Avant-propos

Le Cloud Networking apporte une révolution au domaine des réseaux en proposant un changement complet de paradigme. Le Cloud impose une vision complètement différente de l'Internet avec une forte centralisation des contrôles et une virtualisation de l'ensemble des fonctions nécessaires à la vie d'un réseau, la virtualisation consistant à remplacer les équipements matériels par des équipements logiciels qui s'exécutent dans le Cloud. De ce fait, les réseaux physiques sont remplacés par des réseaux logiques qui s'exécutent dans des centres de données plus ou moins éloignés des nœuds du réseau lui-même.

Les grandes catégories de Clouds, possédant des centres de données allant de l'infiniment grand à l'infiniment petit, sont examinées et décrites en détail. Les centres de données situés à moins de dix kilomètres forment l'Edge et permettent de prendre en charge des applications temps réel avec des latences de moins d'une milliseconde. Cet ensemble de centres de données se concrétise en formant le Cloud Continuum qui devient l'environnement de base du Cloud Networking. Les équipements logiques qui remplacent les équipements physiques doivent être urbanisés dans cet environnement, c'est-à-dire positionnés dans le Cloud Continuum au meilleur emplacement pour que le réseau fonctionne au mieux.

Cet ouvrage introduit ensuite l'infrastructure numérique des années 2020, composée de trois éléments : une antenne qui permet de récupérer les signaux provenant des utilisateurs, une fibre optique transportant le signal et le centre de données qui reçoit ce signal, le traite et exécute l'ensemble des fonctions demandées par l'utilisateur. Tous les boîtiers et les équipements intermédiaires disparaissent physiquement pour réapparaître en tant que machines virtuelles. De ce fait, il est possible d'ajouter simplement de nombreuses fonctions comme le contrôle, la gestion, la sécurité, l'intelligence, l'automatisation, etc.

Une autre caractéristique provient de l'utilisation de logiciels open source, ce qui semble aller de soi puisque toute l'astuce de cette nouvelle génération du Cloud Networking est de pouvoir diminuer les coûts malgré l'augmentation des débits des utilisateurs qui

Cloud et Edge Networking,
par Kamel HADDADOU et Guy PUJOLLE. © ISTE Editions 2024.

doublent chaque année. L'agilité et la flexibilité de cette approche en font une solution incomparable qui est largement introduite dans cet ouvrage.

Cette nouvelle génération de réseaux logiciels se traduit en un certain nombre de produits, décrits en détail, dont le SD-WAN qui forme la principale demande des grandes entreprises avec le vCPE (virtual Customer Premises Equipment) et les fabrics d'accès aux centres de données. L'impact du Cloud Networking est tout aussi important dans les réseaux d'opérateurs et les fournisseurs de réseaux. Il est à la base de la 5G et forme même l'aspect révolutionnaire de cette génération. En effet, ce qui est appelé classiquement la 5G concerne la partie radio mais ce n'est pas là que se trouve l'essentiel. La partie révolutionnaire concerne les centres de données MEC (Multi-access Edge Computing) qui se trouvent sur le bord de l'Edge avec des temps de réponse permettant de prendre en charge tout un ensemble de nouveaux services temps réel comme l'automatisation des réseaux véhiculaires, l'industrie 4.0, les réseaux tactiles, etc.

Le SDN joue un rôle particulier dans le Cloud Networking, il apporte un contrôle automatisé de l'infrastructure numérique grâce justement à la centralisation. Cependant, cette solution ne s'est pas encore complètement imposée par son aspect trop disruptif, en apportant simultanément le contrôle automatisé, des équipements d'une nouvelle génération par la migration de l'intelligence vers le contrôleur central mais avec une centralisation qui peut paraître trop exacerbée.

Enfin, nous introduisons dans cet ouvrage un ensemble de points importants comme la sécurité, la fiabilité, l'intelligence et les accélérateurs. Il se termine par une vision de ce que le Cloud Networking pourrait devenir en particulier avec la 6G. En effet, un retour au matériel est plus que probable pour à la fois améliorer les performances et consommer beaucoup moins d'énergie.

Nous espérons pouvoir remplir les attentes de tous ceux qui s'intéressent au Cloud Networking dans une vision de relativement haut niveau avec l'ensemble des éléments nécessaires pour bien comprendre le cheminement de cette technologie vers la 6G.

1

Introduction au Cloud et à l'Edge Networking

1.1. Introduction à l'infrastructure numérique

Pour les dix années qui viennent, l'infrastructure numérique provenant du Cloud Networking s'est imposée comme le standard. Ce standard est repris par l'ensemble des équipementiers réseau et télécommunications. Elle consiste en quatre éléments : un équipement terminal, une antenne, une fibre optique et un centre de données. Pour comprendre les raisons qui ont mené à cette architecture, il faut démarrer par l'élément de base : la virtualisation.

Le processus de virtualisation est décrit à la figure 1.1. Il provient du passage d'une machine physique à une machine logique. La première étape est d'écrire un code qui réalise exactement la même chose que la machine physique. En supposant que la machine physique est un routeur, le code du routeur virtuel doit réaliser le même routage et envoyer le paquet entrant et traité par le code logique sur la même ligne de sortie que ne le ferait la machine physique.

L'étape suivante est de comparer les performances de la machine physique et de la machine logique en l'exécutant sur le processeur de la machine physique. Sans matériel accélérateur comme des ASIC ou des FPGA, les performances vont facilement tomber par un facteur d'au moins 10 pouvant atteindre 100. Si nous supposons cette perte d'un facteur 20, il faudrait un processeur vingt fois plus puissant pour atteindre la même performance, ce qui n'est pas un problème avec la puissance des centres de données. Cependant, la consommation énergétique étant très approximativement proportionnelle à la puissance du processeur, elle bondit à un niveau élevé.

Cloud et Edge Networking,
par Kamel HADDADOU et Guy PUJOLLE. © ISTE Editions 2024.

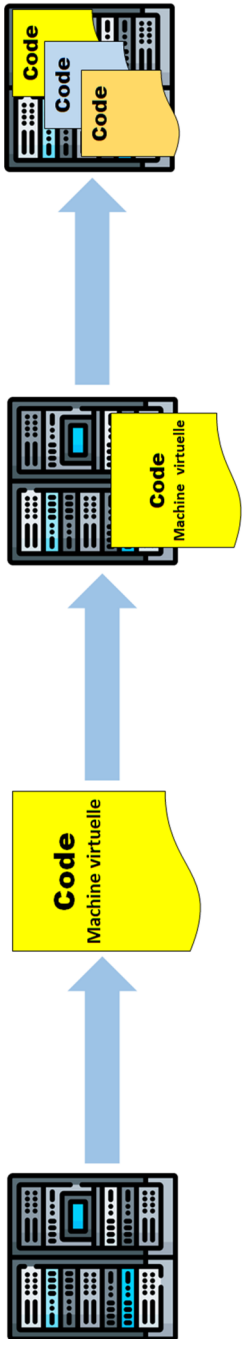

Figure 1.1. *Le processus de virtualisation*

L'étape suivante est d'essayer de minimiser la dépense énergétique. Pour cela, il faut que le processeur de la machine physique supportant la machine logique soit occupé au plus près de 100 %. Comme cela n'est pas vraiment possible, il faut essayer de rester dans les environs de 80 %. Pour y arriver, il faut multiplexer un nombre de machines virtuelles suffisant pour atteindre une très bonne utilisation de l'unité centrale.

La solution utilisée est de regrouper les machines virtuelles pour qu'il y en ait exactement le bon nombre. Si la demande est trop forte, il faut faire migrer des machines virtuelles vers d'autres serveurs et *vice versa* pour maintenir une forte utilisation du processeur. On voit également sur la figure 1.1 que l'utilisation des centres de données est la solution puisque les nombreux serveurs sont soit mis en mode veille s'ils ne sont pas utilisés, soit ils tournent avec un fort taux d'utilisation. L'optimisation de l'énergie consommée est donc réalisée en faisant migrer les machines virtuelles pour que tous les serveurs qui ne sont pas en mode veille soient fortement utilisés. Les migrations de machines virtuelles, c'est-à-dire les déplacements des machines virtuelles d'un serveur à un autre, sont en très grande majorité réalisées dans un même centre de données et beaucoup plus rarement entre centres de données distincts.

La figure 1.2 montre un centre de données avec ses machines virtuelles. Comme indiqué, il y a des migrations continuelles pour optimiser le fonctionnement. Il faut également être capable de donner aux machines virtuelles la puissance dont elles ont besoin pour effectuer la tâche demandée. Pour cela, il faut un orchestrateur des ressources du centre de données qui sont allouées aux machines virtuelles.

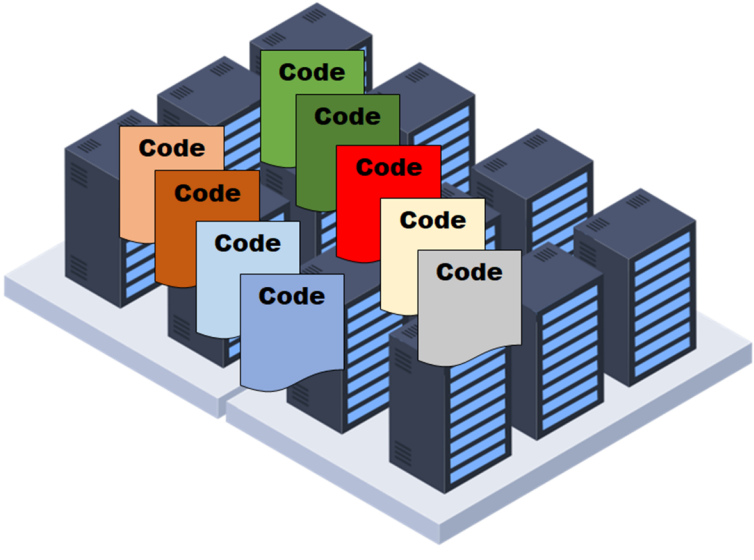

Figure 1.2. *Un centre de données et ses machines virtuelles*

Cette virtualisation logicielle devrait petit à petit être remplacée par une virtualisation matérielle grâce à des processeurs reconfigurables, mais il faudra de nombreuses années avant l'arrivée de cette nouvelle génération qui permettra de consommer beaucoup moins d'énergie et d'augmenter fortement les performances.

On peut se poser la question de savoir quels sont les éléments physiques qui peuvent être virtualisés et ceux qui ne le peuvent pas. En fait, il vaut mieux se pencher sur la deuxième partie de la question puisque tout est virtualisable sauf globalement trois éléments : les capteurs et les cartes de communications hertziennes et filaires. Les capteurs ne sont pas virtualisables car ils doivent capter quelque chose, ce ne peut être réalisé par un code. Par exemple, on ne peut pas mesurer la température dans une pièce en écrivant un code. De même, on ne peut pas capter un signal électromagnétique avec un code ni envoyer une lumière sur une fibre optique toujours par un code. Sinon, tout est virtualisable, un boîtier Wi-Fi, un firewall, une clé, un commutateur, etc.

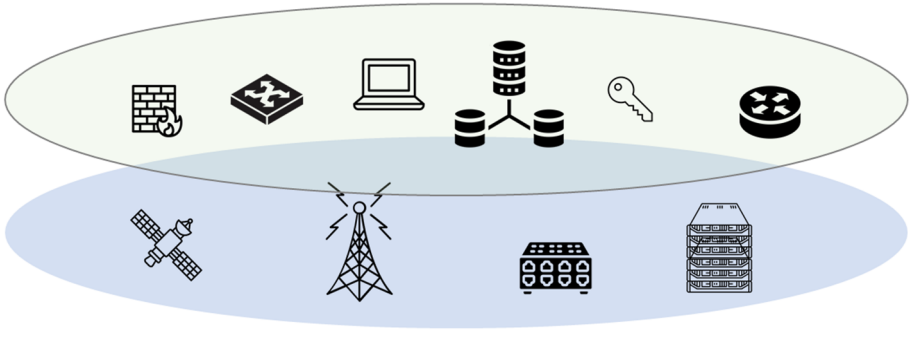

Figure 1.3. *Les équipements virtualisables en haut et non virtualisables en bas*

Le Cloud Networking est justement la solution réseau qui utilise l'infrastructure numérique qui a été décrite au début de ce chapitre, c'est-à-dire à base de quatre éléments, l'équipement terminal, l'antenne, la fibre optique et le centre de données. Nous allons commencer par décrire quelques types de Clouds et leur importance.

Le Cloud, c'est avant tout un mécanisme qui consiste à regrouper les ressources d'une entreprise dans l'Internet plutôt que de les avoir directement dans la société, pour les partager avec d'autres utilisateurs et bénéficier d'un fort multiplexage des ressources et donc d'un coût réduit. Les fournisseurs de Clouds, eux, gagnent également sur le multiplexage en vendant des ressources communes à des utilisateurs qui peuvent se trouver sur des continents différents.

Au début des années 2000, l'utilisation des ressources matérielles, logicielles et de personnels n'était pas optimisée puisque ces ressources n'étaient fortement utilisées qu'aux

quelques heures de pointe et quasiment pas du tout la nuit. Les calculs de moyenne d'utilisation montraient que les ressources étaient utilisées à moins de 20 %. En connectant sur les mêmes ressources communes plusieurs entreprises à des heures de pointe différentes, il est possible d'arriver à des taux d'utilisations de l'ordre de 80 % sans augmenter les ressources.

Le problème qui s'est tout de suite posé concerne les données des entreprises qui se trouvent dans un Cloud public et qui sont de ce fait souvent à la merci d'attaquants ou d'États demandant à leurs fournisseurs des informations pour des raisons de cybersécurité.

Les Clouds privés se sont démocratisés pour tenir compte de cette problématique et sont devenus majoritaires. Ils regroupent en plusieurs points les données et les traitements des grandes entreprises qui ont de nombreux sites ou même des entreprises qui n'ont qu'un site mais ont des départements indépendants.

Aujourd'hui, on trouve différents types de Clouds qui se sont complexifiés pour tenir compte de nouveaux paramètres de diversification et de disponibilité nécessaires aux entreprises.

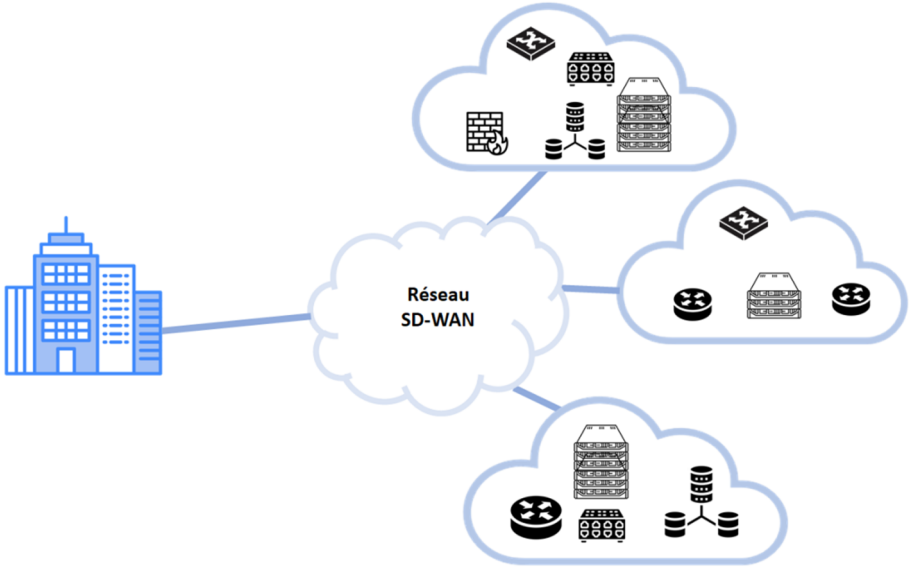

Figure 1.4. *Un Cloud distribué*

Le premier type concerne les Clouds distribués. Comme l'indique la figure 1.4, il s'agit de différents types de Clouds proposés par un même fournisseur : public, privé, près de l'utilisateur, sur la bordure que nous appellerons Edge, abréviation d'Edge Cloud, le

mot anglais étant ici utilisé partout dans le monde, ou bien encore dans le cœur (*core*) du réseau Internet mais beaucoup plus loin de l'utilisateur, que nous appellerons Core Cloud, qui sera abrégé par le seul mot Cloud.

Le fournisseur d'Edge (centre de données sur le bord du réseau cœur) ou de Cloud (centre de données à l'intérieur du réseau cœur) peut proposer plusieurs types de services que nous détaillerons plus loin : une infrastructure, une plate-forme ou directement un logiciel applicatif.

Un autre terme fortement employé est Cloud hybride dans lequel les centres de données peuvent être à la fois privés et publics mais pouvant provenir de fournisseurs de Clouds différents. Le Cloud hybride est donc une solution qui associe un Cloud privé, un ou plusieurs services de Cloud public et un logiciel souvent propriétaire permettant la communication entre chaque service. En optant pour une stratégie de Cloud hybride, les entreprises obtiennent une plus grande flexibilité en déplaçant les charges entre les différents fournisseurs de Cloud Computing au fur et à mesure des besoins et des coûts qui peuvent varier rapidement.

Une illustration de ce type de Cloud est apportée par la figure 1.5 où l'on voit la connexion des deux Clouds réalisée par des accès à de multiples applications portées par la partie publique ou privée.

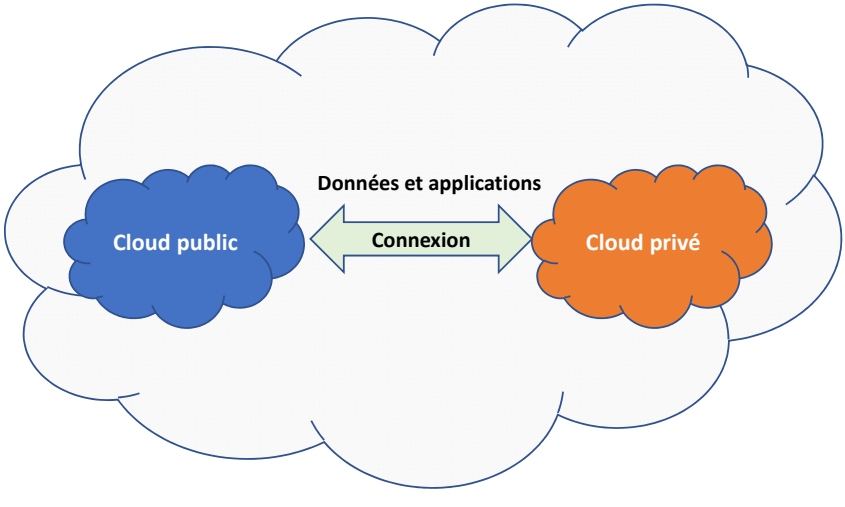

Figure 1.5. *Un Cloud hybride*

D'autres types ont encore été définis comme le multicloud qui rassemble plusieurs fournisseurs pour prendre en charge l'ensemble des services demandés par les entreprises

et qui permet une meilleure disponibilité en cas de surcharge d'un des Clouds de l'environnement. Ces multiclouds regroupent à la fois des Clouds publics et privés et différents types de services, de plates-formes, d'infrastructures et d'applications.

Enfin, le terme omnicloud est le plus général pour prendre en compte la multitude de possibilités d'associations et de structures de Clouds.

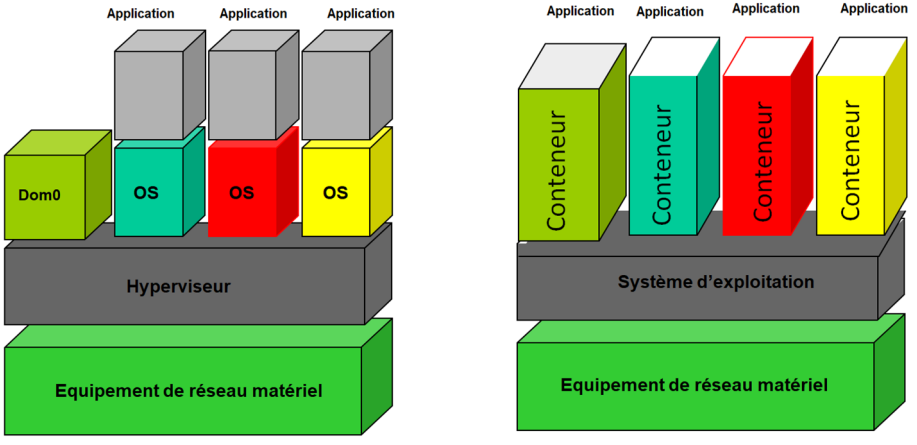

Figure 1.6. *L'hypervision et la conteneurisation*

La figure 1.6 décrit l'architecture interne des serveurs dans un centre de données. Deux grandes possibilités : l'hypervision et la conteneurisation. La première est la plus ancienne, elle concerne le support des machines virtuelles tel qu'il a été conçu à l'origine. La seconde solution remplace petit à petit la première par une architecture plus simple, moins coûteuse et plus flexible.

L'hypervision consiste à utiliser un hyperviseur sur une machine physique standard (*commodity*) qui est un logiciel capable de supporter plusieurs machines virtuelles simultanément par l'intermédiaire d'un ou de plusieurs systèmes d'exploitation (OS, *Operating System*). L'hyperviseur prend en charge des domaines formés d'un système d'exploitation et de la machine virtuelle qui s'exécute dessus. Le domaine 0 ou Dom0 est spécialisé dans le traitement des entrées-sorties des autres domaines sur la machine physique de base.

Il existe différents types d'hyperviseurs. La paravirtualisation demande à ce que les systèmes d'exploitation soient légèrement modifiés pour que tous les traitements demandés par la machine virtuelle puissent se faire nativement sur la machine physique de base. Au contraire, la deuxième solution est d'accepter les systèmes d'exploitation sans modification mais avec l'introduction, au-dessus de l'hyperviseur, d'un logiciel d'émulation capable d'adapter l'exécution de certaines fonctions à la machine physique sous-jacente.

La conteneurisation remplace petit à petit l'hypervision par un découpage des services en microservices qui s'exécutent chacun dans un conteneur. Dans ce cas, on utilise un système d'exploitation unique qui supporte des conteneurs qui sont isolés les uns des autres pour éviter le « jumping » permettant à un utilisateur de passer d'un conteneur à un autre. Chaque microservice s'exécute dans un conteneur et des interfaces applicatives entre microservices permettent de réaliser le service lui-même. Nous étudierons plus en détail dans la suite cette technologie de microservices qui permet de réaliser des mises à jour plus simples, sans arrêter complètement le service, et qui permet également le développement de façon simplifiée des services.

Cette technologie de microservices commence elle-même à être remplacée par une solution à base de fonctions que nous étudierons en détail dans la suite consistant à réaliser des services avec une succession de fonctions. Cette dernière solution est appelée « serverless » pour indiquer que le programmeur qui développe un service par fonctions n'a plus du tout conscience qu'il y a des serveurs sous-jacents.

1.2. Les services Clouds

La figure 1.7 explicite les trois types majeurs de Clouds qui sont complétés par deux nouveaux qui se situent entre les solutions indiquées par la figure 1.7.

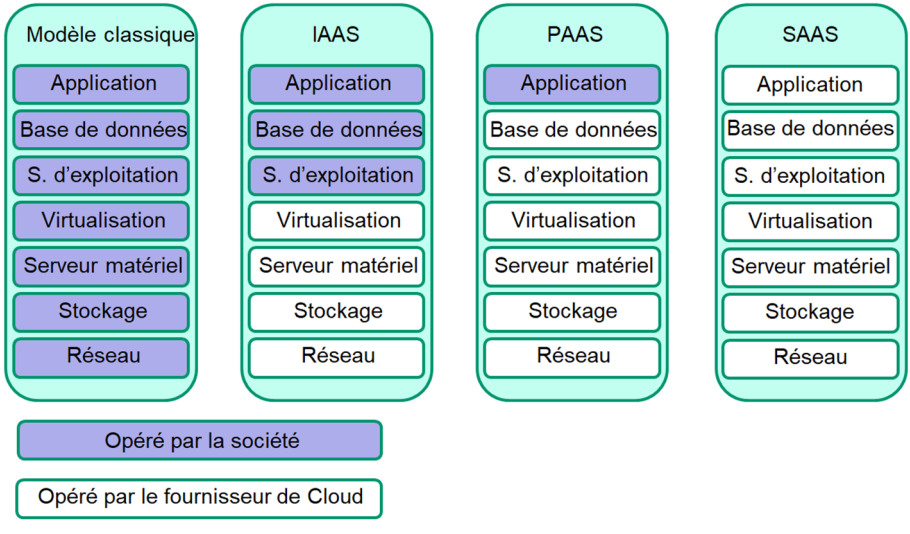

Figure 1.7. *Les trois grands types de Clouds*

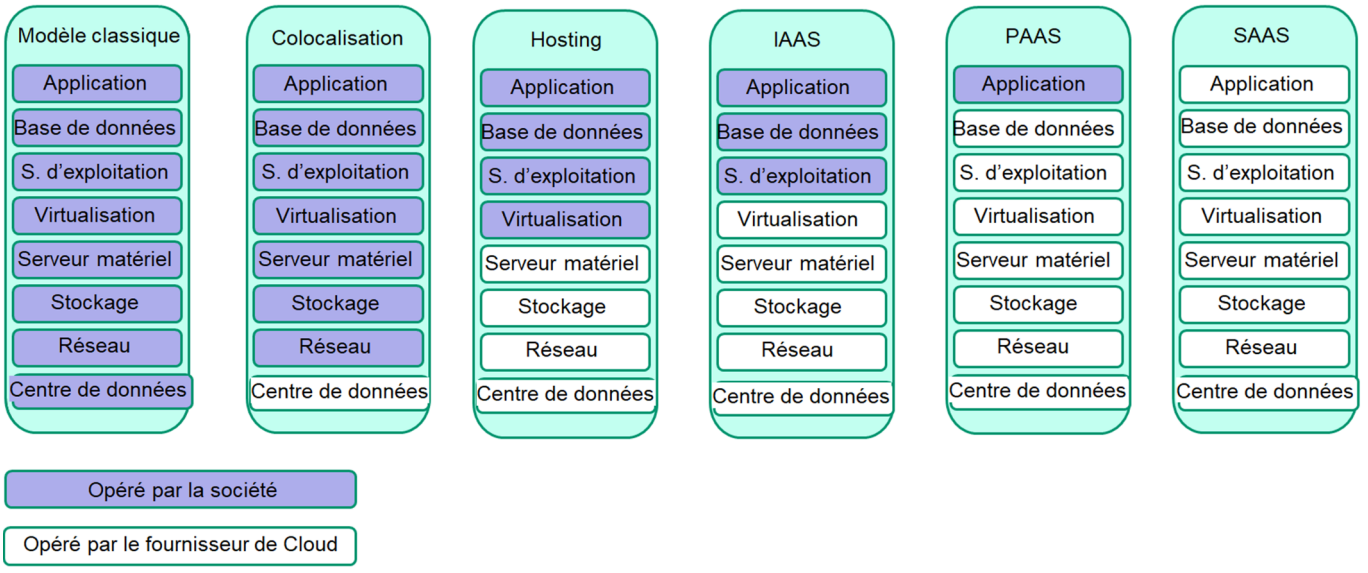

Figure 1.8. *Les premières architectures de services*

La première pile protocolaire de gauche représente le cas où une entreprise possède l'ensemble des ressources pour faire tourner son système d'information. Parmi ces ressources, il faut les éléments de réseau, de stockage et de calcul sur des serveurs matériels. Il faut y ajouter la virtualisation et le système d'exploitation qui prennent en charge les données et les applications de l'entreprise.

En utilisant un fournisseur d'IaaS (Infrastructure as a Service), celui-ci fournit les couches basses correspondant au réseau, au stockage et au calcul avec l'environnement de virtualisation et l'entreprise s'occupe des couches supérieures. Dans la cadre du PaaS (Platform as a Service), le fournisseur s'occupe de tout sauf de l'application et enfin pour le SaaS (Software as a Service), le fournisseur procure l'ensemble de toutes les couches y compris les applications. On peut donner comme exemple de ce dernier cas de figure, le service Office 365 de Microsoft. Le SaaS représente approximativement 50 % des Clouds installés, les deux autres se partageant le reste.

Nous montrons à la figure 1.8 des cas plus classiques utilisés depuis longtemps et remplacés petit à petit par les trois cas que nous avons décrits ci-avant. Il s'agit principalement de la colocalisation et de l'hosting.

Maintenant que nous avons introduit la virtualisation et le Cloud, nous allons pouvoir entrer plus avant dans le Cloud Networking.

1.3. Cloud Networking

Le Cloud Networking regroupe des technologies réseau que l'on obtient à partir du Cloud et donc de la virtualisation. L'infrastructure numérique en est la base : toutes les machines physiques, tous les services de l'infrastructure numérique et les services applicatifs forment le Cloud Networking. Dans un premier temps, définissons les réseaux virtuels qui forment la base du Cloud Networking.

La figure 1.9 représente un ensemble de réseaux virtuels qui sont formés de machines virtuelles, des routeurs, des commutateurs, des firewalls, des serveurs d'authentification, etc., pour réaliser toutes les fonctions nécessaires à la bonne marche d'un réseau. Chaque réseau virtuel est formé de ses propres machines virtuelles qui peuvent être très différentes les unes des autres. Un réseau peut être formé de routeurs IPv6, un autre de commutateurs Ethernet, un troisième de LSR (Label Switch Router) que l'on trouve dans les réseaux MPLS et enfin un quatrième qui possède des équipements tout à fait spécifiques et propriétaires. Ces réseaux utilisent les mêmes centres de données et des infrastructures de câbles ou des liaisons hertziennes entre les équipements virtuels. Le nombre de réseaux virtuels, qui peuvent coexister sur l'infrastructure numérique, dépend de la volonté du gestionnaire de l'environnement et du trafic sur chaque réseau. On appelle ces réseaux

virtuels, des « slices » (tranches) et nous utiliserons surtout ce mot dans le cœur des réseaux 5G. Ces différents réseaux virtuels doivent être indépendants et isolés les uns des autres pour éviter qu'une attaque puisse se propager d'un réseau à un autre.

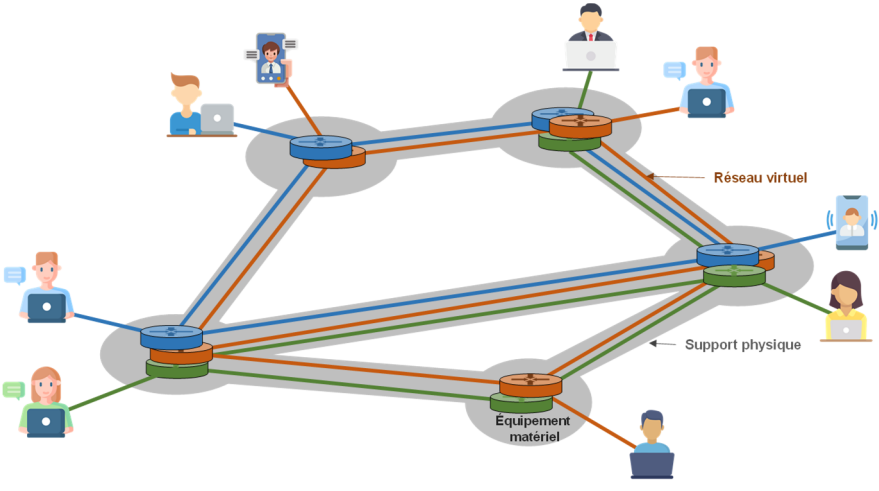

Figure 1.9. *Un ensemble de réseaux virtuels*

La figure 1.10 montre un réseau virtuel réalisé sur des centres de données qui deviennent de ce fait les salles réseau de l'infrastructure numérique. Dans cette figure, en dehors des routeurs ou commutateurs, les équipements ne sont pas virtualisés comme les box Internet chez les utilisateurs ou des boîtiers intermédiaires comme le DPI (Deep Packet Inspection) ou encore le firewall ou un serveur d'authentification.

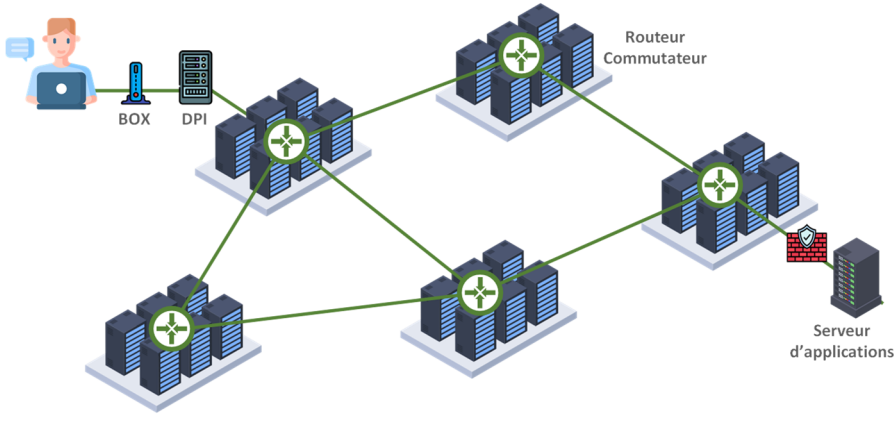

Figure 1.10. *Un réseau virtuel avec des boîtiers non virtualisés*

La figure 1.11 introduit le « slicing », c'est-à-dire comme nous l'avons vu la présence de plusieurs réseaux virtuels sur une même infrastructure physique. Ces réseaux peuvent appartenir à des opérateurs différents, chacun avec sa technologie de transfert de paquets et ses fonctions spécifiques. Le slicing est plus particulièrement introduit dans le réseau cœur du réseau 5GC (5G Core), le réseau qui permet d'interconnecter les antennes entre elles pour permettre le transfert d'informations d'une région à une autre région. Cependant, les slices sont définies de bout en bout, donc se continuent sur les parties accès et radio comme nous le détaillerons au chapitre 8. Au départ, le réseau 5G n'aura qu'une slice pour prendre en charge les connexions des utilisateurs et petit à petit le nombre de slices augmentera pour se spécialiser dans la connexion d'objets, de véhicules, de machines-outils, etc. Ensuite, ces slices pourront devenir des réseaux d'entreprise permettant d'interconnecter les différents sites d'une même entreprise.

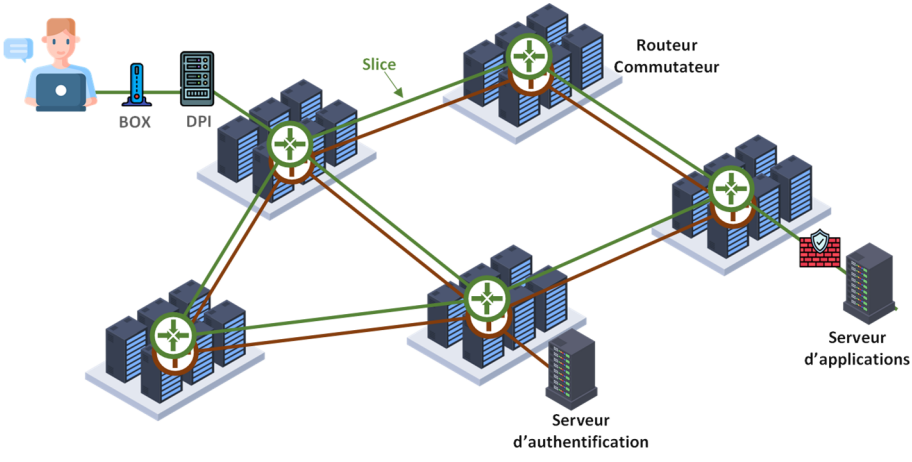

Figure 1.11. *Le slicing*

L'étape explicitée à la figure 1.12 est la virtualisation totale des infrastructures physiques pour permettre l'arrivée de l'infrastructure numérique. Toutes les fonctions sont virtualisées, de celles de la box Internet au DPI en passant par le firewall et le serveur d'authentification pour ne citer que quelques fonctions.

Examinons par exemple le cas de la box Internet qui se met en place chez les opérateurs. L'antenne ne peut pas être virtualisée et donc il faut garder un boîtier qui contiendra l'antenne Wi-Fi et/ou l'antenne 5G. Derrière cette antenne, une fibre optique relie le boîtier de réception des signaux électromagnétiques à un centre de données de l'opérateur. Ces centres sont ceux en cours de déploiement dans le cadre de la 5G. Ils se situeront à

moins de dix kilomètres de l'antenne pour permettre un temps de latence extrêmement court, de l'ordre de la milliseconde. En effet, pour parcourir les vingt kilomètres aller-retour à la vitesse de la lumière sur la fibre optique, il faut 0,1 milliseconde. Avec le temps d'accès à l'antenne et le temps de traitement d'une requête courte, nous sommes dans l'ordre de la milliseconde. Cette valeur correspond au temps de latence maximale pour des applications temps réel comme le contrôle d'un réseau véhiculaire. En effet, entre le début du freinage d'un véhicule et le début du freinage du véhicule suivant, il faut que le délai de la communication passant par le réseau soit le plus court possible. De même, pour le contrôle de robots et de machines-outils, la milliseconde est le temps recommandé. De même, pour réaliser une chirurgie à distance où le chirurgien doit voir ce qu'il fait avec un laps de temps du même ordre de grandeur de la milliseconde. Les fonctions intégrées dans la box Internet sont virtualisées dans le centre de données de l'opérateur. Le nom de ces centres de données des opérateurs 5G est MEC pour « Multi-access Edge Computing », qui a succédé à la première définition, « Mobile Edge Computing », qui ne faisait référence qu'aux réseaux de mobiles alors que la 5G s'intéresse à tous les types de réseaux qu'ils soient fixes ou mobiles. Par exemple, un LAN (Local Area Network) utilisant du Wi-Fi fait partie des systèmes à connecter dans l'univers de la 5G. La fonction de DPI (Deep Packet Inspection), qui analyse les flots bit par bit permettant ainsi de détecter des anomalies en ne reconnaissant pas les signatures de certaines applications, est également virtualisée dans le Cloud. De même, le firewall et le serveur d'authentification sont virtualisés dans un des centres de données MEC de l'opérateur ou éventuellement dans un centre de données d'un fournisseur de Cloud.

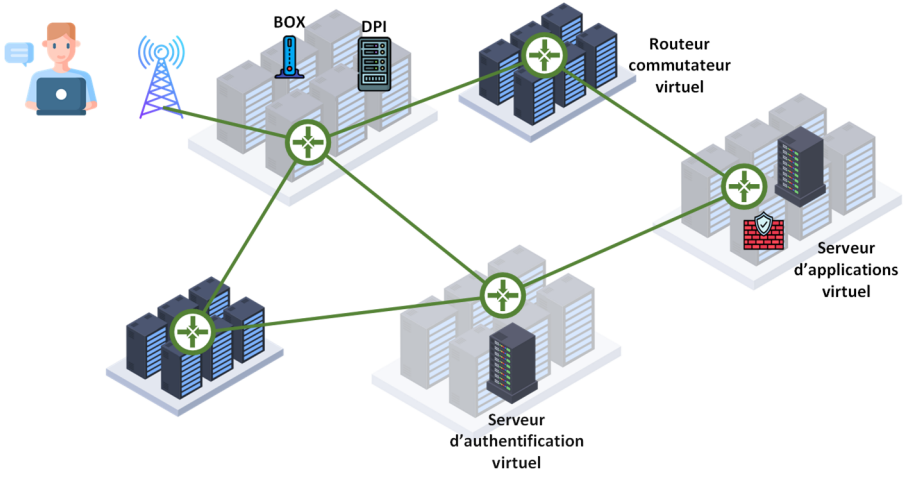

Figure 1.12. *Virtualisation de l'ensemble des fonctions d'un réseau*

Il faut se poser la question du lieu où sont positionnés les machines virtuelles ou les conteneurs. Aujourd'hui, on travaille avec quatre niveaux qui sont représentés à la figure 1.13. Le niveau dénommé Cloud représente les grands centres de données qui sont dans le cœur d'Internet, les core Clouds. Les trois autres niveaux correspondent à l'Edge Cloud qui est abrégé en Edge. L'Edge a lui-même trois niveaux : le MEC, le Fog et l'Edge embarqué (Embedded Edge).

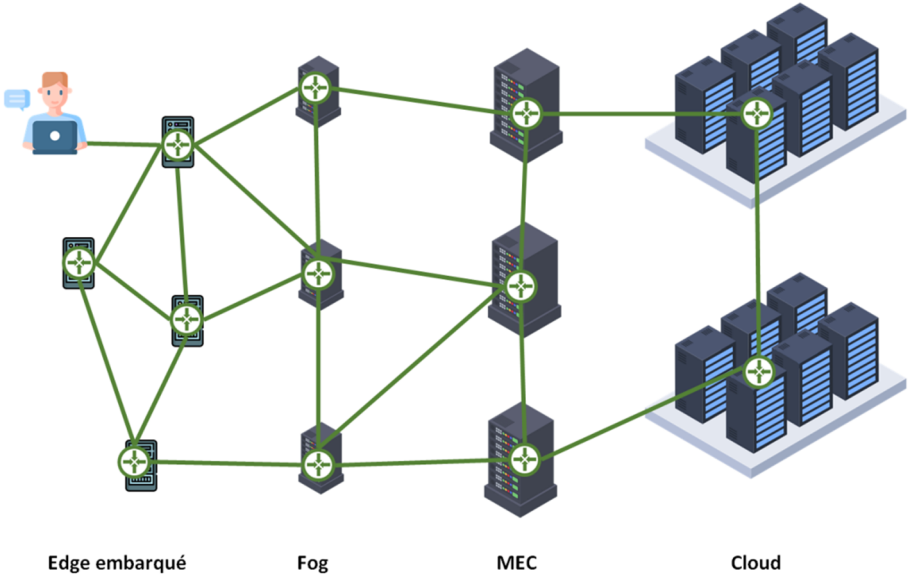

Figure 1.13. *Les quatre niveaux de Cloud et Edge Networking*

Le MEC (Multi-access Edge Computing) est le niveau le plus éloigné du client, spécifié et réalisé dans l'environnement 5G qui introduit ces centres de données pour définir une infrastructure numérique qui sera bâtie sur les centres de données MEC. Les antennes 5G ou les accès terrestres *via* Wi-Fi ou d'autres techniques de LAN sont connectés sur le MEC par l'intermédiaire de fibres optiques. Les centres de données MEC sont destinés à recevoir toutes les machines virtuelles des équipements physiques et des logiciels situés entre l'équipement terminal et le centre de données : le traitement de signal, les fonctions de localisation, les algorithmes de contrôle et de gestion, le pilote automatique de l'environnement, les processus d'intelligence artificielle, les applications métiers, etc. En particulier, on y trouve les applications temps réel avec de fortes contraintes comme par exemple le temps de latence qui doit être de l'ordre de la milliseconde.

La dénomination « Fog Computing » provient de la société Cisco qui en 2012 proposait de connecter les objets en cours d'apparition (capteurs, actionneurs ou autres objets) sur un centre de données intermédiaire pour avoir un prétraitement avant d'acheminer les informations retenues vers un Cloud central pour y être traitées. Ce terme « Fog » est resté et il désigne aujourd'hui les centres de données en entreprise mais plus particulièrement les centres de données sur les campus, très près des utilisateurs, c'est-à-dire quelques centaines de mètres au maximum. L'idée est de remplacer tous les équipements physiques par des machines virtuelles et de rassembler tous les processus métier en tant que machines virtuelles ou conteneurs dans le centre de données Fog. Les temps de latence sont très petits, inférieurs à la milliseconde.

Le troisième niveau porte plusieurs noms possibles comme l'Edge embarqué, que nous utiliserons, ou bien, le Skin, le Mist ou le far Edge. Il se situe tout près de l'utilisateur à portée de Wi-Fi ou de 5G privée qui ont la même portée de quelques dizaines de mètres au maximum puisque ces deux technologies émettent sur les bandes libres avec les mêmes contraintes. Les centres de données sont des calculateurs embarqués de puissance relativement faible dans un premier temps. Cependant, ces équipements embarqués vont être de plus en plus puissants et acceptent les technologies de conteneurisation et de serverless. Les avantages de cette solution sont nombreux : un temps de latence excessivement faible, une sécurisation des données en restant dans le giron de l'utilisateur et une minimisation de l'énergie consommée. L'objectif est également d'avoir un environnement mobile, l'Edge embarqué pouvant se trouver dans un véhicule, dans un objet mobile, dans un smartphone ou dans un équipement spécifique dans une poche de l'utilisateur. L'équipement Edge embarqué est lui-même connecté soit à une antenne plus lointaine comme une antenne 5G d'un opérateur soit à d'autres Edges embarqués pour former un Cloud embarqué. Cette solution permet d'automatiser des réseaux véhiculaires ou bien de prendre en charge le contrôle de robots mobiles et plus globalement de réaliser des environnements mobiles intelligents.

1.4. Network Functions Virtualization (NFV)

La base des technologies qui sont utilisées dans les sections précédentes provient du NFV (Network Functions Virtualization) qui consiste à virtualiser toutes les fonctions des divers boîtiers comme un équipement NAT, un firewall, un DPI, un contrôleur, un routeur, etc. comme ceci est illustré à la figure 1.14.

Le problème du NFV provient de la potentielle non-compatibilité des machines virtuelles entre elles. Du coup, les opérateurs, qui ont été à l'origine de la demande, ont souhaité une normalisation des machines virtuelles pour permettre une interconnexion simple entre opérateurs. La première demande de normalisation a été adressée à l'ETSI, l'organisme de normalisation des télécommunications pour l'Europe et qui à la demande de très nombreuses compagnies du monde entier a ouvert les portes à une normalisation mondiale.

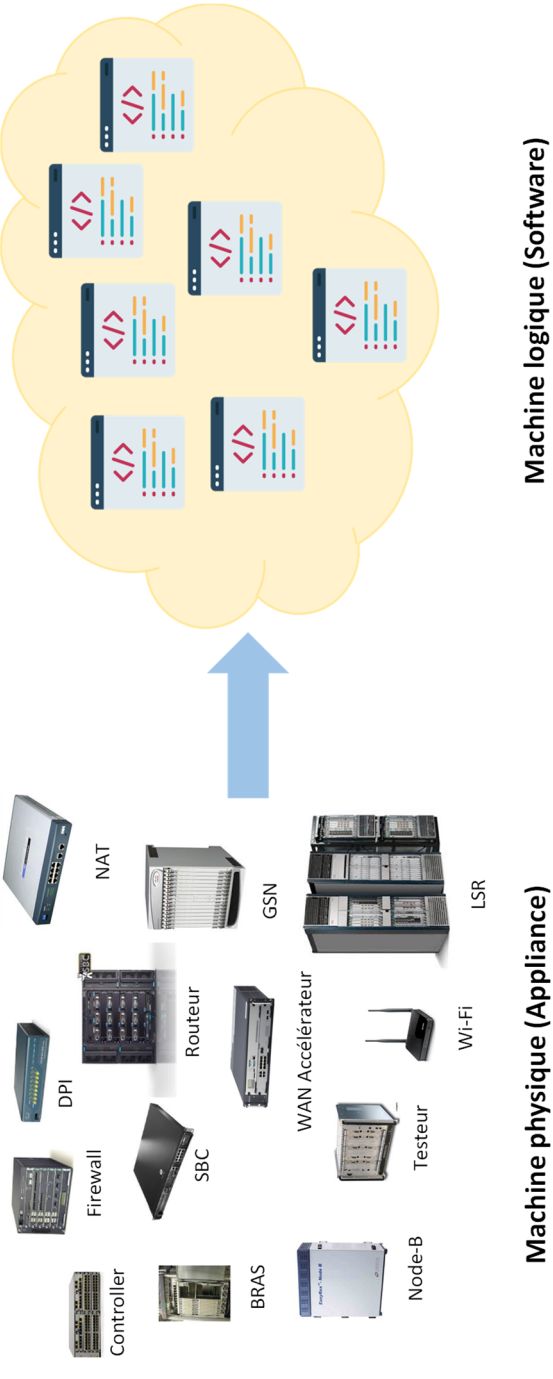

Figure 1.14. *Le NFV (Network Functions Virtualization)*

Pour aller plus loin, l'ETSI s'est adressée à la Linux Foundation pour réaliser un code open source reflétant cette normalisation des machines virtuelles. Mais quelques mois après le démarrage, l'ETSI a souhaité aller plus loin encore en proposant un logiciel open source associé à chaque fonction pour permettre à ces dernières d'interagir entre elles pour réaliser une plate-forme complète et opérationnelle. Cette plate-forme a pris le nom d'OPNFV (Open Platform Network Functions Virtualization) qui représente à sa finalisation au 31 décembre 2021 plus de dix mille personnes × an de travail, c'est-à-dire l'équivalent de dix mille personnes travaillant pendant un an pour réaliser ce logiciel. Le nom final qui lui a été donné est Anuket provenant de la LF-Networking (Linux Foundation Networking). Le projet aura duré six ans avec le développement de releases intermédiaires allant de A à I.

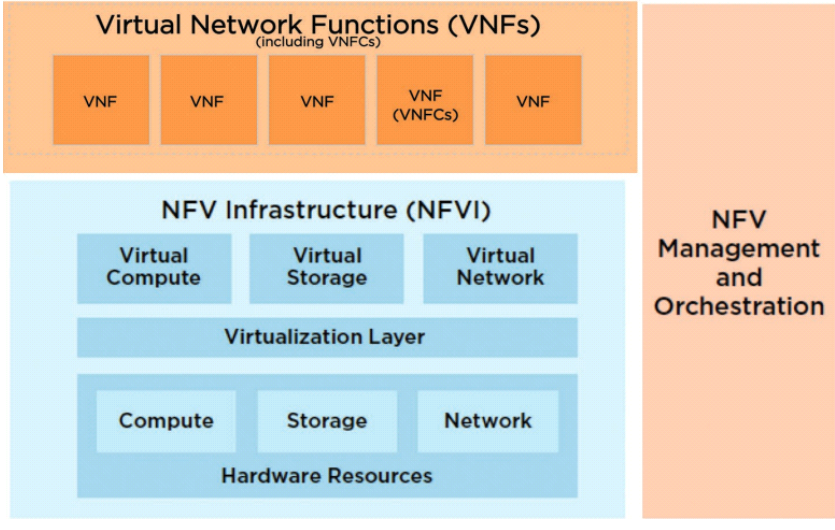

Figure 1.15. *L'architecture de la plate-forme Anuket de la LF-Networking*

On retrouve dans cette plate-forme Anuket de la LF-Networking de nombreux logiciels open source de la Linux Foundation. Le travail principal a été d'agglomérer tous ces logiciels tout en les complétant. La structure générale de cette plate-forme est décrite à la figure 1.15. Elle comporte trois grandes parties :

– NFVI (NFV Infrastructure) qui est l'élément d'infrastructure nécessaire pour exécuter les machines virtuelles non seulement pour le réseau mais également pour le stockage et le calcul ;

– VNF (Virtual Network Function) qui représente toutes les fonctions virtuelles disponibles dans lesquelles va venir piocher le système pour réaliser les services demandés par les utilisateurs ;

– NFV MANO (Management and Orchestration) qui est le pilote automatique de la plate-forme.

L'architecture plus précise de la plate-forme Anuket de la LF-Networking sera décrite au chapitre 4.

1.5. Conclusion

Nous avons vu dans ce premier chapitre l'introduction de l'infrastructure numérique qui simplifie beaucoup les environnements que nous connaissions en réseau par une forte centralisation de toutes les fonctions que ce soit des fonctions d'infrastructure numérique comme le traitement de signal ou le routage, ou des fonctions de service d'infrastructure avec le contrôle, la gestion, l'intelligence, l'automatisation, etc., ou encore des fonctions applicatives correspondant aux grandes applications demandées par les utilisateurs. Ce changement est révolutionnaire et mène au Cloud Networking qui se met en place doucement puisque de nombreuses pièces du puzzle ne sont pas encore vraiment disponibles comme les centres de données MEC ou le slicing.

1.6. Bibliographie

Antonopoulos, N. and Gilla, L. (2017). *Cloud Computing: Principles, Systems and Application*. Springer, New York.

Artasanchez, A. (2021). *AWS for Solutions Architects: Design Your Cloud Infrastructure by Implementing DevOps, Containers, and Amazon Web Services*. Packt Publishing, Birmingham.

Ben Jemaa, F., Pujolle, G., Pariente, M. (2016). Cloudlet- and NFV-based carrier Wi-Fi architecture for a wider range of services. *Annals of Telecommunications*, 71(11–12), 617–624.

Comer, D. (2021). *The Cloud Computing Book: The Future of Computing Explained*. CRC Press, Boca Raton.

Culkin, J. and Zazon, M. (2021). *AWS Cookbook: Recipes for Success on AWS*. O'Reilly, Sebastopol.

Dutt, D. (2019). *Cloud Native Data Center Networking: Architecture, Protocols, and Tools*. O'Reilly, Sebastopol.

Fox, R. and Hao, W. (2017). *Internet Infrastructure: Networking, Web Services, and Cloud Computing*. CRC Press, Boca Raton.

Cette bibliographie est identique à celle de l'ouvrage correspondant en anglais publié par ISTE.

Gessert, F., Wingerath, W., Ritter, N. (2020). *Fast and Scalable Cloud Data Management*. Springer, Cham.

Gray, K. and Nadeau, T.D. (2016). *Network Function Virtualization*. Morgan Kaufmann, Burlington.

Halabi, S. (2019). *Hyperconverged Infrastructure Data Centers: Demystifying HCI*. Cisco Press, Indianapolis.

He, Y., Ren, J., Yu, G., Cai, Y. (2019). D2D communications meet mobile Edge computing for enhanced computation capacity in cellular networks. *IEEE Transactions on Wireless Communications*, 18(3), 1750–1763.

Kraemer, F.A., Braten, A.E., Tamkittikhun, N., Palma, N. (2017). Fog computing in healthcare – A review and discussion. *IEEE Access*, 5, 9206–9222.

Mach, P. and Becvar, Z. (2017). Mobile Edge computing: A survey on architecture and computation offloading. *IEEE Communications Surveys & Tutorials*, 19(3), 1628–1656.

Mao, Y., You, C., Zhang, J., Huang, K., Letaief, K.B. (2017). A survey on mobile Edge computing: The communication perspective. *IEEE Communications Surveys Tutorials*, 19(4), 2322–2358.

Moura, J. and Hutchison, D. (2019). Game theory for multi-access Edge computing: Survey, use cases, and future trends. *IEEE Communications Surveys Tutorials*, 21(1), 260–288.

Mouradian, C., Naboulsi, D., Yangui, C., Glitho, R.H., Morrow, M.J., Polakos, P.A. (2018). A comprehensive survey on fog computing: State-of-the-art and research challenges. *IEEE Communications Surveys Tutorials*, 20(1), 416–464.

Mukherjee, M., Shu, L., Wang, D. (2018). Survey of fog computing: Fundamental, network applications, and research challenges. *IEEE Communications Surveys Tutorials*, 20(3), 1826–1857.

Olaoye, A. (2022). *Beginning DevOps on AWS for iOS Development*. Apress/Springer Nature, Cham.

Perera, C., Qin, Y., Estrell, J.C.A., Reiff-Marganiec, S., Vasilakos, A.V. (2017). Fog computing for sustainable smart cities: A survey. *ACM Computing Surveys*, 50(3), 1–43.

Satyanarayanan, M. (2017). The emergence of Edge computing. *Computer*, 50(1), 30–39.

Shaukat, U., Ahmed, E., Anwar, Z., Xia, F. (2016). Cloudlet deployment in local wireless networks: Motivation, architectures, applications, and open challenges. *Journal of Network and Computer Applications*, 62, 18–40.

Sujata, D., Subhendu, K., Ajith, A., Yulan, L. (2021). *Advanced Soft Computing Techniques in Data Science, IoT and Cloud Computing*. Springer, Cham.

Vaquero, L.M. and Rodero-Merino, L. (2014). Finding your way in the fog: Towards a comprehendsive definition of fog computing. *ACM SIGCOMM Computer Communication Review*, 44(5) 27–32.

Zburivsky, D. and Partnet, L. (2021). *Designing Cloud Data Platforms*. Manning Publications, Shelter Island.