
Système de fichiers distribués - DFS

BTS SIO - Bloc 2 - SISR - Administration des systèmes et des réseaux

1. Introduction	3
2. Le système DFS	3
2.1. Présentation de l'espace de noms DFS	3
2.2. La réplication DFS	3
2.2.1. Que se passe-t-il en cas de conflit ?	4
2.3. Fonctionnement de l'espace de noms	4
2.3.1. Terminologie : racine, dossier et cible	4
2.4. La déduplication de données	4
2.4.1. Principe de fonctionnement de la déduplication	4
2.5. Scénarios DFS	5
2.5.1. Partage de fichiers entre succursales	5
2.5.2. Collecte de données	5
2.5.3. Distribution des données	5
3. Architecture : exemple	6
4. Conclusion : Avantages du DFS	7

1. Introduction

Le système de fichiers est un des points essentiels dans une entreprise, en effet il évolue de façon quotidienne. Mal administré, ce système peut devenir très vite ingérable.

2. Le système DFS

DFS est accessible dans un environnement Microsoft sous Windows Server. L'acronyme **DFS** signifie *Distributed File System* c'est-à-dire **Système de fichiers distribués**.

Ce système de fichier hiérarchisé permet de structurer les fichiers partagés sur différents serveurs du réseau de façon logique. Il permet de référencer un ensemble de partages qu'il faudra rendre accessibles de manière uniforme puis, de centraliser l'ensemble des espaces disponibles sur cet ensemble de partages.

Avec le DFS, l'utilisateur final ne visualise pas le nom du serveur sur lequel il accède pour lire les données comme c'est le cas pour un partage classique (chemin UNC en \\nomServeur\nomPartage), cela est totalement transparent. En cas de remplacement d'un serveur de fichiers, il est nécessaire de procéder à la mise à jour de tous ces noms. Cette dernière étape peut se révéler dans certains cas très longue. Un **espace de noms DFS** permet donc de faciliter la tâche de l'administrateur, le chemin UNC ne contenant pas le nom du serveur concerné.

Derrière un même chemin d'accès DFS peuvent se cacher plusieurs serveurs, contenant les mêmes données avec une synchronisation entre ces serveurs grâce à **DFSR (DFS Replication)**. Cette approche est très intéressante pour de la haute disponibilité de données et de la répartition de charge. Ainsi la tolérance de panne est assurée.

Bien sûr, il est possible d'utiliser un chemin pour un serveur (1 pour 1), où chaque serveur héberge des données différentes. Sachez que le DFS étant de plus en plus présent, désormais il est même utilisé pour la réplication SYSVOL sur un domaine Microsoft entre les contrôleurs de domaine.

Lors de l'installation du rôle DFS, deux services de rôle peuvent être sélectionnés. Il s'agit de :

- **Espace de noms ou DFS-N** : permet l'installation de la console et des outils nécessaires à l'administration de l'espace de noms.
- **Réplication DFS ou DFS-R** : installe un moteur de réplication multi-maître permettant la réplication des différents dossiers contenus dans l'espace de noms. La réplication peut être planifiée avec une utilisation de la bande passante différente en fonction de l'heure. De plus, la compression différentielle à distance est utilisée afin de répliquer uniquement la partie d'un fichier modifié depuis la dernière réplication. La réplication n'est pas obligatoirement associée à l'espace de noms, elle peut fonctionner de manière autonome.

2.1. Présentation de l'espace de noms DFS

Un espace de noms simplifie la gestion d'un système de fichiers en représentant de manière virtuelle les partages réseau. cet espace de noms peut être de type autonome ou basé sur un domaine Active Directory.

2.2. La réplication DFS

La réplication DFS est un mécanisme qui permet la réplication des différents dossiers sur un ou plusieurs serveurs. Ce type de réplication offre l'avantage d'utiliser la compression différentielle à distance, qui est un protocole de type client-serveur qui permet la détection des modifications (ajout/suppression/modification) opérées sur un fichier,

dans le but de répliquer uniquement ce bloc de données modifiées. Ce protocole est utilisé pour des fichiers d'une taille minimale de 64Ko.

Lors de la réplication, un dossier intermédiaire est utilisé, une copie compressée du fichier est créée dans ce dossier intermédiaire, par la suite l'envoi est effectué. Le serveur qui reçoit les données stocke également le fichier dans le dossier intermédiaire. Lorsque le téléchargement est terminé, le fichier est décompressé puis installé dans le dossier adéquat. Ces dossiers temporaires sont généralement présents dans DFSrPrivate\Staging.

2.2.1. Que se passe-t-il en cas de conflit ?

En cas de conflit lors de la réplication, la personne qui a effectué la dernière modification l'emporte. Si le conflit concerne le nom de fichier, le premier utilisateur qui a effectué la modification l'emporte. Un déplacement des fichiers qui ont perdu le conflit est effectué vers DFSrPrivate\ConflictandDeleted.

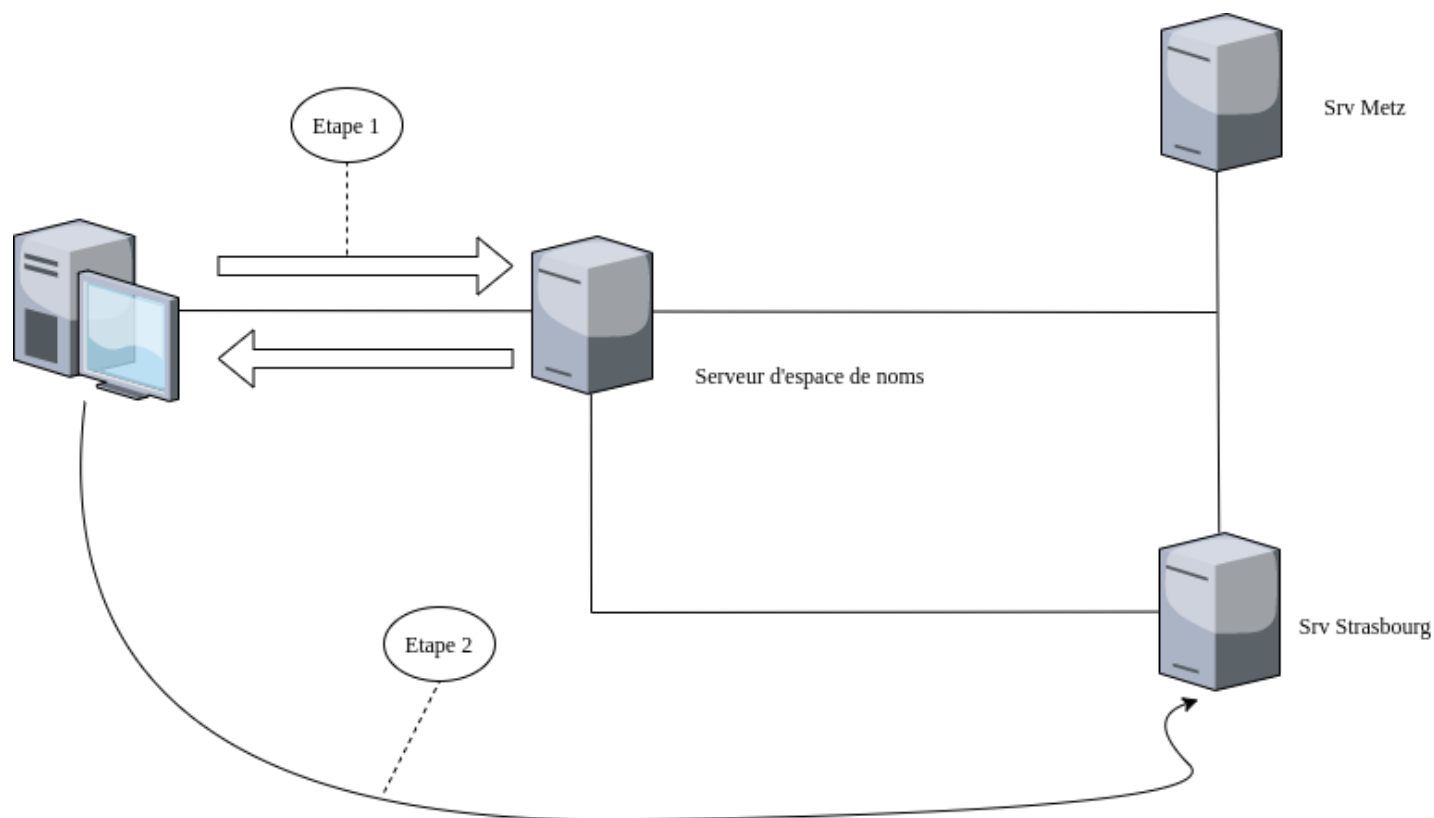
2.3. Fonctionnement de l'espace de noms

Pour faciliter la compréhension du fonctionnement de l'espace de noms, nous allons prendre un exemple.

Un utilisateur nommé Nicolas travaille au siège de l'entreprise Rebirth. L'entreprise est composée d'un siège social à Strasbourg ainsi que d'une agence à Metz. L'espace de noms est utilisé par les utilisateurs pour accéder aux différents partages réseau.

Le poste de l'utilisateur contacte le serveur d'espace de noms (étape 1) qui lui envoie une liste triée (en fonction de la configuration des administrateurs) des serveurs contenant les répertoires partagés (cibles de dossier) auxquels l'utilisateur tente d'accéder. Le poste client essaie d'accéder au premier serveur (étape 2) de la liste (les autres sont contactés uniquement si le premier est HS).

Dans l'étape 2, l'utilisateur a la possibilité d'accéder aux autres serveurs car la réplication entre deux serveurs a été mise en place.



2.3.1. Terminologie : racine, dossier et cible

Lorsque l'on parle de DFS, trois termes sont importants à retenir et à comprendre : Racine DFS, dossier et cible.

- **Racine DFS** : Point d'entrée principal d'un système DFS, la racine DFS contient le chemin d'accès aux différentes liaisons DFS qui lui sont associées. Il existe deux types de racine DFS mais nous verrons cela en détail dans une autre partie.
- **Dossier** : Le dossier sera le nom du partage affiché côté client et dans la configuration du serveur, une liaison sera effectuée entre ce dossier DFS et la cible DFS afin de faire un lien entre les deux éléments. Certains dossiers n'utilisent pas de cible, uniquement dans le but de hiérarchiser les espaces de noms DFS. Les dossiers sont également appelés « **Liaison DFS** ».
- **Cible** : Serveur sur lequel sont situées les données, la cible représente le chemin d'accès vers le dossier partagé situé sur ce serveur.

Ces 3 éléments sont essentiels à la mise en place d'une infrastructure DFS.

2.4. La déduplication de données

Windows Server 2019 offre la possibilité d'activer la déduplication des données. Cette fonctionnalité ne peut pas être utilisée sur une partition système. Le but de cette fonctionnalité est d'optimiser l'espace disque. Ainsi, un bloc identique à plusieurs fichiers n'est stocké qu'une seule fois.

2.4.1. Principe de fonctionnement de la déduplication

- **Analyse des fichiers sur le volume** : le système analyse les fichiers sur la partition
- **Segmente les fichiers en morceaux de taille variable**
- **Identification des blocs identiques**
- **Rangement et compression** : les blocs de taille variable sont rassemblés et compressés en fonction de leur taille.
- **Remplacement du flux de fichiers** : les fichiers sont remplacés par un lien entre les différents blocs qui composent un fichier et le magasin de stockage des blocs variables.

Le rôle de déduplication comprend différents paramètres de configuration en fonction du type de fichiers que l'on souhaite optimiser.

2.5. Scénarios DFS

Plusieurs scénarios peuvent être utilisés avec le système DFS.

2.5.1. Partage de fichiers entre succursales

Les fichiers sont échangés entre deux ou plusieurs sites de l'entreprise. Cette solution permet une réplication bidirectionnelle donc l'assurance d'avoir des serveurs toujours à jour. De plus, les personnes mobiles d'un site à l'autre ont accès aux différents fichiers de manière plus aisée. Il est à noter que les modifications sont répliquées uniquement après la fermeture du fichier.

Ce scénario est déconseillé pour les fichiers de base de données ou un fichier ouvert sur une longue période (exemple : fichier excel de suivi de production ouvert toute la journée).

2.5.2. Collecte de données

Le scénario Collecte de données consiste à récupérer les données d'un site afin de les répliquer sur un autre site. La réplication est de type unidirectionnel. Elle peut consister à répliquer les données sur un site concentrateur afin de pouvoir procéder à la sauvegarde.

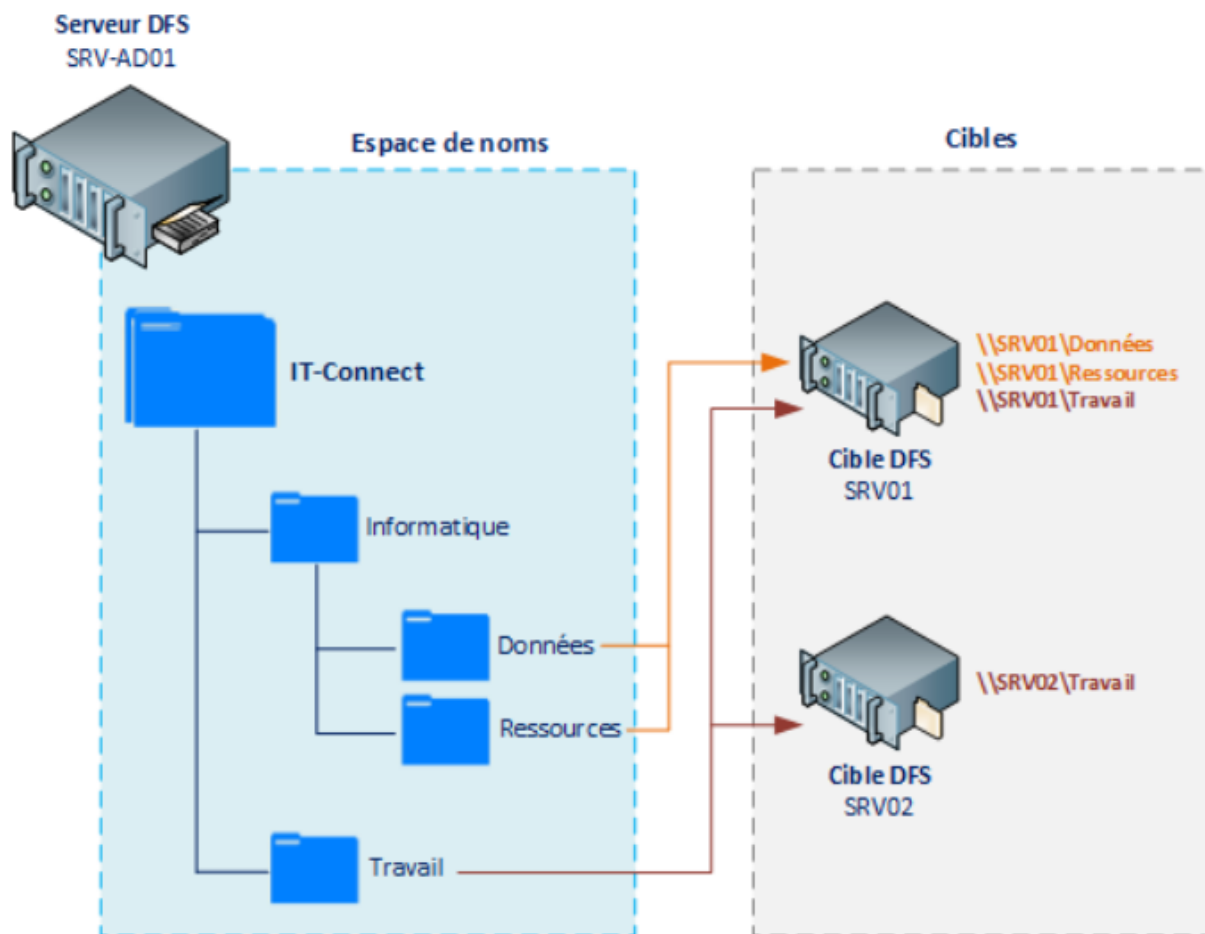
Ainsi les données sont présentes sur deux sites, ce qui permet une tolérance de panne en cas de crash du premier serveur et un coût matériel moindre pour les différents sites distants. Il n'est plus nécessaire de sauvegarder les fichiers sur chaque site puisque ces derniers sont rapatriés sur le site central.

2.5.3. Distribution des données

Cette solution consiste à répliquer des documents sur plusieurs serveurs. Par exemple, un fichier catalogue répliqué de la maison mère vers l'ensemble des sites distants.

3. Un exemple d'architecture

Afin d'illustrer mes propos et de vous permettre de mieux comprendre le fonctionnement de DFS, voici un exemple d'infrastructure DFS :



Sur cette infrastructure, le serveur DFS est « **SRV-AD01** » qui est également contrôleur de domaine du domaine « **it-connect.fr** » et DNS. La racine DFS est « **IT-Connect** ».

Cette racine contient deux sous-dossiers :

- « **Informatique** » qui ne contient pas de cible il est là uniquement pour un aspect organisationnel de l'espace de noms. Ce même dossier contient deux dossiers « **Données** » et « **Ressources** » qui ciblent **SRV01**, ce serveur sera celui qui héberge les données pour ces deux dossiers DFS.
- « **Travail** » contient deux cibles : **SRV01** et **SRV02** qui auront pour but d'héberger les données et d'assurer la haute disponibilité grâce à de la réplication via DFSR.

4. En conclusion : les avantages du DFS

L'utilisation du DFS en entreprise présente plusieurs avantages non négligeables.



- Simplifier l'administration : Si une cible DFS tombe, la liaison DFS peut être déplacée vers un autre serveur contenant une copie des données en changeant le dossier cible sur le serveur DFS. Du côté utilisateur, cela sera totalement transparent car le nom ne changera pas.
- Le client DFS est intégré à Windows ce qui ne nécessite pas d'installation supplémentaire sur les postes clients.
- Un nom unique permet d'accéder à toutes les ressources, il n'est pas nécessaire de mapper une lettre sur chaque ressource.
- Fonction de mise en cache afin d'améliorer les performances.
- Le DFS prend en compte les ACL situées au niveau du système de fichiers.
- Remplacement d'un serveur simplifié car l'espace de noms utilisé côté client n'est pas affecté.
- Équilibrage de charge (si plusieurs cibles par dossier DFS).
- Tolérance aux pannes (si plusieurs cibles par dossier DFS).
- Évolution : Un espace disque supplémentaire peut être ajouté si l'espace disque actuel ne suffit plus.