

Haute école du paysage, d'ingénierie et d'architecture de Genève

Disponibilité d'une architecture virtualisée avec VMware ESX-ESXi

Travail de diplôme

Session 2009

Professeur : LITZISTORF Gérald **Étudiant** : SANDMEIER Loïc Filière Télécommunications Laboratoire de transmission de données

Table des matières

I Introduction	6
II Remerciements	6
III Convention d'écriture	7
1 Étude du matériel et des logiciels utilisés	8
1.1 Composants du réseau	8
1.2 SAN	9
1.3 LUN	10
1.4 VMFS	11
1.5 iSCSI	12
1.5.1 Introduction	12
1.5.2 Adressage	14
1.5.3 Opérations sur un disque iSCSI	16
1.5.3.1 Connexion au disque iSCSI	17
1.5.3.2 Liste des LUNs disponible	
1.5.3.3 Infos globales à propos du disque	19
1.5.3.4 Lecture de la capacité du LUN cible	20
1.5.3.5 Lecture de données	21
1.5.3.6 Écriture de données	22
1.5.3.7 Réservation du LUN suivi d'une écriture	23
1.5.3.8 Logout	23
2 vMotion	24
2.1 Introduction	24
2.2 Fonctionnement	25
2.3 Paramètres	27
2.4 Scénario 1 : vMotion entre serveurs ESXi sur un réseau à 100Mbps	28
2.4.1 Schémas et plan du réseau	28
2.4.2 Méthodologie	28
2.4.3 Mesures	29
2.4.3.1 Échanges entre l'hôte source et l'hôte de destination	31
2.4.3.2 Échanges entre le disque iSCSI et les hôtes ESXi	32
2.4.3.3 Échanges comprenant le VI Serveur / Client	32
2.5 Scénario 2 : vMotion entre serveurs ESXi sur un réseau à 1Gbps	
2.5.1 Schémas et plan du réseau	

2.5.2 Méthodologie	
2.5.3 Mesures	
2.6 Scénario 3 : vMotion entre serveurs ESX sur un réseau à 100Mbps	
2.6.1 Schémas et plan du réseau	
2.6.2 Méthodologie	
2.6.3 Mesures	
2.7 Scénario 4 : vMotion entre ESX sur un réseau à 1Gbps	41
2.7.1 Schémas et plan du réseau	41
2.7.2 Méthodologie	41
2.7.3 Mesures	
2.8 Conclusion	42
3 Distributed Resource Scheduler	43
3.1 Introduction	
3.2 Cluster	
3.3 Fonctionnement	45
3.3.1 Réservations	45
3.4 Paramètres	46
3.5 Topologie et plans du réseau	47
3.6 Scénario 5 : déplacement manuel	
3.6.1 Configuration testée et méthodologie	
3.6.2 Mise en œuvre de la configuration de DRS	
3.6.3 Observations et mesures	50
3.7 Scénario 6 : déplacement automatique, sensibilité maximum	53
3.7.1 Configuration testée et méthodologie	53
3.7.2 Mise en œuvre de la configuration de DRS	53
3.7.3 Observations et mesures	54
3.8 Scénario 7 : déplacement automatique, changement du réglage de la sensibilité	55
3.8.1 Configuration testée et méthodologie	55
3.8.2 Mise en œuvre de la configuration de DRS	55
3.8.3 Observations et mesures	56
3.9 Scénario 8 : 2 VMs doivent rester ensemble	57
3.9.1 Configuration testée et méthodologie	57
3.9.2 Mise en œuvre de la configuration de DRS	57
3.9.3 Observations et mesures	

3.10 Scénario 9 : 2 VMs ne doivent pas être sur le même serveur ESX	60
3.10.1 Configuration testée et méthodologie	60
3.10.2 Mise en œuvre de la configuration de DRS	61
3.10.3 Observations et mesures	61
3.11 Conclusion	61
4 Conclusions	62
A Annexes	63
A.1 Versions des logiciels utilisés	63
A.2 Matériel	63
A.3 Installation d'Openfiler	63
A.4 Installation et configuration de ESX/ESXi	64
A.4.1 ESXi	64
A.4.2 ESX	64
A.5 Virtual Infrastructure Server (VI Server)	65
A.6 Virtual Infrastructure Client (VI Client)	68
A.6.1 Installation	68
A.6.2 Ajout/création d'un Datacenter	68
A.6.3 Ajout d'un host	69
A.6.4 Ajout d'un VMkernel	72
A.6.5 Ajout d'un disque iSCSI	76
A.6.6 Activation de vMotion	83
A.6.7 Déplacer une VM à l'aide de vMotion	85
A.6.8 Création d'un cluster DRS	87
A.7 Tests du switch NETGEAR	92
A.7.1 Schémas et plans du réseau	92
A.7.2 Méthodologie	92
A.7.3 Mesures	92
A.8 Script	93
A.9 Problèmes rencontrés et bugs	94
A.9.1 Bug de déconnection aléatoire	94
A.9.2.1 Symptômes	94
A.9.2.2 Cause	94
A.9.2.3 Solution	94
A.9.2 Problème d'effacement des VMs	94

A.9.2.3 Solutions 94 A.9.3 Problème lors d'usage simultané de plusieurs VMs 95 A.9.3.1 Symptômes 95 A.9.3.2 Solutions et causes 95 A.9.4 Erreur lors d'accès à certaines pages de l'interface web d'Openfiler 95 A.9.4.1 Symptômes 95 A.9.4.2 Cause et solution 96		A.9.2.1 Symptômes	.94
A.9.3 Problème lors d'usage simultané de plusieurs VMs 95 A.9.3.1 Symptômes 95 A.9.3.2 Solutions et causes 95 A.9.4 Erreur lors d'accès à certaines pages de l'interface web d'Openfiler 95 A.9.4.1 Symptômes 95 A.9.4.2 Cause et solution 96		A.9.2.3 Solutions	.94
A.9.3.1 Symptômes	A.	9.3 Problème lors d'usage simultané de plusieurs VMs	.95
A.9.3.2 Solutions et causes		A.9.3.1 Symptômes	.95
 A.9.4 Erreur lors d'accès à certaines pages de l'interface web d'Openfiler		A.9.3.2 Solutions et causes	.95
A.9.4.1 Symptômes	A.	9.4 Erreur lors d'accès à certaines pages de l'interface web d'Openfiler	.95
A.9.4.2 Cause et solution96		A.9.4.1 Symptômes	.95
		A.9.4.2 Cause et solution	.96

I Introduction

Actuellement, les solutions de virtualisation sont de plus en plus utilisées, et encore plus particulièrement celles proposées par VMware. Il y a plusieurs raisons à cela, dont la principale est l'économie que représentent de telles solutions. En effet, là ou il était nécessaire d'avoir de nombreux serveurs physiquement indépendants, seuls 2 ou 3 serveurs hébergeant des machines virtuelles suffisent (ces serveurs sont plus puissant que les serveurs traditionnels, mais cette puissance est mieux exploitée). Il est également très aisé de créer de nouvelles machines virtuelles, ce qui s'avère très pratique en environnement de test.

Cependant, plusieurs serveurs virtuels sur un même matériel impliquent que si ce matériel est défectueux, ou qu'un serveur virtuel mobilise trop de ressources, il peut en résulter une indisponibilité d'une partie ou de la totalité des machines virtuelles. C'est là qu'interviennent les fonctionnalités proposées par VMware pour ses serveurs ESX et ESXi, testées lors de ce projet.

Mon choix s'est porté sur ce projet, car la virtualisation m'intéresse particulièrement pour plusieurs raisons, la principale étant la curiosité. Je connaissais déjà les machines virtuelles, mais le cours dispensé par M. Litzistorf m'a permis de me rendre compte de l'étendue des possibilités offertes par les serveurs ESX et ESXi, que je ne connaissais pas.

Mon travail s'est déroulé en 4 parties, durant une période de 8 semaines :

- 1. Installation et configuration des produits VMware, et correction des problèmes et bugs de fonctionnement. Cette première partie, répartie tout au long du projet, a duré environ trois semaines.
- Comprendre et expliquer le fonctionnement des logiciels et du matériel utilisé. Cette partie explique principalement le fonctionnement des disques iSCSI utilisé pour stocker les machines virtuelles. Cette première partie s'est déroulée sur une période d'une semaine et demi.
- 3. Tester et comprendre le fonctionnement de vMotion, une fonctionnalité qui permet de déplacer une machine virtuelle d'un serveur physique à un autre alors qu'elle est en état de fonctionnement. VMware annonce que lors de ce type de déplacement, aucune interruption de service n'a lieu, l'objectif principal est donc de vérifier cette affirmation. Cette partie fut l'élément central de mon travail, elle s'est déroulée sur une période de 4 semaines.
- 4. Tester plusieurs configurations de DRS, une fonctionnalité qui à pour but d'équilibrer la charge CPU et RAM des serveurs ESX et ESXi faisant partie d'un même groupe. DRS utilise vMotion pour rééquilibrer la charge, c'est pourquoi cette partie n'est étudiée qu'après vMotion. 4 jours ont été consacrés à l'étude de cette fonctionnalité.

II Remerciements

Je tiens à remercier M. Gérald LITZISTORF pour m'avoir proposé ce projet et également pour ses conseils et recommandations. Je remercie aussi M. José TAVARES pour son aide.

7

III Convention d'écriture

Le texte écrit en **gras** est important, celui écrit en *italique* est une citation ou, dans les annexes, le texte sur un bouton à cliquer. Le texte entre « guillemets » est une copie d'une partie du texte d'une boite de dialogue. Le texte blanc sur fond noir correspond aux lignes de commandes.

L'abréviation VM signifie Virtual Machine ou Machine Virtuelle en français, et VI signifie Virtual Infrastructure.

1 Étude du matériel et des logiciels utilisés



1.1 Composants du réseau

Figure 1.1 : plan regroupant les principaux composants réseau

Les composants du réseau sont :

- Le VI Serveur (réseau de gestion)
- Le VI Client (réseau de gestion)
- Le disque iSCSI (réseau de stockage)
- Les serveurs ESX/ESXi (réseau de gestion et réseau de stockage)

Le VI Serveur est un service Windows qui supervise/contrôle un ou plusieurs serveurs ESX/ESXi qui lui ont été attribués. vMotion et DRS exigent sa présence.

Le VI Client est l'interface utilisateur qui se connecte soit directement sur un serveur ESX/ESXi, soit sur le VI Serveur, permettant de contrôler plusieurs serveurs dans une seule fenêtre et d'exploiter les fonctionnalités du VI Serveur. Le VI Client peut être installé sur plusieurs machines, dont celle qui héberge déjà le service du VI Serveur (configuration utilisée lors de mon travail).

Le disque iSCSI contient les fichiers des VMs d'un ou plusieurs serveurs ESX/ESXi. Ces fichiers correspondent :

- au disque dur virtuel, dont la taille varie selon la quantité de données contenues
- à la partition swap, de taille également variable, et uniquement présente lorsque la VM est en fonction
- à la description des caractéristiques matérielles de la VM (CPU, RAM, disque dur virtuel, ...), de taille plus petite que 3kB
- aux logs de la VM, de taille maximum équivalente à 30kB

Comme plusieurs serveurs peuvent y accéder, son utilisation rend possible des fonctions comme vMotion en réduisant la quantité de données à déplacer lors d'une migration. N'ayant pas de disque iSCSI physique, un disque a du être émulé. Pour cette émulation, le premier choix a été d'utiliser FreeNAS (une distribution BSD basée sur FreeBSD), hélas, ce dernier supportait mal les accès multiples, rendant l'ensemble de l'infrastructure inutilisable (voir annexe A.9.3). C'est pourquoi le

choix s'est porté sur Openfiler (une distribution GNU/Linux) qui gère correctement ces accès multiples, mais qui ne permet pas de profiter pleinement des capacités d'un réseau gigabit (plus de 12 minutes pour transférer 3 GB, soit 3.9 MBps, voir annexe A.7).

Les serveurs ESX/ESXi ont pour rôle de faire fonctionner les VMs en leur attribuant les ressources nécessaires. Il s'agit à la base d'un OS GNU/Linux RedHat modifié et allégé le plus possible par VMware. ESXi est la version gratuite et allégée de ESX. Les deux types ont été utilisé pour comparer les différences entre un produit payant et un autre gratuit, tant au niveau des performances que des fonctionnalités.

Les VI Client et Serveur ainsi que les serveurs ESX sont en version 3.5, avec des licences d'évaluations. Cette version était la plus récente au moment ou mon travail de diplôme a débuté, si bien que nous avons convenu de ne pas utiliser la version 4, vSphere, sortie le 21 avril 2009.

1.2 SAN

Selon Wikipedia, <u>http://fr.wikipedia.org/wiki/R%C3%A9seau_de_stockage_SAN</u>, un SAN, Storage Area Network ou réseau de stockage, se différencie des autres systèmes de stockage tel que le NAS (Network attached storage) par un accès bas niveau aux disques. Pour simplifier, le trafic sur un SAN est très similaire aux principes utilisés pour l'utilisation des disques Internes (ATA, SCSI). C'est une mutualisation des ressources de stockage.

Dans le cas du SAN, les baies de stockage n'apparaissent pas comme des volumes partagés sur le réseau. Elles sont directement accessibles en mode bloc par le système de fichiers des serveurs. En clair, chaque serveur voit l'espace disque d'une baie SAN auquel il a accès comme son propre disque dur. L'administrateur doit donc définir très précisément les LUNs (unités logiques), le masking et le zoning, pour qu'un serveur Unix n'accède pas aux mêmes ressources qu'un serveur Windows utilisant un système de fichiers différent.

L'utilisation d'un SAN est nécessaire pour utiliser des fonctionnalités comme vMotion. Le SAN utilisé durant ce projet contient le disque iSCSI émulé par Openfiler ainsi que les serveurs ESX et ESXi. Une fois le disque lié aux serveurs ESX/ESXi, ces derniers y accèdent comme si le disque était une database interne. Il est recommandé dans les bests parctices

(<u>http://www.vmware.com/pdf/vc_technical_best.pdf</u> p.12) d'utiliser un réseau dédié pour le stockage, bien que ce ne soit pas toujours possible, comme je l'explique au début du scénario 1 (chapitre 2.4).

1.3 LUN

Selon Wikipedia, <u>http://fr.wikipedia.org/wiki/LUN</u>, un LUN (**Logical Unit Number**) est, en informatique, un "identifiant d'unité logique", c'est-à-dire un pointeur vers un espace de stockage. Un LUN est caractérisé par sa taille (en Giga octets en général).

Un contrôleur iSCSI peut gérer jusqu'à 8 unités logiques, le LUN est donc codé sur 3 bits.

Le seul moment où l'on voit les LUNs dans l'utilisation des produits VMware, c'est au moment de l'ajout d'un disque iSCSI à un serveur ESX/ESXi. Si ce dernier n'est pas formaté en VMFS, il ne sera pas détecté automatiquement, il faut donc le configurer à la main. Durant cet ajout, tout les LUNs vu par le serveur sont proposé, et il faut choisir lequel sera formaté. Pour plus d'info, voir le chapitre A.6.5.

🖉 Add Storage Wizard 📃 🗆 🔀									
Select Disk/LUN If a device cannot be configured unambiguously, you will be asked to select a partition.									
Disk/LUN Device Location			5	SAN Identifier contains: 👻		Clear			
Current Disk Layout	Device	Capacity	Available	SAN Identifier		LUN			
Properties	vmhba32:3:0	48.84 GB	None	ign.2006-01.com.openfiler	r:tsn.9ed4ca97e2	0			
Ready to Complete									
Help				≤Back	Next ≥	Cancel			

Figure 1.2 : choix du LUN à formater

L'appareil est également identifié par une chaine de la forme vmhba32 :0 :0.

Selon le document

<u>http://publib.boulder.ibm.com/infocenter/dsichelp/ds6000ic/index.jsp?topic=/com.ibm.storage.smr</u> <u>ic.help.doc/f2c_vmesxlunprobe_1m5c8p.html</u>, cet identifiant peut être décomposé en trois parties :

- La première partie, vmhba32, définit le contrôleur iSCSI utilisé. Le chiffre 32 peut varier d'une machine à l'autre, mais vmhba est fixe.
- Le premier 0 identifie le disque iSCSI (la cible) sur lequel se trouve l'unité logique.
- Le second zéro correspond au LUN.

1.4 VMFS

Le VMFS, ou Virtual Machine File System, est le système de fichiers pour les unités de stockage développé par VMware.

La principale raison de la création de ce système de fichiers est la nécessité de garantir l'accès concurrent en lecture et écriture aux mêmes fichiers à plusieurs clients (dans ce cas, les serveurs ESX/ESXi). Cet accès concurrent est vital pour des fonctionnalités telles que vMotion, hors les systèmes de fichiers utilisés, par exemple, par les disques iSCSI ne permettent pas ce genre d'opération.

Par le biais du VI Client, il est possible de formater des unités logiques sur des disques (i)SCSI en VMFS. Le LUN devient ainsi un volume VMFS.

Comme il permet l'accès multiple, le VMFS doit également garantir que l'accès en écriture n'est possible que depuis un client à la fois. Un système de verrous est ainsi utilisé pour garantir qu'une seule écriture à la fois est réalisée sur un fichier donné.

L'utilisation du VMFS est transparente pour l'utilisateur, qui indique les opérations à effectuer depuis le VI Client, mais n'a pas besoin d'intervenir lors du déroulement de ces dernières.

Voici un lien vers la fiche du produit (en anglais) : <u>http://www.vmware.com/pdf/vmfs_datasheet.pdf</u>

1.5 iSCSI

Ce chapitre s'appuie principalement sur le document <u>http://docs.hp.com/en/T1452-90011/T1452-90011.pdf</u>

1.5.1 Introduction

Le SCSI est un protocole qui permet de communiquer avec des unités (disque dur, lecteur CD, imprimante) à l'intérieur d'un SAN. Le protocole iSCSI transporte le primitives SCSI au travers d'un réseau IP. De ce fait, pour comprendre le fonctionnement du iSCSI, il est nécessaire de comprendre celui du SCSI.

Lors d'un échange SCSI, on a d'un côté l'initiateur (le client) et de l'autre, la cible (le serveur). Le disque dur (ou autre périphérique) SCSI se situe sur la cible, qui est composées d'une ou plusieurs unités logiques. Chaque unité logique est indépendante, et peut donc recevoir des commandes de la part de l'initiateur. Pour savoir à quelle unité est destinée une commande, chaque unité se voit attribuer un LUN (voir 1.5.1). Ce LUN est joint à chaque requête.

Le protocole SCSI est prévu pour être utilisé sur des réseaux de données dédiés à haute vitesse (souvent fiber channel). Seulement, ces solutions sont extrêmement couteuses, et par conséquent, les entreprises de petite et moyenne tailles ne peuvent pas se permettre ce type d'investissement. C'est pour cette raison qu'est né le iSCSI.

Le protocole iSCSI, qui est une surcouche du TCP, encapsule les requêtes (les primitives) SCSI afin de les faire transiter sur un réseau IP.



Figure 1.3 : empilement protocolaire du iSCSI

Pour plus de détails, voici un lien qui explique bien et de façon simple le fonctionnement de iSCSI : <u>http://www.supinfo-projects.com/fr/2006/stockres_fr/4/</u>.

Les messages échangés entre un initiateur et une cible sont appelé PDU (iSCSI Protocol Data Unit). C'est les messages PDU qui contiennent les requêtes ainsi que les réponses SCSI. Ces messages ne sont pas limités à la taille disponible dans un paquet TCP, ils peuvent être segmenté et envoyé dans plusieurs paquets TCP. Ils sont de la forme suivante :



Figure 1.4 : message PDU

Seule la partie entête est obligatoire, les commandes ne nécessitant pas toujours un transfert de données. Par exemple, voici l'entête et les données d'une réponse à une demande de lecture (pour plus de détails, voir chapitre 1.5.3.5) :

🖃 iSCSI (SCSI Data In)
Opcode: SCSI Data In (0x25)
🗉 Flags: 0x81
Status: Good (0x00)
TotalAHSLength: 0x00
DataSegmentLength: 0x00000008
LUN: 00000000000000
InitiatorTaskTag: 0x0000000b
TargetTransferTag: 0xffffffff
StatSN: 0x000000c
ExpCmd5N: 0x000000c
MaxCmd5N: 0x000002c
DataSN: 0×0000000
BufferOffset: 0x00000000
ResidualCount: 0x00000000
<u>Request in: 110</u>
Time from request: 0.000112000 seconds
Figure 1.5 : entête d'une réponse à une requête de lecture

```
SCSI Payload (Read Capacity(10) Response Data)
[LUN: 0x0000]
[command Set:Direct Access Device (0x00) ]
[SBC opcode: Read Capacity(10) (0x25)]
[Request in: 110]
[Response in: 111]
LBA: 102432767 (48 GB)
Block size in bytes: 512
```

Figure 1.6 : données d'une réponse à une requête de lecture

1.5.2 Adressage

L'adressage utilisé par Openfiler est de type iqn (iSCSI Qualified Name). Il existe un deuxième type d'adresse (eui ou IEEE EUI-64 format), qui ne sera pas détaillé dans ce document.

Selon http://www.cuddletech.com/articles/iscsi/ar01s02.html, un iqn est de la forme suivante :

iqn.2006-01.com.openfiler:tsn.9ed4ca97e298

iqn est fixe, il permet d'identifier le type d'adressage utilisé.

2006-01 correspond au premier mois entier ou le nom de domaine d'authentification a été en fonction.

com.openfiler est le nom du domaine d'authentification inversé (openfiler.com). C'est ce domaine qui attribue l'adresse iqn.

Ce qui suit les deux points est une chaine de caractères déterminée par le domaine d'authentification ne suivant aucune règle particulière. Chaque domaine est libre de choisir ses propres règles ou de générer aléatoirement cette suite de caractères.

Pour obtenir cette adresse en mode dynamic discovery, il faut établir une session avec le disque iSCSI voulu en se servant de son adresse IP. Cette session ne restera ouverte que le temps d'obtenir l'adresse iqn du disque cible. Voici les échanges qui permettent d'aboutir à son obtention :

→Login Request (annonce du type de session désiré, méthode d'authentification)

←Login Response (connexion acceptée)

 \rightarrow Login Request (indication des méthodes connues pour signer l'échange, proposition de configuration pour divers paramètre)

←Login Response (méthodes acceptées, indication des valeurs retenue pour les paramètres)

→Command (demande de l'envoi de l'adresse des cibles utilisant cette adresse IP)

← Response (adresse des cibles)

→Logout Request (demande de déconnection)

←Logout Response (validation de la demande)

Voici un exemple pratique d'obtention d'adresse ign :

```
🗏 Key/Value Pairs
    KeyValue: InitiatorName=iqn.1991-05.com.microsoft:g11
    KeyValue: SessionType=Discovery
    KeyValue: AuthMethod=None
            Figure 1.7 : annonce du type de session voulu
             Status: Success (0x0000)
           ⊟ Key/Value Pairs
               KeyValue: TargetPortalGroupTag=1
               KeyValue: AuthMethod=None
                   Figure 1.8 : session accepté
       ⊟ Key/Value Pairs
           KeyValue: HeaderDigest=None,CRC32C
           KeyValue: DataDigest=None,CRC32C
           KeyValue: MaxRecvDataSegmentLength=65536
           KeyValue: DefaultTime2Wait=0
           KeyValue: DefaultTime2Retain=60
```

Figure 1.9 : annonce des méthodes de hash connues et proposition de valeurs pour divers paramètres

```
Status: Success (0x0000)

■ Key/Value Pairs

KeyValue: HeaderDigest=None

KeyValue: DataDigest=None

KeyValue: DefaultTime2Wait=2

KeyValue: DefaultTime2Retain=20
```

Figure 1.10 : indications des méthodes de hash choisie et valeurs retenues pour les paramètres

■ Key/Value Pairs KeyValue: SendTargets=All

Figure 1.11 : demande de l'envoi des adresses iqn des cibles

Key/Value Pairs
KeyValue: TargetName=iqn.2006-01.com.openfiler:tsn.9ed4ca97e298
KeyValue: TargetAddress=10.1.1.54:3260,1

Figure 1.12 : envoi de la liste

Le logout ne comportant pas de paramètres particuliers, les captures d'écrans ne sont pas nécessaires.

1.5.3 Opérations sur un disque iSCSI

Pour illustrer les opérations qu'il est possible d'effectué sur un disque iSCSI, le scénario pédagogique suivant va être employé :

- On effectuera en premier lieu une connexion sur le disque iSCSI (p.15)
- On demandera ensuite au disque de nous fournir une liste des LUNs qu'il contient (p.16)
- On lui demandera également des informations plus globales (p.17)
- On cherchera à connaître la capacité du LUN cible (p.18)
- On procèdera à une lecture sur un LUN (p.19)
- Suivra une écriture sur le même LUN (p.20)
- Une réservation du disque sera effectuée, durant laquelle une écriture aura lieu (p.21)
- Pour finir, il faudra se déconnecter du disque (p.21)

On part du principe que lors de ces échanges, l'identifiant iqn de la cible est connu par l'initiateur.

Ces opérations nécessitent une ou plusieurs commandes SCSI. Pour toutes commandes, il faut préciser le disque de destination ainsi que le LUN (voir chapitre 1.3), au cas où le disque hébergerait plusieurs LUNs.

La version des commandes SCSI utilisée est SPC-2, cette information est communiqué au travers du premier Inquiry (voir p.17) de chaque LUN :

Version: Compliance to SPC-2 (0x04)

Figure 1.13 : version des commandes iSCSI utilisées

Les points suivants s'appuient sur la page Wikipedia suivante

<u>http://en.wikipedia.org/wiki/SCSI_command</u> ainsi que les pages qui y sont reliées. Les autres sources seront indiquées au cas par cas.

Les captures Wireshark d'où sont extraites les captures d'écran utilisées pour illustrer ce scénario sont disponibles sur le DVD : iSCSI/iSCSI_rescan.pcap et iSCSI/ech_xp.pcap en utilisant le filtre ip.addr==10.1.1.54 (et 10.1.2.155 pour la 2^{ème} capture). La première capture est obtenue grâce à la commande esxcfg-swiscsi —s, qui lance une découverte des LUNs sur les disques iSCSI qui lui sont liés. La commande est effectuée depuis un serveur ESX, et est issue du document <u>http://www.rtfm-ed.co.uk/docs/vmwdocs/ESX3.x-VC2.x-ServiceConsole-Guide.pdf</u>. La deuxième capture est issue d'un échange de fichiers entre le disque iSCSI et un poste sous Windows XP, et contient également la détection du disque iSCSI au poste XP.

Comme le scénario est pédagogique, il ne suit pas l'ordre des captures Wireshark, les paquets ont été choisit afin d'illustrer le scénario décrit plus haut.

Sandmeier Loïc

1.5.3.1 Connexion au disque iSCSI

Durant la phase de connexion (ou login), le but est d'identifié, avec ou sans mot de passe, les participants (la cible et l'initiateur) des futurs échanges. Pour simplifier, aucun mot de passe ne sera utilisé dans cet exemple.

Les échanges suivants ont lieu durant cette étape (l'initiateur est à gauche, il en sera de même pour tout les échanges suivants) :

→Login Request (annonce du type de session désiré, annonce de la méthode d'authentification)
 ←Login Response (connexion acceptée)

→Login Request (indication des méthodes connues pour signer l'échange, proposition de configuration pour divers paramètre)

←Login Response (méthodes acceptées, indication des valeurs retenue pour les paramètres)

```
⊟ Key/Value Pairs
    KeyValue: InitiatorName=iqn.1991-05.com.microsoft:g11
    KeyValue: SessionType=Normal
    KeyValue: TargetName=iqn.2006-01.com.openfiler:tsn.9ed4ca97e298
    KeyValue: AuthMethod=None
    Figure 1.14 : annonce du type de session voulu et annone de la cible voulue
                Status: Success (0x0000)
              □ Key/Value Pairs
                  KeyValue: TargetPortalGroupTag=1
                  KeyValue: AuthMethod=None
                     Figure 1.15 : session accepté
        ⊟ Key/Value Pairs
             KeyValue: HeaderDigest=None,CRC32C
             KeyValue: DataDigest=None,CRC32C
             KeyValue: ErrorRecoveryLevel=2
             KeyValue: InitialR2T=No
             KeyValue: ImmediateData=Yes
             KeyValue: MaxRecvDataSegmentLength=65536
             KeyValue: MaxBurstLength=262144
             KeyValue: FirstBurstLength=65536
             KeyValue: MaxConnections=8
             KeyValue: DataPDUInOrder=Yes
             KeyValue: DataSequenceInOrder=Yes
             KeyValue: DefaultTime2Wait=0
             KeyValue: DefaultTime2Retain=60
             KeyValue: MaxOutstandingR2T=16
```

Figure 1.16 : annonce des méthodes de hash connues et proposition de valeurs pour divers paramètres

Status: Success (0x0000) ⊟ Key/Value Pairs KeyValue: HeaderDigest=None KeyValue: DataDigest=None KeyValue: ErrorRecoveryLevel=0 KeyValue: InitialR2T=Yes KeyValue: ImmediateData=No KeyValue: MaxBurstLength=262144 KeyValue: FirstBurstLength=65536 KeyValue: MaxConnections=1 KeyValue: DataPDUInOrder=Yes KeyValue: DataSequenceInOrder=Yes KeyValue: DefaultTime2Wait=2 KeyValue: DefaultTime2Retain=20 KeyValue: MaxOutstandingR2T=8 KeyValue: MaxRecvDataSegmentLength=131072

Figure 1.17 : indications des méthodes de hash choisie et valeurs retenues pour les paramètres

1.5.3.2 Liste des LUNs disponible

Cette phase permet d'obtenir une liste des LUNs disponibles sur le disque.

Les échanges suivants ont lieu durant cette étape :

→Report LUNs
←Report LUNs Response (Liste des LUNs)

```
LUN List Length: 16
LUN: 0
LUN: 1
```

Figure 1.18 : liste des LUNs obtenue suite à l'échange

1.5.3.3 Infos globales à propos du disque

Il s'agit ici d'obtenir des informations à propos du disque allant du constructeur jusqu'à la méthode d'accès aux LUNs. Les échanges suivants ont lieu pour chaque LUN durant cette étape :

→Inquiry (LUN, type d'infos attendu en retour)
←Inquiry Response (infos demandées)
→Inquiry (LUN, type d'infos attendu en retour)
←Inquiry Response (infos demandées)
→Inquiry (LUN, type d'infos attendu en retour)
←Inquiry Response (infos demandées)
→Inquiry (LUN, type d'infos attendu en retour)
←Inquiry Response (infos demandées)

Le premier échange est le plus complet, et celui qui contient les informations principales, je parlerai donc uniquement de celui-ci, et non des 3 suivants.

Les informations obtenues sont :

- 1. L'accès au périphérique est direct ou indirect
- 2. Le type de cible (lecteur CD, imprimante, disque dur, ...)
- 3. La cible est-elle retirable à chaud (un CD peut-être retiré du lecteur CD)
- 4. La cible supporte-t-elle les événements asynchrones (p.ex : interruptions asynchrones)
- 5. La version des commandes iSCSI supportées par le LUN
- 6. La cible supporte-t-elle l'adressage relatif (par rapport au LUN et non au disque entier)

```
■ SCSI Payload (Inquiry Response Data)
    [LUN: 0x0000]
    [Command Set:Direct Access Device (0x00) ]
    [SBC Opcode: Inquiry (0x12)]
    [Request in: 94]
   [Response in: 95]
 Peripheral: 0x00, Device type is connected to logical unit, Direct Access Device
  2 000. .... = Qualifier: Device type is connected to <u>logical unit</u> (0x00)
     ...0 0000 = Device Type: Direct Access Device (0x00)
 1
  🗏 Flags: 0x00
 з
     0... .... = Removable: This device is NOT removable
  5 version: compliance to SPC-2 (0x04)
  Flags: 0x52, TrmTsk, Hisup, SPC-2/SPC-3
 4
     0... .... = AERC: Async event reporting capability is NOT supported
     .1.. .... = TrmTsk: Terminate Task management functions are SUPPORTED
     ..... = NormACA: Normaca is NOT supported
     ...1 .... = HiSup: Hierarchical Addressing Mode is SUPPORTED
      .... 0010 = Response Data Format: SPC-2/SPC-3 (2)
   Additional Length: 59
 ∃ Flags: 0x00
  🖃 Flags: 0x02 CmdQue
    0... .... = RelAdr: <u>Relative addressing</u> mode is <u>NOT supported</u>
     ...0 .... = Sync: Synchronous data transfer is NOT supported
     .... 0... = Linked: Linked commands are NOT supported
      .... ..1. = CmdQue: Command queuing is SUPPORTED
   Vendor Id: OPNFILER
   Product Id: VIRTUAL-DISK
   Product Revision Level: 0
```

Figure 1.19 : détail de la réponse contenant les informations principales

1.5.3.4 Lecture de la capacité du LUN cible

Pour connaître la capacité d'un LUN, il faut utiliser la commande Read Capacity, qui se déroule de la manière suivante :

→Read Capacity (LUN cible)
 ←Read Capacity Response (Capacité du LUN, taille des blocs du système de fichier)

Figure 1.20 : commande Read Capacity

Le 10 entre parenthèses indique que le disque utilise un bus de 32 bits.

Figure 1.21 : réponse à un Read Capacity

1.5.3.5 Lecture de données

La lecture de données introduit des échanges particuliers : selon la quantité de données à lire, la réponse ne tient pas dans un seul paquet TCP, c'est pourquoi il sera également échangé de paquets contenant des segments PDU en plus des habituelles commandes et réponses.

Les échanges se déroulent comme suit :

→Read (LUN, adresse et taille des données à lire)
 ←Read Response(validation de la requête, données)
 ←PDU segment + TCP ACK (données)
 ←PDU segment + TCP ACK (données)
 →TCP ACK

Les trois derniers échanges étant facultatifs, et se déroulant autant de fois que nécessaire pour transmettre la totalité des données. La longueur des paquets contenant les segments PDU est de 1514 Bytes, exception faite du paquet contenant le dernier segment PDU du transfert.

Voici le détail de la requête :

```
□ SCSI CDB Read(10)
    [LUN: 0x0000]
    [Command Set:Direct Access Device (0x00) (Using default commandset)]
    [Response in: 49880]
    Opcode: Read(10) (0x28)
  ■ Flags: 0x00
      000. .... = RDPROTECT: 0x00
      ...0 .... = DPO: Disable page out is DISABLED (cache this data)
      .... 0... = FUA: Read from cache if possible
      .... .. 0. = FUA_NV: Read from volatile or non-volatile cache permitted
                                                    Adresse des données
    Logical Block Address (LBA): 3267 -
    ...0 0000 = Group: 0x00
    Transfer Length: 64-
                                                    → Taille des données
    Vendor Unique = 0, NACA = 0, Link = 0
```

Figure 1.22 : commande de lecture

1.5.3.6 Écriture de données

Pour écrire les données, la démarche est un peu plus complexe, car il faut commencer par prévenir le disque que l'on compte faire une écriture, puis attendre son autorisation pour pouvoir commencer à écrire.

Les échanges suivants ont lieu :

→Write (bloc ou commencera l'écriture, taille des données à écrire)
 ←Ready To Transfert (rappel de la taille)
 →Write Request Data (données)
 ←Write Response (indique si l'écriture s'est bien déroulée)

Voici le détail de la demande d'écriture :

```
    SCSI CDB Write(10)
        [LUN: 0x0000]
        [Command Set:Direct Access Device (0x00) (Using default commandset)]
        [Response in: 49001]
        opcode: Write(10) (0x2a)
        Flags: 0x00
        Logical Block Address (LBA): 108816 -> Bloc à partir duquel on veut écrire
        ...0 0000 = Group: 0x00
        Transfer Length: 1 -> Taille des données à écrire
        Vendor Unique = 0, NACA = 0, Link = 0
```

Figure 23 : commande Write

1.5.3.7 Réservation du LUN suivi d'une écriture

Ce passage s'appuie également du la page suivante : <u>http://support.microsoft.com/kb/309186/fr</u>

Dans le cas de certaines écritures, **l'initiateur réserve en premier lieu le LUN**. Cette opération est nécessaire si la zone du disque ou doit avoir lieu l'écriture est également utilisée par un autre initiateur. Dans le cas de vMotion, deux serveurs ESX/ESXi sont amené à gérer conjointement une même VM, et par conséquent, les fichiers qui s'y rapportent sur le disque. C'est pourquoi au moment ou le serveur de destination prend la main sur une partie des données correspondant à la VM, elle réserve le disque afin de modifier en premier lieu un byte qui sert à indiquer au serveur de départ qu'il n'a plus accès en écriture à ces données.

Cette opération implique que seul l'initiateur qui a réservé le disque peut y accéder, les autres requêtes seront ignorées.

Dans cet exemple, l'initiateur effectuera une écriture une fois la réservation effective. Il va de soi que d'autres commandes sont également réalisables, comme une lecture, bien que ces dernières ne nécessitent généralement pas de réservation.

Les étapes se dérouleront de la façon suivante :

→Reserve (LUN à réserver)
←Reserve Response (LUN, validation de la réservation)
→Write (LUN, bloc ou commencera l'écriture, taille des données à écrire)
←Ready To Transfert (LUN, rappel de la taille)
→Write Request Data (LUN, données)
←Write Response (LUN, indique si l'écriture s'est bien déroulée)
→Release (LUN)
← Release (LUN)

 \leftarrow Release Response (LUN, confirmation du release)

L'opération d'écriture se déroule comme la précédente, et les opérations spécifiques à la réservation ne demandant aucun paramètre particulier, il n'est pas nécessaire de montrer des captures de ces paquets.

1.5.3.8 Logout

Le logout s'effectue en une seule commande et réponse :

→Logout Request (demande de déconnection)
 ←Logout Response (validation de la demande)

Comme aucun paramètre particulier n'est requis, une capture d'écran serait inutile.

2 vMotion

2.1 Introduction

Selon VMware <u>http://www.vmware.com/files/fr/pdf/vmotion_datasheet_fr.pdf</u>, vMotion est une fonctionnalité des serveurs ESX et ESXi servant à transférer une VM d'un hôte à un autre sans éteindre cette dernière.



VMware VMotion permet de déplacer des machines virtuelles en fonctionnement depuis un serveur physique vers un autre sans interruption de service.

Figure 2.1 : principe de vMotion

VMware[®] VMotion[™] permet de déplacer en temps réel des VMs en fonctionnement depuis un serveur physique vers un autre sans interruption de service, tout en garantissant une disponibilité continue et une totale intégrité des transactions. VMotion est une technologie clé dans le cadre de la mise en place d'un datacenter dynamique, automatisé et optimisé.

Grâce à VMotion, les utilisateurs peuvent :

• Optimiser et allouer automatiquement des pools entiers de ressources pour améliorer le taux d'utilisation, la flexibilité et la disponibilité des ressources matérielles.

- Effectuer la maintenance des ressources matérielles sans planifier d'interruption de service.
- Migrer proactivement les VMs des serveurs défaillants ou insuffisamment performants.

2.2 Fonctionnement

Ce paragraphe est basé sur l'étude de l'article http://www.vmware.com/pdf/usenix_vmotion.pdf.

La structure d'un système ESX ou ESXi est la suivante :

- 1. La couche matérielle (CPU, RAM, interfaces réseau, disque dur interne et réseau, pour plus de détails, voir l'annexe A.2)
- 2. Le VMkernel, qui gère les connections sur le réseau virtuel et les VMM. Cette fonction est désactivée par défaut sur ESX, mais nécessaire pour utiliser vMotion
- 3. Une couche intermédiaire, VMM (Virtual Machine Monitor), qui gère, entre autre, le matériel « virtuel »
- 4. L'OS « invité », qui doit avoir l'impression de tourner sur du vrai matériel



Figure 2.2 : les différentes couches d'un serveur ESX/ESXi

Une VM est composée en réalité de fichiers qui représentent ses composants et sa configuration : un fichier qui définit les caractéristiques (taille de la RAM, fréquence du processeur, ...), un autre qui contient le disque dur virtuel, etc...

Afin de rendre les VMs encore moins dépendantes du serveur ESX/ESXi qui lui fournit les ressources nécessaires à son fonctionnement, les fichiers composant une VM seront stockés dans un SAN sur un disque iSCSI. Il en va de même pour toutes les VMs fonctionnant sur les serveurs d'un même groupe.

Les principaux fichiers composant une VM sont les .vmdk et le .vmx. Les .vmdk contiennent le disque dur de la VM, et le .vmx la configuration matérielle (taille de la RAM et du disque dur, nombre de processeur, famille de l'OS, etc.). Lors d'une sauvegarde manuelle d'une VM, ce sont les fichiers qu'il faut copier.

De par cette configuration, pour une VM donnée en fonctionnement, la RAM et le processeur seront sur un serveur ESX/ESXi (A sur la figure 2.3), alors que le disque dur virtuel et les périphériques **virtuels** (lecteur de disquette/CD, ...) seront sur le disque de stockage iSCSI (B sur la figure 2.3).



Figure 2.3 : stockage de VMs

Il y a 2 types de données à transférer lors d'une migration par vMotion :

- L'état du matériel comme le processeur, la carte mère, les cartes réseau, les adaptateurs de stockages, les lecteurs de disquettes, la carte graphique, les lecteurs cd, etc.
- La RAM de la VM

Le but étant de transférer ces données de la façon la plus transparente possible, en minimisant le temps d'indisponibilité des machines transférées. La méthode de migration actuelle est composée de **3 étapes** :

- A. Opérations : initialisation de la migration en choisissant la VM à migrer et l'hôte de destination
- B. Opérations : **Pré-copie** de la mémoire vive (RAM) de la VM sur l'hôte de destination, la VM continuant de fonctionner. C'est la partie où le plus de données sont échangées.
- C. Opérations : Mise en pause (quiesed state) de la VM puis envoi des données d'état non-RAM, transfert du contrôle de la VM à l'hôte de destination, envoi des données mémoire modifiées et annulation des dépendances sur l'hôte source et remise en marche de la VM. C'est la seule étape où la VM est éteinte, pour une durée dépendant des paramètres indiqués dans le point suivant.

Les serveurs ESX et ESXi sont contrôlés à distance par le Virtual Infrastructure (VI) Serveur. Ce dernier est lui-même contrôlé à distance ou en local par un ou plusieurs VI Client. L'utilisateur interagit donc avec les serveurs ESX/ESXi et les VMs (VMs) par le biais du VI Client. À noter qu'une interface web peut-être utilisée à la place du VI Client, mais certaines fonctionnalités n'ont pas été portées sur cette interface.



2.3 Paramètres

En connaissant le déroulement du transfert d'une VM d'un hôte à un autre, nous pouvons déduire que les paramètres suivants influencent la durée d'indisponibilité :

- La vitesse du réseau sur lequel on effectue le transfert (le minimum recommandé est de 1Gbps pour le réseau de stockage et 100Mbps pour le réseau de management)
- La taille de la RAM occupée ainsi que l'activité de la machine durant le transfert (RAM modifiée durant la pré-copie)
- La fréquence du processeur et la taille de sa mémoire cache

L'activité de la VM est très importante, car elle définit la quantité de données contenues dans la RAM à retransmettre car modifiée depuis la pré-copie. Si la VM ne fait rien, la quantité de données à retransmettre sera quasiment nulle, et donc le temps nécessaire sera très court. En revanche, si la VM était très active, la quantité de données à transmettre sera importante, le temps de transfert sera donc bien plus conséquent.

2.4 Scénario 1 : vMotion entre serveurs ESXi sur un réseau à 100Mbps



Figure 2.5 : schéma réseau du scénario 1, vMotion sur ESXi à 100Mbps

2.4.2 Méthodologie

Pour ce scénario, le réseau utilisé fonctionne à 100Mbps, ce choix permettant de garantir que les PC Wireshark 1 et 2 puissent capturer tous les paquets sans saturation du buffer. De cette façon, l'analyse effectuée sera plus pertinente, et les diverses étapes seront plus distincte sur un diagramme temporel.

- Les Best Practices (<u>http://www.vmware.com/pdf/vc_technical_best.pdf</u> p.12) recommandent de séparer le réseau de gestion (ESXi et VI Serveur/Client) du réseau de stockage (ESXi et disques de stockage), utilisé par vMotion. Cependant, sur les serveurs ESXi, il n'est pas possible d'ajouter une deuxième console de management, alors qu'il serait nécessaire d'en avoir une pour le réseau de management et une autre pour le réseau de stockage. C'est pourquoi ces deux réseaux sont réunis en un pour ce scénario.
- L'observation des échanges côté ESXi est effectuée par une machine physique, Wireshark 1, dont le port de connexion sur le switch est configuré pour le monitoring. Ce choix permet de faire des captures globales des échanges en utilisant un port de monitoring. Le switch possède 24 ports, de marque Cisco modèle Catalyst 2950.
- 3. Les VM sont connectées au même switch physique que les serveurs ESXi, mais dans un autre VLAN
- 4. Un **client Ubuntu** est déplacé via vMotion entre les deux serveurs ESXi. Durant ces transferts, il exécute un script téléchargeant en boucle une page sur un serveur web (voir annexe A.8). La mémoire RAM de cette machine est de 256MB, dont 247MB sont occupés
- 5. Le **serveur** est une autre VM Ubuntu, tournant sur un serveur ESX qui n'est pas impliqué dans l'échange vMotion.
- 6. Sur ce même hôte ESX, une VM Windows XP (Wireshark 2) monitore les échanges de données entre les deux autres VMs.

- 7. Le vSwitch auquel sont connectés le serveur web Ubuntu et la VM Wireshark 2 est configuré en hub, de sorte à rendre la capture des paquets échangés possible. Grâce à Wireshark 2, il est possible de voir quand et durant combien de temps la machine déplacée par vMotion est indisponible.
- 8. Toutes les VMs sont stockées sur le disque iSCSI.

J'ai choisi de virtualiser le serveur Ubuntu et le pc d'observation XP par besoin d'utiliser un hub pour la capture des échanges entre les VMs. Dans ce scénario, j'aurais pu configurer un autre port de monitoring et utiliser des machines physiques, mais pour le scénario suivant, le switch ne sera pas configurable. Ce choix unifie ainsi la manière de prendre les mesures.

Une comparaison du nombre de paquets capturés via Wireshark avec les compteurs de paquets du switch sera également effectuée afin d'être sûr qu'aucun paquet n'a été perdu par Wireshark. Une autre capture comprenant les réseaux applicatif, de gestion et de stockage sera également effectuée afin de voir de façon plus précise les concordances temporelles entre les captures faites précédemment. Pour ce faire, le switch sera configuré de sorte à ce que le PC physique d'observation puisse monitorer les ports des VMs en plus des ports déjà monitorés pour les premières captures.

Pour plus de précisions, une dizaine de migrations par vMotion sera effectuée.

2.4.3 Mesures

La quantité de paquets capturés par Wireshark 1 et 2 correspond à celle que le switch indique avoir fait transiter par les ports de connexion des dites machines. Les chiffres indiqués sont des moyennes des valeurs obtenues dans les captures scénario 1/physique_X, scénario 1/virtuel/virtuel_X et scénario 1/global.pcap disponible sur le DVD. Pour plus d'information, reportez-vous au fichier READ ME du dossier scénario 1.

Lors des mesures, le temps de transfert moyen indiqué par le VI Client est de **70 secondes**, cependant, les relevés effectués avec Wireshark montrent que la quasi-totalité des échanges ont lieu dans un intervalle d'environ **32 secondes** (A, B, C et D sur la figure de la page suivante). Je pense que cette différence constatée entre les temps indiqués vient du fait que les transferts ne commencent pas tout de suite, il faut d'abord que l'ordre soit donné et que les hôtes concernés libèrent les ressources nécessaires à un transfert rapide. De même, à la fin du transfert, le VI Client attend que la VM soit de nouveau complètement opérationnelle pour considérer le transfert terminé. Or comme on peut le voir sur la figure 2.6, le trafic réseau ne reprend pas instantanément à la fin des échanges sur le réseau de gestion et stockage.

Le temps d'inactivité moyen est de **8.8 secondes** (C et D sur la figure de la page suivante), soit environ 12.5% du temps de transfert annoncé par le VI Client.



Figure 2.6 : les différentes étapes d'un transfert par vMotion

La figure ci-dessus est obtenue à partir du fichier scénario 1/global.pcap disponible sur le DVD. Concernant les filtres à appliquer, reportez-vous au fichier READ ME.txt disponible dans le même dossier.

- **A.** Cette partie des échanges a lieu entre le disque iSCSI et les serveurs ESXi. Le volume de données transféré est de **3.3MB** sur une période de **1.5 à 2 secondes**.
- B. C'est là que sont transférés plus de 90% des données de l'échange par vMotion. Durant cette étape, une pré-copie de la RAM est effectuée entre les 2 serveurs ESXi, alors que la VM est encore en fonction. La durée de ces échanges est de 25 secondes pour une quantité de données échangées valant 290.4MB (taille comprenant également le début de la partie C : envoi des changements de la RAM). Cette taille correspond à la RAM occupée sur la VM plus environ 7% de la valeur de la RAM répartit entre le ré-envoi de la RAM (début C), les entêtes des paquets et les paquets servant à la gestion de l'échanges.

Si des écritures devaient avoir lieux dans une zone de la RAM déjà pré-copiée, elles seraient inscrites dans un fichier bmp qui sera envoyé au début de la **partie C**.

- **C.** Durant cette partie de l'échange, la VM est mise en pause. On peut décomposer cette partie en deux sous partie :
 - l'envoi des changements de la RAM, durant **0.5 à 1 secondes** qui est similaire à la **partie B**, tant en vitesse de transmission que dans le type de paquets transmis.
 - l'envoi des données non RAM, qui dure environ 6 secondes pour un envoi de 12.1MB entre le disque iSCSI et le serveur ESXi sur lequel la VM est déplacée.

D. Durant cette étape, la VM est remise en fonctionnement. Dès qu'elle est résumée, elle envoie une requête ARP contenant sa propre MAC address en broadcast afin de faire connaître son nouvel emplacement sur le réseau. On peut constater que cette requête est émise environ 1 seconde avant que les échanges entre la VM et le serveur de test ne reprennent. Je pense donc que cette requête est envoyée pendant que la machine se remet en fonction, mais avant d'être à nouveau totalement fonctionnelle.

On peut également voir 2 pics de données qui ont lieu entre l'ESXi sur lequel tournait la VM au début de l'échange et le disque iSCSI, qui servent surement à libérer définitivement les ressources sur le serveur ESXi en question. Cependant, je n'ai aucune certitude à ce sujet.

Les échanges impliquant le VI Client / Serveur sont peu importants en volume par pics, mais comme ils ont lieu sur toute la durée du transfert, ils représentent un volume de 5.3MB. Ces échanges ont lieu uniquement avec les hôtes ESXi.

Il est intéressant d'observer que la totalité des échanges a lieu en TCP, l'UDP n'offrant, de par sa conception, pas la sécurité requise pour le déplacement d'une VM.



Figure 2.7 : plan du réseau indiquant les flux entre les machines

2.4.3.1 Échanges entre l'hôte source et l'hôte de destination

Ces échanges se déroulent durant la partie B et le début de la partie C de la figure 2.6. Ils consistent en **2 sessions TCP** simples (pas de surcouches protocolaires entre l'application et le TCP) entre les serveurs ESXi. Pour rappel, ces échanges se produisent durant **25 secondes**, pour un volume de données de **290.4MB**.

La session 1 représente 99.99% des paquets échangés. C'est le serveur ESXi qui possède la VM au début de l'échange qui l'initialise, avec un port de départ compris entre 2050 et 5000, à destination du port 8000 de l'autre serveur ESXi. Le port 8000 permet l'accès à la console de service qui permet de contrôler le serveur ESXi cible. Je pense donc que le serveur possédant la VM pilote ainsi celle qui doit en prendre le contrôle pour lui dire ce qu'elle doit télécharger et quand. On peut constater que

la moitié des paquets a une taille comprise entre 40 et 159 Bytes, alors que l'autre moitié se situe entre 640 et 1514 Bytes par paquet. Cela indique qu'il y a autant de paquets servant au contrôle de l'échange qu'à l'échange à proprement parler.

La session 2 est la même, mais dans le sens inverse. Je pense qu'elle permet au serveur de destination d'orienter le serveur ESXi de départ dans les différentes étapes en le tenant informé de l'état de charge dans lequel il se trouve. Je ne me suis cependant pas trop attardé sur cette session, vu le faible pourcentage des transferts qu'elle représente.

2.4.3.2 Échanges entre le disque iSCSI et les hôtes ESXi

Ces échanges sont présents dans les parties A et C de la figure 2.6, soit un volume total de 15.3 MB.

On constate deux paires de partenaires pour ces échanges :

- Le serveur ESXi sur lequel on veut déplacer la VM et le disque iSCSI (92.6% des échanges du iSCSI)
- Le serveur ESXi sur lequel est la VM et le disque iSCSI (7.4% des échanges du iSCSI)

Dans les deux cas, les échanges utilisent la même structure. Le iSCSI est un protocole qui permet de faire des échanges de données/envoi de commande SCSI au travers d'un réseau IP. iSCSI se présente comme une surcouche au protocole TCP.

J'ai pu remarquer **3 types** d'opérations lors des transferts (voir les fichiers Wireshark physique_X et global indiqué au début du point 2.4) :

- Lecture de la capacité du disque
- Lecture de données
- Écriture de données

Pour plus d'informations, veuillez vous référer au chapitre 1.4.

2.4.3.3 Échanges comprenant le VI Serveur / Client

Ces échanges ont lieu tout le long de la transmission de la VM par vMotion et leur volume est de **5.3MB**. Ils ont uniquement lieu avec les serveurs ESXi, principalement avec celui qui héberge la VM au début de l'échange. J'ai pu constater qu'une fois la pré-copie terminée, ils deviennent insignifiants. Comme la console distante était ouverte sur le PC qui héberge le VI Serveur, je présume que la plupart de ces derniers servent à rafraichir l'affichage de la console distante. Le reste des échanges sert à la transmission d'informations comme l'avancement du vMotion, les logs, etc...

Les protocoles utilisés sont **TCP simple** à hauteur de **93.6%** et **SSL version 3** à hauteur de **6.65%** des paquets échangés.

2.5 Scénario 2 : vMotion entre serveurs ESXi sur un réseau à 1Gbps



2.5.1 Schémas et plan du réseau

Figure 2.8 : schéma réseau du scénario 2, vMotion sur ESXi à 1Gbps

2.5.2 Méthodologie

- L'étape précédente a permis d'observer en détail tous les échanges ayant lieu durant un transfert vMotion, autant du côté de la VM déplacée, que du côté des hôtes.
- Cependant, la vitesse du réseau était inférieure au minimum recommandé, c'est pourquoi ce scénario a été mis en place.
- Les switch gigabit utilisés n'étant pas configurables, un monitoring du côté des serveurs ESXi est exclu, mais reste possible du côté de la VM. En effet, la machine faisant tourner Wireshark 2 pour le réseau applicatif étant virtuelle et sur le même vSwitch que le serveur recevant les requêtes de la machine déplacée, une fois ce vSwitch configuré en hub, elle peut monitorer les paquets échangés entre ces deux autres VMs.
- Ces captures permettront d'établir le temps d'indisponibilité, comme à l'étape précédente.

La méthodologie sera donc la même que lors du scénario 1, mais sans la partie monitoring du réseau de gestion et stockage. Le but de ce scénario est donc limité à la mesure du temps d'indisponibilité de la VM.

Pour ces captures, la charge moyenne de la **RAM** de la VM déplacée est de **245MB** sur les 256MB disponibles.

2.5.3 Mesures

Comme on peut le voir sur la figure 2.9, les temps d'indisponibilité sont bien au dessus des 1 à 2 secondes annoncées par VMware. Ce temps s'explique par les performances médiocres du serveur Openfiler sur un réseau gigabit (plus de 12 minutes pour un fichier de 3GB, voir annexe A.7 : tests du switch NETGEAR), et par le bridage des performances de vMotion sur les serveurs ESXi (voir scénario 3, qui donne de meilleurs résultats en étant sur un réseau à 100Mbps).

Test n°	1	2	3	4	5	6	7	8	9	10	moyenne
Temps total [s]	37	43	36	67	35	66	35	66	37	67	48.9
Temps d'indisponibilité [s]	6.157	9.456	6.338	9.173	6.265	6.239	9.171	6.042	9.459	9.307	7.7607

Figure 2.9 : temps de transfert et d'indisponibilité

2.6 Scénario 3 : vMotion entre serveurs ESX sur un réseau à 100Mbps



2.6.1 Schémas et plan du réseau

Figure 2.10 : schéma réseau du scénario 3, vMotion sur ESX à 100Mbps

2.6.2 Méthodologie

La méthodologie de ce scénario sera très similaire à celle du scénario 1, car le but principal est de voir si les échanges se déroulent de la même façon entre des serveurs ESX qu'entre des serveurs ESXi.

- Les Best Practices recommandent de séparer le réseau de gestion (ESX et VI Serveur/Client) du réseau de stockage (ESX et disques de stockage), utilisé par vMotion. Contrairement aux serveurs ESXi, les serveurs ESX permettent la création d'une seconde console de service, et donc la séparation de ces deux réseaux. Ces réseau seront sur le même switch, mais dans un VLAN différent.
- L'observation des échanges côté ESX est effectuée par une machine physique, Wireshark 1, dont le port de connexion sur le switch est configuré pour le monitoring. Cette machine effectuera l'observation aussi bien des réseaux de gestion et de stockage.
- Les VM sont connectées au même switch physique que les serveurs ESX, mais dans un autre VLAN.
- 4. Un client Ubuntu stocké sur le disque iSCSI (Openfiler) est déplacé via vMotion entre les deux serveurs ESX. Durant ces transferts, il exécute un script (voir annexes) téléchargeant en boucle une page sur un serveur web. La mémoire RAM de cette machine est de 256MB, dont 245MB sont occupé.
- 5. Le **serveur** est une autre VM Ubuntu, tournant sur un serveur ESXi qui n'est pas impliqué dans l'échange vMotion et stocké en local.
- 6. Sur ce même hôte ESXi, une VM Windows XP (Wireshark 2) monitore les échanges de données entre les deux autre VMs.

- 7. Le vSwitch auquel son connectés le serveur web Ubuntu et la VM Wireshark 2 est configuré en hub, de sorte à rendre la capture des paquets échangés possible. Grace à Wireshark 2, il est possible de voir quand et durant combien de temps la machine déplacée par vMotion est indisponible.
- 8. Contrairement au serveur ESXi, les serveurs ESX offrent la possibilité d'accéder à des logs plus nombreux et plus complets. Parmi ces logs, 1 contient des informations temporelles intéressantes à propos de vMotion : /var/log/vmkernel. Je me servirai donc de ces logs pour les comparer aux relevés de Wireshark 1.

Une comparaison du nombre de paquets capturés via Wireshark avec les compteurs de paquets du switch sera également effectuée afin d'être sûr qu'aucun paquet n'a été perdu par Wireshark. Une autre capture comprenant tous les réseaux sera à nouveau effectuée afin de voir de façon plus précise les concordances temporelles entre les captures faites précédemment. Pour ce faire, le switch sera configuré de sorte à ce que le PC physique d'observation puisse monitorer les ports des VMs en plus des ports déjà monitorés pour les premières captures.

2.6.3 Mesures

La quantité de paquets capturés par Wireshark 1 et 2 correspond à celle que le switch indique avoir fait transiter par les ports de connexion des dites machines. Les chiffres indiqués sont des moyennes des valeurs obtenues dans les captures scénario 3/physique_X et scénario 3/virtuel/virtuel_X disponibles sur le DVD. Pour plus d'information, reportez-vous au fichier READ ME du dossier scénario 3.

Lors des mesures, le temps de transfert moyen indiqué par le VI Client est de **53 secondes**, cependant, les relevés effectués avec Wireshark montrent que la quasi-totalité des échanges ont lieu dans un intervalle d'environ **27 secondes**.

Le temps d'inactivité moyen est de **3.3 secondes**, soit environ **6%** du temps de transfert annoncé par le VI Client.
À partir des logs (disponible sur le DVD : scénario 3/log/vmkernel 15X.txt) on peut calculer **28.548 secondes** entre la première et la dernière entrée (log des 2 serveurs confondu). Le temps d'interruption est de **1.744 secondes**, soit environ la moitié du temps observé à l'aide de Wireshark. Cette différence est due au fait que les logs indiquent la fin de la pré-copie et la remise en marche de la VM, ce qui ne comprend pas le temps de remise en fonction des connections réseau. Voici les logs pour un transfert :

Log de la machine de départ (10.1.2.151)
Dun 10 13:01:05 G3 vmkernel: 0:22:49:27.154 cou0:1139)Migrate: vm 1140: 7383: Setting migration info
ts = 1244631644672670, src ip = <10.1.2.151> dest ip = <10.1.2.152> Dest wid = 1130 using SHARED swap Jun 10 13:01:05 G3 vmkernel: 0:22:49:27.154 cpu0:1139)world: vm 1147: 901: Starting world migSendHelper-1140 with flags 1
Jun 10 13:01:05 G3 vmkernel: 0:22:49:27.154 cpu0:1139)World: vm 1148: 901: Starting world migRecvHelper-1140
Jun 10 13:01:05 G3 vmkernel: 0:22:49:27.155 cpu0:1077)MigrateNet: vm 1077: 854: Accepted connection from <10.1.2.152>
Jun 10 13:01:28 G3 vmkernel: 0:22:49:50.609 cpu1:1140)Migrate: 7296: 1244631644672670: Stopping pre-copy: only 977 pages were modified
Jun 10 13:01:28 G3 vmkernel: 0:22:49:50.621 cpu0:1139)DevF5: 2300: Unable to find device: 25010-Serveur1 1-000001-delta.vmdk
Jun 10 13:01:28 G3 vmkernel: 0:22:49:50.624 cpu0:1139)DevFS: 2300: Unable to find device: 1000d-Serveur1-000001-delta.vmdk
l og de la machine de destination (10.1.2.151)
<u>Jun 10 13:01:04</u> G4 vmkernel: <u>[0:04:10:59.457]</u> cpu0:1034)world: vm 1129: 901: Starting world vmware-vmx with flags 4 Jun 10 13:01:05 G4 vmkernel: 0:04:10:59.959 cpu1:1129)world: vm 1130: 901: Starting world vmm0:Serveur153 with flags 8
Jun 10 13:01:05 G4 vmkernel: 0:04:10:59.959 cpu1:1129)Sched: vm 1130: 5333: adding 'vmm0:Serveur153': oroup 'bost/user': cpu: sbares=-3 min=0 min!imit=-1 max=-1
Jun 10 13:01:05 G4 vmkernel: 0:04:10:59.959 cpu1:1129)Sched: vm 1130: 5352: renamed group 33 to vm.1129
Jun 10 13:01:05 G4 vmkerne]: 0:04:10:59.959 cpul:1129)sched: vm 1130: 5366: moved group 33 to be under group 4
Jun 10 13:01:05 G4 vmkernel: 0:04:11:00.247 coul:1129)Maprate: vm 1130: 2109: extending swap to 26:144 KB
ts = 1244631644672670, src ip = <10.1.2.151> dest ip = <0.0.0.0> Dest wid = -1 using SHARED swap
Jun 10 13:01:05 G4 vmkernel: 0:04:11:00.247 cpu1:1129)world: vm 1131: 901: Starting world migSendHelper-1130
Jun 10 13:01:05 G4 vmkernel: 0:04:11:00.247 cpu1:1129)world: vm 1132: 901: starting world migRecvHelper-1130 with flags 1
Jun 10 13:01:05 Ğ4 vmkernel: 0:04:11:00.366 cpu1:1071)MigrateNet: vm 1071: 854: Accepted connection from <10.1.2.151>
Jun 10 13:01:29 G4 vmkernel: 0:04:11:24.033 cpu0:1131)A]loc: vm 1130: 15678: Regular swap file bitmap checks out.
Jun 10 13:01:29 G4 vmkernel: 0:04:11:24.426 cpul:1132)Migrate: 3978: 1244631644672670: DONE paging in, status 0
Jun 10 13:01:30 GA VMMKernel: 0:04:11:23.323 cpu0:1123/VSCS1: 4060; creating virtual bevice for world 1130 vScS10:0 (handle 8204)
Jun 10 13:01:30 G4 vmkernel: 0:04:11:25.323 cpu0:1129)V5C5I: 4060: Creating Virtual Device for world 1130 vscsi0:1 (bandle 8205)
Jun 10 13:01:30 G4 vmkernel: 0:04:11:25.383 cpu1:1129)Migrate: 6441: 1244631644672670: No page at vpn 0x110 offset: 1114112 size: 16384 bvtesLeft: 1332
Jun 10 13:01:30 G4 vmkernel: 0:04:11:25.446 cpu1:1129)World: vm 1133: 901: Starting world vmware-vmx with flags 44
Jun 10 13:01:30 G4 vmkernel: 0:04:11:25.446 cpu0:1133)World: vm 1134: 901: starting world vmware-vmx with flags 44
Jun 10 13:01:30 G4 vmkernel: 0:04:11:25.446 cpu0:1133/World: vm 1135: 901: Starting world vmware-vmx with flags 44
Jun 10 13:01:30 G4 vinkernel: 0:04:11:25.483 cpu1:1130/Migrate: 7601: 1244631644672670: Waiting to receive all
changed pages before resuming
Jun 10 13:01:30 G4 vmkernel: 0:04:11:25.483 cpu1:1130)woglate: 7622: 1244631644672670: Received all changed pages. Jun 10 13:01:20 G4 vmkernel: 0:04:11:25.483 cpu1:1130)woglate: 7622: 1244631644672670: Received all changed pages
flags 1 Jun 10 13:01:50 G4 Vinkernel: 0:04:11:25:484 cpu1:1126)Gupp: vm 1120: 7020: Starting eppertupictic cupper for the
migration swap file
Dun 10 13:01:30 G4 vmkernel: 0:04:11:25.484 cpu1:1130)Migrate: 3871: 1244631644672670: Resume handshake successful
Jun 10 13:01:30 G4 vmkernel: 0:04:11:25.489 cpu0:1130/0prink: 2491; setting capabilities 0x20 for device vmnici Jun 10 13:01:30 G4 vmkernel: 0:04:11:25.551 cpu0:1136/Swap: vm 1130: 8133: Finish swapping in migration swap file. (faulted in 0 pages). Success.
Référence temporelle du VI Client Fin de la pré-copie / début de l'interruption
Références temporelles de serveurs ESX Remise en marche de la VM
Ligne se produisant au même moment sur les 2 serveurs

Figure 2.11 : extrait des logs du VMkernel

Comme on peut le voir, les références de temps du VI Client ne sont pas assez précises pour être intéressantes, ce qui n'est pas le cas des références temporelles des serveurs ESX. Malheureusement, ces temps ne sont pas synchronisés. Le décalage de temps peut-être calculé à partir de l'entrée qui indique les paramètres de migration, et apparaissant dans les 2 logs au même moment.

Les entrées les plus intéressantes sont :

- Celle qui indique la fin de la pré-copie
- Celle qui indique la remise en marche de la VM

Ces lignes donnent une indication sur le temps d'indisponibilité de la VM indiqué plus haut.

La ligne indiquant la fin de la pré-copie donne également le nombre de pages de RAM modifiées et donc à retransmettre. Si on connait la taille des pages, cette information nous permet de calculer le volume de données à retransmettre.



Figure 2.12 : transfert par vMotion

La figure ci-dessus est obtenue à partir du fichier scénario 3/global.pcap disponible sur le DVD. Concernant les filtres à appliquer, reportez-vous au fichier READ ME.txt disponible dans le même dossier.

- A. Cette partie des échanges a lieu entre le disque iSCSI et les serveurs ESXi. Le volume de données transféré est de **1.8MB** sur une période de **1.5 à 2 secondes**.
- B. La pré-copie dure **24 secondes** pour une quantité de données échangées valant **426.5MB** (taille comprenant également le début de la **partie C** : envoi des changements de la RAM).
- C. Durant cette partie de l'échange, la VM est mise en pause. On peut décomposer cette partie en deux sous partie :
 - l'envoi des changements de la RAM, durant **0.5 à 1 secondes.** qui est similaire à la **partie B**, tant en vitesse de transmission que dans le type de paquets transmis.
 - l'envoi des données non RAM, qui dure environ **1 seconde** pour un envoi de **6.8MB** entre le disque iSCSI et le serveur ESXi sur lequel la VM est déplacée.

Les échanges impliquant le VI Client / Serveur représentent un volume de 6.5MB.

Aucun changement concernant la nature des flux par rapport au scénario 1. En revanche, le volume des flux à bien changé, pour une raison inconnue.



Figure 2.13 : flux de transfert

Le déplacement s'effectue à un débit bien plus élevé entre les serveurs ESX qu'entre les serveurs ESX (scénario 1). La figure ci-dessous montre les 2 types de transferts sur une même échelle de temps, mais avec des Bytes au lieu des paquets sur l'axe vertical :



Figure 2.14 : comparaison de l'évolution temporelle d'un vMotion entre ESX ou ESXi

La trace noire est extraite de la figure 2.12 (trace verte plus trace rouge) ce qui correspond à un échange entre serveur ESX et la trace rouge est extraite de la figure 2.6 (trace rouge), ce qui correspond à un échange entre serveur ESXi.

Bien que le débit soit plus élevé entre les ESX, le seul moment où l'on voit clairement un gain de temps correspond au point C et D de la figure 2.12.

Les 2 captures ayant été effectuées sur un réseau fonctionnant à 100Mbps, j'en conclus que les performances de la fonction vMotion sont **bridées en débit** sur les **serveurs ESXi**.

Cependant, la **différence de données échangées** durant la pré-copie est **très importante** (426MB sur les ESX et 290MB sur les ESXi). Aucune raison ne semble justifier une telle différence.

2.7 Scénario 4 : vMotion entre ESX sur un réseau à 1Gbps



2.7.1 Schémas et plan du réseau

Figure 2.15 : schéma réseau du scénario 1, vMotion sur ESX à 1Gbps

2.7.2 Méthodologie

- L'étape précédente a permis d'observer en détail tous les échanges ayant lieu durant un transfert vMotion, autant du côté de la VM déplacée, que du côté des hôtes.
- Cependant, la vitesse du réseau était inférieure au minimum recommandé, c'est pourquoi ce scénario a été mis en place.
- Le switch gigabit utilisé n'étant pas configurable, un monitoring du côté des serveurs ESXi est exclu, mais reste possible du côté de la VM. En effet, la machine faisant tourner Wireshark 2 pour le réseau applicatif étant virtuelle et sur le même vSwitch que le serveur recevant les requêtes de la machine déplacée, une fois ce vSwitch configuré en hub, elle peut monitorer les paquets échangés entre ces deux autres VMs.
- Ces captures permettront d'établir le temps d'indisponibilité, comme à l'étape précédente.

La méthodologie sera donc la même que lors du scénario 1, mais sans la partie monitoring du réseau de gestion et stockage. Le but de ce scénario est donc limité à la mesure du temps d'indisponibilité de la VM.

Pour ces captures, la charge moyenne de la RAM de la VM déplacée est de **245MB** sur les 256MB disponibles.

2.7.3 Mesures

Comme lors du scénario 2, on peut constater que les temps d'indisponibilité dépassent les 2 secondes indiquées par VMware. Cette fois, je pense que ce dépassement est dû uniquement à Openfiler qui ne gère pas bien le gigabit.

Test n°	1	2	3	4	5	6	7	8	9	10	moyenne
Temps total [s]	18	19	17	17	19	18	18	16	18	17	17.7
temps d'indisponibilité [s]	3.745	2.885	1.552	4.285	4.277	4.296	2.817	4.365	2.805	1.597	3.3524
Figure 2.16 : temps de transfert et d'indisponibilité											

L'utilisation d'un vrai disque iSCSI donnera sûrement des temps inférieurs à 2 secondes, je n'ai malheureusement pas de tel équipement pour effectuer des tests.

2.8 Conclusion

Comme on a pu le constater, les performances annoncées ne sont pas atteinte, mais ceci est plutôt imputable au matériel utilisé. Pour un environnement de production un SAN physique est souhaitable en lieu et place de solution comme Openfiler, qui peuvent satisfaire un environnement de test peu regardant sur les performances du disque iSCSI.

L'utilisation du VI Server (qui est payante) est obligatoire pour mettre en œuvre cette fonctionnalité, et celle des serveurs ESX (payant eux aussi) est souhaitable à celle de serveurs ESXi (qui sont gratuits), et ce pour des raisons de débits non bridé, ce qui réduit le temps d'indisponibilité.

Cette fonctionnalité est prometteuse, reste à savoir si VMware l'améliorera dans le futur pour atteindre un temps d'indisponibilité le plus petit possible, voir même nul.

3 Distributed Resource Scheduler

3.1 Introduction

Distributed Resource Scheduler, ou DRS dans la suite de ce document, est une fonctionnalité permettant la gestion automatique de la répartition de la charge entre les serveurs ESX/ESXi faisant partie d'un même cluster (groupe de serveurs ESX/ESXi mettant leurs ressources en commun, voir chapitre suivant). DRS utilise vMotion pour déplacer les VMs

Selon VMware (<u>http://www.vmware.com/files/pdf/drs_datasheet.pdf</u>), VMware DRS dynamically balances computing capacity across a collection of hardware resources aggregated into logical resource pools, continuously monitoring utilization across resource pools and intelligently allocating available resources among the virtual machines based on pre-defined rules that reflect business needs and changing priorities. When a virtual machine experiences an increased load, VMware DRS automatically allocates additional resources by redistributing virtual machines among the physical servers in the resource pool. VMware DRS allows IT organizations to:

- Prioritize resources to the highest value applications in order to align resources with business goals
- Optimize hardware utilization automatically and continuously to respond to changing conditions

• Provide dedicated resources to business units while still profiting from higher hardware utilization through resource pooling

• Conduct zero-downtime server maintenance

DRS est une fonctionnalité très intéressante si vous virtualisez des serveurs dont la charge est amenée à changer régulièrement, par exemple, un serveur de VOIP (pique de charge entre 10h et 11h le matin) ou un site web très sollicité à certaines heures (p.ex : en fin de journée) et peu ou pas du tout à d'autres (p.ex : la nuit).

3.2 Cluster

Un cluster est un groupe de serveurs physiquement distincts qui mettent leurs ressources en commun. Un cluster sera perçu comme étant une seule entité physique ayant les capacités de tous les serveurs la composant. Le principal avantage est qu'en cas de panne d'un serveur, le cluster continue à fonctionner. La charge processeur et RAM est répartie de façon équitable entre les serveurs.

Dans le cas de VMware, un cluster est composé d'un groupe de serveurs ESX et/ou ESXi ainsi que d'un groupe de VMs. Les VMs sont hébergées par un serveur quelconque du cluster et non par un serveur spécifique. Le but est de répartir la charge générée par les VMs le plus équitablement possible entre les différents serveurs ESX/ESXi à l'aide de DRS.

Il est recommandé (<u>http://www.kingston.com/branded/pdf_files/HAnDRSCapacity_WP.pdf</u>) que ces serveurs aient une configuration matérielle identique afin d'éviter qu'à charge égale un serveur soit à 90% de ses capacité et un autre à 50%.



Figure 3.1 : cluster de serveurs ESX vu depuis le VI Client

3.3 Fonctionnement

Comme indiqué dans l'introduction, l'utilisation de DRS nécessite la mise en place de cluster (voir annexe A.6.8).

La suite de ce chapitre sera basée sur le document http://www.kingston.com/branded/pdf_files/HAnDRSCapacity_WP.pdf.

Le VI Server vérifie en continu la charge CPU et RAM de chaque serveur faisant partie d'un cluster utilisant DRS. Selon les paramètres définis (voir chapitre suivant), si un serveur se retrouve avec une charge trop importante (un trop grand nombre de VMs ou une VMs qui consomme trop de ressources), DRS déplacera automatiquement ou suggèrera de déplacer manuellement des VMs vers un autre serveur moins chargé en utilisant vMotion.

Il est recommandé d'utiliser du matériel le plus homogène possible, voire même identique.

3.3.1 Réservations

La partie 2 du cours VMware de M. Litzistorf (<u>http://www.tdeig.ch/vmware/Plaquette.pdf</u>) explique la possibilité de réserver des ressources CPU et RAM pour une VM, ainsi que celle de limiter l'accès à ces mêmes ressources.

Par défaut, les ressources sont partagées entre les VMs selon leurs besoins respectifs. Si ces ressources ne sont pas suffisamment nombreuses, chaque VM aura le droit à une part égale. Cette situation n'est pas acceptable pour certaines VMs, c'est pourquoi il est possible de leur attribuer manuellement une part des ressources.

De même, une VM utilisant un système d'attente active consommera bien plus de temps processeur que celui consommé par les besoins réels de la même VM. Il est donc utile de limiter les ressources disponibles pour cette VM de sorte à les conserver pour d'autres VMs qui en auraient vraiment l'utilité.

Selon le document cité au début du chapitre 3.3, il est également possible de réserver des ressources pour un groupe de VMs, par exemple, celles qui sont contenues dans un même cluster DRS. Il faut savoir qu'un serveur ESX/ESXi peut faire partie d'un cluster et également avoir d'autre VMs/groupes de VMs qui ne sont pas inclues dans le cluster.

Cette fonctionnalité ne sera pas exploitée dans la suite de ce document, car les serveurs ESX ont été entièrement intégrés au cluster, mais il peut s'avérer utile, selon votre configuration, de savoir que cette possibilité existe.

3.4 Paramètres

Comme DRS utilise vMotion, les paramètres déterminant le temps d'indisponibilité d'une VM lors d'un déplacement sont identiques à ceux indiqués dans le paragraphe 2.3.

La configuration de DRS permet de régler différents paramètres tels que :

- Le mode de déplacement (manuel, semi-automatique ou automatique) global (pour toutes les VMs) et individuel (pour une VM spécifique, attribuer un mode de déplacement différent de celui défini globalement)
- La sensibilité : rééquilibrage rapide de la charge entre les serveurs ou attendre un délai
- Des règles sur la disposition des VM, afin que les VMs désirées soient toujours sur le même serveur ESX/ESXi, ou au contraire, toujours sur un serveur différent
- Les options avancées, pour lesquels je n'ai malheureusement trouvé aucune documentation

Une partie de ces paramètres a pu être réglée à la création du cluster, mais selon les scénarios qui suivront, nous verrons comment modifier ces règles.

Remarque : Il est indiqué dans à la page 46 dans le document

https://140.110.8.180/king/modules/userupload/upload/ets2007/vmware.13.pdf que DRS fonctionne mieux avec des petites VMs (faible consommation RAM et CPU) qu'avec des grosses VM. En effet, un déplacement d'une VM générant une grosse charge CPU et RAM aura d'une part un temps d'indisponibilité plus long, mais laissera également derrière lui une quantité importante de ressources libres, ce qui pourrait créer un nouveau déséquilibre de charge avec un autre serveur ESX/ESXi non impliqué dans le premier déplacement. Le déplacement en lui-même provoque également une consommation supplémentaire de ressources, réduisant ainsi les ressources disponibles pour les VMs ne se déplaçant pas, ce qui gêne le fonctionnement général. En opposition, une petite VM aura un déplacement rapide, perturbant ainsi moins longtemps le fonctionnement général, et produira une différence de charge bien moins importante sur les serveurs ESX/ESXi suite à son déplacement, réduisant ainsi la probabilité de devoir à nouveau déplacer une VM.

3.5 Topologie et plans du réseau

Contrairement aux scénarios de test de vMotion qui nécessitaient des changements de topologie du réseau, les scénarios de test de DRS seront tous effectués avec la même topologie réseau. Pour des raisons de performances, j'ai choisi d'utiliser un réseau de stockage à 1Gbps ne comprenant que des serveurs ESX. La topologie sera donc très proche de celle employée lors du scénario 4.



Figure 3.2 : plan du réseau utilisé pour les tests de DRS

- Conformément aux best practices (<u>http://www.vmware.com/pdf/vc_technical_best.pdf</u>), les réseaux de stockage et de gestion sont séparés. Le réseau de stockage est sur un switch NETGEAR à 1Gbps et le réseau de gestion est sur le VLAN 1 du switch Cisco (switch 1).
- 2. Comme indiqué plus haut, seuls les serveurs ESX seront utilisés.
- 3. Les VMs possèdent leur réseau séparé qui se trouve sur le VLAN 2 du switch Cisco.
- 4. Les VMs seront réparties à l'aide de DRS entre les 2 serveurs ESX. Plus de détails seront donnés concernant la disposition des VMs dans chaque scénario.
- 5. Comme pour le scénario 1, toutes les VMs sont stockées sur le disque iSCSI émulé par le serveur Openfiler.
- 6. Le tout étant toujours géré par le VI Serveur/Client.

3.6 Scénario 5 : déplacement manuel

3.6.1 Configuration testée et méthodologie

Pour commencer, afin de pouvoir mieux voir à quel moment ont lieu les déplacements, le mode utilisé sera manuel. Aucune autre règle ne sera mise en place pour ce scénario.

3 VMs Windows XP seront utilisées (la VM Wireshark 2, des scénarios précédant et 2 clones de cette dernière). Mon choix s'est porté sur ces VMs car Windows XP a une charge CPU et RAM suffisamment importante pour déclencher un déplacement même au repos (contrairement aux VMs Ubuntu utilisées précédemment, qui ont une charge au repos presque nulle car moins de services sont activés au démarrage), tout en restant considérable comme des petites VMs.

action	Serveur ESX 1	Serveur ESX 2	figure
mise en route de la 1e VM			
dialogue de demande du serveur de destination,			27
aucune préférence indiquée			3.7
choix effectué	VM1		
mise en route de la 2e VM	VM1		
dialogue de demande du serveur de destination,	V/N 4.1		2.0
préférence indiquée pour le serveur 2	VIVIT		5.8
choix effectué en suivant les recommandations	VM1	VM2	
mise en route de la 3e VM	VM1	VM2	
dialogue de demande du serveur de destination,	V/N 4 1		2.0
aucune préférence indiquée	VIVII	VIVIZ	3.9
choix effectué	VM1 + VM3	VM2	3.10 et
		01012	3.11
arrêt de la 2e VM	VM1 + VM3		
proposition de déplacement de la 3e VM	VM1 + VM3		3.13
déplacement accepté	VM1	VM3	3.14

Les étapes suivantes seront effectuées :

Figure 3.3 : tableau des opérations à effectué lors du scénario 5

3.6.2 Mise en œuvre de la configuration de DRS

Cette configuration n'étant pas celle par défaut (configurée dans l'annexe A.6.8), il va falloir éditer les paramètres. Pour ce faire, faites un clic droit sur le cluster dans le panneau de gauche du VI Client, puis cliquez sur *Edit Settings...*



Figure 3.4 : ouverture de la fenêtre d'édition des paramètres du cluster DRS

Dans le menu à gauche de la fenêtre qui vient de s'ouvrir, cliquez sur VMware DRS.



Figure 3.5 : écran principal de la fenêtre d'édition des paramètres

Sur ce nouvel écran, sélectionnez « Manual ». Cliquez sur *OK*, les changements de configuration vont s'effectuer.

DRS Settings	
General VMware DRS Rules Virtual Machine Options Power Management VMware EVC Swapfile Location	Automation Level Manual WirtualCenter will suggest migration recommendations for virtual machines. Partially automated WirtualCenter will be automatically placed onto hosts at power on and VirtualCenter will suggest migration recommendations for virtual machines. Fully automated Wirtual machines will be automatically placed onto hosts when powered on, and will be automatically migrated from one host to another to optimize resource usage. Migration threshold: Conservative Aggressive Apply recommendations with three or more stars. VirtualCenter will apply recommendations that promise at least good improvement to the cluster's load balance.
	Advanced Options
Help	OK Cancel

Figure 3.6 : choix du mode de fonctionnement et de la sensibilité de DRS

3.6.3 Observations et mesures

Lors du lancement des VMs, le choix du serveur ESX a bien été laissé à l'utilisateur. Un système d'étoiles (de 1 à 5 étoiles) indique à l'utilisateur quel serveur est le plus optimal pour héberger la VM (plus il y a d'étoiles, plus le serveur est recommandé).

Au lancement de la première VM, aucun serveur n'est occupé, tous ont donc le même nombre d'étoiles, à savoir 1. Mon choix s'est porté sur le serveur n°2 (10.1.2.154), sélectionné par défaut.

Remarque : Observateur2 est le nom de la VM Wireshark 2 dans le VI Client.

🖉 Host Recommendations for Observateur 2								
VirtualCen ready for p	irtualCenter recommends the following host(s) for your virtual machine. Selecting a host may select other actions that VirtualCenter must take to make that host eady for power-on. Select a host from the list below by clicking on its action group.							
Priority	Recommendation	Reason						
*	Place Observateur2 at host 10.1.2.153	Power on virtual machine						
*	Place Observateur2 at host 10.1.2.154	Power on virtual machine						
1								
		Power On Cancel Help						

Figure 3.7 : lancement de la première VM, choix du serveur ESX

Au lancement de la seconde VM, comme le serveur 2 héberge déjà une VM, le nombre d'étoiles n'est plus le même entre les 2 serveurs : 3 étoiles pour le serveur 2, alors que le serveur 1 (10.1.2.153) en a 5. Mon choix se porte donc sur le serveur 1, qui est recommandé et sélectionné par défaut.

🖉 Host F	🖉 Host Recommendations for XP1							
VirtualCento ready for p	er recommends the following host(s) for your virtu ower-on. Select a host from the list below by click	al machine. Selecting a host may select other actions that VirtualCenter must take to make that host ng on its action group.						
Priority	Recommendation	Reason						
***	Place XP1 at host 10.1.2.153	Power on virtual machine						
*	Place XP1 at host 10.1.2.154	Power on virtual machine						
1								
		Power On Cancel Help						

Figure 3.8 : lancement de la seconde VM, choix du serveur ESX

Au lancement de la troisième VM, comme les 2 serveurs ESX ont une charge égale, chacun se voit attribuer une étoile, le choix du serveur a donc peu d'importance, je laisse donc le serveur 1 comme proposé par défaut.

Host R VirtualCente ready for po	ecommendations for XP2 r recommends the following host(s) for your virtu wer-on. Select a host from the list below by click	al machine. Selecting a host may select other actions that VirtualCenter must take to make that host ig on its action group.
Priority	Recommendation	Reason
*	Place XP2 at host <u>10.1.2.154</u>	Power on virtual machine
*	Place XP2 at host <u>10.1.2.153</u>	Power on virtual machine
1		
		Power On Cancel Help

Figure 3.9 : lancement de la troisième VM, choix du serveur ESX

Une fois que les 3 VMs ont démarré, il est possible en allant dans l'onglet « Virtual Machines » de chaque serveur ESX d'observer quelles sont les VMs en fonctionnement sur ce serveur ainsi que leur occupation du processeur et de la RAM.

🖃 📂 Hosts & Clusters	10.1.2.154 ¥Mware ESX Server, 3.	5.0, 153875 Evaluat	ion (55 day(s) remaining)						
🖃 🏬 New Datacenter										
🖃 🏨 DRS	Getting Started Summary Virtual	Machines Performan	ce 🔪 Configui	ration 🥄 Tasks & Eve	ents 🔪 Alarms 🔪 P	ermissions Maps				
10.1.2.153										
10.1.2.154										
Client2	Name	State	Status	Host CPU - MHz	Host Mem - MB	Guest Mem - % Note	es			
🔂 Nostalgia1	🚯 Observateur2	Powered On	000	63	187	75				
Mostalgia2	👜 Nostalgia1	Powered Off	000	0	0	0 Nost	talç			
Mostalgia3	👜 Nostalgia2	Powered Off	000	0	0	0 Nost	calç			
Mashalaia5	🖆 Serveur153	Powered Off	000	0	0	0				
Discaligida	💼 Nostalgia4	Powered Off	000	0	0	0 Nost	calç			
Serveur153	💼 Nostalgia5	Powered Off	000	0	0	0 Nost	calç			
A XP1										
	1									

Figure 3.10 : vue sur les VMs du serveur 2

Hosts & Clusters New Datacenter Nes Nors 10.1.2.153 10.1.2.154	10.1.2.153 VMware ESX Se Getting Started Summary	rver, 3.5.0, 153875 Virtual Machines	Evaluation (47 day Performance Configu	(s) remaining) Iration Tasks & Eve	ents Alarms F	Permissions Maps	
🔂 Client2	Name	State	Status	Host CPU - MHz	Host Mem - MB	Guest Mem - %	Notes
Nostalgia1 Nostalgia2 Nostalgia3 Nostalgia4 Nostalgia5 Nostalgia5 Serveur153 Serveur153 Maximum XP1 Maximum XP2	Client2 XP1 XP2 Nostalgia3	Powe Powe Powe	rred Off OO	0 583 719 0	0 1058 172 0	0	Nostalgia

Figure 3.11 : vue sur les VMs du serveur 2

En allant sur l'onglet « DRS Recommandations » du cluster, on peut voir les déplacements suggérés. Comme on vient de démarrer les 3 VMs, leur charge est égale, aucun déplacement n'est proposé.

 Hosts & Clusters Image: New Datacenter Image: DRS 	DRS Getting Starled Summary Virtual Machines Hotte DRS Recommendations Percource Allocation Performance Tasks & Eventil Alarma	Permetion Mape	
10.1.2.153	DRS Recommendations:		
10.1.2.154	Priority Recommendation	Reason	Apply
P Client2	No DRS recommendations at this time.		
Nostalgia2			
Nostalgias			
Nostalgia5			
Observateur2			
Serveur153			
10 XP1			
UD 052	I		
	Override suggested DRS recommendations	Generate Recommendations	Apply Recommendations

Figure 3.12 : déplacements suggérés par DRS

On éteint la VM qui est seule sur le serveur ESX n°2. Après un instant (ou un clic sur le bouton « Generate Recommandations »), il nous est recommandé avec 2 étoiles de déplacer une des VMs restante sur l'autre serveur ESX. Ce déplacement est accepté en cliquant sur *Apply Recommandations*.

DRS Reco Priority	mmendations: Recommendation	Reason	Apply
**	Migrate XP2 from 10.1.2.153 to 10.1.2.154	Balance average memory loads	M
Overri	de suggested DRS recommendations	Generate Recommendations	Apply Recommendations

Figure 3.13 : proposition de déplacement d'une VM par DRS

Comme on l'a vu lors des scénarios précédents, il est ensuite possible de suivre la progression du déplacement par vMotion sur la partie basse de la fenêtre du VI Client :

Recent Tasks								
Name	Target	Status	Initiated by	Time Time	Start Time	Complete Time		
Apply Recommendation	DRS DRS	S Completed	Administrator	19.06.2009 14:01:44	19.06.2009 14:01:44	19.06.2009 14:01:44		
🖄 Migrate Virtual Machine	🖆 XP2	10%		19.06.2009 14:01:44	19.06.2009 14:01:44			
Refresh Recommendations	DRS DRS	Completed	Administrator	19.06.2009 14:01:17	19.06.2009 14:01:17	19.06.2009 14:01:17		

Figure 3.14 : déplacement de la VM par vMotion

3.7 Scénario 6 : déplacement automatique, sensibilité maximum

3.7.1 Configuration testée et méthodologie

Cette fois, il s'agit d'automatiser les migrations. Le mode sera donc automatique, avec la sensibilité maximum, ce qui aura pour effet de déplacer les VMs là ou le mode manuel se contentait d'afficher une recommandation.

Comme la configuration est très similaire, la méthodologie employée sera la même que celle du scénario 5. Pour plus de détails, reportez-vous à ce scénario.

3.7.2 Mise en œuvre de la configuration de DRS

La mise en œuvre de la configuration se déroule de la même manière que celle du scénario précédent, mais quand vous arrivez sur cet écran :

🕑 DRS Settings	
General VMware DRS Rules Virtual Machine Options Power Management VMware EVC Swapfile Location	Automation Level Manual VirtualCenter will suggest migration recommendations for virtual machines. Partially automated VirtualCenter will suggest migration recommendations for virtual machines. Image: The state of the submatically placed onto hosts at power on and virtual machines will be automatically placed onto hosts when powered on, and will be automatically migrated from one host to another to optimize resource usage. Migration threshold: Conservative Apply all recommendations. VirtualCenter will apply recommendations that promise even a slight improvement to the cluster's load balance.
	Advanced Options
Help	OK Cancel

Figure 3.15 : configuration de DRS en mode automatique

Choisissez « Fully automated » à la place de « Manual », puis déplacer le curseur sur « Aggressive ».

3.7.3 Observations et mesures

Comme attendu, au démarrage des VMs, le serveur ESX a été choisi en fonction de la charge disponible.

🖃 🌮 Hosts & Clusters	10.1.2.1.F2.VMmara EEV. Compare 2	E.O. 1E207E Euclust	ion (47 dau)	(c) romaining)		
🗖 🖶 New Datacenter	10.1.2.133 THWAIE ESA SELVER, 5	.3.0, 133073 E¥aluau		s) remaining)		
n die Des	Getting Started Summary Virtua	al Machines Performand	ce Configu	ration Tasks & Ev	ents Alarms P	ermissions Maps
				,	, , , , , , , , , , , , , , , , , , , ,	
10.1.2.153						
10.1.2.154		,		,		,
dient2	Name	State	Status	Host CPU - MHz	Host Mem - MB	Guest Mem - %
🔂 Nostalgia1	A XP2	Powered On	000	34	950	92
🐴 Nostalgia2	- Nostalgia3	Powered Off	000	0	0	
Nostalgia3	inoscalgias	POWEIEd OIT	000	0	0	0
Mostalgia4						
- Nostalgia						
Nustaigas						
UDservateur2						
p Serveur153						
👘 XP1						
👘 XP2						
	1					
	Figure 3.16 : VM	ls tournant sur	le serve	ur 1		
	0					
🖃 🃂 Hosts & Clusters	10 1 2 154 VMware ESX Server 3	5.0. 153875 Evaluati	ion (55 davi	(s) remaining)		
🗖 📠 New Datacenter	10.1.2.134 fi indie E5x 5ci (ci, 5	.5.0, 155015 Etaldad	1011 (55 447)	sy remaining/		
E EB DRS	🖓 🖬 DRS 🛛 🛛 Getting Started Summary Virtual Machines Performance Configuration Tasks & Events Alarms Permissions Maps					
□ III 10 1 2 153				```	· · · · · ·	, , , , , , , , , , , , , , , , , , , ,
10.1.2.154						
[] [] [] [] [] [] [] [] [] [] [] [] [] [Mama	Chake	Chakus	Heat COLL MUs	Heat Man MD	Cuest Mars 9/
		Julie	Juatus	THOSE CHO FINITZ	Those Mein - Mb	duest menn - 76
mostalgia1	📋 🛗 Client2	Powered Off	000	0	0	0
m Nostalgia2	👘 Nostalgia2	Powered Off	000	0	0	0
Mostalgia3	🚹 Serveur153	Powered Off	000	0	0	0
Mostalgia4	🖶 Nostalgia4	Powered Off	000	0	0	0
👘 Nostalgia5	A Nostalgia5	Powered Off	000	0	0	0
👘 Observateur2		Demored C=	000	27	012	
Serveur153	Observatedf2	🔽 Fowered On		Z 1	012	66
	I R H H H H	- 1.00				66
TA XP1	🗿 Nostalgia1	Powered Off	000	0	0	66
TP1	 Nostalgia1 XP1 	Powered Off Powered On	000	0 26	0 753	66

Figure 3.17 : VMs tournant sur le serveur 2

Après avoir éteint la VM se trouvant seule sur un serveur ESX, une des autres a été migrée par vMotion. La migration a été initiée moins de 5 minutes après l'extinction de la VM isolée, et non immédiatement.

約 Migrate Virtual Machine	Ð	Observateur2	37% 📃

Figure 3.18 : migration d'une des 2 VMs restante

La réaction de DRS est bien similaire à celle attendue expliquée au début de ce scénario.

3.8 Scénario 7 : déplacement automatique, changement du réglage de la sensibilité

3.8.1 Configuration testée et méthodologie

Les déplacements se feront toujours de façon automatique, mais la sensibilité sera modifiée.

La méthodologie est celle employée dans les scénarios 5 et 6. Une fois le test fini, la sensibilité est réduite d'un cran, puis le test est recommencé jusqu'à ce que la migration d'une des deux VMs se trouvant sur le même serveur ESX après l'extinction de la troisième ne se produise plus.

Normalement, cela devrait arriver quand la sensibilité sera de 3 étoiles ou plus (position centrale du curseur). En effet, comme vu lors du scénario 5, après l'extinction de la troisième VM, la recommandation portait 2 étoiles.

3.8.2 Mise en œuvre de la configuration de DRS

Ouvrez la fenêtre de configuration de DRS (clic droit sur le cluster -> Edit settings... puis VMware DRS dans le menu de gauche). À chaque nouveau test, déplacez d'un cran sur la gauche le curseur de sensibilité, puis cliquez sur *OK*.

🖉 DRS Settings		X
General VMware DRS Rules Virtual Machine Options Power Management VMware EVC Swapfile Location	Automation Level Manual VirtualCenter will suggest migration recommendations for virtual machines. Partially automated Witual machines will be automatically placed onto hosts at power on and virtual machines. Fully automated Witual machines will be automatically placed onto hosts when powered on, and will be automatically migrated from one host to another to optimize resource usage. Migration threshold: Conservative Apply all recommendations. VirtualCenter will apply recommendations that promise even a slight improvement to the cluster's load balance.	
Help	OK Cance	

Figure 3.19 : configuration de la sensibilité de DRS en mode automatique

3.8.3 Observations et mesures

Comme prévu, au niveau 2 (2 étoiles minimum sont nécessaires pour initier la migration), la migration a bien lieu à l'extinction de la troisième VM, une des 2 VMs restante est migrée.

Passage au niveau 3, à l'extinction de la troisième VM, aucune migration n'est initiée. En redescendant au niveau 2, la migration se déclenche.

Les tests des niveaux 4 et 5 n'ont pas été effectués, car si le niveau 3 ne fonctionne pas, les niveaux 4 et 5 qui sont plus restrictifs ne fonctionneront pas.

DRS a, à nouveau, réagi comme prévu au début du scénario.

3.9 Scénario 8 : 2 VMs doivent rester ensemble

3.9.1 Configuration testée et méthodologie

Le mode sera à nouveau automatique et la sensibilité maximum. Cependant, une règle obligera 2 VMs à rester sur le même serveur. Ce cas de figure peut se présenter quand une VM est utilisée comme firewall pour une autre VM. Il va de soi qu'il serait aberrant de déplacer la VM qui doit être protégée par le firewall sans déplacer ce dernier.

La méthodologie sera la même que précédemment, mais une fois la troisième VM éteinte (celle qui n'est pas soumise à la règle), si les 2 VMs restantes ne migrent pas, la règle sera retirée, ce qui devrait provoquer la migration.

3.9.2 Mise en œuvre de la configuration de DRS

Pour commencer, répétez la mise en œuvre de la configuration du scénario 6, mais ne cliquez pas sur OK.

🕝 DRS Settings		
General VMware DRS Rules Vritual Machine Options Power Management VMware EVC Swapfile Location	Use this page to create rules for virtual machines within this cluster. Rules will at to virtual machines only while they are deployed to this cluster and will not be retained if the virtual machines are moved out of the cluster.	petails
Help	ОК	Cancel

À la place, cliquez sur Rules dans le menu de gauche, puis sur Add...

Figure 3.20 : ajout d'une règle à DRS

Dans la nouvelle fenêtre, entrez un nom pour votre règle (ici : « scénario 8 »), et laissez le type d'origine (« Keep ... »). Cliquez sur *Add...*

Virtual Machine Rule	X
Give the new rule a name and choose its type from the menu below. Then, select the virtual machines to which this rule will apply. Name scénario 8	
Type	
Virtual Machines	
Add	
<u></u> K <u>Cancel</u>	

Figure 3.21 : choix de la règle

Cochez la case à coté du nom des VMs auxquelles la règle doit s'appliquer, puis cliquez 2 fois sur OK.

🕝 Virtual Machines		
Virtual machine list		
 Client2 Nostalgia1 Nostalgia3 Nostalgia4 Nostalgia5 Observateur2 Serveur153 ✓ XP1 ✓ XP2 		
	All	None
	ОК	Cancel

Figure 3.22 : choix des VMs concernées par la règle

Votre règle apparaît sur la fenêtre de configuration. Si vous désirez la désactiver, il faut simplement décocher la case à gauche de la règle. Cliquez sur *OK* pour appliquer la configuration.

🕝 DRS Settings		$\overline{\mathbf{X}}$
Constrained by the second seco	Use this page to create rules for virtual machines within this cluster. Rules to virtual machines only while they are deployed to this cluster and will not retained if the virtual machines are moved out of the cluster.	will apply t be
	Add Remove Edit	Details
Help	ОК	Cancel

Figure 3.23 : apparition de la règle dans la fenêtre de configuration

3.9.3 Observations et mesures

Après l'extinction de la troisième VM, les 2 autres restent sur le même serveur ESX. Après quelques *Generate Recommendations*, aucune migration n'est déclenchée.

Il reste donc à désactiver la règle, pour voir si c'est bien cette dernière qui empêche la migration. En effet, après le premier *Generate Recommendations* qui suit le changement de règle, la migration est déclenchée.

Toutes les étapes s'étant déroulées comme prévues, on peut en conclure que les règles établies pour forcer des VMs à rester ensemble fonctionnent bien.

3.10 Scénario 9 : 2 VMs ne doivent pas être sur le même serveur ESX

3.10.1 Configuration testée et méthodologie

Le mode sera à nouveau automatique et la sensibilité maximum. Cependant, une règle obligera 2 VMs à rester sur le même serveur.

Contrairement au scénario précédant, on ne cherchera pas à garder 2 VMs ensemble, mais à les empêcher d'être ensemble.

La méthodologie sera différente des scénarios précédents. Au niveau de VMs, il y aura toujours 2 Windows XP utilisés lors des précédents scénarios, mais la troisième VM sera une machine Nostalgia.

Nostalgia est une collection de vieux jeux DOS, qui utilisent un système d'attente active, ce qui implique une charge processeur très élevée. Cette occupation du processeur bien supérieure à celle des VMs XP cumulées aura pour effet d'isoler la VM Nostalgia sur un serveur ESX. Seule une règle imposant au VMs XP d'être sur des serveurs ESX différents pourra provoquer le déplacement d'une VM XP sur le même serveur ESX que la VM Nostalgia.

La VM Nostalgia sera lancée en premier. Viennent ensuite les 2 VMs Windows, qui se retrouveront donc sur le même serveur ESX, le serveur où se trouve la VM Nostalgia étant saturé.

action	serveur ESX 1	serveur ESX 2	
démarrage de la VM Nostalgia			
démarrage de la VM XP1		VM Nostalgia	
démarrage de la VM XP2	VM XP1	VM Nostalgia	
toutes les VMs sont allumées,	VMs XP1 et	VM Nostalgia	
aucun déplacement n'a lieu	XP2	vivi nostalgia	
application de la règle d'exclusion	VMs XP1 et		
application de la règle d'exclusion	XP2	vivi Nostalgia	
la VM XP2 est déplacée	VM XP1	VMs Nostalgia et XP2	

Figure 3.24 : méthodologie du scénario 9

Ce type de règle sera utilisé, par exemple, sur une VM faisant office de serveur web et une seconde VM servant de miroir à la première. Si on crée un miroir, c'est pour éviter que le serveur web se retrouve indisponible suite à une panne matérielle (entre autres). Si les 2 VMs sont sur le même serveur ESX ou ESXi lors de la panne, le miroir ne pourra pas remplir son rôle, et ne sert donc à rien.

3.10.2 Mise en œuvre de la configuration de DRS

La mise en œuvre est la même que celle du scénario 8, à l'exception du choix du type de règle, ou il faut choisir xxx à la place de « keep ... », comme sur la capture ci-dessous.

Name			
scénario 9			
Туре			
Separate Virtual Mac	hines		
<u>V</u> irtual Machines —			
1			

Figure 3.25 : choix du type de règle empêchant des VMs d'être sur le même serveur ESX

3.10.3 Observations et mesures

Le démarrage s'est déroulé comme prévu, la VM Nostalgia s'est retrouvée sur un serveur ESX et les VMs Windows sur l'autre. Plusieurs *Generate Recommendations*, les VMs sont toujours dans la même configuration.

La règle est alors activée, et un *Generate Recommendations* effectué. Une des 2 VMs Windows se retrouve déplacée sur le même serveur ESX que la VM Nostalgia, et ce malgré la charge processeur très élevée de ce dernier.

Cette configuration fonctionne donc comme prévu.

3.11 Conclusion

La granularité des règles est très satisfaisante, les règles pouvant être appliquée VM par VM si cela devait s'avérer nécessaire. Bien que les règles soient peu nombreuses, elles permettent de remplir l'objectif premier de DRS, à savoir répartir la charge entre les serveurs ESX/ESXi.

Je regrette cependant le manque de documentation concernant le fonctionnement de DRS et particulièrement à propos des paramètres avancés. La documentation ne manque pas concernant la configuration, qui est très intuitive, mais le fonctionnement n'est expliqué que dans peu de documents (en contraste avec vMotion, dont des documents détaillés sur le fonctionnement ont pu être trouvé).

Je recommande néanmoins l'activation de DRS si vos VMs sont amenées à avoir une charge variable, car DRS permettra à ces dernières de fonctionner dans un environnement le plus favorable possible. Faites toutefois attention à ne pas régler les déplacements automatique trop sensiblement, ce qui génèrerait un important trafic vMotion, nuisible pour les performances des serveurs ESX/ESXi.

4 Conclusions

Suite à ce projet, j'ai pu répondre à certaines de mes questions concernant le fonctionnement des produits de virtualisation de VMware, particulièrement celui de vMotion. Les objectifs fixés ont tous pu être atteints.

On retiendra de vMotion des temps d'interruption très court dans les conditions minimales recommandées (3.3 secondes avec un réseau de stockage à 1Gbps), et en utilisant que des serveurs ESX (les serveurs ESX étant apparemment bridé en débit réseau). Cependant, le temps annoncé par VMware pour les interruptions est de 2 secondes maximum, ce qui n'a pas été le cas lors de mes tests. Cette différence est imputable au serveur Openfiler qui émule le disque iSCSI plus qu'aux produits VMware. Il faudrait donc idéalement refaire une série de mesures avec un disque iSCSI réel et non pas émulé.

Il reste également une question en suspend : pourquoi les serveurs ESX envoient un nombre de données bien plus élevé pour le pré-copie de la RAM et la copie du fichier BMP, contenant les changements apportés à cette dernière durant son transfert, que les serveurs ESXi ? Pour rappel, les serveurs ESXi échangent environ 290MB de données là ou les serveurs ESX échangent 425MB de données. Cette différence n'est pas tolérable, d'autant plus qu'elle n'est apparemment pas justifiée.

DRS, quand à lui, remplit entièrement son rôle (équilibrer la charge entre les serveurs ESX/ESXi d'un même cluster en déplaçant des VMs via vMotion ou en suggérant à l'utilisateur de le faire), et toutes les règles testées fonctionnent comme prévu au début des scénarios de test. Le seul point noir de DRS est le manque de documentation sur son fonctionnement et sur les paramètres avancés.

Ces fonctions apportent un vrai plus aux produits proposé par VMware et permettent de rendre la virtualisation plus performante. Elles offrent la possibilité de faire de la maintenance matériel sans pour autant devoir interrompre le fonctionnement des VMs, ce qui n'est pas le cas des serveurs traditionnels. En prenant également en compte l'argument économique, je pense que la progression de l'utilisation des VMs n'est pas prête de s'arrêter.

A Annexes

A.1 Versions des logiciels utilisés

Voici une liste des logiciels utilisés dans le cadre de mon travail de diplôme :

- Windows XP SP2 32bits Build 2600
- VMware ESX Server version 3.6 update 4 64bits avec licence d'évaluation de 60 jours
- VMware ESXi version 3.6 update 3 64bits Build 123629
- FreeNAS version 0.69 pour plateforme x86 (héberge le disque iSCSI utilisé par ESX/ESXi)
- Openfiler version 2.3 pour plateforme x86 (héberge le disque iSCSI utilisé par ESX/ESXi)
- VMware vCenter Server (VI Server) version 2.5.0 update 4 Build 147697
- VMware VI Client 2.5.0 Build 147704
- Wireshark Version 0.99.7 (SVN Rev 23910)

A.2 Matériel

Voici la liste du matériel utilisé (pc, serveurs et switch) :

- PC Gigabyte (serveur ESX/ESXi et FreeNAS), processeur dualcore 3GHz 64bits, 4GB de RAM (2GB pour Openfiler)
- PC Dell (poste de contrôle -> VI Server & Client), processeur pentium 4 2.8GHz 32 bits, 1GB de RAM
- Switch Cisco Catalyst 2950 à 100Mbps full duplex
- Switch NETGEAR à 1Gbps full duplex

A.3 Installation d'Openfiler

Je vous recommande d'utiliser la distribution GNU/Linux pour l'émulation de disque iSCSI si vous n'en possédez pas de physique (voir annexe A.9.3), malgré des problèmes de performances sur un réseau gigabit (voir annexe A.7).

Voici deux tutoriels, un pour l'installation et l'autre pour la configuration d'Openfiler (les deux sont en anglais) expliquant bien les diverses étapes :

- Installation : <u>http://www.openfiler.com/learn/how-to/graphical-installation</u>
- Configuration : <u>http://www.techhead.co.uk/how-to-configure-openfiler-v23-iscsi-storage-for-use-with-vmware-esx</u>

A.4 Installation et configuration de ESX/ESXi

A.4.1 ESXi

L'installation se déroule en mode texte, la seule opération préalable consiste à accepter l'EULA, tout le reste se fait automatiquement. Une fois l'installation terminée, la machine redémarre, et à ce moment nous pouvons modifier la configuration. Voici une liste non-exhaustive des paramètres modifiable :

- Mot de passe d'administration (root)
- Adresse IP
- Disposition du clavier (le choix étant restreint, une bonne solution consiste à utiliser le clavier allemand qui est très similaire au clavier suisse romand)
- ...

A.4.2 ESX

Contrairement à ESXi, ESX propose une installation en mode graphique en plus de l'installation en mode texte, j'ai donc utilisé l'installation en mode graphique (attention, cette dernière nécessite une souris). Cette fois ci, la configuration a lieu avant l'installation, voici les différentes étapes de cette dernière ainsi que quelques recommandations :

- Test facultatif du CD d'installation si ce dernier a été gravé et non pressé
- Message de bienvenue
- Choix de la disposition du clavier (cette fois ci, la disposition suisse romande est disponible)
- Choix du modèle de souris (si la souris fonctionne bien, vous pouvez conserver le choix par défaut)
- Un message d'erreur si le disque n'est pas vierge, ce message est à ignorer
- Lecture et acceptation de l'EULA
- Partitionnement, le mode automatique est recommandé et si votre disque dur n'est pas vierge, je vous recommande de retirer le vu de l'option « keep virtual machine ... » car ESX ne supportant que les disques SCSI et non IDE ou SATA, il ne sert à rien de conserver ces partitions
- Apparition d'un Warning si le disque dur n'était pas vierge informant que les partitions existantes seront formatées (cliquez sur « Yes »)
- Apparition d'un Warning si le disque dur est de type IDE ou SATA indiquant que ce type de disque n'est pas supporté (cliquez sur « Yes »)
- Un écran qui montre le partitionnement automatique proposé qu'il est possible de modifié manuellement s'il ne vous convient pas
- Réglage des options avancées, tel que les options de boot
- Configuration réseau (ip, sous-réseaux, passerelle par défaut, serveur DNS, hostname, ...)
- Sélection du fuseau horaire de l'emplacement du serveur (pour la Suisse, choisir Zurich)
- Choix du mot de passe administrateur (root) et création d'utilisateurs supplémentaires
- Récapitulatif des paramètres choisis
- Installation
- Ejection du CD d'installation et reboot

A.5 Virtual Infrastructure Server (VI Server)

Le VI Server permet de gérer des groupes de machines ESX et ESXi, rendant possible la mise en œuvre de fonctions comme vMotion.

Lancez l'installation, puis choisissez la langue d'installation (le Français n'est pas disponible) et cliquez sur OK.

Choose	Setup Language 🛛 🛛 🕅
Ð	Select the language for this installation from the choices below.
	English (United States)
	OK Cancel

Figure A.1 : VI Server installation, choix de la langue

Un texte de bienvenue apparait, suivi d'un texte d'introduction. Cliquez 2 fois sur Next>.

Choisissez maintenant les composants que vous désirez installer. Je vous recommande de cocher au moins les 3 premières cases si elles ne sont pas grisée (ce qui signifierai que le composant en question est déjà installé), Puis cliquez sur :

🗒 VMware Infrastructure Management 🛛 🛛 🔀
VMware Infrastructure Components Selection
Select the VMware Infrastructure components that you want to install on this computer.
WMware Infrastructure Client
₩ Wware VirtualCenter Server
VMware Update Manager
VMware Converter Enterprise for VirtualCenter Server
InstaliShield Cancel

Figure A.2 : VI Server installation, choix des composants

Le VI Server nécessitant une base de données, il vous est proposé soit d'en créer une nouvelle, soit d'en utiliser une existante. À moins de savoir ce que vous faites, choisissez la création d'une nouvelle base de données. Cliquez sur *Next>*.

VMware Infrastructure Management VMware VirtualCenter Server Deployment Options - Step 1 Select a database for VirtualCenter Server				
VMware VirtualCenter require	es either a Microsoft SQL Server or Oracle database server.			
● Install Microsoft SQL Se	rver 2005 Express (suitable for small scale deployment)			
OUse an <u>e</u> xisting databas	se server			
Data Source Name (DSN)	VMware VirtualCenter			
Login User Name				
Password				
stallShield	< Back Next > Can	cel		

Figure A.3 : VI Server installation, création d'une base de données

Si vous ne possédez pas de licence, laissez le vu dans la case « I want to evaluate VirtualCenter Server », sinon, décochez cette case et cochez « Use an Existing License Server » puis modifiez le chemin en cliquant sur *Browse…* si nécessaire. Une fois fait, cliquez sur *Next*>.

Configure a license server for Virt	Deployment Options - Step 2
VirtualCenter License Server	
VirtualCenter Server retrieves licen license server, we will install one fo	nses from a license server. If you do not have a
	-
I want to eyaluate virtualCente	er Server
Use an Existing License Ser	rver
C:\Program Files\VMware\VM	Nware License Server\Licenses\vmware Browse
C:\Program Files\VMware\VM	Iware License Server\Licenses\vmwari Browse VirtualCenter Management Server V
C:\Program Files\VMware\VM	Mware License Server Licenses (vmwari Browse) VirtualCenter Management Server
C:\Program Files\VMware\VM	Mware License Server\Licenses\vmwari Bro <u>w</u> se VirtualCenter Management Server ♥

Figure A.4 : VI Server installation, choix de la licence

Si vous avez choisi d'utiliser la version d'évaluation, un message vous indiquera que vous disposez de 60 jours de tests. Cliquez sur *OK*.

Vous pouvez maintenant régler les ports qui seront utilisés par les différents services du VI Server. Si vous ne devez pas utiliser de ports spécifiques, laissez ceux par défaut. Cliquez sur *Next>*.

VMware Infrastructure Management VMware VirtualCenter Server Deployment Options - Step 3 Customize port settings for VMware VirtualCenter Server.					
VirtualCenter Ports	80	HTTPS Web Service	443		
Heartbeat (UDP)	902	Web Server Port:	8086		
		< Back Next >	Cancel		

Figure A.5 : VI Server installation, choix des ports

Vous pouvez maintenant changer le répertoire d'installation en cliquant sur *Change…* ou conserver celui par défaut. Une fois choisi, cliquez sur *Next>*.

Cliquez sur *Install* pour débuter l'installation, puis patientez un moment. Une fois l'installation terminée, cliquez sur *Finish*.

A.6 Virtual Infrastructure Client (VI Client)

A.6.1 Installation

L'installation du VI Client est intégrée dans l'installation du VI Server. Pour vous connecter, entrez dans le champ adresse IP *localhost* (si le VI Client est sur la même machine que le VI Serveur, sinon, mettez l'adresse IP correspondante), et complétez les identifiants en utilisant ceux du compte d'un administrateur local.



Figure A.6 : fenêtre de connexion au VI Client

A.6.2 Ajout/création d'un Datacenter

Cette étape est nécessaire uniquement dans le cas où l'on se connecte sur un VI Server, afin de pouvoir lui attribués des machines « physiques » à gérer. Il faut se connecter au VI server à l'aide du VI Client, puis dans la partie gauche de la fenêtre, faire un clic droit sur le dossier « Hosts & Clusters ». Dans le menu qui s'ouvre, sélectionner *New Datacenter*, ce dernier ira s'ajouter dans l'arborescence en dessous de « Hosts & Clusters », en vous proposant d'entrer un nom. Une fois le nom entré (ou si vous souhaitez conserver le nom d'origine), appuyez sur enter.



Figure A.7 : création d'un Datacenter

A.6.3 Ajout d'un host

Cette opération consiste à lier/attribuer une machine à un Datacenter pour profiter de fonction comme vMotion qui ne peuvent se faire qu'entre membre du même groupe. Cette opération est à répéter pour toutes les machines que vous souhaitez ajouter à votre VI Server.

Pour commencer, sélectionnez dans la partie de droite le Datacenter auquel vous désirez ajouter une machine. Dans le menu qui s'ouvre, choisissez *Add Host...*



Figure A.8 : ajout d'un hôte, sélection de la Database

Dans la nouvelle fenêtre, entrez l'adresse IP de la machine à ajouter, ainsi que les identifiants administrateur correspondant à la machine. Cliquez sur *Next>*.

Add Host Wizard Specify Connection Setting Type in the information use	gs ed to connect to this host.
Connection Settings Host Summary Virtual Machine Location Ready to Complete	Connection Enter the name or IP address of the host to add to VirtualCenter. Host name: 10.1.2.154 Authorization Enter the administrative account information for the host. VMware Infrastructure Client will use this information to connect to the host and establish a permanent account for its operations. Username: root Password: root
< >	
Help	≤Back Next ≥ Cancel

Figure A.9 : ajout d'un hôte, entrée des identifiants de l'hôte

L'écran suivant donne un récapitulatif des caractéristiques de la machine hôte choisie, cliquez sur *Next* >.

🕝 Add Host Wizard						
Host Information Review the product infor	nation for the specific	ed host.				
Connection Settings	 You have chose	You have chosen to add the following host to VirtualCenter:				
Host Summary Virtual Machine Location Ready to Complete	Name: Vendor: Model: Version: Virtual Machine	10.1.2.154 Gigabyte Technology Co., Ltd. G33M-52L VMware ESX Server 3.5.0 build-153875 s:				
<						
Help		≤Back Next ≥ Cancel				

Figure A.10 : ajout d'un hôte, récapitulatif matériel

Ce nouvel écran propose de choisir la database à laquelle ajouter l'hôte, par défaut, la database sélectionnée est celle choisie au début (voir figure A.8). Cliquez sur *Next>*.

🖉 Add Host Wizard		
Virtual Machine Location Select a location in the Virtua	alCenter inventory for the host's virtual machines.	
Connection Settings Host Summary	Select a location for this host's virtual machines.	
Yirtual Machine Location Ready to Complete	New Datacenter	
	, ≤Back Nex	t ≥ Cancel



La dernière fenêtre propose un récapitulatif des caractéristiques logicielles (ESX) de l'hôte, cliquez sur Finish pour que l'ajout s'effectue.

🕝 Add Host Wizard		- 🗆 🗙
Ready to Complete Review the options you have	selected and click Finish to add the host.	
Connection Settings Host Summary	Review this summary before finishing the wizard.	
Ready to Complete	Host: 10.1.2.154 Version: VMware ESX Server 3.5.0 build-153875 Networks: VM Network	
Help	≤ Back	Cancel

Figure A.12 : ajout d'un hôte, récapitulatif logiciel

A.6.4 Ajout d'un VMkernel

Le VMkernel permet, entre autre, d'utiliser vMotion ainsi que de connecter des disques iSCSI. Cette opération n'est pas nécessaire sur des hôtes ESXi qui ont le VMkernel automatiquement installé, mais devra être répétée sur tous les hôtes ESX individuellement.

Dans le VI Client, sélectionnez la machine à laquelle vous voulez ajouter le VMkernel, puis rendezvous à l'onglet « Configuration », partie « Networking ». Cliquez sur *Add Networking…*

🖉 localhost - VMware Infrastru	cture Client						_ 2 🛛
Eile Edit View Inventory Administr	ation Plugins Help						
Inventory Scheduled Tasks	 Weight Stratic Events Administratic 	an Maps C	Consolidation				60 Days Remaining
🌩 🔶 🗗 🤕							
New Datacenter 10.1.2.153 10.1.2.154	10.1.2.154 VMware ESX Getting Stated Summ The VMware ESX To run vitue machines, c Note: If you plan to use it To add storage now, click Hardware Health Status Processors Memory Storage Networking Storage Adapters Network Adapters Software Licensed Features Time Configuration DK5 and Routing Software Software Software Sum Configuration DK5 and Routing Sum Configuration DK5 and Routing Sum Configuration DK5 and Routing Software State Resource Alk Advanced Settings	Server, 3.5.0, 15387 Server, 3.5.0, 15387 Server does mail Machine Scale at least one data CCSI or a network file server here to create a datast	75 Evaluation (60 day(s) rem Resource Allocation Perform ot have persistent stor store for maintaining vitual machine vystem (VFS), ensure that your store tore Networking Vitual Switch: VSwitch0 Vitual Switch: VSwitch0 Vitual Switch: VSwitch0 Vitual Switch: Console Port Service Console Port Service Console Port VSWID : 10.1.2.154	atining) nce Configuration age. ss and other system files. ge adapters and network Remov Physical Adapt Physical Adapt	Connections are properly of e Properties es 100 Full p	Permissions Maps	Refresh Add Networking
Recent Tasks							×
Name Target	Status	Initiated by	Time	Start Time	Complete Time		
Reboot Host 11	0.1.2.153 S Comple 0.1.2.153 Comple	eted Administrato	or 05.05.2009 11:33:22 or 05.05.2009 11:30:32	05.05.2009 11:33:22 05.05.2009 11:30:32	05.05.2009 11:33:22 05.05.2009 11:30:34		
🔄 Tasks 💇 Alarms							Administrator

Figure A.13 : ajout d'un VMkernel, choix de la machine et ouverture de la fenêtre d'ajout de réseaux
🕝 Add Network Wizard Connection Type Networking hardware can be partitioned to accommodate each service requiring connectivity. Connection Type -Connection Types Network Access Connection Settings O Virtual Machine Summary Add a labeled network to handle virtual machine network traffic. • VMkernel The VMkernel TCP/IP stack handles traffic for the following ESX server services: VMware VMotion, iSCSI, and NFS. Service Console Add support for host management traffic. Help <u><</u> Back Next \geq Cancel

Dans la nouvelle fenêtre, sélectionnez « VMkernel », puis cliquez sur Next>.

Figure A.14 : ajout d'un VMkernel, sélection du réseau « VMkernel »

Dans mon cas, je n'ai pas créé de vSwitch en supplément de celui créé par défaut, et je ne désire pas le faire. Je vais donc sélectionner « Use vSwitch0 » dans la liste proposée. On peut constater que dans la partie basse de la fenêtre, un aperçu du vSwitch après modification est proposé. Cliquez sur *Next>*.

🖁 Add Network Wizard				_ 0 >
VMkernel - Network Ac The VMkernel reache:	c cess s networks through uplink adapters attached	to virtual switches.		
Connection Type Network Access	Select which virtual switch will handle t using the unclaimed network adapters	the network traffic for th listed below.	is connection. You may also create a new	virtual switch
Connection Settings Summary	Create a virtual switch	Speed	Networks	
Samilary	🖂 🔛 vmnic1	down		
	mnic2	down	10.1.0.1-10.1.3.254	
	O Use vSwitch0	Speed	Networks	
	📈 🔜 vmnic0	100 Full	10.1.0.1-10.1.7.254	
	Preview:	Physical Adapt	teiz	
	Virikernei - Virtual Machine Port Group		cu	
	- Service Console Port Service Console vswif0 : 10,1,2,153	Q •		
Help			≤ Back Next ≥	Cancel

Figure A.15 : ajout d'un VMkernel, choix du réseau

Complétez les champs de la partie « IP Settings », en faisant toutefois attention d'utiliser une adresse IP disponible. Vous pouvez également modifier les valeurs de « Port Group Proporties », mais ce n'est pas obligatoire et pas forcément nécessaire. On peut constater que la partie basse de la fenêtre propose à nouveau un aperçu de la configuration qu'aura le vSwitch une fois les modifications appliquées. Cliquez sur *Next*>.

🕝 Add Network Wizard			_ 🗆 🛛
YMkernel - Connection Se Use network labels to ide	ettings entify VMkernel connections while ma	naging your hosts and datacenters.	
Connection Type Network Access Connection Settings Summary	Port Group Properties Network Label: VLAN ID (Optional):	VMkernel	
	IP Settings IP Address: Subnet Mask:	10 , 1 , 2 , 152 255 , 255 , 0 , 0	
	Preview:	Physical Adapters	
	VMkernel 10.1.2.152 - Virtual Machine Port Group	Image: Second	
	Service Console Port Service Console vswif0 : 10.1.2.154	<u>®</u> .	
Help		<u>≤</u> Back Next ≥	Cancel

Figure A.16 : ajout d'un VMkernel, saisie des paramètres réseau

La dernière fenêtre propose elle aussi un aperçu du vSwitch après modification. Cliquez sur *Finish* pour appliquer les changements.

Si votre hôte n'a pas d'adresse de passerelle par défaut spécifiée, une fenêtre supplémentaire s'ouvrira afin que vous l'ajoutiez. Dans la majorité des cas, vous pouvez réutiliser celle utilisée par le « Service Console » indiquée juste au-dessus.

🕗 DNS and Routing Configuration 🛛 🛛 🔀					
DNS Configuration Routing					
Service Console					
Default gateway:	10 . 1 . 0 . 1				
Gateway device:	vswif0				
VMkernel					
Default gateway:	10 . 1 . 0 . 1				
	OK Cancel He	lp			

Figure A.17 : ajout d'un VMkernel, configuration de la passerelle par défaut

A.6.5 Ajout d'un disque iSCSI

Cette opération permet d'ajouter un espace disque partitionné en VMFS3 à un hôte ESX ou ESXi. Cette étape est nécessaire pour pouvoir utiliser vMotion, car cette fonctionnalité requiert que les hôtes de départ et de destination aient accès à l'espace de stockage de la VM à transférer.

Pour commencer, sélectionnez l'host auquel vous désirez ajouter le disque iSCSI, puis rendez-vous sur l'onglet « Configuration », partie « Storage Adapters ». Sélectionnez ensuite la ligne « iSCSI Software Adapter » de la partie « Storage Adapters » de la fenêtre. Une fois que cette étape est effectuée, cliquez sur *Proporties...*

🖉 localhost - VMware Infrastructure Client 💷 🗐					d 🔀					
Ele Edit View Toyentory Administration Blugins Help										
Inventory Scheduled Tasks	الا Events ب	Administration	🐣 Maps	Consolidation					<mark>60</mark> Days Remaining	æ
🗢 🔶 🗗 🤩										
Morts & Clusters Mew Datacenter Morts 0.10.2.153 10.1.2.154	10.1.2.153 VM Getting Starte The VMw To run virtual Note: If you p To add storag Hardware Health St. Processor	tware ESX Serv are ESX Se machines, create wachines, create alan to use ISCSI of e now, click here abus s	rer, 3.5.0, 15 Virtual Machin rver does at least one d or a network fil to create a da	S875 Evaluation (60 day(s Resource Allocation Prot have persistent s atastore for maintaining virtual m e system (NFS), ensure that you tastore Storage Adapters Device ISSIS Software Adapter	y remaining) vformance Configuration storage, achines and other system files. v storage adapters and network	Taska & Events Alarm connections are propert Type	Permissions Maps y configured before continuing, SAN Identifier		Reso	an
	Memory Storage Networkir • Storage # Network #	ng Adapters Adapters		SCSI Software Adapter		ISCSI				
Recent Tasks Name Target	Liensed Time Con DNS and Victual M Victual M Society F System, Advanced	Features figuration Routing Achine Startup/Sh Ichine Swapfile Lo Profile Secure Allocation I Settings	utdown cation 1	Model: ISCSI Name: ISCSI Allas:	Time Start Time	Complete Time		IP Address: Discovery Methods: Targets:	Properties.	
🚰 Tasks 💇 Alarms									Administ	rator

Figure A.18 : ajout d'un disque iSCSI, ouverture des propriétés de l'adaptateur iSCSI

77

Dans la nouvelle fenêtre, cliquez sur configure...

🔗 iSCSI Initiator (vmhba32)	Properties	
General Dynamic Discovery St	atic Discovery CHAP Authentication	
iSCSI Properties		
iSCSI name:	iqn.1998-01.com.vmware:G3-69fb3ed2	
iSCSI alias:	G3.localdomain	
Target discovery methods:	Send Targets	
Software Initiator Properties		
Status:	Enabled	
		Configure
	Clos	e Help

Figure A.19 : ajout d'un disque iSCSI, ouverture de la configuration

Ajoutez le vu dans la case à coté de « Enabled » puis cliquez sur Ok.

🕜 General Prope	rties 🛛 🔀
Status F Enabled	
iSCSI Properties	ian. 1998-01. com. vmware: a3-69fb3ed2
iSCSI Alias:	G3.localdomain
	OK Cancel Help

Figure A.20 : ajout d'un disque iSCSI, activation du contrôleur

Allez sur l'onglet « Dynamic Discovery » puis cliquez sur Add...

Complétez l'adresse IP de votre disque iSCSI, puis, si nécessaire, modifiez le numéro du port de ce dernier. Cliquez sur *Ok*.

🖉 Add Send Targets Server 🛛 🔀
Send Targets
iSCSI Server: 10 . 1 . 1 . 54
Port: 3260
Authentication may need to be configured before a session can be established with any discovered targets.
OK Cancel Help

Figure A.21 : ajout d'un disque iSCSI, entrée des informations de connections

De retour sur la fenêtre principale, cliquez sur *rescan...* Patientez quelques minutes, le temps que l'opération se termine. Si tout s'est déroulé normalement, vous verrez votre disque en cliquant sur la ligne en dessous de « iSCSI Software Adapter ».



Figure A.22 : ajout d'un disque iSCSI, fin du paramétrage de la connexion du disque

Toujours dans l'onglet « Configuration », sélectionnez la partie « Storage ». Si le disque contient déjà une partition formatée en VMFS3, elle apparaitra automatiquement, sauf dans le cas d'ESXi qui demande que la source soit vierge pour la mettre dans sa liste. Si votre cible est reconnue, vous aurez un écran similaire à celui de la figure A.17. Attention également, si votre machine possède un espace de stockage local, celui-ci apparait également dans cette liste. Si le disque iSCSI n'y est pas, cliquez sur *Add Storage...*



Figure A.23 : ajout d'un disque iSCSI, liste des emplacements de stockage

Plusieurs solutions vous sont proposées, choisissez « Disk/LUN » puis cliquez sur Next>.

🔗 Add Storage Wizard	
Select Storage Type Do you want to format a ne	w volume or use a shared folder over the network?
■ Disk/LUN Device Location Current Disk Layout Properties Formatting Ready to Complete	Storage Type Disk/LUN Choose this option if you want to create a datastore or other volume on a Fibre Channel, iSCSI or local SCSI disk. Network File System Choose this option if you want to use a shared folder over a network connection as if it were a Whare datastore. A mount point must be created on the host before it is added as a datastore. Diagnostic No diagnostic datastore is configured or none is accessible. Choose this option to reserve space for server fault data.
Help	≤ Back Cancel

Figure A.24 : ajout d'un disque iSCSI, choix du type de stockage

Attendez quelques instants que la liste s'actualise, puis sélectionnez votre disque dans la liste, puis cliquez sur *NEXT*>.

2	Add Storage Wizard						
	Select Disk/LUN						
	If a device cannot be configur	ed unambiguously, you	i will be asked to) select a pa	rtition.		
	Disk/LUN Device Location				SAN Identifier contains: 👻 🗌		Clear
	Current Disk Layout	Device	Capacity	Available	SAN Identifier	LUN	
	Properties	vmhba32:0:0	19.53 GB	None	iqn.1994-04.org.netbsd.isc.	. 0	
	Formatting Ready to Complete	vmhba32:1:0	48.83 GB	None	iqn.1994-04.org.netbsd.isc.	0	
	Ready to complete						
-		J					
	Help				<u><</u> Back	Next <u>></u>	Cancel

Figure A.25 : ajout d'un disque iSCSI, choix du disque

Vous vous trouvez maintenant devant un récapitulatif du contenu de votre disque. S'il n'est pas vierge, un petit warning s'affiche vous indiquant que toutes les données du disque seront perdues lors du formatage. Cliquez sur *Next>*.

🕝 Add Storage Wizard				_ 🗆 🔀
Current Disk Layout You can partition and form	at the entire device, all free space, or	a single block of free s	space.	
	Review the current disk layout:			
Current Disk Layout Properties	Device /vmfs/devices/disks/	Capacity 48.83 GB	Target Identifier vmhba32:1:0	LUN O
Ready to Complete	Primary Partitions VMFS	Capacity 48.82 GB	Description	
	Warning: The current disk lay	yout will be destroyed	I. All file systems and data will t	be lost permanently
Help			\leq Back Next \geq	Cancel

Figure A.26 : ajout d'un disque iSCSI, récapitulatif du contenu du disque

Entrez Le nom que vous désirez donner à votre disque. Ce nom apparaitra sur l'écran de la figure A.17 pour toutes les machines y ayant accès. Cliquez sur *Next>*.

🕝 Add Storage Wizard		
Disk/LUN - Properties Labels provide stable access	to VMFS volumes that is not affected by hardware variations	
Disk/LUN Device Location Current Disk Layout Properties Formatting Ready to Complete	Datastore Name	
Help	≤Back Next ≥	Cancel

Figure A.27 : ajout d'un disque iSCSI, choix du nom

Il faut maintenant choisir la taille des blocs de votre disque iSCSI. Je vous recommande de laisser les valeurs par défaut. Cliquez sur *Next>*.

🕝 Add Storage Wizard		
Disk/LUN - Formatting The format of your file syste	m determines which class of virtual machines it will be able to support.	
Disk/LUN Device Location Current Disk Lavout Properties Formatting Ready to Complete	Maximum file size Large files require large block size; the minimum disk space used by any file is equiblock size. These values are adjusted by VMF5-3 file systems on demand. [256 GB , Block size: 1 MB Capacity If Maximize capacity	al to the file system
Help	≤ Back Nex	t ≥ Cancel

Figure A.28 : ajout d'un disque iSCSI, choix de la taille des blocs

Vous vous retrouvez devant un récapitulatif général des options que vous venez de régler, cliquer sur *Finish*.

🖉 Add Storage Wizard				
Ready to Complete Review this summary befo	ore completing the wizard.			
Disk/LUN Beadu to Complete	Review the proposed disk layout:			
Ready to complete	Device /vmfs/devices/disks/	Capacity 48.83 GB	Target Identifier vmhba32:1:0	LUN O
	Primary Partitions VMFS	Capacity 48.82 GB	Description	
	The following VMware file system will	be created:		
	Datastore name: Datastore			
	Formatting File system: VMFS-3			
	Block size: 1 MB Maximum file size: 256 GB			
Help			<u>≤</u> Back <u>F</u> inish	Cancel

Figure A.29 : ajout d'un disque iSCSI, récapitulatif final

Vous pouvez suivre l'avancement de l'opération dans la partie basse de la fenêtre principale du VI Client. Une fois terminé, votre disque devrait apparaître comme sur la figure A.23.

L'ajout du disque iSCSI peut également se faire en ligne de commande depuis le serveur ESX auquel vous désirez l'ajouter (je n'ai pas testé cette procédure sur les serveurs ESXi qui n'offrent pas de

console). Le document <u>http://www.rtfm-ed.co.uk/docs/vmwdocs/ESX3.x-VC2.x-ServiceConsole-</u> <u>Guide.pdf</u> explique la procédure en détail, ce qui suit est donc un résumé de ce document.

Pour commencer, activez le iSCSI avec la commande suivante : esxcfg-swiscsi –e , cette étape correspond à la figure A.20.

Ensuite, tapez la commande vmkiscsi-tool -D -a 10.1.1.54 vmhba32, ou l'adresse ip est celle de votre disque iSCSI. Le numéro qui suit vmhba peut changer, le document utilisant le n°40, et mes serveurs utilisant le n°32. Cette commande a le même effet que l'action de la figure A.21.

Il faut ensuite forcer le serveur à scanner le périphérique iSCSI. Pour ce faire, utiliser la commande esxcfg-swiscsi –s, qui à le même effet que le clic sur le *Rescan…* à la figure A.22.

Quand j'ai effectué la configuration en ligne de commande, mon disque iSCSI était déjà formaté en VMFS. Si ce n'est pas votre cas, vous pouvez le faire graphiquement comme expliqué ci-dessus, ou en ligne de commande en suivant les instructions du document d'où proviennent les 3 lignes de commandes précédentes.

A.6.6 Activation de vMotion

VMotion est la fonction qui permet de déplacer une VM d'un hôte physique vers un autre. Cette fonctionnalité, disponible aussi bien sur un hôte ESX qu'un hôte ESXi, n'est pas activée par défaut.

Pour activer cette fonctionnalité, rendez-vous dans l'onglet « Configuration », partie « Networking ». Cliquez sur *Proporties…* en haut à droite du vSwitch pour lequel vous désirez activer vMotion.



Figure A.30 : activation de vMotion, choix du vSwitch

Dans la fenêtre qui vient de s'ouvrir, sélectionnez le VMkernel dans la liste puis cliquez sur Edit...

🔗 vSwitch0 Properties			
Ports Network Adapters			
Configuration	Port Properties		<u> </u>
the second summary	Network Label:	VMkernel iSCSI	
VM Network Virtual Machine	VLAN ID:	None	
Service Console Service Console	VMotion:	Disabled	
VMkernel iSCSI VMotion and IP		0.00000	
	IP Settings		
	IP Address:	10.1.7.22	
	Subnet Mask:	255.255.0.0	=
	MAC Address:	00:50:56:74:fa:c8	
	Effective Policies		
	Security		
	Promiscuous Mode:	Reject	
	MAC Address Changes:	Accept	
	Forged Transmits:	Accept	
	Teoffic Chaping	(teaps	
	Trainc Snaping		
	Average Bandwidth:	N/A	
	Peak Bandwidth:	N/A	
Add Edit Remove	Burst Size:	N/A	
	Esilouse and Load Palan	rina	
		Close	Help

Figure A.31 : activation de vMotion, ouverture des paramètres du VMkernel

Sur la ligne « vMotion », cochez la case à côté de « Enabled », puis cliquez successivement sur *Ok* et *Close*.

Port Properties	ig NIC Teaming
Network Label:	VMkernel ISCST
VLAN ID (Optional):	None
VMotion:	Frabled
IP Settings	
IP Address:	10 . 1 . 7 . 22
Subnet Mask:	255 , 255 , 0 , 0
VMkernel Default Gateway:	10 , 1 , 0 , 1 Edit

Figure A.32 : activation de vMotion, activation effective

Source : http://communities.vmware.com/thread/62993

A.6.7 Déplacer une VM à l'aide de vMotion

Avant toutes choses, pour un échange de VM via vMotion, il faut que la configuration des vSwitch de départ et de destination soit la même (vMotion activé, le réseau de la VM doit aussi être sur le vSwitch de destination, règles de sécurité identiques...). Si toutes ces règles sont respectées, vous pouvez passer aux étapes suivantes.

Sur le panneau à gauche de la fenêtre principale du VI Client, cliquez sur la VM à déplacer. Maintenez le bouton enfoncé, puis déplacez la VM sur la machine hôte de destination. Une fois sur cette dernière, relâchez le bouton.



Figure A.33 : déplacement avec vMotion, choix du déplacement

Il est possible qu'une fenêtre contenant un ou plusieurs warnings apparaisse. Dans la plupart des cas elle apparaît si la vitesse du réseau est inférieure à 1Gbps ou qu'une légère différence de configuration est détectée entre la machine source et la machine de destination. Si la différence de configuration est trop importante, la fenêtre contiendra des erreurs et il vous sera impossible de passer à la suite des étapes. Dans ce cas, vérifiez que les points évoqués au début de ce sous-chapitre sont bien respectées. Si la fenêtre ne contient que des warnings, cliquez sur *Yes* pour passer à l'étape suivante.

Errors	and Warnings 🛛 🛛 🔀		
	ubuntu_i		
A	Migration from 10.1.7.2: No guest OS heartbeats are being received. Either the guest OS is not responding or VMware tools is not configured properly.		
	Migration from 10.1.7.2 to 10.1.2.153: VMotion interface 'Management Network' on the source host '10.1.7.2' does not have the recommended capacity (full duplex / 1000 Mbps) to properly support VMotion.		
Δ	Migration from 10.1.7.2 to 10.1.2.153: The VMotion interface of the destination host uses network 'VMkernel', which differs from the network 'Management Network' used by the VMotion interface of the source host.		
	Migration from 10.1.7.2 to 10.1.2.153: VMotion interface 'VMkernel' on the destination host '10.1.2.153' does not have the recommended capacity (full duplex / 1000 Mbps) to properly support VMotion.		
	Are you sure you want to continue?		
	Yes No		

Figure A.34 : déplacement avec vMotion, warnings

Il existe deux différentes priorités lors d'un transfert : « high » et « low ».

La priorité « high » permet un transfert dit « sans perte de service », mais n'autorise le transfert que si la machine de destination possède suffisamment de ressources pour accueillir la VM sans que cette dernière ne subisse de perte de ressources.

La priorité « low » garanti que le transfert aura bien lieu, peu importe l'état de l'hôte de destination, cependant, si cet hôte ne possède pas suffisamment de ressources, la VM sera temporairement inaccessible, impliquant ainsi une interruption de service.

Choisissez « High Priority », puis cliquez sur Next>.

🕝 Migrate Virtual Machin	ne Wizard 💷 🗆 🖂 🔤 🔀
Migration Priority Set the priority of the m	igration, relative to other operations on destination host.
Migration Priority Ready to Complete	 High Priority Reserve resources on both the source and destination hosts to maintain virtual machine availability during the migration. High priority operations will not proceed if the resources are unavailable. Low Priority Low priority operations will always proceed, but the virtual machine may become briefly unavailable if sufficient host resources are unavailable.
Help	≤Back Next ≥ Cancel

Figure A.35 : déplacement avec vMotion, choix de la priorité

Vous êtes maintenant face à un récapitulatif des paramètres choisis pour le transfert. Cliquez sur *Finish* pour lancer le transfert.

🕝 Migrate Virtual Machin	e Wizard	_ 🗆 🔀
Summary Review this summary bef	ore completing the wizard.	
Migration Priority Ready to Complete	Review this summary before finishing the wizard. Host: 10.1.2.153 ResourcePool: Resources DataStore: Current Location Migration Priority: High Priority	
Help		Cancel

Figure A.36 : déplacement avec vMotion, validation du transfert

Vous pouvez suivre l'avancement du transfert dans la partie basse de la fenêtre principale du VI Client.

Recent Tasks						
Name	Target	Status	Initiated by	▽ Time	Start Time	Complete Time
🌮 Migrate Virtual Machine	👜 ubuntu_i	10% 🚺 🗌 🗌	Administrator	07.05.2009 08:23:44	07.05.2009 08:23:44	
🎽 Migrate Virtual Machine	🛍 ubuntu_i	🥝 Completed	Administrator	07.05.2009 08:16:48	07.05.2009 08:16:48	07.05.2009 08:17:07
🖄 Migrate Virtual Machine	💼 ubuntu_i	🥝 Completed	Administrator	07.05.2009 08:15:24	07.05.2009 08:15:24	07.05.2009 08:15:46
🕗 Update Virtual Switch	10.1.2.154	🥝 Completed	Administrator	07.05.2009 08:15:15	07.05.2009 08:15:15	07.05.2009 08:15:17
🚰 Tasks 🞯 Alarms						

Figure A.37 : déplacement avec vMotion, suivit du transfert

A.6.8 Création d'un cluster DRS

Afin de pouvoir utiliser DRS, il est nécessaire que les VMs soient sur un vSwitch qui gère le vMotion (ce qui est déjà le cas si vous avez suivi les étapes de configuration précédentes) et que les serveurs ESX/ESXi soient dans un même cluster.

Dans le panneau à gauche du VI Client, faites un clic droit sur le datacenter dans lequel vous voulez ajouter le cluster, puis cliquez sur *New Cluster...*



Figure A.38 : lancement de la création d'un cluster

Dans la fenêtre qui s'ouvre, choisissez un nom pour votre cluster (ici, j'ai choisi de l'appeler DRS) et cochez la case à côté de « VMware DRS ». Cliquez sur *Next>*.

PNew Cluster Wizard	
Cluster Features What features do you want	to enable for this cluster?
Cluster Features VMware DRS VM Swapfile Location Ready to Complete	DR5
	Cluster Features Select the features you would like to use with this cluster. VMware HA Wmware DRS Wmware DRS Wmware DRS enables VirtualCenter Server to manage the assignment of virtual machines to hosts automatically, suggesting placement when virtual machines are powered on, and migrating running virtual machines to balance load and enforce resource allocation policies.
Help	≤Back Next ≥ Cancel

Figure A.39 : choix du nom du cluster et activation de DRS

Sur cet écran, il est possible de configurer l'attitude de DRS (déplacement automatique ou simple suggestion), ainsi que la sensibilité à la différence de charge RAM et CPU déclenchant un déplacement (Conservative = peu sensible et Aggressive = très sensible, uniquement utile en mode automatique). Pour le moment, laissez les paramètres par défaut, il est possible d'y revenir plus tard, c'est ce que je vais faire pour les scénarios de la partie 3. Cliquez sur *Next*>.

🕝 New Cluster Wizard	
VMware DRS What level of automation do y	vou want this cluster to use?
Cluster Features YMware DRS VM Swapfile Location Ready to Complete	Automation level Automation level Automation level Automation level Automation level Automation Automation
Help	≤Back Next ≥ Cancel

Figure A.40 : réglage du mode et de la sensibilité de DRS

Il est possible de séparer les swapfiles des disques durs virtuels en les stockant sur un autre disque iSCSI (ou une autre partition du même disque). Cette fonctionnalité n'est pas activée par défaut et n'est pas recommandée lors de l'utilistation de vMotion, et donc, celle de DRS. Laissez donc le choix par défaut (le premier) et cliquez sur *Next>*.

🖉 New Cluster Wizard	
Virtual Machine Swapfile Location Which swapfile location policy should	d virtual machines use while in this cluster?
Cluster Features YMware DRS YM Swapfile Location Ready to Complete	Swapfile Policy for Virtual Machines Store the swapfile in the same directory as the virtual machine. Recommended Store the swapfile in the datastore specified by the host. If not possible, store the swapfile in the same directory as the virtual machine. A host specified datastore may degrade VMotion performance for the affected virtual machines.
Help	_≤Back Next ≥ Cancel

Figure A.41 : choix de l'emplacement du swapfile

La configuration est terminée, cet écran présente un récapitulatif des paramètres choisis. Cliquez sur *Finish*.

🕝 New Cluster Wizard			_ 🗆 🗙
Ready to Complete Review the options you have selected	d for this cluster and click Finish to c	omplete.	
Cluster Features VMware DRS VM Swapfile Location Ready to Complete	The cluster will be created with the Cluster Name: VMware DRS i VMware DRS migration threshold: Virtual Machine Swapfile Location:	following options: DRS Enabled Fully Automated Apply recommendations with three or more stars. Same directory as the virtual machine	
Help		<u>≤</u> Back <u>Fi</u> nish	Cancel

Figure A.42 : récapitulatif de la création du cluster

Une fois la création effectuée, le cluster apparaît dans le panneau gauche du VI Client. Cependant, ce nouveau cluster est vide :



Figure A.43 : cluster dans le VI Client juste après sa création

Il faut donc déplacer les serveurs ESX dans ce cluster. Pour ce faire, cliquez en maintenant le bouton enfoncé sur le serveur que vous désirez ajouter à votre cluster, et déplacez-le sur le cluster, puis relâchez le bouton. Une nouvelle fenêtre s'ouvre. Elle vous demande de choisir si toutes les VMs de ce serveur doivent être déplacées dans le cluster, ou si vous désirez créer un « ressource pool » sur le serveur qui sera dans le cluster. Les ressources pools sont des groupes de VMs auxquels on peut attribuer des règles spécifiques. Ce choix est à faire selon votre configuration. Dans mon cas, je n'ai pas créé de ressource pool et je ne compte pas le faire par la suite, je laisse donc le choix par défaut (le premier). Cliquez sur *Next*>.

Add Host Wizard Choose the Destination Re Choose where to place this	source Pool host's virtual machines in the resource pool hierarchy.
Choose Resource Pool Ready to Complete	Virtual Machine Resources What would you like to do with the virtual machines and resource pools for this Put all of this host's virtual machines in the cluster's root resource pool. Resource pools currently present on the host will be deleted. Create a new resource pool for this host's virtual machines and resource pools. This preserves the host's current resource pool hierarchy. Name: Gratted from 10.1.2.153
Help	≤ Back Next ≥ Cancel

Figure A.44 : ajout d'une VM au cluster, choix de la configuration des ressources pools

Le nouvel écran vous indique dans quel cluster votre server va être déplacé. Cliquez sur Finish.

🖉 Add Host Wizard		
Ready to Complete Review the options you have selected and click Finish to add the host.		
Choose Resource Pool Ready to Complete	Review this summary before finishing the wizard.	
	Resources Destination: DR5	
Help		

Figure A.45 : ajout d'une VM au cluster, récapitulatif

Répétez l'opération pour tous les serveurs que vous désirez joindre au cluster. Une fois fini, dans le panneau de gauche du VI Client, la hiérarchie à changer. Toutes les VMs et tous les serveurs sont au même niveau, juste en dessous du cluster :

🖃 📂 Hosts & Clusters			
🖃 📑 New Datacenter			
🗆 💼 DF	25		
	10.1.2.153		
	10.1.2.154		
	Client2		
	Observateur2		
	Serveur153		

Figure A.46 : panneau de gauche du VI Client une fois les serveurs dans le cluster

A.7 Tests du switch NETGEAR

A.7.1 Schémas et plans du réseau

A.7.2 Méthodologie

- Un poste sous Windows XP (Wireshark 1 lors des scénarios) télécharge une image 2.86 GB sur un disque iSCSI émulé via FreeNAS (disque iSCSI Openfiler lors des scénarios).
- Une première série de 10 mesures est effectuée alors que les 2 postes sont reliés directement entre eux.
- Une seconde série de 10 mesures est effectuée, mais en passant par le switch, afin de voir si ce dernier ralenti la transaction.
- On remplace le serveur FreeNAS par un serveur Openfiler, et on reprend quelques mesures avec le switch.
- Les mesures sont effectuées à la main à l'aide d'un chronomètre car le fichier copié est trop gros pour une analyse Wireshark, ce qui génère un arrêt prématuré de ce dernier avec message d'erreur « Out of Memory ».

A.7.3 Mesures

Sans le switch, les temps de téléchargement de l'image varient entre **60 et 65 secondes**, ce qui donne des vitesses de transfert comprises entre **45.1 et 48.8 MBps**.

Avec le switch, les temps sont dans les mêmes valeurs, le switch ne ralenti donc pas le transfert.

En remplaçant FreeNAS par Openfiler, le temps de transfert devient bien plus important. Pour le même fichier, il aura fallu plus de **12 minutes et 20 secondes de copie**, soit un débit de **3.86 MBps**. J'en conclus donc que les temps mesurés sur le réseau gigabit (scénarios 2 et 4) sont plus élevés que ce que l'on pourrait attendre d'un tel réseau.

A.8 Script

Afin de pouvoir observer combien de temps la VM déplacée est indisponible, il était nécessaire que cette dernière envoie régulièrement des paquets sur le réseau. Le système le plus simple est de lui faire envoyer des requêtes ICMP en boucle, cette fonctionnalité étant disponible de base. Cependant, le temps entre les envois ainsi que le temps d'attente des réponses, même réglés au minimum, étaient trop importants pour que les mesures soient exploitables.

Comme mon serveurs virtuel Ubuntu a un serveur apache d'installé, j'ai décidé de créer un script qui irait en boucle télécharger une page web sur ce serveur. Les requêtes engendrent un échange de paquets pour le téléchargement entier de la page, et la requête suivante est lancée immédiatement après la fin de cet échange.

Le temps entre 2 requêtes est de 5.715 millisecondes, soit environ 175 requêtes/seconde.

```
#!/bin/bash
# auteur : Sandmeier Loïc
# date de création : 10.05.2009
# date de modification : 10.05.2009
# but : envoyer en boucle des requêtes wget à un serveur web afin de faire
        des mesures de disponibilités
echo "Ce script va envoyer en boucle des requêtes wget à un serveur
cible."
echo "Pour l'arreter, appyuez sur [ctrl]+c"
echo -n "Quelle est l'adresse ip du serveur cible? "
read IP CIBLE
echo -n "Quel est le nom de la page/du fichier à télécharger? "
read NOM CIBLE
while [ '1' ]
do
       wget $IP CIBLE/$NOM CIBLE
       rm $NOM CIBLE
done
exit 0
```

A.9 Problèmes rencontrés et bugs

A.9.1 Bug de déconnection aléatoire

A.9.2.1 Symptômes

Ce bug est présent dans la version 2.5 update 3 et 4 du VI Server. Dans l'update 3, des déconnections arrivent à des moments vraiment aléatoires, alors que ces dernières arrivent environ 2 minutes après le démarrage du VI Server dans l'update 4.

A.9.2.2 Cause

Il semblerait que la cause de ce bug provient de l'écoute des ports par le VI Server qui passerai certains ports de connexion en écoute uniquement de l'adresse IP 127.0.0.1, rendant impossible la connexion des serveurs ESX et ESXi distant.

A.9.2.3 Solution

Pour l'heure, aucun correctif n'est proposé. Dans certains cas, une réinstallation du VI Server suffit, dans d'autres, le problème réapparaît après une réinstallation. Ce bug est très ennuyant, car malgré le fait que les machines restent disponibles et que les VMs continuent de fonctionner normalement, beaucoup de fonctions deviennent inutilisables, comme les transferts via vMotion qui prennent souvent plus de 2 minutes et nécessitent une supervision de la part du VI Server.

Source : <u>http://communities.vmware.com/thread/191265</u>

A.9.2 Problème d'effacement des VMs

A.9.2.1 Symptômes

Lors de suppressions de VMs stockée sur un périphérique iSCSI, il arrive que la machine soit effectivement supprimée, mais que l'espace disque correspondant ne soit pas libéré.

Certaines fois, en parcourant le contenu du disque, on peut voir que des fichiers et des dossiers restent. Il est alors parfois possible de les supprimer manuellement, mais le reste du temps, une erreur apparaît lors de la tentative de suppression.

Dans le reste des cas, les fichiers ne sont pas visibles lors d'un parcours manuel du disque, il est donc impossible de les effacer.

A.9.2.3 Solutions

Une solution efficace mais lourde consiste à formater le disque en question. Avant de formater, pour conserver les machines, on crée des « appliances », que l'on réinstalle sur le disque une fois ce dernier reformaté. Cependant, cette solution implique une interruption de service pouvant se révéler longue, car il faut compter le temps de création des appliances, celui du formatage et de la reconversion des appliances, les transferts des appliances étant en lui-même déjà conséquent.

Une autre solution permettant de minimiser le temps d'indisponibilité consiste à stocker les VMs sur un second disque connecté aux mêmes hôtes que celui qu'il faut formater. Les VMs pourront continuer de fonctionner sur ce second disque durant le formatage du premier, le temps d'indisponibilité se limitant au temps de transfert, on gagne la durée du formatage en disponibilité de service.

A.9.3 Problème lors d'usage simultané de plusieurs VMs

A.9.3.1 Symptômes

Le stockage des VMs se fait sur un disque iSCSI commun à tous les hôtes. N'ayant pas de vrai disque iSCSI, j'utilise un disque émulé via FreeNAS, une version allégée de FreeBSD, qui se contrôle via une interface web. Tant que je ne fais tourner qu'une VM à la fois, tout fonctionne très bien (démarrage, arrêt, redémarrage, fonctionnement général, etc.). Cependant, si je désire faire tourner plusieurs VMs en même temps, plus rien ne va. Que la VM soit une distribution GNU/Linux ou une version de Windows, le problème est le même : impossible d'utiliser la ou les machines. Soit j'obtiens une erreur directement au lancement dans le VI Client, soit la machine amorce son démarrage et se fige sur un écran ou retourne une erreur :



Figure A.47 : problème de démarrage, ligne ou bloque Ubuntu



Figure A.48 : problème de démarrage, erreur système de Windows

J'ai également eu quelques « kernel panic » avec les VMs Ubuntu

A.9.3.2 Solutions et causes

Ne trouvant aucune solution, j'ai ouvert une discussion sur le forum *VMware Communities*, dans laquelle il m'a été suggéré de remplacer FreeNAS par Openfiler, une distribution GNU/Linux permettant également d'émuler un disque iSCSI. Après avoir effectué ce remplacement, le problème ne s'est plus manifesté. Il semblerait donc que ce soit FreeNAS qui ne supporte pas les accès de plusieurs clients simultanément sur la même cible iSCSI.

Source : http://communities.vmware.com/message/1254669

A.9.4 Erreur lors d'accès à certaines pages de l'interface web d'Openfiler

A.9.4.1 Symptômes

Lorsque je tentais d'accéder à la page « volumes » de l'interface web d'Openfiler, la page tentait indéfiniment de se charger. Si je regardais directement la console d'Openfiler durant ces tentatives d'ouverture, 2 messages apparaissaient en boucle :

- end_request: I/O error, dev fd0, sector 0
- end_request: I/O error, dev fd0, sector 0

A.9.4.2 Cause et solution

Une recherche sur Google m'a conduit sur le blog d'une personne ayant rencontré le même problème : le pc faisant tourner Openfiler n'a pas de lecteur disquette, la recherche des espaces de stockages bloque car elle n'arrive pas à accéder à ce dernier, ce qui est normal. En suivant un lien fourni dans ce blog, il est indiqué que la commande *modprobe –r floppy* corrige ce problème, ce qui fut effectivement le cas. L'effet de cette commande est la désactivation du lecteur disquettes.

Sources : <u>http://michaelellerbeck.com/2009/03/27/installing-openfiler-or-in-otherwords-how-i-lost-two-hours-of-my-life-to-a-floppy-drive/</u> (le blog)

http://www.cyberciti.biz/faq/linux-end_request-ioerror-dev-fd0-sector0/