Chapitre 3
Fabric réseau

1. Considérations préalables

Dans ce chapitre nous discuterons des topologies physiques et des protocoles des centres de données, appelées ***fabrics*** (tissus) en raison de la complexité de l'infrastructure réseau qui relie les différents composants du centre de données. Ce terme est devenu courant pendant les années 2010, mais, puisque des topologies traditionnelles sont encore présentes dans certains contextes, nous allons commencer par les topologies de la fin du XX^{e} siècle.

Au fil du temps plusieurs producteurs ont proposé des solutions différentes. Pas toutes ont survécu le passage de la théorie au déploiement dans de conditions réelles, toutefois les échecs ont fourni l'expérience nécessaire pour atteindre le niveau actuel.

Pour ces raisons, nous avons abordé le problème avec une approche pragmatique, en évitant des protocoles très intéressants, mais qui sont plutôt orientés vers les opérateurs télécoms que vers les centres de données. Dans la première partie, nous détaillerons les technologies de base de la couche 2, qui datent des années 1980 et 1990 et qui étaient adaptées aux réseaux de production de l'époque. Ces technologies n'ont pas disparu, car elles sont encore valables pour la plupart des réseaux de bureau et pour les réseaux de gestion des centres de données.

Malgré de nombreux efforts et des améliorations continues, nous expliquerons les contraintes encore insurmontables de cette couche en montrant un exemple de topologie physique et logique afin d'en mettre en évidence les limites pour les environnements des grands centres de données. Puis, nous passerons principalement aux technologies de la couche 3, en expliquant les avantages qu'elles introduisent. Même si le futur est orienté vers la couche réseau, plusieurs organisations sont encore obligées d'utiliser certaines technologies de couche 2, et par conséquent, dans le chapitre Sécurité des réseaux de centre de données, nous montrerons une alternative qui s'appuie sur la superposition pour virtualiser les réseaux de couche 2.

2. Modèle physique à 3-tiers et protocoles de couche 2

Au début, les applications étaient développées comme des entités monolithiques, dont toutes les fonctionnalités et tous les services étaient regroupés et fonctionnaient comme service unique, généralement hébergé sur des serveurs centralisés appelés *mainframes*. Au fil du temps, les architectures fondées sur les ordinateurs centraux ont été remplacées par le modèle client-serveur, dans lequel la logique de l'application est répartie entre un serveur (qui gère généralement le traitement et le stockage des données) et un client (généralement le terminal personnel de l'utilisateur).

Pendant les années 1990, l'architecture des réseaux et des applications a convergé vers le modèle **3-tiers**. En conséquence, les applications ont été divisées en trois composants : la présentation, qui assure l'interface utilisateur, le traitement, qui exécute les opérations logiques, et l'accès aux données, où les informations sont stockées. En parallèle, les réseaux ont été distribués dans les niveaux ***core*** (cœur), ***distribution*** (agrégation) et ***access*** (accès). À ce moment-là, les architectes réseau se concentraient principalement sur les flux nord-sud, c'est-à-dire le trafic entre l'Internet et l'intérieur du centre de données.

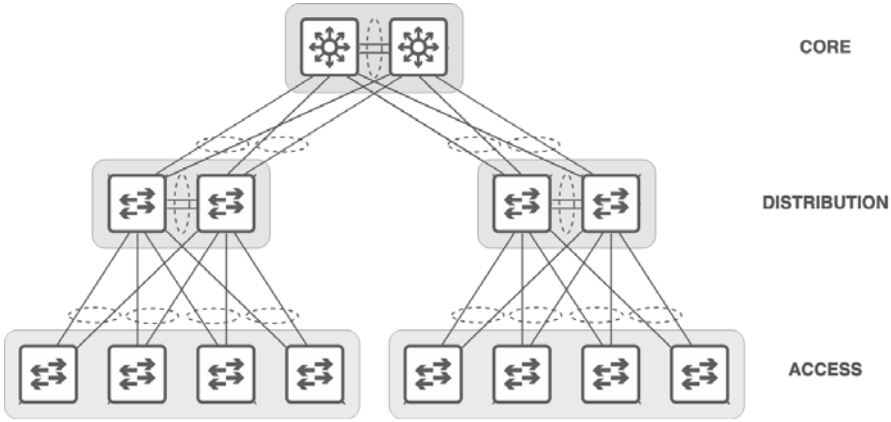
Par rapport au modèle précédent, limité à deux niveaux, ce modèle offre une approche structurée de la gestion du flux de données tout en garantissant que le réseau peut s'étendre plus simplement. Cette conception facilite la redondance, améliore les performances et garantit un comportement plus prévisible du réseau.

Avec les progrès technologiques et l'évolution des besoins des centres de données, des modèles différents ont été adoptés. Pourtant, la topologie 3-tiers est toujours présente dans certains centres de données.

2.1 Architecture physique

Les interconnexions entre les couches mentionnées de *core*, *distribution* et *access* impliquent des chemins redondants pour éviter les points de défaillance uniques, garantissant que même si un chemin ou un dispositif devient indisponible, le réseau reste opérationnel. Cette topologie ressemble à une pyramide, au sommet de laquelle se trouve une paire de commutateurs cœurs connectés entre eux, généralement avec deux connexions physiques redondantes.

Au niveau de la couche d'agrégation, il y a plusieurs paires de commutateurs redondants, également connectés via LACP (entre les commutateurs au même niveau pour l'agrégation et le cœur) et appariés via MLAG (entre les commutateurs de niveaux différents, pour toutes les couches). Les commutateurs des couches centrale et d'agrégation sont généralement situés dans les *racks* (baies) réseau à l'extrémité de la rangée de la cage. Il s'agit généralement de commutateurs de châssis, avec souvent la possibilité d'ajouter de nouvelles lames au fur et à mesure de l'expansion de la topologie. Les commutateurs des différentes couches sont connectés dans une topologie maillée complète redondante.



Exemple simplifié de topologie physique 3-tiers

La figure ci-dessus ne montre pas simplement un exemple de topologie de réseau de production traditionnelle. Vu que le trafic de gestion est quasi exclusivement nord-sud, cette topologie est encore valable pour son trafic. Il suffit d'utiliser des commutateurs moins chers.

2.1.1 Access layer

L'*access layer* (couche d'accès) est formée par l'ensemble des commutateurs ToR qui connectent les serveurs au réseau. Les VLAN isolent le trafic en fonction des exigences fonctionnelles. Si nécessaire, la couche d'accès gère les politiques liées à la hiérarchisation du trafic.

Ces commutateurs se connectent en amont à la couche de distribution ou d'agrégation. Des fonctionnalités telles que le LACP permettent de combiner plusieurs liaisons physiques en une seule liaison logique, afin d'améliorer la bande passante et de garantir que les appareils peuvent toujours se connecter au réseau en cas de défaillance d'une voie.

2.1.2 Distribution layer

La *distribution layer* (couche d'agrégation) sert de pont entre la couche centrale et la couche d'accès, introduisant potentiellement un premier niveau de mise en œuvre des politiques réseau. Ce rôle devient particulièrement important dans les situations où il y a beaucoup de communications entre des serveurs situés dans des racks différents mais dans le même VLAN, ce qui allège la charge sur le réseau central, qui reste libre pour le routage inter-réseaux à grande vitesse.

En établissant des voies multiples et en exploitant le LACP, qui s'adapte dynamiquement aux changements ou aux défaillances du réseau, cette couche garantit qu'il n'y a pas de point unique de défaillance physique.

2.1.3 Core layer

La *core layer* (couche cœur) sert de colonne vertébrale et se concentre sur le transport rapide des données à travers le réseau. Son rôle principal est de ne jamais tomber en panne et de soutenir une grande quantité de trafic entre les commutateurs de distribution et d'autres parties du réseau telles que les connexions ***peering*** (connexions vers les sociétés partenaires et les opérateurs télécoms majeurs) ou l'Internet au sens large.

Le cœur de réseau doit être conçu en tenant compte à la fois de la vitesse et de la résilience. En déployant des dispositifs et des chemins redondants, il assure qu'il n'y a pas de point de défaillance unique, ce qui permet de maintenir le flux de données même en cas de dysfonctionnement du matériel.

Le cœur du réseau est moins axé sur les services que sur le transport et doit se concentrer sur la livraison des paquets aussi rapidement que possible. Une approche minimaliste au niveau des fonctionnalités avancées (telles que les ACL ou des politiques de routage complexes) permet de réduire le temps de latence et d'assurer un transfert de données à grande vitesse.

En matière de matériel, les commutateurs et les routeurs déployés au cœur du réseau sont généralement les plus robustes et les plus puissants dans l'environnement du centre de données. Ces dispositifs sont des châssis modulaires de grande taille qui peuvent être étendus en cas de besoin, optimisant ainsi l'investissement initial et permettant des extensions au fur et à mesure. Ces dispositifs sont souvent équipés de fonctions avancées de redondance, telles que des alimentations doubles et des composants remplaçables à chaud, ce qui renforce l'accent mis sur la disponibilité continue du service.

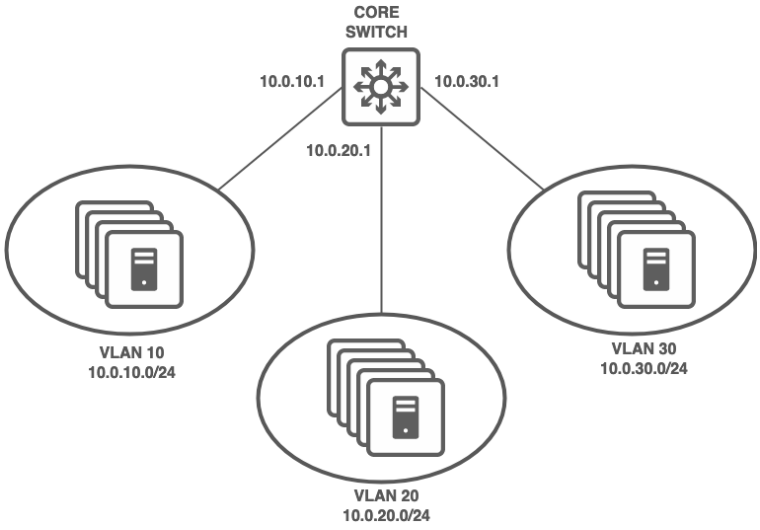
Le protocole de routage choisi pour le cœur de réseau est essentiel pour ses performances. Dans ce type d'architecture, le protocole OSPF est choisi dans la plupart des cas pour garantir qu'en cas de défaillance, le réseau peut rapidement se reconfigurer pour acheminer les données via un chemin alternatif.

2.2 SVI

Pour permettre la communication d'un VLAN vers l'extérieur, il faut une interface virtuelle (SVI, *Switch Virtual Interface*) dotée d'une adresse IP. Vu qu'elle n'est pas liée à une interface physique spécifique, l'état opérationnel d'une SVI reste actif tant qu'il existe au moins un port opérationnel dans le VLAN associé. Cela signifie que même si des liens physiques individuels rencontrent des problèmes, la connectivité logique fournie par la SVI reste intacte.

Du point de vue de la configuration, les SVI prennent en charge toutes les caractéristiques d'une interface physique L3 classique, ne se limitant pas à l'adressage IP, mais incluant les listes de contrôle d'accès et la configuration du protocole de routage, si nécessaire.

Normalement, les SVI sont configurés sur la couche centrale, de manière à ce que le cœur prenne essentiellement en charge le rôle de routage inter-VLAN.



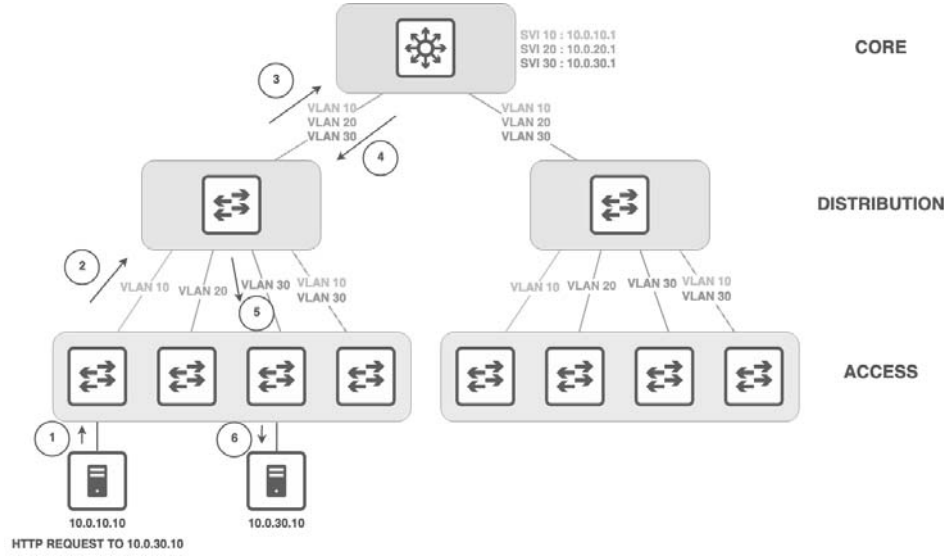
SVI configurées sur le commutateur cœur

Remarque

Notez la différence avec la topologie physique, visible dans la première image de ce chapitre.

Il convient de noter que ce type de conception, où les VLAN s'étendent sur plusieurs blocs de distribution et où les SVI sont centralisées dans le cœur, présente des inconvénients. Il peut introduire un risque de *tromboning* du trafic, c'est-à-dire que le trafic peut remonter inutilement jusqu'au cœur et redescendre, même si la source et la destination se trouvent dans le même bloc de commutateurs de distribution. Les meilleures pratiques suggèrent de limiter le rayon d'action et le domaine de défaillance en gardant les VLAN localisés à la couche d'agrégation, si possible.

Examinons l'image suivante : un serveur installé dans le rack 01 et configuré dans le VLAN 10 veut communiquer avec un autre serveur installé dans le rack 03 et configuré dans le VLAN 30. Puisque la source et la destination de la communication se trouvent derrière la même paire de commutateurs d'agrégation, d'un point de vue physique, le trafic n'aurait pas besoin d'effectuer les étapes 3 et 4, mais les SVI respectives sont configurées sur les commutateurs centraux, donc il est forcé à un parcours plus long :



Flux de trafic sous-optimal

2.3 MSTP

2.3.1 MSTI

Comme nous l'avons vu dans le chapitre Introduction aux protocoles de base, le protocole MSTP offre une approche équilibrée pour garantir des topologies sans boucle dans les réseaux Ethernet, en particulier lorsqu'il s'agit d'un grand nombre de VLAN dans une structure de réseau complexe à trois niveaux. Ce protocole gère efficacement les différentes instances de l'arbre de recouvrement et garantit une sélection optimale des chemins pour les différents types de trafic.

Nous savons que le MSTP divise les VLAN en plusieurs instances de *spanning tree* (*Multiple Spanning Tree Instance*, MSTI), permettant à différents groupes de VLAN d'avoir des topologies distinctes, même s'ils fonctionnent sur le même réseau physique. Le MSTP constitue un compromis entre le STP traditionnel et le RSTP (*Rapid Spanning Tree Protocol*). Ce dernier crée une instance séparée pour chaque VLAN, solution théoriquement idéale mais qui gaspille les ressources CPU des commutateurs, de manière proportionnelle au nombre de VLAN. S'adaptant à la réalité de la topologie physique, qui assure une redondance mais ne peut pas offrir un chemin séparé pour chaque VLAN, le MSTP permet d'utiliser tous les chemins physiques disponibles optimisant l'utilisation des ressources.

En matière de conception, vu que les incohérences peuvent entraîner des boucles de réseau et d'autres comportements indésirables, il est important de veiller à ce que les VLAN soient mappés de manière cohérente aux MSTI sur tous les commutateurs. Regrouper les VLAN ayant des modèles de trafic ou des exigences de chemin similaires sous le même MSTI permet de conserver une conception simple et efficace. En plus, il est important que le *root bridge* pour chaque instance soit un des commutateurs où la SVI est configurée.