

À LA CARTE, OU LA VIRTUALISATION AVEC VMWARE ESX



LAURENT.KLING@epfl.ch, FACULTÉ STI

LA PRODUCTION POUR L'INFORMATIQUE

L'informatique vue par les administrateurs système consiste à gérer des triplets: **matériel, logiciels et usagers**. Si les utilisateurs finaux sont l'objectif même de notre travail, les services offerts nécessitent la gestion de serveur basé sur:

- un matériel
- un système et ses pilotes
- les logiciels.

L'évolution des besoins évoluant avec le temps, le parc matériel installé n'est pratiquement jamais homogène.

L'administrateur doit gérer les relations de compatibilité suivantes pour chacun de ces serveurs:

- matériel - version de microprogrammation (*firmware*),
- version de microprogrammation - pilotes,
- pilotes - système d'exploitation,
- système d'exploitation - applications,
- applications - services offerts à l'utilisateur.

Parfois, un logiciel particulier nécessite une version de pilote spécifique entraînant la mise à jour de la microprogrammation du serveur.

S'ajoutant à ces contraintes élémentaires d'adéquation entre les composants, le dimensionnement d'un serveur est loin d'être une science exacte.

Dans le monde Microsoft Windows, la règle suivie est:

- Un service = un serveur.
- Un serveur = une machine capable de répondre à 100 % des demandes.
- 100 % des demandes = suffisamment de puissance de calcul, de mémoire vive et d'espace de stockage.
- Un serveur = une machine spécialisée pour la fiabilité.
- la fiabilité =
 - ▷ deux alimentations, dont une sur un circuit secours (UPS);
 - ▷ des disques redondants physiquement, RAID 1 - RAID 5;
 - ▷ un suivi de l'ensemble des paramètres physiques: température, tension d'alimentation, vitesse des ventilateurs;
 - ▷ mémoire avec correction/détection d'erreurs (ECC);
 - ▷ une sauvegarde assurée sur un support amovible indépendant;
 - ▷ un environnement climatisé;
 - ▷ un suivi des modifications apportées au système;
 - ▷ une permanence de gestion assurée par deux administrateurs;

Le respect de ces contraintes entraîne un coût important, donc une limitation du nombre de serveurs et comme conséquence perverse, un serveur trop bien dimensionné.

Parfois, l'équipement choisi ne respecte pas les contraintes de fiabilité décrites ci-dessus, au choix:

- la corruption de données entraînée par une alimentation électrique défectueuse;
- la confiance excessive dans un RAID 5 dont 2 des disques sont défectueux, la machine et le système d'exploitation ayant signalé le défaut d'un disque depuis 1 mois sans intervention humaine;
- la combinaison des deux épisodes ci-dessus, mais dont la sauvegarde sur bande ne fonctionne pas, car la cassette initiale a été recyclée.

Naturellement, l'arrivée d'un tel événement fâcheux entraîne par la suite le respect des règles de fiabilité énoncées ci-dessus.

L'idée séduisante d'externaliser ces services n'est pas une solution, car l'entreprise prestataire ne remplit peut-être pas complètement ces conditions!

LA VIRTUALISATION

Sous ce néologisme se cache la méthode consistant pour VMware à:

- une virtualisation du matériel à travers des composants standardisés;
- une machine correspondante aux
 - fichier(s) disque(s)
 - fichier de configuration;
- un noyau de virtualisation;
- un ensemble de commutateurs de réseaux virtuel.

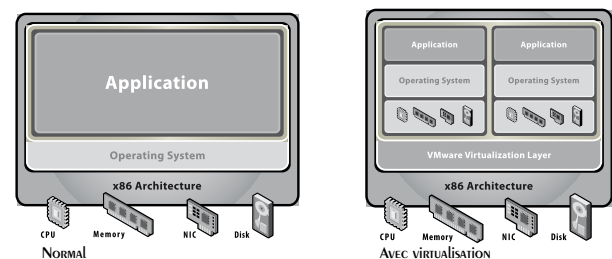


fig. 1 – SERVEUR VS. VIRTUALISATION

Cet environnement existe sous différentes formes, intégrant tout ou partie, la méthode de virtualisation VMware:

- Player: gratuit, uniquement pour lire une machine virtuelle.
- Workstation: avec licence, logiciel sur un système d'exploitation Windows ou Linux, intègre une configuration préétablie de 3 commutateurs virtuels: isolé, NAT, partage de la connexion de l'hôte.

- Serveur: gratuit, similaire à Workstation orienté vers la production, mais sans optimisation de la gestion de la mémoire et des processus d'ESX.
- GSX: un rare exemple d'un produit payant remplacé par un produit gratuit, Server le remplace !
- ESX: avec licence, le sujet de cet article.

Pour éviter la gestion d'un parc matériel important, j'utilise depuis plus de 7 ans des machines virtuelles pour simuler des développements systèmes.

En particulier, quatre projets sont basés sur un environnement virtuel:

- gestion Active Directory et synchronisation,
- déploiement d'installations avec RIS,
- packaging d'applications,
- (in)sécurité avec NTLM.

Pour réaliser ces projets, l'utilisation de VMware Workstation avec un serveur virtuel Windows 2003 avec les services:

- DNS
- DHCP
- RIS

et parfois les services ou applications suivantes

- WINS
- GHOST
- REMBO

L'espace mémoire disponible sur mon poste de travail ne me permet d'utiliser uniquement que deux, parfois trois machines virtuelles. Ceci explique l'apparente contradiction avec la règle d'allouer un serveur par service.

La séparation avec le mode réel par l'utilisation d'un commutateur réseau virtuel permet de disposer d'un clone de l'ensemble du domaine Active Directory STI. Les copies de cet environnement dans différentes périodes de son existence permettent de simuler l'évolution de la réalité. Cette ressemblance entraîne parfois la situation comique de confondre deux accès à distance de serveur réel et virtuel, et de réaliser en virtuel une demande d'un usager réel !

Un des avantages de l'utilisation de machines virtuelles est la facilité de gestion des différentes configurations en pointant sur des fichiers, et la richesse quasi infinie des différents modes d'utilisation du disque.

VMWARE ESX

L'application de la règle d'un serveur par service entraîne une augmentation constante des machines à gérer. La complexité de leur gestion et le temps consacré augmentant parallèlement, une rationalisation s'impose.

L'utilisation d'un serveur ESX représente un nouveau paradigme, car la gestion du couple serveur - service devient virtualisée.

Pour un serveur, les avantages sont multiples:

- indépendance du matériel,
- indépendance de la configuration,
- indépendance du stockage,
- analyse détaillée du fonctionnement des serveurs.

L'INSTALLATION

Pour s'assurer que la base matérielle correspond aux attentes d'un serveur de virtualisation, seulement une gamme restreinte de serveurs est certifiée pour ESX.

Bien que le système d'exploitation du serveur ESX soit basé sur BSD et RedHat, cette restriction s'explique aisément par les besoins physiques:

un minimum réaliste

- serveur certifié
- 2 noyaux de calcul
- 4 GB de mémoire vive
- 73 GB de disque sécurisé, local ou au travers d'un SAN
- 2 cartes réseau Gigabit.

la configuration que j'utilise

- Serveur certifié: DELL 2800
- 2 processeurs bi-cœur = 4 noyaux de calcul (8 avec Hyper-Threading activé)
- 8 GB de mémoire vive
- 300 GB Raid 1, 1.2 To Raid 5
- 2 cartes réseau Gigabit.

au maximum

- serveur certifié
- 16 noyaux de calcul
- 64 GB de mémoire
- au maximum 9 TB de disque virtuel
- 16 cartes réseau Gigabit

La configuration utilisée peut paraître importante, mais la virtualisation avec un serveur ESX a pour objectif de remplacer une série de serveurs existants.

Dans mon cas, j'utilise en phase de développement la licence incluse avec le *VMware Technology Network*. ESX offre l'utilisation de processeur multicœur avec la base d'une licence par processeur. L'activation de l'option SMP ouvre la possibilité d'utiliser deux processeurs pour une machine virtuelle.

L'utilisation d'une base matérielle réduite permet une installation rapide. La seule précaution est de ne pas utiliser la configuration automatique des partitions du disque de démarrage. Par rapport à une installation de la même machine avec Windows 2003, le gain est évident (pas de touche F8 au démarrage!).

UN NOUVEL ENVIRONNEMENT DE TRAVAIL

Un serveur orienté pour la production

Le serveur ESX n'est pas un poste de travail, en conséquence son interface sécurisée avec l'extérieur se limite à une console Web et un accès en mode ligne de commande.

À ce titre, une connexion Gigabit du serveur est obligatoirement réquisitionnée par ESX. Cette contrainte n'est pas exagérée, car l'ensemble du trafic de contrôle, de transfert d'image virtuelle et de sauvegarde passe par cette connexion. Un débit élevé est nécessaire, car il est particulièrement désagréable de perdre le contrôle des machines, virtuelles.

Pour accéder à distance aux machines virtuelles comme

avec un Keyboard Vidéo Monitor, KVM, il est possible d'installer un client sur un poste de contrôle distant Windows ou Linux.

La deuxième connexion Gigabit est utilisée pour les serveurs qui accèdent au monde extérieur.

De la même manière, il est déconseillé d'être avare pour la mémoire allouée pour la console. En pratique, l'espace maximum proposé de 800 Mo est une valeur adéquate.

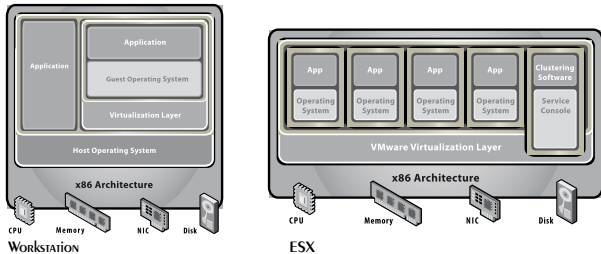


fig. 2 – ARCHITECTURE DE VIRTUALISATION

Contrairement à une idée largement répandue, un serveur ESX n'est pas un serveur Linux/Unix optimisé pour les machines virtuelles, c'est un serveur dont le cœur est le noyau de virtualisation qui exécute une console RedHat utilisée comme interface. Dans la version actuelle, ESX 2.53, la console et le noyau sont indissociables. Dans la prochaine version, ESX 3.0, la séparation entre console et noyau sera réalisée.

Si l'installation est particulièrement rapide, il est nécessaire de respecter les conditions de fonctionnement d'un serveur:

- dans un premier temps, il faut synchroniser son temps avec l'extérieur pour pouvoir s'incorporer aux ressources modernes, comme Kerberos;
- ensuite, l'installation d'un client d'un logiciel de sauvegarde permet d'assurer l'existence des données du serveur sur la durée.

Ces deux opérations s'effectuent dans la console du serveur ESX, cette console est basée sur RedHat 7.4 et permet d'utiliser les outils et logiciels standards du monde Linux. Une certaine limitation des ajouts extérieurs doit être observée pour éviter la perturbation du serveur de virtualisation. Si la console est utilisée de manière intensive, par exemple pour une sauvegarde, elle entraîne l'utilisation du serveur!

VMFS

Sur un poste de travail avec VMware Workstation, un dossier contient le fichier de configuration et l'ensemble des fichiers disques.

Sur ESX une séparation existe, car le système de fichiers disques est spécifique: VMFS

VMFS possède la propriété de nous permettre de remonter le temps:

- pas de hiérarchie
- 255 fichiers au maximum

Ces limites peuvent paraître réductrices, en pratique, on conserve uniquement des disques, un fichier = un disque, et cela est largement suffisant. Par contre, établir des règles de désignation des fichiers-disques devient nécessaire!

Une propriété de VMFS est d'être partagé par défaut. Cette propriété devient particulièrement intéressante quand deux serveurs ESX accèdent au même LUN d'un SAN.

Deux espaces VMFS peuvent être joints. Cela est dé-

conseillé, car la perte d'un des espaces VMFS entraînerait la perte de l'ensemble, propriété que les aficionados du RAID 0 connaissent.

Commutateur réseaux virtuel ou Virtual Switch

Cette fonctionnalité est une des clés du succès de VMware par rapport à la concurrence. En VMware Workstation, trois commutateurs sont configurés par défaut:

VMnet0 : un lien partagé avec la carte réseau du poste client

VMnet1 : un réseau isolé avec DHCP, mais qui peut accéder à la ressource du poste client

VMnet8 : un réseau isolé avec une translation d'adresse NAT avec l'hôte.

Avec ESX, au départ, uniquement un commutateur réseau virtuel, équivalent au monde extérieur existe. Pour recréer un environnement similaire à celui d'un VMware Workstation, il faut le recréer. Quatre commutateurs sont utilisés:

Réal: connecter à la 2e carte réseau du serveur, l'accès au monde extérieur

Local: réseau isolé avec serveur DHCP

et deux réseaux isolés avec une translation d'adresse NAT

Nat

VMnet1

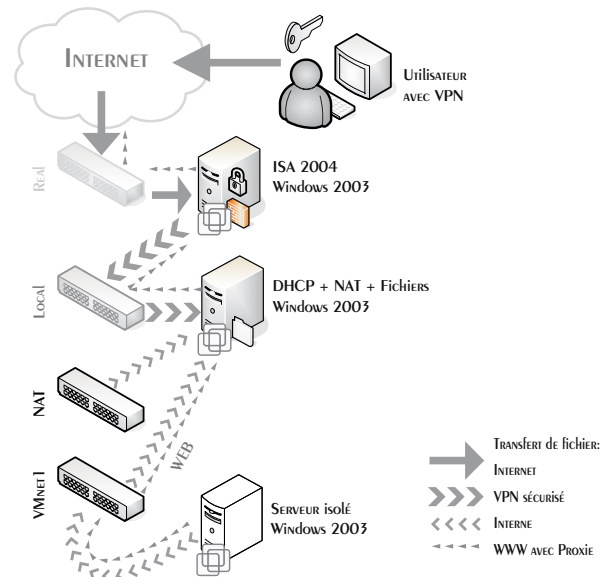


fig. 3 – RÉSEAU VIRTUEL

La création des commutateurs virtuels ne suffit pas à recréer l'environnement d'un poste de travail avec VMware Workstation. Il est nécessaire de réaliser les connexions entre chaque commutateur virtuel.

Par souci d'homogénéisation, 2 serveurs virtuels Windows 2003 réalisent ces fonctions:

1. Un pare-feu avec ISA 2004: l'isolation réalisée par ce serveur cloisonne et filtre les communications entre le monde réel et les réseaux internes. Il offre également les fonctions de serveur Proxy Web et de serveur VPN
2. Un serveur NAT et DHCP: pour permettre une gestion indépendante des réseaux internes, il réalise une translation d'adresse NAT entre lui-même et les réseaux isolés. Le serveur DHCP est utilisé par le pare-feu pour les

connexions VPN.

Pour limiter le nombre de serveurs de service, le serveur NAT abrite également un espace de stockage accessible depuis trois sources:

- le monde extérieur avec l'accès VPN
- les ordinateurs dans le réseau local
- les ordinateurs des réseaux isolés

Ainsi, l'isolation des réseaux internes est assurée tout en permettant un accès au Web par l'intermédiaire du serveur Proxie Web contenu dans ISA 2004.

Un respect strict des contraintes de pilotes et de l'outil VMware

Le serveur ESX est calibré pour de la production, les contraintes pour les machines virtuelles sont:

- uniquement des disques SCSI (l'IDE n'est pas supporté pour les disques)
- uniquement le bon contrôleur SCSI
- de préférence, la carte réseau adaptée, autrement, on bénéficie d'un débit de 10 Mbit/s
- un fichier pour le disque entièrement alloué

L'absence de respect d'une de ces contraintes entraîne soit des performances limitées, soit l'impossibilité de démarrer la machine virtuelle.

L'installation de l'outil VMWare n'est pas une option avec VMX, c'est une obligation! En particulier, lui seul permet de bénéficier de la technologie de l'allocation-mémoire par ballon, notion qui sera expliquée dans la suite de cet article. Il permet également de synchroniser l'horloge de la machine virtuelle avec le serveur, à condition que celui-ci soit lui-même synchronisé.

UTILISATION

Après l'installation et la configuration du serveur, on peut mettre en place les différents serveurs assumant l'infrastructure. Naturellement, l'idée de copier le serveur virtuel utilisé sur son poste de travail avec VMware Workstation émerge. Ce serveur virtuel a déjà subi deux changements de version et une copie sur un autre disque pour pouvoir être mis à jour en Windows 2003.

Le premier transfert sur le serveur ESX a échoué. Il faut respecter les contraintes:

- **Mise à jour du pilotes SCSI**
Heureusement, la machine utilisait des disques SCSI, par contre le contrôleur utilisé était obsolète, incompatible avec ESX. L'ajout temporaire d'un 2e contrôleur SCSI, puis la configuration du pilote ont résolu ce problème.
- **Mise à jour du pilote réseau**
Dans ce cas, la mise à jour est simple, car le changement de la configuration matériel n'est pas problématique avec le démarrage de Windows 2003.
- **Utilisation de disque préalloué**

Pour économiser de la place sur le poste de travail, les disques des machines sont configurés pour ne pas être préalloués, en conséquence la taille du fichier du disque virtuel correspond à l'utilisation. Cette méthode ne pose pas de problème, la conversion en VMFS sur ESX va engendrer la création d'un fichier égal à la capacité

maximum, entraînant parfois quelques surprises !

■ Vérifier le niveau de développement des produits

Deux générations successives de produits cohabitent chez VMware:

- ▶ Noyau 2.0: ESX 2.53 et Workstation 4.0
- ▶ Noyau 3.0: Workstation 5.5.

En conséquence, le format de fichier de Workstation 5.5 est incompatible avec ESX 2.53. Le résultat, mon image n'est plus compatible avec ESX!

Pour éviter de recréer le serveur sur ESX, la solution apparaît avec la version bêta d'un utilitaire de VMware, Virtual Machine Importer 2.0, qui est capable de rétrograder dans une génération antérieure une machine virtuelle. L'utilitaire permet également de convertir des images provenant de nombreuses sources:

- ▶ Microsoft Virtual Server & Virtual PC
- ▶ Symantec Ghost 9 & LiveState

Ainsi corrigé, ce serveur virtuel a facilement été transmis par SFTP sur le serveur ESX, puis converti en format VMFS. Son utilisation de RIS a permis de créer l'ensemble des serveurs et poste de travail utilisé sur ce serveur ESX.

Contrairement à l'idée véhiculée sur Internet, il est facile d'intégrer les pilotes natifs de VMware permettant la création d'une installation Windows 2003 ou XP entièrement sans intervention. L'outil VMware étant intégré dans la phase de postinstallation de Windows. Les lecteurs intéressés peuvent se reporter à l'article que j'ai écrit sur ce sujet, http://ditwww.epfl.ch/SIC/SA/SPIP/Publications/article.php?id_article=155.

LA MISE EN PRODUCTION ET L'OPTIMISATION

La première migration concerne le serveur responsable de la mise à jour de la sécurité pour Windows, SUS ou Software Update Services. La machine actuelle ne correspondant pas aux caractéristiques requises par le remplaçant de SUS, WSUS ou Windows Server Update Services.

WSUS en quelques mots



fig. 4 – WUS SUR UNE CONSOLE DE CONTRÔLE

Le principal avantage de WSUS est de concerner l'ensemble des produits de Microsoft, pas uniquement les rustines des composants systèmes. L'autre élément marquant est sa capacité de fournir un état détaillé des mises à jour appliquées

à un parc de machine ou d'analyser l'état d'un ordinateur particulier. Sur le plan du réseau, après avoir téléchargé 30 Go en un jour depuis Microsoft pour réaliser la copie des mises à jour, la mise à jour des 900 PC client ne nécessite plus de téléchargement depuis Microsoft!

La capacité de trier les ordinateurs permet d'isoler les ordinateurs posant problème. Par exemple, pour une rustine de sécurité les ordinateurs qui n'ont pas pu être mis à jour, représentent un signe évident de dysfonctionnement.

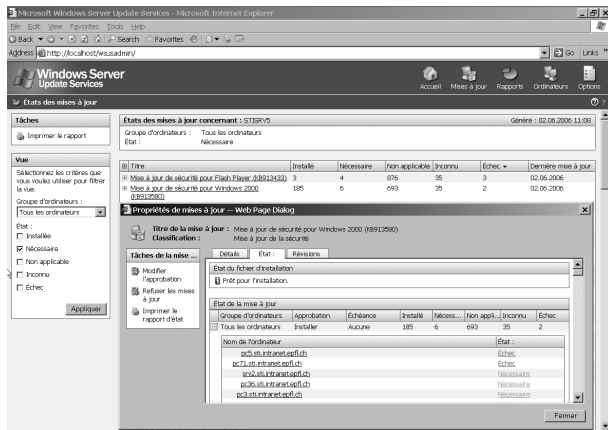


fig. 5 – UN SIGNE DE DYSFONCTIONNEMENT

WSUS présente un avantage évident par rapport aux produits commerciaux, il est gratuit !

P2V

Le transfert d'un serveur physique vers un serveur virtuel entraîne la création d'une machine virtuelle. L'utilisation de RIS intégré dans le serveur virtuel interne facilite cette opération. En particulier par l'adaptation automatique des pilotes utilisés pour la configuration choisie. Quand il s'agit de migrer plusieurs dizaines de serveurs, il devient impossible de recréer chacun des serveurs individuellement. Particulièrement quand les services offerts sont interdépendants et que la documentation des modifications n'est pas forcément à jour.

VMware P2V permet de résoudre cette problématique, de convertir un système avec des disques IDE, des RAID matériels en image virtuelle avec des disques SCSI et utilisant les composants virtuels standard. Malheureusement, cette démarche est limitée par les matériels supportés. Uniquement des modèles récents de serveurs et quelques postes de travail sont inclus dans la liste de compatibilité: http://www.vmware.com/pdf/p2v_hardware.pdf.

Le matériel du serveur SUS à migrer n'étant pas inclus, installer un nouveau serveur est indispensable.

CRÉATION D'UNE MACHINE POUR VMWARE ESX

La démarche de création d'un serveur est identique à celle utilisée dans le monde réel. Quatre éléments nécessitent des commentaires:

Un démarrage accéléré

Le respect des règles d'un serveur fiable entraîne une multiplication de périphériques intelligents. Dans le cas du

serveur de Virtualisation, le déroulement des processus de découverte et d'auto configuration prend un certain temps, plusieurs minutes. La seule solution pour diminuer ce délai est de supprimer des tests ou leur affichage. Cette suppression est vivement déconseillée, car c'est justement ces affichages qui permettent de diagnostiquer les problèmes au démarrage, en particulier pour les contrôleurs de disque dur! Cette phase de vérification du matériel n'existe pas pour les serveurs virtuels, c'est un gain de temps appréciable.

Des composants matériels standardisés

La couche d'abstraction matérielle présente avec VMware permet d'éviter de s'occuper du matériel présent dans le serveur. Ainsi, la machine virtuelle devient portable, indépendante du serveur hôte. Elle peut s'exécuter dans une grande variété d'environnement matériel.

La virtualisation du contenu des disques

Sur un serveur réel, deux activités sont coûteuses:

- Modifier la répartition des disques, en particulier quand Partition Magic Workstation ne peut gérer un serveur utilisant un RAID physique.
- La récupération de l'état précédent d'un disque avec une sauvegarde, en particulier si la sauvegarde n'existe pas.

Dans le monde virtuel, les solutions sont simples et efficaces:

- Un fichier VMFS peut être agrandi ou réduit en taille, l'utilisation de plusieurs partitions pour utiliser l'espace physique n'est plus nécessaire.
- La sauvegarde peut être remplacée par la propriété du disque en *Undoable*, à l'extinction de la machine virtuelle VMware nous demandera si on désire conserver ce fichier de modification, le supprimer ou l'appliquer sur le disque.

La virtualisation des réseaux

Le déplacement d'un espace de test à l'espace de production demande uniquement le choix du réseau virtuel. Au redémarrage, le serveur de test devient un serveur de production sans manipulation physique.

Définir une règle de désignation cohérente

La capacité de modifier l'appartenance d'un serveur à sa connexion réseau peut vite aboutir à une mélasse virtuelle.

La règle suivante est appliquée pour les noms des fichiers de configurations et disques:

	fichiers de configuration	disques
monde virtuel	reseauinterne-os-role	reseauinterne-os-role-partition
monde réel	nomdns-os-role	nomdns-os-role-partition

Si un serveur réel doit être créé dans un environnement isolé, son nom correspond à son état final. Cette règle évite de nombreuses collisions entre monde réel et monde virtuel.

Afficher les statistiques des machines

Sur un serveur réel, la mesure des performances est un processus qui doit être réalisé dans le système d'exploitation.

4194304512	Jun 14 12:54	local-wxp.vmdk
6291456512	Jun 14 13:02	stisrv5-w2k3srv-wus-bootC.vmdk
10737418291	Jun 14 13:02	stisrv5-w2k3srv-wus-patchE.vmdk
8388608512	Jun 14 13:02	stisrv9-w2k3srv-sql2005-bootC.vmdk
8388608512	May 22 18:56	stisrv9-w2k3srv-sql2005-dbE.vmdk
4194304512	May 12 10:08	vmnet1-w2k3srv-tstdriver.vmdk
8589935104	Jun 14 12:55	vmnet1-w2k3srv-risghost-bootC.vmdk
8589935104	May 18 14:31	vmnet1-w2k3srv-risghost-ghostH.vmdk
1073742336	May 18 14:31	vmnet1-w2k3srv-risghost-netE.vmdk
8589935104	May 18 14:31	vmnet1-w2k3srv-risghost-remboG.vmdk
8589935104	Jun 14 12:55	vmnet1-w2k3srv-risghost-risF.vmdk
8589935104	May 18 14:31	vmnet1-w2k3srv-risghost-tmpJ.vmdk
6291456512	Jun 14 12:55	vmnet1-w2k3srv-wus-bootC.vmdk
31457280512	Jun 14 12:55	vmnet1-w2k3srv-wus-patchE.vmdk
4194304512	Apr 28 15:05	vmnet1-wxp-pxe.vmdk
4194304512	May 17 17:36	vmnet1-wxp-test_tools.vmdk

fichiers VMFS AVEC LA RÈGLE APPLIQUÉE

Sur un serveur ESX, la mesure est extérieure à la machine virtuelle, elle est commune pour l'ensemble des machines. Ainsi, la mesure des performances et leur optimisation deviennent la norme plutôt que l'exception.

L'utilisation d'Apache sur la console comme serveur Web rend possible la disponibilité d'outil complémentaire de rapport. Mis gratuitement à la disposition de la communauté, **Vmktree**, de Lars Trøen est particulièrement précieux: <http://larstr.tihlde.org/vmktree/>.

Avec cet outil, il est possible d'afficher l'état du système et l'état de chaque machine virtuelle, mais également de zoomer sur une portion significative des données. La capacité de mettre en relations ces données sur différentes bases temporelles est utile.

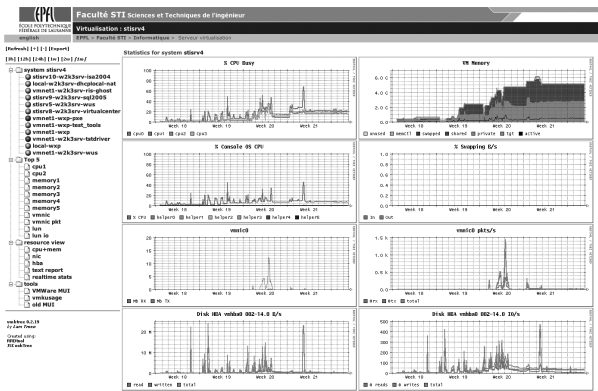


fig. 7 – utilisation du serveur sur un mois

OPTIMISATION

Dans le monde réel, l'analyse de performance d'un serveur n'est entreprise que si celui-ci ne donne pas entière satisfaction. Si une machine fonctionne correctement, il paraît normal de la laisser tranquille.

Avec ESX, cette exception devient la norme, mais pour optimiser, il est nécessaire de comprendre les mécanismes sous-jacents au serveur.

La lecture d'un article de Carl Waldspurger, qui a obtenu la distinction du meilleur papier à la conférence Operating Systems Design and Implementation en 2002: **Memory Resource Management in VMware ESX Server**, <http://www.waldspurger.org/carl/research.html>.

Cet article permet de comprendre les mécanismes internes d'un serveur ESX, deux d'entre eux méritent de s'y

attarder:

La pagination de la mémoire partagée

Avec VMware Workstation, la mémoire allouée est réservée dans la mémoire disponible du poste de travail, limitant le nombre de machines virtuelles. Dans un serveur ESX, le mécanisme utilisé est plus subtil. Par défaut, le serveur va réserver 50% de l'espace mémoire alloué à un espace commun et 50% à un espace réservé.

Pour déterminer la mémoire réellement utilisée par l'espace commun, en bleu foncé sur les graphiques, ESX calcule un code de hachage sur les différentes zones-mémoire occupées dans la machine virtuelle, si un code de hachage est identique entre deux zones-mémoire, avec le contrôle d'un checksum, le bloc sera considéré

comme partagé et paginé dans la mémoire commune du serveur ESX.

Cette capacité de *récupérer* la mémoire commune explique certainement les bonnes performances de ESX.

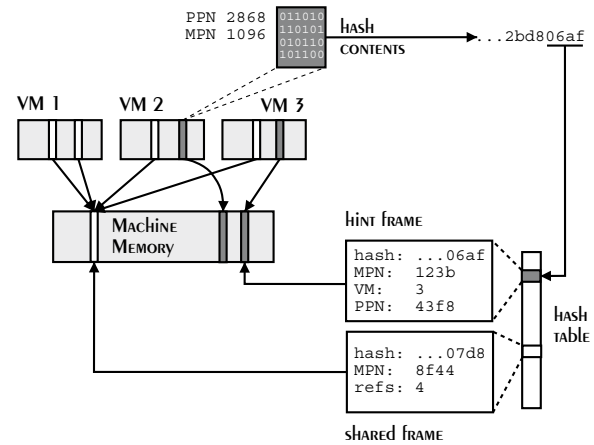


fig. 8 – utilisation d'un code de hachage

L'utilisation des ballons de baudruche

L'utilisation de la mémoire partagée peut entraîner que l'ensemble de la mémoire vive du serveur ESX soit occupé. Que se passe-t-il si un serveur virtuel augmente sa consommation?

Sur une machine normale, le processus de gestion de la mémoire virtuelle va *swapper* le contenu le moins actif sur le disque dur. Les adeptes du canotage parleront de serveur qui rame.

Pour éviter ce processus catastrophique pour les performances, et particulièrement quand on ne gère pas les processus internes à chaque machine virtuelle, on doit trouver un mécanisme capable de répartir la charge sur l'ensemble des machines virtuelles.

L'installation de l'outil VMware résout ce problème. Encapsulé dans celui-ci, il existe un processus quasi invisible.

Si le cas du dépassement de la capacité-mémoire réelle apparaît, ce processus va commencer à gonfler sa consommation de mémoire dans les autres machines virtuelles. Celles-ci vont naturellement commencer à utiliser le mécanisme du système d'exploitation interne de swap; paginer les processus

moins prioritaires. L'espace mémoire libéré va être utilisé pour le serveur gourmand. Quand la charge de travail est passée, la situation va redevenir normale et les ballons vont se dégonfler.

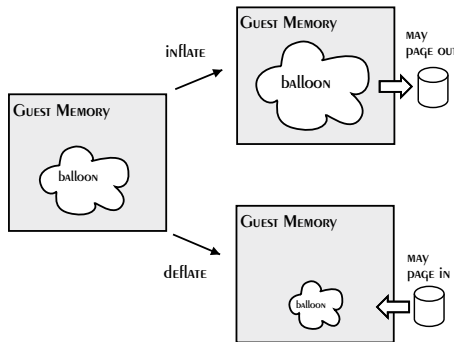


fig. 9 – LES BALLONS EN ACTION

En pratique, il est recommandé de ne jamais utiliser ce mécanisme, et il est souvent plus simple d'arrêter les serveurs de test qui peuvent résider dans le serveur ESX.

Optimiser la mémoire - processeur - réseau - espace disque.

Comme déjà décrit, on n'optimise pas un serveur réel qui fonctionne correctement. Le prédimensionnement d'un serveur est généralement fourni par le constructeur du logiciel, on multiplie les valeurs par un facteur de sécurité, et généralement, le résultat est bon. Avec un serveur de virtualisation, on peut mesurer, et surtout agir sur l'ensemble des paramètres physiques du serveur virtuel:

- la mémoire
- le nombre de processeurs, leur réservation et leurs priorités
- la ressource réseau
- l'espace disque utilisé.

Dans le cas du serveur WSUS, la recommandation est d'allouer 1 Go de mémoire, un processeur > 1 GHz et 4 Go de disque système et 8 Go de disque pour les rustines. Il est également mentionné la possibilité d'utiliser une base de don-

VMware ESX Server 2.5.3 build-22981 | root@stisrv4

Status Monitor CPU Memory Disk Network Hardware Options Users and Events Refresh | Help | Close

Last updated Wed Jun 14 2006 12:34:01 GMT+0200 (CEST)

stisrv5-w2k3srv-wus
Powered on | PID 15833 | VMID 308

Performance		Resources	
Used Memory			
Total Allocated	640.0 M	Resource Settings	Edit...
Recently Used	97.3 M	Minimum	320 M
Shared	143.7 M	Maximum	640 M
Reclaimed Memory		Shares	6400
Written to Swap File	0		
Reclaimed by VMMemCtl Module	0		
Virtualization Overhead			
Total Overhead	60.8 M		

VMware ESX Server 2.5.3 build-22981 | root@stisrv4

Status Monitor CPU Memory Disk Network Hardware Options Users and Events Refresh | Help | Close

Last updated Wed Jun 14 2006 12:34:01 GMT+0200 (CEST)

stisrv5-w2k3srv-wus
Powered on | PID 15833 | VMID 308

Performance		Resources	
Read Bandwidth			
vmhba0:1:0	15.0 KBps	Number of Shares	Edit...
Write Bandwidth		vmhba0:1:0	1000
vmhba0:1:0	55.0 KBps		

VMware ESX Server 2.5.3 build-22981 | root@stisrv4

Status Monitor CPU Memory Disk Network Hardware Options Users and Events Refresh | Help | Close

Last updated Wed Jun 14 2006 12:31:51 GMT+0200 (CEST)

stisrv5-w2k3srv-wus
Powered on | PID 15833 | VMID 308

Performance		Resources	
Receive Bandwidth			
00:0c:29:23:1a:57	16.0 Kbps	Traffic Shaping	Edit...
Transmit Bandwidth		Enable Traffic Shaping	No
00:0c:29:23:1a:57	440.0 Kbps		

fig. 10 – RÉGLAGE ET UTILISATION POUR LA MACHINE VIRTUELLE WSUS: MÉMOIRE – DISQUE – RÉSEAU

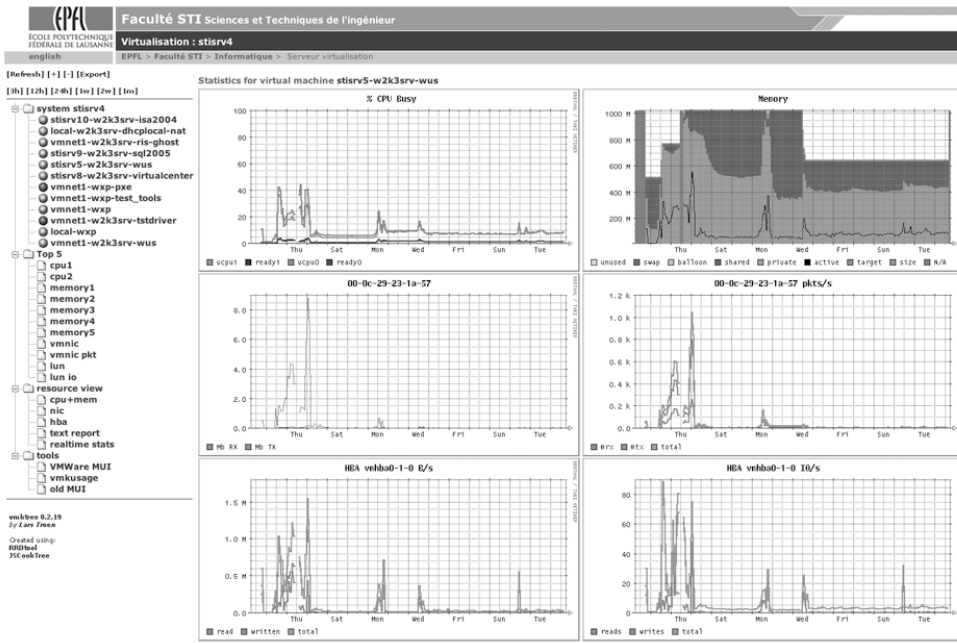


fig. 11 – WSUS vue du côté ESX

nées externe pour conserver les données de mises à jour.

La mise en place du serveur WSUS s'est déroulée en 2 époques:

- en environnement protégé pendant sa création.
- en environnement réel, dans sa phase de production.

En enfance, les étapes suivantes ont été réalisées:

- création du serveur avec RIS, 512 Mo de mémoire, 4 Go de disque système, 30 Go de disque pour les rustines.
- refus d'installation de WSUS,
- augmentation de la taille de la mémoire à 1024 Mo,
- installation de WSUS,
- utilisation d'une base de données externalisée,
- baisse des performances,
- ré-installation de WSUS (utilisation de la propriété Undoable du disque),
- augmentation de la capacité des disques, 8 Go système, 100 Go pour les rustines.

Dans le monde adulte:

- changement de la configuration réseau,
- analyse de l'utilisation-mémoire, décision de passer à 640 Mo,
- fonctionnement correct avec 900 usagers.

Si les modifications des propriétés physiques d'un serveur sont remarquables. Le dimensionnement des disques est fabuleux. Pour passer d'un disque de 30 Go à 100 Go, la commande est:

```
wmkfstools -X 102400m /vmfs/wrk/stisrv5-w2k3srv-wus-patchE.vmdk
```

La durée d'exécution de la commande est inférieure à dix secondes! Pour que le système d'exploitation puisse utiliser ainsi l'espace alloué, il reste à utiliser un utilitaire .

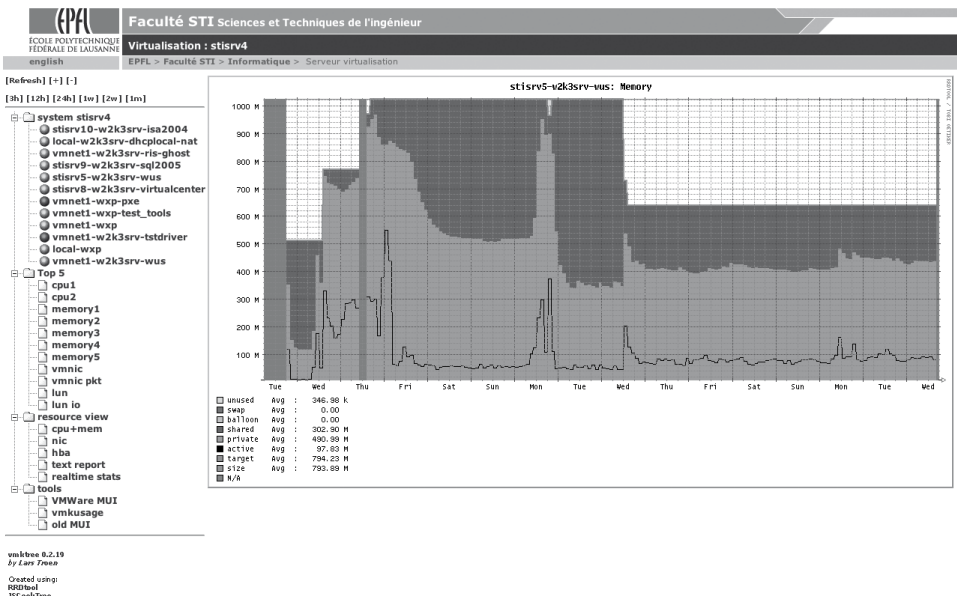


fig. 12 – Evolution de la MÉMOIRE de WSUS

Dans le cas des tests effectués aujourd'hui, la possibilité de définir des priorités entre les différents serveurs n'a pas encore été utilisée, de même pour la capacité de définir des limites pour l'utilisation du réseau et des accès disques.

VMOTION ET VIRTUAL CENTER

Le risque de la virtualisation est d'oublier la base physique. Si le serveur physique s'arrête, il entraîne l'arrêt de l'ensemble des serveurs virtuels hébergés.

Pour assurer une fiabilité suffisante en production, deux serveurs ESX sont nécessaires. La cerise sur le gâteau est la disponibilité de VMotion.

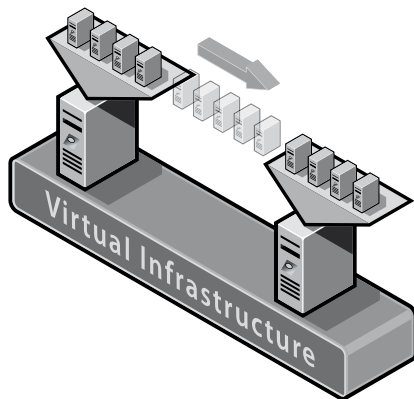


fig. 13 – PRINCIPLE DE VMOTION

Avec VMotion, le déplacement à chaud de serveur virtuel d'un serveur physique ESX sur un autre serveur physique ESX devient réalité.

On ne peut être qu'époustouffé par la vision d'un serveur virtuel qui migre d'un serveur ESX sur un autre serveur ESX tout en étant connecté sur le serveur qui se déplace. Cette magie s'explique certainement par le mécanisme de pagination de la mémoire d'un serveur ESX et la capacité de VMFS d'accepter des accès concurrents. Les personnes intéressées peuvent énumérer les brevets déposés autour de ces technologies.

VMotion représente un changement supplémentaire de paradigme, le serveur physique avec VMware ESX devient lui-même virtualisé !

Cet enthousiasme doit être compensé par l'environnement nécessaire pour VMotion:

- 2 serveurs certifiés avec une architecture de processeur identique (Intel ou AMD)
- Une connexion réseau Gigabit dédiée pour le réseau VMotion sur chaque serveur.
- Un espace disque sur un SAN avec une connexion des deux serveurs sur le même LUN.
- Le logiciel Virtual Center
- Et deux licences de VMotion!

Ne disposant pas de cet environnement, je n'ai pas pu encore tester cette fonctionnalité.

L'AVENIR

Le marché de la virtualisation est maintenant mature, la concurrence entre les entreprises devient féroce. Quatre solutions me semblent émerger du marché:

Xen

Cette solution open Source semble prometteuse, elle va être intégrée dans le noyau d'environnement Linux. Son inconvénient réside dans l'absence de virtualisation des interfaces d'entrée sortie.

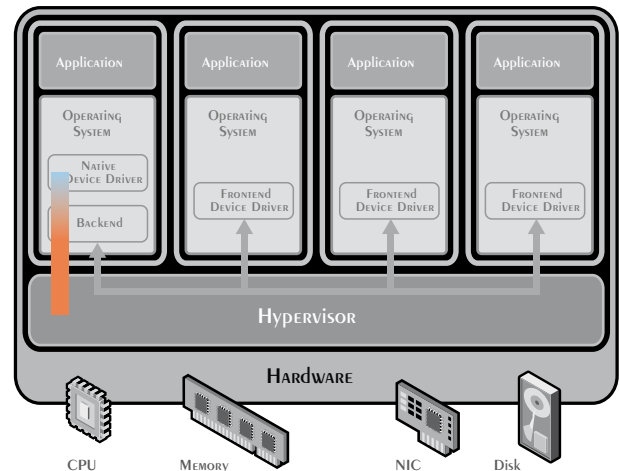


fig. 14 – XEN I/O

Microsoft Virtual Server & Virtual PC

Virtual PC provoque une certaine nostalgie, car mon premier travail sur Active Directory utilisait Virtual PC sur un Macintosh G4 !

Racheté et développé par Microsoft, le produit est mature, la disponibilité de Virtual Server 2005 en bêta gratuite est intéressante.

VMware

Inventeur de la virtualisation, qui est passée des laboratoires universitaires au monde commercial. Après avoir refusé un mariage avec Microsoft, VMware a été rachetée par EMC, mais reste indépendante.

Quatre développements modifient la situation:

- La mise à disposition gratuite de VMware Serveur, qui possède des caractéristiques avancées comme la virtualisation avec SMD, la possibilité de gérer des machines 64 bits a profondément modifié le débat.
- L'arrivée de VMplayer, également gratuit, rend possible le scénario d'application virtuelle portable, Virtual Appliance.
- La publication du format du fichier de la machine virtuelle amplifie les mouvements autour de VMware.
- Le trio ESX - VMotion - Virtual Center unique sur le marché et qui reste une solution incomparable.

les nouveautés

Les annonces du 5 juin 2006 réorganisent les produits autour de trois pôles de virtualisation:

- Centre de calcul: VMware Infrastructure regroupant ESX 3.0, VMFS, SMP, Virtual Center 2.0 et VMotion 2.0. Deux nouveautés sont prometteuses:
 - ▶ VMware High Availability, HA: permet la gestion automatique de parc de serveur et leur reprise en cas de problème.
 - ▶ VMware Distributed Resource Scheduler, DRS: gère automatiquement l'allocation de ressource pour répartir les charges

- Développeur: contient Workstation et VMTN
- Entreprise Desktop: contient une solution d'intégration pour les postes clients (ACE, Assured Computing Environment) et une infrastructure virtuelle de reprise (VDI, Virtual Desktop Infrastructure).

Si Microsoft devient un concurrent sérieux, l'avance technologique de VMware est impressionnante. Il est difficile de prévoir l'évolution, mais je reste persuadé que l'humain sera le vainqueur.

NAP

JAVA permet de ne pas dépendre d'une configuration système par l'utilisation d'environnement virtuel. Pour optimiser les performances, une nouvelle entreprise a réalisé l'évolution logique, matérialisez la virtualité: Azul Systems.

Cette technologie ajoute un nouvel acronyme: Network Attached Processing, NAP. Par l'intégration de 4x24 processeurs 64 bits dans un serveur, cette technologie offre une nouvelle manière d'envisager le déploiement d'applications JAVA. Il est naturel que le virtuel rejoigne le matériel.

CONCLUSIONS

La mise en œuvre d'un serveur ESX n'est pas insignifiante, en particulier pour obtenir une configuration correspondant à ses désirs. Une fois cette étape franchie, la question de la pertinence de cette solution, par rapport aux systèmes de gestion antérieure n'existe plus, on veut virtualiser!

Le changement est similaire à celui de l'arrivée de la couleur; on regarde avec intérêt, on évalue le prix et on attend une baisse de prix. Dès que l'achat devient économiquement possible, on achète et l'idée de revenir à la situation antérieure devient totalement impensable.

J'ai vécu cette transition pour: la télévision, la pellicule photographique, les terminaux, les imprimantes, les portables et les PDA.

De manière amusante, la même démarche s'est chaque fois répétée!

ICONOGRAPHIE EXTERNE

- fig. 1, virtualisation, ©VMware 2006, www.vmware.com/pdf/virtualization.pdf
- fig. 2, server-vs-esx, ©VMware 2006, www.vmware.com/pdf/virtualization.pdf
- fig. 8, hash-shared-memory, ©Carl Waldspurger 2002, www.waldspurger.org/carl/papers/esx-mem-osdi02.pdf
- fig. 9, balloons, ©Carl Waldspurger 2002, www.waldspurger.org/carl/papers/esx-mem-osdi02.pdf
- fig. 13, VMotion, ©VMware 2006, www.vmware.com/pdf/vmotion_datasheet.pdf
- fig. 14, xen-io, ©VMware 2006, www.vmware.com/pdf/virtualization_considerations.pdf ■