennes	Faibles	Faibles	Faibles	Moyennes	Moyennes	Moyennes
Performances en lecture (mode dégradé)	S/O	Moyennes	Elevées	Faibles	Faibles	Faibles	Elevées	Moyennes	Moyennes
Performances en écriture (mode dégradé)	S/O	Elevées	Elevées	Faibles	Faibles	Faibles	Elevées	Moyennes	Faibles
Utilisation de la capacité	100 %	50 %	50 %	67 % - 94 %	50 % - 88 %	50 % - 88 %	50 %	67 % - 94 %	50 % - 88 %
Applications types	Stations de travail de haut de gamme, journalisation de données, rendu en temps réel, données très transitoires	Système d'exploitation, bases de données transactionnelles	Système d'exploitation, bases de données transactionnelles	Entreposage de données, mise en oeuvre de serveurs Web, archivage	Entreposage de données, mise en oeuvre de serveurs Web, archivage	Archivage de données, sauvegarde sur disque, solutions à haute disponibilité, serveurs gourmands en capacité	Base de données à accès rapide, serveurs d'applications	Bases de données volumineuses, serveurs de fichiers, serveurs d'applications	Archivage de données, sauvegarde sur disque, solutions à haute disponibilité, serveurs gourmands en capacité

Tableau 3: Tableau comparatif des solutions des RAID

III. Mise en œuvre

3.6. Présentation de la solution

Un SAN est un stockage basé sur des blocs et une architecture à haute vitesse qui connecte des serveurs à leurs unités logiques (LUN). Un LUN est une plage de blocs extraite d'un pool de stockage partagé et présentée au serveur en tant que disque logique. Le serveur partitionne et formate ces blocs, généralement à l'aide d'un système de fichiers, pour stocker les données sur le LUN comme s'il s'agissait d'un disque local.

Les clusters de basculement offrent également la fonctionnalité de volume partagé de cluster (CSV) qui procure un espace de noms distribué et cohérent que les rôles en cluster peuvent utiliser pour accéder au stockage partagé à partir de tous les nœuds. La fonctionnalité de clustering de basculement garantit ainsi aux utilisateurs des interruptions de service minimales.

La disponibilité du stockage est un atout très important dans une entreprise. Donc, on a trouvé utile d'utiliser la solution qu'on vient de vous présenter au-dessus.

3.6.1 Architecture du déploiement

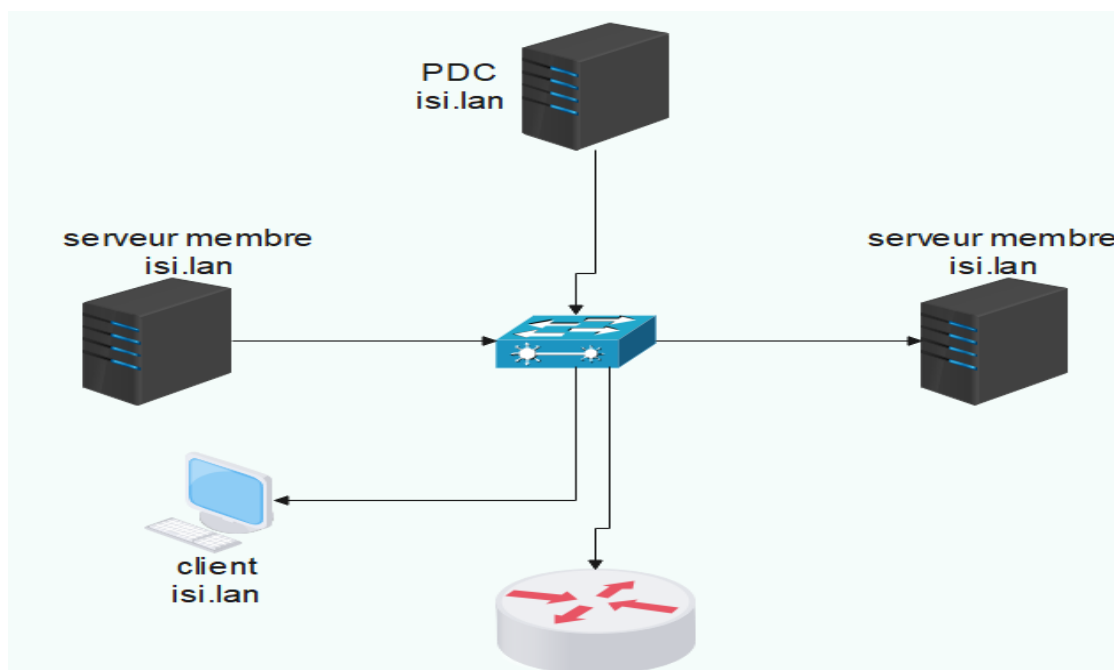


Figure 20:Architecture du déploiement

3.7. Implémentation

Dans ce chapitre nous allons procéder à la configuration du stockage SAN et un cluster.

3.7.1 prérequis

Pour la mise en œuvre de notre mémoire, on aura besoins de 4 machines physique ou virtuelle pour l'installation des services. Mais pour notre étude nous ferons usage d'utiliser des machines virtuelles et d'un outil de virtualisation appelé VMWare Workstation.

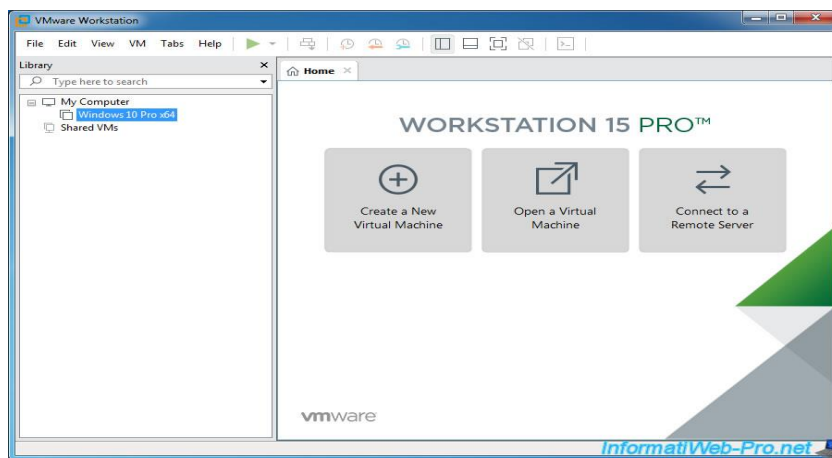


Figure 21: photo du logiciel VMWARE

Sur les 4 machines virtuelle que je vais installer. L'un est le PDC, le 2 autres seront les serveurs membre et la dernière est le client.

Le PDC aura :

- Un **CPU** : 2ghz
- Une **RAM** : 1Go
- Deux **disques** : 60Go et 120Go
- Une **carte réseau**
- Une **carte graphique**
- Un **lecteur CD**

Les deux serveurs membre auront :

- Un **CPU** : 2ghz
- Une **RAM** : 1Go
- un **disque** : 60Go
- Deux **carte réseau**
- Une **carte graphique**
- Un **lecteur CD**

Le client aura :

- Un **CPU** : 2ghz
- Une **RAM** : 1Go
- un **disque** : 60Go
- Une **carte réseau**
- Une **carte graphique**
- Un **lecteur CD**

3.7.2 Configuration du stockage SAN

Pour la configuration du stockage SAN, on devrait installer un rôle qui s'appelle serveur cible ISCSI sur le PDC.

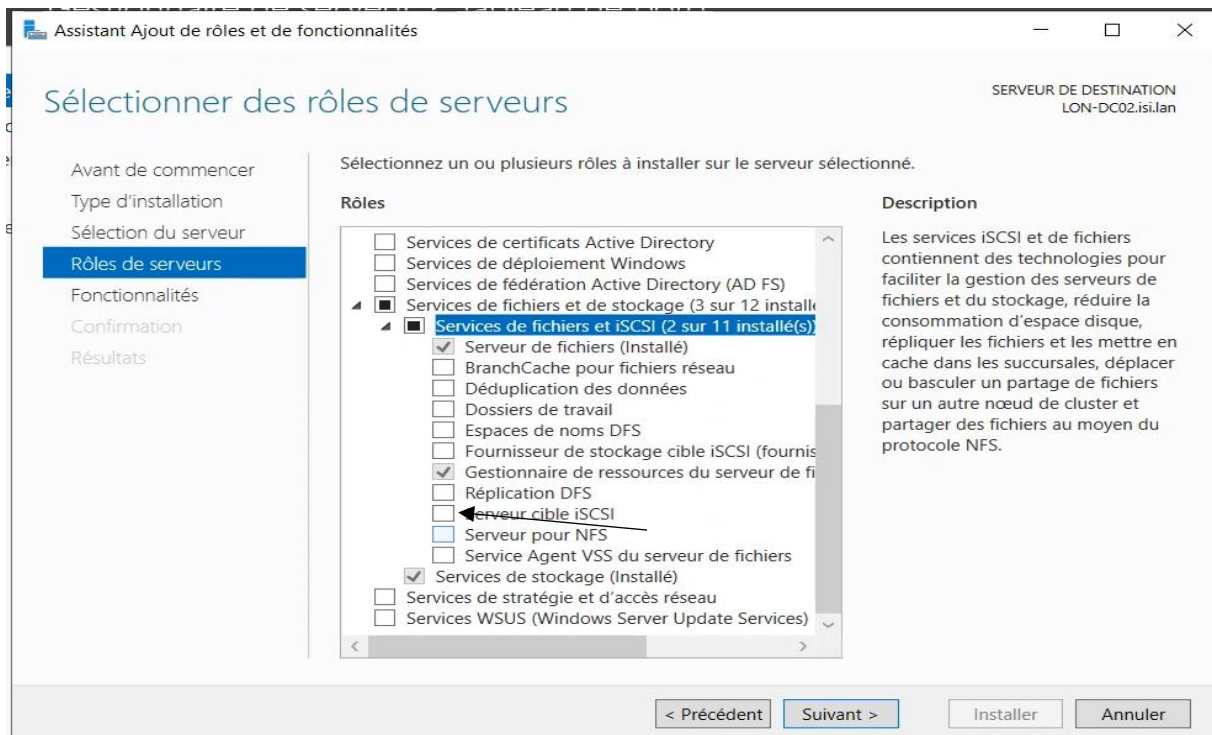


Figure 22: installation de rôle d'ISCSI

Après l'installation du serveur cible iscsi, on crée un nouveau disque virtuel iscsi.

Comme vous voyez dans la photo ci-dessous. On a créé deux disques virtuels avec comme initiateur les deux serveur membre.

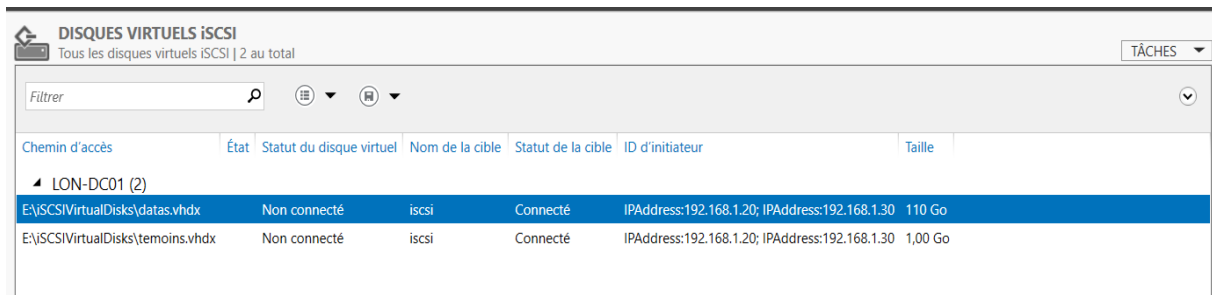


Figure 23: création des disques virtuels iscsi

On initie les disques virtuels qu'on vient de créer sur les deux serveurs membre.

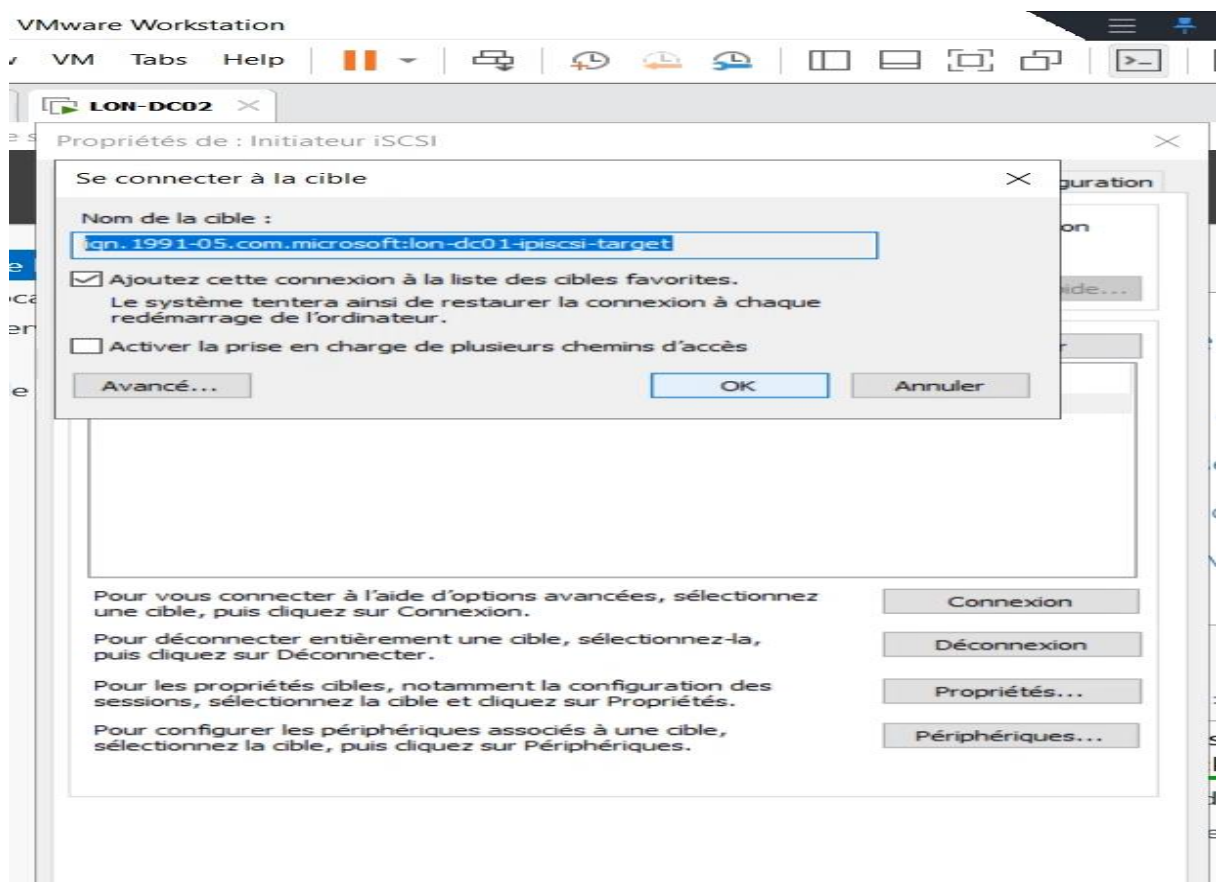
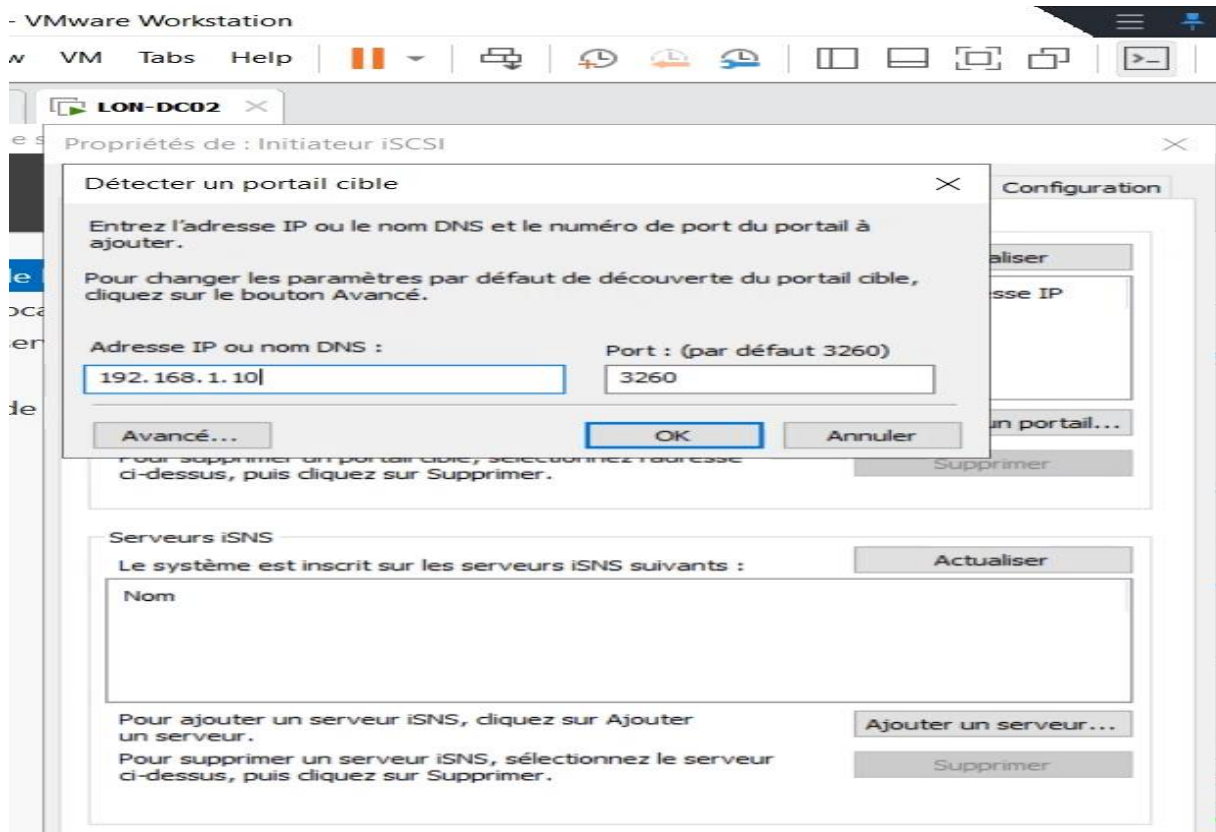


Figure 24: initiation des disques virtuels

3.7.3 Configuration de clustering de basculement

Un cluster de basculement est un groupe d'ordinateurs indépendants qui travaillent conjointement pour accroître la disponibilité et l'extensibilité des rôles en cluster.

Pour la configuration du cluster de basculement, on devrait installer un rôle qui s'appelle clustering de basculement sur les deux serveurs membre.

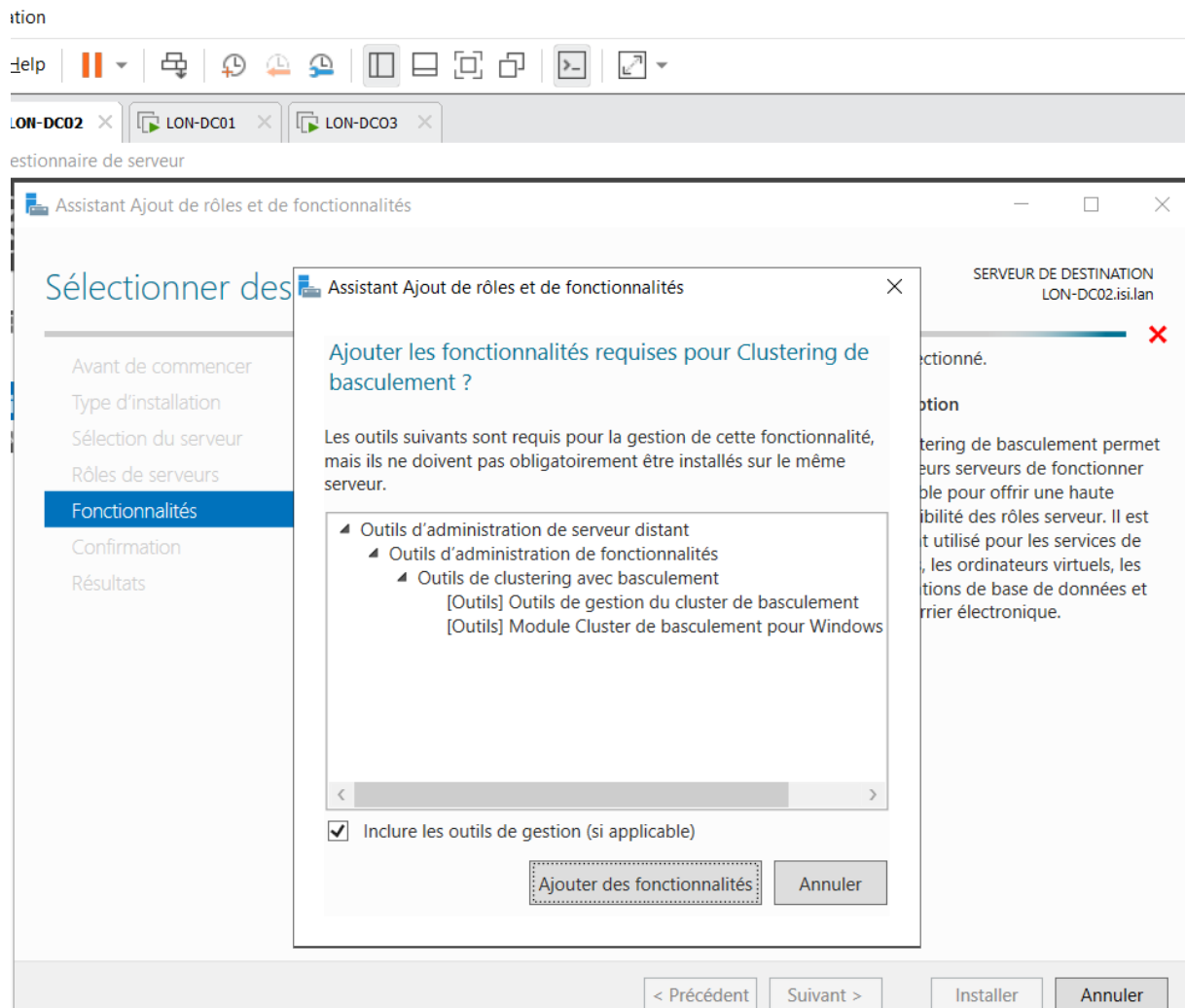


Figure 25: installation du rôle de clustering de basculement

Après l'installation du rôle de clustering, on va passer à la configuration de ce rôle

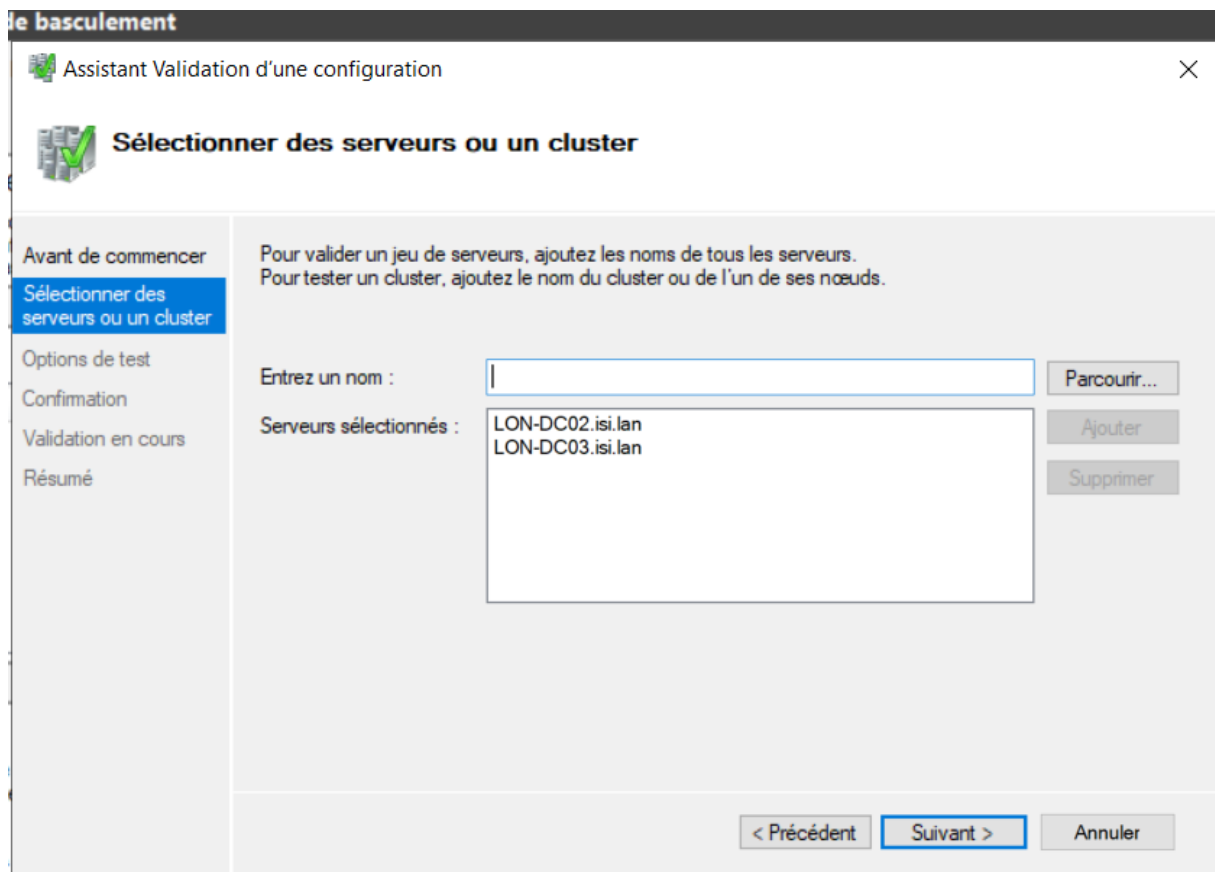


Figure 26: ajout des serveurs pour le cluster

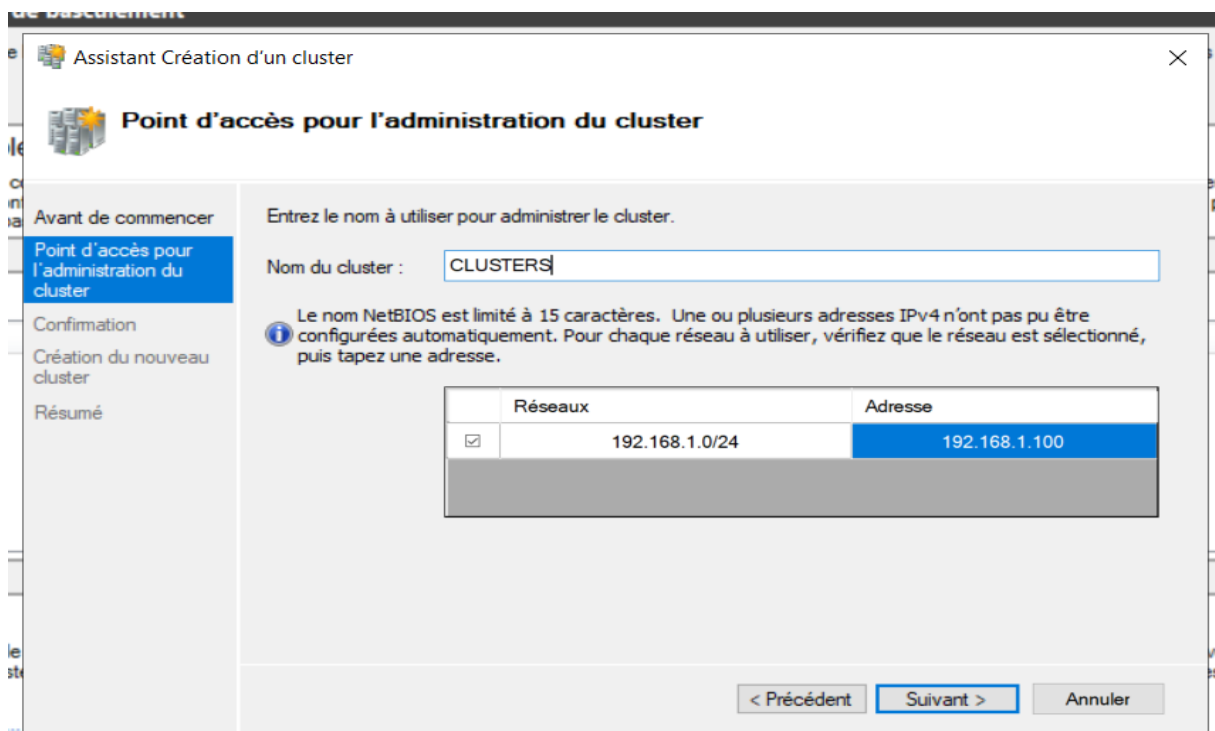


Figure 27: création d'un cluster

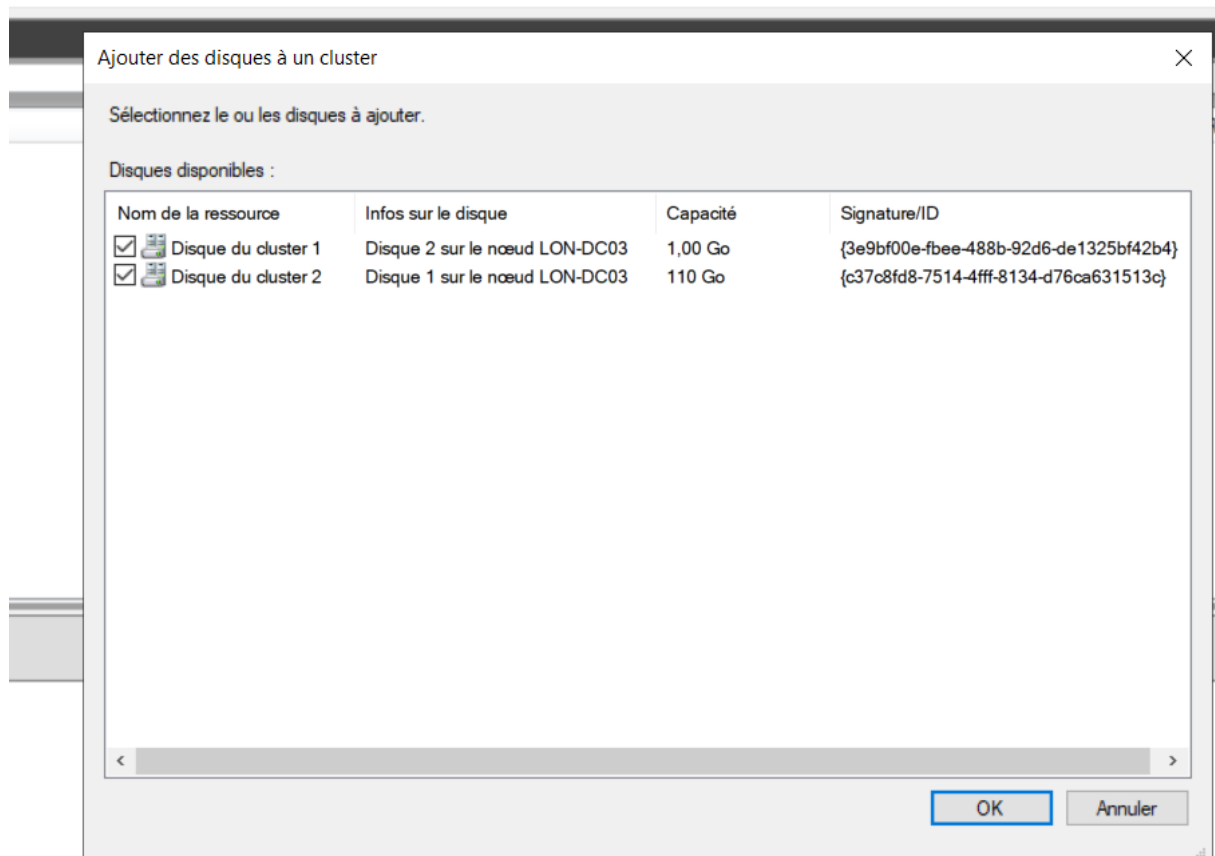


Figure 28: ajout les disques au cluster

Après l'ajout des disques de stockage SAN. On prend l'un de disque pour qu'il soit le témoin et l'autre disque de partage (là où on stockera nos données).

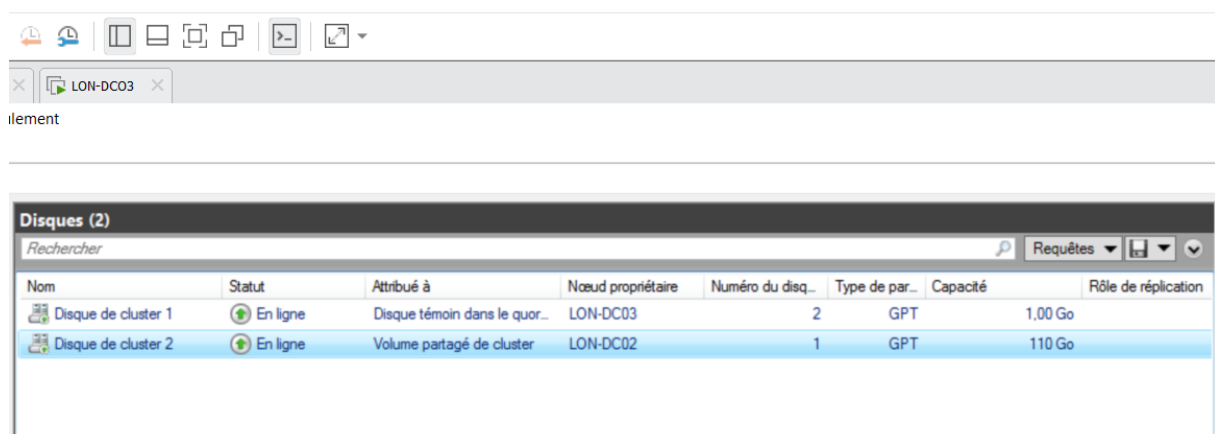


Figure 29: les disques au cluster avec les configurations

3.8 Test

Pour ce chapitre on va configurer un serveur de fichier pour que le client ait les données.

3.8.1 Configuration du serveur de fichier dans le cluster

Pour la configuration du serveur de fichier, on devrait d'abord installer le rôle serveur de fichier sur les deux serveurs membre.

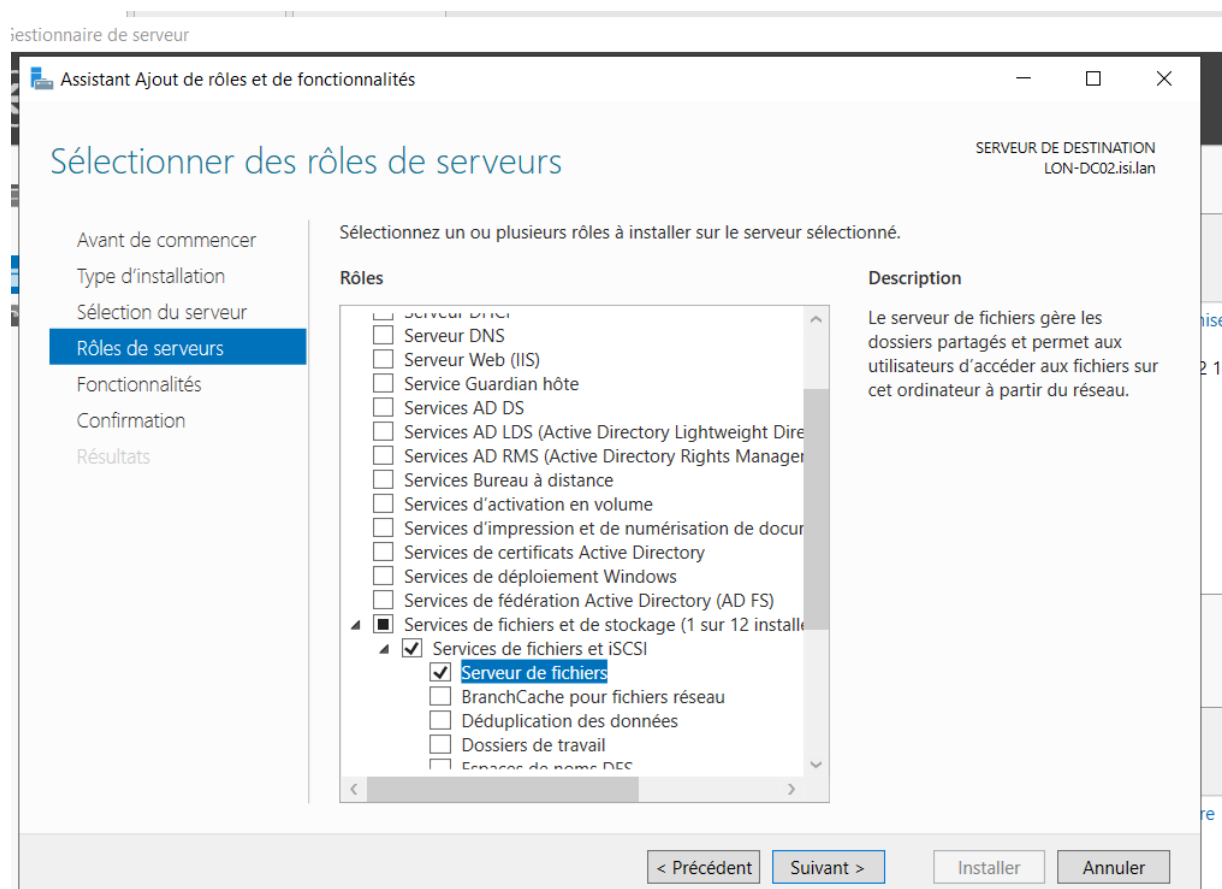


Figure 30:installation du serveur de fichier

Après l'installation, on ajoute le rôle sur le cluster.

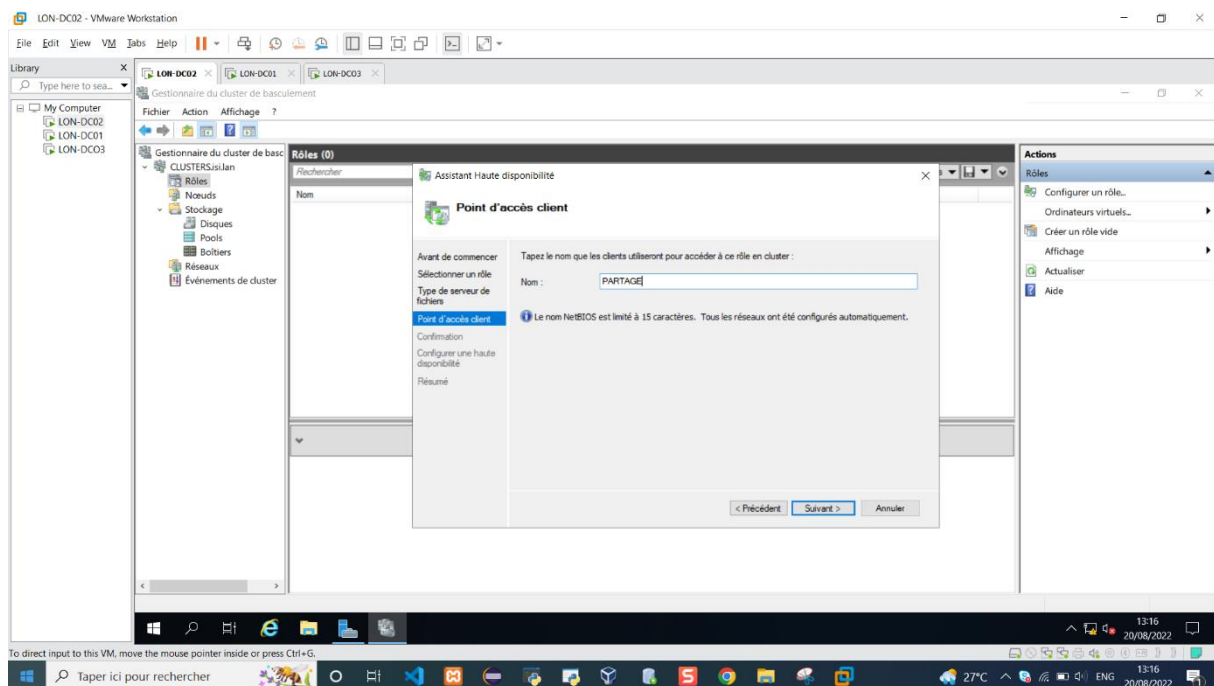
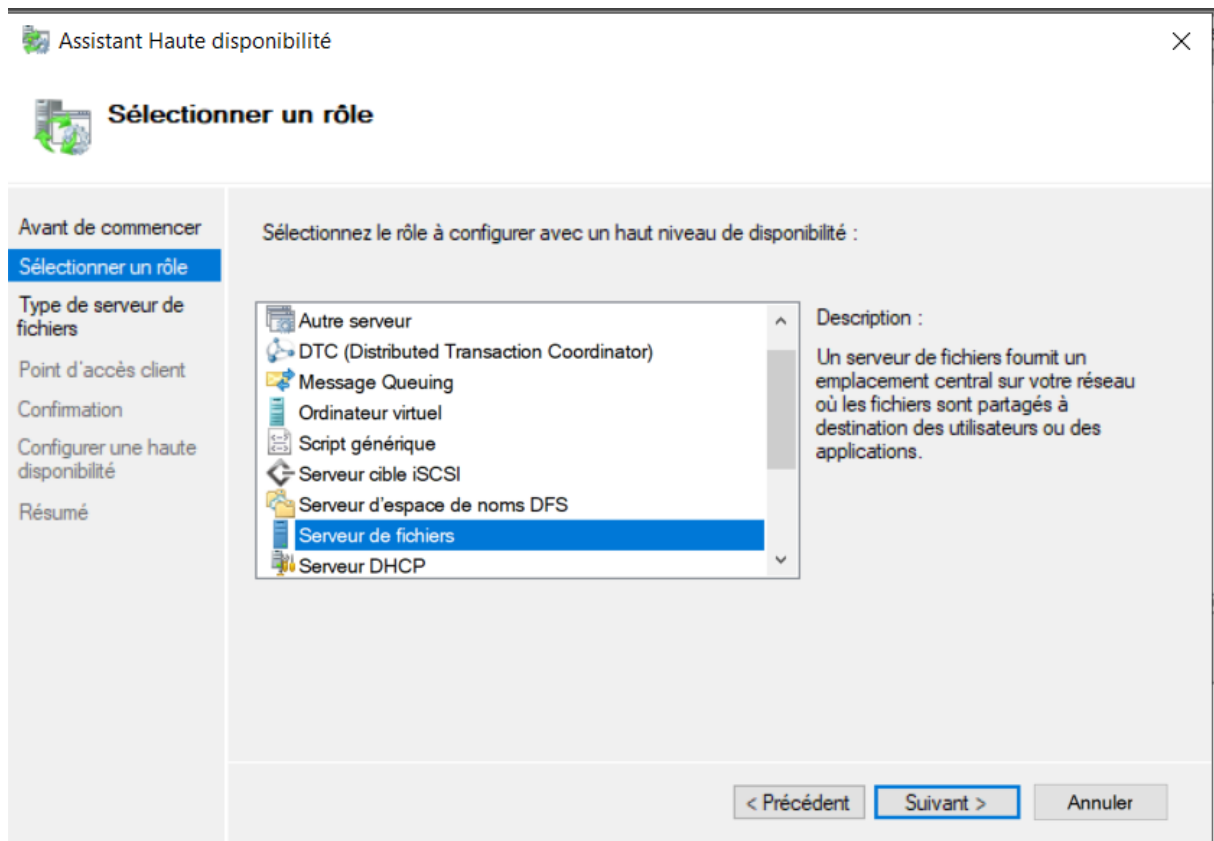


Figure 31: ajout le rôle au cluster

3.8.2 Configuration du partage

Le partage de fichiers est une technique de transfert de fichier consistant à distribuer ou à donner accès, à distance, à des données numériques à travers un réseau informatique

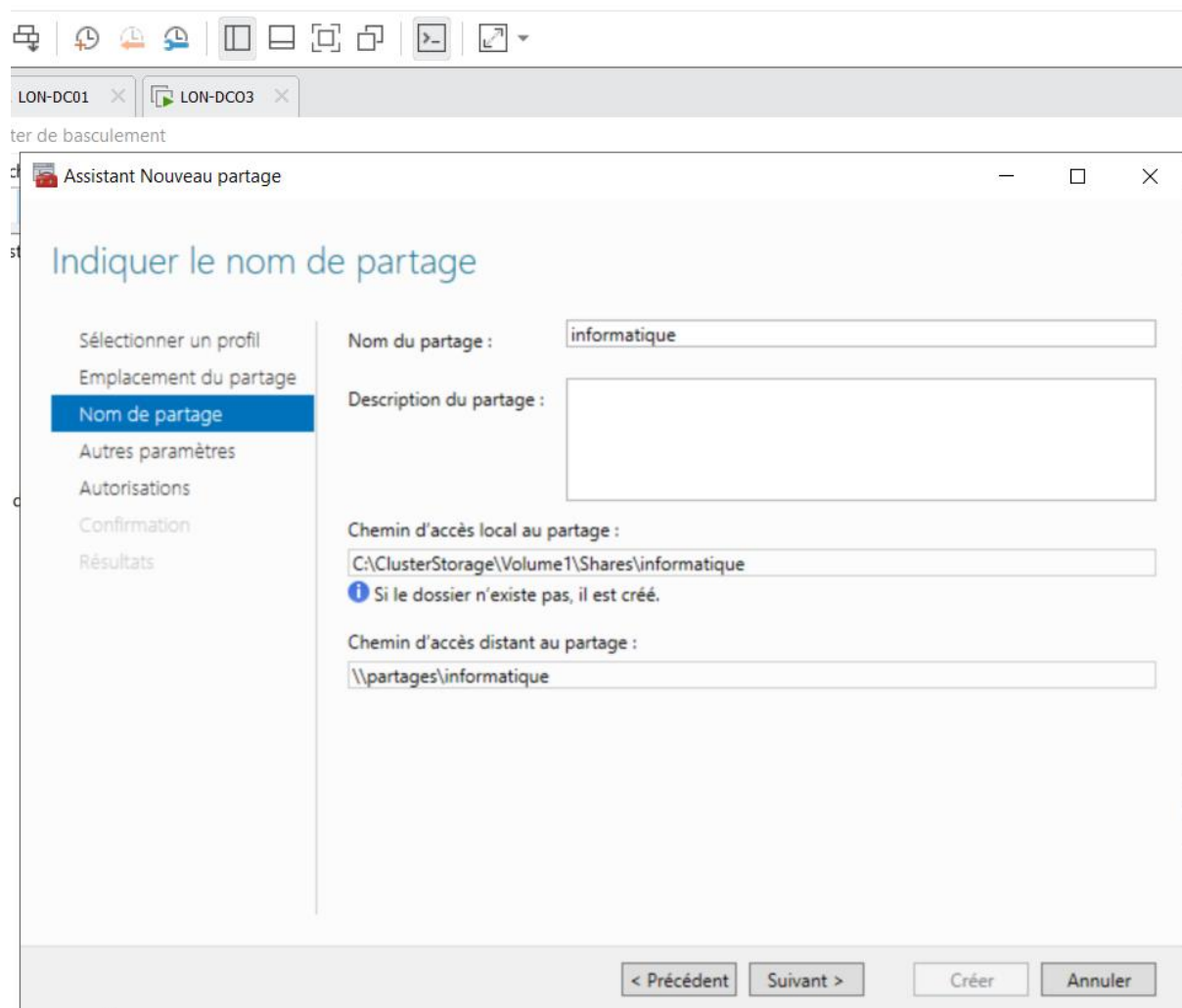


Figure 32: création du fichier de partage

3.8.3 Testé la connectivité sur le client

Pour tester si le partage qu'on vient de faire marche, on entre dans la machine cliente avec le même domaine que les serveur membre.

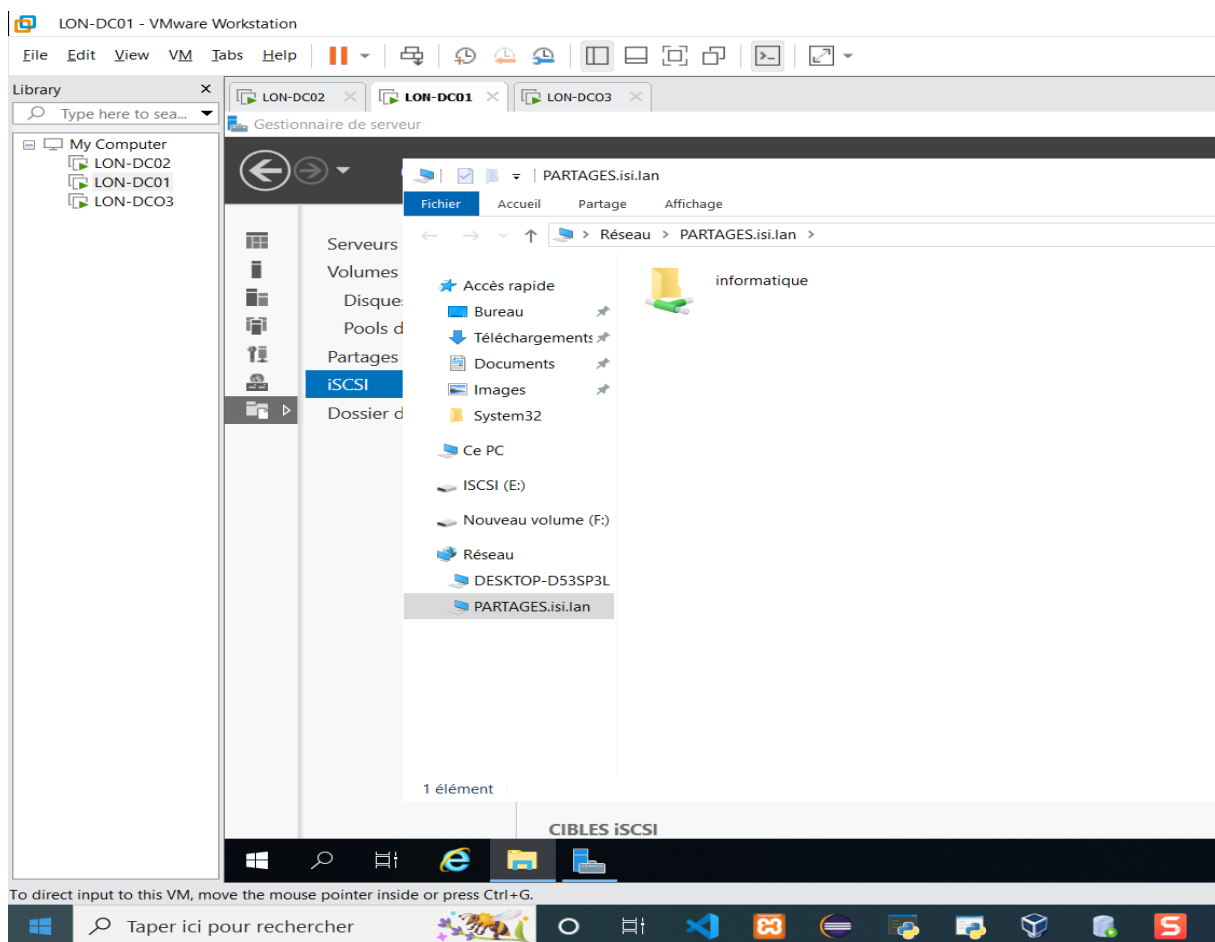
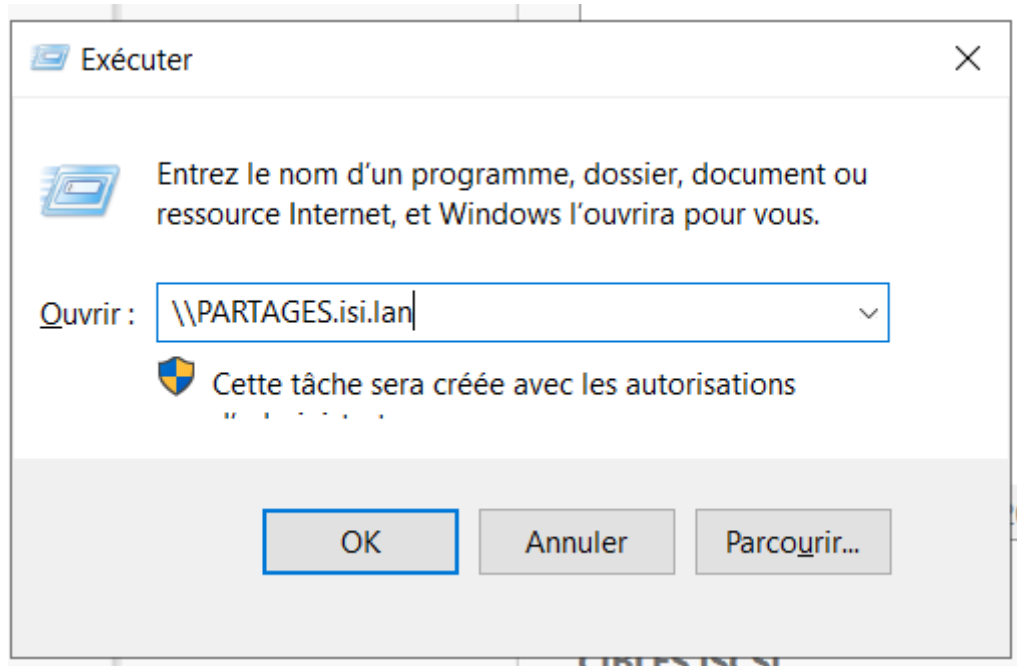


Figure 33:image du partage

Testons le cluster de basculement

Comme vous voyez dans la figure ci-dessus le rôle de partage est associé avec le serveur LON-DC03.

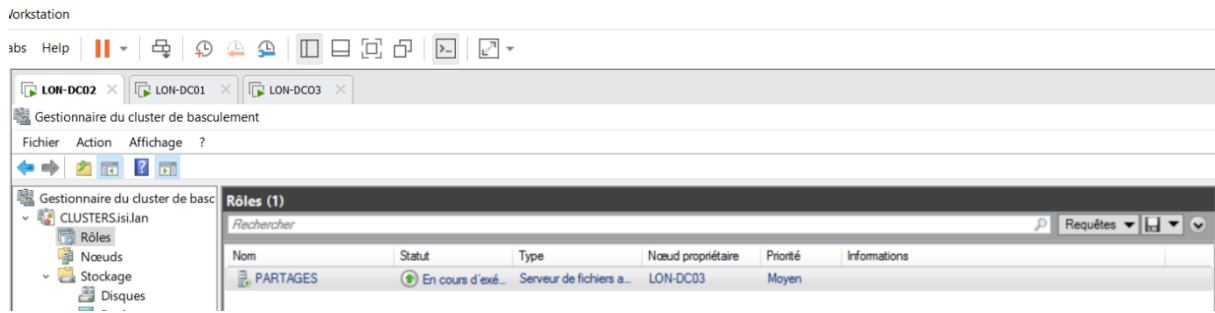


Figure 34: rôle du serveur de fichier

Maintenant, on va essayer de stopper le serveur LON-DC03 pour voir s'il va basculer vers le LON-DC02.

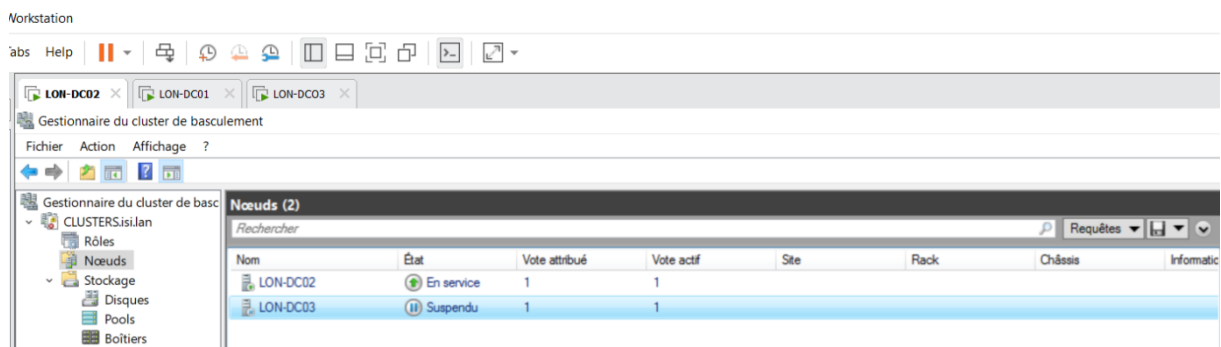


Figure 35: image de suspension d'un des serveurs

On voit maintenant dans la figure ci-après que le rôle de serveur de fichier est basculé vers le serveur LON-DC02

LON-DC02 LON-DC01 LON-DC03

Gestionnaire du cluster de basculement

Fichier Action Affichage ?

Navigation icons

Gestionnaire du cluster de basculement

- CLUSTERS.isi.lan
 - Rôles
 - Nœuds
 - Stockage
 - Disques
 - Pools

Rôles (1)

Rechercher [Icon] Requête [Dropdown] [Dropdown]

Nom	Statut	Type	Nœud propriétaire	Priorité	Informations
PARTAGES	En cours d'exé...	Serveur de fichiers a...	LON-DC02	Moyen	

3.9 Amélioration de la solution

Pour l'amélioration de notre solution, on a trouvé nécessaire d'utiliser la réplication des données pour que si l'un de serveur est tombé en panne, on trouvera parfaitement nos données dans l'autre serveur.

3.9.1 configuration de la réplication

La réplication est le processus de partage d'informations pour assurer la cohérence des données entre plusieurs sources de données redondantes afin d'améliorer la fiabilité, la tolérance aux pannes ou la disponibilité des données.

Pour la configuration répliquée, nous devons installer ces deux rôles : DFS Namespace et DFS Réplication dans les deux serveurs membres.

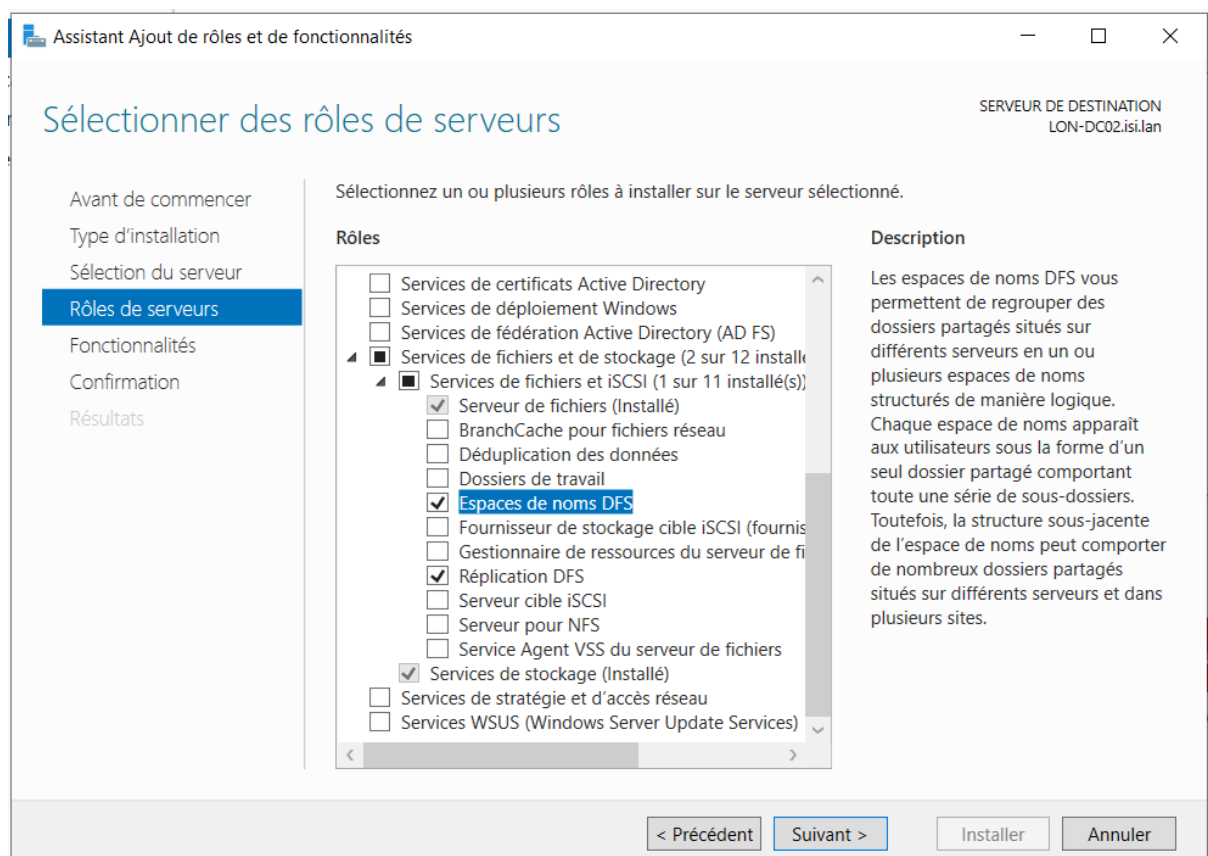


Figure 36: installation de 2 rôles pour la réplication

Après l'installation des rôles, on crée un nouveau groupe de réplication.

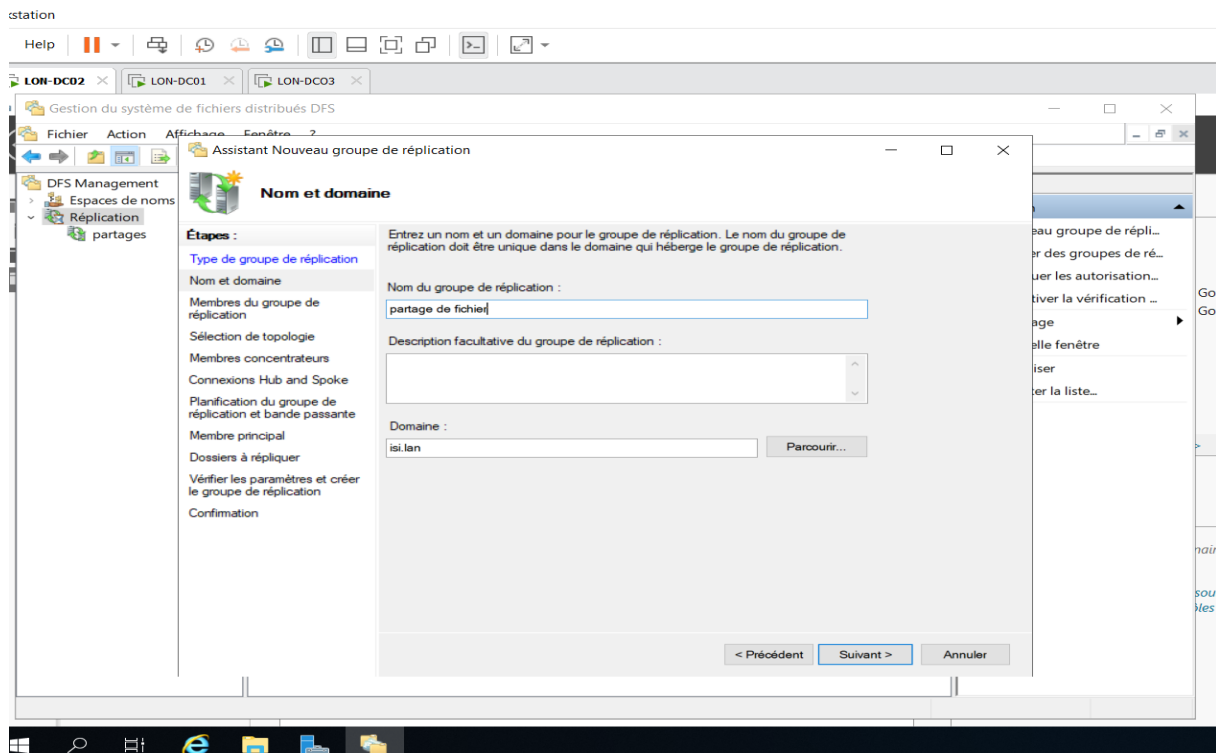


Figure 37: groupe de réplication

Maintenant ajoutons les deux serveurs membres pour qu'on réplique les données sur l'autre.

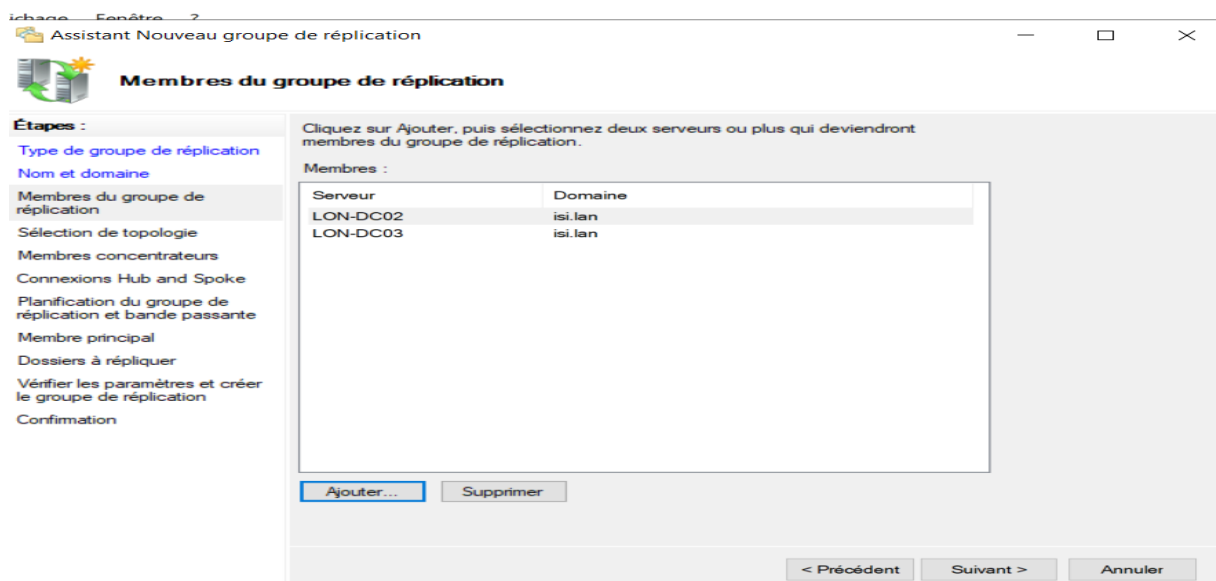


Figure 38:ajout des serveurs

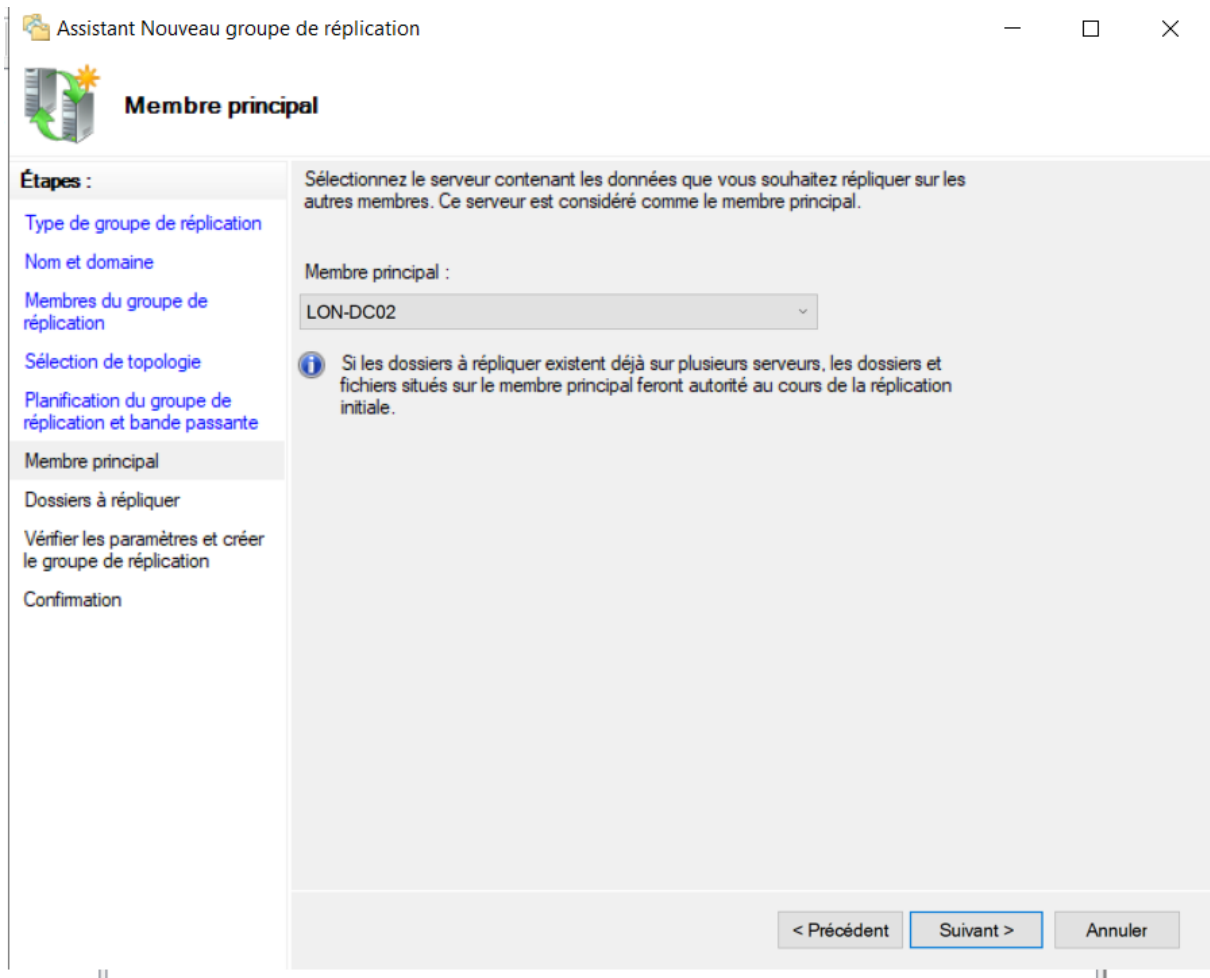


Figure 39:choix du serveur principal

Choisissons le dossier qu'on veut répliquer. Comme nous on a fait un partage de fichier, on va répliquer le dossier du partage.

Après avoir fait cela on choisit là où on veut le stocker sur l'autre serveur membre

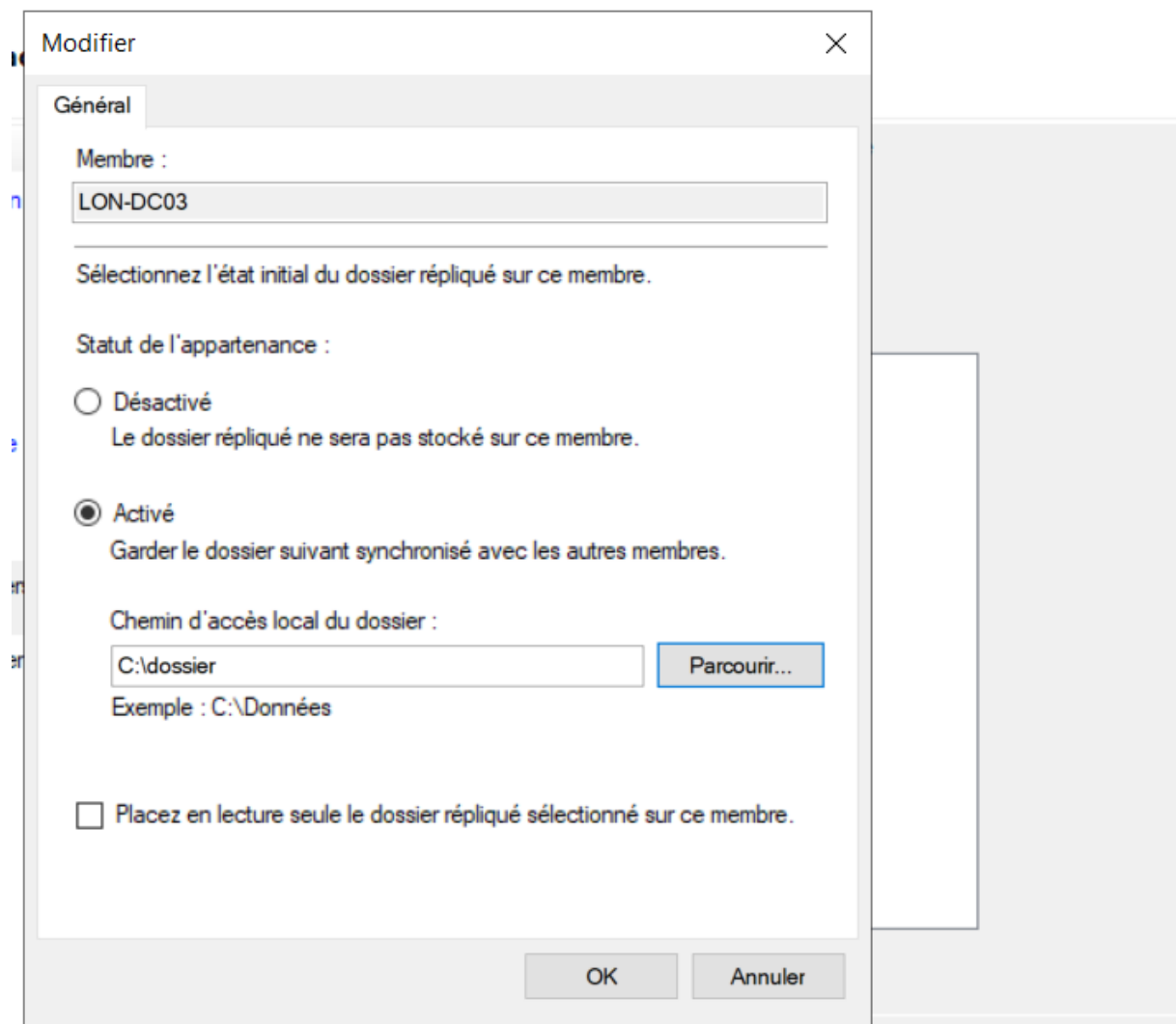


Figure 40:choix d'emplacement du fichier répliqué

3.9.2 Teste de la solution

On va d'abord créer un document dans le dossier principal pour voir si on va le voir dans l'autre serveur après la réplication.

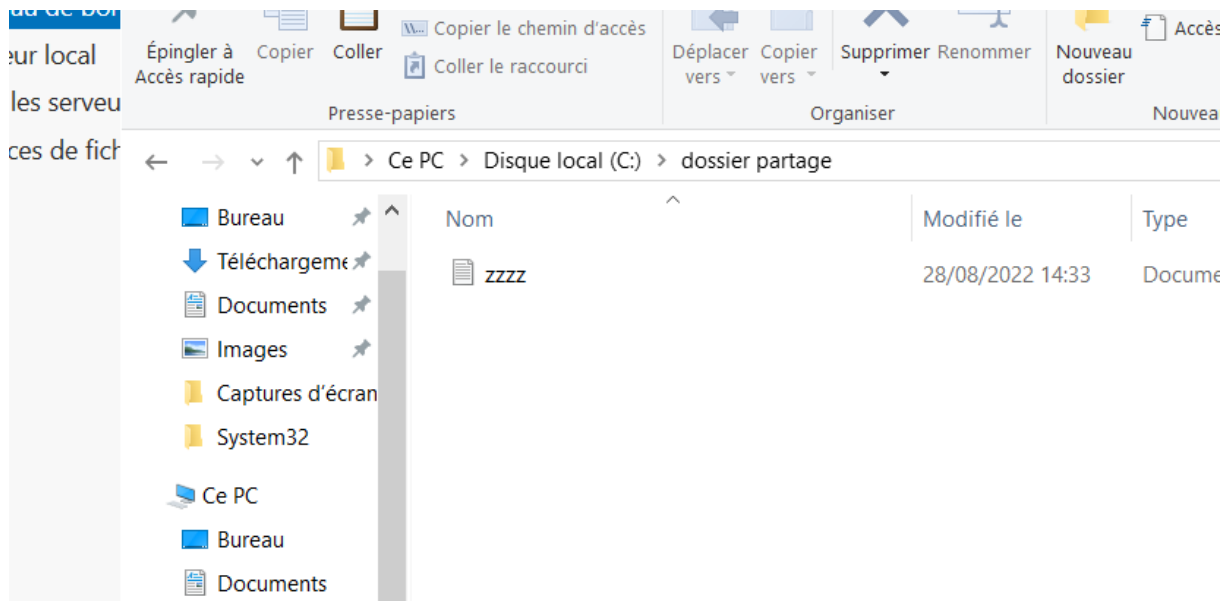


Figure 41:creation du document dans le serveur principal

Maintenant on va essayer de répliquer les données.

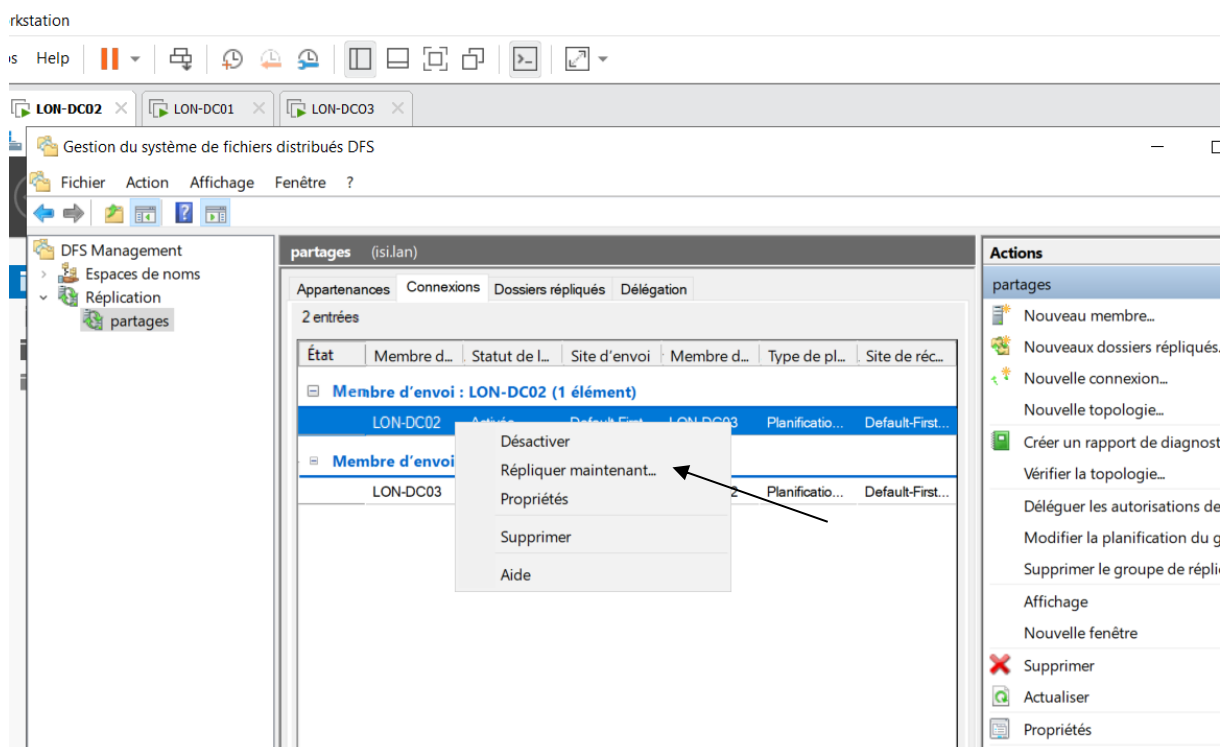


Figure 42:réplication des données



Comme vous voyez dans cette image ci-dessous les données sont répliquées dans l'autre serveur membre (LON-DC03).

Conclusion générale

Au terme de notre analyse, les données sont la ressource la plus précieuse de toute entreprise et leur perte peut avoir de graves conséquences. De plus, si votre entreprise est soumise à certaines normes réglementaires, la perte d'un serveur pourrait vous mettre hors conformité et risquer une amende de plusieurs milliards d'euro. C'est pour cela chaque entreprise a un intérêt de garder ces données en toutes sécurité et hautement disponible afin de préserver leur activité. Pour minimiser ces défaillances au sein des entreprise, nous utiliseront la solution de clustering de basculement. Ce dernier est une solution de qui rend les données hautement disponibles.

En effet, il est très important pour les entreprises de grande taille de chercher un système de stockage de hautement disponible pour qu'il ne perd jamais ces données, au cas où le pire arriverait, bien avant qu'il n'arrive. Le système de stockage de haute disponibilité dont le clustering de basculement qui fonctionnent bien utilisent des serveurs qui sont connectés par des câbles physiques et par des logiciels. En cas de défaillance d'un ou plusieurs nœuds, d'autres nœuds prennent le relais pour fournir les services requis. Comme avec cette solution, la perte de données sera minimisée

Depuis le début de notre travail, nous avons rencontré de nombreuses difficultés, notamment pour trouver des informations précises sur notre sujet sur Internet. on voit parfois beaucoup d'informations sur notre sujet en anglais, on est obligé de les traduire et de les comparer à d'autres pour aboutir à des informations justes et bonnes.

En plus de clustering de basculement, il y a le rôle de réplication de données, qui offrent des opportunités supplémentaires pour améliorer notre sujet afin d'assurer la disponibilité des données.

Bibliographie

I) Mémoire

M. Dione Aliou, *Mise en œuvre d'une solution de sauvegarde et de restauration*, ISI, 2019-2020, 83 pages.

Webographie

<http://www.ctmsolutions.com/www/news-integration/4085.html> 24/05/2022 à 20H

<https://www.techno-science.net/glossaire-definition/RAID-informatique-page-4.html>
30/05/2022 à 20H

<https://www.redhat.com/fr/topics/data-storage/network-attached-storage> 25/06/2022 à 20H

<https://ryax.tech/fr/san-nas-das-differences/> 1/07/2022 à 20H

<https://cisco.goffinet.org/ccna/fondamentaux/modeles-tcp-ip-osi/#13-mod%C3%A8le-tcpip>
19/07/2022 à 20H

https://fr.wikibooks.org/wiki/Les_r%C3%A9seaux_informatiques/Les_mod%C3%A8les_OSI_et_TCP#:~:text=Le%20mod%C3%A8le%20TCP%20est,physiques%20et%20liaison%20d'OSI. 1/08/2022 à 20H

<https://docs.microsoft.com/fr-fr/windows-server/failover-clustering/failover-clustering-overview> 20/08/2022 à 20H

<https://docs.microsoft.com/fr-fr/learn/modules/implement-windows-server-file-server-high-availability/2-explore-windows-server-file-server-high-availability-options> 1/09/2022 à 20H

<https://public.dalibo.com/exports/formation/manuels/modules/w7/w7.handout.html>
15/09/2022 à 18H

Table des matières

Dédicace	I
Remerciements	II
Avant-propos	III
Glossaire	IV
Liste des figures	V
Liste des tableaux	VII
Sommaire	VIII
Résumé	IX
Abstract	X
Introduction	1
I. Les Cadres théoriques et méthodologique	2
1.1. Cadre théorique	2
1.1.1. Problématique	2
1.1.2. Objectifs	2
1.1.3. Hypothèse de recherche	3
1.2. Cadre méthodologique	4
1.2.1. Délimitation du champ de l'étude	4
1.2.2. Les techniques de recherche	4
1.2.3. Les difficultés rencontrées	4
II. Cadre conceptuel	6
2.3. Rappel sur le réseau	6
2.3.1. Modèle TCP/IP et OSI	6
2.3.1.1. Modèle TCP/IP	6
2.3.1.2. Modèle OSI	7
2.3.2. L'environnement client/serveur	9
2.4. Généralité sur le stockage	10
2.4.1. Définition du stockage	10
2.4.2. Les différentes architectures de stockage	11
2.4.2.1. Le DAS	11
2.4.2.2. Le NAS	12
2.4.2.3. Le SAN	14
2.4.3. Comparaison	16
2.4.4. Présentation de la solution	18
2.5. Rappel sur la haute disponibilité	18

2.5.1	Présentation de la haute disponibilité	18
2.5.2	Les technologies des RAIDS	19
2.5.2.1	Les technologies RAID standard	20
2.5.2.2	Les technologies RAID combiné.....	27
2.5.3	Comparaison	31
III.	Mise en œuvre	33
3.6.	Présentation de la solution.....	33
3.6.1	Architecture du déploiement	33
3.7.	Implémentation.....	34
3.7.1	prérequis	34
3.7.2	Configuration du stockage SAN	35
3.7.3	Configuration de clustering de basculement	38
3.8	Test	41
3.8.1	Configuration du serveur de fichier dans le cluster	41
3.8.2	Configuration du partage	43
3.8.3	Testé la connectivité sur le client	43
3.9	Amélioration de la solution	47
3.9.1	configuration de la réplication	47
3.9.2	Teste de la solution	50
	Conclusion générale	53
	Bibliographie	ix
	Webographie	x
	Table des matières	viii