

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE  
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET LA RECHERCHE  
SCIENTIFIQUE

.....  
Université Mouloud Mammeri de Tizi-Ouzou Algérie  
Faculté de génie électrique et d'informatique  
Département d'informatique  
Master 2, RMSE



Mémoire de fin d'études en vue de l'obtention de diplôme de :

## Master en Informatique

Spécialité : Réseaux, Mobilités et Systèmes embarqués

### Thème

# LA MIGRATION EN TEMPS REEL DES MACHINES VIRTUELLES DANS LE CLOUD COMPUTING

Présenté par :

Mlle. MANIRAKIZA Nina Marlène

&

Mr. RAKOTONIRINA Zarasoa Ernest

Encadré par :

Mme. LAGAB Aoudjit

Devant les membres de jury :

Présidente : Mme. BELKADI

Examinatrice : Mme. HADAOUI

Examineur : Mr. DAOUI MEHAMMED

PROMOTION : 2016–2017

# Remerciements

*Nous remercions Le Dieu tout puissant de nous avoir donné la force et la capacité qui nous ont menées à ce niveau.*

*Nos plus vifs remerciements vont à notre Promotrice, Madame AOUDJIT LAGAB pour nous avoir encadrés et guidés tout au long de notre projet, pour ses conseils judicieux et minutieusement prodigués, pour sa disponibilité ainsi que ses critiques constructives.*

*Nous tenons à exprimer notre gratitude aux membres du jury pour avoir accepté d'honorer par leur jugement notre travail.*

*Nous tenons aussi à exprimer notre profonde reconnaissance à l'égard de ceux qui, par leurs apports multiformes ont contribué à l'aboutissement et à la réussite de ce travail.*

*M. Nina Marlène & R. Zarasoa Ernest*

## *Dédicaces*

\*\*\*\*\*

*Je dédie ce travail à mes très chers parents qui n'ont jamais cessé de me montrer amour et soutien tout au long de ma vie.*

*A mes sœurs pour leur soutien infaillible*

*A mon binôme pour sa bonne collaboration et son dévouement*

*Aux Pères Blancs de Tizi-Ouzou qui ont mis à notre disposition des moyens matériels nécessaires à la réussite de notre mémoire*

*A mes amis(es) qui n'ont jamais cessé de croire en mes capacités et pour leurs encouragements.*

*M. Nina Marlène*

*Pour ma famille d'Algérie : ma binôme, mes amis, mes compatriotes, la communauté pères blancs de Tizi-Ouzou qui m'ont toujours soutenu.*

*Pour mes parents, mes sœurs et mon frère, mes cousines, mes nièces et neveux, sans vos appuis et prières je ne serai pas là où je suis actuellement.*

*Zaraso.*

# Sommaire :

|  |           |
|--|-----------|
| <b>SOMMAIRE :</b> .....  | <b>IV</b> |
| <b>INTRODUCTION GENERALE</b> .....   | <b>1</b>  |
| <b>CHAPITRE I : VIRTUALISATION ET CLOUD COMPUTING</b> .....  | <b>3</b>  |
| I.1. INTRODUCTION .....  | 3         |
| I.2. LA VIRTUALISATION .....   | 3         |
| I.2.1. Historique de la virtualisation .....   | 3         |
| I.2.2. Définition de la virtualisation .....   | 3         |
| I.2.2.1. La virtualisation des applications. ....  | 4         |
| I.2.2.2. La virtualisation des réseaux. ....   | 4         |
| I.2.2.3. La virtualisation du stockage. ....   | 4         |
| I.2.2.4. La virtualisation des postes de travail. ....   | 5         |
| I.2.2.5. La virtualisation des systèmes d'exploitation. ....   | 5         |
| I.2.3. Les types de virtualisation [1.4] .....   | 5         |
| I.2.3.2. La virtualisation complète .....  | 5         |
| I.2.3.1. La paravirtualisation .....   | 6         |
| I.2.3.3. L'émulation en informatique .....   | 7         |
| I.2.3.4. La simulation .....   | 7         |
| I.2.4. Les Techniques de virtualisation [1.3] .....  | 7         |
| I.2.4.1. La virtualisation au niveau d'une application. ....   | 8         |
| I.2.4.2. La virtualisation au niveau du noyau. ....  | 8         |
| I.2.4.3. La virtualisation par l'emploi d'un émulateur .....   | 9         |
| I.2.4.4. La virtualisation par l'utilisation d'un moniteur ou hyperviseur de machines virtuelles. .... | 9         |
| I.2.4.5. La virtualisation au niveau du processeur. ....   | 10        |
| I.2.5. Les avantages de la virtualisation. ....  | 10        |
| I.2.5.1. Le temps moyen de restauration .....  | 10        |
| I.2.5.2. L'infrastructure vieillissante .....  | 10        |
| I.2.5.3. Les capacités physiques insuffisantes .....   | 11        |
| I.2.5.4. La sous-utilisation .....   | 11        |
| I.3. LE CLOUD COMPUTING .....  | 11        |
| I.3.1. Introduction .....  | 11        |
| I.3.1.1. Avant le cloud computing .....  | 11        |
| I.3.1.2. Avec le cloud computing .....   | 12        |
| I.3.2. Historique .....  | 12        |
| I.3.3. Définitions .....   | 13        |
| I.3.4. Les caractéristiques principales .....  | 13        |
| I.3.5. Les modèles de service .....  | 14        |
| I.3.5.1. Le Software as a Service (SaaS) .....   | 14        |
| I.3.5.2. La Platform as a Service (PaaS) .....   | 15        |
| I.3.5.3. L'Infrastructure as a Service (IaaS) .....  | 17        |
| I.3.5.4. La différence entre l'IaaS, le PaaS et le SaaS .....  | 18        |
| I.3.6. Les modèles de déploiement du cloud computing .....   | 19        |
| I.3.6.1. Le Cloud Public .....   | 19        |
| I.3.6.2. Le Cloud Privé .....  | 21        |
| I.3.6.3. Le Cloud Hybride .....  | 23        |
| I.3.6.4. Le Cloud Communautaire .....  | 23        |
| I.3.7. Les avantages et les inconvénients du cloud computing .....                                     | 23        |
| I.3.7.1. Les avantages .....   | 23        |
| I.3.7.2. Les inconvénients .....   | 25        |
| I.3.8. Les risques du Cloud Computing .....  | 27        |
| I.4. CONCLUSION .....  | 28        |
| <b>CHAPITRE II : L'ALLOCATION DE RESSOURCES ET LA MIGRATION DES VMS</b> .....                          | <b>30</b> |

|   |  |           |
|---|--|-----------|
| II.1.   | INTRODUCTION.....  | 30        |
| II.2.   | L'ALLOCATION DES RESSOURCES DANS LE CLOUD .....                  | 30        |
| II.2.1.   | Présentation du contexte .....                                   | 31        |
| II.2.2.   | Définitions .....  | 32        |
| II.2.3.   | Allocation des VMs .....   | 33        |
| II.2.4.   | Techniques d'allocation des ressources .....                     | 33        |
| II.2.4.1.   | Allocation statique des ressources .....                         | 33        |
| II.2.4.2.   | Allocation dynamique des ressources.....                         | 33        |
| II.3.   | LA CONSOMMATION ENERGETIQUE AU NIVEAU DU CLOUD .....             | 33        |
| II.3.1.   | Le Datacenter.....   | 34        |
| II.3.2.   | La consommation d'énergie dans le data center .....              | 34        |
| II.4.   | LA MIGRATION DES VMs.....  | 35        |
| II.3.1.   | Les avantages et les inconvénients de la migration des VMs ..... | 35        |
| II.3.1.1.   | Les avantages .....  | 35        |
| II.3.1.2.   | Les inconvénients .....  | 36        |
| II.3.2.   | Les différentes méthodes de migration des VMs .....              | 37        |
| II.3.2.1.   | La migration à froid.....  | 37        |
| II.3.2.2.   | La migration à chaud ou la migration en temps réel.....          | 37        |
| II.3.2.3.   | Amélioration du processus de migration de VMs .....              | 42        |
| II.3.3.   | Les modèles de performance de migration à chaud .....            | 43        |
| II.3.3.1.   | Modèles linéaires.....   | 43        |
| II.3.3.2.   | Fréquence d'écriture des pages mémoires .....                    | 43        |
| II.5.   | LES DIFFERENTES APPROCHES DE MIGRATION DE VMs .....              | 44        |
| II.5.1.   | L'approche basée sur la qualité de service.....                  | 44        |
| II.5.2.   | L'approche basée sur la consommation énergétique.....            | 44        |
| II.5.2.1.   | La détection de la surcharge de l'hôte .....                     | 45        |
| II.5.2.2.   | La détection de la Charge réduite de l'hôte .....                | 45        |
| II.5.2.3.   | La sélection de la machine virtuelle et la migration .....       | 45        |
| II.5.3.   | Quelques travaux et algorithmes sur la migration de VMs .....    | 45        |
| II.6.   | CONCLUSION .....   | 47        |
| <b>CHAPITRE III : DESCRIPTION ET CONCEPTION DE L'APPROCHE PROPOSEE.....</b> |  | <b>49</b> |
| III.1.  | INTRODUCTION.....  | 49        |
| III.2.  | L'APPROCHE « THRESHOLD » .....                                   | 49        |
| III.2.1.  | Calcul de la consommation d'énergie .....                        | 49        |
| III.2.2.  | La phase de sélection .....                                      | 50        |
| II.2.2.1.   | Le seuil supérieur .....   | 50        |
| II.2.2.2.   | Le seuil inférieur.....  | 51        |
| II.2.2.3.   | L'algorithme et le diagramme d'activité .....                    | 51        |
| II.2.3.   | La phase d'allocation .....                                      | 53        |
| II.2.3.1.   | Le principe de l'algorithme Best Fit Decreasing (BFD).....       | 53        |
| II.2.3.2.   | L'algorithme et le diagramme d'activité .....                    | 54        |
| III.3.  | ARCHITECTURE .....   | 55        |
| III.4.  | CONCLUSION .....   | 57        |
| <b>CHAPITRE IV. IMPLEMENTATION ET RESULTAT .....</b>                        |  | <b>59</b> |
| IV.1.   | INTRODUCTION.....  | 59        |
| IV.2.   | OUTILS DE TRAVAIL .....  | 59        |
| IV.2.1.   | Le langage de programmation Java .....                           | 59        |
| IV.2.2.   | L'environnement de développement Eclipse .....                   | 60        |
| IV.2.3.   | Le simulateur CloudSim.....                                      | 60        |
| IV.2.3.1.   | Définition.....  | 60        |
| IV.2.3.2.   | L'architecture du CloudSim .....                                 | 61        |
| IV.2.3.3.   | Les classes principaux du CloudSim .....                         | 62        |
| IV.2.3.4.   | L'installation .....   | 64        |
| IV.3.   | TEST.....  | 68        |

|  |           |
|--|-----------|
| IV.4. CONCLUSION .....                           | 72        |
| <b>CONCLUSION GENERALE ET PERSPECTIVES .....</b> | <b>73</b> |
| <b>LISTE DES FIGURES : .....</b>                 | <b>75</b> |
| <b>LISTE DES ABREVIATIONS ET SIGLES.....</b>     | <b>76</b> |
| <b>SOURCES : .....</b>                           | <b>77</b> |

# Introduction générale

Ces dernières années, des investissements considérables ont été réalisés dans les technologies de l'information et de la communication (TIC). Le web est aujourd'hui l'un des piliers de ces investissements.

Initialement utilisé en tant que simple moyen de partage d'information, le Web devient aujourd'hui un outil indispensable pour mieux gérer et organiser les différentes missions d'une entreprise. En effet, le contexte économique incite de plus en plus les entreprises aux fusions, acquisitions et à l'externalisation de nombreuses activités. Cette politique d'ouverture porte le nom d'entreprise étendue. De plus, ces dernières sont confrontées aux problèmes de communications entre les systèmes d'information distribués et hétérogènes.

Pour pallier à ce problème, les chercheurs ont mis en place le Cloud computing (ou informatique dans les nuages) qui est l'une des technologies spécifiée pour fournir un service au lieu d'un produit. Des cervidés comme le calcul, le logiciel, les applications, l'accès aux données et le stockage sont fournis à un utilisateur sans que ce dernier ne connaisse exactement l'emplacement du serveur fournissant ce service. Les Datacenters du cloud contenant les serveurs et d'autres équipements réseaux ainsi que les refroidisseurs utilisent de grandes quantités d'énergie.

## Problématique

La demande croissante d'infrastructure dans les environnements de cloud computing augmente l'énergie consommée dans les Datacenters. On se retrouve ainsi avec une mauvaise distribution de machines virtuelles opérantes dans les différents Datacenter. Certains d'entre ces derniers étant surutilisés ce qui aboutit à une très grande consommation d'énergie et à la dégradation continue du matériel, d'autres moyennement utilisés tandis que dans d'autres encore on remarque une sous-utilisation qui aboutit à un gaspillage d'énergie : le Datacenter est allumé alors qu'il n'a pas une grande charge à traiter

## Solution proposée

Dans ce travail, nous avons fait une étude approfondie sur une approche de la migration de machines virtuelles dans les Datacenter.

Nous inspirant de travaux connexes, nous avons étudié une technique d'optimisation de l'énergie consommée dans les Datacenter. Cette technique appelée « Threshold » se base sur des seuils de consommation énergétique dans les CPU et comprend trois phases à savoir la phase de sélection de machine virtuelle, la phase d'allocation et pour finir la phase de migration de VMs. Cette approche permet de repérer les Datacenter surutilisé et de migrer certaines de leurs machines virtuelles vers d'autres Datacenter.

---

# CHAPITRE 1 :

---

# VIRTUALISATION ET CLOUD COMPUTING

# Chapitre I : Virtualisation et Cloud Computing

## I.1. Introduction

Alors que les données informatiques augmentent de façon exponentielle, et que les entreprises font de plus en plus appel aux processus informatiques pour gagner en productivité et en compétitivité, la possible réduction des coûts de gestion des infrastructures informatiques est une des principales priorités des entreprises.

Ces dernières années, plusieurs moyens sont apparus pour aborder cette réduction des coûts, parmi lesquels, la virtualisation, et le Cloud Computing. La virtualisation et le Cloud Computing sont deux concepts différents, mais pourtant complémentaires. Le cloud computing est le fruit des évolutions récentes des technologies de l'information, notamment la virtualisation, et de l'augmentation de la capacité des réseaux. Si le cloud utilise en général les technologies de virtualisation, il incarne surtout une nouvelle approche des infrastructures matérielles et logicielles. Nous allons dans ce chapitre nous attacher à vous présenter et vous définir ces deux notions, ainsi que leurs avantages et les usages possibles pour les entreprises. Nous allons d'abord aborder la virtualisation, nous allons parler de son historique, de sa définition, des types, techniques ainsi que ses avantages. Deuxièmement, on va présenter le cloud computing : nous allons parler de son introduction, suivi de son historique puis de sa définition, ses caractéristiques principales, de ses modèles de service ainsi que de déploiement, et à la fin de ses avantages et inconvénients et sans oublier ses risques.

## I.2. La virtualisation

### I.2.1. Historique de la virtualisation

La virtualisation, contrairement à ce que pensent de nombreuses personnes, n'est pas née en 1999 avec la sortie du premier produit VMware. Ses débuts remontent à environ 40 ans sur la plate-forme de superordinateurs (mainframe) d'IBM ; les machines virtuelles s'appelaient alors des « pseudo-machines ». À l'origine, l'ordinateur central utilisait un programme de contrôle pour allouer des ressources et isoler les différentes instances des pseudo-machines les unes des autres [1.1].

La version contemporaine du programme de contrôle s'appelle un « hyperviseur ».

Un hyperviseur est un logiciel qui constitue une couche d'abstraction entre la machine virtuelle (système invité) et la plateforme sous laquelle elle tourne (s'exécute) qui est alors appelée « système hôte ».

### I.2.2. Définition de la virtualisation

La définition formelle de la virtualisation fait référence à l'abstraction physique des ressources informatiques [1.2].

Sont donc candidats à la virtualisation tout ce qui utilise la charge matérielle disponible peut être virtualisé avec succès :

- Les serveurs web
- Les serveurs de courrier électronique

- Les autres serveurs réseau (DNS, DHCP, NTP...)
- Les serveurs d'applications
- Les serveurs de bases de données

Nombreuses sont aussi les applications de la virtualisation [1.3] :

- Virtualisation des applications
- Virtualisation des réseaux
- Virtualisation du stockage
- Virtualisation des postes de travail
- Virtualisation des systèmes d'exploitation et des services (serveurs)

### **I.2.2.1. La virtualisation des applications.**

La virtualisation d'application implique de pouvoir exécuter un logiciel sans toutefois l'installer physiquement sur le système auquel l'utilisateur est connecté, avec tout ce que cela implique en termes d'économies (de processus, de déploiement, d'altération, de mise à jour, de test, de compatibilité, etc.). La virtualisation d'application correspond à ce que l'on appelle «client léger».

### **I.2.2.2. La virtualisation des réseaux.**

Il s'agit du domaine le plus ambigu et le plus spécifique des approches de virtualisation. Il comprend les tables de routage, les réseaux locaux virtuels (VLAN), VPN etc. L'exemple le plus simple de virtualisation de réseau est le VLAN, réseau informatique logique indépendant. De nombreux VLANs peuvent coexister sur un même commutateur réseau.

L'introduction des groupes réseaux constitue le cœur de la virtualisation des réseaux. Ces groupes sont représentés par un ensemble d'utilisateurs qui ont le même profil et partagent les mêmes ressources. Les connexions correspondantes forment un groupe auquel on peut attribuer des privilèges et des autorisations sur un ensemble de ressources ; cette méthode de gestion contribue énormément dans l'implémentation de la sécurité.

### **I.2.2.3. La virtualisation du stockage.**

C'est une forme de virtualisation qui a fait ses preuves, elle se compose en deux catégories principales : la virtualisation de blocs et la virtualisation de fichiers. La virtualisation de blocs correspond plus précisément aux technologies de réseau de stockage NAS (Network Attached Storage), la virtualisation de fichiers correspond aux réseaux SAN (Storage Area Network).

Au niveau du NAS, la virtualisation de l'espace de stockage se rencontre d'abord dans les unités de stockage, avec l'introduction il y a plusieurs années, de la première forme de virtualisation du stockage est sous le nom RAID. Le RAID (Redundant Array of Independant Disks) constitue une couche d'abstraction matérielle et logicielle qui se trouve entre le système d'exploitation et les disques durs ; ainsi, le système ne voit pas un ensemble de disques, mais un disque logique ou « Virtuel ».

La virtualisation de fichiers rend la couche virtuelle plus accessible à l'échelle de l'utilisateur. La plupart des technologies de virtualisation de fichiers sont associées à des réseaux de stockage (SAN) et permettent de suivre la localisation et la répartition des fichiers et répertoires dans le dispositif de stockage.

#### **I.2.2.4. La virtualisation des postes de travail.**

La virtualisation de poste de travail est une approche qui consiste à fournir aux utilisateurs leurs environnements de travail à partir de n'importe quel emplacement dans le réseau, et en utilisant des moyens différents (ordinateur, terminal, téléphone...). Ainsi les données de l'utilisateur sont à l'abri des failles logicielles et matérielles.

#### **I.2.2.5. La virtualisation des systèmes d'exploitation.**

Forme de virtualisation la plus répandue, les systèmes d'exploitation virtualisés, ou machines virtuelles, constituent un composant incontournable de l'infrastructure informatique, permettant de faire fonctionner simultanément des systèmes standards, comme Windows ou Linux, sur la même plate-forme matérielle.

### **I.2.3. Les types de virtualisation [1.4]**

Il existe 4 types de virtualisation à savoir :

- La virtualisation complète
- La paravirtualisation
- L'émulation en informatique
- La simulation

#### **I.2.3.2. La virtualisation complète**

La virtualisation complète permet de faire fonctionner n'importe quel type de système d'exploitation (Windows ou linux) étant qu'invité dans une machine virtuelle. Elle est très simple à mettre en place et à utiliser.

Dans cette génération, l'hyperviseur est installé au-dessus du système d'exploitation hôte, comme le montre la figure 1, qui considère l'environnement de virtualisation comme une vulgaire application. Une particularité de cette génération de virtualisation est l'utilisation d'hyperviseur dit hyperviseur de type 2. Un hyperviseur dit de type 2 est un logiciel que l'on installe sur un système d'exploitation hôte pour bénéficier d'une interface de virtualisation. Les hyperviseurs de type 2 créent un environnement virtuel simulant un nouvel ordinateur complet avec du « faux matériel ». Le système d'exploitation invité ne communique qu'avec ce faux matériel. L'hyperviseur de type 2 communique avec les ressources physiques de l'ordinateur hôte via le système d'exploitation hôte qui traduit les requêtes. Ceci a pour effet de baisser les performances du système.

Il faut noter que le système invité peut ne pas être du même type que le système hôte.

#### **Exemples d'Hyperviseur de type 2 :**

- VMware Workstation, VMware Fusion de VMware.
- L'hyperviseur open source QEMU
- VirtualBox d'Oracle
- Virtual PC, Virtual Server de Microsoft.
- Parallels Workstation et Parallel Desktop de Soft, etc.

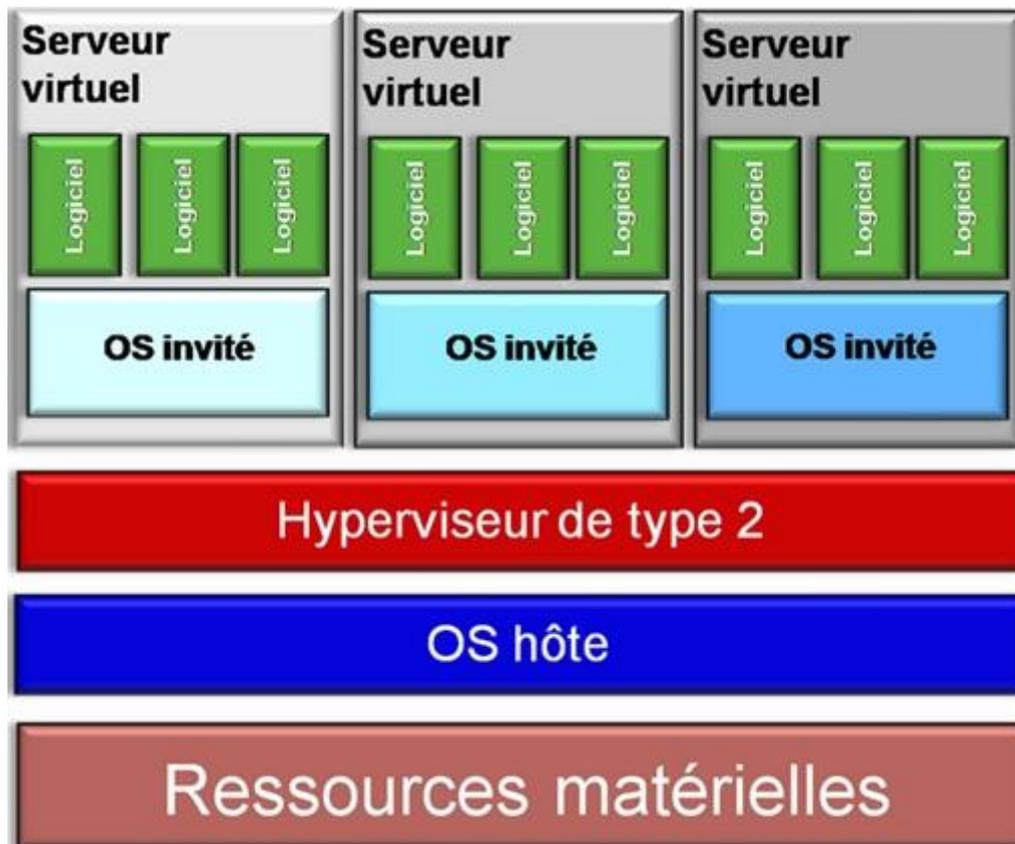


Figure 1: La virtualisation complète

### I.2.3.1. La paravirtualisation

Dans ce type de virtualisation, l'hyperviseur et le système d'exploitation hôte coopèrent. En virtualisation complète, expliqué précédemment, l'hyperviseur et le système d'exploitation hôte sont deux couches bien distinctes et que le système d'exploitation hôte traduit les requêtes de communication entre le matériel et l'hyperviseur. Dans le cas de la paravirtualisation par contre, le système d'exploitation hôte et l'hyperviseur ne forment plus qu'une même couche. On parle donc ici d'hyperviseur de type 1. L'hyperviseur de type 1 encore appelé hyperviseur natif est un noyau hôte allégé et optimisé pour la virtualisation qui s'exécute directement sur une plateforme matérielle. Il n'est plus question de traduire les appels systèmes entre l'hyperviseur et le matériel mais simplement de transfert, ce qui rend cette génération de virtualisation plus performante. La paravirtualisation offre des performances optimales et nécessite un système d'exploitation modifié et conscient qu'il évolue dans un environnement de virtualisation.

Elle permet aux Moniteurs de Machines Virtuelles (MMV) d'être plus simples et aux machines virtuelles fonctionnant dessus d'atteindre un niveau de performance proche du matériel réel.

Dans ce type de virtualisation, les machines virtuelles n'ont pas besoin d'un système d'exploitation hôte pour être opérationnelles, seul un hyperviseur (généralement lancé sur un Live-CD) est utilisé pour les installer (ou éventuellement les déployer).

#### Exemples de hyperviseur de type 1 :

- ESXi Server de VMWare.
- Hyper-V Server de Microsoft.
- XEN d'Oracle.

- Logical Domains de SUN, etc.

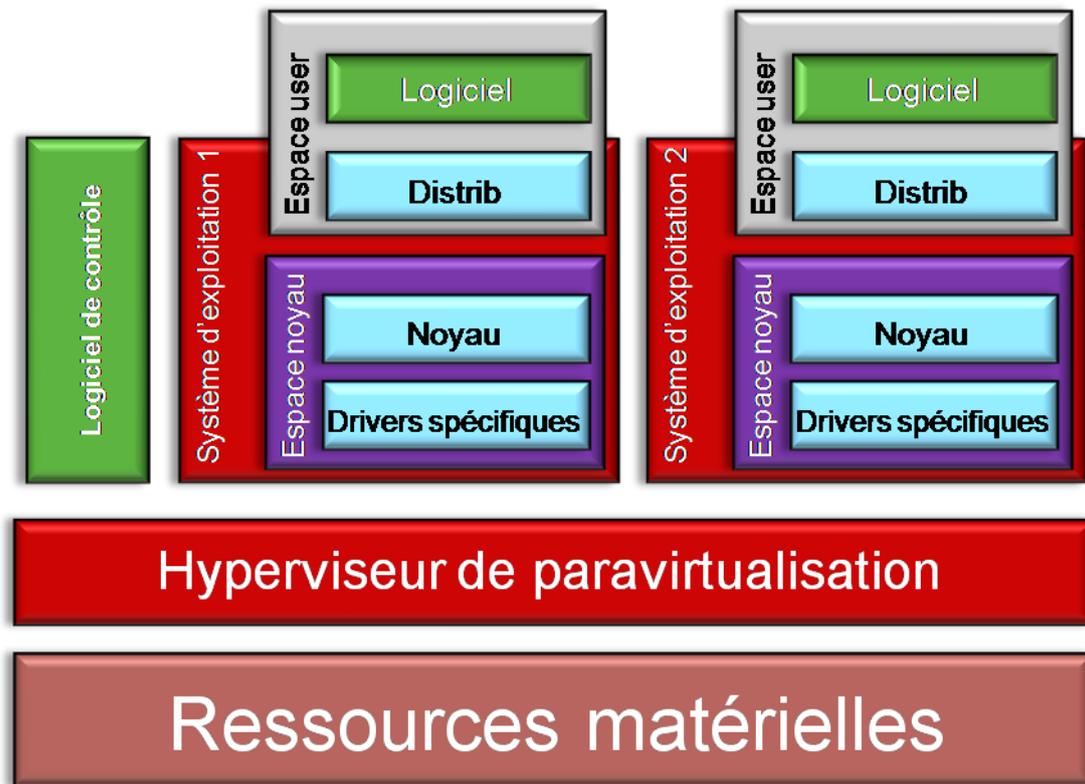


Figure 2: La paravirtualisation

### I.2.3.3. L'émulation en informatique

L'émulation consiste à substituer un élément de matériel informatique -tel un terminal, un ordinateur ou une console de jeux- par un logiciel.

Emuler c'est «chercher à imiter». Il faut voir dans l'émulation une imitation du comportement physique d'un matériel par un logiciel, et ne pas la confondre avec la simulation, laquelle vise à imiter un modèle abstrait.

L'émulation présente des performances moindres par rapport aux deux types cités précédemment, car les instructions de la machine invitée (émulées) ne sont pas exécutées directement par le CPU de la machine hôte. Cela a un impact important sur son comportement. Il faut cependant noter que l'émulation est plus simple à mettre en œuvre que la virtualisation proprement dite.

### I.2.3.4. La simulation

La simulation consiste à reproduire le comportement d'un système qui n'a pas encore été élaboré, ou qui n'est pas réel afin de le tester et de pouvoir interagir avec les résultats qu'on obtient. C'est une technique utilisée en générale dans la phase de pré-production.

## I.2.4. Les Techniques de virtualisation [1.3]

Elles existent sous de nombreuses formes :

- Au niveau d'une application.
- Au niveau du noyau.

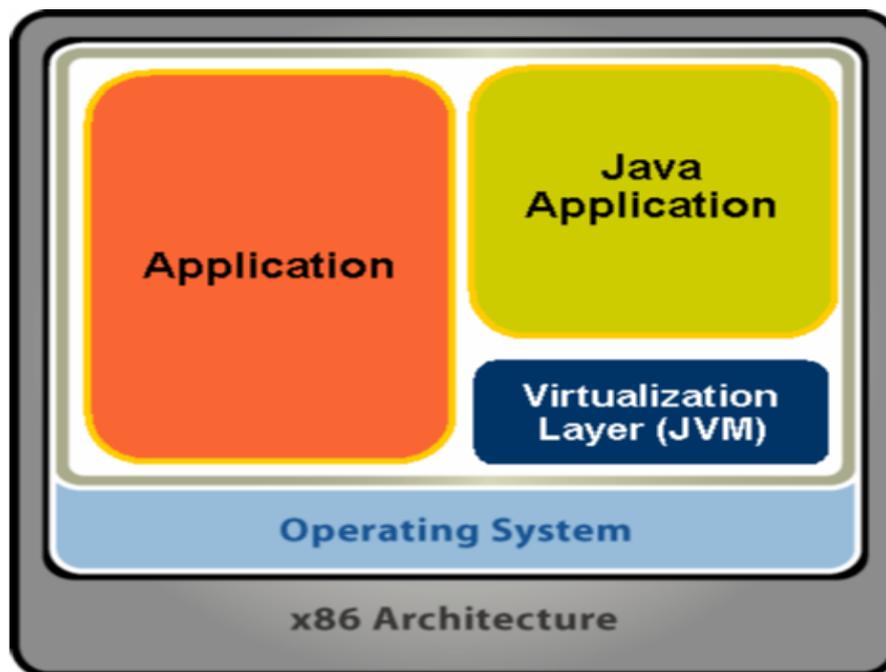
- Par l'emploi d'un émulateur.
- Par l'utilisation d'un moniteur ou hyperviseur de machines virtuelles.
- Au niveau du matériel, par l'intégration de commandes dédiées dans le processeur.

#### **I.2.4.1. La virtualisation au niveau d'une application.**

Elle simule la présence de plusieurs services.

Exemples :

- Hôtes virtuels avec le serveur Web Apache.
- Domaines virtuels avec Postfix.
- Les performances sont optimales.



*Figure 3: La virtualisation au niveau d'une application*

#### **I.2.4.2. La virtualisation au niveau du noyau**

Elle permet de faire croire à la présence de plusieurs machines.

- Les applications sont séparées et regroupées dans des « cages ».
- Le noyau répartit les ressources.
- Exemples : Linux-VServer, OpenVZ.
- Les performances sont excellentes.
- Il n'est pas possible d'utiliser des noyaux différents en même temps.

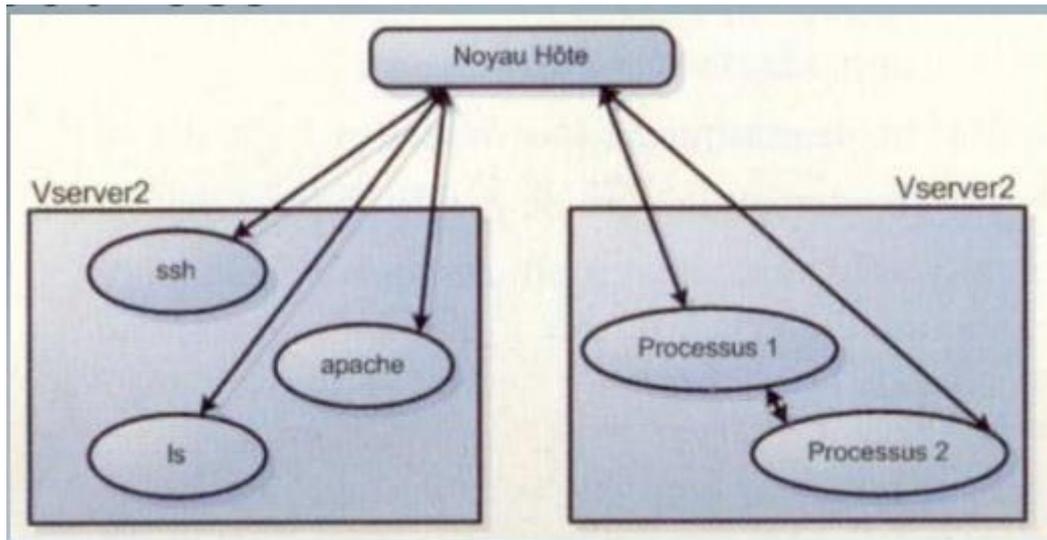


Figure 4: La virtualisation au niveau du noyau

### I.2.4.3. La virtualisation par l'emploi d'un émulateur

Une application peut simuler un processeur ou une machine complète (mémoire, disque, réseau, etc.). Elle intercepte les instructions de la machine virtuelle pour les traduire à destination de la machine hôte. Cela permet d'exécuter des applications prévues pour d'autres architectures mais souvent avec des performances moyennes.

Exemples : QEMU, VMware Workstation, Virtual Box.

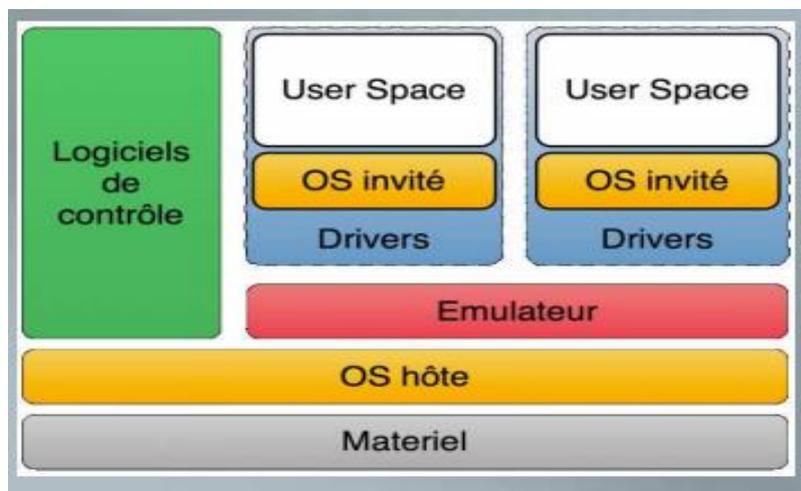


Figure 5: La virtualisation par l'utilisation d'un hyperviseur

### I.2.4.4. La virtualisation par l'utilisation d'un moniteur ou hyperviseur de machines virtuelles.

Un hyperviseur fonctionne directement au-dessus du matériel et gère des machines virtuelles. Celles-ci fonctionnent directement avec le matériel sans interception des instructions. Les performances sont très bonnes. Il est possible de faire cohabiter des noyaux et des systèmes différents.

Exemples : Xen, Hyper-V.

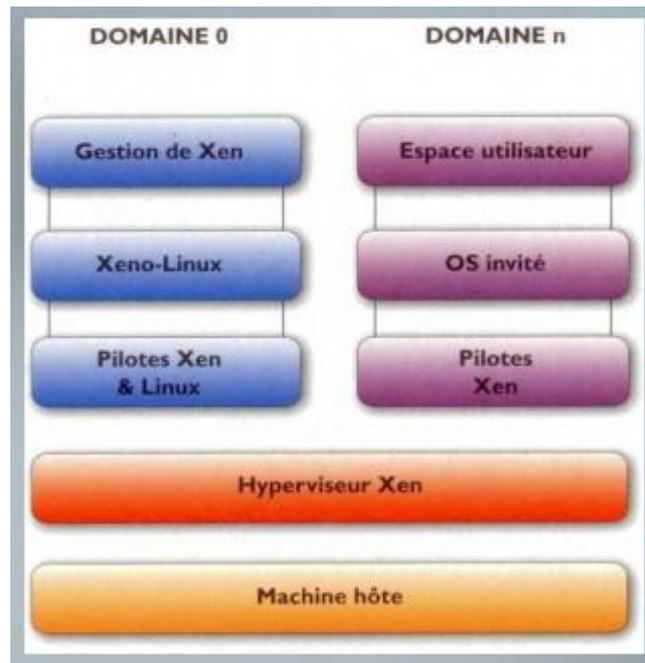


Figure 6: la virtualisation par l'emploi d'un hyperviseur de machines virtuelles

#### I.2.4.5. La virtualisation au niveau du processeur

Le matériel se charge de virtualiser les accès mémoire et de protéger le processeur des accès bas niveau. Il n'est plus nécessaire de surveiller et de traduire les instructions des machines virtuelles. Les performances de la virtualisation logicielle sont nettement améliorées. Les performances sont très bonnes.

Exemples : AMD-V, Intel VT.

#### I.2.5. Les avantages de la virtualisation

Les points suivants sont les cinq raisons principales de consolider des services :

- Le temps de restauration moyen est trop long.
- L'infrastructure souffre d'un matériel vieillissant.
- Les capacités de l'infrastructure ne suffisent plus.
- Les systèmes sont sous-utilisés.
- La virtualisation permet d'économiser.

##### I.2.5.1. Le temps moyen de restauration

Le temps moyen de restauration est le temps qu'il faut pour restaurer un système aux clients lorsqu'il est hors ligne. La virtualisation diminue ce temps en mettant à disposition des instantanés ou des sauvegardes de machines virtuelles complètes pour restaurer les services en erreur.

##### I.2.5.2. L'infrastructure vieillissante

La vie du matériel est courte. L'espérance de vie d'une infrastructure matérielle est de 3 à 4 ans, car le taux d'échec augmente énormément après cette durée. La virtualisation est la solution de ce dilemme ; les machines virtuelles sont indépendantes du matériel, de ce fait, on peut facilement les migrer vers d'autres matériels, même à chaud!, puis, mettre à jour leur configuration (RAM, etc...)

### **I.2.5.3. Les capacités physiques insuffisantes**

Un système d'exploitation a des limites de capacité fixes ; par exemple, si notre système ne prend en charge que 4 Go de RAM, on ne peut pas en ajouter plus que cela. Ces limitations n'affectent pas les machines virtuelles. Si la machine hôte dispose des capacités suffisantes et que le logiciel de virtualisation le permet, on peut améliorer les capacités de nos machines virtuelles.

### **I.2.5.4. La sous-utilisation**

Comme vu précédemment, les systèmes sous-utilisés sont parfaits pour la virtualisation et la consolidation, non pas parce que les machines virtuelles sont plus utilisées que la machine physique, mais parce que la machine virtuelle, à la différence de la machine physique, ne consomme pas d'électricité, n'a pas besoin de refroidissement et ne prend pas de place.

## **I.3. Le Cloud Computing**

### **I.3.1. Introduction**

Bien que l'anglicisme Cloud computing soit largement utilisé en français, on rencontre également les francisations :

- Informatique virtuelle,
- Informatique dans le nuage, informatique en nuage
- Informatique dématérialisée
- Stockage dans les nuages,
- Stockage à distance
- ou encore Infonuagique

#### **I.3.1.1. Avant le cloud computing**

Les salariés de l'entreprise accèdent aux ressources informatiques (serveurs, applications, espaces de stockage...) via le réseau interne de l'entreprise.



*Figure 7: Le réseau interne d'une entreprise sans le cloud*

Les applications d'entreprise traditionnelles sont toujours trop compliquées et onéreuses. Le nombre et la variété des logiciels et des matériels requis pour leur exécution est écrasante. Une équipe d'experts est nécessaire pour en assurer l'installation, la configuration, les tests, l'exécution, la sécurité et la mise à jour.

### I.3.1.2. Avec le cloud computing

Les salariés de l'entreprise accèdent à des infrastructures informatiques mises à leurs dispositions par un ou des prestataires de Cloud via Internet. Ces infrastructures sont mutualisées entre plusieurs entreprises pour être proposées au meilleur coût.

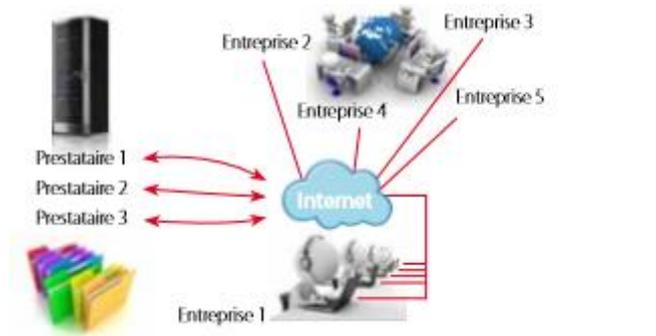


Figure 5 : Infrastructures informatiques avec le cloud

Dans ce cas, vous ne gérez aucun matériel, ni logiciel. L'infrastructure partagée signifie qu'il fonctionne comme un utilitaire : vous ne payez que ce dont vous avez besoin, les mises à niveau sont automatiques et l'évolution du système est facile

### I.3.2. Historique

Le cloud computing est le genre de concept dont on n'arrive pas à attribuer précisément la paternité. Certains l'attribuent au père fondateur d'Internet (Arpanet), technologie sans laquelle l'informatique dans les nuages n'existerait pas. D'autres pensent que la discipline n'est apparue que plus tard, dans les années 2000, telle une solution à un problème rencontré par Amazon : en 2006, Amazon a été le premier à proposer ce type de services, presque malgré lui. En 2002, le géant investit dans un énorme parc informatique pour éviter des surcharges des serveurs à l'approche de Noël, et se retrouve avec une grande quantité de ressources inutilisées. Il décide alors de créer une plateforme hautement disponible et virtualisée pour louer ces ressources à des entreprises, profitant d'Internet pour rentabiliser des équipements.

Nous ne trancherons pas ici cette querelle quasi-théologique, mais reconnaitrons que la première personne à employer l'expression de cloud computing fut le professeur de système d'information, Ramnath Chellappa. En poste à l'université du Texas à Austin, il présenta en 1997 un paradigme informatique, le cloud, où les limites de l'informatique seraient définies par ses ressources économiques plus que par ses limites techniques.

Salesforces<sup>1</sup> fut le premier, en 1999, à transformer ce concept en business avec le logiciel de gestion de la relation client éponyme. Amazon lui emboîte le pas en 2002. Mais c'est sans doute grâce à IBM que tout le monde en parle aujourd'hui. En 2007, le Big Blue décida de faire de ce nuage l'une des lignes de force de sa stratégie [1.5].

---

<sup>1</sup> Salesforce est un éditeur de logiciels, basé à San Francisco aux États-Unis. Il distribue des logiciels de gestion basés sur Internet et héberge des applications d'entreprises. L'entreprise est surtout connue au niveau international pour ses solutions en gestion de la relation client.

### I.3.3. Définitions

Le cloud computing est un domaine vaste et dynamique, si bien qu'à l'heure actuelle personne ne prétend en posséder une définition claire et absolue. En effet, la plupart des définitions attribuées à cette discipline semblent se concentrer seulement sur certains aspects de la technologie. Nous donnons ci-dessous quelques définitions.

Pour wikipédia [1.6] : « Le Cloud computing est un concept qui consiste à déporter sur des serveurs distants des traitements informatiques traditionnellement localisés sur le poste client de l'utilisateur ».

Selon la définition du National Institute of Standards and Technology (NIST) : « Le cloud computing est l'accès via un réseau de télécommunications, à la demande et en libre-service, à des ressources informatiques partagées configurables. Il s'agit donc d'une délocalisation de l'infrastructure informatique ».

Pour génération NT : « Le cloud computing est un concept d'organisation informatique qui place Internet au cœur de l'activité des entreprises, il permet d'utiliser des ressources matérielles distantes pour créer des services accessibles en ligne ».

### I.3.4. Les caractéristiques principales

Le cloud computing a généralement ces caractéristiques principales suivantes [1.7] :

- **Un accès en libre-service à la demande** : C'est-à-dire les services sont fournis de façon automatique, sans nécessiter d'interaction humaine.
- **Un accès ubiquitaire au réseau** : Services facilement accessibles au travers du réseau, par le biais de mécanismes standard qui permettent une utilisation depuis différents types de terminaux.
- **Une mise en commun des ressources (Pooling)** : Les ressources peuvent être regroupées pour servir des utilisateurs multiples, pour lesquelles les ressources physique et virtuelles sont automatiquement attribuées. Généralement ces mécanismes se font totalement dans la transparence.
- **Une élasticité rapide** : Des ressources peuvent être automatiquement mises à la disposition des utilisateurs en cas d'accroissement de la demande et libérées lorsqu'elles ne sont plus indispensables.
- **Paiement à l'usage** : La consommation des ressources dans le cloud s'adapte au plus près aux besoins de l'utilisateur. Le fournisseur est capable de mesurer de façon précise la consommation (en durée et en quantité) des différents services (CPU, stockage, bande passante,...) ; cela lui permettra de facturer l'utilisateur selon sa réelle consommation [1.7].
- **Basé-SLA<sup>2</sup>** : Les Cloud sont dynamiquement en fonction des contrats d'accord de niveau de service entre le fournisseur et l'utilisateur. Le SLA définit des politiques,

---

<sup>2</sup> Il s'agit de clauses basées sur un contrat définissant les objectifs précis attendus et le niveau de service que souhaite obtenir un client de la part du prestataire et fixe les responsabilités.

telles que les paramètres de livraison, les niveaux de disponibilité, la maintenabilité, la performance, l'exploitation, ou bien autres attributs du service, comme la facturation et des sanctions en cas de violation de contrats.

### I.3.5. Les modèles de service

Le cloud computing dispose de trois principaux modèles de service [1.6] :

- Le SaaS : Software as a Service
- La PaaS : Platform as a Service
- L'IaaS : Infrastructure as a Service

#### I.3.5.1. Le Software as a Service (SaaS)

##### ➤ Définition

Le SaaS est l'utilisation d'un logiciel, comme d'un service. Il se réfère à tout service Cloud permettant aux clients d'avoir accès à des applications logicielles sur Internet. Ces applications sont hébergées dans le Cloud et peuvent être utilisées pour un large éventail de tâches, tant par les personnes privées que par les organisations. Google, Twitter, Facebook et Flickr sont tous des exemples de SaaS, où les utilisateurs peuvent avoir accès aux services via n'importe quel appareil disposant d'une connexion à Internet. Généralement, les usagers d'une entreprise utilisent des applications pour remplir de nombreuses fonctions, incluant la comptabilité, la facturation, le suivi des ventes, la planification, le monitoring des performances et les communications (y compris la messagerie web et la messagerie instantanée).

##### ➤ Description

On emploie souvent l'expression de logiciel à la demande pour parler des services SaaS, car l'utiliser équivaut au fait de louer un logiciel plutôt que de l'acheter. En présence d'applications logicielles traditionnelles, l'utilisateur achète d'abord le logiciel et l'installe ensuite sur son ordinateur. La licence du logiciel peut limiter le nombre d'utilisateurs et/ou d'équipements où il peut être déployé. En revanche, les utilisateurs ayant recours à une offre de SaaS souscrivent un abonnement au logiciel souhaité, généralement sur une base mensuelle, plutôt que de l'acheter. L'achat et l'utilisation des applications se fait en ligne, et les fichiers sont sauvegardés dans le Cloud plutôt que sur un ordinateur individuel.

##### ➤ Les avantages

Il existe de nombreuses raisons pour lesquelles SaaS est avantageux tant pour les organisations que pour les individus :

- **Pas de coûts d'équipements supplémentaires** : la puissance de traitement nécessaire pour faire fonctionner les applications est fournie par le fournisseur Cloud.
- **Pas de frais d'installation** : les applications sont prêtes à l'emploi dès que l'utilisateur a souscrit à l'abonnement.
- **Paiement à l'utilisation** : si l'utilisateur n'a besoin du logiciel que pour une période limitée, il ne paie que cette période et généralement, il est possible de mettre fin aux abonnements en tout temps.

- **Utilisation évolutive** : si un utilisateur a besoin par exemple de davantage de stockage ou de services supplémentaires, il peut y accéder à la demande sans avoir besoin d'installer de nouveaux logiciels ou du nouveau matériel.
- **Mises à jour automatiques** : dès qu'une mise à jour est disponible, l'utilisateur peut y accéder directement en ligne, souvent gratuitement. Aucun nouveau logiciel n'est nécessaire, comme il arrive souvent avec d'autres types d'applications, et les mises à jour sont généralement déployées automatiquement par le fournisseur Cloud.
- **Inter-compatibilité entre équipements** : il est possible d'accéder aux applications SaaS depuis n'importe quel équipement disposant d'une connexion à Internet, ce qui est idéal pour ceux qui utilisent de nombreux appareils, comme des téléphones et des tablettes dotés d'une connexion à Internet et pour ceux qui n'utilisent pas toujours le même ordinateur.
- **Accessible depuis partout** : pour peu que l'on dispose d'un équipement doté d'une connexion à Internet, il est possible d'accéder à l'application depuis n'importe où, sans être limité aux ordinateurs individuels.
- Les applications peuvent être personnalisées et faire l'objet d'une marque blanche ; il est possible de personnaliser certains logiciels, ce qui veut dire qu'ils peuvent être modifiés pour correspondre aux besoins et à la marque d'un client déterminé.

En résumé, avec le SaaS, il n'y a aucun frais d'installation, comme c'est souvent le cas avec d'autres applications. Il est extensible, avec des mises à jour disponibles à la demande. Il est possible d'accéder au Software as a Service depuis n'importe quel équipement disposant d'une connexion à Internet. Pour peu qu'une connexion à Internet soit disponible, il est possible d'accéder aux applications depuis n'importe où.

### **I.3.5.2. La Platform as a Service (PaaS)**

#### ➤ **Définition**

La plateforme as a service est l'utilisation d'une plateforme, comme d'un service. La PaaS est une catégorie de services de Cloud computing qui fournit la plateforme et l'environnement informatique nécessaire aux développeurs pour mettre en place leurs différents services et applications sur Internet. Les services PaaS sont hébergés dans le Cloud et les utilisateurs y accèdent simplement, par leur navigateur web.

#### ➤ **Description**

Les services PaaS permettent aux utilisateurs de créer des applications logicielles en utilisant les outils fournis par le fournisseur. Ils peuvent prendre la forme de fonctionnalités préconfigurées auxquelles les clients peuvent souscrire, en ne choisissant que celles qui conviennent à leurs exigences. Cela signifie que les packs PaaS peuvent aller de la simple offre "point-and-click", où le client n'a pas besoin d'avoir des connaissances particulières en hébergement, à la fourniture d'options d'infrastructure pour développement avancé.

L'infrastructure et les applications sont gérées par le fournisseur et les clients peuvent accéder à des services de support si nécessaire. Les services sont constamment mis à jour, avec l'amélioration des fonctionnalités existantes et ajout de fonctionnalités additionnelles. Comme

avec la plupart des offres Cloud, les services PaaS sont facturés sur la base d'un système d'abonnement, où les clients ne paient que ce qu'ils utilisent effectivement.

### ➤ **Exemples des services offerts**

On trouvera ci-dessous certaines des fonctionnalités pouvant être incluses dans une offre PaaS :

- Système d'exploitation
- Environnement de script serveur
- Système de gestion de bases de données
- Logiciel serveur
- Support
- Stockage
- Accès réseau
- Outils de design et de développement
- Hébergement

Les développeurs d'applications et les entreprises web peuvent bénéficier des services PaaS. Ils mettent en place une application qu'ils souhaitent vendre sur Internet ou un logiciel prêt à l'emploi, les développeurs de logiciels peuvent tirer profit d'une solution PaaS. Par exemple, les développeurs web peuvent utiliser les environnements PaaS à chaque stade du processus, du développement au test et, en dernier lieu, d'hébergement de leurs sites web.

### ➤ **Les avantages**

On trouvera ci-dessous certains des avantages des services PaaS pour les développeurs d'applications et les clients du cloud computing :

- Ils n'ont pas besoin d'investir dans une infrastructure physique.
- Le fait de "louer" une infrastructure virtuelle est avantageux à la fois en termes de coûts et d'un point de vue pratique.
- Ils n'ont pas besoin d'acheter eux-mêmes le matériel ou d'avoir recours au personnel capable de le gérer. Cela leur permet de se concentrer sur le développement de leurs applications.
- Plus encore, les clients ne doivent que louer les ressources dont ils ont besoin, plutôt que d'investir dans une capacité fixe qu'ils n'utiliseront pas, engendrant du gaspillage.
- Rend le développement accessible aux "non-experts" ; grâce à l'offre PaaS, chacun peut développer une application. Ils peuvent le faire simplement par leur navigateur web, en utilisant la fonctionnalité one-click. Un bon exemple est constitué par les logiciels d'installations blog one-click comme WordPress.
- Flexibilité : les clients ont le contrôle sur les outils installés sur leurs plateformes et sont libres de créer une plateforme qui corresponde à leurs exigences spécifiques. Ils peuvent choisir les fonctionnalités qu'ils estiment nécessaires.
- Adaptabilité : il est possible de modifier les fonctionnalités en fonction des circonstances.

- Les équipes situées à différents emplacements peuvent travailler ensemble ; étant donné que la seule chose nécessaire est une connexion internet et un navigateur web, les développeurs dispersés sur différents sites peuvent travailler ensemble à la même application.
- Sécurité : le service PaaS inclut également la sécurité, y compris la sécurité des données, le back up et la récupération.

En résumé, l'offre PaaS fournit un environnement d'exploitation pour le développement d'applications. En d'autres mots, elle fournit l'architecture et l'infrastructure globales nécessaires pour supporter le développement d'applications. Cela inclut le réseau, le stockage, le support logiciel et les services de gestion. Par conséquent, l'offre PaaS est idéale pour développer de nouvelles applications destinées au web ou aux équipements mobiles et aux autres terminaux.

### **I.3.5.3. L'Infrastructure as a Service (IaaS)**

#### **➤ Définition**

L'infrastructure as a service est l'utilisation d'une infrastructure, comme d'un service. L'IaaS est l'une des trois principales catégories de services de Cloud computing, avec Platform as a Service (PaaS) et Software as a Service (SaaS). Comme tous les services de Cloud computing, ce service donne accès aux ressources informatiques dans un environnement virtualisé, le "Cloud", à travers une connexion publique, généralement Internet.

#### **➤ Description**

Dans le cas de l'IaaS, la ressource fournie est de matériel informatique virtualisé ou, en d'autres termes, une infrastructure informatique. La définition du service comprend des offres telles que l'espace serveur, des connexions réseau, de la bande passante, des adresses IP et des load-balancers<sup>3</sup>. Physiquement, les ressources hardware proviennent d'une multitude de serveurs et de réseaux régulièrement distribués à travers de nombreux Datacenter, que le fournisseur de services Cloud a la responsabilité d'entretenir. Parallèlement, l'accès aux composants virtualisés est donné à l'entreprise cliente afin que celle-ci puisse construire ses propres plateformes IT.

A l'instar des deux autres formes d'hébergement Cloud, l'IaaS peut être utilisée par les entreprises clientes pour créer des solutions informatiques à des coûts avantageux et facilement extensibles, dans la mesure où la complexité et les coûts inhérents à la gestion du matériel informatique sous-jacent sont externalisés au fournisseur Cloud. Ainsi, les entreprises ont tout loisir d'utiliser les ressources Cloud IaaS lorsqu'elles en ont besoin, au gré de l'évolution de leurs opérations et de leur développement, plutôt que d'acheter, d'installer et d'intégrer elles-mêmes le matériel.

#### **➤ Exemples d'utilisations de services IaaS par une entreprise :**

Infrastructure de l'entreprise : pour ses réseaux internes, comme les réseaux de Cloud privés et les réseaux locaux virtuels, qui utilisent des pools de serveurs et des ressources réseau et sur

---

<sup>3</sup> Le **Load-balancing/ Load-balancer** (équilibrage de charge de charge) est une technique spécifique aux serveurs et connexions réseaux qui permet de répartir le trafic entre plusieurs ordinateurs ou lignes physiques.

lesquels l'entreprise peut stocker ses données et faire fonctionner les applications dont elle a besoin pour ses activités quotidiennes. Les entreprises en expansion peuvent augmenter leur infrastructure en fonction de leur croissance, et leurs Cloud privés (auxquels seule l'entreprise a accès) peuvent protéger le stockage et le transfert de leurs données sensibles.

Hébergement Cloud : hébergement de sites web sur des serveurs virtuels fondés sur un pool de ressources tirées de serveurs physiques sous-jacents. Par exemple, un site web hébergé dans le Cloud peut bénéficier de la redondance fournie par un vaste réseau de serveurs physiques et d'une extensibilité sur demande pour affronter les demandes inattendues dont il fait l'objet.

Virtual Data Centers (VDC) : un réseau virtualisé de serveurs virtuels interconnectés pouvant être utilisé pour offrir davantage de capacités d'hébergement Cloud ou d'infrastructure informatique ou pour intégrer l'ensemble de ces opérations dans une implémentation de Cloud privé ou de Cloud public.

#### ➤ **Les avantages**

Une offre typique d'IaaS peut fournir les fonctionnalités et avantages suivants :

- **Extensibilité** : les ressources étant disponibles dès que le client en a besoin, il n'y a donc pas de délais pour étendre ses capacités ni gaspillage de capacités non utilisées
- **Pas d'investissement en matériel** : le matériel physique sous-jacent supportant un service IaaS est installé et maintenu par le fournisseur Cloud, économisant ainsi du temps et de l'argent pour le client qui n'a pas besoin d'y veiller lui-même
- **Modèle de prix basé sur l'utilisation** : le service est accessible à la demande et le client ne paie que pour les ressources qu'il a effectivement utilisées
- **Indépendance de l'emplacement** : généralement, il est possible d'accéder au service depuis n'importe quel emplacement tant qu'il dispose d'une connexion Internet et que la sécurité du Cloud le permet
- **Sécurité physique des sites de data centres** : services accessibles par le Cloud public, ou Cloud privés hébergés à l'externe auprès du fournisseur Cloud, bénéficiez de la sécurité physique des serveurs hébergés dans un data centre
- **Aucun point de défaillance unique** : en cas de panne d'un serveur ou d'un commutateur réseau, le service n'est pas affecté grâce à la multitude de ressources de matériel et de configurations de redondance. Pour de nombreux services, si un data centre entier devait se retrouver hors ligne, le service IaaS continuerait à fonctionner sans problèmes.

#### **I.3.5.4. La différence entre l'IaaS, le PaaS et le SaaS**

Le tableau ci-après résume la différence entre les trois principales catégories de cloud computing.

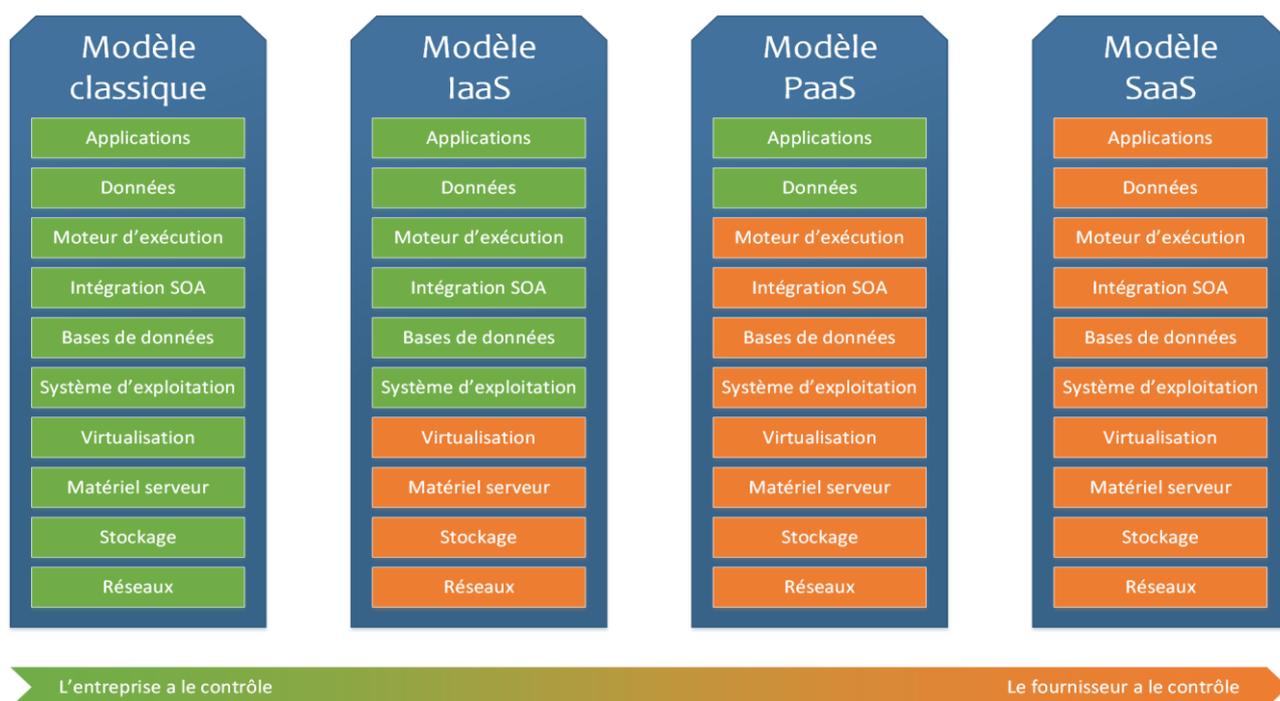


Figure 8: Différence entre IaaS, PaaS et SaaS

La figure 6 nous montre les différentes parties gérées par les fournisseurs et la gestion est à la disposition d'un utilisateur pour chaque modèle de service qui fait la différence entre ces trois modèles.

### I.3.6. Les modèles de déploiement du cloud computing

Le cloud computing dispose quatre modèles de déploiement :

- Le Cloud Public
- Le Cloud Privé
- Le Cloud Communautaire
- Le Cloud Hybride

#### I.3.6.1. Le Cloud Public

Un service dans le cloud public est l'équivalent d'un service sur Internet. L'infrastructure et les applications du client sont hébergées chez un prestataire de services dans le cloud, à l'endroit où celui-ci est installé. Le client n'a aucune visibilité ni aucun contrôle sur l'endroit où les services sont hébergés dans le cloud. L'infrastructure principale est partagée par plusieurs entreprises, mais les données et l'usage des applications sont séparés logiquement, de manière à ce que seuls les utilisateurs autorisés y aient accès. Le service dans le cloud public intéresse les décideurs car il réduit la complexité et les délais de mise en œuvre, imputables aux tests et au déploiement de nouvelles applications. En outre, cette solution est généralement moins coûteuse, car les dépenses en capital sont moindres, voire nulles.

##### ➤ Les avantages des services dans le cloud public

Face au choix entre Cloud public ou Cloud privé, l'entreprise doit prendre en considération certains avantages et inconvénients de chaque solution.

Les avantages des Cloud publics sur les Cloud privés sont les suivants.

- **La simplicité et l'efficacité** sont les avantages principaux d'un cloud public. Les services de cloud public sont des prestations généralement proposées par le biais d'une connexion Internet. Un prestataire tiers externe héberge et gère le système. Les utilisateurs se connectent au système par Internet. Le paiement du service de cloud public est généralement mensuel ou annuel et selon l'usage.
- **Un faible coût** : avec un service de cloud public, les entreprises peuvent réduire leur budget informatique du fait qu'elles n'ont pas à acheter de matériel physique, ce qui, de plus, réduit les coûts énergétiques, puisque les serveurs sont virtuels et hébergés par un tiers. Les entreprises peuvent personnaliser leur service dans le cloud public en choisissant le nombre d'utilisateurs par exemple, de manière à ne payer que ce dont elles ont besoin (modèle de paiement à l'usage ou « pay-as-you-go »). Le cloud public étant hébergé par un tiers, l'entreprise n'a pas à employer une personne pour surveiller le système, qui est pris en charge par l'hébergeur.
- **Temps de travail réduit** : les serveurs sur site ont besoin de maintenance. S'il s'avère nécessaire de modifier les configurations matérielles ou logicielles ou si un serveur tombe en panne ou doit être redémarré, plusieurs heures ou journées de travail selon le cas sont nécessaires. Dans le cas d'un service dans le cloud public, tout étant virtualisé, la reconfiguration peut ne prendre que quelques minutes. De plus, les serveurs étant hébergés dans le cloud, si l'un d'entre eux tombe en panne, un autre serveur peut être démarré et la durée d'interruption est moins longue.
- **Aucune maintenance** : du fait de l'hébergement hors site du service dans le cloud public, le service informatique n'est plus responsable de la maintenance du système. L'ensemble étant géré par l'hébergeur, les ajouts technologiques ou les mises à jour peuvent être réalisés par les utilisateurs bien plus rapidement. Le service dans le cloud public élimine la maintenance du matériel physique, qui peut être effectuée à partir d'un simple écran de configuration.
- **Aucun contrat** : le modèle de paiement à l'usage « pay-as-you-go » supprime les engagements à long terme. Lorsque l'abonnement mensuel ou annuel se termine, vous n'avez aucune obligation de continuer à utiliser votre service dans le cloud public.

En général, le service dans le cloud public est plus avantageux que celui dans le cloud privé en termes d'extensibilité, de polyvalence, de simplicité d'utilisation et de prix. La possibilité d'utiliser tous les services, y compris l'infrastructure, sur la base d'un paiement selon l'usage et l'élimination des problèmes liés à leur gestion quotidienne sont les avantages les plus significatifs que reconnaissent les entreprises clientes des services dans le cloud public.

#### ➤ **Les Inconvénients des services dans le cloud public**

Avant de prendre leur décision, les entreprises doivent considérer les inconvénients que présentent les services dans le cloud public.

- **Le manque de contrôle** : beaucoup d'entreprises ressentent un manque de contrôle sur

leurs données propres dans le cas d'un service dans le cloud public, puisque c'est l'hébergeur qui est responsable du système de données.

- **Une vitesse faible** : les services dans le cloud public se font par des connexions Internet. La vitesse de transfert des données est donc limitée à celle que propose le fournisseur d'accès Internet. Pour une entreprise qui doit stocker et transférer de grands volumes de données, le service dans le cloud public n'est peut-être pas la meilleure solution. Les serveurs du cloud public doivent également assurer un accès rapide aux informations sur demande de leurs clients, et bon nombre d'entre eux ne toléreront pas que les temps de téléchargement de pages augmentent.
- **Le manque d'investissement** : bien que cette méthode permette de faire des économies importantes en réduisant les investissements initiaux, louer un service auprès d'un fournisseur externe signifie également des frais, qui se déduisent de ces économies en capital. Posséder des équipements réseau et des serveurs sur site peut rapporter sur le long terme, si l'on considère les avantages fiscaux calculés sur les actifs.
- **La perception d'une sécurité moindre** : l'un des inconvénients majeurs serait de ne pas se sentir en complète sécurité en utilisant le cloud public. Cela ne signifie pas que les Cloud publics ne présentent aucune sécurité, la plupart d'entre eux appliquent des règles excellentes, mais pour les clients dont les informations sont sensibles (comme les institutions financières par exemple), le fait de confier ces informations à des tiers est souvent ressenti comme intolérable et considéré comme une charge.

### **I.3.6.2. Le Cloud Privé**

On parle de service dans le cloud privé, ou encore de cloud interne ou de cloud d'entreprise, lorsque l'infrastructure informatique est hébergée sur une plateforme privée, dans le centre de données du client. Cette infrastructure est dédiée à l'entreprise et n'est pas partagée avec d'autres. La virtualisation est la clé de la mise en œuvre d'un cloud privé par les entreprises. Elle permet aux entreprises d'économiser sur les coûts en utilisant l'infrastructure matérielle existante, sans avoir à acquérir d'équipement supplémentaire, comme dans le cas du cloud public. La grande différence, bien sûr, est que le cloud privé est situé dans les locaux du client et que celui-ci peut mieux contrôler l'infrastructure. Il faut garder à l'esprit que le cloud privé offre également des capacités à la demande, grâce auxquelles des services peuvent être ajoutés aussi rapidement qu'il est nécessaire.

#### **➤ Les avantages des services dans le cloud privé**

Le service dans le cloud privé présente un certain nombre d'avantages qui en font une solution plus viable que celle du cloud public.

- **Un meilleur contrôle** : le matériel étant sur site, les entreprises contrôlent mieux leurs données. C'est l'entreprise qui est responsable de la surveillance et de la maintenance de ses données et elle en garde le contrôle.
- **Plus de sécurité** : le cloud privé étant dédié à une seule entreprise, le matériel, le stockage des données et le réseau peuvent être conçus de manière à garantir une sécurité

de haut niveau, d'autres clients ne pouvant accéder au même centre de données. Cela ne veut pas dire que les services dans le cloud public ne sont pas sûrs, mais certaines entreprises se sentiront plus en sécurité en sachant que les données se trouvent dans les locaux mêmes de l'entreprise. Une autre raison de préférer un cloud privé peut être l'existence de contraintes réglementaires locales. Dans certains pays, le centre de données hébergeant un service de cloud public doit se trouver dans le pays où résident ses utilisateurs. Lorsqu'il n'existe pas de cloud public dans le pays, un cloud privé devient l'unique solution.

- **Des performances plus élevées** : le cloud privé est déployé derrière un pare-feu sur l'intranet de l'entreprise et l'augmentation des vitesses de transfert est spectaculaire. De plus, il n'est plus nécessaire de se préoccuper des temps d'accès aux pages comme cela peut être le cas avec un cloud public.
- **Une conformité plus complète** : les Clouds publics peuvent fournir les données nécessaires à la conformité aux règles<sup>4</sup> Sarbanes Oxley, PCI DSS et HIPAA, mais parfois ces données ne sont pas assez détaillées ou personnalisables. Avec un cloud privé, la configuration du matériel, du stockage et du réseau correspond à un seul client et les contraintes de conformité sont plus faciles à respecter.
- **Personnalisable** : les performances du matériel, du réseau et du stockage peuvent être spécifiées et personnalisées dans le cloud privé, puisqu'il appartient à l'entreprise.

Comme on peut le constater, le client contrôle mieux le cloud privé et se préoccupe donc moins des aspects de sécurité. En transférant dans le cloud son système informatique matériel classique, le client peut toujours bénéficier d'extensibilité, de flexibilité et d'une meilleure productivité, mais sans rien perdre de sa responsabilité envers la sécurité des données, comme le cloud public peut en donner parfois l'impression.

#### ➤ **Les avantages des services dans le cloud privé**

Il semble qu'une meilleure sécurité et un meilleur contrôle soient associés à la solution du cloud privé, mais ce dernier présente tout de même des inconvénients.

- **Coût plus élevé** : les services dans le cloud privé sont en général plus coûteux que dans le cloud public car ils requièrent du matériel et du personnel de maintenance. Pour mettre en place un cloud privé, l'entreprise doit investir dans du matériel ou utiliser des systèmes existants, alors que le cloud public est utilisé hors site. Les Clouds privés requièrent aussi des administrateurs système et augmentent les coûts associés.
- **Maintenance sur site** : le cloud privé étant hébergé dans les locaux de l'entreprise, celle-ci doit se charger de fournir l'alimentation électrique, la climatisation et la maintenance adéquates. L'entreprise doit aussi assumer le risque de perte de données en cas de dommages matériels (incendie, surtension, dégâts des eaux). En outre, si l'entreprise possède plusieurs centres de données comportant chacun un cloud privé, la maintenance sur site et les coûts associés peuvent augmenter de manière

---

<sup>4</sup> Cette loi a pour objectif d'accroître la responsabilité des entreprises, de rendre la communication de l'information financière plus fiable et de lutter contre les comportements déviant et frauduleux des entreprises.

significative.

- **Plafonnement des capacités** : les limitations du matériel physique dans le centre de données de l'entreprise vont obligatoirement entraîner un plafonnement des capacités. La place disponible dans l'entreprise pour le déploiement des serveurs matériels est nécessairement limitée.

Sans aucun doute la première critique qui peut être faite du cloud privé est que l'entreprise doit acquérir, configurer et maintenir l'infrastructure du système ou de sa virtualisation. Alors que l'utilisateur d'un service de cloud public devra simplement acheter un système prêt à démarrer et peu coûteux, l'utilisateur de cloud privé devra investir considérablement plus au départ pour acquérir un système souvent hébergé sur site et devoir en plus gérer son développement. C'est malheureusement la contrepartie des niveaux de sécurité et de contrôle plus élevés qu'offre le cloud privé.

### **I.3.6.3. Le Cloud Hybride**

Un service de Cloud hybride combine le cloud privé et le cloud public. Dans le cas du Cloud hybride, l'entreprise garde certaines de ses opérations dans le Cloud privé sur site et utilise les services de Cloud public chez un hébergeur pour les autres. Par exemple, pour une entreprise qui a des besoins variables et travaille à la fois avec des données sensibles et non sensibles, le cloud hybride donne le meilleur des deux configurations de cloud. L'approche hybride permet de bénéficier de l'extensibilité et de l'efficacité du cloud public, tout en conservant les applications importantes sur un cloud privé.

En déployant un cloud hybride, l'entreprise peut garder chacun des types d'activité dans l'environnement le plus efficace possible. L'inconvénient est d'avoir à effectuer le suivi de plusieurs plateformes sécurisées différemment et de garantir la communication entre les différents types d'activités

### **I.3.6.4. Le Cloud Communautaire**

L'infrastructure d'un cloud communautaire est partagée par plusieurs organisations indépendantes et est utilisée par une communauté qui est organisée autour des mêmes besoins, vis-à-vis de son utilisation. Par exemple, dans le projet Open Cirrus<sup>5</sup>, le cloud communautaire est partagé par plusieurs universités dans le cadre d'un projet scientifique commun. Son infrastructure peut être gérée par les organisations de la communauté qui l'utilise ou par un tiers et peut être située, soit au sein des dites organisations, soit chez un prestataire de service.

## **I.3.7. Les avantages et les inconvénients du cloud computing**

Toute chose ayant des avantages et des inconvénients, le cloud computing ne se dérobe pas à la règle. Voici une liste des avantages et des inconvénients qu'on a pu recenser :

### **I.3.7.1. Les avantages**

Voici les dix avantages du Cloud computing :

---

<sup>5</sup> C'est un projet qui est fait pour aider les entreprises ou les fournisseurs à profiter des opportunités émergentes dans le cloud engineering, le cloud computing et l'hébergement en nuage.

### ➤ **Flexibilité**

À la seconde où une entreprise a besoin de plus de bande passante que d'ordinaire, un service de cloud computing peut instantanément répondre à sa demande, grâce à la grande capacité des serveurs distants. En fait, cette flexibilité est si cruciale que 65 % des personnes interrogées dans le cadre d'une enquête InformationWeek ont déclaré que « la capacité à répondre rapidement aux besoins de l'entreprise » constituait une des raisons principales de leur migration vers le cloud computing.

### ➤ **Récupération des données**

Lorsque les entreprises commencent à s'appuyer sur les services de cloud computing, elles n'ont plus besoin de programmes complexes de récupération des données. Les fournisseurs de cloud computing se chargent de la plupart de ces tâches et ils le font plus vite. Une étude menée par Aberdeen Group a montré que les entreprises qui utilisaient le cloud parvenaient à résoudre leurs problèmes en 2,1 heures en moyenne, soit presque quatre fois plus rapidement que les entreprises qui n'y avaient pas recours (8 heures). La même étude a également démontré que les petites et moyennes entreprises bénéficiaient du meilleur temps de récupération de données, mettant presque deux fois moins de temps que les grandes entreprises !

### ➤ **Mises à jour logicielles automatiques**

En 2010, les entreprises britanniques ont dédié 18 jours ouvrables par mois à la gestion de la seule sécurité sur site. Bien au contraire, les fournisseurs de cloud computing se chargent de la maintenance du serveur, y compris des mises à jour de sécurité, permettant ainsi à leurs clients d'allouer le temps et les ressources économisés à d'autres tâches plus stratégiques.

### ➤ **Absence de frais d'investissement**

Les services de cloud computing sont généralement fournis selon un modèle de « paiement à l'utilisation ». Ils n'engendrent donc aucun frais d'investissement (CAPEX). Par ailleurs, comme les services de cloud computing sont bien plus rapides à déployer, les entreprises supportent des coûts de démarrage minimes et des frais d'exploitation prévisibles sur le long terme (OPEX).

### ➤ **Collaboration accrue**

Le cloud computing améliore la collaboration en permettant à l'ensemble des employés, où qu'ils se trouvent, de synchroniser leurs actions et de travailler sur des documents et des applications partagées simultanément. Ils peuvent aussi suivre leurs collègues et leurs enregistrements afin recevoir des mises à jour critiques en temps réel. Une enquête menée par Frost & Sullivan a montré que les entreprises qui ont investi dans les technologies de collaboration ont bénéficié d'un retour sur investissement de 400 %.

### ➤ **Mobile benefits-revised-thumb**

À partir du moment où les employés disposent d'un accès Internet, ils peuvent travailler de n'importe quel endroit. Cette flexibilité a une influence positive sur l'équilibre entre vie

professionnelle et vie privée des employés, ainsi que sur la productivité. Une étude a montré que 42 % des adultes actifs seraient prêts à renoncer à une partie de leur salaire s'ils pouvaient travailler à domicile et accepteraient en moyenne une réduction de salaire de l'ordre de 6 %.

#### ➤ **Contrôle des documents**

Selon une étude, « 73 % des utilisateurs en entreprise travaillent avec des collaborateurs de différents fuseaux horaires et de différentes régions au moins une fois par mois ».

Si une entreprise n'utilise pas le cloud, les employés doivent sans cesse renvoyer leurs fichiers par e-mail. Ainsi, seule une personne peut travailler sur un fichier à la fois et un même document peut exister sous divers noms et formats.

Le cloud computing conserve tous les fichiers dans un emplacement central et tout le monde travaille à partir d'une copie centralisée. Les employés peuvent même utiliser la messagerie instantanée lorsqu'ils effectuent des modifications. L'ensemble de ce processus renforce la collaboration, ce qui augmente l'efficacité et améliore les résultats financiers d'une entreprise.

#### ➤ **Sécurité**

Quelque 800 000 ordinateurs portables sont perdus chaque année rien que dans les aéroports. Cela peut avoir de graves conséquences sur le plan financier, mais lorsque toutes les données sont stockées dans le cloud, elles restent toujours accessibles quoi qu'il arrive à ces ordinateurs portables.

#### ➤ **Compétitivité**

Avec le cloud, les PME-PMI ont accès à des technologies utilisées par des multinationales. Il permet aux petites entreprises d'agir plus rapidement que leurs concurrents établis ou ayant une plus grande envergure. Une étude portant sur la récupération des données a conclu que les entreprises qui n'utilisaient pas le cloud devaient compter sur les méthodes de sauvegarde sur bande et des procédures compliquées pour récupérer leurs données... Autant de manœuvres lentes et laborieuses dont les utilisateurs du cloud computing n'ont pas à se préoccuper. David l'emporte une nouvelle fois sur Goliath.

#### ➤ **Respect de l'environnement**

Les entreprises qui ont recours au cloud computing utilisent uniquement l'espace dont ils ont besoin sur le serveur, ce qui diminue leur empreinte carbone. L'utilisation du cloud computing entraîne une diminution de 30 % de la consommation énergétique et des émissions de carbone par rapport aux serveurs sur site. Là encore, ce sont les PME et les PMI qui tirent le mieux leur épingle du jeu : pour les petites entreprises, la baisse de la consommation énergétique et des émissions de carbone serait de l'ordre de 90 %.

### **I.3.7.2. Les inconvénients**

Voici une liste des dix inconvénients que présente le cloud computing :

#### ➤ **Le cadre légal**

Les données transférées dans le cloud ne sont pas forcément présentes sur le territoire national : elles peuvent l'être, comme elles peuvent être dans un autre pays européen. Par conséquent, sauf mention contraire du prestataire de service, on ne sait pas précisément à quel endroit sont stockées les données. De plus, on n'a aucun accès physique à ces données.

#### ➤ **La connexion internet**

Le cloud utilisant de manière intensive le transfert de données, il faut avoir une connexion très performante. Plusieurs cas peuvent faire que le Cloud sera inadapté à votre entreprise :

- Si elle se situe dans un lieu éloigné
- Si la connexion ne dispose pas d'un débit garanti, une coupure peut survenir, privant l'entreprise de tous les accès au Cloud, et donc à toutes vos applications et données.

#### ➤ **Le coût du cloud**

Beaucoup d'entreprises ne regardent que les frais de stockage, mais il faut également prendre en compte les frais de transferts, qui peuvent s'avérer être importants, selon l'utilisation que l'entreprise faite du cloud.

#### ➤ **L'optimisation des applications**

Malgré une connexion internet rapide, avec un débit garanti, certaines applications web peuvent s'avérer être très lentes. Elles peuvent s'avérer être plus limitées que des applications fonctionnant sur les propres ordinateurs de l'entreprise.

#### ➤ **La sécurité du cloud**

Plusieurs points sont à étudier :

- La sécurité vis-à-vis du stockage : si les données sont conservées dans un seul disque, ou si elles sont entre plusieurs unités de stockage.
- La sécurité et la confidentialité des données : si le fournisseur de service assure des tests portant sur sa sécurité informatique et si de tes tests sont faits de façon régulières.
- La sécurité des locaux : sont-ils inaccessibles pour des personnes malintentionnées ?

#### ➤ **Le piratage**

Certaines applications comme Facebook et Twitter sont très sujets aux attaques. Le piratage d'un compte d'entreprise pourrait avoir des conséquences néfastes pour la réputation de l'entreprise, tandis que l'utilisation imprudente des applications par un salarié pourrait offrir aux cybercriminels l'opportunité d'entrer dans le réseau et de soustraire des données des clients.

#### ➤ **La pérennité du service**

Il est nécessaire de se demander si l'hébergeur cloud va durer dans le temps. Cet élément est important à prendre en compte car un changement d'hébergeur peut prendre du temps, et peut nécessiter un recodage des applications.

### ➤ **La productivité des salariés**

Il est nécessaire que les employés de l'entreprise sachent se servir du cloud. En effet, même si l'entreprise fait des économies sur le stockage et le traitement d'informations, si les employés passent plus de temps pour leurs tâches à cause du cloud l'entreprise risque d'y perdre plus que d'y gagner.

### ➤ **La plateforme**

Il faut vérifier que l'hébergeur est capable de supporter de multiples plateformes car si ce n'est pas l'entreprise devra gérer plusieurs Cloud ce qui peut s'avérer vite très complexe.

### ➤ **Les conditions de service**

Il faut vérifier que les conditions de services sont conformes aux exigences de l'entreprise. Vu que ces contrats sont très détaillés, mieux vaut lire tout en détail avant de prendre une décision.

Le cloud computing n'est une mauvaise chose, bien au contraire, mais il est nécessaire avant de le mettre en place de peser le pour et le contre. En effet, certaines PME, de par leur activité, leur localisation, mais également par leurs applications utilisées, pourraient y percevoir plus de désavantages que de bénéfices.

## **I.3.8. Les risques du Cloud Computing**

Si le Cloud présente de nombreuses opportunités pour les entreprises, de nombreux problèmes peuvent freiner son expansion et compromettre son succès. En effet, plusieurs risques liés à l'utilisation du Cloud ont été déjà identifiés et des référentiels comme la norme ISO 27000 aident à les prévoir et à les gérer. La nature de ces risques dépend de nombreux facteurs et peut être liée au modèle de déploiement utilisé (SaaS, PaaS, IaaS). Les grandes catégories de risques liées à l'adoption du Cloud se déclinent selon les points suivants :

- **Confidentialité, intégrité et disponibilité** : les préoccupations majeures lors de la migration des applications vers le Cloud sont la confidentialité, l'intégrité et la disponibilité des données. En effet, il est primordial pour une entreprise de prévoir et de prévenir les risques liés à la perte et la confidentialité de ses données. Des clients sont de plus en plus exigeants en ce qui concerne les normes de sécurité et imposent à leurs fournisseurs des certificats comme la norme ISO 27000 ou SAS 70.  
Un autre point de défiance vis-à-vis du Cloud est celui de la garantie de l'intégrité des données. Les données doivent être fiables et converties au bon format, sinon de mauvaises interprétations sont susceptibles de se produire. Quant à la disponibilité, les données déployées dans le Cloud doivent être en permanence accessibles même lorsque des opérations de maintenances sont effectuées par les fournisseurs (l'une des solutions utilisées est la répliquion des données). Par ailleurs, il est indispensable de garantir aux utilisateurs la disponibilité des ressources en cas de problèmes (défaillance par exemple) autres que le cas de non disponibilité.
- **Localisation des données** : la dématérialisation des données sur différents sites physique de stockage peut conduire à un éclatement des données et leur répartition sur différentes zones géographiques (par exemple pays).

- **Malveillance** : les architectures du Cloud sont gérées par des personnes ayant des privilèges très élevés qui sont donc à risque très élevé.
- **Usurpation** : deux types d'usurpation peuvent être distingués : (i) usurpation de service offert par l'architecture et (ii) usurpation d'identité d'utilisateurs (clients) de services.

## I.4. Conclusion

À travers ce chapitre, nous avons pu développer le vaste sujet sur la généralité en rapport avec le cloud computing et la virtualisation

Le cloud computing est donc un moyen de délivrer un service informatique ciblé et quantifié à une clientèle précise sans que cette dernière n'ait à investir dans un système d'information dédié à ce service. Selon cette définition, nous avons vu des modèles tels que le PaaS, SaaS, IaaS. Ces modèles reposent sur des technologies de virtualisation qui permettent d'atteindre la flexibilité requise à la réalisation de ces modèles. La virtualisation est une technologie clé du cloud computing.

La virtualisation permet d'assurer une très grande souplesse au niveau des ressources allouables à une solution ou à un client. L'indépendance des solutions matérielles et logicielles permet de donner toute la puissance requise à la bonne exécution du service.

Le cloud est une technologie jeune et innovante dont l'attractivité est en hausse. Ce procédé peut changer les méthodes de travail, réduire les coûts, modifier les interactions client/fournisseur ainsi que changer le modèle économique.

Le marché du cloud computing est encore en plein essor. Selon le cabinet Gartner, le marché du cloud computing a été de 68,8 milliards de dollars en 2010 et pourrait s'élever à 148,8 milliards de dollars en 2014. Une des raisons de ce succès est le fait que le marché se soit adapté, permettant aux fournisseurs de transmettre des services à travers le réseau. Ainsi il est important de prêter une attention particulière à ces technologies qui vont bouleverser la manière de gérer les systèmes d'information.

---

# CHAPITRE 2 :

---

## L'ALLOCATION DE RESSOURCES ET LA MIGRATION DES MACHINES VIRTUELLES

# Chapitre II : L'allocation de ressources et la migration des VMs

## II.1. Introduction

Le cloud computing est de plus en plus reconnu comme une nouvelle façon d'utiliser, à la demande, les services de calcul, de stockage et de réseau de manière transparente et efficace. Les Datacenter du cloud computing émergent comme de nouveaux candidats pour remplacer les Datacenter traditionnels qui grandissent rapidement en nombre et en capacité pour rencontrer un nombre grandissant pour la virtualisation des ressources et des stockages. A cette expansion résulte clairement une consommation énergétique exponentielle de la part de ces centres de données. C'est dans cette logique que s'inscrit ce chapitre décrit les notions de l'allocation des ressources et la migration des machines virtuelles dans le cloud computing, notions qui aident dans la diminution de la consommation énergétique des Datacenter.

La migration de machines virtuelles est une des fonctionnalités les plus significatives introduite par les infrastructures virtualisées qui est beaucoup plus basée sur les techniques d'allocation de ressources notamment les ressources virtuelles (Machines virtuelles). Le chapitre suivant est divisé en trois principales sections. Dans la première section nous parlerons de l'allocation des ressources du cloud computing en particulier nous présenterons le contexte, suivi de la définition, des techniques de l'allocation des ressources pour enfin parler de l'allocation en elle-même. Quant à la deuxième nous allons parler de la consommation énergétique au niveau du cloud en passant de la description d'un Datacenter à la consommation proprement dite de l'énergie. Pour la troisième section nous allons définir migration, étudier les différentes techniques de la migration notamment la migration en temps réel (à chaud) après avoir cité les différents avantages et désavantages de migrer une machine virtuelle.

## II.2. L'allocation des ressources dans le Cloud

Les Ressources de Cloud peuvent être vues comme n'importe quelle ressource (physique ou virtuelle) que les utilisateurs peuvent demander du cloud, Par exemple, les utilisateurs peuvent demander des exigences de réseau, telles que la bande passante et les délais, et des exigences computationnelles, telles que le processeur, la mémoire et le stockage. En général, les ressources sont situées dans un centre de données qui est partagé par plusieurs clients, et doivent être attribués et ajustés dynamiquement en fonction de la demande [2.0].

Les ressources de cloud consistent en ressources physiques et virtuelles. Les ressources physiques sont partagées entre plusieurs demandes grâce à la virtualisation et le provisioning. La demande des ressources virtualisées est décrite par un ensemble de paramètres détaillant les besoins de traitement, de mémoire et de disque. Le provisioning satisfait la demande de ressources par le mapping (la correspondance) les ressources virtualisées sur les ressources physiques. Les ressources matérielles et logicielles sont allouées aux applications cloud sur la base de la demande" [2.1].

L'allocation des ressources est un sujet qui a été abordé dans de nombreux domaines informatiques, tels que les systèmes d'exploitation, le Grid Computing<sup>6</sup>, la gestion des Datacenters (centre de données) et le cloud computing. Les objectifs les plus importants de cloud est d'intégrer les ressources inutilisées pour créer des pools de ressources partagées et virtualisées, rendre l'accès à la demande des utilisateurs aux ressources commodément, améliorer l'utilisation des ressources, etc. Cependant, les ressources disponibles des fournisseurs et les exigences des consommateurs en ressources sont à la fois variées dynamiquement. Par conséquent, définir un mécanisme d'allocation des ressources aux utilisateurs d'une manière souple, dynamique et fiable est l'un des principaux enjeux dans le Cloud Computing.

### II.2.1. Présentation du contexte

De nos jours, il y a eu une augmentation dramatique dans la popularité des systèmes du cloud computing qui offrent des ressources informatique à la demande, basent sur la facturation à l'usage, afin que les utilisateurs puissent augmenter ou diminuer leur taux de consommation de ressources en fonction de leurs besoins. Ces environnements peuvent multiplexer de nombreux utilisateurs sur la même infrastructure physique. Dans les environnements de cloud computing, il y a deux acteurs : les fournisseurs et les utilisateurs de cloud computing. D'une part, les fournisseurs détiennent des ressources informatiques massives dans leurs grands centres de données et louent ces ressources à des utilisateurs. D'autre part, il y a les utilisateurs qui ont des applications avec des charges variées et louent des ressources de la part des fournisseurs pour exécuter leurs applications. La Figure suivante montre l'interaction entre les fournisseurs et les utilisateurs.

Tout d'abord, l'utilisateur envoie une demande contenant ses besoins en ressources au fournisseur. Lorsque ce dernier reçoit la demande, il cherche des ressources pour satisfaire la demande et alloue ces ressources à l'utilisateur demandeur, généralement sous forme des machines virtuelles (VM). Ensuite, l'utilisateur utilise les ressources assigné a lui pour exécuter leur applications et paie les ressources qui a utilisé. Lorsque l'utilisateur termine avec ces ressources, il les retourne au fournisseur. L'un des aspects intéressants du cloud computing est que ces acteurs ont leurs propres intérêts. En règle générale, l'objectif des fournisseurs est de maximiser le plus possible leurs revenus avec un investissement minimum. À cet effet, les fournisseurs veulent maximiser l'utilisation de leurs ressources informatiques, à titre d'exemple, par l'hébergement du plus grand nombre que possible de machines virtuelles sur chaque machine. D'autre part, les utilisateurs veulent accomplir leur travail à un coût minimal ou, en d'autres termes, ils veulent maximiser leur performance économique. Chacun de ces deux partis ne veut pas partager ces informations avec l'autre, ce qui rend l'allocation optimale des ressources plus difficile. Par exemple, les fournisseurs ne veulent pas exposer combien et quel genre de machines qu'ils ont et comment ils sont connectés. De même, les utilisateurs ne veulent pas exposer les détails de leur charge de travail, y compris les codes sources et les ensembles de leurs propres données à d'autres personnes, y compris les fournisseurs. Par conséquent, les utilisateurs ne peuvent pas exprimer leurs demandes de ressources de sorte que les ressources allouées soient optimales, car ils ne savent pas exactement ce qui est disponible. De même, les fournisseurs ne peuvent pas allouer des ressources de manière plus appropriée aux demandes des utilisateurs, car ils n'ont pas d'informations sur les charges de travail de ces derniers.

---

<sup>6</sup> C'est le fait de mettre en commun des ressources logicielles et matérielles distribuées (ensemble que l'on appelle la « grille ») afin de fournir une puissance de calcul importante

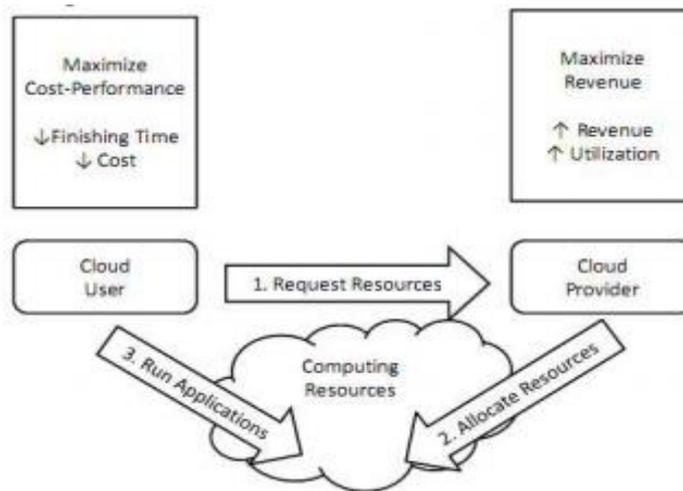


Figure 9: Utilisation des ressources du Cloud

Avec un nombre limité de ressources physiques disponibles, l'allocation des ressources devient un défi pour les fournisseurs des services cloud.

Les fournisseurs de service cloud doivent décider combien de ressources virtuelles doivent être créées basé sur les requêtes des utilisateurs du cloud. Aussi ils doivent décider quelle VM doit être mappée à quelle machine qui doit la prendre en charge ; à quel moment la migration des VMs doit être faite. L'ultime but du fournisseur cloud est donc de maximiser le profit, donc l'utilisation des ressources. Par contre, celui de l'utilisateur cloud sera de minimiser le paiement en louant les ressources.

Ces paramètres sont à prendre en considération quand on fait l'allocation des ressources. En allouant des ressources à l'utilisateur cloud, la sous-utilisation des ressources due au sur-provisionnement et la surutilisation des ressources dues au sous-provisionnement devraient être évités.

L'allocation des ressources devrait aussi considérer des paramètres importants comme les paramètres de qualité de service. Nous pouvons citer : Le temps de réponse, La performance, la fiabilité, la disponibilité, la sécurité

- **Le temps de réponse** : Pour les applications interactives, le temps est un facteur très important. Le système doit bien répondre pour ce genre d'applications.
- **La performance** : Pour certaines applications le système doit être performant pour fournir un service digne de l'attente de l'utilisateur.
- **La fiabilité** : Le système doit être fiable pour la bonne utilisation par l'utilisateur. Il ne devrait pas être un casse-tête pour ce dernier.
- **La disponibilité** : Le fournisseur cloud doit être en mesure d'allouer les ressources peu importe quand celles-ci sont demandées
- **La sécurité** : Pour des applications critiques comme les applications de transaction en ligne, le système doit être

## II.2.2. Définitions

Dans le cloud computing, l'allocation des ressources est le processus d'attribution des ressources disponibles pour les requises des applications de cloud. Les ressources cloud peuvent être provisionnées à la demande à grains fins, d'une manière multiplexée. Dans le cloud l'allocation des ressources est basée sur l'infrastructure en tant que service (IaaS) [2.2].

Dans le cloud computing, l'allocation des ressources est le processus d'attribution des ressources disponibles requises pour des applications de cloud via Internet. L'allocation des ressources affame des services si l'allocation n'a pas été gérée avec précision

### **II.2.3. Allocation des VMs**

L'allocation de VM (provisioning) est le processus de création d'instances VM sur les hôtes qui correspondent aux caractéristiques critiques (stockage, mémoire), aux configurations (environnement logiciel) et aux exigences (zone de disponibilité) du fournisseur SaaS. Autrement dit, elle est l'action de rotation et fourniture de VM sur un hôte physique de Datacenter dans lequel la VM peut contenir des plusieurs caractéristiques telles que la CPU, RAM, Stockage, VLAN ou Adresse IP, etc.

Le placement de VM (placement de VMs sur les PMs) est un sujet largement étudié. Les principaux objectifs de la majorité de ces études sont de consolider les VMs sur les serveurs pour l'efficacité énergétique et l'utilisation du serveur. Nécessairement, notre travail se focalisera sur le placement des VMs afin de réaliser la migration de ces dernières ; deux sujets qui sont largement dépendants.

### **II.2.4. Techniques d'allocation des ressources**

Il existe deux principaux types d'allocation des ressources : L'allocation statique et l'allocation dynamique

#### **II.2.4.1. Allocation statique des ressources**

Pour l'allocation statique, l'utilisateur cloud doit effectuer une demande à l'avance pour ces ressources. En réalisant cela, l'utilisateur sait quelles ressources sont requises et combien d'instances de ressources sont nécessaires avant d'utiliser le système.

L'inconvénient de ce cas est qu'il conduit à la sous-utilisation ou à la surutilisation des ressources dépendamment du temps pendant lequel l'application s'exécutera.

#### **II.2.4.2. Allocation dynamique des ressources**

Les ressources cloud sont demandées par l'utilisateur au fur et à mesure que l'application en aura besoin. Ici la sous-utilisation et la surutilisation sont évitées au maximum possible.

Le fournisseur Cloud doit allouer les ressources de la part des autres Datacenter participant au cloud. C'est dans cette optique que la migration des machines virtuelles (VM) est enclenchée pour équilibrer les charges des Datacenter. Des Datacenter surutilisés vers les Datacenter sous-utilisés pour éviter ces cas de figure.

## **II.3. La consommation énergétique au niveau du Cloud**

Comme toute infrastructure basée sur les matériaux informatiques, le domaine du Cloud computing a un grand souci de la consommation d'énergie. Cela est un autre défi pour le fournisseur du Cloud et le gouvernement qui ont un objectif de produire un service sophistiqué qui répond aux attentes des clients tout en réduisant la consommation d'énergie.

### II.3.1. Le Datacenter

Un data center dans le cloud computing, appelé aussi centre de traitement de données est un site physique sur lequel se trouvent regroupés des équipements constituant le système d'information d'une entreprise ou d'un fournisseur de cloud (ordinateur puissant, serveurs rangés dans les baies, baies de stockage, équipement réseaux, etc...) [2.4].

Les grands Datacenter sont généralement construits pour soutenir la demande croissante de calcul et de stockage de données d'entreprise en plein élargissement. Il y a environ 23 000 Datacenter dans le monde. En 2014, le marché devrait croître à environ 343,4 milliards de dollars américains [2.5]. Nous prenons par exemple le data center de Facebook qui se trouve en Suède, à l'intérieur, le bâtiment est divisé en quatre salles contenant des dizaines de milliers de serveurs, rangés sur de hautes étagères et baignant dans une lumière bleue futuriste. Leur nombre exact et leur puissance de calcul sont des secrets industriels [2.6]. Ces centres de données consomment généralement une énorme quantité d'énergie électrique qui entraîne des coûts opérationnels élevés et des émissions de CO<sub>2</sub>.

### II.3.2. La consommation d'énergie dans le data center

Des applications à forte intensité de données pour le traitement de grandes données sont hébergées dans le cloud par le data center. Étant donné que l'environnement cloud fournit des ressources virtualisées pour le calcul et les applications nécessitent une communication entre les nœuds informatiques, le placement des VMs. La majorité des travaux de recherche rapportés dans la littérature actuelle tiennent compte de la sélection des hôtes physiques pour placer les données et les machines virtuelles comme des problèmes indépendants.

La consommation d'énergie conséquente de Datacenter s'explique généralement par deux raisons. D'une part, les serveurs de stockage ne peuvent pratiquement jamais être arrêtés pour ne pas perdre les données confiées par les clients des Datacenter ou du cloud. Ils fonctionnent presque 24 heures sur 24, toute l'année. D'autre part, ces ordinateurs puissants qui accueillent de VMs sont des appareils électroniques qui produisent énormément de chaleur. Leur bon fonctionnement est assuré par des systèmes de climatisation très énergivores.

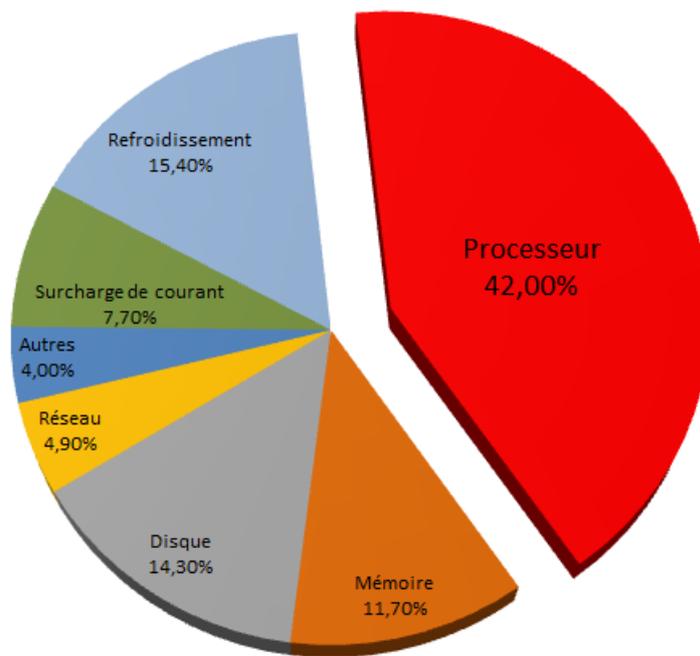


Figure 10: Répartition de consommation énergétique d'un centre de données [2.6]

Cette figure nous montre que les processeurs sont les plus gourmands dans la consommation énergétique des centres de données. Sachant qu'un processeur ne fonctionne jamais à 100 % du temps, la consommation d'énergie pendant les périodes de repos est importante en raison du courant de fuite. De plus, le fait de rendre inactif un processeur n'est pas toujours le bon choix si son mode de veille nécessite un temps de réveil trop long, voire plus long que sa période de repos.

Notre travail se focalise plus précisément sur l'utilisation de CPU (processeur) et la bande passante (Réseau) qui consiste à réduire une surutilisation au niveau d'une PM hôte par le biais de la migration de VMs. Nous verrons les différentes techniques de migration de VM dans la section qui suit.

## II.4. La migration des VMs

La migration de VM signifie déplacer l'état de la machine virtuelle, c'est-à-dire sa mémoire, son CPU, les entrées/sorties... d'un hôte physique vers un autre hôte de destination via un réseau. Elle est exploitée quotidiennement par les gestionnaires de machines virtuelles pour répartir la charge, améliorer les performances, réduire la consommation énergétique ou préparer la maintenance de serveurs en production.

### II.3.1. Les avantages et les inconvénients de la migration des VMs

#### II.3.1.1. Les avantages

Les différents avantages de la migration des machines virtuelles dans le Cloud computing sont généralement les suivants :

- **Réduction de la consommation d'énergie**

Ces serveurs du cloud utilisent généralement jusqu'à 70% de leur consommation énergétique maximale, même avec une faible utilisation des ressources. Par conséquent, l'utilisation des techniques de migration conservent l'énergie consommée par les serveurs par une utilisation optimale des ressources.

➤ **Tolérance aux pannes**

Avec la nouvelle méthode de migration à chaud, il est possible de déplacer une VM d'un serveur physique vers un autre, sans même qu'on ait besoin de la stopper.

Cette fonctionnalité est un élément clé pour améliorer le taux de disponibilité de vos services et applications. L'utilisation de deux serveurs physiques permet ainsi de doubler aisément une infrastructure virtualisée (redondance). En cas de panne d'un des deux serveurs, les VM seront automatiquement déplacées vers le second.

➤ **Répartition de charge**

La migration des machines virtuelles, particulièrement la migration à chaud, entre serveurs physiques a un autre avantage : elle permet de répartir la charge de travail entre les serveurs. Lorsqu'une VM monte en charge de façon extrême, les autres pourront se replier sur un serveur physique moins sollicité.

Les tâches critiques pourront également fonctionner au sein d'une VM disposant de plus de cœurs de CPU (virtuels), de mémoire (virtuelle) et d'espace disque (virtuel) que les autres. Il est ainsi possible de moduler la taille des VM en fonction des tâches qu'elles devront accomplir.

➤ **Avantages économiques et financiers**

Du point de vue économique et financier, cette migration en temps réel des VMs peut avoir des retombées économiques positives pour les fournisseurs et les responsables des services informatiques. Parce qu'elle permet de générer des flux monétaires supplémentaires dus aux économies d'échelle résultant de la création et l'hébergement simultanés de plusieurs VMs sur une ou plusieurs machines physiques plus performantes; de plus, elle permet également de meilleures répartitions de charge en migrant des VM (hébergées sur des machines physiques hôtes peu performants) vers de nouveaux hôtes plus performants. Ainsi, une entreprise peut diminuer ses coûts d'immobilisations corporelles, augmenter son chiffre d'affaires grâce à l'amélioration de la qualité de ses services, augmentant ainsi sa marge bénéficiaire nette par rapport à ses ventes brutes. Les coûts d'exploitation, liés aux coûts d'immobilisations corporelles dans les divisions des services informatiques des entreprises, représentent généralement une grande part du budget du fonctionnement d'une entreprise, et ils sont dus aux loyers, achats et aux amortissements des équipements informatiques physiques et logiciels, aux loyers des locaux de ces machines physiques, etc...

### **III.3.1.2. Les inconvénients**

L'inconvénient majeur de la migration réside dans la nécessité d'arrêter ou de mettre en pause la VM pendant toute la durée de son déplacement. Les applications et services qui s'exécutent sur les VMs sont généralement critiques et il est donc difficilement envisageable de les arrêter même pendant une courte période.

Lorsqu'une VM est suspendue, ses connexions réseaux actives sont arrêtées et la durée d'interruption nécessaire à sa migration est trop élevée pour permettre la reprise des connexions

dans un état cohérent. Il est ainsi généralement nécessaire de redémarrer entièrement les services s'exécutant sur la VM pour pouvoir reprendre un fonctionnement normal.

## II.3.2. Les différentes méthodes de migration des VMs

Les techniques de migration de machines virtuelles sont classées globalement en trois parties. La migration à froid, la migration à chaud et la migration hybride.

### II.3.2.1. La migration à froid

Cette stratégie est la plus basique, il est nécessaire de mettre la machine virtuelle hors tension sur la machine source et son exécution avec sa mémoire sont migrées vers une autre machine de destination. Une fois toutes les pages mémoires transférées, la VM est activée sur la machine de destination. Dans cette migration, le temps total de migration est égal au temps d'arrêt de la VM.

La décision de migration se fait sur base d'un calcul des requêtes entrantes, de l'utilisation du processeur, des ressources disponibles sur l'hôte physique ainsi que de la consommation énergétique.

L'avantage principal de cette méthode est qu'elle n'impliquera aucune faute lors de la migration de la mémoire. La machine étant arrêtée, les services ne seront plus disponibles et la mémoire ne sera pas altérée sur l'hôte source. La machine une fois migrée pourra reprendre son activité dans le même état que lorsqu'elle s'est arrêtée.

Elle comporte par contre un désavantage certain. À partir du moment où la machine est arrêtée, les services fonctionnant sous celle-ci ne seront plus disponibles durant toute la durée de la migration. Ce qui pose évidemment des problèmes majeurs pour des applications nécessitant une haute disponibilité. De plus, la mémoire étant transférée entièrement sur l'hôte cible, la migration à froid provoque un ralentissement sur tout le réseau du cloud.

### II.3.2.2. La migration à chaud ou la migration en temps réel

La migration à chaud (Live Migration) est un outil utile pour la migration des systèmes d'exploitation complets ainsi que leurs applications entre les machines physiques éloignées des Data Centers. C'est une technologie impressionnante qui facilite l'équilibrage de charge, la gestion des pannes, la maintenance du système et la réduction de la consommation d'énergie. La migration à chaud des machines virtuelles est le processus de transférer une ou plusieurs machines virtuelles d'une machine physique ou d'un espace de stockage vers un autre sans interrompre les autres VMs.

Le coût de la migration à chaud, du point de vue de l'administrateur de la VM, est ainsi principalement lié à l'utilisation des ressources réseaux du centre de données. Du point de vue de l'utilisateur final de la VM, une diminution temporaire de performance est le principal coût imposé par la migration à chaud. Cette perte de qualité de service est matérialisée par une faible durée d'interruption de la VM ainsi que d'une possible diminution de la bande passante réseau allouée pendant la durée de migration.

Parmi les différentes approches de migration à chaud, on retrouve deux principaux algorithmes, « l'algorithme de pré-copie » et « l'algorithme de post-copie ». Les termes pré- et post-copie font référence à la phase de mise en pause de la VM. La pré-copie consiste à transmettre les

données de la VM sur l'hôte destination avant son réveil alors que la post-copie effectue la transmission des données seulement après le réveil sur l'hôte destination

### ➤ L'algorithme de pré-copie

L'algorithme de pré-copie représenté sur la figure ci-dessous, est l'implémentation la plus répandue de la migration à chaud. Il consiste principalement à transmettre l'ensemble des pages mémoire de la VM vers l'hôte de destination tout en garantissant une faible durée d'interruption.

L'algorithme de pré-copie est un processus de transmission itératif. La première étape consiste à envoyer l'ensemble des pages mémoire utilisées vers le serveur de destination. Ce dernier réceptionne les pages et les stocke en mémoire en reconstituant ainsi l'état mémoire de la VM tel que capturé sur le serveur source. La VM étant toujours en cours d'exécution, certaines de ses pages mémoire sont modifiées par les applications durant le transfert et doivent donc être renvoyées. C'est ainsi que le processus itératif de l'algorithme de pré-copie se dévoile, en effet les pages mémoire modifiées durant chaque transfert sont envoyés lors de la prochaine itération. Ainsi, la quantité de pages mémoire à transférer diminue au fur et à mesure des itérations jusqu'à atteindre la dernière étape d'arrêt et copie.

L'étape d'arrêt et copie consiste à mettre en pause la VM sur le serveur source, transférer les dernières pages modifiées ainsi que les données annexes, et enfin réveiller la VM sur le serveur destination. Cette étape intervient sous différentes conditions, en fonction des contraintes d'interruption définies par l'administrateur ou encore de l'implémentation faite dans l'hyperviseur utilisé. Dans la majorité des cas, l'administrateur fixe un temps d'interruption maximal autorisé pour la migration d'une VM. Ainsi, au fur et à mesure des itérations, l'hyperviseur estime la durée nécessaire pour envoyer les nouvelles pages modifiées. La migration passe alors en phase d'arrêt et copie si le temps nécessaire estimé pour l'envoi des pages restantes et des données annexes est inférieur à la durée d'interruption maximale exigée.

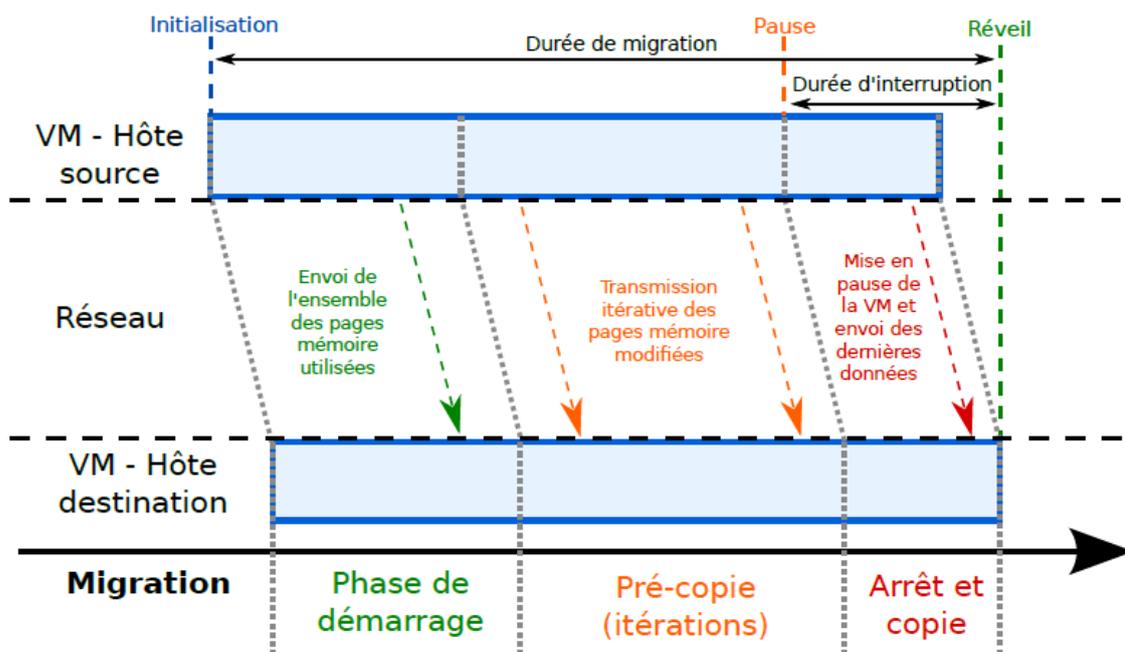


Figure 11: Algorithme de pré-copie

- Phase de démarrage :

Durant cette phase, l'hyperviseur copie toutes les pages mémoire de l'hôte source à l'hôte cible alors que la machine virtuelle est toujours opérationnelle. Si une ou plusieurs pages mémoire sont modifiées, elles seront recopiées d'un hôte à un autre durant la phase itérative suivante jusqu'au moment où le taux de pages sales deviendra inférieur au taux de pages copiées.

- Phase pré-copie itérative :

Durant cette phase, l'hyperviseur copie les pages de mémoire sur l'hôte distant via TCP. Lors de ce transfert, si une page de mémoire est modifiée sur l'hôte source, elle est marquée comme "sale". Cette étape est répétée, on ne copie plus que les pages qui ont été modifiées durant la pré-copie précédente. Le cycle s'arrête lorsqu'un nombre d'itérations maximum a été atteint (29 fois par défaut pour Xen) ou bien lorsque ce sont toujours les mêmes pages qui sont sales. En moyenne, l'hyperviseur Xen itère 9 fois avant de passer à l'étape suivante.

- Phase d'arrêt et copie :

Lors de cette étape, la machine virtuelle est arrêtée sur l'hôte source et les pages sales restantes sont copiées sur l'hôte cible. C'est lors de cette étape que la machine virtuelle ne sera plus accessible. Elle ne sera active sur aucun hôte. Cette période est appelée « **downtime** », c'est ce temps-là qu'il est important de minimiser.

Cette approche possède des avantages et des inconvénients. L'avantage majeur réside dans l'assurance d'un temps d'interruption contrôlé et borné pour la VM, ce qui permet par exemple de garantir un taux de disponibilité fiable entre un fournisseur de service et ses clients. Le problème majeur intervient lorsque la durée d'interruption maximale est trop faible pour pouvoir envoyer les dernières pages modifiées au serveur destination. Dans ce cas la migration peut durer indéfiniment.

➤ **II.4.2.2. L'algorithme de post-copie**

L'algorithme de post-copie, à l'inverse de la pré-copie, débute par l'arrêt immédiat de la machine virtuelle sur le serveur source puis transfère uniquement les données annexes avant de réveiller la VM sur le serveur de destination. Il est représenté sur la figure suivante :

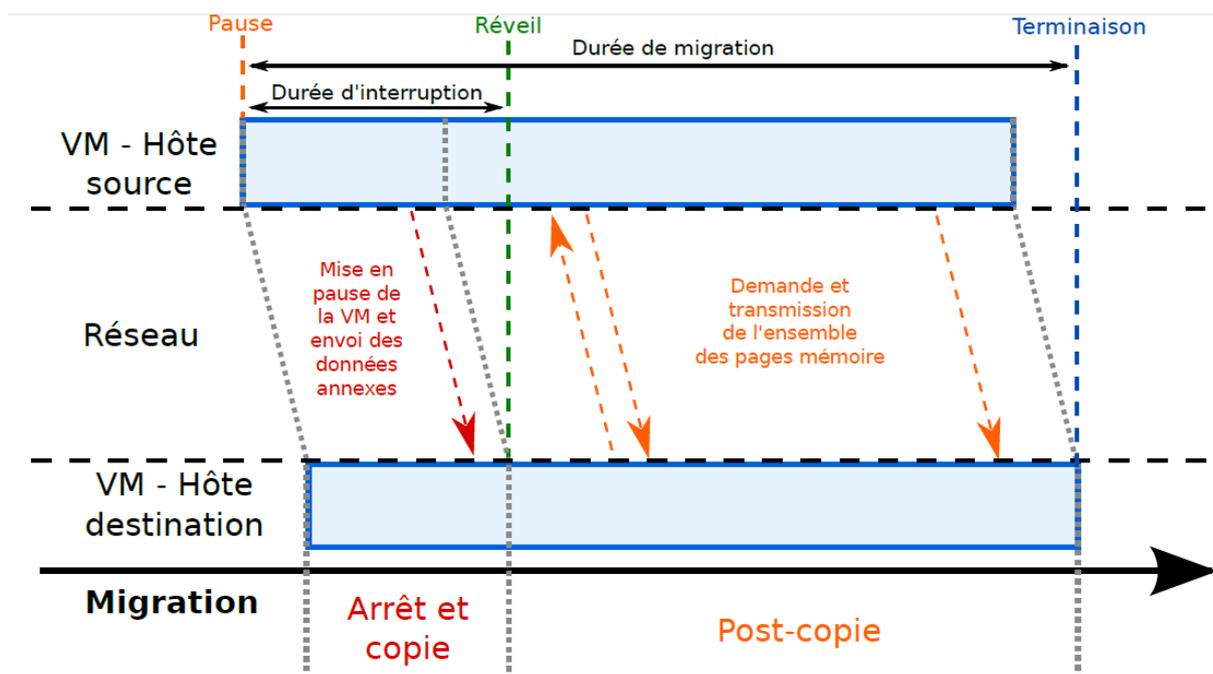


Figure 12: Algorithme de post-copie

La durée d'interruption de la VM est très faible en comparaison de l'algorithme de pré-copie puisqu'elle ne dépend que de la seule transmission des données annexes vers le serveur destination. Ainsi, la fréquence de modification des pages mémoire n'impacte plus la durée d'interruption de la VM à migrer. Cependant, la VM est réveillée sur l'hôte cible sans aucune donnée mémoire. Elle doit alors récupérer les pages mémoire manquantes au travers du réseau selon différents procédés entraînant une perte de performance plus ou moins importante. En effet, la perte de performance occasionnée peut engendrer des interruptions de connexions réseaux et ainsi nuire fortement aux services s'exécutant sur la VM. Plusieurs techniques sont utilisées pour récupérer les pages mémoire manquantes, ces techniques peuvent être combinées et utilisées conjointement pour limiter la perte de performance occasionnée à la VM. Ces techniques sont les suivantes :

- La récupération à la demande

Le serveur de destination initie la demande de récupération d'une ou plusieurs pages mémoire manquantes au serveur source lorsque la VM en a besoin. Cette méthode est contraignante car elle induit un délai d'attente entre le moment où la VM a besoin d'une page mémoire et le moment où elle est récupérée sur l'hôte de destination au travers du réseau. Ce comportement provoque un ralentissement du fonctionnement de la VM.

- La poussée active

Cette méthode permet de transférer en continu les pages mémoire restantes vers le serveur de destination à l'initiative du serveur source. Plus important, la poussée active permet d'envoyer en priorité les pages mémoire ayant une dépendance résiduelle avec les pages déjà demandées. Dès lors les risques de ralentissements liés à la récupération à la demande sont réduits. La durée totale de migration est ainsi minimisée et le temps de ralentissement de la VM réduit en faisant parvenir l'ensemble des pages mémoire manquantes le plus rapidement possible au serveur de destination.

- La pré-pagination

Cette dernière méthode permet d'améliorer la poussée active en prédisant les prochaines pages mémoire qui ont le plus de chances d'être accédées rapidement par la VM. En effet, en se basant sur les répétitions et schémas d'accès aux pages mémoire, cette méthode réduit le risque d'erreurs réseau mais aussi la durée de la dernière phase de la post-copie.

Les avantages de la post-copie sont :

- La durée d'interruption est fixe et très faible quelle que soit l'activité mémoire à transférer.
- La migration des VM ayant un fort taux d'écriture mémoire est avantageuse en garantissant une durée d'interruption minimale.

L'inconvénient majeur par rapport à la post-copie repose sur la robustesse de l'approche. En effet, étant donné que la VM est directement réveillée sur l'hôte de destination dans un état inconsistant, la défaillance de l'un des deux hôtes, source ou destination, en cours de migration entraîne la perte inévitable de l'intégrité de l'état mémoire de la VM. À l'inverse, en pré-copie, la défaillance de l'hôte de destination en cours de migration n'a aucun impact sur la VM qui continue alors son exécution normale sur l'hôte source sans aucune perte de données.

- L'algorithme de post-copie hybride

Afin de réduire les problèmes de performances de la post-copie, une variante nommée post-copie hybride a été définie. Cet algorithme est considéré comme une post-copie hybride car il reprend les principales caractéristiques de la post-copie en ajoutant une étape de la pré-copie comme illustré par la figure qui suit :

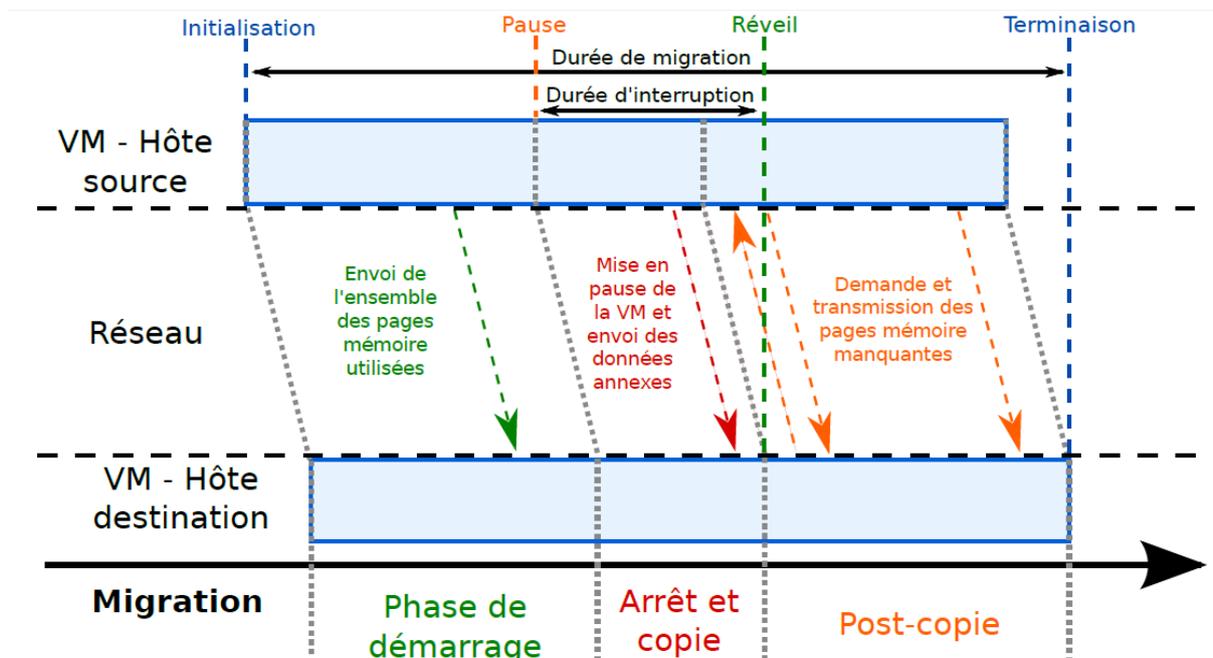


Figure 13: Algorithme de post-copie hybride

En post-copie hybride, une seule phase de pré-copie est effectuée avant la phase de post-copie, de cette manière l'hôte cible dispose déjà de toute la mémoire, seules les pages sales devront

être transférées en post-copie, ce qui restreint la perte de performance provoquée à la VM. Pendant la copie des pages mémoire d'un hôte à un autre, la machine virtuelle continue de fonctionner sur l'hôte source. Une fois la copie effectuée, la machine virtuelle est stoppée, l'état du processeur et les pages modifiées sont copiées à leur tour sur la machine cible. La machine virtuelle est ensuite redémarrée sur l'hôte cible et la phase de post-copie commence.

Ce système fonctionne bien pour les migrations de VM nécessitant un grand nombre de lectures mémoire, tout comme la migration par post-copie. Il fournit également un temps de migration déterministe pour les VMs nécessitant un grand nombre d'écritures mémoire.

### **II.3.2.3. Amélioration du processus de migration de VMs**

Parmi les techniques mises au point pour améliorer et parfaire le processus de migration à chaud, on retrouve globalement l'idée de restreindre la quantité de données à transmettre sur le réseau afin de réduire la durée de migration et la durée d'interruption de la VM.

#### **➤ L'auto-ballonnement dynamique**

Elle consiste à allouer dynamiquement de la mémoire à une machine virtuelle à la demande en fonction de ses besoins. L'allocation dynamique de mémoire permet d'abord d'augmenter la consolidation dans les centres des données virtualisées grâce au sur-engagement des ressources mémoire (memory overcommitment). Il est ainsi possible d'héberger une quantité plus importante de VMs en ajustant l'allocation mémoire de manière à pouvoir absorber les pics de consommation temporaire de l'ensemble de VMs sans saturer les ressources matérielles des hôtes. Cette fonctionnalité n'est cependant pas sans risques car elle suppose que toutes les VMs hébergées sur un hôte physique n'atteindront jamais leurs consommations maximales en même temps.

Au niveau de la migration à chaud, l'auto-ballonnement dynamique permet de limiter la transmission des pages à la quantité de mémoire réellement utilisée par la machine virtuelle, et donc de réduire sa durée de migration en évitant l'envoi de pages vides.

#### **➤ La compression de page mémoire**

Elle consiste à compresser les pages mémoire sur l'hôte source avant leur transfert, puis les décompresser une fois réceptionnées sur l'hôte destination. L'inconvénient de cette approche étant le temps et les ressources nécessaires à la (dé)compression des pages mémoire. Afin d'améliorer cette technique, les auteurs ont ensuite décidé d'utiliser différents algorithmes de compression de manière dynamique en fonction du taux de similarité des pages pour obtenir un meilleur ratio compression/performance.

#### **➤ La compression du delta**

C'est une méthode consistant à ne transmettre que les différences entre les versions des pages mémoire sous forme compressée dans le but de réduire la quantité de données transférées. Une page mémoire contient des données sous forme binaire. Un XOR est alors appliqué entre la version courante d'une page et celle de l'itération précédente pour obtenir son delta. Le delta a la même taille que la page mémoire d'origine mais peut être compressée bien plus efficacement car de manière générale la proportion des modifications faites sur les pages mémoire, entre deux itérations, reste faible.

### ➤ La déduplication de données

C'est une des méthodes performantes permettant l'amélioration du processus de migration à chaud. Certaines pages mémoire pouvant être identiques, la déduplication de données consiste à identifier les données redondantes pour ne les transmettre qu'une seule fois. Des tables de hachage indexant les parties communes (par portions de 64 octets) des pages mémoire sont alors créées et maintenues sur les deux hôtes source et destination. Comme pour la compression de données, la déduplication utilise des algorithmes d'encodage pour éliminer les données redondantes. Les algorithmes les plus performants consommant beaucoup de ressources, des compromis doivent être faits entre l'efficacité de l'encodage et le temps de calcul requis pour pouvoir réellement réduire les durées de migration.

### II.3.3. Les modèles de performance de migration à chaud

La performance de la migration à chaud est principalement caractérisée par la durée de l'opération, la durée d'interruption de la VM, mais aussi de la quantité de données transmises sur le réseau pendant la migration. Plus récemment, la consommation énergétique de la migration est devenue un nouveau critère de performance.

De nombreux modèles de prédiction permettant d'estimer les performances de la migration à chaud ont été définis. Parmi les différentes approches existantes, la distinction majeure se situe au niveau de la caractérisation de l'activité mémoire de la machine virtuelle et ainsi de la modélisation de la quantité de pages mémoire modifiées durant le processus de migration. En effet, le taux d'écriture des pages mémoire varie plus ou moins fortement en fonction des applications en cours d'exécution, il est donc difficile à modéliser.

#### II.3.3.1. Modèles linéaires

Selon les principes de la migration à chaud, la durée de migration  $D_{mig}$  évolue linéairement par rapport à la quantité de mémoire à transmettre  $M_{mig}$ , et la bande passante disponible pour la migration  $BP$ . La durée d'interruption occasionnée à la VM  $I_{mig}$ , généralement très faible, n'a que très peu d'impact sur la durée de migration.

La formule qui suit nous donne la durée de migration estimée :

$$D_{mig} = \frac{M_{mig}}{BP} + I_{mig}, \quad (f.1)$$

$BP$  : Bande passante

$D_{mig}$  : Durée de migration

$I_{mig}$  : Interruption de migration

$M_{mig}$  : Quantité de mémoire à transmettre

Certains outils de simulation tel que CloudSim [2.7] par exemple considèrent que  $M_{mig}$  correspond simplement à la quantité de mémoire utilisée par la VM en début de migration. Cependant, la quantité de mémoire  $M_{mig}$  transmise lors de la migration est difficilement prédictible car elle varie en fonction de la vitesse d'écriture des pages mémoire de la VM et de la bande passante.

#### II.3.3.2. Fréquence d'écriture des pages mémoires

Afin d'améliorer les performances des approches linéaires classiques, des études plus récentes tentent de modéliser la quantité de pages chaudes des VMs qui joue un rôle important dans la durée de migration mais aussi la durée d'interruption des VMs.

Dans une première étude, Liu et al. [2.8] statuent que la quantité de pages chaudes est approximativement proportionnelle au nombre total de pages modifiées lors de chaque itération de pré-copie. En s'appuyant sur les statistiques d'écriture des pages mémoire issues de plusieurs types d'applications, les auteurs utilisent une méthode d'apprentissage par régression linéaire pour généraliser une formule permettant de déduire la quantité de pages chaudes d'une VM en fonction du taux moyen de réécriture de ses pages mémoire. Ainsi pour estimer la durée de migration d'une VM, le taux d'écriture moyen de ses pages mémoire est d'abord calculé à partir de l'analyse de son activité au sein de l'hyperviseur Xen. La durée de migration est ensuite estimée à l'aide d'un algorithme itératif simulant l'implémentation de la pré-copie dans l'hyperviseur Xen ou autre.

Une seconde étude de Zheng et al. [2.9], reposant cette fois sur l'hyperviseur KVM, propose un modèle analytique permettant de prédire le temps de fin d'une migration. Leur solution consiste en une approche en temps réel qui permet d'estimer la proportion de pages chaudes en calculant un ratio d'accès pour chacune des pages mémoire de la VM. Le ratio d'accès d'une page repose sur sa fréquence de modification et représente la contribution de la page par rapport à l'ensemble des pages mémoire modifiées lors de la capture de l'état mémoire de la VM. La solution proposée repose sur une modification de l'hyperviseur KVM permettant d'analyser l'activité mémoire des VMs et d'agir directement sur le processus de migration. À l'aide d'un algorithme d'échantillonnage, leur solution consiste à calculer puis maintenir périodiquement à jour le ratio d'accès de chacune des pages mémoire afin d'améliorer la précision de l'estimation au fur et à mesure de la migration.

## II.5. Les différentes approches de migration de VMs

Généralement il existe deux approches de migration de VMs qui sont les suivantes :

- Approche basée sur la qualité de service
- Approche basée sur la consommation énergétique

### II.5.1. L'approche basée sur la qualité de service

La qualité de service (QoS) dénote le degré de performance, de fiabilité et de disponibilité offert par une application et par la plateforme ou l'infrastructure qui l'héberge. La qualité de service est fondamentale aux utilisateurs du cloud, qui attendent des fournisseurs de délivrer les caractéristiques annoncés et pour les fournisseurs cloud, qui ont besoin de trouver le bon compromis entre les degrés de la qualité de service et le coût opérationnel. Trouver le compromis optimal n'est pas une décision facile parce qu'elle implique la « Service Level Agreements (SLAs)» qui spécifie les cibles de la QoS et les pénalités économiques associées aux violations de la SLAs.

Les fournisseurs de service doivent remplir des contrats SLA qui déterminent les revenus et les pénalités sur la base du degré de réussite de la performance [2.10].

### II.5.2. L'approche basée sur la consommation énergétique

Cette approche est beaucoup plus utilisée pour réduire le coût énergétique très élevé dans les Datacenter. Les hôtes qui ont un très petit nombre de machines virtuelles pourront donc être éteints. Elle comprend trois phases habituellement.

### **II.5.2.1. La détection de la surcharge de l'hôte**

La technique de planification doit avancer un seuil limite dans le but de décider quand un certain hôte/serveur est surutilisé. Cette limite peut être appelée « Seuil chaud » ou « Hot Threshold » en anglais, et quand cette limite est dépassée, certaines des machines virtuelles de l'hôte doivent être migrées vers d'autres hôtes. [2.11]

### **II.5.2.2. La détection de la Charge réduite de l'hôte**

Si un certain serveur est sous-utilisé, c.-à-d. qu'il a atteint sous le « Seuil-chaud » ou en anglais « Cold Threshold », le scénario est juste contraire de celui de l'hôte surchargé, le but est d'identifier ce serveur et de migrer tous ses machines virtuelles vers d'autres hôtes actifs. Pour ce, l'hôte sous utilisé est libéré et peut être éteint pour économiser l'énergie. [2.11]

### **II.5.2.3. La sélection de la machine virtuelle et la migration**

Les candidats appropriées (les VMs) sont sélectionnées soit des hôtes sur ou sous-utilisés pour la migration. Les machines virtuelles sélectionnées dans les étapes précédentes sont alors placées sur une autre machine physique en fonction des critères de mappage adéquats. [2.11]

## **II.5.3. Quelques travaux et algorithmes sur la migration de VMs**

Le tableau suivant résume tous les travaux dans la migration de VMs par rapport aux différents aspects :

| Algorithme                                | Basé sur                           | Ressource considérée               | Nouvel aspect  | Puissance   | Faiblesse   | Performant que :  |
|---|------------------------------------|------------------------------------|--|---|---|---|
| Allocation exacte et Migration [t.1]      | Programmation par contrainte       | CPU                                | Fonctions objectives pour l'optimum                    | Réduction de nombre d'AMPs et le coût de la migration           | Exécution légèrement lent                               | Best-fit Heuristic  |
| Min-Cut hierarchical clustering [t.2]     | Programmation par contrainte       | CPU et BP                          | Optimisation de MLU et réutilisation de VM             | Réduction de la consommation énergétique et le réseau de trafic | Coût élevé de la migration                              | BFD et Random algorithm                                       |
| MFR (Measure-Forecast Ramap) [t.3]        | Programmation stochastique entière | CPU                                | Longueur de l'intervalle de temps « t »                | Réduction de nombre d'AMPs                                      | Besoin d'extension de nombreuses ressources             | Static algorithm  |
| Algorithme de paquetage de groupe [t.4]   | Emballage stochastique             | CPU et BP                          | Variable aléatoire pour prédire le future BP           | Réduction de nombre d'AMPs                                      | Besoin d'extension de nombreuses ressources             | First-fit, FFD and Harmonic algorithm                         |
| Algorithme de dimensionnement de VM [t.5] | Programmation stochastique entière | CPU et Mémoire                     | Probabilité « p » de débordement de serveur            | Réduction de nombre d'AMPs et O(l)-approximation                | Besoin d'extension de nombreuses ressources             | FFD algorithm   |
| VectorDot [t.6]                           | Emballage stochastique             | CPU, mémoire et réseau : E/S de BP | Métrique EVP, Serveur intégré et Stockage de migration | Equilibrage de charge dynamique, gestion de nœuds surchargés    | Besoin de suivre les modèles prédictifs et statistiques | Best-fit, First-fit, Worst-fit and Relaxed-Bestfit heuristics |

Table 1: Exemples d'algorithmes de migration

### ➤ Allocation exacte et migration

C'est une approche de Bin-Packing étendue grâce à l'inclusion de conditions valides exprimées sous forme de contraintes ou d'inégalités. L'objectif est de charger des éléments (VM dans notre cas) dans un ensemble de bacs (serveurs ou nœuds hébergeant le Machines virtuelles) caractérisées par leurs consommations d'énergie. Cette optimisation est soumise à un certain nombre de contraintes linéaires reflétant les limites de capacité des serveurs et des faits évidents tels qu'une machine virtuelle ne peut être affectée qu'à un serveur ou à un serveur ne peut héberger que des VM virtuelles en fonction de la quantité de ressources restantes.

### ➤ Min-Cut hierarchical clustering

Cette méthode construit les clusters en divisant récursivement les instances selon leur capacité et les classifie afin d'ordonner sur quel hôte une classes d'instance (machine virtuelle) peut être migré.

### ➤ **MFR**

C'est un algorithme de gestion de migration des VMs afin de minimiser le nombre de PMs requis pour supporter une charge de travail à un taux de violation SLA spécifié. Il utilise le modèle estimé pour faire des prédictions sur la future demande des ressources de manière standard, et MFR mappe l'ensemble des VMs aux PMs sans considérer le nombre de migration.

### ➤ **Algorithme de paquetage de groupe**

L'algorithme de regroupement est un algorithme qui regroupe des objets de forme ou de taille similaire, et il est développé sur la base d'une méthodologie comparative basée sur la forme totale pour trouver des formes et des tailles similaires.

### ➤ **Algorithme de dimensionnement de VM**

L'approche sert à estimer le montant des sources qui devraient être attribuées à un VM dans le but de faire correspondre les rendements de VMs à sa charge de travail. Le dimensionnement de VM doit être effectué de manière à éviter la violation de SLA résultant du sous-provisionnement des ressources.

### ➤ **VectorDot**

Un algorithme dynamique VectorDot est proposé pour Load-Balancing. Cet algorithme prend en charge la structure en couches du centre de données et gère également la multi-dimensionnalité des commutateurs réseau, la charge de ressources dans les serveurs et le stockage dans un centre de données afin d'équilibrer le charge au niveau de Datacenter en appliquant de migration des VMs.

## **II.6. Conclusion**

Dans ce chapitre, une des moyens d'économie de l'énergie des Datacenter a été étudiée, il s'agit de la migration des machines virtuelles du Datacenter surutilisé et sous-utilisé vers les autres Datacenter du cloud en question. Ceci est rendu possible par la mesure de la consommation énergétique et de l'allocation des machines virtuelles.

Dans le chapitre suivant on va voir un cas pratique de cette théorie par rapport à l'approche de migration qu'on a choisi : il s'agit de la méthode threshold.

---

# CHAPITRE 3 :

---

## DESCRIPTION ET CONCEPTION DE L'APPROCHE PROPOSEE

# Chapitre III : Description et conception de l'approche proposée

## III.1. Introduction

Pour réaliser notre travail sur la migration en temps réel de VMs au niveau de Cloud computing, nous nous basons sur la consommation d'énergie dans le Cloud, précisément dans le Datacenter. Une contrainte qui est beaucoup évoquée ces dix dernières années pour le « green computing » ou l'informatique verte qui combat pour la protection de l'environnement. Pour atteindre cet objectif, nous nous référons à la méthode de seuil (threshold) qui conduit à la migration de VMs et à la baisse de charge sur chaque PMs qui accueille ces dernières.

Le présent chapitre présentera l'approche « threshold » d'une manière détaillée. Puis il décrit les outils de travail éventuellement la plateforme CloudSim. Ensuite il décrit la modélisation de la migration ainsi que son architecture générale.

## III.2. L'approche « threshold »

D'après les auteurs de [3.1], la méthode de threshold est distinguée des deux phases ; la phase de migration et la phase de placement. Elle est basée sur un seuil dynamique ou statique sous l'influence de la consommation d'énergie et ce seuil est obtenu en utilisant la formule mathématique.

Notre modèle est défini comme étant un Datacenter composé d'un ensemble de PMs. Chaque hôte physique peut contenir  $m$  VMs hétérogènes. Ces PMs et VMs possèdent plusieurs ressources physiques telles que : l'utilisation de processeur, la capacité de la RAM, le MIPS, les seuils supérieur d'utilisation de ressources, etc.

Le but de notre démarche est de proposer une stratégie basée sur la migration des VMs. Cette proposition nous permet de sélectionner un nombre de VMs à migrer afin de réduire l'énergie consommée par les Datacenter. L'avantage général de cette stratégie est d'augmenter sensiblement le taux d'utilisation des ressources physiques et de réduire la demande d'énergie et les émissions de CO<sub>2</sub>.

### III.2.1. Calcul de la consommation d'énergie

Comme la méthode threshold est généralement basée sur la consommation énergétique nous devons d'abord calculer la consommation de l'énergie.

Il existe plusieurs méthodes pour calculer la consommation d'énergie, parmi lesquels nous avons choisi la méthode « puissance par rapport à l'utilisation ». De nombreuses études [3.2] ont montré que la consommation d'énergie par les hôtes/serveurs peut être décrite par une relation linéaire entre la consommation d'énergie et l'utilisation du processeur. Ces études confirment qu'une puissance moyenne consommée par un serveur inactif est de 70% de par rapport à un serveur pleinement utilisé. D'après [3.3], si l'utilisation du processeur est supérieure à 30%, la valeur inférieure est toujours 0.3. Donc, ils ont défini la consommation d'énergie  $P(u)$  par la formule suivante :

$$P(u) = Pmax * (0.7 + 0.3u) \quad (3. f1)$$

Avec ; **Pmax** = 250 W pour des serveurs modernes.

**0.7** : est la constante de puissance moyenne consommée par un serveur inactif.

**u** : est l'utilisation du processeur.

Comme l'utilisation de CPU peut changer avec le temps en raison de la variabilité de la charge de travail, il s'agit d'une fonction du temps  $u(t)$ . Par conséquent, pour définir la consommation d'énergie totale par un serveur, nous utilisons l'équation suivante :

$$Es = \int_t P(u(t))dt \quad (3. f2)$$

Donc la consommation d'énergie par rapport à un Datacenter est calculée par la formule qui suit [3.4] :

$$Es = \frac{\sum_{j=1}^m P(u)j}{m} \quad (3. f3)$$

Avec ; **m** : nombre de machines physiques dans un Datacenter.

### III.2.2. La phase de sélection

Le principal objectif de cette phase est de déterminer quand est ce qu'on doit migrer les VMs afin de réduire la consommation énergétique. Cette phase est basée principalement sur deux seuils d'utilisations des ressources physiques. Chaque PM possède des seuils inférieur et supérieur d'utilisation de CPU. Dans ce cas :

- Si l'utilisation du processeur par la PM est plus élevée que le seuil supérieur, alors nous devons transférer quelques VMs pour obtenir un équilibre et réduire la consommation d'énergie et le risque des violations SLAs.
- Si l'utilisation du processeur par la PM est inférieur au seuil inférieur, alors nous devons migrer toutes les VMs et mettre l'hôte à l'état « OFF ».

Les seuils supérieur et inférieur de l'utilisation du CPU sont fixes durant toute la simulation. Nous présentons dans la section qui suit comment choisir et calculer les différents seuils dans les deux approches.

#### II.2.2.1. Le seuil supérieur

Le seuil supérieur  $T_{U_i}$  est calculé par la formule (3.f6) [3.5].

$$U_i = \sum_{j=1}^m u_j \quad (3. f4)$$

$$S_{U_i} = \sqrt{\sum_{j=0}^m u_j^2} \quad (3. f5)$$

$$T_{U_j} = 1 - \left( (P_{uu} * S_{U_i}) + U_i \right) - \left( (P_{ul} * S_{U_i}) + U_i \right) \quad (3.f6)$$

Avec;  $u_j$  : L'utilisation du CPU de la VM  $j$ .

$m$  : Le nombre de VMs.

$U_i$  : L'utilisation du CPU de la PM  $i$ .

$P_{uu}$  : égale à 95% [3.5].

$P_{ul}$  : égale à 90% [3.5].

### II.2.2.2. Le seuil inférieur

Le seuil inférieur  $T_1$  est calculé selon la formule (3.f9) [3.5].

$$U_i = \frac{1}{m} \sum_{j=1}^m u_j \quad (3.f7)$$

$$S_{U_i} = \sqrt{\left( \sum_{j=0}^m u_j - U_i \right)^2_j} \quad (3.f8)$$

$$T_1 = \begin{cases} U_i - (P_l * S_{U_i}) & , \text{ Si l'utilisation du CPU} < 0.3 \\ 0.3 & , \text{ Si l'utilisation du CPU} > 0.3 \end{cases} \quad (3.f9)$$

Avec,  $P_l$  : égale à 90% [3.5].

Les autres notations sont les mêmes que celles qui sont citées ci-dessus.

### II.2.2.3. L'algorithme et le diagramme d'activité

#### ➤ L'algorithme

La première partie se situe entre les lignes 1 et 13. Cette section est exécutée lorsque l'utilisation du CPU de l'hôte est supérieure au seuil supérieur. Elle permet de choisir les meilleures VMs pour faire une migration afin de réduire l'utilisation de la PM en dessous de ce seuil. L'algorithme choisit les VMs possédant une grande utilisation de CPU en comparaison avec les autres VMs de la même PM.

La seconde partie (lignes 15-18) permet de migrer toutes les VMs de la PM dans le cas où l'utilisation de CPU de la PM est en dessous du seuil inférieur. La PM doit être en état hors tension afin de ne pas gaspiller de l'énergie.

### Algorithme 1 : Algorithme de sélection des machines virtuelles.

Entrées : hostList, VmList

Sorties : migrationList

```
1 pour chaque h dans hostList faire
2   **Sélectionner la VM avec la plus grande utilisation de CPU.
3   hostUtilCPU ← h.utilCPU();
4   tant que hostUtilCPU > h.upThreshCPU() faire
5     bestVmUtilCPU ← h.BestVmUtilCPU();
6     bestVm ← Null;
7     pour chaque vm dans VmList faire
8       si vm.utilCPU=bestVmUtilCPU alors
9         bestVm ← vm;
10        break;
11    hostUtilCPU ← hostUtilCPU - bestVm.utilCPU();
12    migrationList.add(bestVm);
13    vmList.remove(vm);
14  **Migration de toutes les VMs et mettre en état Off la PM.
15  si hostUtilCPU < h.lowThreshCPU() alors
16    migrationList.add(h.getVmList());
17    vmList.remove(h.getVmList());
18    h.OffState();
19 retourner migrationList;
```

Figure 14: Algorithme de la phase de sélection

#### ➤ Le diagramme d'activité

Avec l'aide de l'UML diagramme nous présentons ci-dessous le diagramme d'activité de l'algorithme de la sélection dans l'approche threshold.

Par définition, l'UML Diagramme et appelé aussi langage de modélisation unifié, est un langage de modélisation graphique à base de pictogrammes conçu pour fournir une méthode normalisée pour visualiser la conception d'un système. Il est couramment utilisé en développement logiciel et en conception orientée objet [3.9].

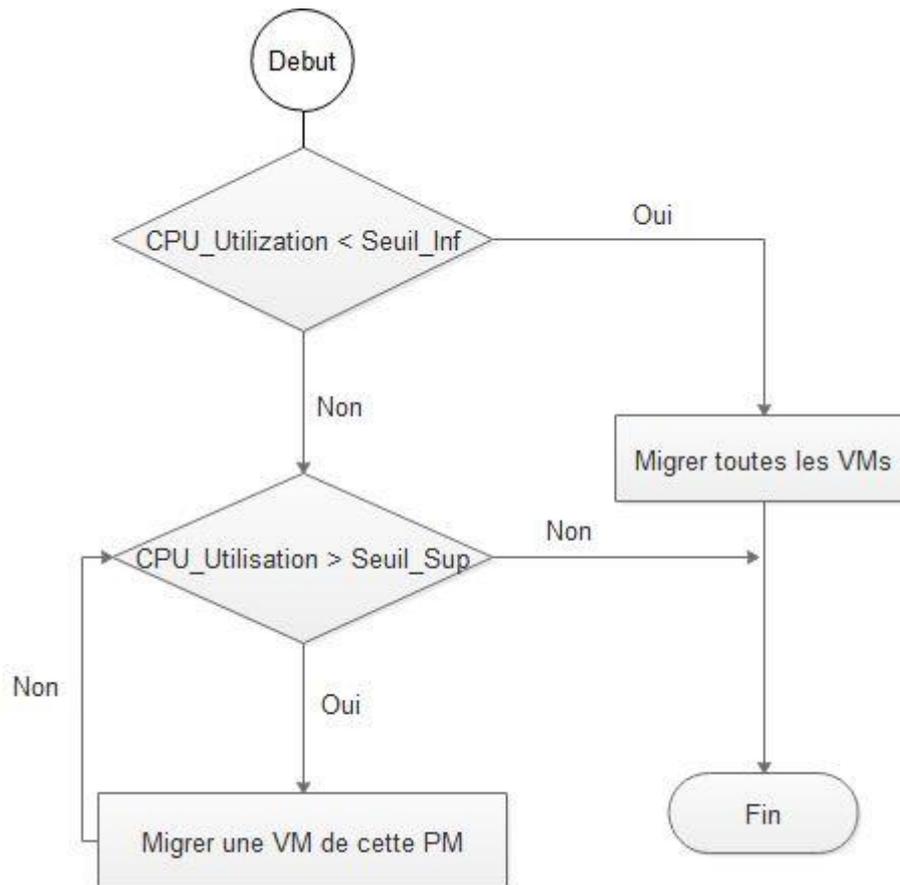


Figure 15: Diagramme d'activité de la phase sélection

## II.2.3. La phase d'allocation

Cette phase a pour rôle de placer les VMs dans les hôtes moins utilisés en se basant sur l'algorithme de MBFD (Modified Best Fit Decreasing) [3.6] [3.8].

### II.2.3.1. Le principe de l'algorithme Best Fit Decreasing (BFD)

Le problème d'optimisation Bin Packing peut être appliqué à un grand nombre de secteurs industriels et informatiques comme par exemple le remplissage des camions avec des produits sous la contrainte de poids, le rangement des fichiers dans un support informatique en prenant en considération la capacité de stockage.

Par exemple, on considère un ensemble des sacs identiques ' $m$ ' et un ensemble d'objets ' $n$ '. Pour chaque objet ' $i$ ', nous avons un poids ' $pi$ '. Sachant que les sacs peuvent supporter qu'un poids maximum ' $pmax$ ', combien faudra-t-il au minimum de sac pour ranger l'ensemble des objets considérés.

Afin de résoudre le problème de Bin Packing, des méthodes heuristique ont été développées tels que la Next Fit Heuristic, la First Fit Decreasing, la Best Fit Decreasing, etc. [3.7]. Ces algorithmes ne sont pas optimaux mais ils permettent d'obtenir de très bons résultats en pratique. Ils ont pour but de trouver une solution avec un bon compromis entre la qualité de la solution et le temps de calcul.

### II.2.3.2. L'algorithme et le diagramme d'activité

L'algorithme de l'allocation de VMs est globalement composé des deux boucles imbriquées, la première boucle (ligne 2) permet de parcourir la liste de VMs à faire migrer. La seconde boucle (ligne 5) permet de parcourir la liste de toutes les PMs pour placer les VMs. Pour chaque VM dans la liste des VMs à faire migrer, l'algorithme vérifie si l'hôte a suffisamment de ressources pour cette VM. Il vérifie également si l'utilisation de CPU de la PM est inférieure au seuil supérieur de l'utilisation de CPU. Si c'est le cas, l'énergie consommée par l'hôte après allocation de la VM est estimée et la VM est allouée dans la PM offrant une consommation minimale.

#### ➤ L'algorithme

---

Entrées : *hostList*, *MigratedVmList*, *UpperCPU\_UtilizationThreshold*  
Sorties : *allocatedHost*

```
1 MigratedVmList.sortDecreasingUtilization();
2 pour chaque vm dans MigratedVmList faire
3   minPower ← MAX;
4   allocatedHost ← NULL;
5   pour chaque host dans hostList faire
6     si (host a suffisamment de ressources pour vm) alors
7       si (Host_Utilization_CPU after allocation <
8         UpperCPU_UtilizationThreshold) alors
9         power ← estimatePower(host, vm);
10        si power < minPower alors
11          allocatedHost ← host;
12          minPower ← power;
13   si allocatedHost ≠ NULL alors
14     Allouer vm à allocatedHost;
15 retourner allocatedHost (Allouer une VM dans une PM);
```

---

Figure 16: Algorithme de l'allocation de VM

#### ➤ Le diagramme d'activité

La figure suivante présente le diagramme d'activité qui résume le fonctionnement de la phase d'allocation de notre approche.

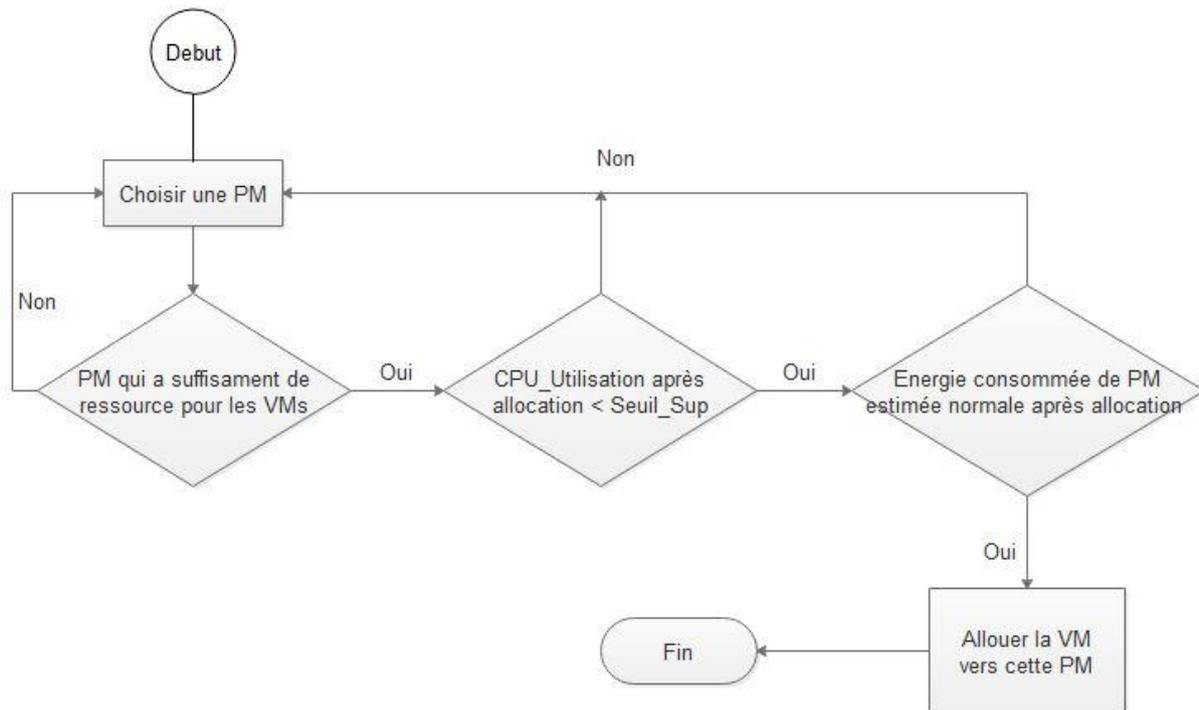


Figure 17: Diagramme d'activité de la phase d'allocation

### III.3. Architecture

Globalement l'architecture de notre approche est constituée par les différents éléments de base suivants.

- **Datacenter** : c'est une entité informatique composée d'un ensemble de machines physiques. Il est caractérisé par le nombre et la vitesse des processeurs, la capacité de la mémoire et de stockage, la bande passante ainsi que le coût de chaque utilisation de ressources.
- **DatacenterBroker** : il assure la gestion des VMs dans plusieurs Datacenters et le routage du trafic vers les Datacenters appropriés. Il choisit également le Datacenter qui fournit le meilleur service pour les demandes envoyées par chaque utilisateur. Il est responsable de la communication entre les utilisateurs et les Datacenters.
- **Hôte (Physical Host)** : une machine physique est caractérisée par le nombre des VMS, la capacité de stockage, le nombre des requêtes arrivées, la vitesse du CPU et le nombre des cores qu'elle a.
- **VMs** : la machine virtuelle consiste à créer plus environnement d'exécutions sur une seule machine physique. Elle fournit à chaque utilisateur un service selon la demande.
- **VM Manager (VMM)** : il assure le suivi de la disponibilité des Vms et leur utilisation des ressources. Il est chargé du provisionnement des nouvelles VMs ainsi que la réaffectation des VMs d'une hôte physique à l'autre afin d'adapter le placement.

- **Local managers** : les gestionnaires locaux se trouvent sur chaque nœud comme un module du VMM. Leur objectif est le suivi continu d'utilisation du processeur d'un nœud physique, le redimensionnement de la VM en fonction de leurs besoins en ressources et la prise de décision quant à la migration des VMs.
- **Global manager** : Il réside sur un nœud maître et recueille des informations auprès des gestionnaires locaux pour maintenir la vue globale d'utilisation des ressources.

e

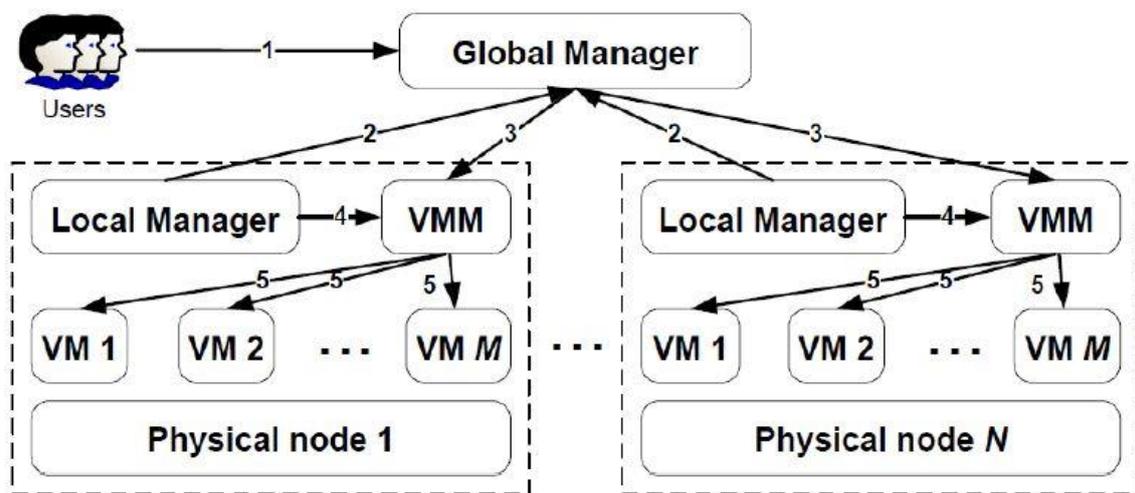


Figure 18: Architecture générale du système

Les gestionnaires locaux se trouvent sur chaque nœud en tant que module du moniteur de la VM (VMM). Leur objectif est le suivi continu d'utilisation du processeur d'un nœud, le redimensionnement de la VM en fonction de leurs besoins en ressources et la prise de décision (quand et où et les VMs doivent être migrées à partir du nœud d'accueil ?) (4). Ils permettent également de recevoir des requêtes à partir des clients (1). Le gestionnaire global réside sur un nœud maître et recueille les informations provenant des responsables locaux pour le maintenir la vue d'ensemble de l'utilisation des ressources (2). Il émet aussi des commandes pour l'optimisation du placement des VMs (3). Le VMM gère les redimensionnements réels et la migration des VMs ainsi que des changements dans les états d'alimentation des nœuds (5).

### **III.4. Conclusion**

A travers ce chapitre nous avons trouvé la description de l'approche que nous avons proposée avec les deux phases que comporte cette technique ainsi que le mode de calcul de seuil. Dans la première phase qui est la sélection qui consiste en général à indiquer quand est-ce que les VMs doivent être migrées. Quant à la deuxième phase, l'allocation, elle permet de choisir dans quelle hôte future nous pourrions migrer les VMs. Cette migration est globalement déclenchée par l'augmentation de consommation d'énergie, notamment l'utilisation de CPU, au niveau d'un hôte physique. Nous avons vu également l'architecture globale de l'approche threshold à la fin de ce chapitre avec ses différents éléments de base.

Dans le chapitre suivant, nous présenterons l'implémentation de notre approche dans le CloudSim ainsi que quelques résultats de notre travail.

---

# CHAPITRE 4 :

---

## IMPLEMENTATION ET RESULTAT

# Chapitre IV. Implémentation et résultat

## IV.1. Introduction

Pour bien finir notre travail, nous avons fait une simulation de notre projet sous CloudSim. Les résultats de cette simulation seront bien détaillés à la fin de ce chapitre.

En ce qui concerne le chapitre actuel, il est divisé en quatre sections : il y a tout d'abord l'introduction, suivi des outils de travail. Cette section comprend quant à elle trois parties à savoir le langage de programmation utilisé, l'environnement de développement et à la fin le simulateur CloudSim. Pour ce qui est de la troisième section, il s'agit du test, c'est-à-dire des résultats de la simulation sous CloudSim. La section qui termine le chapitre est la conclusion.

## IV.2. Outils de travail

- Une machine avec un processeur Intel® Core™ i7 -2670QM CPU @ 2.2 GHz.
- Un IDE Eclipse version Luna.
- Un simulateur CloudSim version 3.0.3 développé à base de Java.
- Le langage de programmation Java.

### IV.2.1. Le langage de programmation Java

Java est à la fois un langage de programmation informatique orienté objet et un environnement d'exécution portable. Il est créé par James Gosling et Patrick Naughton employés de Sun Microsystems avec le soutien de Bill Joy (co-fondateur de Sun Microsystems en 1982), présenté officiellement le 23 Mai 1995 au SunWorld.

La particularité et l'objectif central de Java est que les logiciels écrits dans ce langage doivent être très facilement portables sur plusieurs systèmes d'exploitation tels que UNIX, Windows, Mac OS ou GNU/Linux, avec peu ou pas de modifications. Pour cela, divers plateformes et frameworks associés visent à guider, sinon garantir, cette portabilité des applications développées en Java [4.1].

Les applications Java peuvent être exécutées sur tous les systèmes d'exploitations pour lesquels a été développée une plateforme Java dont le nom technique est JRE (Java Runtime Environment – Environnement d'exécution Java). Cette dernière est constituée de JVM (Java Virtual Machine), le programme qui interprète le code java et le traduit en code natif. Mais le JRE est surtout constitué d'une bibliothèque standard à partir de laquelle doivent être développées tous les programmes en Java. C'est la garantie de portabilité qui a fit la réussite de Java dans l'architecture client-serveur en facilitant la migration entre serveurs, qui est très difficile pour les gros systèmes [4.2].

Java est devenue actuellement une direction incontournable dans le monde de la programmation. Parmi les différentes caractéristiques qui sont attribuées à son succès, nous avons [4.1] :

- L'indépendance de la plateforme : c'est-à-dire le code reste indépendant de la machine sur laquelle il s'exécute. Il est possible d'exécuté des programmes Java sur tous les environnements qui possèdent une JVM.
- La portabilité : il permet la simulation d'être distribuée facilement sans avoir à recompiler le code pour les différents systèmes.

- Le code est structuré dans plusieurs classes dont chacune traite une partie différente de la simulation.
- Java assure la gestion dynamique de la mémoire
- Il est multitâche : qui veut dire qu'il permet l'utilisation de Threads<sup>7</sup> qui sont des unités d'exécution isolées.

## IV.2.2. L'environnement de développement Eclipse

Eclipse est une communauté open-source dont les projets visent à fournir une plateforme de développement ouverte, comprenant des espaces de travail modulaires, des outils et environnements d'exécution, pour construire, déployer et gérer des applications sur tout leur cycle de vie. [4.3].

Eclipse est surtout reconnu pour son IDE Java mais son champ d'action est devenu beaucoup plus large. Les projets Eclipse offrent des outils et environnements qui couvrent tout le cycle de développement : outils de modélisation, développement, déploiement, reporting, manipulation de données, tests, etc. Eclipse IDE est principalement écrit en Java (à l'aide de la bibliothèque graphique SWT, d'IBM), et ce langage, grâce à des bibliothèques spécifiques, est également utilisé pour écrire des extensions.

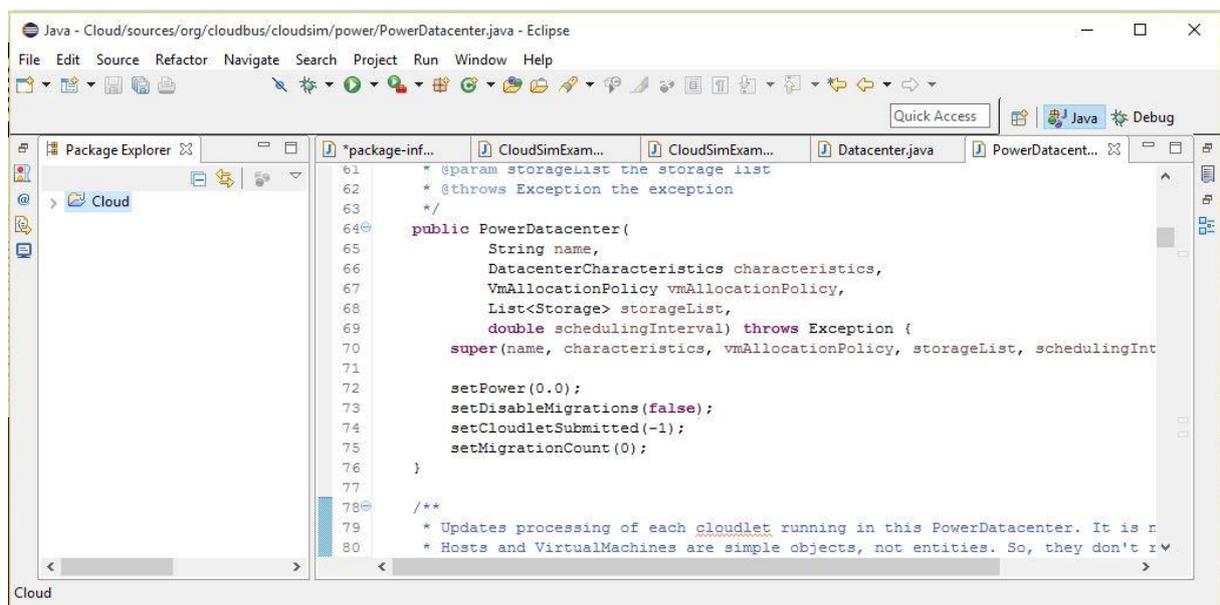


Figure 19: Environnement de développement Eclipse

Il est téléchargeable directement de son site officiel <https://www.eclipse.org/downloads/> et disponible sous Windows, Linux et MacOS. Il est puissant et compatible avec toutes les nouvelles technologies Java (Java EE, Bases de données UML, XML, etc.)

## IV.2.3. Le simulateur CloudSim

### IV.2.3.1. Définition

CloudSim est une boîte à outils (bibliothèque) pour la simulation de scénarios informatiques en nuage. Il fournit des classes de base pour décrire les centres de données, les machines virtuelles,

<sup>7</sup> Un thread ou fil (d'exécution) ou tâche est similaire à un processus car tous deux représentent l'exécution d'un ensemble d'instructions du langage machine d'un processeur.

les applications, les utilisateurs, les ressources informatiques et les politiques pour la gestion de diverses parties du système [4.4].

Ces composants peuvent être mis en place pour que les utilisateurs évaluent de nouvelles stratégies dans l'utilisation de Clouds (politiques, algorithmes de planification, de cartographie et d'équilibrage de charge, etc.). Il peut également être utilisé pour évaluer l'efficacité des stratégies à partir de différentes perspectives, du coût / profit pour accélérer le temps d'exécution de l'application. Il soutient également l'évaluation des politiques d'informatique verte.

CloudSim est un nouveau général, et extensible cadre de simulation qui permet la modélisation, la simulation et l'expérimentation de nouvelles infrastructures du cloud computing et de services d'applications.

Dans le cas Cloud Computing, l'outil de Simulations comme CloudSim offre d'importants avantages aux clients et fournisseurs. Pour les clients, il permet de tester leurs services dans un environnement contrôlable avec exempt du coût et de vérifier la performance avant d'utiliser les vrais nuages. Tandis que pour les fournisseurs, CloudSim leur permet de vérifier les types d'allocation en fonction de divers prix et de la charge ;

En outre, cela permettra d'optimiser le coût de l'accès aux ressources et l'amélioration des bénéfices. Sans ces outils, les clients et des fournisseurs devons s'appuyer sur des évaluations imprécises ou sur des approches essai-erreur, ces approches peuvent conduire à l'inefficacité performance des services et de réduire la génération de revenus.

CloudSim aide donc les chercheurs et développeurs à tester la performance d'un service d'application développé dans un convenable environnement qui de plus est facile à installer.

Il y a de nombreux avantages de l'utilisation CloudSim :

- L'efficacité de temps : il prend moins de temps et d'efforts pour mettre en œuvre des applications de cloud computing.
- La flexibilité : les développeurs peuvent facilement modéliser et tester les performances de leurs applications et ses services dans des environnements hétérogènes.

#### **IV.2.3.2. L'architecture du CloudSim**

La structure logicielle de CloudSim et ses composants est représentée par une architecture en couche comme montré par la figure suivante.

Au niveau le plus bas se trouve le moteur de simulation des événements discrets SimJava, qui implémente les fonctionnalités de base requises, tels que les files d'attente, les traitements des événements, la création des entités du système Cloud (services, hôtes, Datacenter, Broker, VM...), la communication entre les composants et la gestion d'horloge de simulation.

CloudSim supporte la modélisation et la simulation de l'environnement de Datacenter basé sur le Cloud, tel que des interfaces de gestion dédiées aux VMs, la mémoire, le stockage et la bande passante. Elle gère l'instanciation et l'exécution des entités de base (VM, hôtes, Datacenters, applications) au cours de la période de simulation.

Dans la couche plus haute de la pile de simulation, on trouve le code de l'utilisateur qui expose la configuration des fonctionnalités liées aux hôtes (nombre de machines, leurs spécifications), les politiques d'ordonnancement de Broker, applications (nombre de tâches et leurs besoins), VM, nombre d'utilisateurs. Elle contient les informations de base nécessaires au bon

fonctionnement des hôtes et des applications comme le nombre de machines, leurs spécifications, le choix des stratégies d’ordonnancement du broker, etc.

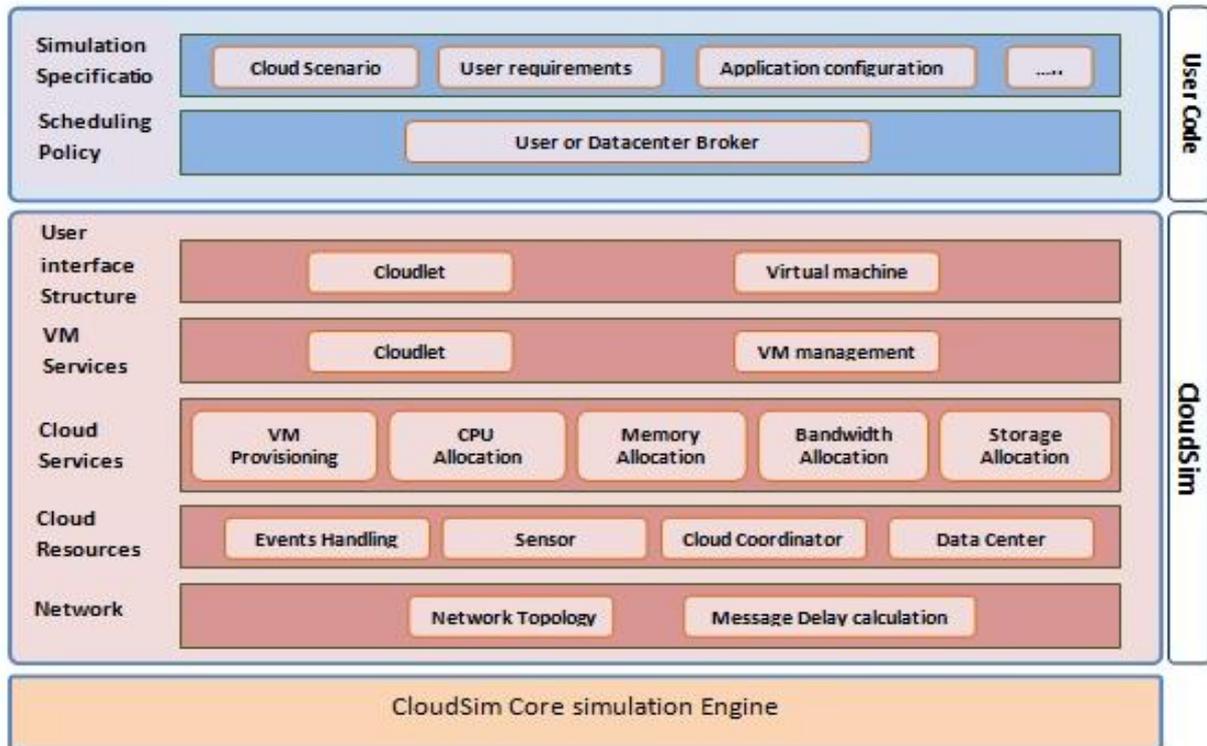


Figure 20: Architecture du CloudSim

### IV.2.3.3. Les classes principaux du CloudSim

CloudSim est composé de plusieurs classes dont nous pouvons classifier en deux catégories : les classes qui modélisent les entités comme le Datacenter, le Broker, etc. et les classes modélisant les politiques d’allocation. Dans la figure 19 nous présentons globalement le diagramme de classe du simulateur CloudSim, pour laquelle nous avons exposé les classes principales.

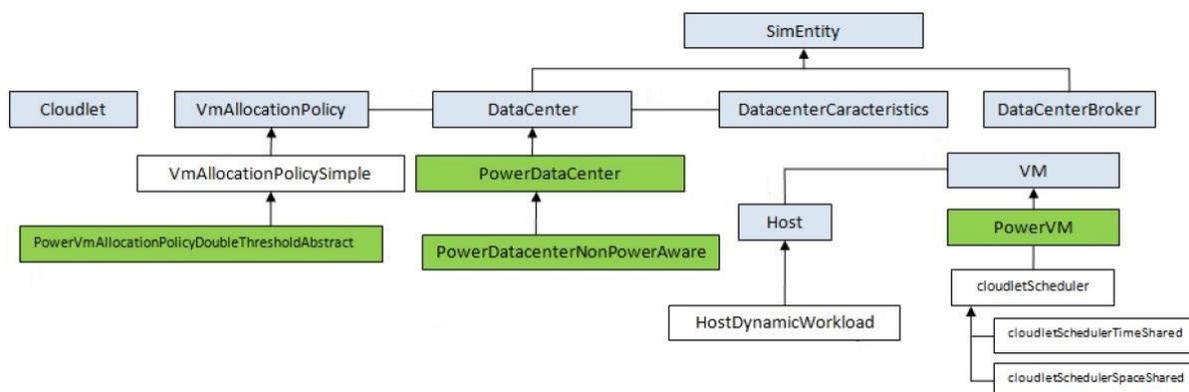


Figure 21: Diagramme de classe du simulateur CloudSim.

#### ➤ **SimEntity**

Il s'agit d'une classe abstraite, elle représente l'entité de simulation qui est capable d'envoyer des messages à d'autres entités et de gérer les messages reçus ainsi les événements.

Toutes les entités doivent étendre cette classe et redéfinir ses trois méthodes : **StartEntity()**, **processeEvent()** et **shutdownEntity()**. Ces méthodes définissent les actions pour l'initialisation, le traitement des événements et la destruction de l'entité.

#### ➤ **Cloudlet**

C'est une classe qui représente une tâche. Elle modélise les services d'application du Cloud (comme la livraison, réseaux sociaux et les sites d'affaires). Cette classe peut aussi être étendue pour supporter la modélisation des performances et d'autres paramètres de composition pour les applications telles que les transactions dans les applications orientées bases de données (Oracle, SQL).

#### ➤ **Datacenter :**

Cette classe modélise l'infrastructure du noyau (matériel, logiciel) offert par des fournisseurs de service dans un environnement de Cloud Computing. Il encapsule un ensemble de machines de calcul qui peuvent être homogènes ou hétérogènes en ce qui concerne leur configurations de ressources (mémoire , noyau ,capacité et stockage ) .En outre ,chaque composant du Datacenter instancie un composant généralisé d'approvisionnement en ressources qui implémente un ensemble de politiques d'allocation de bande passante, de mémoire et des dispositif de stockage.

La classe PowerDataCenter utilisée dans le package « power » hérite de la classe Datacenter et elle permet la simulation des centre de données en calculant leurs énergies consommées.

#### ➤ **DatacenterBroker**

Cette classe modéliser le courtier, qui est responsable de la médiation entre les utilisateurs et les prestataires de service selon les conditions de QoS des utilisateurs et il déploie les tâche de service à travers les Clouds. Le Broker agissant au nom des utilisateurs identifie les prestataires de service appropriés du cloud par le service d'information du cloud CIS (cloud information services) en négocie avec eux pour une allocation des ressources qui répond aux besoins de QoS des utilisateurs.

#### ➤ **DatacenterCharacteristic**

C'est la classe qui contient les informations sur la configuration des ressources des Datacenter. Parmi eux le système d'exploitation, le VMM, la liste des hôtes, le coût par seconde d'utilisation des ressources, etc.

#### ➤ **Host**

Cette classe représente un serveur informatique physique dans un Cloud. Le Host exécute des actions liées à la gestion des machines virtuelles et a une politique définie pour l'approvisionnement mémoire et bande passante, ainsi que d'une politique de répartition des PE (Processeur Element) à des machines virtuelles. Un hôte est associé à un Datacenter. Il peut héberger un ou plusieurs VMs.

#### ➤ **VM**

Cette classe modéliser une instance de machine virtuelle, qui est géré et hébergé pendant son cycle de vie par le composant Cloud Host. Chaque composant de VM a accès à un composant qui stocke les caractéristiques liées à elle telles que : l'accès mémoire, le processeur, la capacité de stockage et les politiques de provisionnement interne de la machine virtuelle.

Dans le but de réduire l'énergie consommée par les Datacenters, la classe VMPower héritant le classe VM enregistre l'historique de l'utilisation de CPU. Cet historique est utilisé dans les politiques de sélection et d'allocations des VMs.

#### ➤ **VMAllocationPolicy**

C'est une classe abstraite qui représente la politique de provisionnement des hôtes aux machines virtuelles dans un Datacenter. Pour notre politique nous avons utilisé la classe PowerVmAllocationPolicyDoubleThresholdAbstract qui hérite la classe VmAllocationPolicy.

### **IV.2.3.4. L'installation**

Pour l'installation du CloudSim au niveau de l'IDE Eclipse, il faut télécharger ces fichiers suivants au préalable :

- CloudSim 3.0.3 toolkit : <https://github.com/Cloudslab/cloudsim/releases/tag/cloudsim-3.0.3>
- CloudSim Commons-Maths : <http://redrockdigimark.com/apachemirror//commons/math/binaries/commons-math3-3.6.1-bin.zip>

**Étape 1 :** Ouvrez Eclipse Java IDE et ouvrez Fichier> Nouveau> Projet Java. Il ouvrira une boîte de dialogue intitulée «Nouveau projet Java».

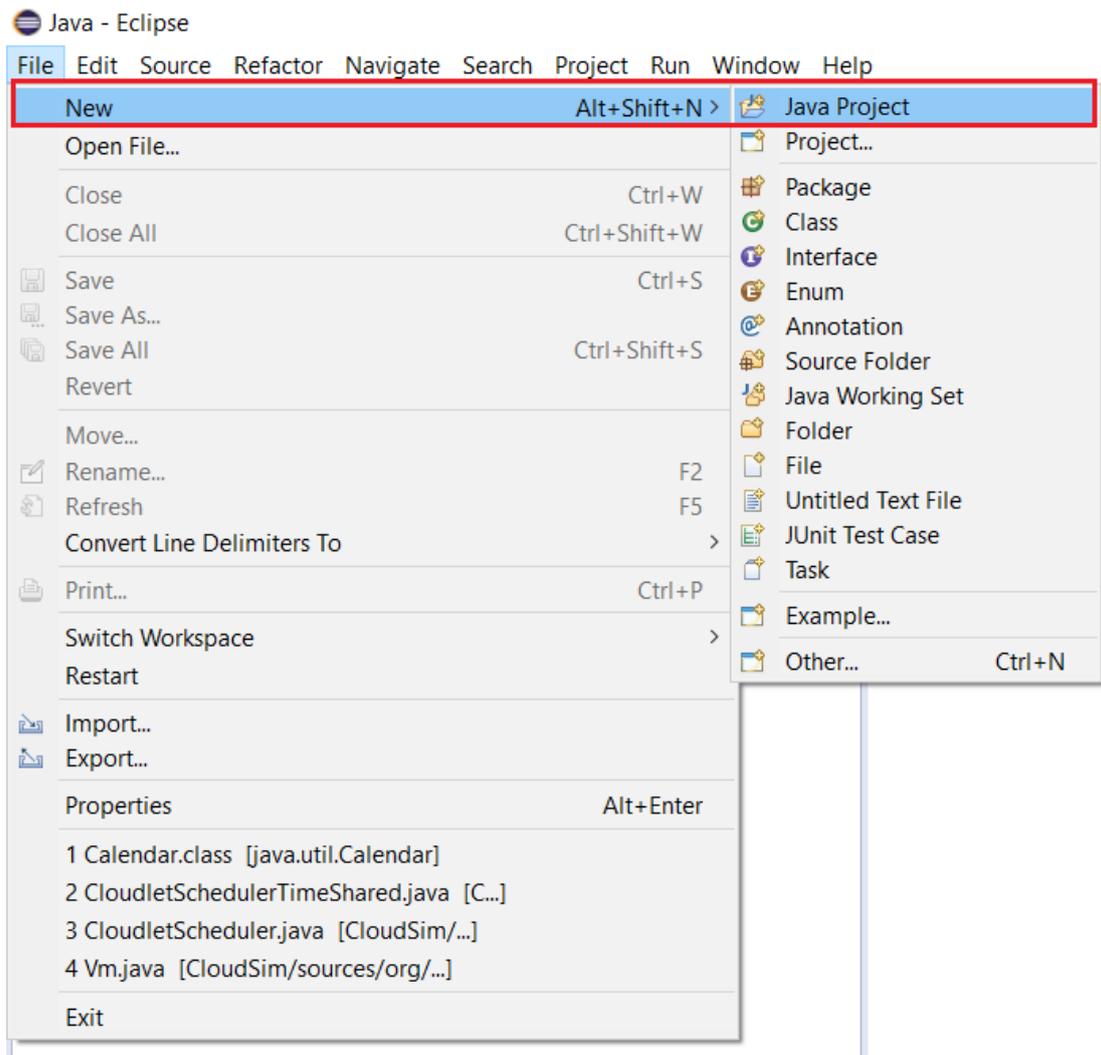


Figure 22: Ajout de CloudSim dans Eclipse

**Étape 2 :** Dans la boîte de dialogue "Nouveau projet de Java", entrez 'Nom du projet'> désélectionnez 'Utiliser l'emplacement par défaut' et parcourez le chemin d'accès auquel vous avez extrait le dossier 'Cloudsim-3.0.3' puis cliquez sur NEXT.

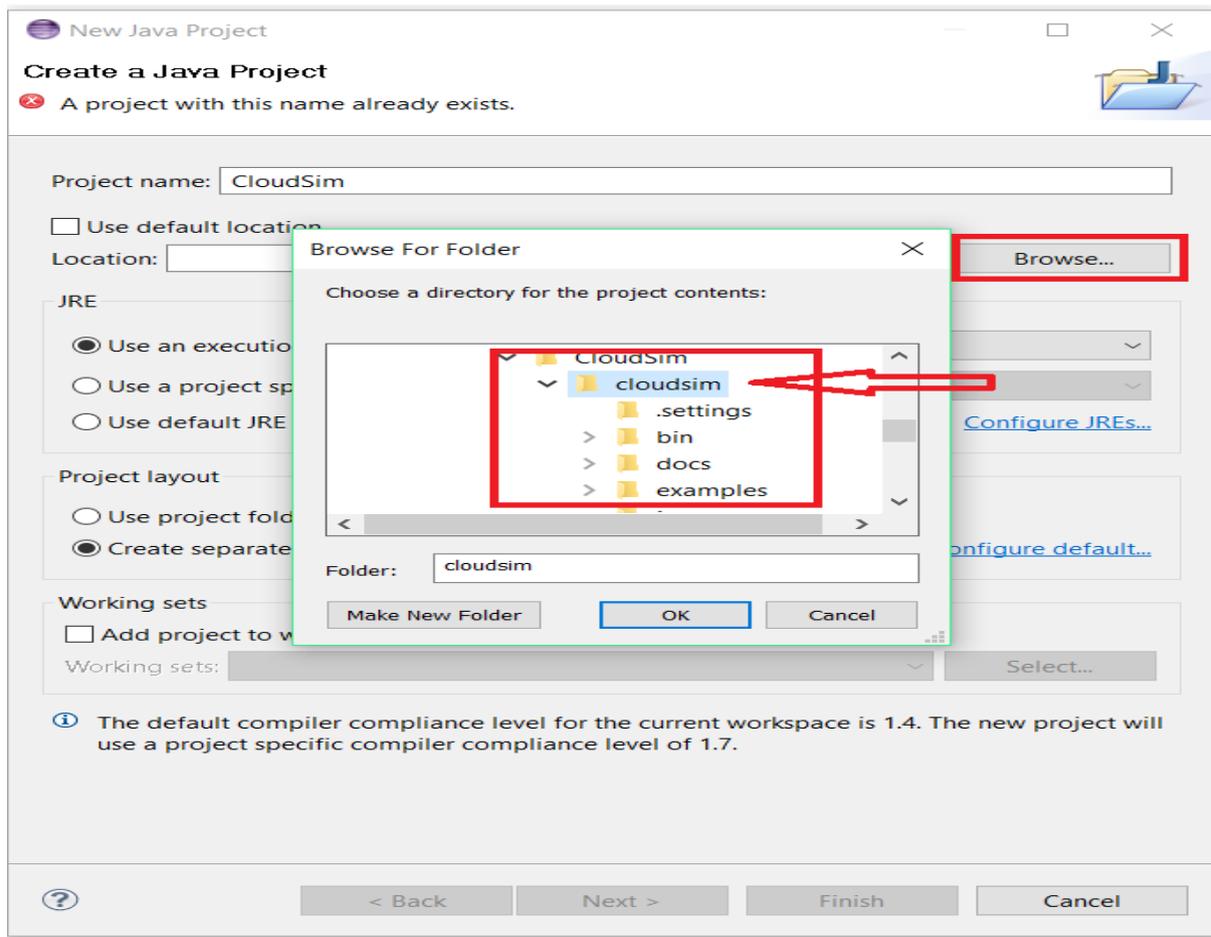


Figure 23: Création de projet java pour CloudSim

**Étape 3 :** Une fois que vous cliquez sur la boîte de dialogue suivante, allez dans l'onglet 'Bibliothèques'. Ici, pour ajouter «commons-math3-3.6.1.jar», nous devons cliquer sur «Ajouter un jet externe», puis parcourir le chemin sur lequel vous avez téléchargé et décompressé le fichier «commun-math3-3.6.1.jar». Ajoutez-le à la liste en cliquant sur Ouvrir.

Enfin achever l'installation on clique sur le bouton « Finish » et CloudSim pourra fonctionner dans l'environnement Eclipse.

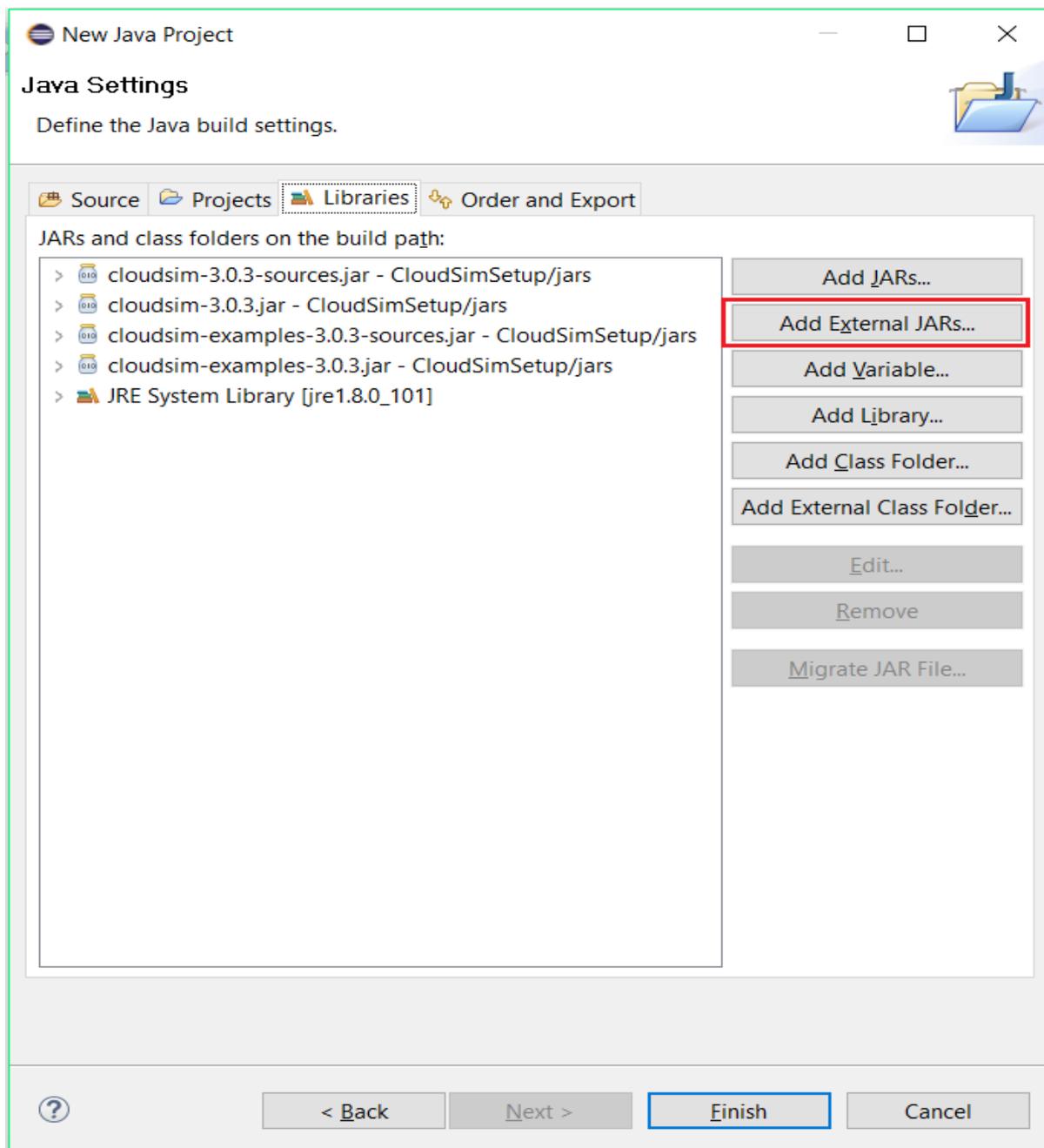


Figure 24: Ajout de fichiers JAR

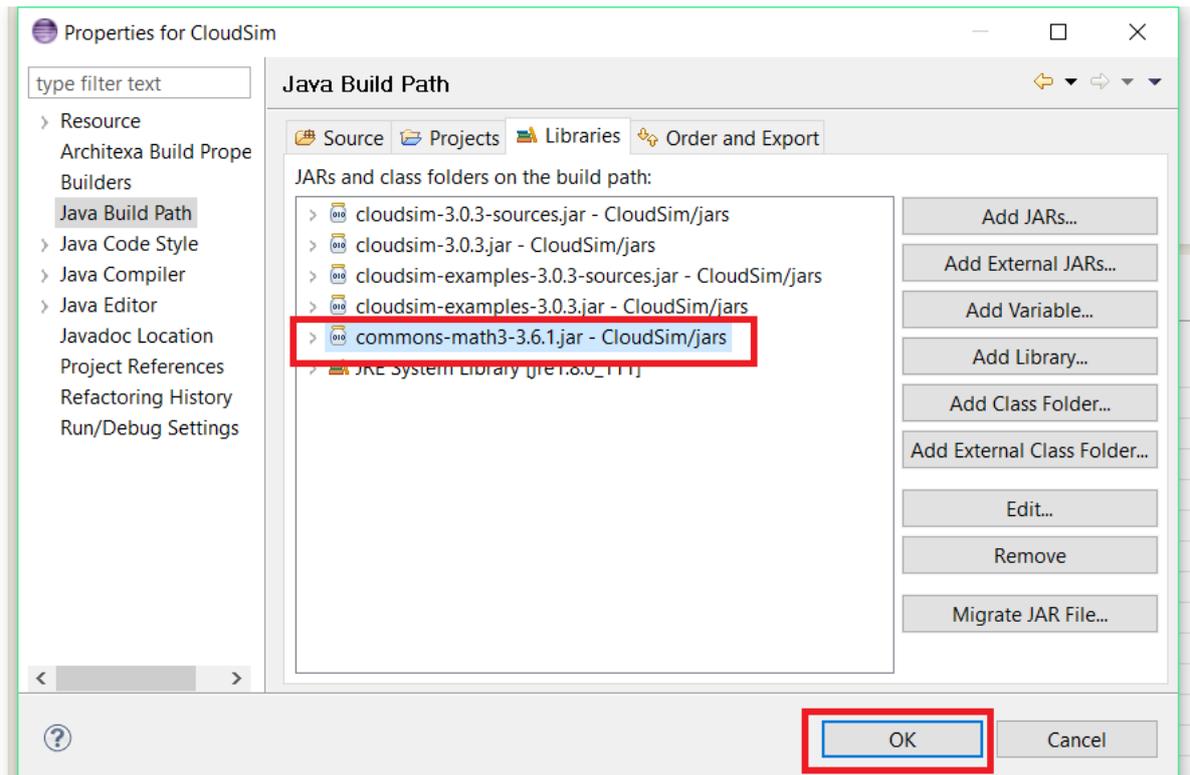


Figure 25: Ajout de Commons maths .JAR

De plus, le projet commencera à construire (compilation) automatiquement, ce qui peut prendre de 2 à 10 minutes selon la configuration de votre machine. Une fois le processus de construction terminé, vous pouvez commencer à travailler avec n'importe quel exemple CloudSim fourni dans le dossier '/ examples'.

### IV.3. Test

Dans cette section, nous montrons le rendu de la simulation de notre projet dans CloudSim. La classe contenant la méthode main de notre projet a été intitulée NotreTravail nous l'avons mise dans le package `org.cloudbus.cloudsim.examples.power.random`

Pour faire la simulation de notre projet, on suit le chemin suivant : Eclipse > Cloudsim > Exemples > `org.cloudbus.cloudsim.examples.power.random` après position de la souris sur la Classe NotreTravail ce qui nous donnera le résultat suivant :

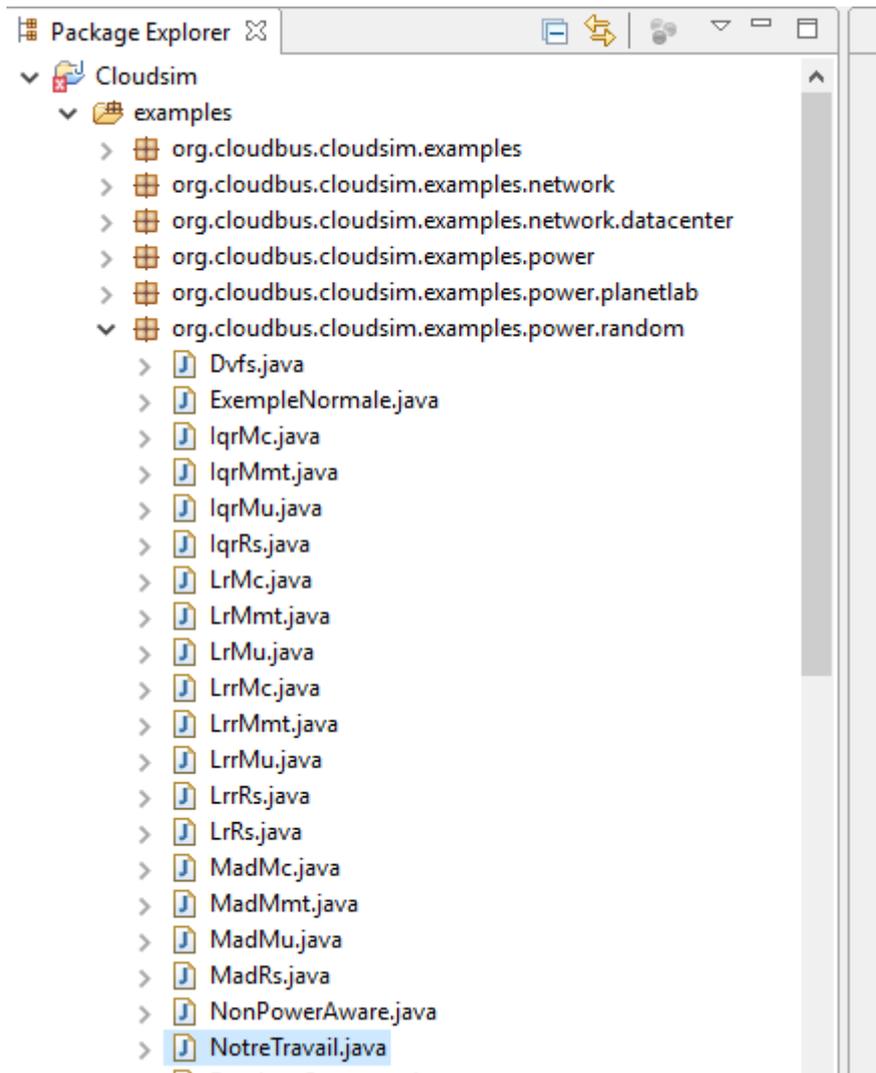


Figure 26: La classe NotreTravail dans CloudSim

Un clic droit sur la classe Notre travaille nous donnera des choix à faire, on choisit Run As > 1 java application ce qui démarre la simulation de notre projet.

### ➤ **Exemple de simulation**

Voici quelques-uns des résultats de Simulation :

```

Console [D] NotreTravail.java
<terminated> NotreTravail [Java Application] C:\Program Files\Java\jre1.8.0_131\bin\javaw.exe (20 sept. 2017 14:48:18)

86109,91: [Host #14] utilization is 0,00%

86109,91: [Host #15] Total allocated MIPS for VM #2 (Host #15) is 1891,42, was requested 1891,42 out of total 2500,00 (75,66%)
86109,91: [Host #15] MIPS for VM #2 by PEs (2 * 2660.0). PE #0: 1891,42.
86109,91: [Host #15] Total allocated MIPS for VM #38 (Host #15) is 298,93, was requested 298,93 out of total 1000,00 (29,89%)
86109,91: [Host #15] MIPS for VM #38 by PEs (2 * 2660.0). PE #0: 298,93.
86109,91: [Host #15] Total allocated MIPS for VM #20 (Host #27) is 154,52, was requested 1545,16 out of total 2000,00 (77,26%)
86109,91: [Host #15] MIPS for VM #20 by PEs (2 * 2660.0). PE #0: 154,52.
86109,91: [Host #15] VM #20 is being migrated to Host #15
86109,91: [Host #15] utilization is 44,08%

86109,91: [Host #16] utilization is 0,00%

86109,91: [Host #17] Total allocated MIPS for VM #6 (Host #17) is 337,87, was requested 375,41 out of total 2500,00 (15,02%)
86109,91: [Host #17] MIPS for VM #6 by PEs (2 * 2660.0). PE #0: 337,87.
86109,91: [Host #17] Under allocated MIPS for VM #6: 37,54
86109,91: [Host #17] VM #6 is in migration
86109,91: [Host #17] utilization is 7,06%

86109,91: [Host #18] utilization is 0,00%

86109,91: [Host #19] Total allocated MIPS for VM #24 (Host #19) is 332,35, was requested 332,35 out of total 2000,00 (16,62%)
86109,91: [Host #19] MIPS for VM #24 by PEs (2 * 2660.0). PE #0: 332,35.
86109,91: [Host #19] Total allocated MIPS for VM #28 (Host #19) is 559,77, was requested 559,77 out of total 1000,00 (55,98%)
86109,91: [Host #19] MIPS for VM #28 by PEs (2 * 2660.0). PE #0: 559,77.
86109,91: [Host #19] Total allocated MIPS for VM #9 (Host #19) is 762,65, was requested 847,39 out of total 2500,00 (33,90%)
86109,91: [Host #19] MIPS for VM #9 by PEs (2 * 2660.0). PE #0: 762,65.
86109,91: [Host #19] Under allocated MIPS for VM #9: 84,74
86109,91: [Host #19] VM #9 is in migration
86109,91: [Host #19] utilization is 32,70%

```

Figure 27: Résultat

Sur la console CloudSim s'affiche les résultats de notre simulation ci-dessus :

- Premier paragraphe, on travaille sur l'hôte #15.
- Sur la première ligne on voit l'utilisation de l'hôte #14 qui est de 0,00%.
- L'utilisation des hôtes est obtenue par la division entre les mips requis et le nombre total des mips de l'hôte.
- Sur la deuxième ligne il y a le nombre de mips alloués pour la machine virtuelle #2 (appartenant à l'hôte #15) est de 1891,42 il avait été demandé pour la même machine 1891,42 pour un total de 2500. Ce total est obtenu par la somme des mips de tous les VMs de l'hôte et les mips demandés pour cette machine virtuelle. A la fin on a entre parenthèses l'utilisation de l'hôte.
- Sur la cinquième ligne du premier paragraphe, le nombre de mips alloués pour la machine virtuelle #20 (appartenant à l'hôte #27) est de 154,52 il avait été demandé 1545,16. Cette machine est donc migrée à l'hôte #15
- Au deuxième paragraphe on travaille sur l'hôte #17, le nombre de mips alloués pour la machine virtuelle #6 (appartenant à l'hôte #17) est de 337,87 il avait été demandé pour la même machine 375,41 pour un total de 2500.
- Il y a donc eu une sous allocation des ressources ce qui fait que la MV #6 va faire sa migration dans un autre hôte. Ceci va donner un même scénario pour la machine virtuelle #9 de l'hôte #19

Notre simulation c'est déroulé selon l'environnement et selon le rendu suivant :

```
Number of hosts: 50
Number of VMs: 50
Total simulation time: 86400,00 sec
Energy consumption: 41,81 kWh
Number of VM migrations: 4839
SLA: 0,03048%
SLA perf degradation due to migration: 0,23%
SLA time per active host: 12,99%
Overall SLA violation: 3,25%
Average SLA violation: 12,81%
Number of host shutdowns: 1424
Mean time before a host shutdown: 929,70 sec
StDev time before a host shutdown: 1348,87 sec
Mean time before a VM migration: 16,82 sec
StDev time before a VM migration: 7,67 sec
Execution time - VM selection mean: 0,00006 sec
Execution time - VM selection stDev: 0,00024 sec
Execution time - host selection mean: 0,00025 sec
Execution time - host selection stDev: 0,00044 sec
Execution time - VM reallocation mean: 0,00227 sec
Execution time - VM reallocation stDev: 0,01128 sec
Execution time - total mean: 0,00563 sec
Execution time - total stDev: 0,01701 sec
```

Figure 28 : Environnement de travail

**Légende :**

Nombre d'hôtes : 50

Nombre de VMs : 50

Temps total de simulation : 86400,00 sec

Consommation de l'énergie : 41,81 kWh

Nombre de migrations de VM : 4839

SLA : 0,03048%

La dégradation de la SLA due à la migration : 0.23%

Le temps de la SLA par hôte actif : 12,99%

Total de violation de la SLA : 3,25%

Moyenne de violation de la SLA : 12,81%

Nombre d'hôtes éteints : 1424

Temps nécessaire pour qu'un hôte soit éteint : 929,70 sec

Le temps standard Deviation pour qu'un hôte soit éteint : 1348,87sec

Le temps nécessaire avant la migration d'une VM : 16,82sec

Le temps standard Deviation avant la migration d'une VM : 7,67sec

Temps d'exécution - sélection aboutie de Vm : 0,00006 sec

Temps d'exécution - sélection de Vm standard Deviation: 0.0024 sec

Temps d'exécution – sélection aboutie de l'hôte : 0.00025 sec

Temps d'exécution - sélection standard Deviation de l'hôte : 0.00044 sec

Temps d'exécution - Réallocation aboutie de Vm : 0.00227 sec

Temps d'exécution - Réallocation standard Deviation de Vm : 0.01128 sec

Temps d'exécution – total abouti : 0,00563 sec

Temps d'exécution – total standard Deviation : 0,01701 sec

## **IV.4. Conclusion**

Dans ce chapitre, nous avons vu les outils de notre simulation ainsi que le langage utilisé pour le faire. A la fin nous avons mis en évidence un test de notre projet réalisé sur CloudSim.

CloudSim est un outil très intéressant qui nous a permis de faire une simulation de notre cloud pour des résultats clairs et limpides. Chaque situation donnée pouvant être interprétée. Il est certes en anglais mais permet de faire un test et nous donne à la fin un résultat détaillé de la simulation avec à la fin des statistiques de tout ce qui s'est passé durant la simulation.

## Conclusion générale et perspectives

Le Cloud computing est devenu au cours de ces dix dernières années un paradigme important dans le monde informatique. En général, son principe est de fournir des services décentralisés aux clients et de consommer les ressources dans un modèle pay-as-you-go. La demande croissante pour ce type de service oblige les fournisseurs des services Clouds à augmenter la taille de leur infrastructure au point que la consommation de l'énergie devient très importante.

Des études récentes visent à concevoir et développer des technologies Cloud tout en se focalisant sur la définition de nouvelles méthodes, des politiques et des mécanismes efficaces de gestion des infrastructures, d'allocation des ressources et surtout la gestion de consommation énergétique. Cette dernière fait beaucoup parler des nombreux grands pays sur la protection de l'environnement et l'application de l'informatique verte ou green computing. Ce concept vise à réduire l'empreinte écologique, économique, et sociale des technologies de l'information et de la communication. Il s'agit d'une manière globale et cohérente de réduire les nuisances rencontrées dans le domaine des équipements informatiques et ce, durant l'ensemble de la durée de vie de chaque équipement : soit aux différents stades de fabrication, d'utilisation.

Comme tout autre système distribué, les centres de données du Cloud consomment une énorme quantité d'énergie électrique. Cette énergie ne provoque pas seulement la diminution des bénéfices ou un ralentissement au niveau des Datacenters, mais aussi l'émission d'une grande quantité de dioxyde de carbone [A]. Afin de avoir une meilleure utilisation des ressources et de réduire cette consommation énergétique, nombreuses sont les techniques utilisées par les fournisseurs des services tels que la virtualisation, l'ordonnancement des tâches, la migration etc. Ce mémoire a pour objectif d'étudier et de proposer une approche de migration à chaud de machines virtuelles dans le Cloud computing. Dont le but est de réduire la consommation de l'énergie au niveau de Datacenters de Cloud en procédant la migration des machines virtuelles.

La réalisation de ce travail nous a conduits à suivre les étapes suivantes :

- En premier lieu, nous avons parlé de la virtualisation dans laquelle on trouve les différents types de virtualisation, les techniques diverses ainsi que ses avantages. Nous avons également présenté le Cloud computing dans ce premier chapitre en donnant ses définitions, ses caractéristiques les modèles des services et de déploiement ainsi que différents avantages et quelques risques.
- Subséquemment, nous avons défini l'allocation des ressources avec ces techniques : allocation dynamique et statique ainsi que l'allocation des machines virtuelles. Dans cette même section, nous avons décortiqué la migration des machines virtuelles en expliquant d'une manière détaillée ses différentes méthodes.
- Par la suite, nous avons proposé notre approche de migration de machines virtuelles dans le Cloud par rapport à la consommation d'énergie. Cette méthode comprend deux phases : la phase de sélection et la phase d'allocation et nous nous basons sur l'utilisation de CPU par les Datacenters pour détecter la surcroissance d'énergie au niveau d'un hôte physique. Nous avons terminé ce paragraphe par une conclusion et l'architecture générale de l'approche proposée.

- Enfin, dans la quatrième partie, nous avons représenté les outils de développement qui sont l'Eclipse IDE Java et le simulateur de Cloud CloudSim. Ce dernier est une plateforme pour modéliser et simuler les infrastructures et les services de Cloud. Il a été conçu par le laboratoire de l'Université de Melbourne, Australie dans un cadre de recherche académique et il est développé à base du langage Java.

Dans la continuité de notre travail, nous envisageons beaucoup des perspectives. Nous projetons dans nos futurs travaux de pouvoir étudier et appliquer cette étude dans un environnement réel de Cloud, d'améliorer notre approche en considérant différentes contraintes comme la capacité de stockage et le flux de la bande passante. Nous voudrions également perfectionner cette approche avec des critères tels que la détection d'augmentation de chaleur et la surcroissance de trafic réseaux.

## Liste des figures :

|   |    |
|---|----|
| Figure 1: La virtualisation complète .....  | 6  |
| Figure 2: La paravirtualisation .....   | 7  |
| Figure 3: La virtualisation au niveau d'une application.....                          | 8  |
| Figure 4: La virtualisation au niveau du noyau .....                                  | 9  |
| Figure 5: La virtualisation par l'utilisation d'un hyperviseur .....                  | 9  |
| Figure 6: la virtualisation par l'emploi d'un hyperviseur de machines virtuelles..... | 10 |
| Figure 7: Le réseau interne d'une entreprise sans le cloud .....                      | 11 |
| Figure 8: Différence entre IaaS, PaaS et SaaS .....                                   | 19 |
| Figure 9: Utilisation des ressources du Cloud.....                                    | 32 |
| Figure 10: Répartition de consommation énergétique d'un centre de données [2.6] ..... | 35 |
| Figure 11: Algorithme de pré-copie .....  | 38 |
| Figure 12: Algorithme de post-copie .....   | 40 |
| Figure 13: Algorithme de post-copie hybride .....                                     | 41 |
| Figure 14: Algorithme de la phase de sélection .....                                  | 52 |
| Figure 15: Diagramme d'activité de la phase sélection .....                           | 53 |
| Figure 16: Algorithme de l'allocation de VM .....                                     | 54 |
| Figure 17: Diagramme d'activité de la phase d'allocation.....                         | 55 |
| Figure 18: Architecture générale du système .....                                     | 56 |
| Figure 19: Environnement de développement Eclipse.....                                | 60 |
| Figure 20: Architecture du CloudSim .....   | 62 |
| Figure 21: Diagramme de classe du simulateur CloudSim.....                            | 62 |
| Figure 22: Ajout de CloudSim dans Eclipse .....                                       | 65 |
| Figure 23: Création de projet java pour CloudSim .....                                | 66 |
| Figure 24: Ajout de fichiers JAR .....  | 67 |
| Figure 25: Ajout de Commons maths .JAR .....  | 68 |
| Figure 26: La classe NotreTravail dans CloudSim .....                                 | 69 |
| Figure 27: Résultat .....   | 70 |
| Figure 28 : Environnement de travail.....   | 71 |

## Liste des abréviations et Sigles

|       |   |
|-------|---|
| BP    | Bande Passante                                    |
| CAPEX | Capital Expenditure                               |
| CO2   | Dioxyde de Carbone                                |
| CPU   | Central Processing Unit                           |
| DHCP  | Dynamic Host Configuration Protocol               |
| DNS   | Domain Name System                                |
| IaaS  | Interface as a Services                           |
| IBM   | International Business Machine                    |
| IDE   | Integrated Development Environment                |
| IT    | Information Technology                            |
| JRE   | Java Runtime Environment                          |
| JVM   | Java Virtual Machine                              |
| KVM   | Kernel-based Virtual Machine                      |
| MBFD  | Modified Best Fir Decreasing                      |
| MIPS  | Millions Instructions Per Second                  |
| MMV   | Moniteur Machines Virtuelles                      |
| NAS   | Network Attached Storage                          |
| NT    | Nouvelles technologies                            |
| NTP   | Network Time Protocol                             |
| OPEX  | Operational Expenditure                           |
| PaaS  | Platform as a Services                            |
| PM    | Physical Machines                                 |
| PME   | Petite et Moyenne Entreprise                      |
| PMs   | Machines Physiques                                |
| QoS   | Qualité de service                                |
| RAID  | Redundant Array of Independant Disks              |
| RAM   | Random Access Memory                              |
| SaaS  | Software as a Services                            |
| SAN   | Storage Area Network                              |
| SLA   | Service Level Agreements                          |
| TIC   | Technologies d'Information et de la Communication |
| UML   | Unified Modeling Language                         |
| VDC   | Virtual Data Center                               |
| VLAN  | Virtual Local Area Network                        |
| VM    | Virtual Machine                                   |
| VMM   | Virtual Machine Manager                           |
| VMs   | Machines Virtuelles                               |
| VPN   | Virtual Private Network                           |
| XML   | eXtensible Markup Language                        |

## Sources :

- [1.1] : <https://fr.wikipedia.org/wiki/Virtualisation/>
- [1.2] : <http://www.wisper.io/fr/les-grands-principes-de-la-virtualisation/>
- [1.3] : Les techniques de virtualisation par Thierry DOSTES et Maurice LIBES ; Aux éditions JT SIARS septembre 2009.
- [1.4] : <https://wapiti.telecom-lille.fr/commun/ens/peda/options/st/rio/pub/exposes/exposesrio2009/COLIN-DESBUREAUX/La-Virtualisation-Différents-types-de-virtualisation.html>
- [1.5] : <https://missarte.wordpress.com/lhistoire-du-cloud-computing/>
- [1.6] : [https://fr.wikipedia.org/wiki/Cloud\\_computing](https://fr.wikipedia.org/wiki/Cloud_computing)
- [1.7] : Mell, P. et Grance, T. The NIST Definition of Cloud Computing, Septembre 2011.
- [1.8] : Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R. Konwinski, A., Lee, G. Patterson, D., Rabkin, A. and Stoi, I. Above the clouds: A Berkeley View of cloud computing, Technical Report, UCB/EECS-2009-28, UC Berkeley Reliable Adaptive Distributed Systems Laboratory, EECS Department, University of California, Berkeley, February, 2009.
- [2.0] : N. Asha ET G. Raghavendra Rao, “A Review on Various Resource Allocation Strategies in Cloud Computing”, International Journal of Emerging Technology and Advanced Engineering (IJETA), July 2013.
- [2.1] : V. Vinothina, S. Lecturer ET al, “A Survey on Resource Allocation Strategies in Cloud Computing”, International Journal of Advanced Computer Science and Applications (IJACSA) 2012.
- [2.2] : N. Krishnaveni ET G. Sivakumar, “Survey on Dynamic Resource Allocation Strategy in Cloud Computing Environment”, International Journal of Computer Applications Technology and Research (IJCAT), 2013.]
- [2.3] : [https://fr.wikipedia.org/wiki/Centre\\_de\\_données](https://fr.wikipedia.org/wiki/Centre_de_données)
- [2.4] : <http://www.sciencedirect.com/science/article/pii/S1876610215028337>
- [2.5] : [http://www.lemonde.fr/pixels/article/2016/06/03/les-datas-du-grand-froid\\_4932566\\_4408996.html#5Z31I6DyHYqjBQaj.99](http://www.lemonde.fr/pixels/article/2016/06/03/les-datas-du-grand-froid_4932566_4408996.html#5Z31I6DyHYqjBQaj.99)
- [2.6] : [https://fr.wikipedia.org/wiki/Gestion\\_énergétique\\_des\\_centres\\_de\\_calcul](https://fr.wikipedia.org/wiki/Gestion_énergétique_des_centres_de_calcul)
- [2.7] : Rodrigo N Calheiros, Rajiv Ranjan, Anton Beloglazov, César AF De Rose et Rajkumar Buyya. \_CloudSim : « a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms » In : Software Practice and Experience 41.1 (2011), p. 23\_50 (Cité en pages 24, 89).

- [2.8] : Haikun Liu, Cheng-Zhong Xu, Hai Jin, Jiayu Gong et Xiaofei Liao. « Performance and energy modeling for live migration of virtual machines ». In : HPDC. Sous la dir. d'Arthur B. Maccabe et Douglas Thain. ACM, 2011, p. 171\_182 (Cit  en pages 25, 52, 76, 88, 94).
- [2.9] : Jie Zheng, TS Ng, Kunwadee Sripanidkulchai et Zhaolei Liu. \_Pacer : « A Progress Management System for Live Virtual Machine Migration in Cloud Computing ». In : IEEE Transactions on Network and Service Management 10.4 (d c. 2013), p. 369\_382 (Cit  en pages 25, 41).
- [2.10] : D. Ardagna, B. Panicucci, M. Trubian, L. Zhang, « Energyaware autonomic resource allocation in multitier virtualized environment », IEEE Transactions on Services Computing, pp 1-18, Volume : 5, Issue : 1, 2012.
- [2.11] : Buyya, R., Beloglazov, A. and Abawajy, J., 2010 : « Energy-efficient management of data center resources for cloud computing : a vision, architectural elements, and open challenges ».
- [3.1] : J. Cao, Y. Wu et M. Li, « Energy efficient allocation of virtual machine in cloud computing environment based on demand forecast » : in Advance in Grid and Prevasive Computing, vol 7296, pp 137-151, Springer Berlin Heidelberg, 2012
- [3.2] : X. Fan, W-D. Weber, et L. A. Barroso, « Power provisioning for a warehouse-sized computer », dans Proceeding of the 34th Annual Intl. Symp. On Computer Architecture, (New York, NY, USA) pp. 13-23, ACm June 2007.
- [3.3] : R. Sinha, N. Purohit, et H. Diwanji, « Energy effecient dynamic integretion of thresholds for migration at cloud data », Special Issue of International Journal of Computer Applications on Communicaton ant Network, no. 11, pp. 44-49, 2011.
- [3.4] : K. L. T. Lee et H. Huang, « Dynamic resource management for energy saving in the cloud computing environement », in Proceeding of the International Conference on Information Science and Industrial Applications, vol. 4, pp. 176-181, May 2012.
- [3.5] : R. Sinha, N. Purohit, et H. Diwanji ; « Power aware live migration for data centers in cloud computing using dynamic threshold » dans International Journal of Computer Technology and Applications, vol 2, no. 6, pp. 2041-2046, 2011.
- [3.6] : A. Beloglazov, J. Abawajy, et R. Buuya ; « Energy aware resource allocation heuristics for efficient management of data centers for cloud computing », dans Future Generation Computer Systemes, vol. 28, no. 5, p. 755-768, 2012.
- [3.7] : J. Kig et S. Park ; « Algorithmes for the variable sized bin packing problem », dans European Journal of Operational Research, vol. 147, no. 2, pp. 365-372, 2003
- [3.8] : D. Dad, E. Yagoubi et G. Belalem ; « Energy efficient VM live migration and allocation at cloud data centers », dans International Journal of Cloud Application and Computing (IJCAC), vol. 4, no. 4, pp. 55-63, 2014.
- [3.9] : [https://fr.wikipedia.org/wiki/UML\\_\(informatique\)](https://fr.wikipedia.org/wiki/UML_(informatique))
- [4.1] : [https://en.wikipedia.org/wiki/Java\\_\(programming\\_language\)](https://en.wikipedia.org/wiki/Java_(programming_language))
- [4.2] : <http://ipeti.forumpro.fr/t21-definition-de-langage-java-java-script>

[4.3] : <https://www.opensourcemacssoftware.org/developpement/eclipse-ide-java-mac.html>

[4.4] : <https://anupinders.gitbooks.io/up-and-running-with-cloudsim/content/models-of-cloud-compnents.html/>

[t.1] : W. Song, Z. Xiao, Q. Chen, H. Luo : « Adaptive resource provisioning for the cloud using online bin packing ». Computers, IEEE Transactions on, 63 (11) (2014), pp. 2647-2660

[t.2] : Ghribi, C., Hadji, M. and Zeghlache, D., 2013, May. « Energy efficient vm scheduling for cloud data centers : Exact allocation and migration algorithms. In Cluster, Cloud and Grid Computing (CCGrid) », 2013 13th IEEE/ACM International Symposium on (pp. 671-678). IEEE.

[t.3] : J. Dong, H. Wang, S. Cheng : « Energy-performance tradeoffs in IaaS cloud with virtual machine scheduling ». Communications, China, 12 (2) (2015), pp. 155-166

[t.4] : Bobroff, N., Kochut, A. and Beaty, K., 2007, May. « Dynamic placement of virtual machines for managing sla violations ». In Integrated Network Management, 2007. IM'07. 10th IFIP/IEEE International Symposium on (pp. 119-128). IEEE.

[t.5] : Wang, M., Meng, X. and Zhang, L., 2011, April. « Consolidating virtual machines with dynamic bandwidth demand in data centers », In INFOCOM, 2011 Proceedings IEEE (pp. 71-75). IEEE.

[t.6] : Chen, M., Zhang, H., Su, Y.Y., Wang, X., Jiang, G. and Yoshihira, K., 2011, May. Effective VM sizing in virtualized data centers. In Integrated Network Management (IM), 2011 IFIP/IEEE International Symposium on (pp. 594-601). IEEE.

[A] : [https://fr.wikipedia.org/wiki/Informatique\\_durable/](https://fr.wikipedia.org/wiki/Informatique_durable/)