Réseaux virtuels

1. Introduction

Le réseau est un des composants critiques de votre architecture de virtualisation.

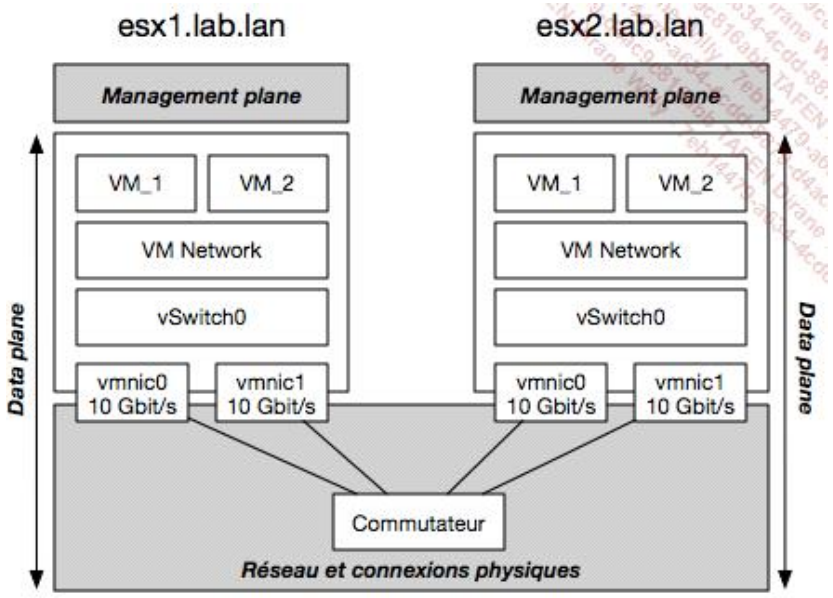
Pour vous en convaincre, imaginez une architecture avec une connectivité réseau défaillante. Vos machines virtuelles ne peuvent pas communiquer. Vos hyperviseurs ne peuvent pas communiquer entre eux et sont incapables d'organiser la haute disponibilité de vos ressources et de vos données.

Ce chapitre sera l'occasion d'aborder les commutateurs virtuels dans leurs différentes formules au sein de vSphere, l'agrégation de liens ainsi que la solution avancée de virtualisation de VMware concernant le réseau et la sécurité, NSX.

2. vSphere Standard Switch

À l'issue de l'installation d'un hôte vSphere, vous disposez par défaut d'un commutateur virtuel local (vSwitch) auquel est reliée l'interface de gestion de l'hyperviseur (VMkernel).

Le schéma ci-dessous décrit comment fonctionne un vSS (*vSphere Standard Switch*).



Dans notre exemple, nous avons ici deux hôtes, esx1.lab.lan et esx2.lab.lan.

Comme vous pouvez le constater, chaque hôte dispose de son propre réseau et connecte les machines virtuelles (VM_1, VM_2) à un réseau local virtuel.

Les machines virtuelles peuvent se connecter à un réseau virtuel via un adaptateur réseau virtualisé (vNIC) installé au sein de la machine virtuelle, lui-même connecté à un groupe de ports, ici « VM Network ».

Enfin, la liaison entre des groupes de ports est assurée par un commutateur, lui aussi virtualisé. Vous remarquerez ici que chaque hyperviseur dispose de son propre commutateur virtuel (vSwitch0).

Si la machine virtuelle « VM_1 » s'exécutant sur esx1.lab.lan souhaite communiquer avec un hôte extérieur ou une machine virtuelle portée par esx2.lab.lan, l'hyperviseur fera appel à ses ports réseau physiques que l'on appelle

également liaisons montantes (*uplink* en anglais). D'ailleurs, ces ports sont caractérisés par leurs noms (vmnic0, vmnic1) et leurs vitesses (10 Gbit/s) pour chacun des ports des hôtes. Vous l'avez deviné, la vitesse de ses ports réseau influence indirectement la performance réseau des machines virtuelles et du système en cas de connexion hors de l'hyperviseur. Mais nous y reviendrons.

Enfin, il est utile de préciser le sens des sections « Management plane » et « Data plane » précisées dans le schéma.

Le « Management plane » permet la gestion, la configuration, la supervision de la couche réseau des hôtes. Ici, concernant le réseau porté par un vSwitch standard, la configuration est exclusivement locale, ce qui veut dire aussi que la configuration réseau (vSwitch, groupes de ports, politiques...) doit être répliquée manuellement ou automatiquement (via des profils d'hôtes pour ce dernier). La gestion dans ce scénario peut être simple vu le nombre d'hôtes à manager. Imaginez maintenant que vous ayez à configurer ou mettre à jour une configuration sur 100 hôtes ou plus. Cette solution n'est pas efficiente et nous verrons qu'il existe une autre solution plus intéressante dans ces cas.

Le « Data plane » décrit la façon dont les communications entre entités s'effectueront et comment l'information va s'échanger sur le réseau. Dans notre cas, cela dépend :

- Si une donnée est envoyée de VM_1 à destination de VM_2 au sein du même hyperviseur, la donnée ne sera pas envoyée sur les liaisons montantes.
- Si une donnée est envoyée de VM_1 à destination de VM_2 hébergé au sein de deux hyperviseurs différents, les liaisons montantes seront sollicitées.
- Si une donnée est envoyée de VM_1 à destination d'un hôte extérieur, les liaisons montantes seront sollicitées.
- Si les deux hyperviseurs souhaitent communiquer entre eux ou avec l'extérieur, les liaisons montantes seront sollicitées.

En outre, les vSwitches standards supportent le protocole CDP (*Cisco Discovery Protocol*) permettant d'envoyer et de recevoir des informations sur les pairs compatibles du même réseau. S'il est activé, les trois options disponibles sont :

- transmission
- réception
- transmission et réception

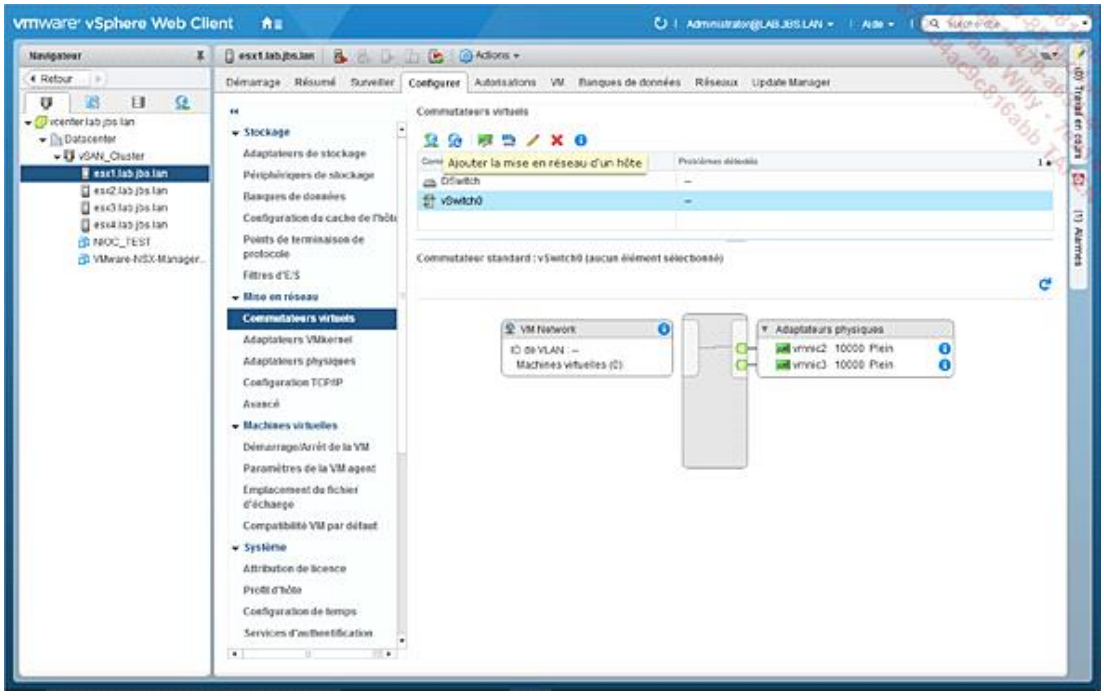
Attention : si le protocole CDP est activé sur le vSwitch, l'hyperviseur envoie les caractéristiques du vSwitch y compris sa version logicielle (qui correspond à la version du serveur hôte) via ce protocole.

Abordons désormais la création d'un vSwitch standard.

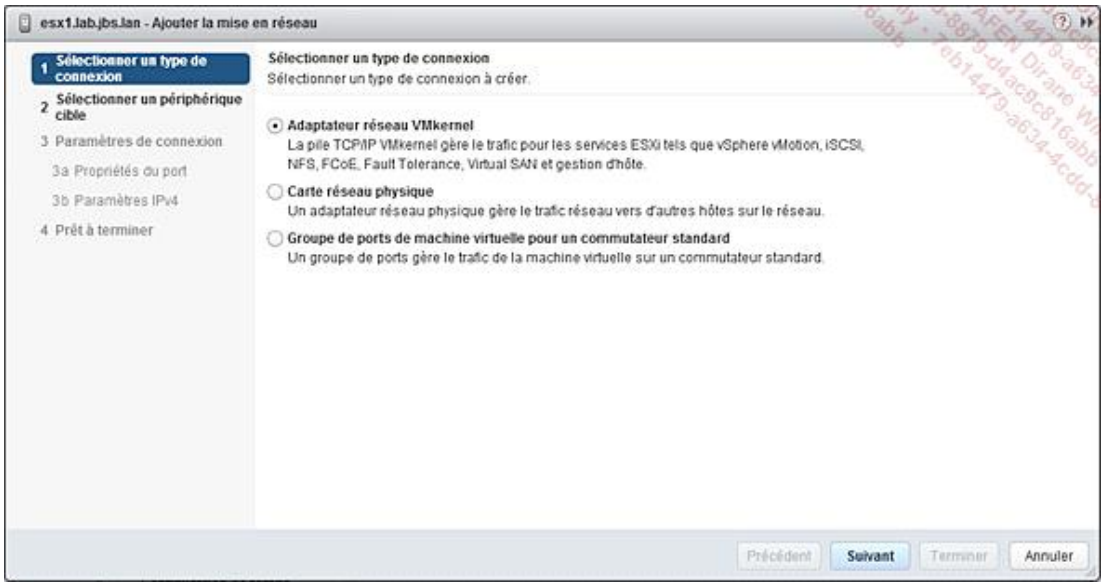
a. Création

Nous allons créer ici un vSwitch standard, à l'aide de vSphere Web Client.

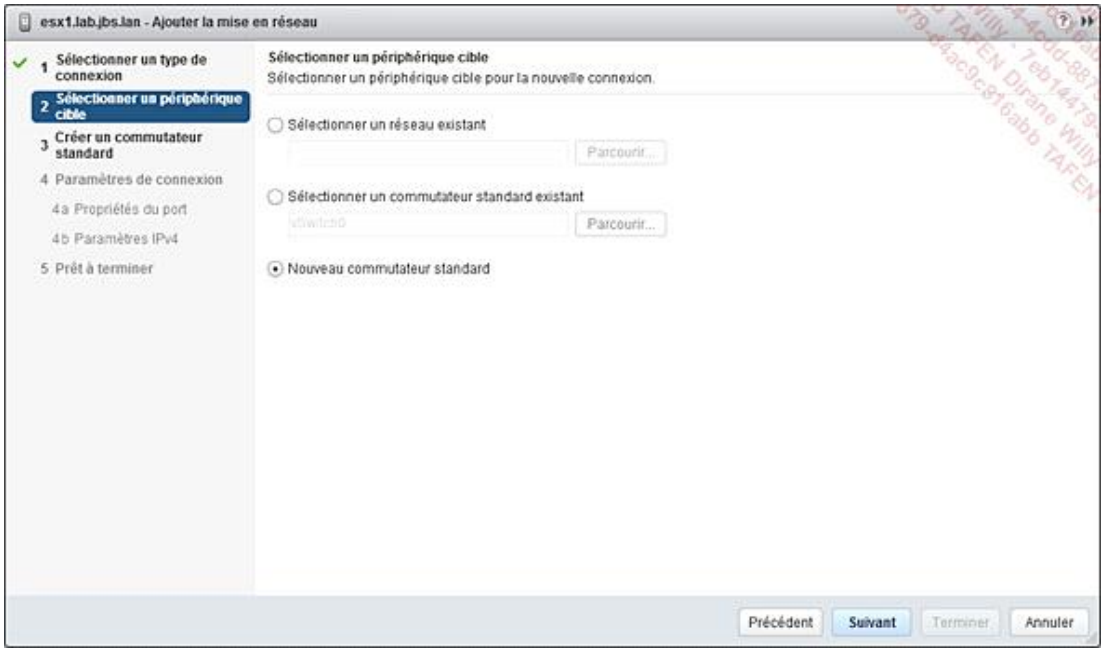
La configuration étant locale, la première étape consiste à se positionner sur l'hôte puis à aller dans la partie **Commutateurs virtuels**, sous l'onglet **Configurer**. Puis, choisissez l'option **Ajouter la mise en réseau d'un hôte** à l'aide du bouton associé doté d'un **+**.



Une fenêtre s'ouvre pour vous guider dans la création du vSwitch standard. Nous allons ici créer un vSwitch avec un port VMkernel. Notez que le choix du type de connexion apparaît avant de choisir la création d'un nouveau vSwitch.

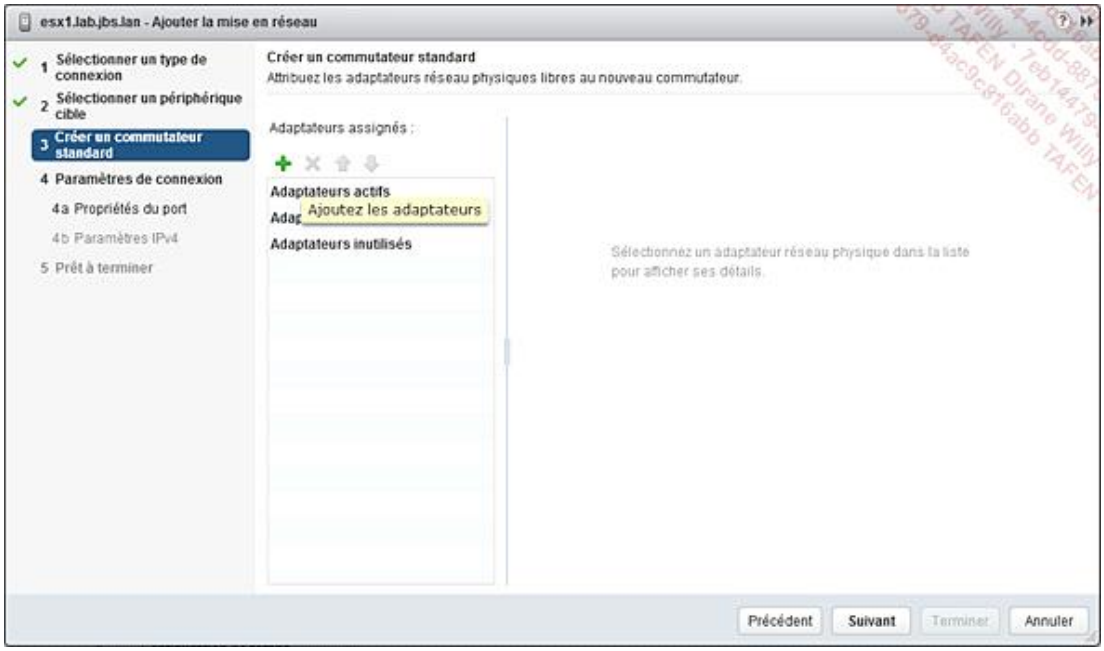


Choisissez ici la création d'un nouveau vSwitch, à l'aide de l'option **Nouveau commutateur standard**.

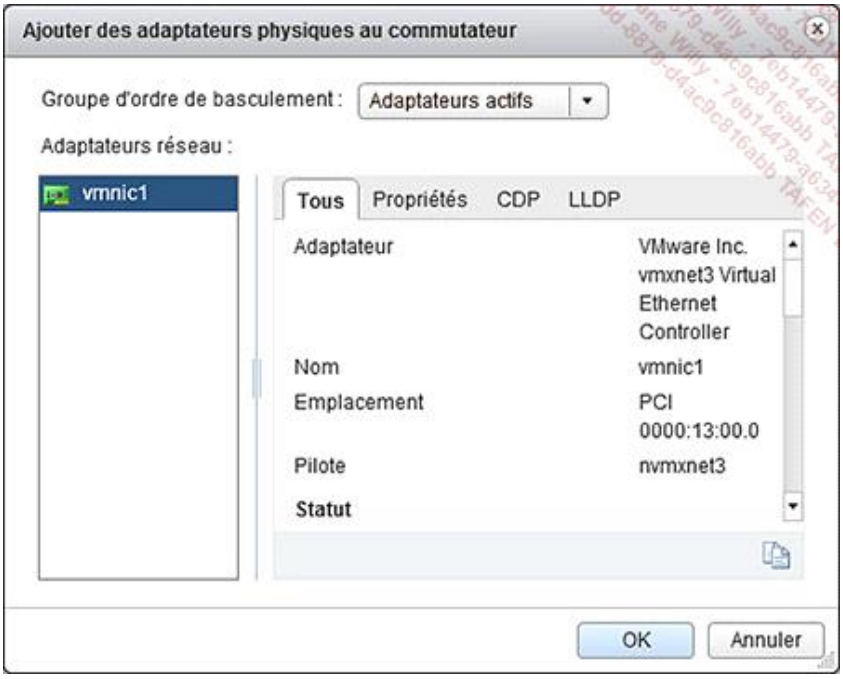


Précisons que si l'administrateur avait souhaité créer un port VMkernel sur un vSwitch existant, le premier choix aurait été approprié. Dans notre cas, nous souhaitons créer un nouveau commutateur virtuel.

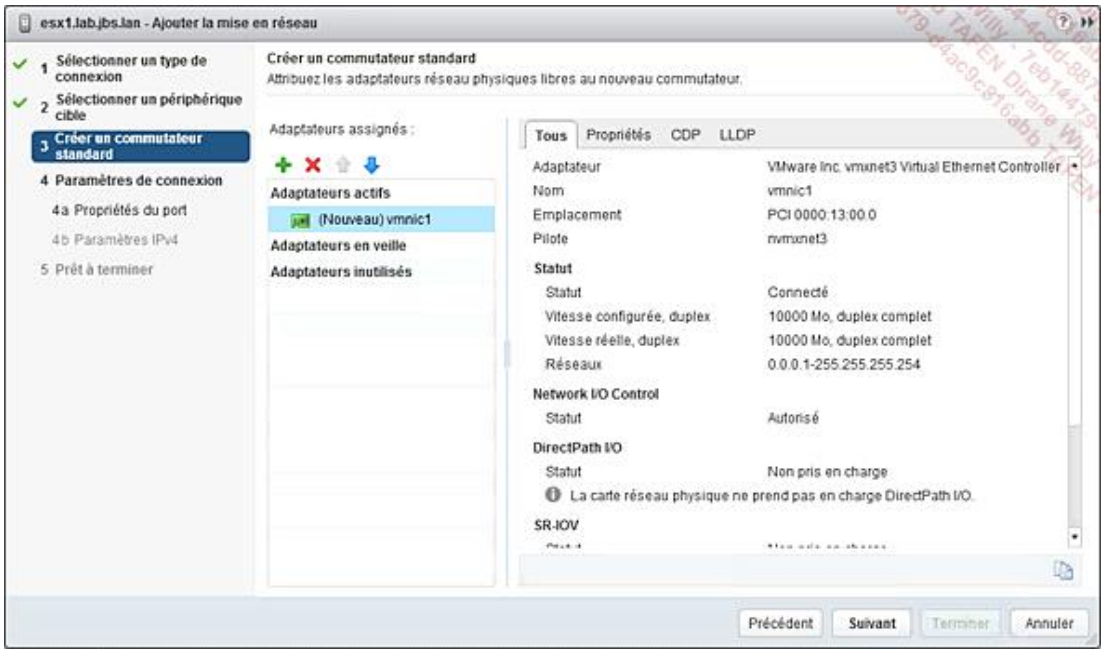
Sélectionnez les adaptateurs physiques qui seront liés à votre nouveau commutateur à l'aide du bouton **+**.



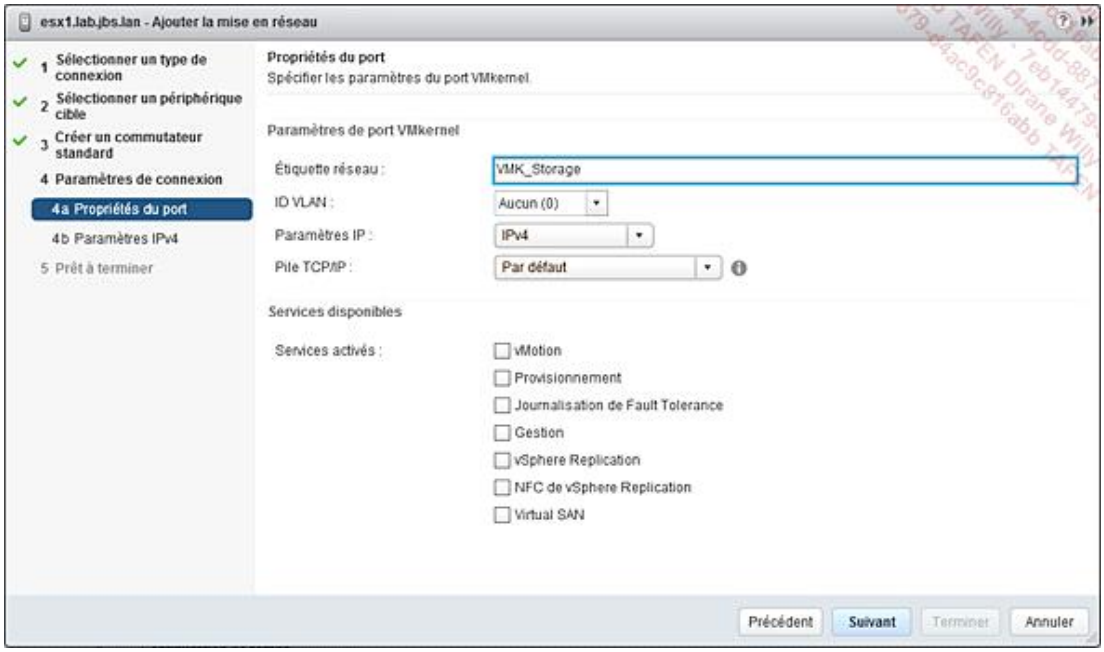
Puis sélectionnez l'adaptateur de votre choix.



On choisit les cartes réseau physiques associées au vSwitch. Seules les cartes non réclamées (c'est-à-dire non associées à une autre entité) sont disponibles. Notez que cela n'est pas obligatoire, un commutateur virtuel peut très bien être créé sans liaison montante.



La carte vmnic1 va être ajoutée à la configuration du vSwitch standard. Les paramètres de la carte réseau sont affichés. Vérifiez bien sous quel nom l'ESXi voit la carte, cela correspond au pilote qui est chargé. Certaines incompatibilités ou certains comportements imprévus peuvent provenir de certains modèles de cartes réseau. Pour éviter les latences liées à la gestion d'un même trafic réseau via des pilotes de cartes différentes, veillez à utiliser des cartes physiques du même type pour un même vSwitch.



On donne un nom au port VMkernel - ici VMK_Storage car on l'utilisera pour l'accès à une baie SAN iSCSI ou un NAS (en NFS).

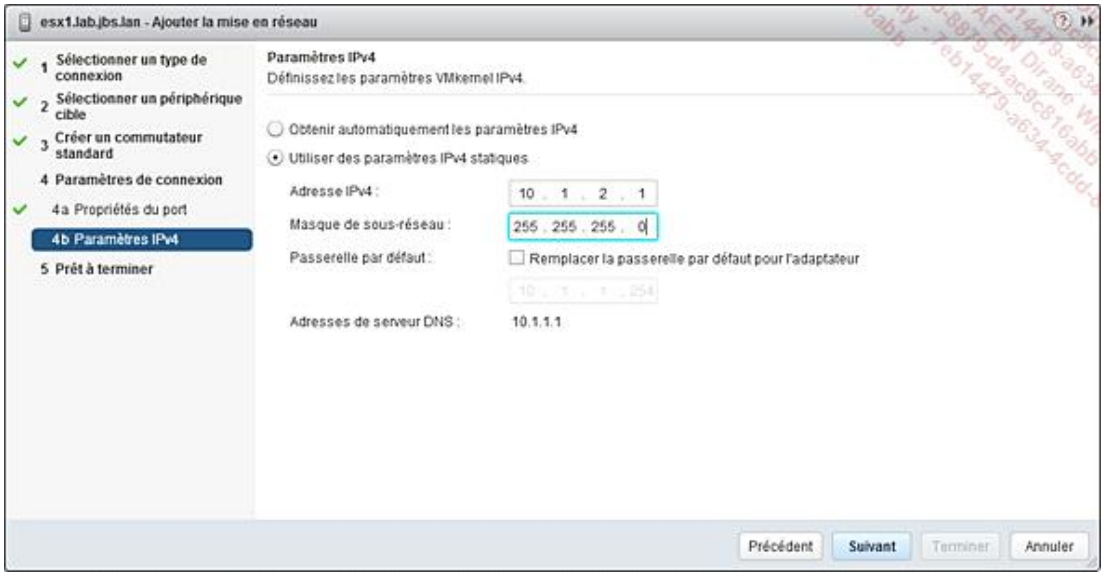
Un port VMkernel permet de créer une interface de gestion et de rendre possible (entre autres) les communications entre l'hyperviseur vSphere et le milieu extérieur (vCenter, baie de stockage, passerelle réseau...).

On décide du service à activer selon la fonction que l'on doit mettre en place. Grâce à cette séparation fine des types de trafic, l'administrateur a la capacité de réserver une ou plusieurs liaisons montantes à un usage particulier. Dans notre exemple, pour accéder à du stockage réseau en iSCSI, il n'y a pas besoin de sélectionner un service particulier.

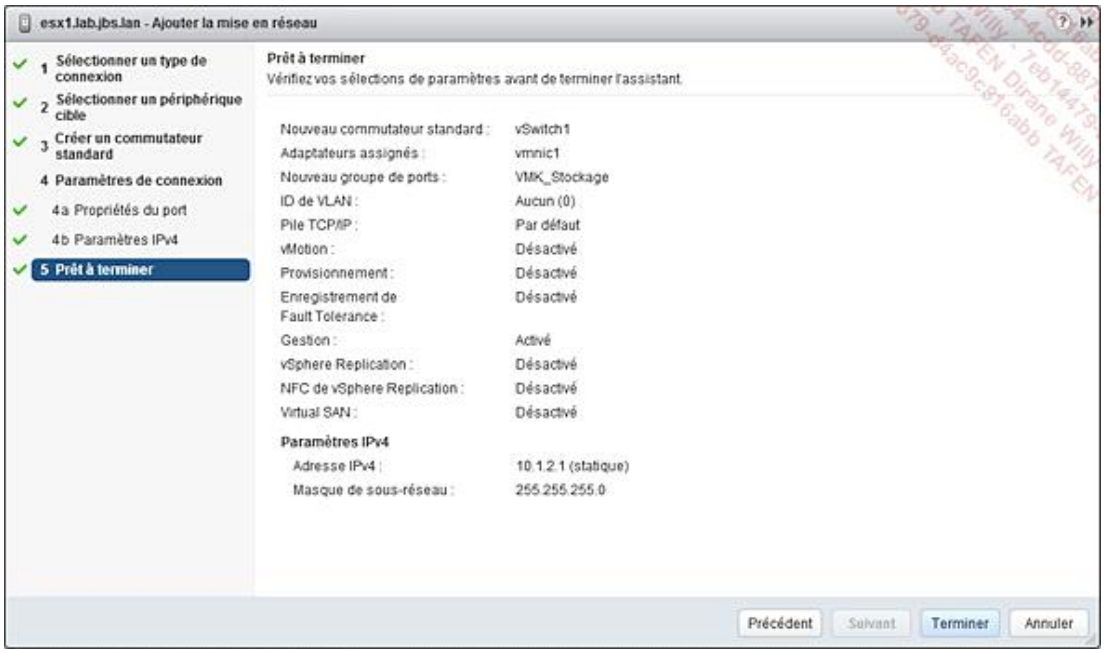
Pour information, voici les différents services disponibles :

- **vMotion** : interface pour la migration du contexte d'exécution de la machine virtuelle d'un serveur vSphere à un autre, à chaud.
- **Provisionnement** : interface concernant le trafic de migration à froid des machines virtuelles, le clonage ou la migration de snapshots.
- **Journalisation de Fault Tolerance** : interface de synchronisation entre la machine primaire et secondaire, dans le cadre d'un déploiement Fault Tolerance dans un cluster.
- **Gestion** : interface liée au trafic d'administration du serveur, flux entre les hyperviseurs et le serveur vCenter, trafic entre hôtes pour les mécanismes HA.
- **vSphere Replication** : interface prenant en charge le trafic sortant de réplication de l'hôte à destination du serveur vSphere Replication.
- **NFC de vSphere Replication** : interface prenant en charge le trafic entrant de réplication sur le site de réplication cible.
- **Virtual SAN** : interface utilisée dans le cas d'un déploiement Virtual SAN (vSAN), le produit d'hyper convergence de stockage de VMware.

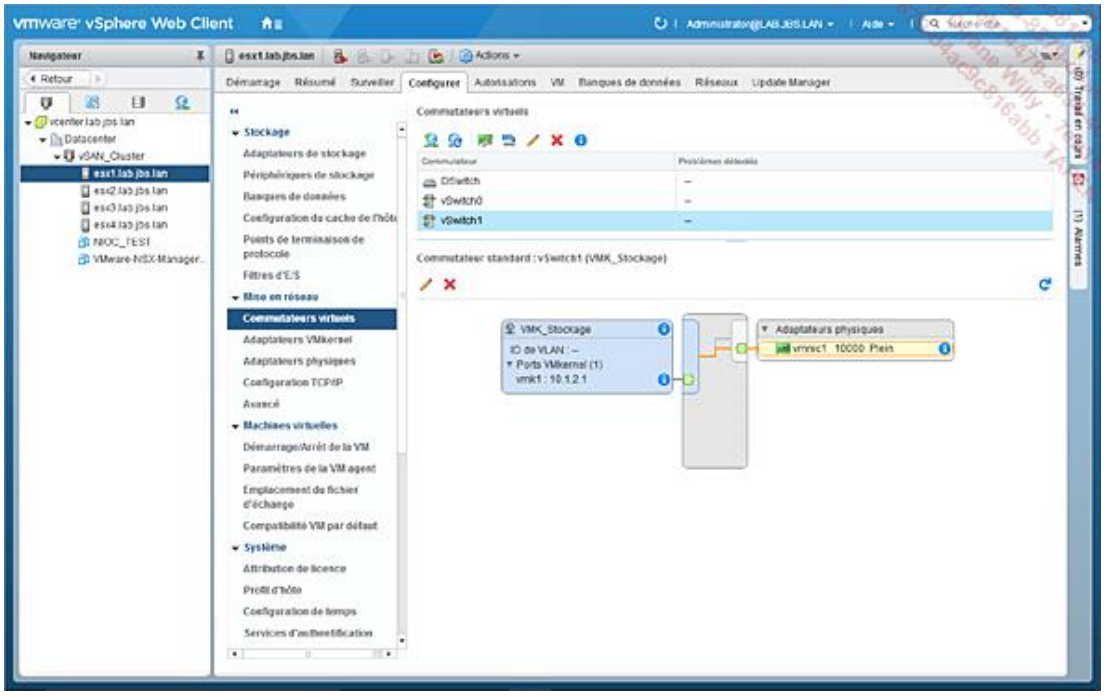
On indique ensuite dans le prochain écran les paramètres réseau du port.



Puis un récapitulatif apparaît. Validez pour terminer la création du nouveau commutateur virtuel et de l'interface VMkernel.



Le vSwitch est créé, vous pouvez visualiser son statut et sa configuration dans la section **Commutateurs virtuels** de la configuration de l'hôte.

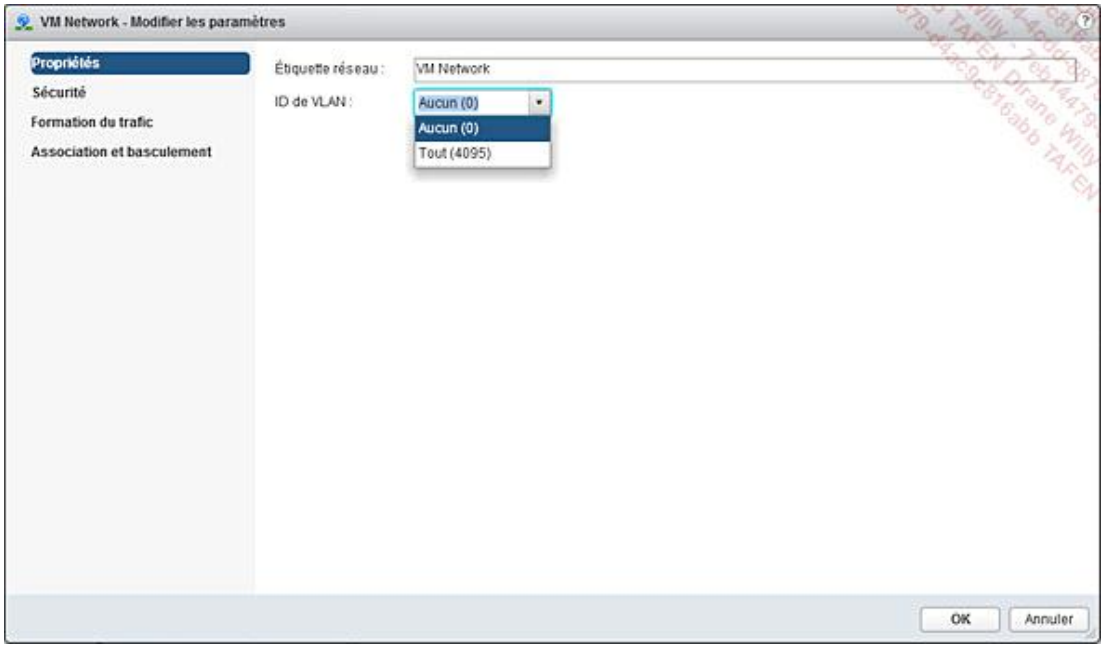


b. VLAN

Il est possible de connecter les machines virtuelles à un ou plusieurs VLAN tout simplement en les connectant à un ou plusieurs groupes de ports de machine virtuelle (*Virtual Machine Port Group* ou VMPG). Il est donc recommandé de créer plusieurs groupes de ports avec chacun des VLAN.

C'est d'ailleurs sur ces groupes de ports que l'on configure (ou pas) les identifiants de VLAN. En ce qui concerne les ports VMkernel, cela peut être effectué à leur création (mais cette option reste modifiable par la suite).

La capture suivante vous montre comment configurer l'identifiant de VLAN pour un groupe de ports, une fois l'interface de paramétrage du groupe de ports affichée.



Plusieurs options sont possibles ici :

- **Aucun (0)** : le vSwitch ne transmet pas de trafic étiqueté, en entrant ou sortant de la perspective de la machine virtuelle. Cependant, le commutateur physique est libre d'étiqueter le trafic non étiqueté par l'hyperviseur.
- **VLAN spécifié manuellement** : le groupe de ports appartient à un VLAN en particulier. Les numéros de VLAN permis vont de 1 à 4094.
- **Tout (4095)** : l'administrateur permet à la machine virtuelle d'étiqueter les trames directement. Il faut alors positionner le VLAN ID à 4095 ou choisir l'option mentionnée.

Dans le cas où nous spécifions un VLAN ID entre 1 et 4094, le vSwitch identifie le trafic qui est étiqueté avec le VLAN paramétré. Le cas échéant, il transmet le trafic à la machine virtuelle connectée après avoir retiré l'étiquette de VLAN (tag). La machine virtuelle ne reçoit pas de trame étiquetée. De même, quand la VM communique, les trames sont étiquetées au niveau du vSwitch.

Bien sûr, si vous décidez d'avoir plusieurs groupes de ports avec plusieurs VLAN, il faudra que vos liaisons montantes côté hyperviseur soient de type Trunk.

En termes de terminologie, sachez que l'on peut également employer les termes suivants :

- **EST** (*External Switch Tagging*) lorsque la première option avec l'identifiant de VLAN à 0 est choisie.
- **VST** (*Virtual Switch Tagging*) lorsque l'administrateur choisit un identifiant de VLAN de 1 à 4094.
- **VGT** (*Virtual Guest Tagging*) lorsque l'option **Tout (4095)** ou l'identifiant de VLAN 4095 est configuré.

c. Nombre de ports

Tout comme un switch physique, un vSwitch comporte un nombre de ports limité. En utilisant vSphere 6.5, vous disposez d'un nombre maximum de 4088 ports par commutateur vSwitch standard.

Jusqu'à vSphere 5.5, un vSwitch disposait par défaut de 120 ports. Ce nombre pouvait être changé (de 64 à 4 088) mais cela nécessitait le redémarrage du serveur hôte pour être pris en charge.

La capture suivante montre le nombre de ports dans une installation vSphere 6.5. Vous voyez que la valeur n'est pas modifiable par l'administrateur et est « élastique », c'est-à-dire dynamiquement allouée en fonction de la demande, dans la limite du maximum cité plus haut.

d. MTU

Le MTU (*Maximum Transmission Unit*) est la taille maximale d'un paquet qui peut être transmis en une seule fois sans fragmentation sur un réseau Ethernet. Parfois, il peut être bon de l'augmenter car plus la fragmentation est présente, plus les performances déclinent (on baisse très rarement ce paramètre).

Le MTU varie selon le type de réseau, par exemple 1564 en Frame Relay, de 1500 à 9000 en Ethernet.

Par défaut le MTU est à 1 500 sur les réseaux Ethernet. Augmenter le MTU revient à permettre les « jumbo frames ». Dans ce cas, on positionne en général le MTU à 9 000 octets.

L'administrateur peut changer cette valeur à l'aide du même écran décrit ci-dessus.

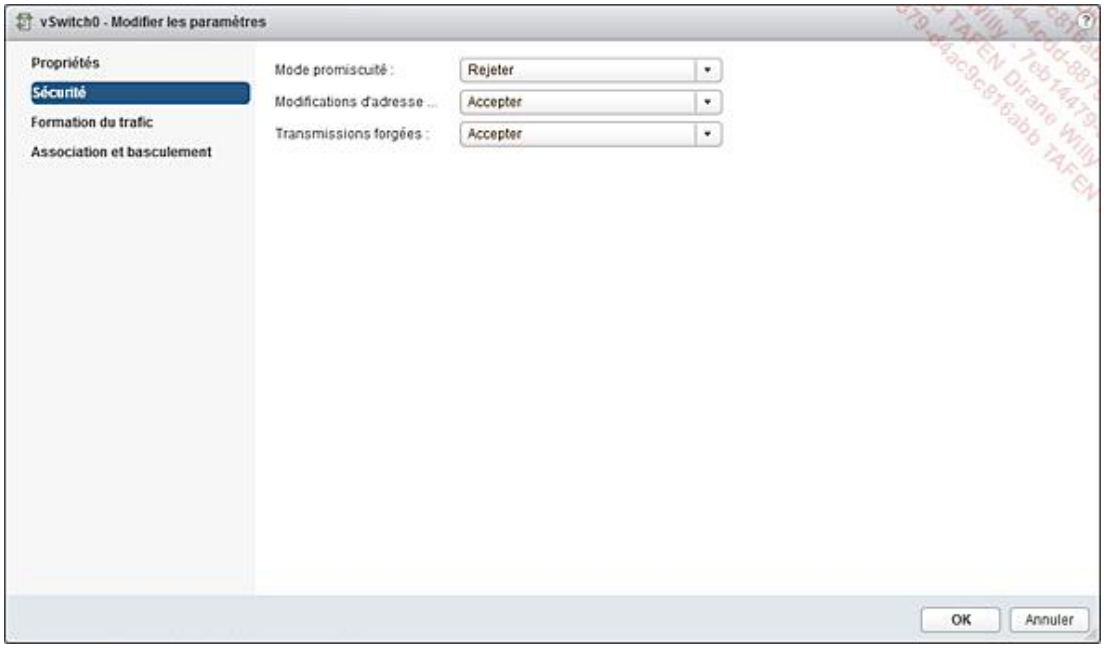
e. Sécurité

Les options de sécurité sont :

- La première appelée **mode espion** ou plus communément **Promiscuous mode** (mode de promiscuité) facilite généralement l'écoute de trafic réseau puisque cela permet à une machine virtuelle via sa carte réseau de recevoir du trafic qui ne lui est pas destiné initialement, notamment le trafic d'un groupe de ports, relié à un vSwitch. De la perspective du système invité, le principe est le suivant : chaque trame envoyée inclut l'adresse MAC de la carte réseau destinataire. Quand une carte réseau reçoit une trame, celle-ci vérifie si elle est bien la destinataire du paquet. Si elle ne l'est pas, elle élimine la trame (drop). En mode espion, la carte réseau n'élimine pas les trames et il est donc possible d'analyser toutes trames reçues, y compris celles dont l'adresse MAC de destination ne correspond pas à cette de la carte réseau du destinataire. Cette option est utile pour les analyses mais aussi les serveurs mandataires (proxy) ou filtrage d'URL / protocoles tels que Olfeo (http://www.olfeo.com) ou Raytheon|Websense (https://www.forcepoint.com/fr) ou enfin les sondes réseau, notamment IDS. Si le paramètre peut être utile dans certains cas, vous devez rester conscient que n'importe quelle machine sur le segment configuré pourra potentiellement demander à recevoir la copie du trafic de ce même segment. De plus, le fait de permettre ce mode au niveau d'un vSwitch autorise cette utilisation sur l'ensemble des groupes de port reliés à ce dernier (sauf paramètre contraire sur un port group). C'est ce qui explique pourquoi il est conseillé, sauf besoin spécifique, d'activer ce paramètre là ou vous en avez besoin (sur un groupe de port directement) lorsque vous utilisez un vSwitch standard. Le problème ne se pose pas avec l'utilisation d'un vDS, puisque dans ce cas, la configuration n'est possible qu'au niveau d'un groupe de port.
- La deuxième **modification d'adresse MAC** demande à l'hyperviseur vSphere d'accepter ou de rejeter les demandes de modification de l'adresse MAC d'un système invité. Cette option est particulièrement intéressante lorsque l'on souhaite arrêter des attaques type « MAC Spoofing ». Lorsque cette option est activée (avec l'option **Rejeter**), toute trame entrante dans la machine virtuelle où l'adresse MAC a été changée subira purement et simplement un « drop ». Une VM qui change donc son adresse MAC est alors considérée comme hostile et ne recevra aucun trafic. Cette option est par défaut désactivée car il existe des cas où l'adresse MAC doit être modifiée dans le Guest OS (le cas d'un système de licences par adresse MAC étant le plus répandu).
- La dernière option **transmission forgée** permet d'accepter ou de rejeter des trames qui auraient été modifiées de quelque manière que ce soit en sortie. Cette option considère qu'une trame forgée en sortie est dangereuse. Couplée à la deuxième option, on évite la plupart des problèmes de MAC Spoofing.

Par défaut le mode de promiscuité est paramétré sur **Rejeter**, tandis que la modification d'adresse MAC et les paquets forgés sont « acceptés ».

La figure suivante montre comment modifier ces paramètres, dans les préférences du vSwitch ou du groupe de ports.



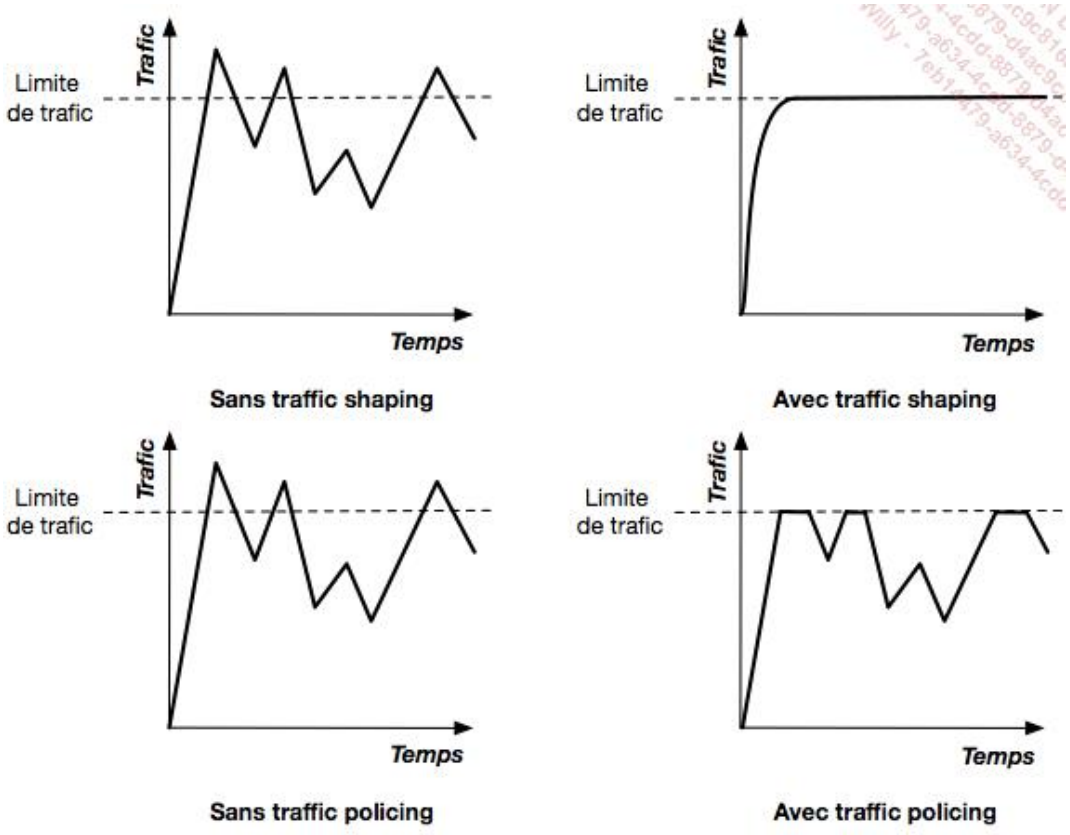
Sachez qu'il est requis d'accepter le mode de promiscuité et les transmissions forgées lorsque vous mettez en place de la virtualisation d'hyperviseurs vSphere (nested).

f. Traffic shaping

Le traffic shaping est un mécanisme de QoS (*Quality of Service*) bien connu. Il consiste à lisser le trafic pour forcer le respect d'une politique d'usage, via l'utilisation d'une mémoire tampon pour retarder l'envoi des données.

Le plus souvent, l'administrateur réseau cherche à effectuer du shaping pour forcer la conformité de la vitesse d'envoi quand du policing est mis en place du côté réception (par un opérateur par exemple). Le policing consiste le plus souvent à réguler le trafic en entrée et peut refuser tout trafic qui dépasse une certaine limite (contractuelle dans la majorité des cas), sans le retarder comme le ferait le shaping. Cela fait donc sens pour l'administrateur d'appliquer du shaping en émission dans ce scénario.

Pour illustrer ce propos, regardez ci-dessous les deux schémas qui récapitulent ces explications sur le shaping et le policing :



Le shaping est configurable sur un groupe de ports ou directement sur un vSwitch standard. Quand l'administrateur utilise un vSwitch standard, cette restriction s'applique **uniquement** sur le trafic sortant (*outgoing*).

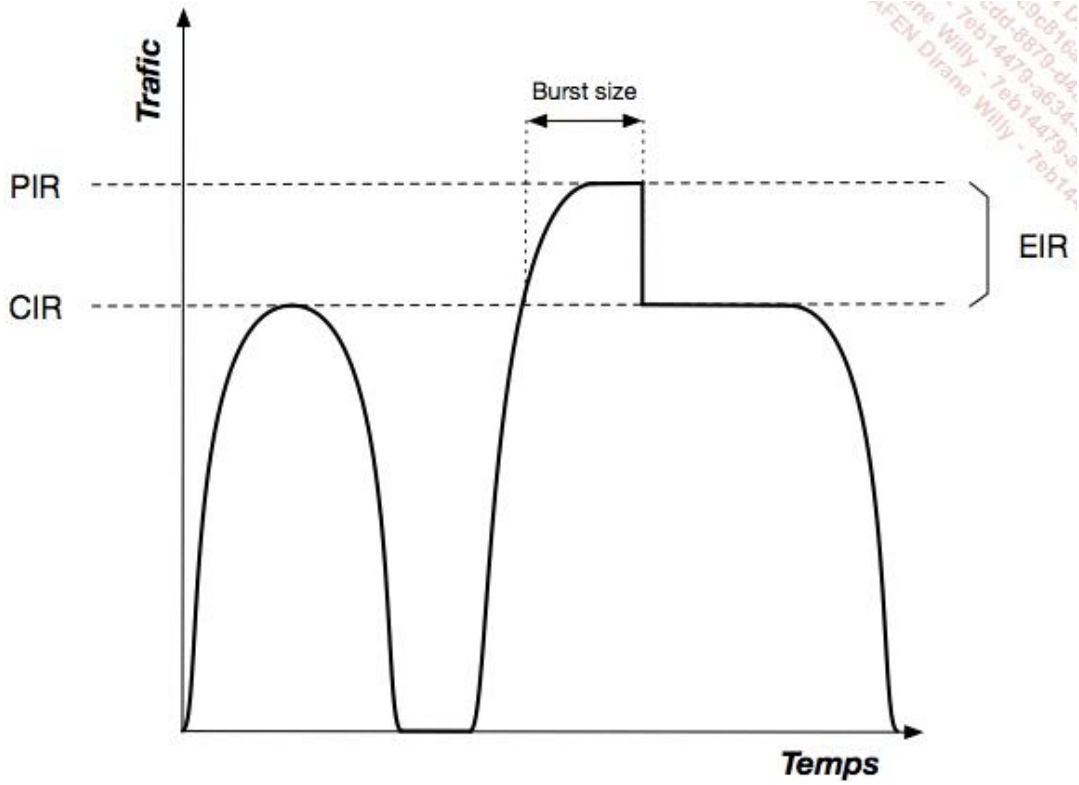
Les paramètres sont les suivants :

- **Bande passante moyenne** (*average bandwidth*) : le nombre de kbit par secondes à autoriser, ramené à une moyenne sur le temps pour calculer la bande passante.
- **Bande passante maximale** (*peak bandwidth*) : le nombre de kbit par secondes maximum à autoriser lors d'une rafale.
- **Taille de rafale** (*burst size*) : la taille maximale en Ko d'une rafale. Une rafale ne peut pas dépasser cette taille.

Pour donner plus de sens à ces différents paramètres, faisons un parallèle avec ce qui s'applique dans le monde du réseau et en particulier des routeurs. Dans ce but, nous devons définir plusieurs termes :

- CIR : il s'agit du **Commited Information Rate**, la vitesse moyenne de trafic, garantie contractuellement par un opérateur.
- PIR : il s'agit du **Peak Information Rate**, la vitesse atteignable en cas de rafale. Pour le calculer, il faut additionner le CIR et l'EIR (*Excess Information Rate*). Ce dernier représente la bande passante qui peut être utilisée lors de rafales. Ces deux éléments sont également précisés dans le contrat de service avec l'opérateur.

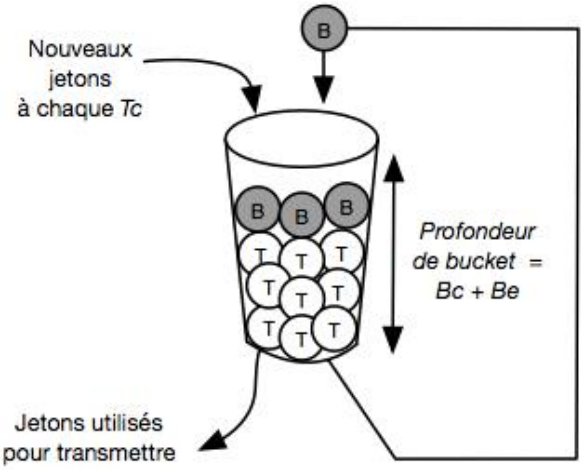
Le schéma ci-dessous illustre ces explications.



Comme vous l'avez vu plus haut, le trafic maximal atteignable en cas de rafale correspond au PIR. La rafale ne pourra pas utiliser une bande passante plus élevée.

Un autre paramètre est à prendre en compte, la taille de rafale. Ainsi, la transmission pourra bénéficier du PIR pour envoyer une somme de données équivalente à la taille de la rafale. Le temps de rafale est fonction de cette taille et de la vitesse PIR.

Remarquez le mécanisme qui a permis la rafale. Le trafic garanti (CIR) n'a pas été utilisé entièrement pendant un moment, permettant ainsi d'accumuler des jetons pour permettre une rafale en cas de besoin. La figure suivante vous explique comment fonctionne ce mécanisme :

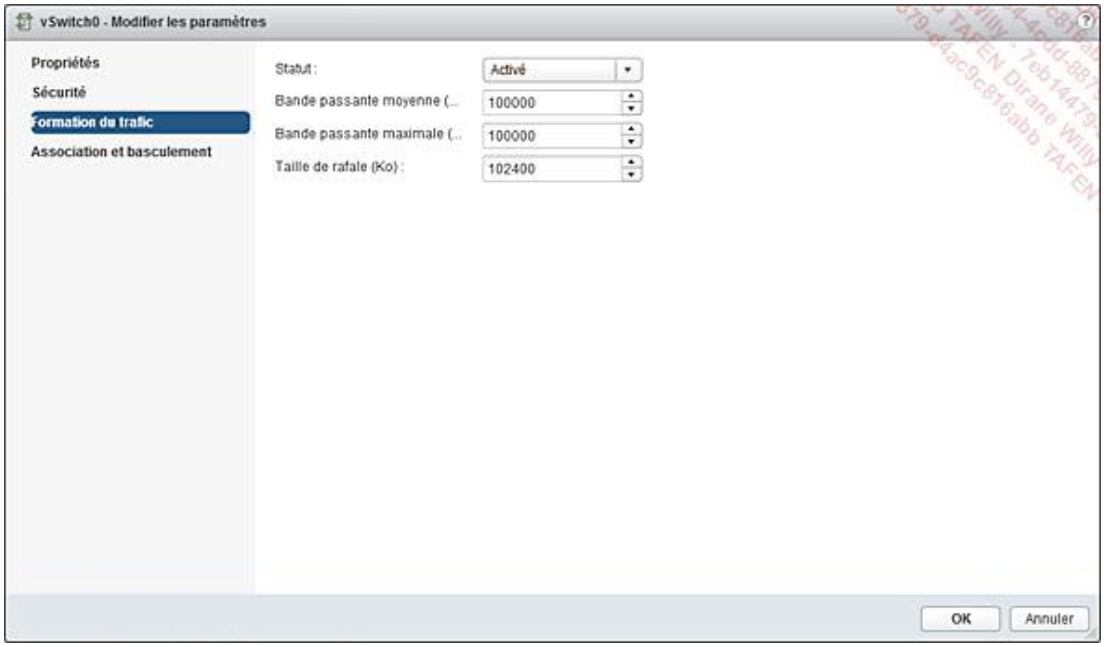


À chaque intervalle d'envoi (Tc), le seau (bucket) est rempli de l'équivalent de jetons permettant un envoi au CIR (Bc - les jetons T). Les jetons non utilisés sont conservés et accumulés pour servir en cas de rafale (Be - les jetons B), dans la limite de la taille de rafale configurée. Un envoi constant au CIR ne générera aucun jeton B, donc aucune possibilité de rafale. Si le PIR est égal au CIR et donc l'EIR de zéro, il ne sera pas non plus possible d'effectuer de rafale.

Prenons une configuration où les paramètres sont les suivants : CIR (ou bande passante moyenne) à 128 000 kbit/s, PIR (ou bande passante maximale) à 192 000 kbit/s et une taille de rafale égale à 100 Mo. L'EIR sera donc égal ici à 64 000 kbit/s. La rafale au PIR pourra durer un peu plus de quatre secondes si l'on considère que le sceau est plein (jetons Bc + Be).

$$Temps\ de\ rafale = \frac{Taille\ de\ rafale}{PIR} = \frac{100 * 1024 * 8}{192000} = 4{,}266\ s$$

Voici l'écran de configuration du traffic shaping sur un vSwitch standard.



Remarquez que dans la configuration par défaut, les bandes passantes moyenne et maximale sont identiques, ce qui interdit toute rafale. La taille de rafale par défaut est ici équivalente à 100 Mo.

Le shaping est désactivé par défaut sur un vSwitch standard.

g. Le NIC teaming

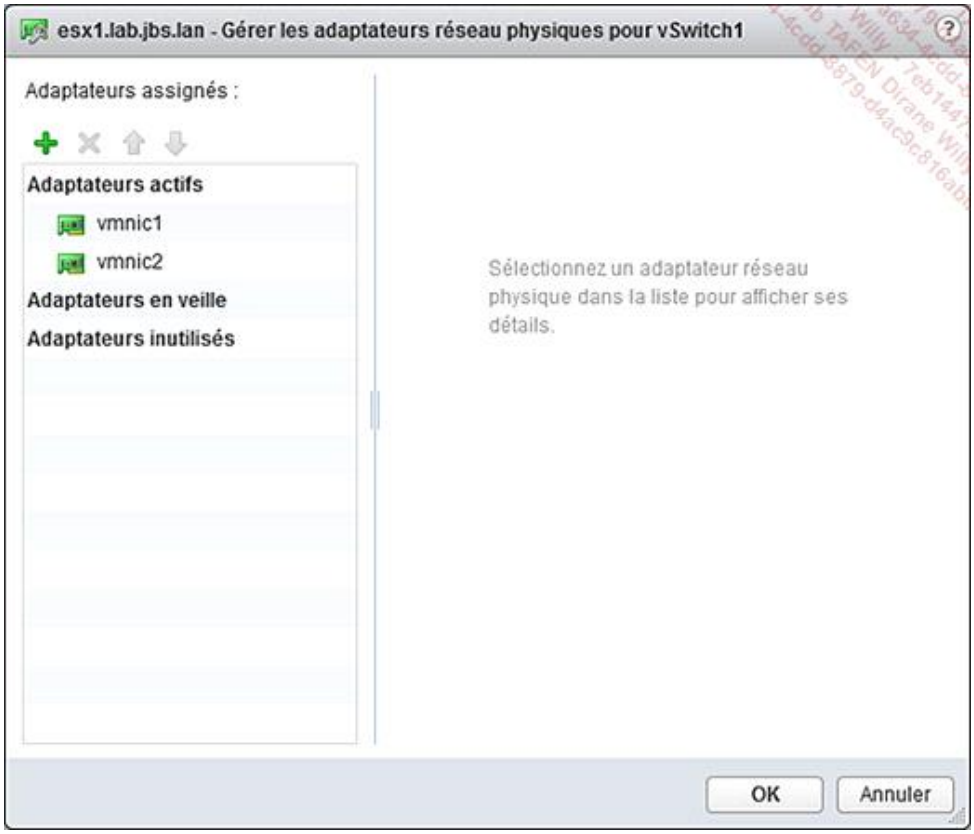
Il s'agit des politiques liées à l'utilisation de plusieurs cartes réseau physiques (vmnic) au niveau d'un même vSwitch.

L'acronyme ASU est utilisé pour décrire les trois états possibles d'une liaison montante au sein d'un vSwitch :

- **Actif** (*Active*) : la carte réseau est utilisée par l'hyperviseur afin de recevoir / envoyer des informations.
- **En attente** (*Standby*) : la carte réseau est inactive mais deviendra active à partir du moment où une des cartes actives dysfonctionnera.
- **Inactif** (*Unused*) : carte réseau ne pouvant pas participer au NIC Teaming. Les raisons peuvent être multiples (maintenance sur la carte, problème détecté sur la carte et investigation en cours, etc.). Il peut y avoir de nombreuses raisons pour lesquelles une carte ne doit pas devenir active...

On peut aussi tout simplement ne pas l'inclure ou la retirer du teaming si la carte ne doit jamais être active.

L'hyperviseur permet de sélectionner également l'ordre de basculement (*Failover Order*). Cette option est particulièrement utile pour le déclenchement d'alerte au niveau réseau, car l'utilisation de certains liens (dits liens de backup) peut alors faire retentir des alarmes programmées au niveau de la supervision.

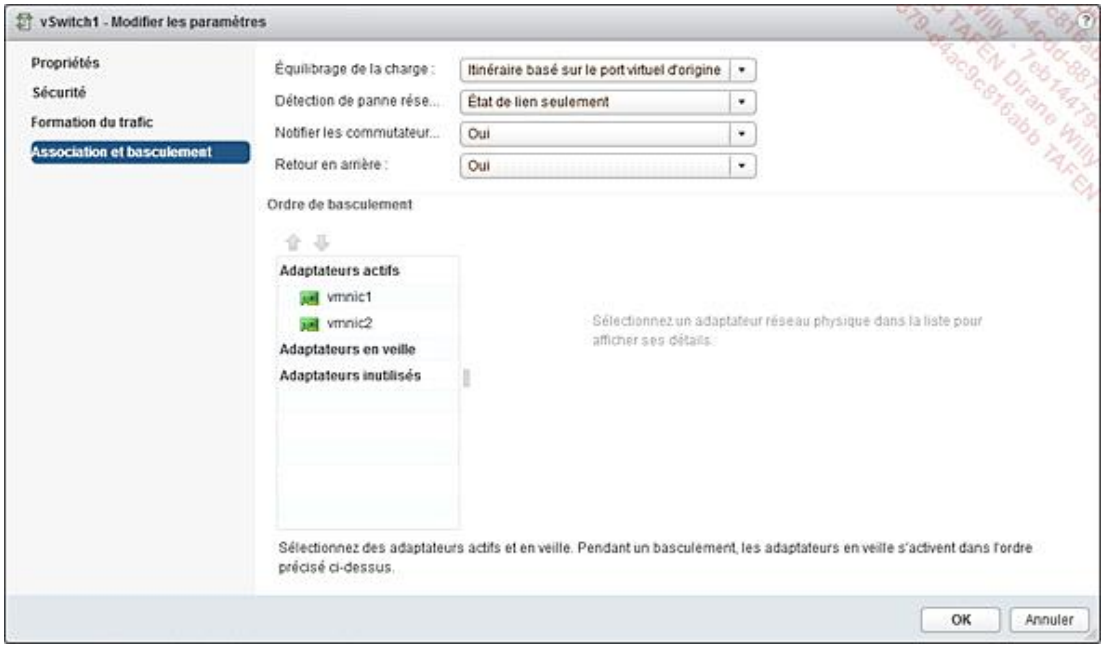


h. Répartition de charge

Pour le trafic réseau des machines virtuelles, la répartition de charge se fait selon trois algorithmes ou méthodes :

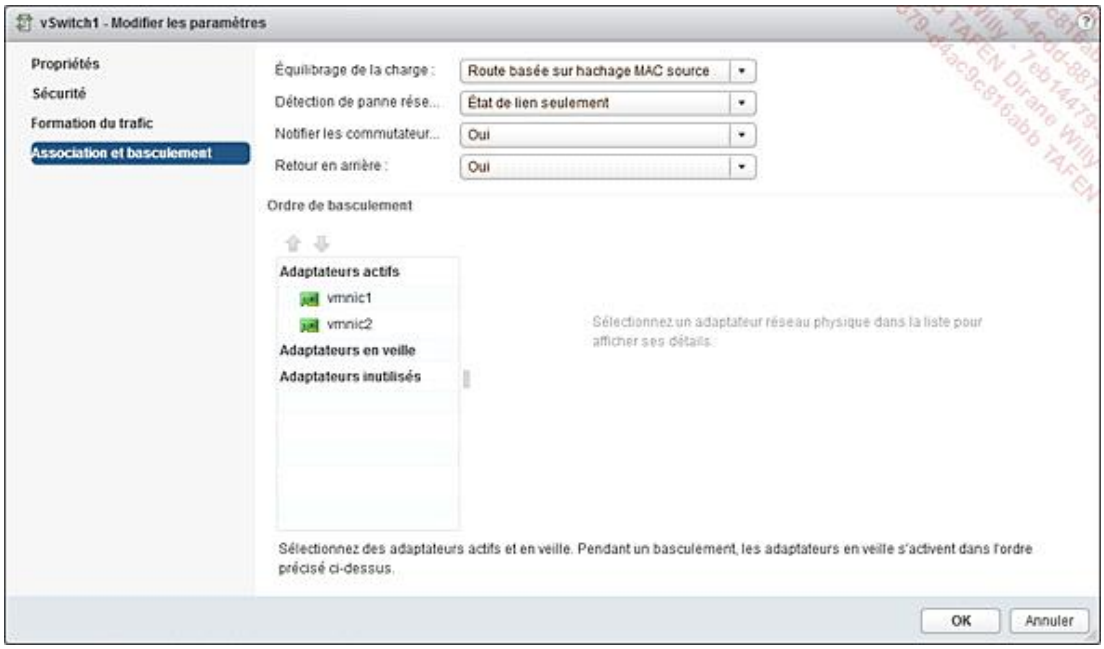
- **Route basée sur ID port virtuel d'origine** (*Virtual port Based*)

Cela permet au VMkernel de choisir sur quelle carte physique envoyer les données en se basant sur le Virtual Port ID. Le Virtual Port ID est un élément assigné à chaque carte réseau virtuelle lorsque celle-ci se connecte à un commutateur virtuel. Cette méthode est compatible avec tout type de commutateur physique. C'est la méthode sélectionnée par défaut.



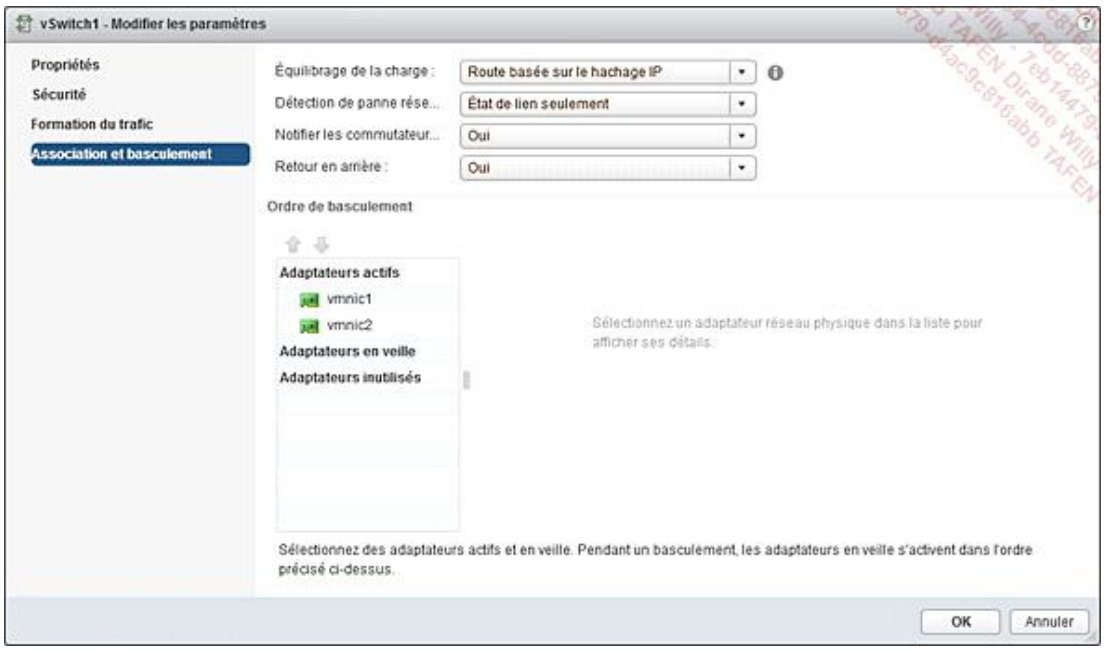
- **Route basée sur hachage MAC source** (*Source MAC Based*)

Au lieu de se baser sur le Port ID, la répartition sur les cartes réseau physiques se fait en évaluant l'adresse MAC source à l'aide d'un algorithme de sélection interne à l'ESXi.



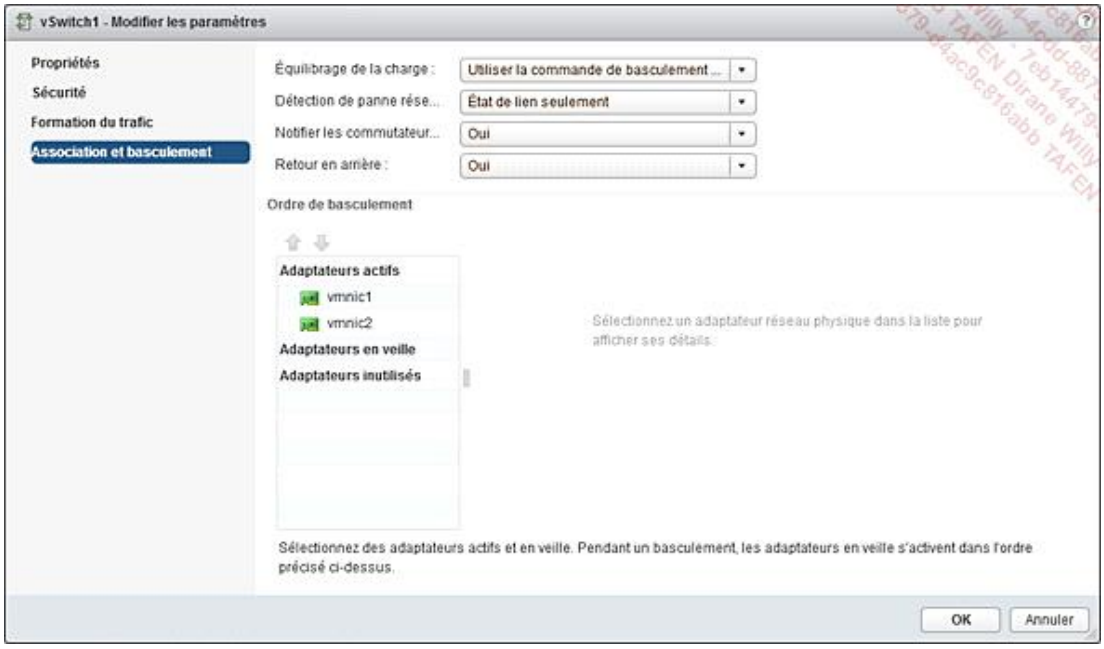
- **Route basée sur le hachage IP** (*IP Hash Based*)

Cette autre méthode consiste à répartir les charges en se basant sur un hash de l'adresse IP source ET de l'adresse IP de destination. Cette méthode est bien meilleure en termes de redondance et de répartition équitable mais nécessite que le commutateur auquel est connecté l'hyperviseur soit correctement configuré. Le protocole LACP (802.3ad) ou une configuration Etherchannel (Cisco) peut être nécessaire pour utiliser un agrégat utilisé dans ce cadre. Attention, bien que le trafic puisse être réparti sur plusieurs liaisons montantes, seule une pNIC sera utilisée par session. De plus, cette méthode engendre une surcharge plus élevée que les deux autres méthodes car le hash doit être calculé et comparé régulièrement.



- **Utiliser la commande de basculement explicite**

Aucune répartition de charge ne s'applique ici. L'hyperviseur choisira toujours la première liaison montante active de la liste des adaptateurs. Dans le cas où une liaison montante active dysfonctionne, il choisit le deuxième et ainsi de suite. En dernier recours, l'hyperviseur aura recours aux liaisons montantes en veille.



i. Détection des défaillances réseau

Les défaillances au niveau des cartes réseau peuvent être détectées selon deux méthodes :

État de lien seulement (Link Status Only)

Dans cette configuration, la détection est très simple. La carte réseau physique détecte s'il y a toujours un lien

opérationnel entre elle et le commutateur réseau physique. Cette méthode fonctionne dans la majorité des cas.

Cette méthode :

- peut détecter un câble débranché entre le serveur (ESXi) et le switch physique.
- peut détecter une coupure de courant du commutateur physique.
- ne peut pas détecter une mauvaise configuration du switch (mauvais VLAN par exemple).
- ne peut pas détecter une panne indirecte comme derrière le switch (câble débranché entre le switch d'accès et un switch de distribution par exemple).

Sondage balise (Beacon probing)

Il existe un cas où la ressource réseau n'est pas accessible bien que le lien direct entre la liaison montante et le commutateur physique soit toujours en bon état ; par exemple, lorsque le commutateur physique subit une panne de liaison montante provoquée par une panne d'un lien ou d'un commutateur de distribution ou de cœur de réseau. Dans ce cas précis, la supervision active de ce mode va être plus qu'utile.

L'hyperviseur envoie des trames (*beacons*) pour détecter un problème réseau quelconque. Chaque carte envoie une trame vers les autres cartes réseau. Dans ce cadre, il est attendu que la trame envoyée d'une interface réseau soit reçue sur les autres interfaces. Si une carte ne reçoit pas trois trames de suite, elle est désactivée car considérée comme défectueuse.

- Cela nécessite au minimum trois cartes réseau en teaming sur un vSwitch.
- Si seules deux cartes réseau physiques sont présentes, le système va rencontrer un split brain : il va détecter le problème mais sera incapable de savoir laquelle des deux cartes est défectueuse. Dans ce cas il vaut mieux rester sur la méthode « État de lien ».

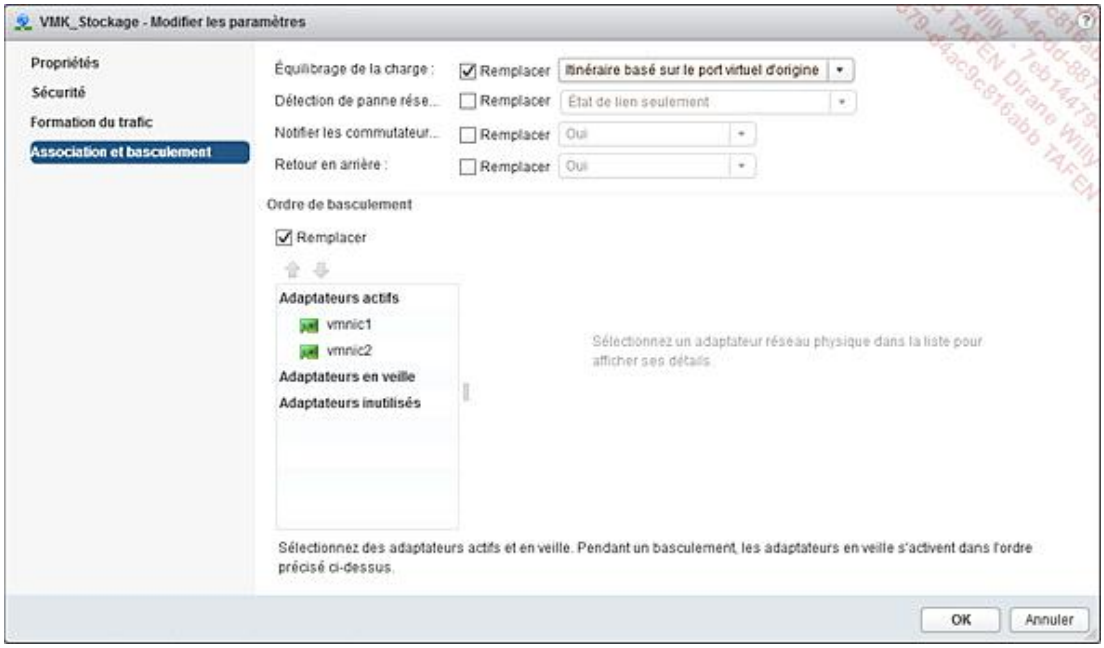
Les deux dernières options présentes sur l'écran sont **Notifier les commutateurs** et **Retour en arrière**.

- **Notifier les commutateurs** permet d'indiquer à l'hyperviseur de communiquer avec les commutateurs physiques le statut des machines virtuelles et de les prévenir si un changement devait avoir lieu (démarrage, vMotion, perte d'une liaison montante).
- **Retour en arrière** ordonne à l'hyperviseur de réutiliser une liaison montante qui a dysfonctionné et de la considérer à nouveau comme active.

Ces deux options sont positionnées à **Oui** par défaut.

j. Exceptions de configuration

Il est possible de configurer des « overrides » ou exceptions permettant de modifier sur un groupe de ports particulier la configuration générale du vSwitch, comme le montre la capture suivante :



Tout ce que nous avons abordé jusqu'ici reste de la configuration purement locale de la perspective de l'hyperviseur. Découvrons maintenant un moyen plus efficace de gérer le réseau des hyperviseurs !

3. vSphere Distributed Switch

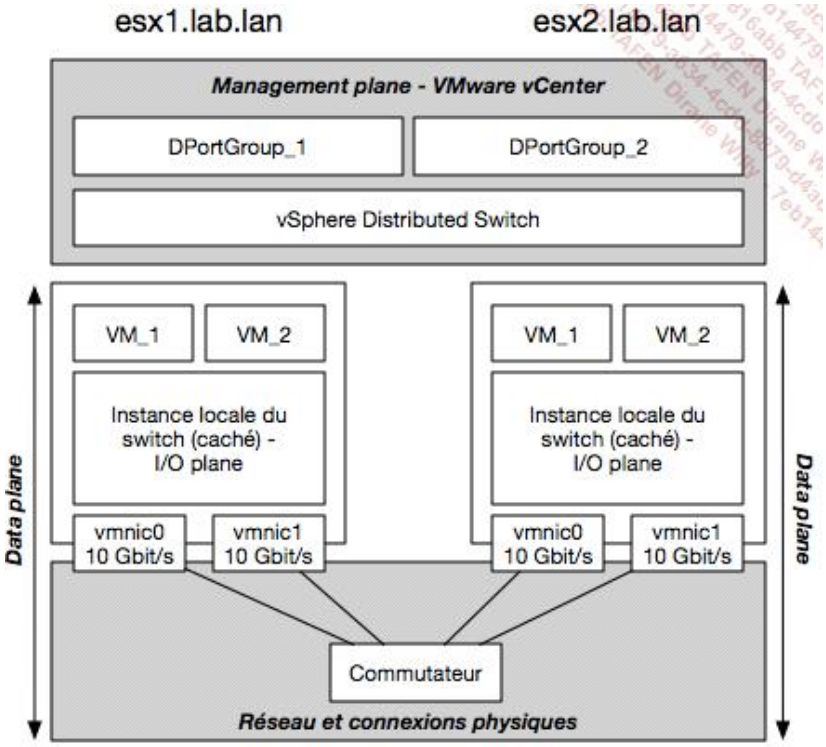
Le commutateur distribué ou vSphere Distributed Switch (vDS) simplifie grandement la gestion de votre réseau au sein des hyperviseurs.

En utilisant un vDS, l'administrateur permet l'extension d'un switch (et de ses groupes de ports) au-delà d'un seul et même hôte et potentiellement de tous les membres d'une ferme de serveurs. La configuration gagne donc en homogénéité et en simplicité. Les erreurs de déploiement (souvent humaines) sont moins présentes.

En termes de maintenance, les opérations sont également simplifiées. L'administrateur souhaite joindre ou retirer un hôte du vDS ? Aucun problème, toutes les opérations d'ajout et de suppression seront gérées sans difficultés.

De plus, vDS apporte des fonctionnalités non présentes dans les vSwitches standards comme Netflow, le mirroring de ports (SPAN/RSPAN), le shaping en entrée **et** en sortie et bien d'autres encore.

Décrivons à l'aide d'un schéma les différents composants de cette architecture :



Comme pour le scénario décrivant le vSwitch standard, nous disposons de deux hôtes, esx1.lab.lan et esx2.lab.lan.

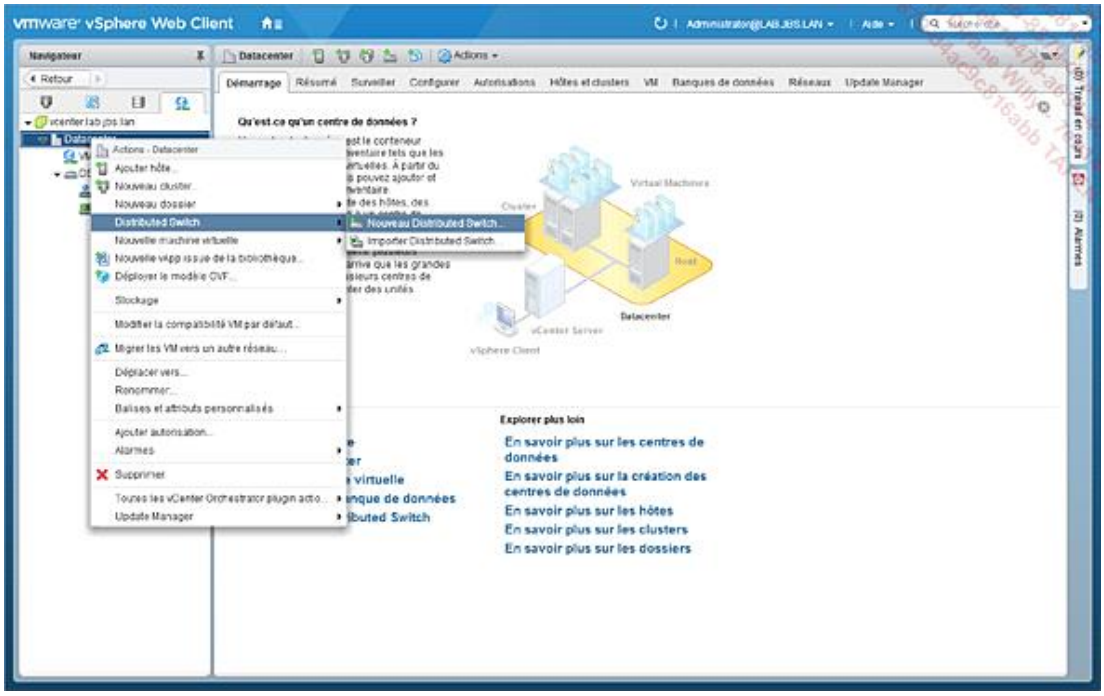
Vous remarquez que le premier composant qui est présent est ici vCenter. En effet, vous aurez besoin d'un serveur vCenter pour configurer, maintenir et superviser votre vDS. Vous pouvez alors vous poser la question de savoir ce qu'il se produit si le vCenter n'est plus disponible et si cet événement impactera ou non les opérations de votre vDS. La réponse est double ici. L'administrateur perdra en effet la capacité d'administrer ou de superviser le vDS. En revanche, le fonctionnement de celui-ci (envoi/réception de données, démarrage de machines virtuelles dont le port réseau a déjà été attribué sur un port group, ...) au sein des hôtes **ne sera pas impacté**.

Le Management Plane (gestion, supervision...) est ici assuré **entièrement** par VMware vCenter. Concernant le Data Plane, les données s'échangent via une instance locale de commutateur cachée, s'exécutant sur chaque hyperviseur.

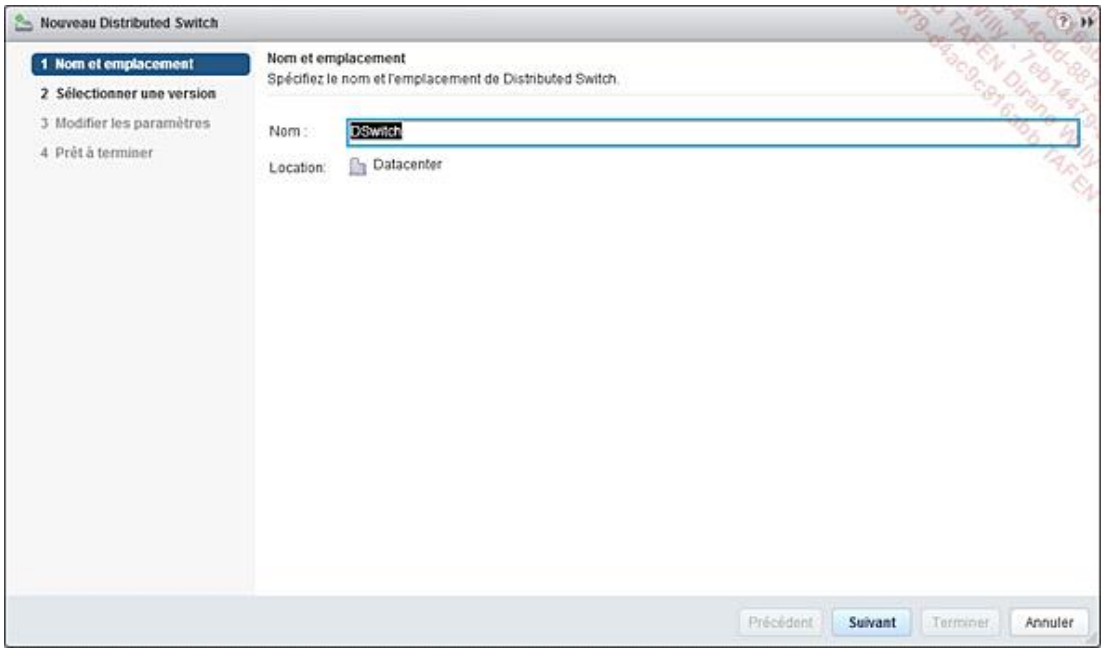
Commençons désormais la configuration d'un vSwitch distribué !

a. Création du vSwitch distribué

La première étape consiste à se positionner sur la partie réseau du menu latéral, puis effectuer un clic droit au niveau du Datacenter (sur lequel vous disposez bien entendu des droits nécessaires) puis sélectionner **Nouveau Distributed Switch** dans le menu **Distributed Switch**.



Saisissez le nom que vous souhaitez attribuer à votre vSwitch, nous le laissons ici à la valeur par défaut, **DSwitch**.



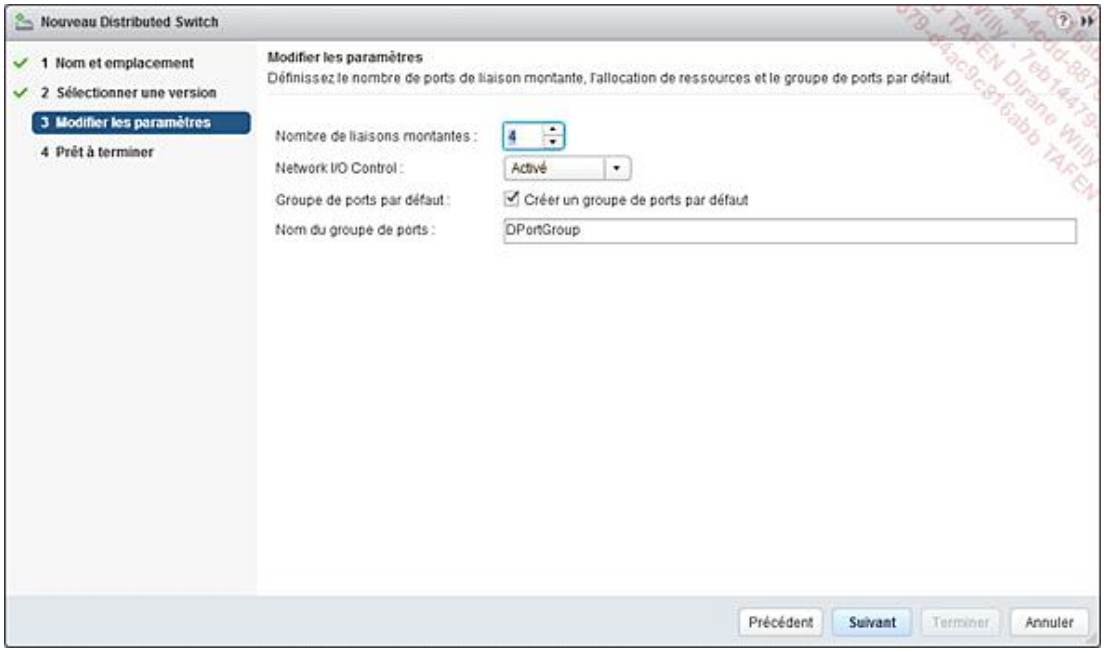
Sélectionnez la version de vDS que vous souhaitez déployer.

Préférez la version la plus récente systématiquement qui offre plus de fonctionnalités. Rappelez-vous que la version que vous pouvez choisir ici est influencée par la version de vos hyperviseurs. Ayant ici une infrastructure avec du vSphere et un vCenter 6.5, nous pouvons créer un vDS dans sa version la plus récente.

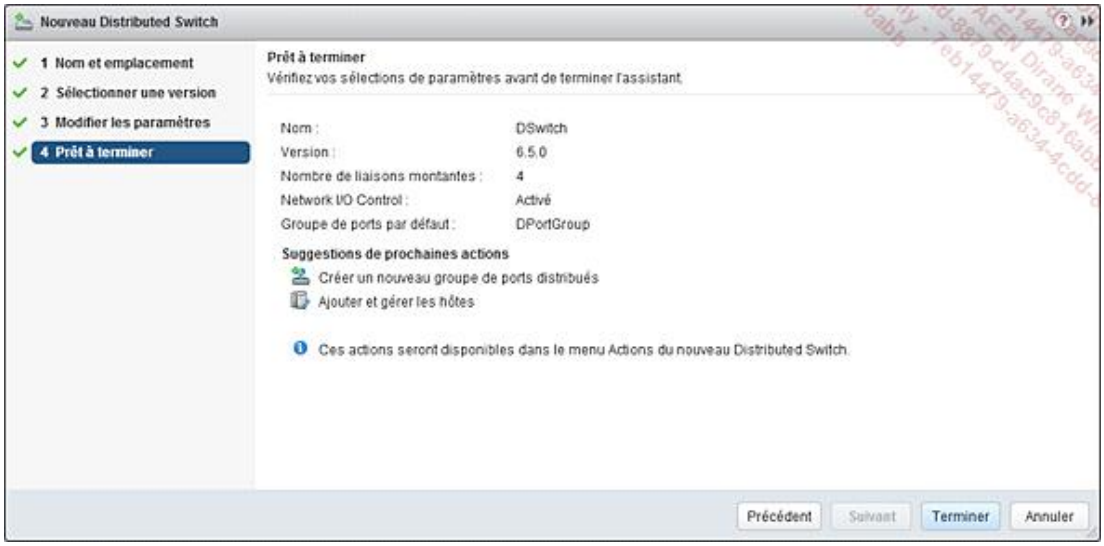
Selon le nombre de cartes réseau physiques prévues par hyperviseur, on choisit le nombre de liaisons montantes (uplinks). La création d'un groupe de ports est proposée par défaut. En cas de refus, on peut en créer plus tard.

Attention, les groupes de ports sont génériques. La création ou la migration de ports VMkernel et ports de VM sont toujours nécessaires.

Network I/O Control sera traité dans la partie optimisation du réseau.

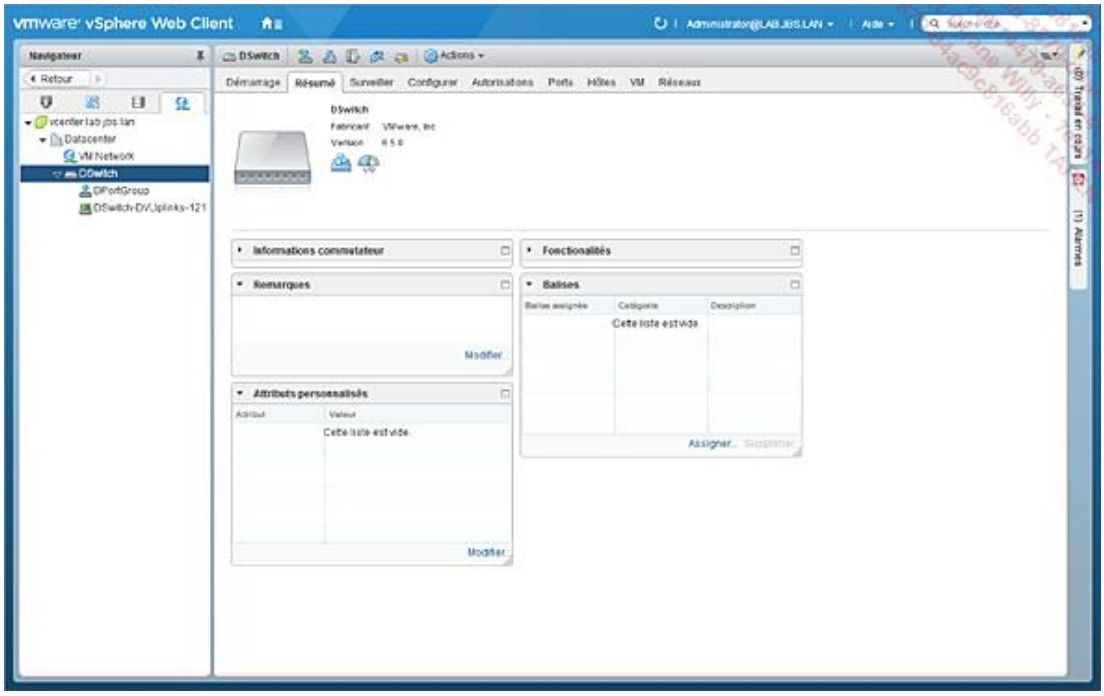


Le résumé des paramètres est présenté pour vérification avant la création du vSwitch distribué en cliquant sur le bouton **Terminer**.



Avant la création d'un vSwitch distribué, il convient de vérifier les configurations réseau et notamment celles concernant les liaisons montantes des hyperviseurs concernés. Afin d'éviter les déconnexions une fois les machines virtuelles migrées ou créées sur le vSwitch distribué, attention à bien identifier les vmnic à utiliser.

Le vSwitch distribué apparaît dans l'inventaire.

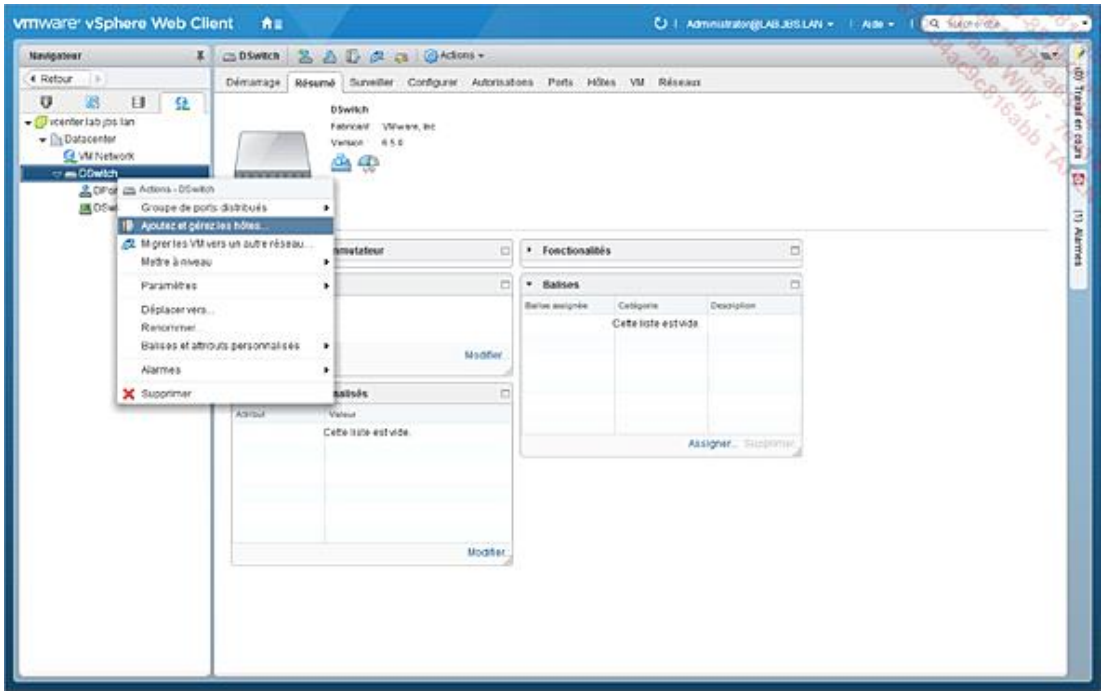


Nous pouvons désormais passer à la configuration basique de notre vDS en y ajoutant les quatre hôtes de notre cluster.

b. Ajout des hôtes au sein d'un vSwitch distribué

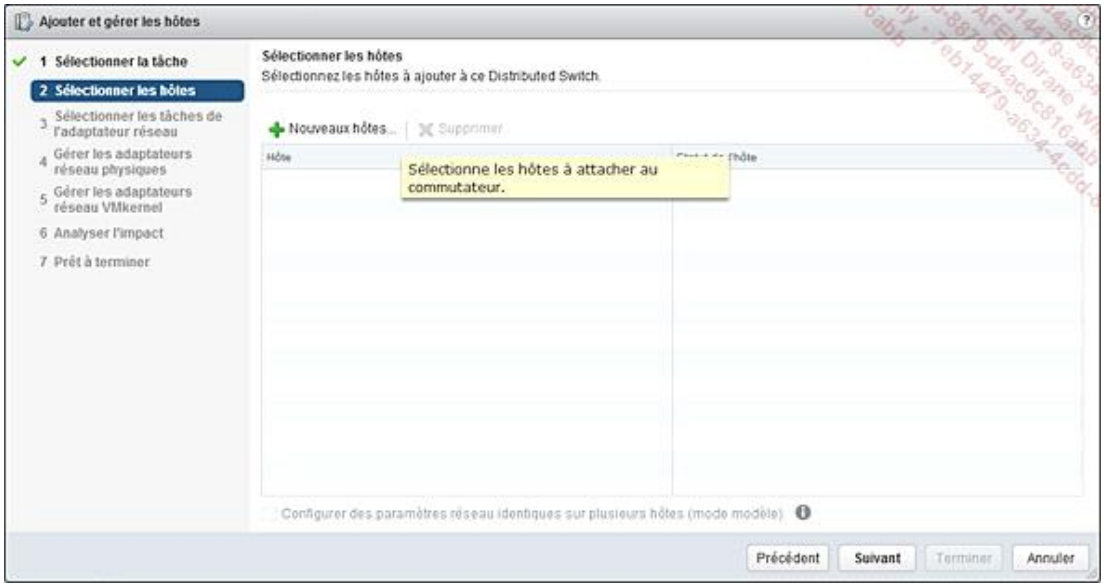
L'ajout d'hôtes au sein d'un vSwitch distribué est simple, voici comment procéder.

La première étape consiste à effectuer un clic droit sur le vSwitch distribué récemment créé puis à sélectionner l'option **Ajoutez et gérez les hôtes**.

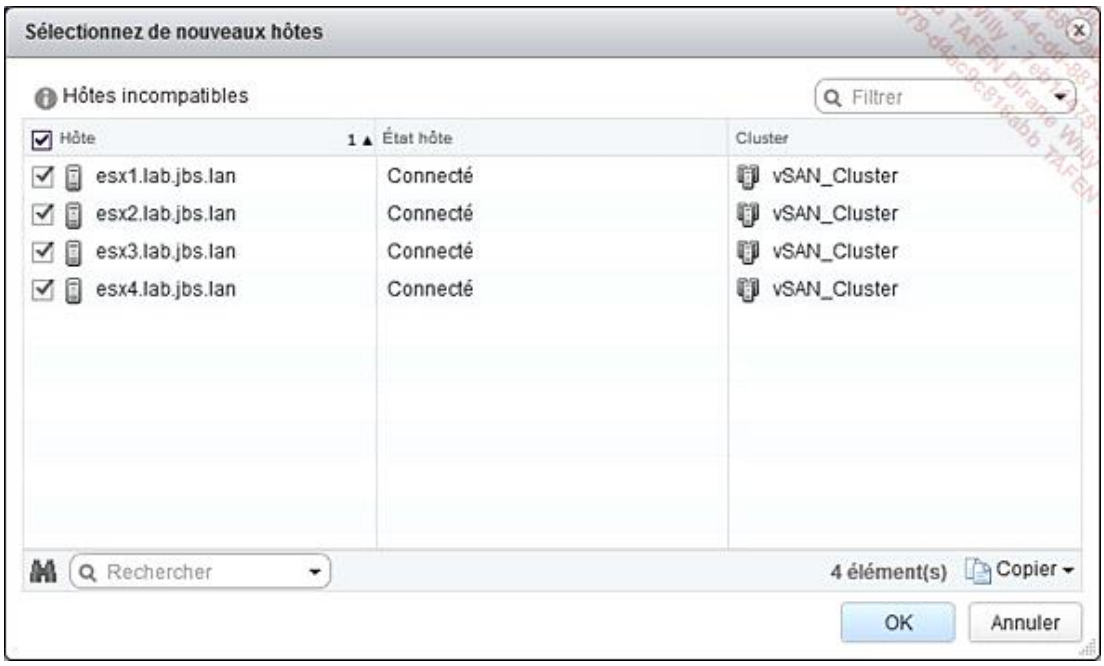


Nous sélectionnons ce que nous souhaitons faire, ici ajouter des hôtes au sein de notre vSwitch distribué.

À l'aide du bouton **Nouveaux hôtes**, lancez l'ouverture de la fenêtre pour sélectionner les hôtes à ajouter.

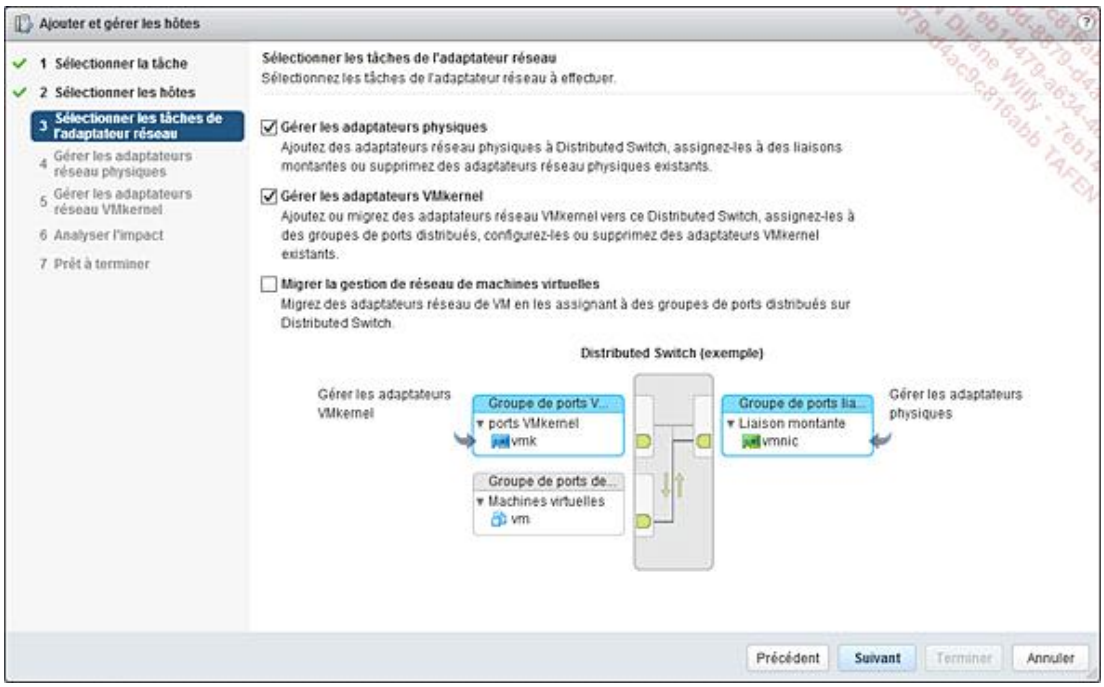


Nous sélectionnons nos quatre hyperviseurs vSphere, par ailleurs membres d'un même cluster comme vous pouvez le constater.



Vous disposez ici de trois options :

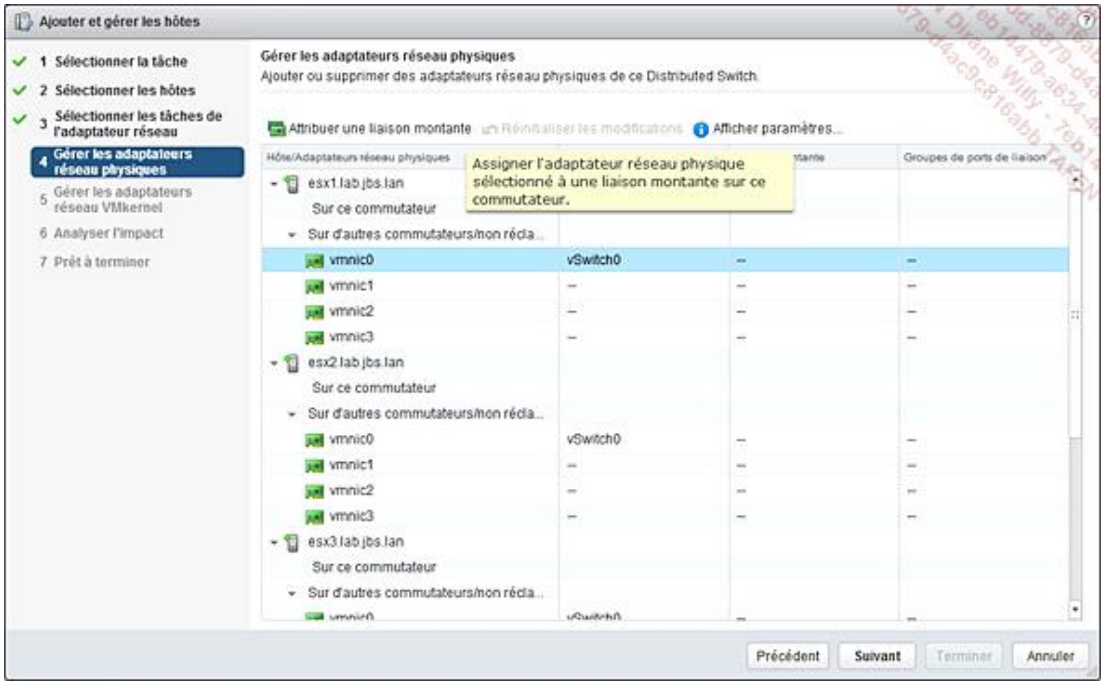
- **Gérer les adaptateurs physiques** pour associer des ports physiques au vSwitch distribué.
- **Gérer les adaptateurs VMkernel** pour l'ajout, la migration, la gestion d'interfaces VMkernel sur le vSwitch distribué.
- **Migrer la gestion de réseau de machines virtuelles** pour migrer les interfaces réseau virtuelles des machines virtuelles sur le vSwitch distribué.



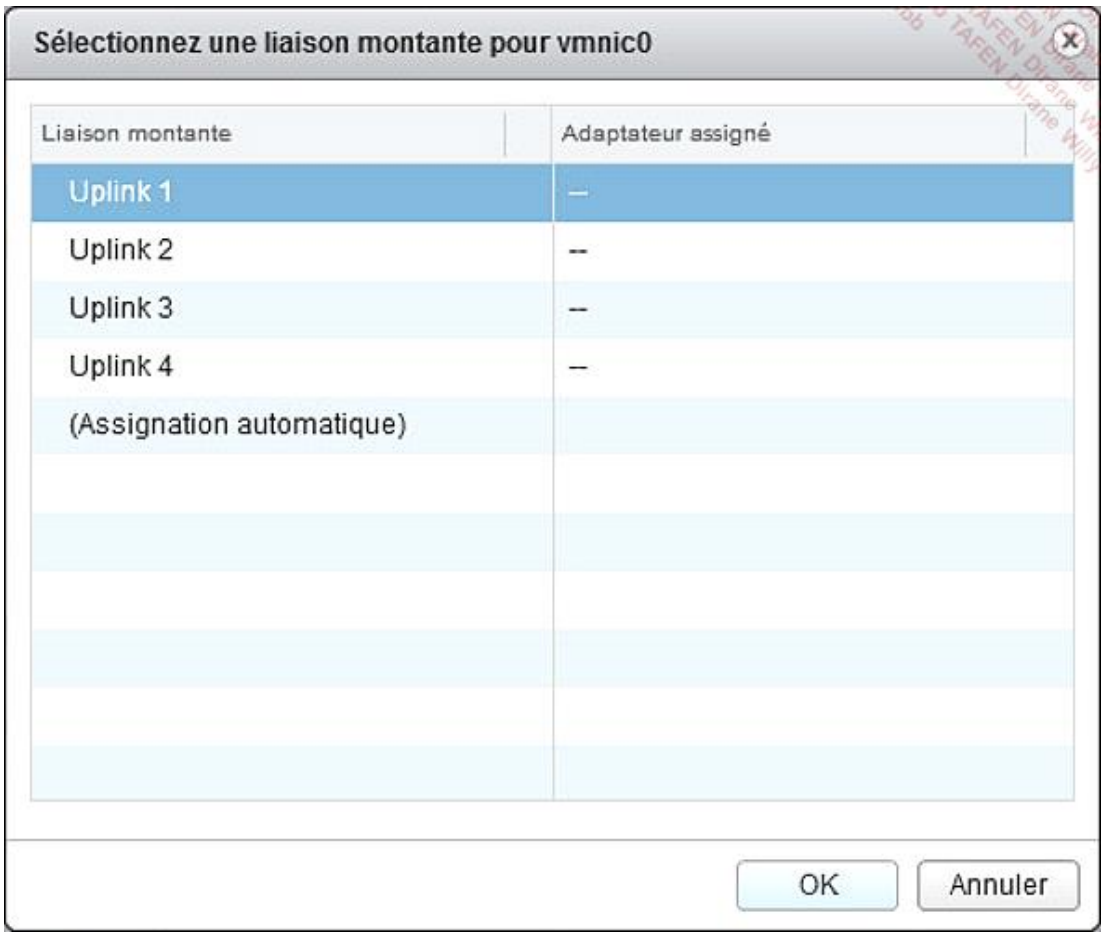
Nous cochons ici les deux premiers choix pour nous permettre de configurer dans l'assistant, à la fois les liaisons montantes ainsi que les adaptateurs VMkernel.

Dans l'écran suivant, il s'agit d'associer une (ou plusieurs) carte réseau pour la joindre au vSwitch distribué.

Ici nous souhaitons joindre l'adaptateur réseau vmnic0 de chaque hyperviseur au vSwitch distribué. Pour ce faire, sélectionnez le port vmnic0 puis cliquez sur le bouton **Attribuer une liaison montante**.

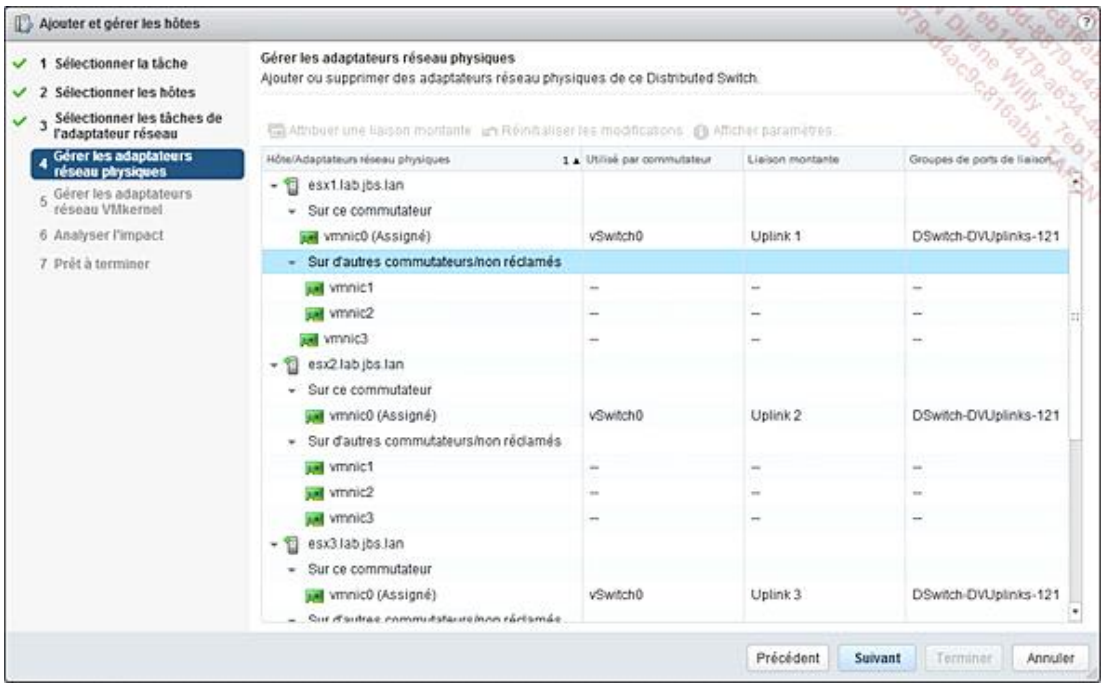


Assignez une liaison montante au port physique puis validez.

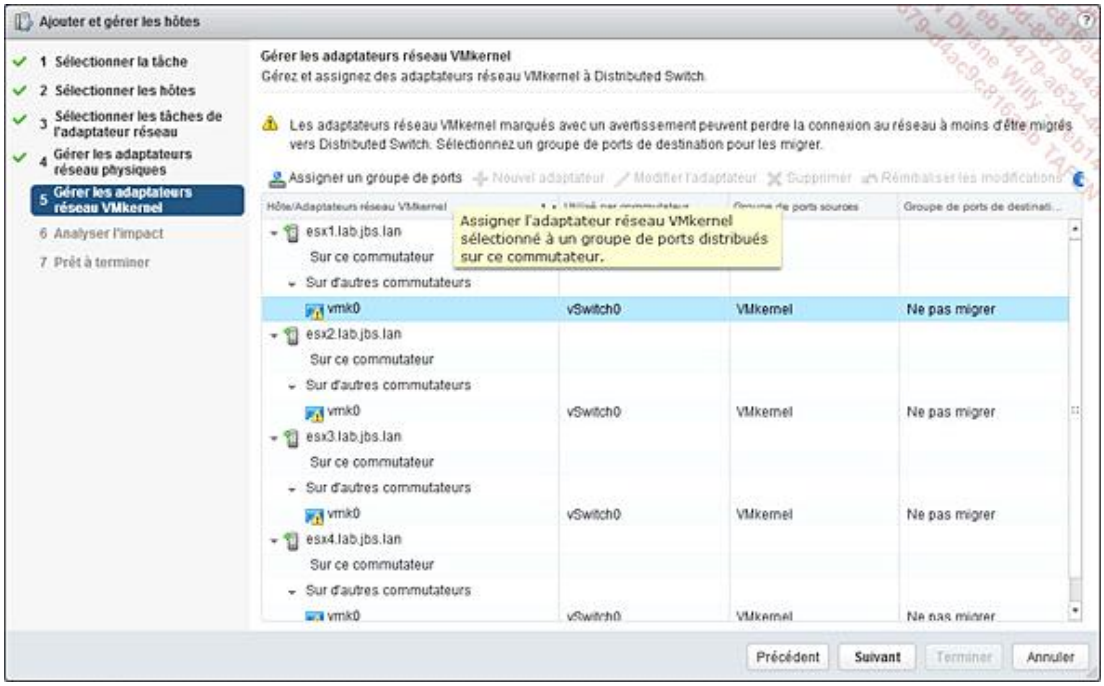


Les noms par défaut des liaisons montantes peuvent être changés après création.

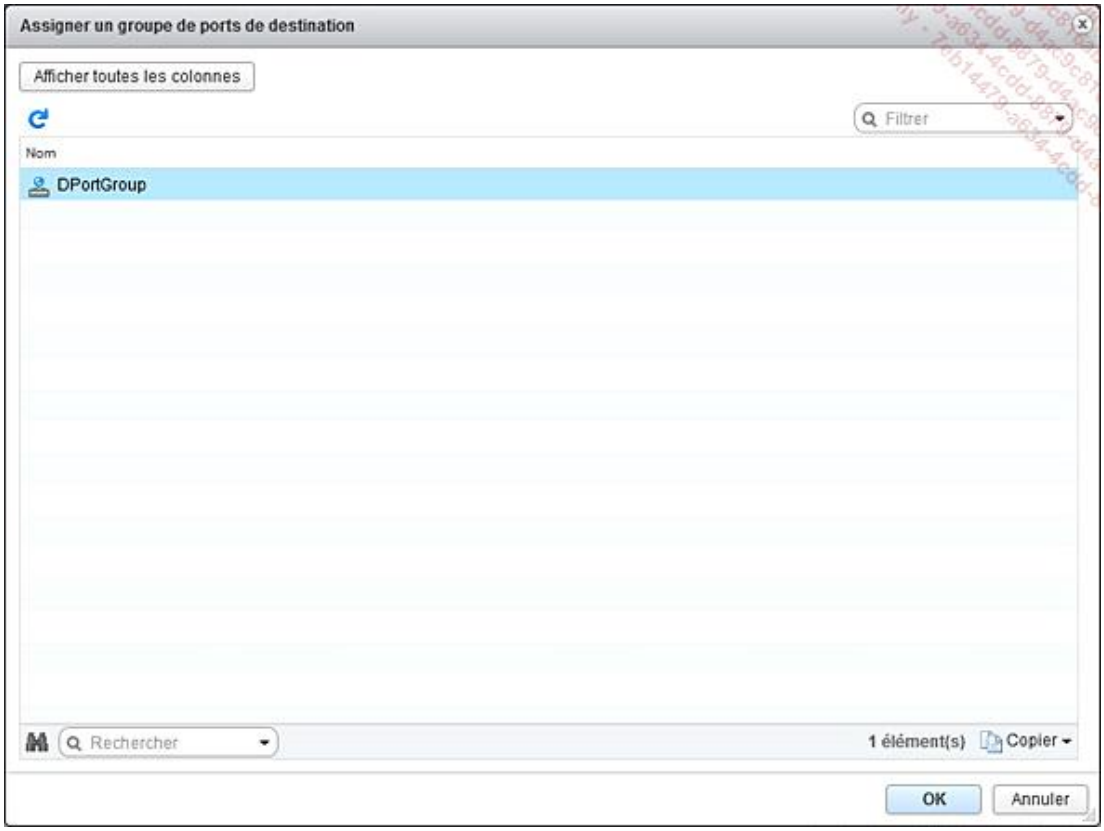
Répétez l'opération pour chaque hyperviseur à configurer. Voyez l'assignation réussie du port vmnic0 aux liaisons montantes de chaque hyperviseur sur la capture ci-dessous.



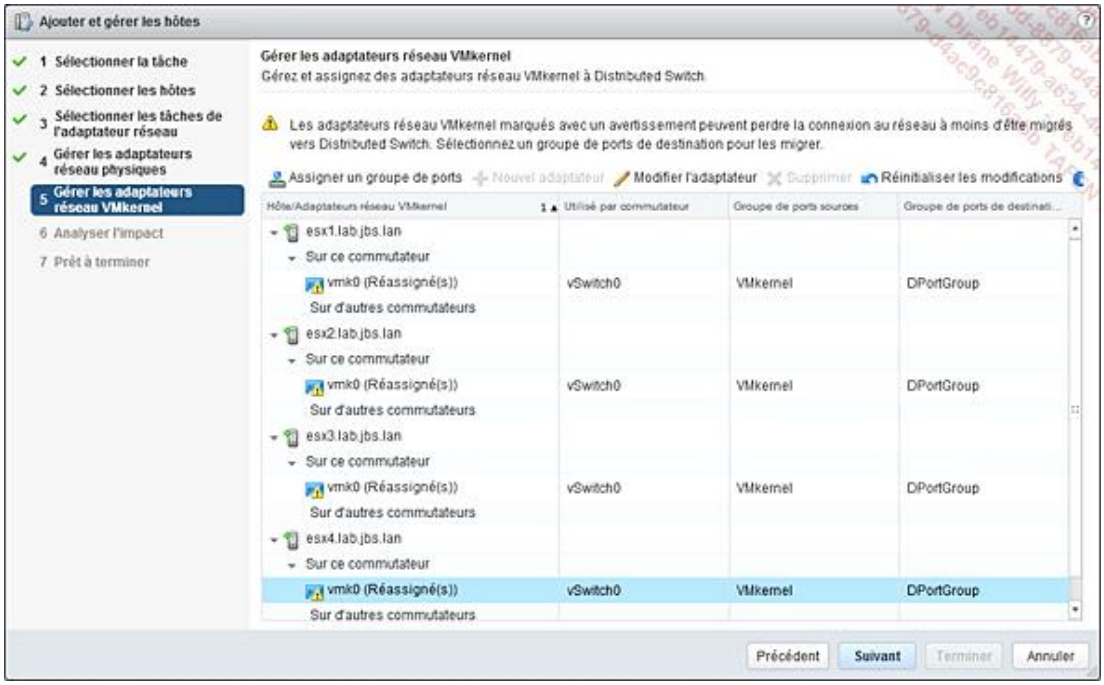
Sur le prochain écran, il s'agit de migrer les interfaces VMkernel vers le groupe de ports par défaut créé dans le vSwitch distribué, « DPortGroup ». Pour chaque hyperviseur, assignez un groupe de ports à l'aide du bouton **Assigner un groupe de ports**.



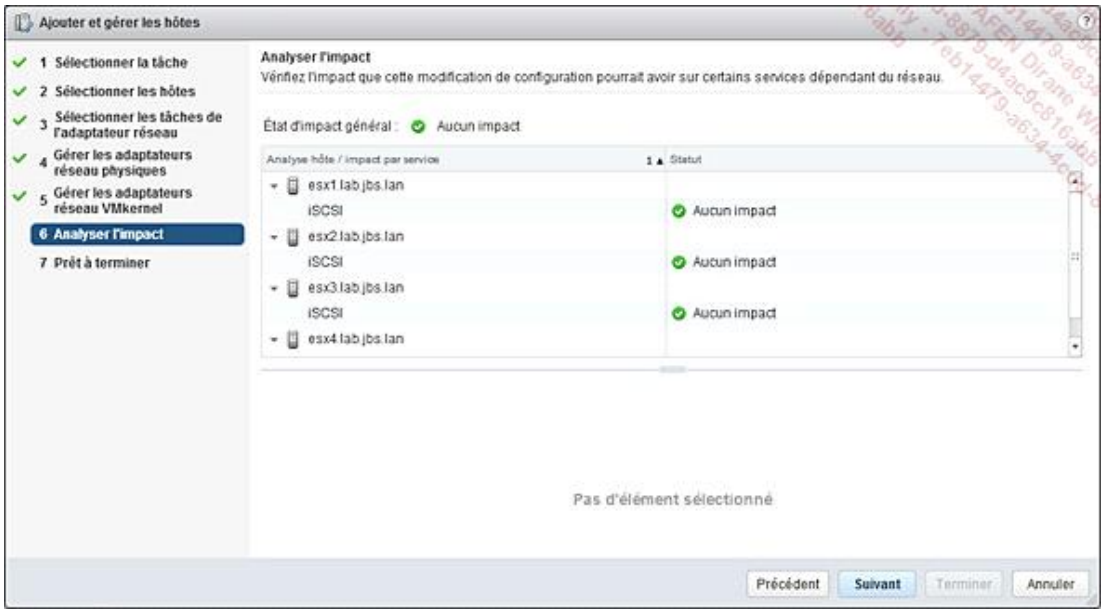
Puis désignez le groupe de ports.



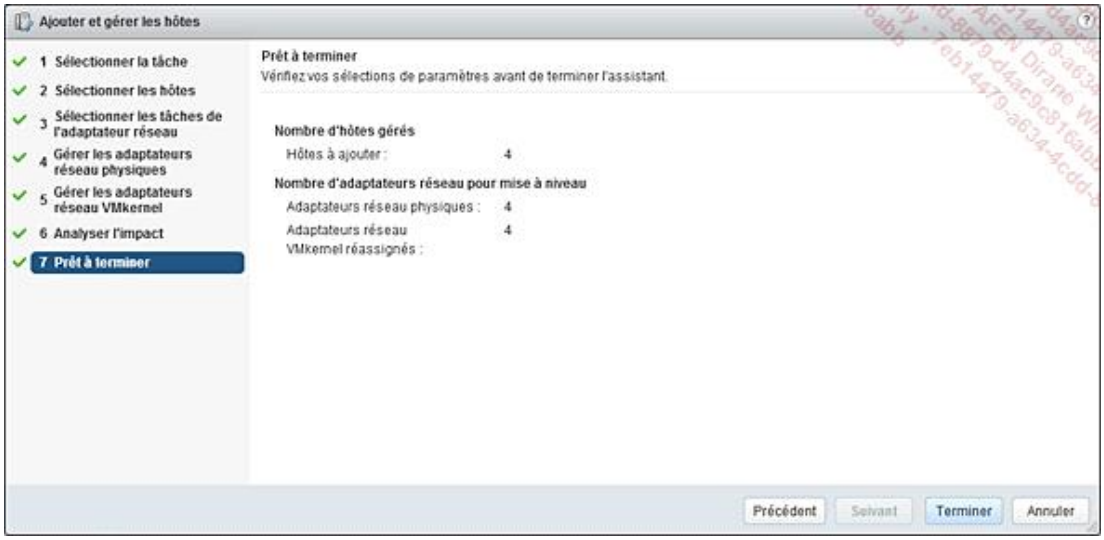
Voici l'écran que vous devriez obtenir après la configuration des hôtes.



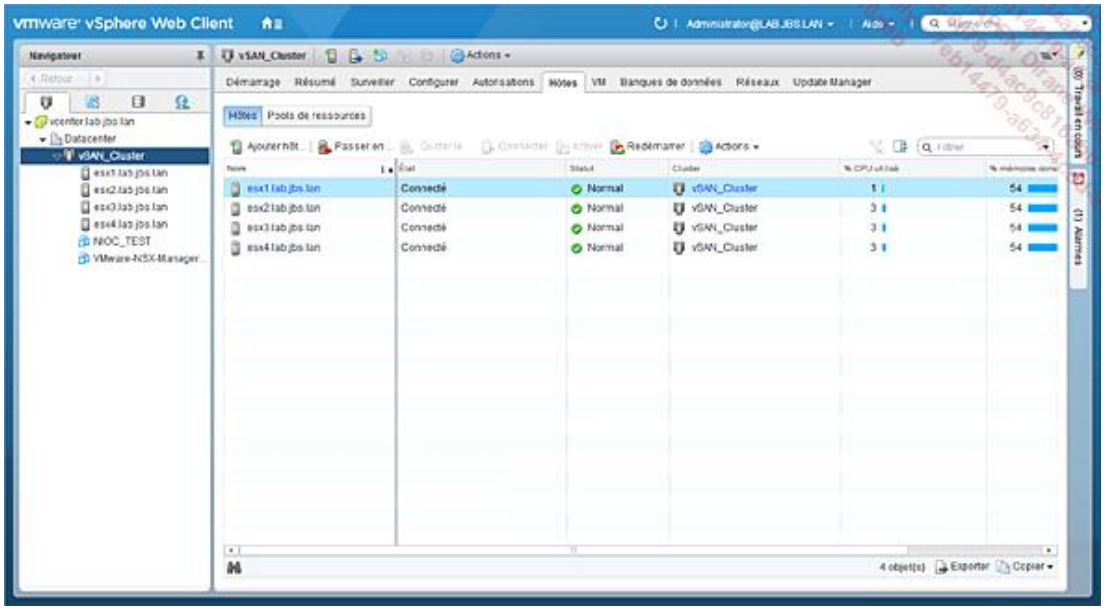
Dans l'avant-dernier écran, vSphere vérifie que la migration n'aura pas d'impact sur les liaisons iSCSI. Aucun impact n'est ici prévu, nous pouvons donc continuer à l'aide du bouton **Suivant**.



Vérifiez vos changements dans l'écran final qui affiche le récapitulatif des actions qui vont être entreprises à la validation.



À la fin de la création, l'administrateur peut visualiser les hôtes joints et leur statut **Normal**, qui confirme que la migration s'est passée comme prévu.



Sur chaque serveur hôte faisant partie d'un vSwitch distribué, il y a en fait la partie I/O plane ou gestion d'entrées-sorties réseau gérée par un vSwitch caché dit proxy. Ce vSwitch proxy a des caractéristiques différentes selon la version du serveur ESXi associé. Pour les versions vSphere 5.5 et ultérieures le vSwitch proxy a un nombre de ports qui varie dynamiquement selon les services réseau et les machines virtuelles en fonctionnement. Pour vSphere 5.1 et antérieur, le nombre de ports sur le vSwitch proxy est fixe. Cela génère une consommation de ressources supplémentaires si le nombre de ports est trop important. La ressource est réservée pour la gestion des ports du vSwitch proxy, qu'ils soient utilisés ou non. C'est pourquoi quand on parle de vSwitch proxy pour serveurs hôtes hérités (legacy ou vSphere 5.1 et antérieur), il convient de limiter le nombre de ports au strict minimum.

c. VLAN

Concernant la gestion des VLAN s'appliquant sur un vSwitch distribué, plusieurs scénarios sont possibles :

- **VLAN** : un seul et unique VLAN est configuré sur un groupe de ports (comme pour un vSwitch standard). Le modèle reste le même qu'avec un vSwitch standard, le **VST** (*Virtual Switch Tagging*).
- **VLAN Trunking** : on autorise la machine virtuelle à former un trunk avec un ou plusieurs VLAN. On propage donc

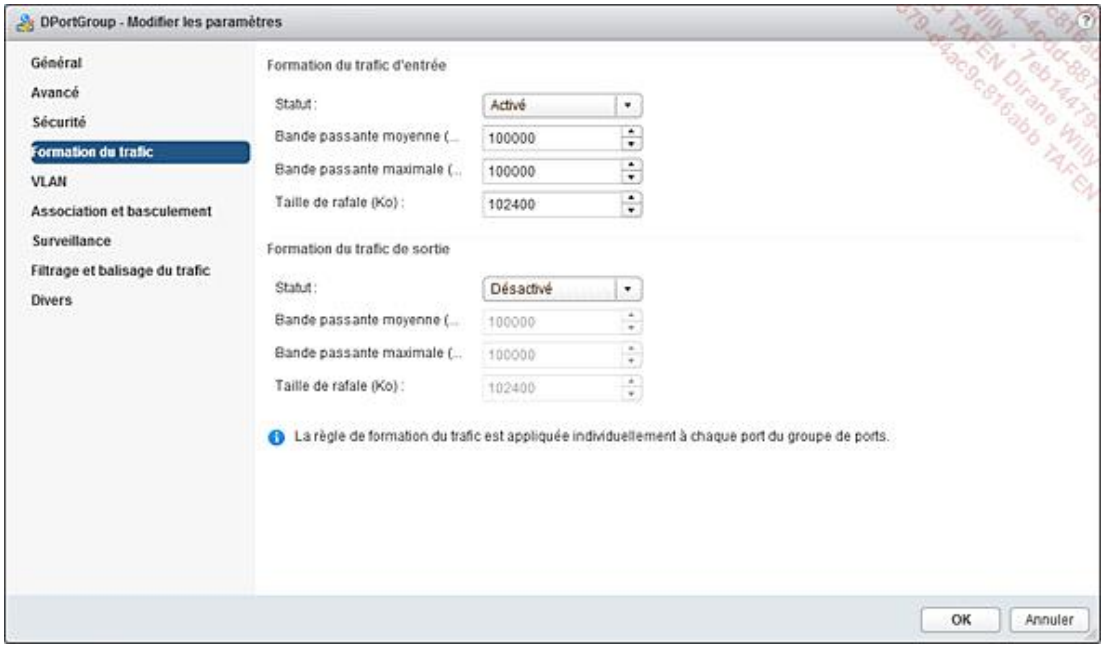
les ID de VLAN des trames jusqu'à la machine virtuelle qui se chargera de les traiter. Bien entendu, la machine virtuelle étiquettera directement les trames en sortie. L'administrateur est libre de saisir une plage ou liste de VLAN qu'il souhaite fournir à la machine virtuelle. Le modèle reste le même qu'avec un vSwitch standard, le VGT (*Virtual Guest Tagging*).

- **Private VLAN** : l'administrateur associe un groupe de ports avec un « Private VLAN » existant. Dans les réseaux informatiques, cette fonctionnalité est très utilisée pour offrir une isolation de niveau 2 plus fine tout en permettant l'utilisation d'un seul domaine de niveau 3. Trois modes sont usuellement définis : promiscious, isolated, community.
  - **Promiscious** représente généralement la liaison montante, notamment un routeur. Ce port peut communiquer avec les deux autres modes.
  - De la perspective des clients (c'est-à-dire ceux qui se connectent au réseau), le mode **isolated** n'autorise un hôte à ne communiquer qu'avec un port promiscious, tandis que le mode community permet la communication entre les ports du même type (community) et le port promiscious.
  - Du point de vue de la sécurité, il peut être très utile pour un administrateur d'utiliser le mode isolated pour empêcher les hôtes de communiquer directement entre eux et les forcer à passer par un mécanisme de routage. Cependant, si deux postes ne peuvent pas communiquer directement ensemble, prêtez attention à ce que les mécanismes de découverte locaux comme ARP soient toujours fonctionnels via un routeur (proxy ARP).

d. Traffic shaping

Contrairement aux vSwitches standards, le lissage de trafic est géré en sortie **et** en entrée du vSwitch.

Les options à configurer restent les mêmes que pour un vSwitch standard à savoir : **Bande passante moyenne**, **Bande passante maximale** et **Taille de rafale**.



Les paramètres de traffic shaping en entrée et en sortie peuvent être gérés de manière indépendante, ce qui offre une flexibilité appréciable à l'administrateur.

4. Fonctions avancées

a. Utilisation des cartes 10 GB/s et plus

Après l'avènement des cartes Gigabit Ethernet il y a quelques années, on peut maintenant les considérer comme banales. Banales au point que désormais, la vitesse Gigabit est applicable dans un grand nombre d'endroits, y compris chez vous ! La plupart des « box » domestiques supportent cette vitesse sur leurs parties de commutation.

Il est évident que pour les serveurs et les réseaux en entreprise, il y a eu des évolutions plus flagrantes. Parmi les nouveaux standards utilisés en entreprise, on peut compter les vitesses 10, 25, 40, 50 et même 100 Gbit/s ! Apparus de façon plus récente, les commutateurs Cisco mGig (multigigabit) vous permettent de pouvoir monter à des vitesses jusqu'à 10 Gbit/s, sans avoir à changer toute votre architecture (y compris votre câblage).

Aujourd'hui, vSphere 6.5 supporte plusieurs types de cartes selon le fabricant et la vitesse voulue. En voici une liste non exhaustive :

Vitesse	Driver	Nombre maximal d'adaptateurs supportés par hyperviseur
1 Gbit/s	igbn (Intel)	16
10 Gbit/s	ixgbe (Intel)	16
25 Gbit/s	qedentv (Qlogic)	4
50 Gbit/s	qedentv (Qlogic)	4
100 Gbit/s	qedentv (Qlogic)	2

b. Network I/O Control (NIOC)

vSphere permet à un administrateur de contrôler l'utilisation et la réservation des ressources. Dans un cluster classique, les pools de ressources permettent une régulation de l'utilisation processeur et mémoire au profit des machines virtuelles. Mais qu'en est-il concernant un autre composant essentiel dans un cluster, le réseau ?

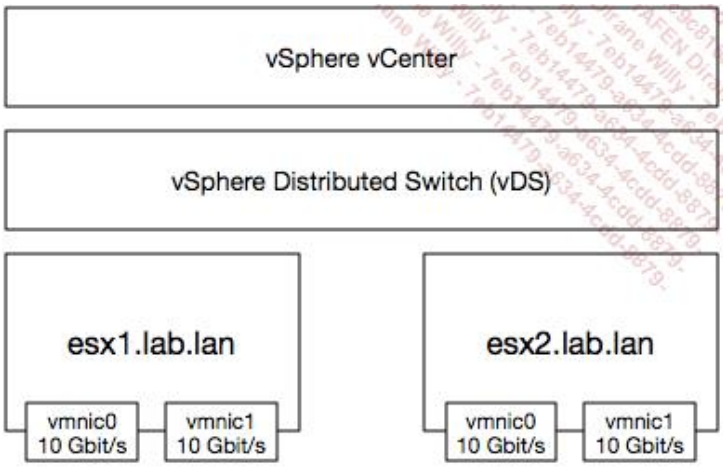
VMware propose pour réguler l'utilisation réseau dans un cluster un mécanisme appelé Network I/O Control (souvent appelé par son acronyme - NIOC). NIOC permet, comme les pools de ressources classiques, d'instaurer une hiérarchie d'utilisation des ressources réseaux par rapport à la demande des machines virtuelles et de leurs accès concurrents. NIOC permet en outre de garantir (par un système de réservation) la bande passante pour les services système.

Notez que vous ne pouvez utiliser NIOC que si et seulement si vous disposez d'une architecture utilisant un vSwitch distribué (vDS).

vSphere 6.0 (et donc 6.5) propose une nouvelle version du NIOC, la version 3 qui sépare deux types de trafic. L'utilisation liée aux activités système (Fault Tolerance, vMotion, iSCSI, vSAN, Management,…) et celle des machines virtuelles à proprement parler.

Concernant les activités systèmes, cette nouvelle version permet de réserver la bande passante dédiée aux services systèmes en fonction de la capacité des adaptateurs réseau installés sur les hôtes.

Prenons un schéma pour illustrer notre propos :



Nous avons deux hôtes, esx1.lab.lan et esx2.lab.lan. Les hôtes sont dotés de deux adaptateurs réseau, d'une vitesse de 10 Gbit/s chacun. La réservation concernant les services système peut se faire comme suit :

Service système	Réservation de bande passante (en Gbit/s)
Management	1.0
iSCSI	1.0
Fault Tolerance	1.0
vMotion	2.0
Machine virtuelle	0.5
vSAN	1
Capacité restante	1
Total	7.5

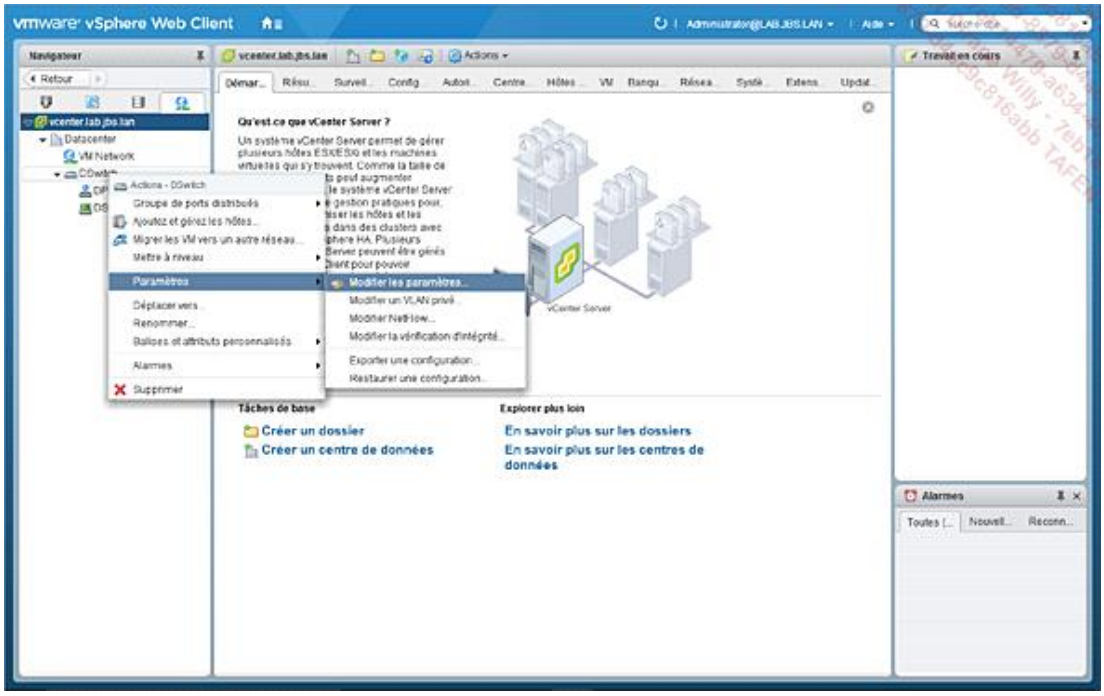
Cette réservation peut se commenter en plusieurs points.

Vous avez peut-être remarqué que la réservation cumulée est ici de 7.5 Gbit/s alors qu'un adaptateur fonctionne à la vitesse de 10 Gbit/s. Cela s'explique par le fait que l'administrateur peut réserver au maximum 75 % de la capacité d'un adaptateur.

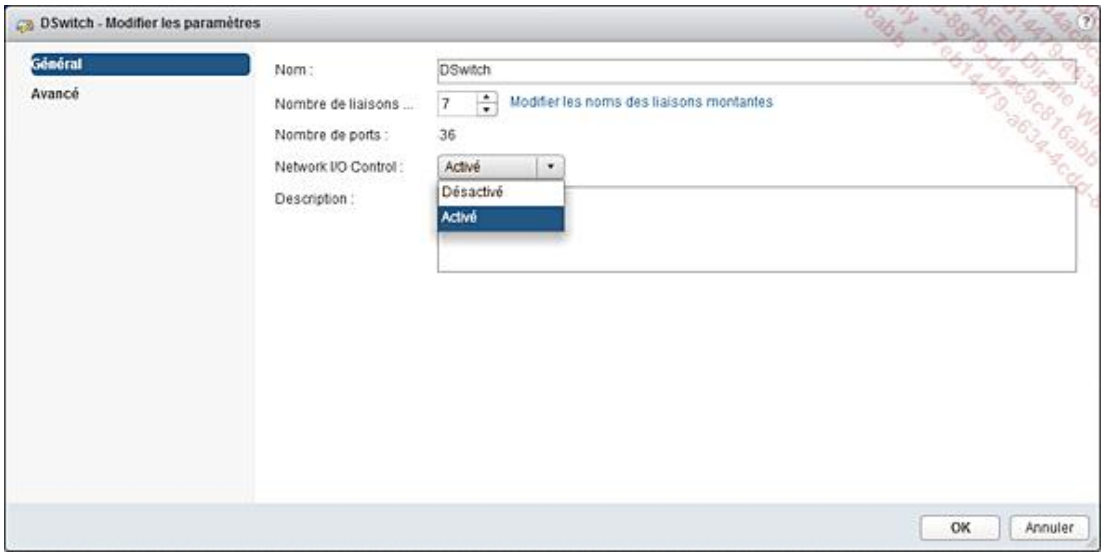
Le trafic système « Machine virtuelle » peut vous interpeller à cause de son classement comme service système. Le trafic système attribué à cette entité sera le trafic maximum réservé pour l'usage réseau des machines virtuelles.

Enfin, la réservation s'appliquant sur chaque adaptateur réseau, nous formulons ici nos réservations en nous adaptant à ce paramètre. Nous n'aurions pas le droit de réserver 15 Gbit/s pour cette raison.

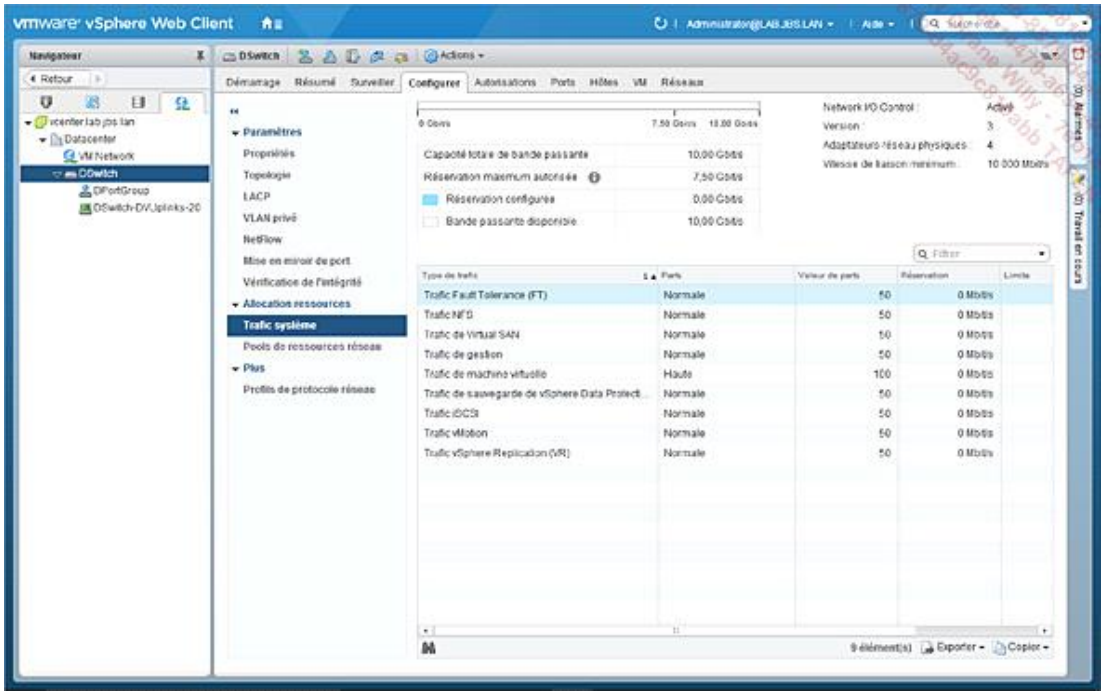
La mise en œuvre est assez aisée et s'effectue comme suit. Tout d'abord, activez NIOC en allant sur les paramètres du vSwitch distribué.



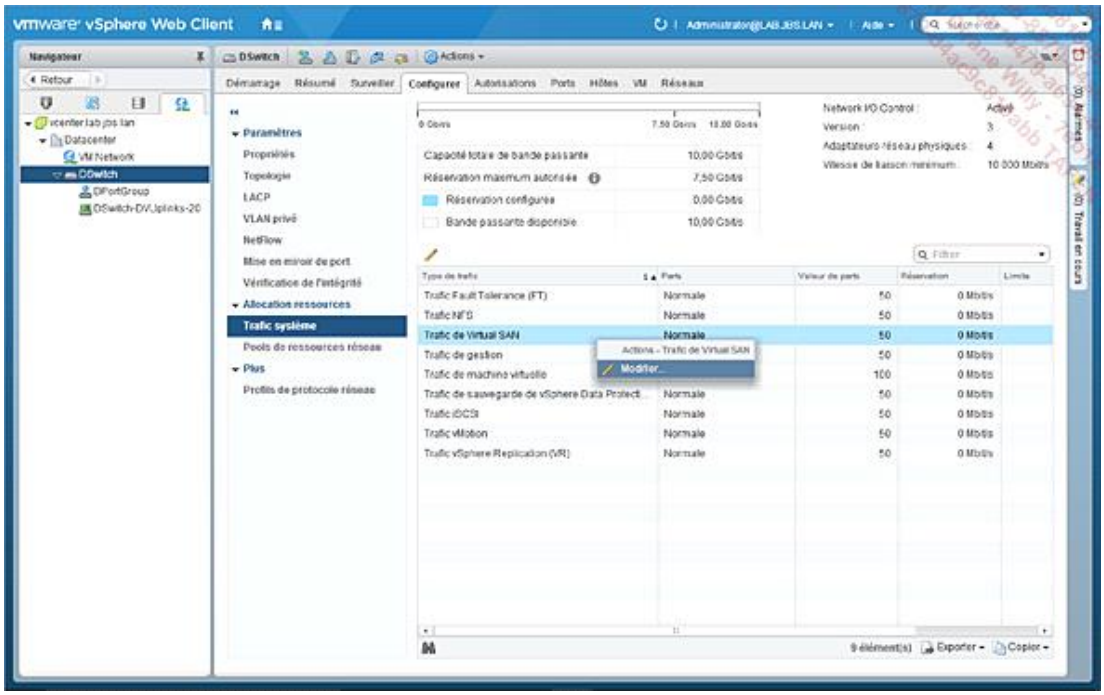
Puis, sélectionnez l'option pour activer NIOC et validez.



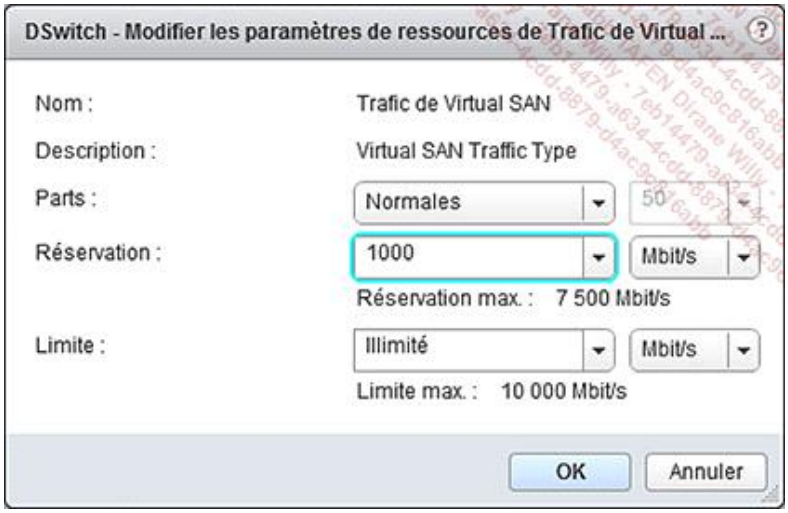
Puis, rendez-vous dans l'onglet **Configurer** du vSwitch distribué présentant les parts, valeurs de parts et réservations par défaut.



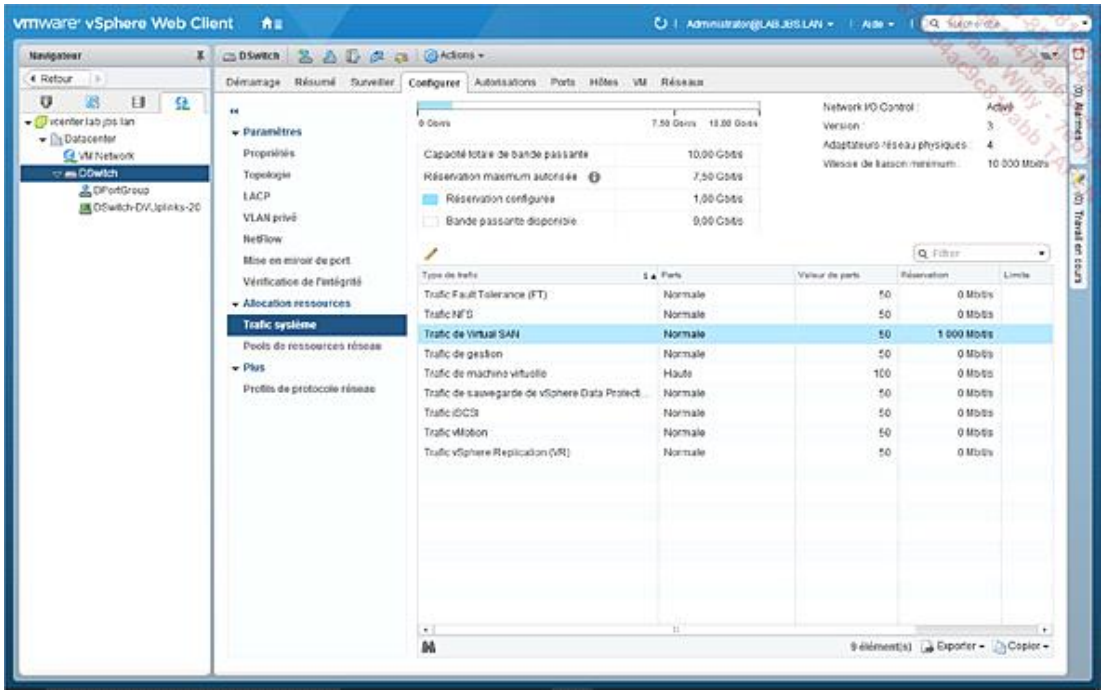
Changeons ici la réservation de trafic pour le cluster vSAN à l'aide d'un clic droit sur la ligne visée.



Nous pouvons alors changer les paramètres pour cette catégorie de trafic et y associer une réservation de l'équivalent de 1 Gbit/s. Puis validez à l'aide du bouton **OK**.

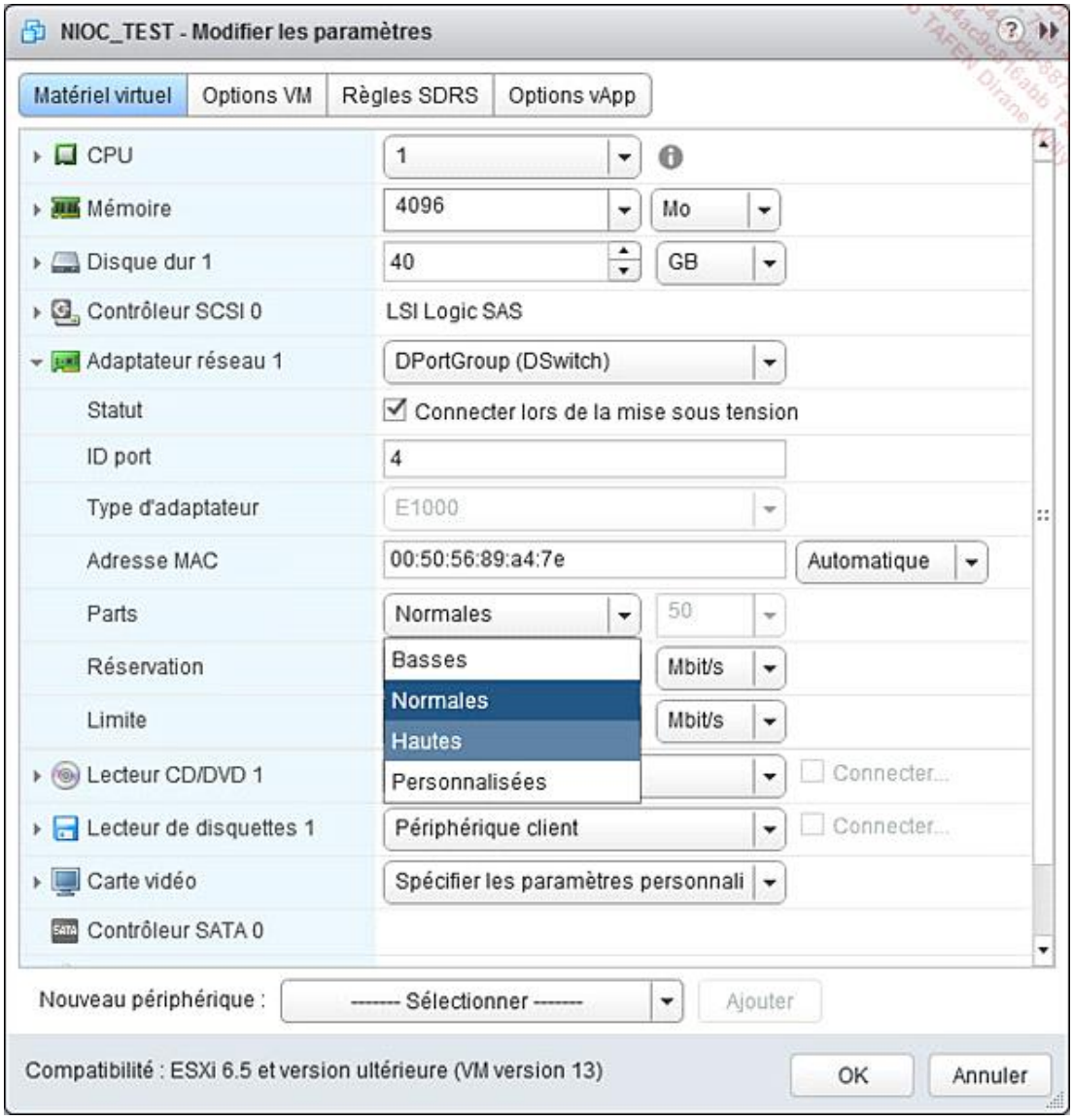


Le trafic réservé apparaît alors dans le tableau de bord.



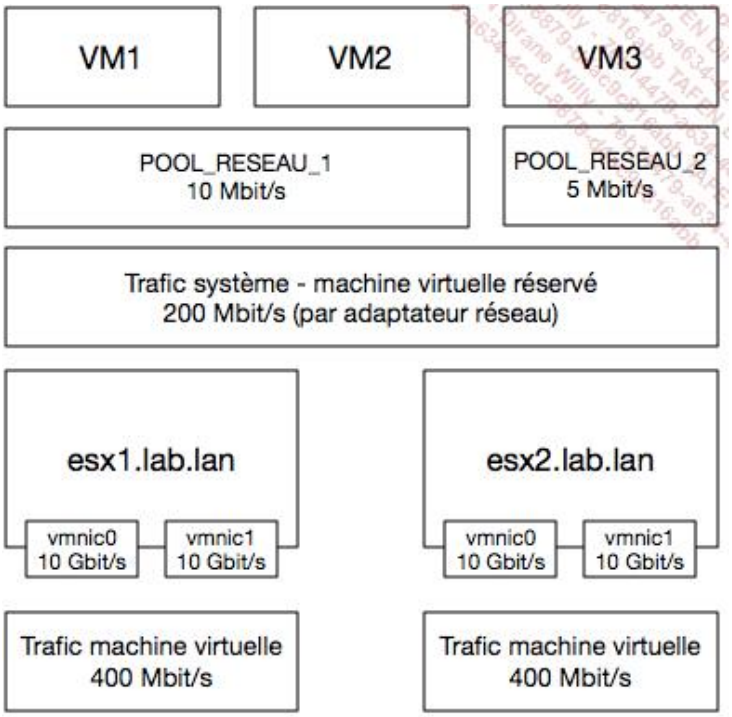
NIOC 3.0 permet à l'administrateur de spécifier la réservation de bande passante mais également sa part (shares en anglais) au niveau d'une machine virtuelle individuelle, comme vous le feriez pour les ressources processeur et mémoire.

La capture ci-dessous vous montre comment modifier les parts concernant le réseau dans la configuration d'une machine virtuelle :



De plus, ces paramètres sont appliqués au niveau du vSwitch distribué et non plus de l'hôte comme c'était le cas dans la version 2.0.

Pour illustrer notre propos avec une configuration à portée plus globale, prenons les pools de ressources réseau en exemple et l'impact de cette configuration à l'aide de la figure suivante :



Nous avons ici trois machines virtuelles, liées à deux pools de ressources par le biais de leurs connexions à un groupe de ports distribué. Ces deux pools réservent chacun 10 et 5 Mbit/s de bande passante.

Ainsi, les machines virtuelles VM1 et VM2 se partageront le débit garanti de 10 Mbit/s tandis que la VM3 disposera d'un débit garanti de 5 Mbit/s.

Le trafic système défini au niveau du vSwitch distribué est ici fixé à 200 Mbit/s (rappelez-vous que cette valeur s'applique par adaptateur réseau et par hôte). Il spécifie la réservation maximale qui pourra être consommée par les pools de ressources réseau, mentionnés plus haut. En outre, cette valeur de trafic système réservera 400 Mbit/s par hôte concernant le trafic de type machine virtuelle.

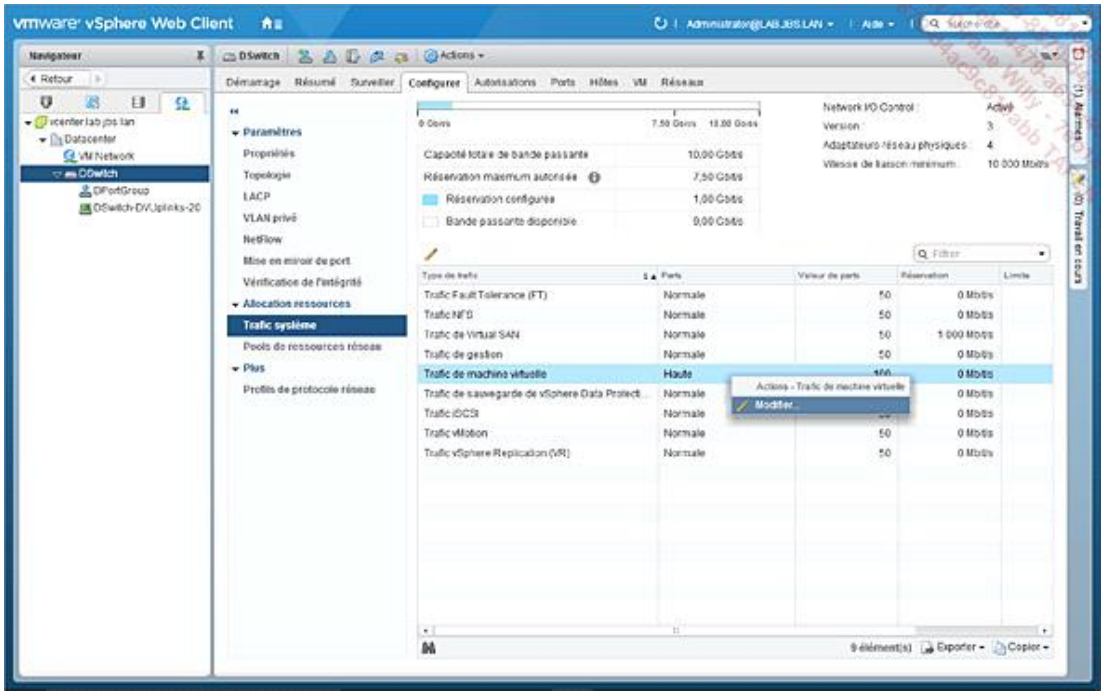
La formule à utiliser pour calculer la réservation globale (parmi tous les hôtes du cluster) réservée est la suivante :

$$Nombre\ de\ pNIC\ (adaptateurs\ réseau) * réservation\ configurée$$

$$4 * 200\ Mbit/s = 800\ Mbit/s$$

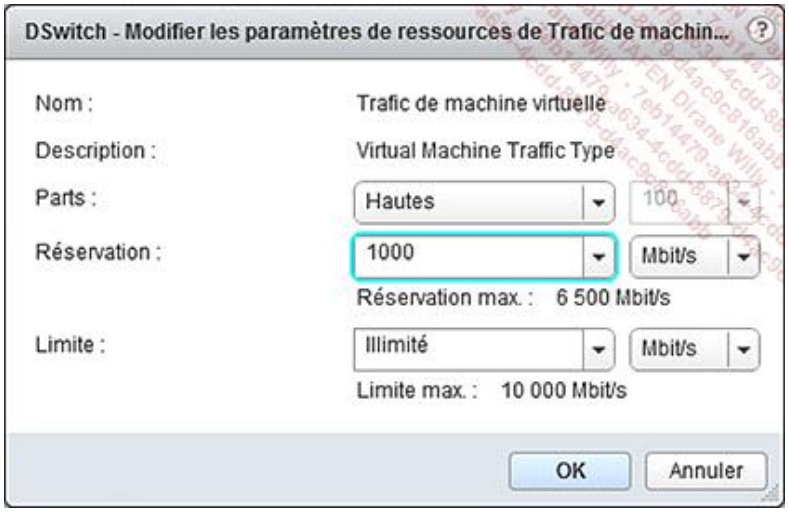
La partie implémentation est simple et est décrite ci-dessous. Dans cet exemple, notre cluster est doté de 4 hôtes avec un adaptateur réseau de 10 Gbit/s sur chacun de ceux-ci.

Dans un premier temps, spécifions au niveau du vSwitch distribué le trafic réservé aux machines virtuelles. Pour ce faire, effectuez un clic droit sur la ligne visée puis sélectionnez l'option **Modifier**.

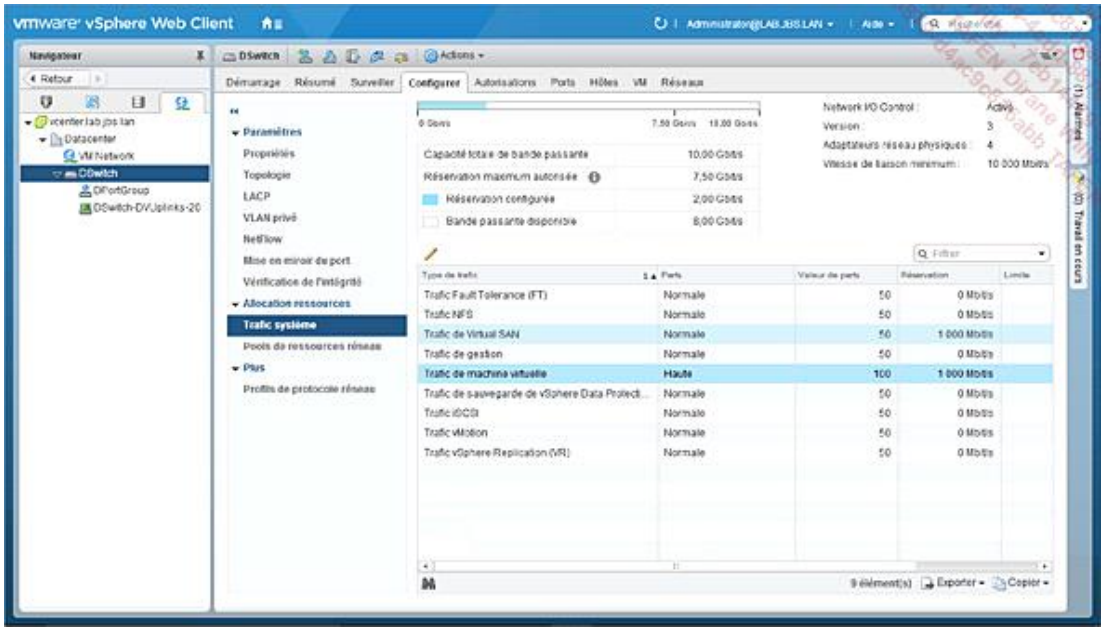


Nous fixons ici la réservation à hauteur de 1 Gbit/s.

Souvenez-vous que cette valeur est associée par hôte et par adaptateur réseau présent sur ce dernier. Ainsi, le cluster offrira une réservation globale égale à 4 Gbit/s aux machines virtuelles. Une fois la valeur spécifiée, validez à l'aide du bouton **OK**.



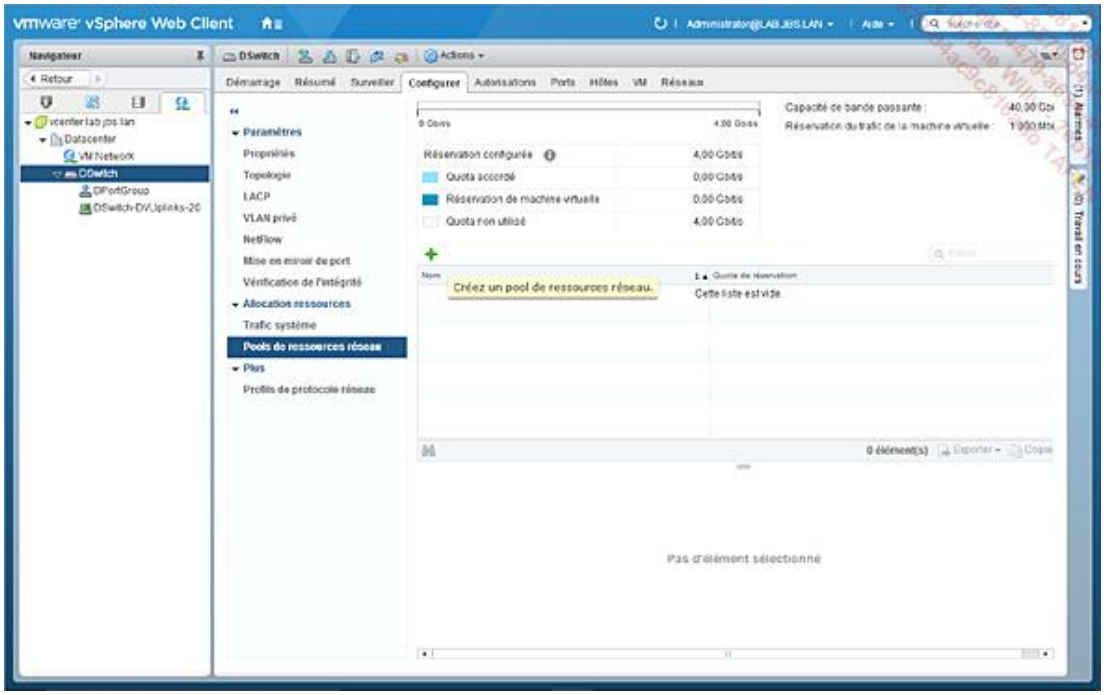
Nous pouvons consulter la mise en place de votre réservation grâce à la mise à jour du tableau de bord.



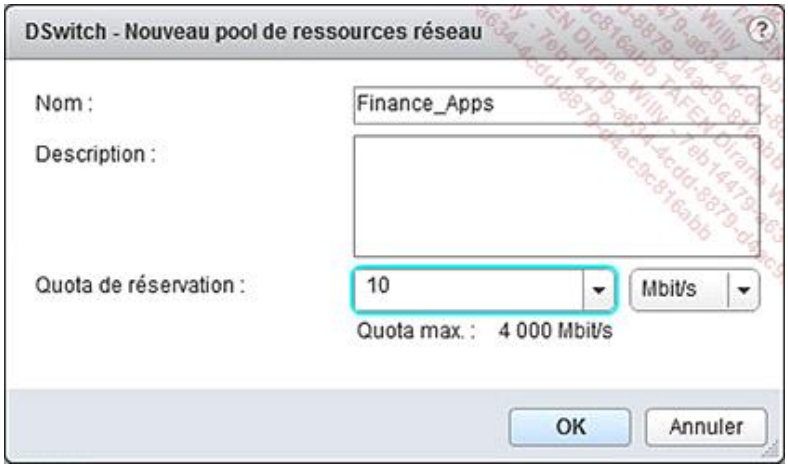
Rendez-vous dans la section **Pools de ressources réseau**.

Constatez ici la réservation opérationnelle de 4 Gbit/s grâce à notre réservation destinée aux machines virtuelles de 1 Gbit/s. Remarquez que la réservation vSAN configurée auparavant n'apparaît pas dans ce calcul, étant remise à zéro pour montrer uniquement dans notre cas, l'impact de la réservation du trafic machine virtuelle.

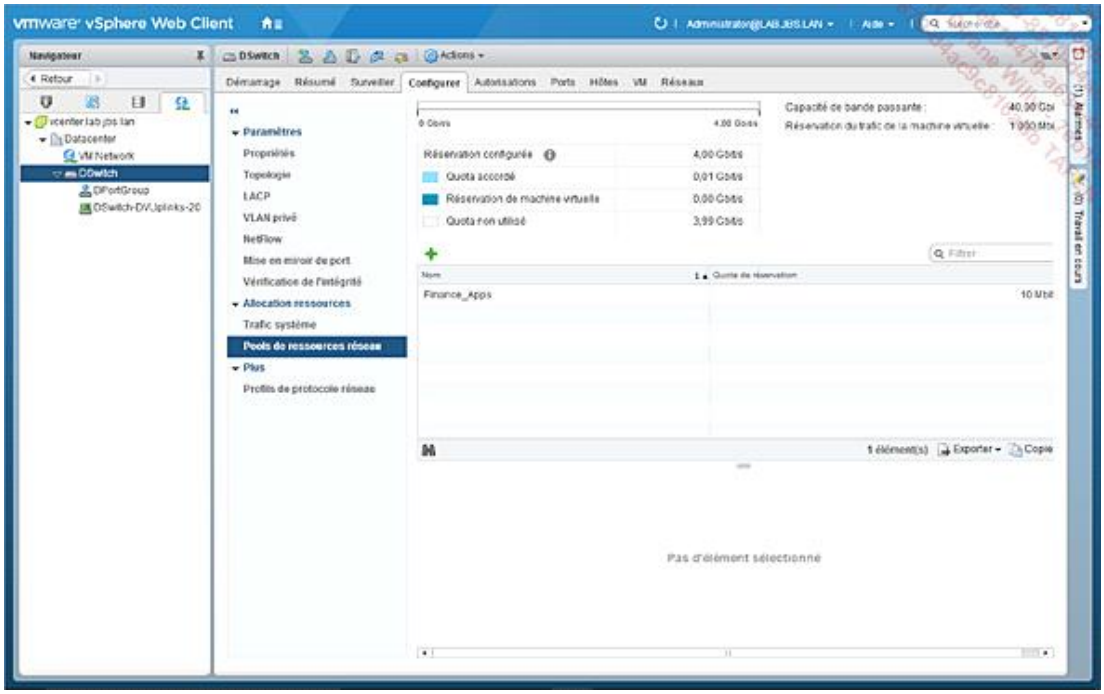
Créons ensuite un pool de ressources réseau qui pourra être consommé par des machines virtuelles, à l'aide du bouton **+**.



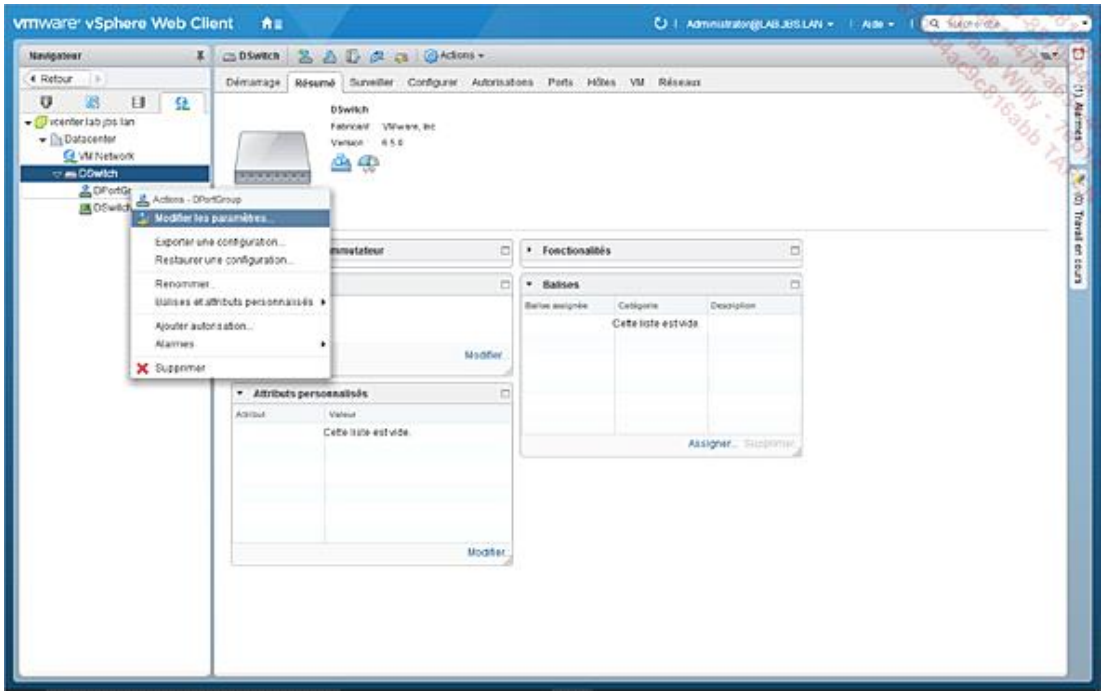
Nous nommons notre pool « Finance_Apps » et fixons une réservation à 10 Mbit/s au niveau de ce dernier.



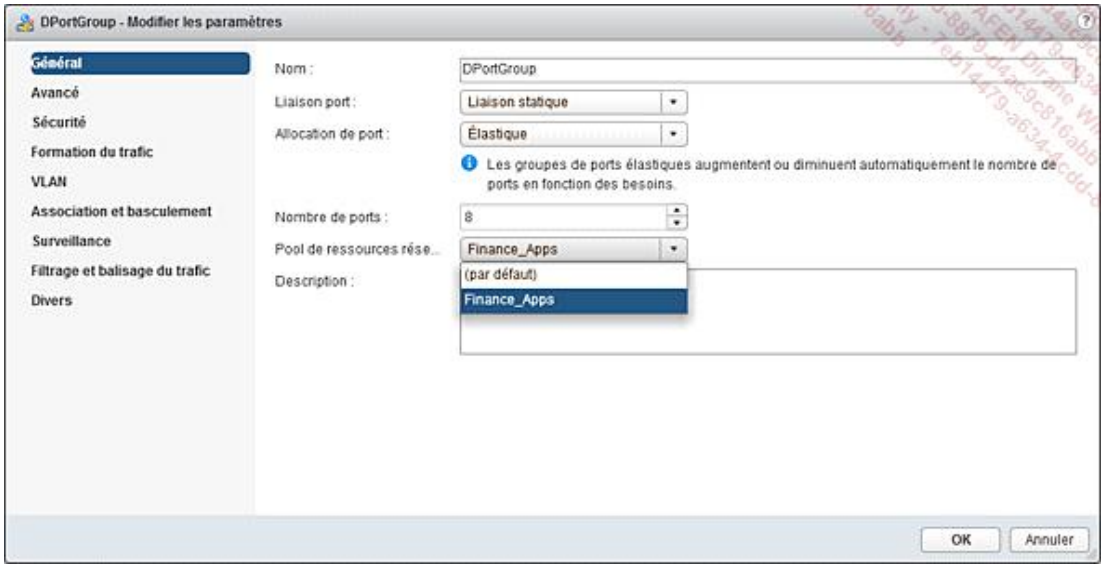
La validation permet de constater que la réservation a été mise en place.



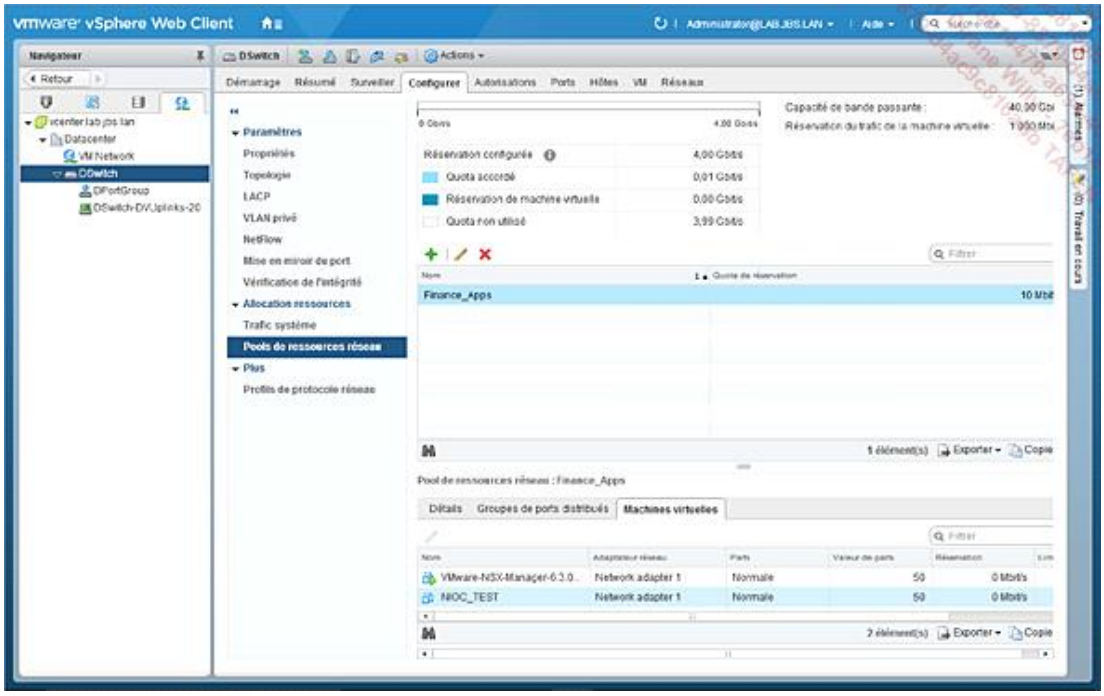
Il reste désormais à associer un pool de ressources à un groupe de ports, ici nous sélectionnerons celui de base. Modifions les paramètres de ce groupe de ports à l'aide d'un clic droit sur l'élément.



Sur l'onglet **Général**, nous pouvons définir à quel pool de ressources réseau appartiendra le groupe de ports. Ici, nous sélectionnons le pool que nous avons créé.



Enfin, nous pouvons visualiser sur l'écran des pools de ressources les machines virtuelles liées à notre profil.

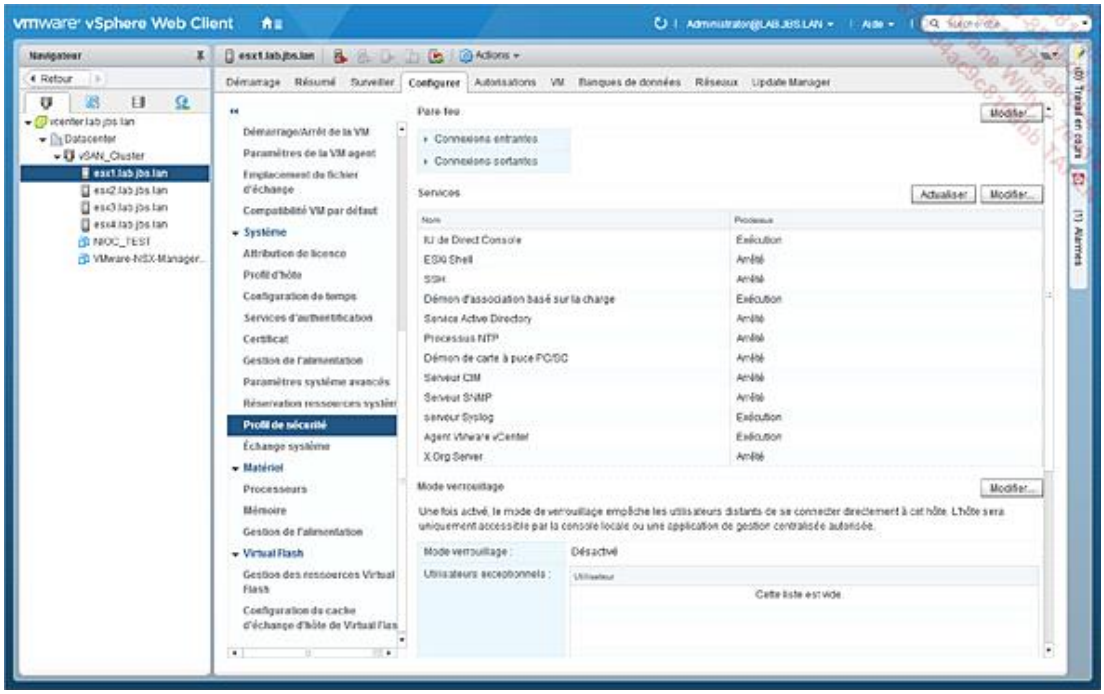


Notez qu'au-delà des éléments évoqués sur la réservation de bande passante et plus particulièrement concernant les parts attribuées, le mode de calcul reste similaire à ce que vous avez vu dans la partie pools de ressources processeur / mémoire.

c. Blocage de protocoles/ports

L'hyperviseur vSphere permet, via son pare-feu intégré, un filtrage efficace des connexions sortantes et entrantes du trafic de gestion. Ce pare-feu se positionne entre l'interface de gestion de l'hyperviseur et le réseau.

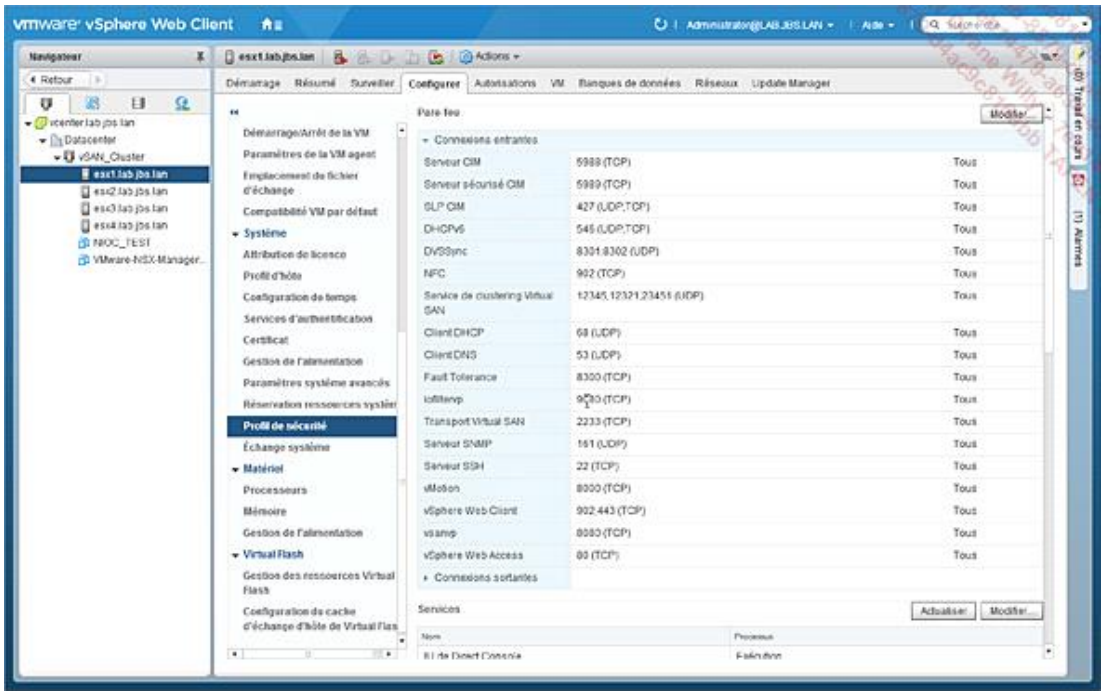
Il est possible de changer ces paramètres à l'aide de la section **Profil de sécurité**.

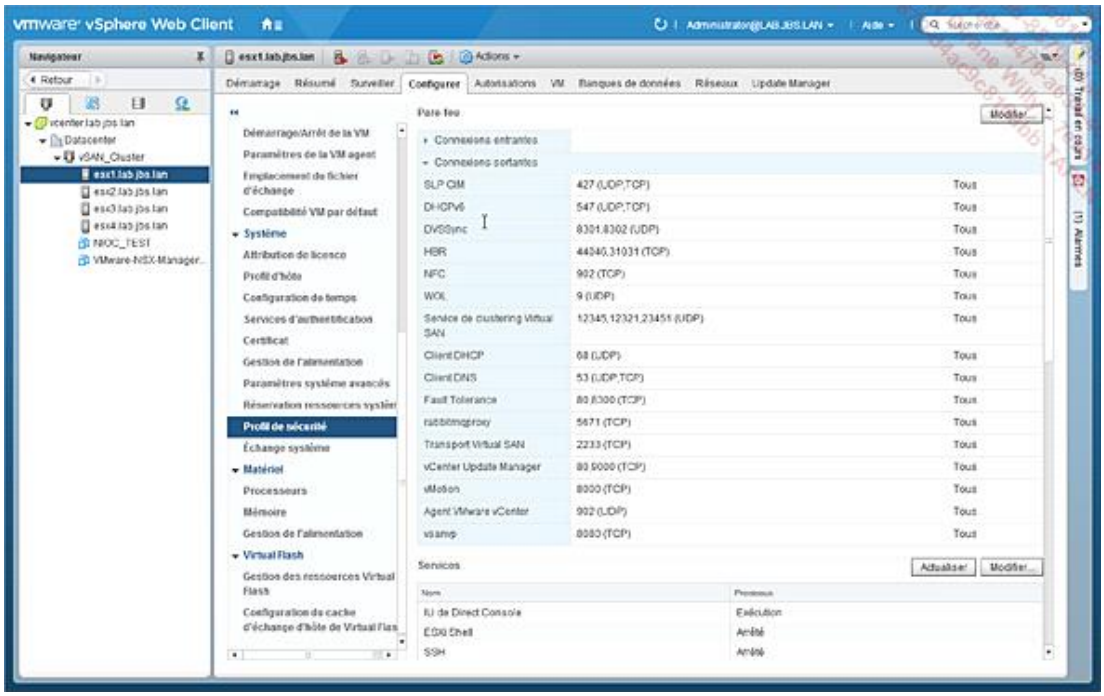


Lorsque vous dépliez les connexions entrantes et sortantes, l'ensemble des connexions possibles ainsi que le

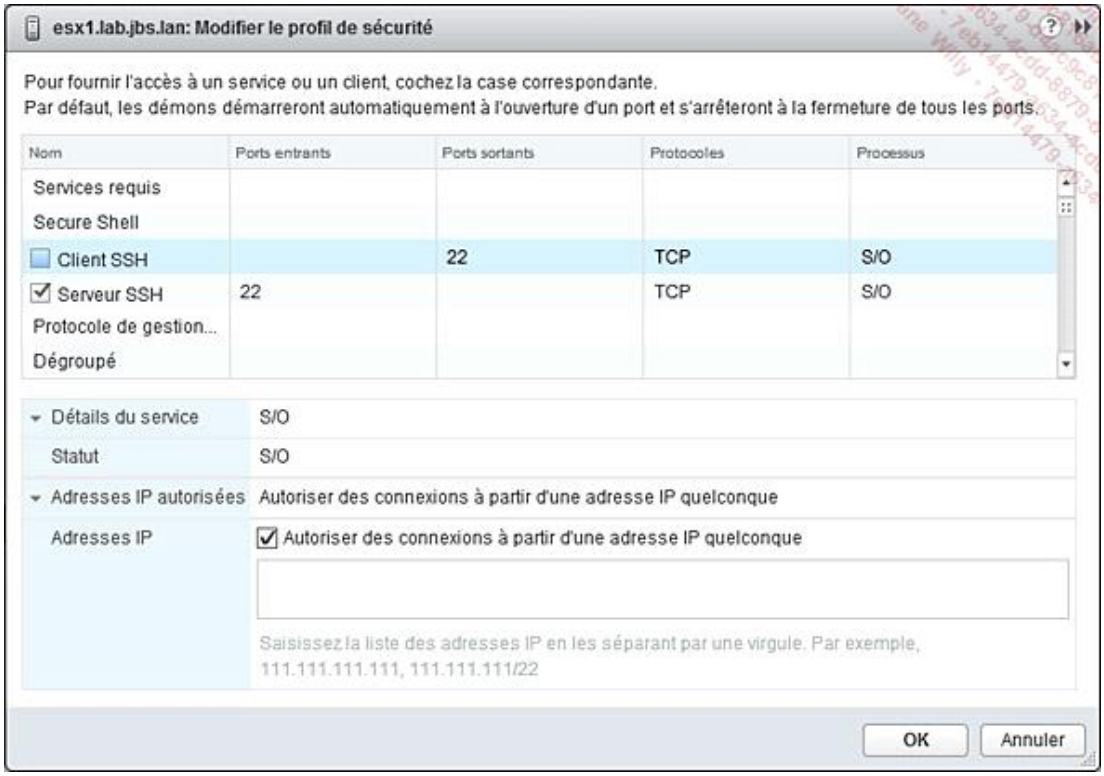
service lié s'affichent.

Il est utile (en complément de la documentation officielle) de prendre note de ces informations pour ouvrir les flux nécessaires au fonctionnement nominal de votre hyperviseur. Les captures suivantes montrent les connexions typiques entrantes et sortantes d'un hyperviseur vSphere 6.5 :





Pour modifier la configuration des connexions entrantes et sortantes, il suffit de cliquer sur le bouton **Modifier**. Vous obtiendrez un écran comme celui ci-dessous :



Il suffit alors à l'administrateur de cocher les cases dont il veut permettre les flux entrants ou sortants ou de les décocher pour les interdire. L'administrateur peut également spécifier dans cet écran les adresses IP permises pour le flux concerné.