

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement supérieur et de la Recherche scientifique  
Université Mohammed Seddik Ben Yahia – jijel  
Faculté des Sciences Exactes et Informatique  
Département d'Informatique



MÉMOIRE DE FIN D'ÉTUDES

pour l'obtention du diplôme de **Master**

en : **Informatique**

Options : **Système D'Information Et Aide À la Décision (S.I.A.D)**  
Thème

---

# Migration de VMs (Virtual Machines) dans le Cloud Computing : Étude des approches d'optimisation de l'énergie basées sur des seuils

---

Réalisé par :

**Aymene SADOON**

**Nazim MESTAR**

Encadré par : **Mme. Messaouda AYACHI**

— Année académique 2024/2025 —

# Dédicace

À nos chers parents, pour leur amour incommensurable, leur patience et leurs sacrifices silencieux. Sans leur soutien moral, leurs prières et leur confiance inébranlable, rien de tout cela n'aurait été possible.

À nos familles, pour leur présence bienveillante, leurs encouragements constants et leur compréhension durant les moments les plus exigeants de ce parcours.

À nos amis proches, avec qui nous avons partagé non seulement des instants de joie, mais aussi les doutes, les efforts, et les nuits blanches. Votre amitié fut un moteur essentiel dans notre cheminement.

Ce mémoire vous est dédié avec toute notre reconnaissance et notre affection. Puissent ces quelques pages être à la hauteur de ce que vous avez semé en nous.

# Remerciements

Au terme de ce travail, nous tenons à exprimer notre plus sincère gratitude à toutes les personnes qui nous ont soutenus, guidés et inspirés durant la réalisation de ce mémoire.

Nous adressons nos remerciements les plus chaleureux à notre encadrante, **Madame Messaouda AYACHI**, pour sa disponibilité constante, sa patience, ainsi que la qualité de ses orientations pédagogiques et scientifiques. Son encadrement rigoureux et bienveillant a été déterminant dans la progression de ce projet.

Nous exprimons également notre reconnaissance à l'ensemble de l'équipe pédagogique du **département d'informatique**, pour l'enseignement riche et structurant qu'ils nous ont dispensé tout au long de notre parcours universitaire. Leur implication et leur professionnalisme ont grandement contribué à notre formation académique et personnelle.

Nous n'oublions pas de remercier nos familles respectives, pour leur soutien inconditionnel, leur encouragement quotidien et leur confiance indéfectible, sans lesquels nous n'aurions pu surmonter les défis rencontrés.

Nos pensées les plus sincères vont aussi à nos amis et camarades, pour leur esprit de solidarité, leur entraide précieuse, et les moments partagés tout au long de cette aventure universitaire.

À toutes et à tous, **merci du fond du cœur.**

# Résumé

Le Cloud Computing est aujourd'hui une technologie incontournable permettant l'accès à des ressources informatiques à la demande. Grâce à la virtualisation, il est possible d'exécuter plusieurs machines virtuelles (VMs) sur un même hôte physique, améliorant ainsi l'utilisation des ressources. Cependant, cette efficacité est contrebalancée par une consommation énergétique importante des centres de données.

Ce mémoire s'intéresse à l'optimisation énergétique dans le cloud via la migration dynamique des VMs. Il permet de présenter le Cloud Computing, ses modèles de service et de déploiement, puis détailler les mécanismes de virtualisation et les techniques de migration à froid et à chaud. Un état de l'art des approches existantes est ensuite présenté. L'étude expérimentale repose sur des simulations menées avec CloudSim, où cinq stratégies de migration ont été mises en œuvre : sans consolidation (**SSTP** (Static Single Threshold Policy)), **DSTP** (Dynamic Single Threshold Policy)), et avec consolidation (**SDTP** (Static Double Threshold Policy)), **DDTP** (Dynamic Double Threshold Policy)), **HDTP** (Hybrid Double Threshold Policy)).

Les résultats montrent que les politiques avec consolidation, notamment SDTP et HDTP, permettent une réduction significative de la consommation énergétique, du nombre de migrations et des hôtes actifs, par rapport à une politique sans migration (WNMP) ou l'approche dynamique (DDTP).

**Mots-clés :** Cloud Computing, virtualisation, migration de machines virtuelles, consommation énergétique, CloudSim, consolidation, optimisation, machine virtuelle (VM), consommation d'énergie, hôte surchargé, hôte sous-utilisé, statique, dynamique, hybride, seuil, seuil unique, double seuil.

# Abstract

Cloud Computing has become an essential technology, enabling on-demand access to computing resources. Thanks to virtualization, it is possible to run multiple virtual machines (VMs) on a single physical host, thereby improving resource utilization. However, this efficiency is counterbalanced by the high energy consumption of data centers.

This thesis focuses on energy optimization in cloud environments through the dynamic migration of virtual machines. It begins by presenting the fundamentals of Cloud Computing, including its service and deployment models, and then describes virtualization mechanisms as well as cold and live migration techniques. A state-of-the-art review of existing approaches is then introduced.

The experimental study is based on simulations conducted using CloudSim Plus, in which five migration strategies were implemented. Two strategies rely on single-threshold policies without consolidation, namely the Static Single Threshold Policy (SSTP) and the Dynamic Single Threshold Policy (DSTP). The other three strategies are based on double-threshold policies with consolidation : the Static Double Threshold Policy (SDTP), the Dynamic Double Threshold Policy (DDTP), and the Hybrid Double Threshold Policy (HDTP).

**Keywords :** Cloud Computing, virtualization, virtual machine migration, energy consumption, CloudSim, consolidation, optimization, Virtual Machine (VM), power consumption, overloaded host, underloaded host, static, dynamic, hybrid, threshold, single threshold, double threshold.

# Liste des acronymes

<b>DC</b>	Datacenter – Centre de données.
<b>VM</b>	Virtual Machine – Machine virtuelle.
<b>PM</b>	Physical Machine – Machine physique.
<b>CPU</b>	Central Processing Unit – Processeur central.
<b>RAM</b>	Random Access Memory – Mémoire vive.
<b>BW</b>	Bandwidth – Bande passante.
<b>Wh</b>	Watt-hour – Unité de consommation d'énergie.
<b>mig</b>	Migration – Déplacement d'une machine virtuelle d'un hôte à un autre.
<b>SLA</b>	Service Level Agreement - Accord de niveau de service
<b>nbr</b>	Nombre – Abréviation utilisée pour "nombre" (ex. : nbr de migrations).
<b>SSTP</b>	Static Threshold Policy – Politique à seuil statique.
<b>DSTP</b>	Dynamic Threshold Policy – Politique à seuil dynamique.
<b>HDTP</b>	Hybride Double Treshold Policy
<b>SDTP</b>	Static Double Treshold Policy
<b>DDTP</b>	Dynamic Double Threshold Policy – Politique dynamique à double seuil.
<b>T<sub>upper</sub></b>	Upper Threshold – Seuil supérieur de surcharge.
<b>T<sub>lower</sub></b>	Lower Threshold – Seuil inférieur de sous-utilisation.
<b>IaaS</b>	Infrastructure as a Service – Infrastructure en tant que service.
<b>Power Aware</b>	Power Aware Energy Module For Virtuelle Machine Migration
<b>PaaS</b>	Platform as a Service – Plateforme en tant que service.
<b>SaaS</b>	Software as a Service – Logiciel en tant que service.
<b>DCM</b>	Datacenter Manager – Gestionnaire de centre de données.
<b>BFD</b>	Best Fit Decreasing
<b>NIST</b>	National Institute of Standards and Technology – Institut national des normes et de la technologie.
<b>IEEE</b>	Institute of Electrical and Electronics Engineers – Institut des ingénieurs en électricité et en électronique.
<b>AWS</b>	Amazon Web Services – Services Cloud d'Amazon.

# Table des matières

<b>1</b>	<b>Cloud Computing</b>	<b>15</b>
1.1	Introduction . . . . .	15
1.2	Définition, caractéristiques et évolution du cloud computing . . . . .	15
1.2.1	Définition du cloud computing . . . . .	15
1.2.2	Caractéristiques fondamentales du cloud computing . . . . .	16
1.2.3	Évolution du Cloud Computing . . . . .	17
1.3	Les modèles de service du Cloud Computing . . . . .	18
1.3.1	Infrastructure as a Service (IaaS) . . . . .	18
1.3.2	Platform as a Service (PaaS) . . . . .	19
1.3.3	Software as a Service (SaaS) . . . . .	19
1.4	Les modèles de déploiement du Cloud . . . . .	20
1.4.1	Cloud Public . . . . .	21
1.4.2	Cloud Privé . . . . .	21
1.4.3	Cloud Hybride . . . . .	21
1.4.4	Cloud Communautaire . . . . .	21
1.4.5	Comparaison des modèles de déploiement . . . . .	21
1.5	Composants et architecture du Cloud Computing . . . . .	22
1.5.1	Les composants du cloud computing . . . . .	22
1.5.2	L'architecture référentielle du NIST . . . . .	23
1.6	Le marché du cloud computing . . . . .	25
1.6.1	Les fournisseurs majeurs du Cloud Computing . . . . .	25
1.6.2	Analyse du marché du cloud computing . . . . .	26
1.7	Avantages et inconvénients du cloud computing . . . . .	27
1.7.1	Avantages du cloud . . . . .	27
1.7.2	Inconvénients du cloud . . . . .	27
1.8	Défis du Cloud Computing . . . . .	28
1.9	Conclusion . . . . .	28
<b>2</b>	<b>La virtualisation et la migration de VMs</b>	<b>29</b>
2.1	Introduction . . . . .	29
2.2	La virtualisation . . . . .	29
2.2.1	Définition et caractéristiques . . . . .	29
2.2.2	Machine virtuelle (VM Virtual Machine) . . . . .	31
2.2.2.1	Concept de machine virtuelle . . . . .	31
2.2.2.2	Applications des machines virtuelles . . . . .	31
2.2.3	L'hyperviseur (VMM Virtual Machine Monitor) . . . . .	31
2.2.3.1	Définition . . . . .	31
2.2.3.2	Hyperviseur de type II . . . . .	32

2.2.3.3	Hyperviseur de type I . . . . .	32
2.2.4	Techniques de virtualisation . . . . .	33
2.2.4.1	Émulation . . . . .	33
2.2.4.2	Virtualisation assistée par le matériel . . . . .	34
2.2.4.3	Virtualisation complète . . . . .	34
2.2.4.4	Para-virtualisation . . . . .	34
2.2.4.5	Conteneurisation (virtualisation légère) . . . . .	35
2.2.4.6	Tableau récapitulatif des techniques de virtualisation . . .	35
2.2.5	Les types de virtualisation . . . . .	36
2.2.5.1	Virtualisation de serveurs . . . . .	36
2.2.5.2	Virtualisation de réseau . . . . .	36
2.2.5.3	Virtualisation de stockage . . . . .	37
2.2.5.4	Virtualisation de données . . . . .	37
2.2.5.5	Virtualisation de postes de travail . . . . .	37
2.2.5.6	Virtualisation d'application . . . . .	37
2.3	La migration des machines virtuelles . . . . .	37
2.3.1	Définition de la migration de VMs . . . . .	37
2.3.2	Technique de migration à froid «Stop and Copy Migration» . . . .	38
2.3.2.1	Définition de la migration à froid . . . . .	38
2.3.2.2	Avantages de la migration à froid . . . . .	39
2.3.2.3	Inconvénients de la migration à froid . . . . .	39
2.3.3	Technique de migration à chaud «Live Migration» . . . . .	39
2.3.3.1	Définition de la migration à chaud . . . . .	39
2.3.3.2	Migration par pré-copie . . . . .	40
2.3.3.3	Migration par post-copie . . . . .	41
2.3.3.4	Migration hybride . . . . .	41
2.3.4	Ressources d'exécution de la migration des machines virtuelles . . .	42
2.3.4.1	Ressources partagées (Shared Resources) . . . . .	42
2.3.4.2	Ressources dédiées (Dedicated Resources) . . . . .	43
2.3.4.3	Vers une approche équilibrée . . . . .	43
2.4	Conclusion . . . . .	43
<b>3</b>	<b>État de l'art sur la migration des VMs dans le cloud computing</b>	<b>45</b>
3.1	Introduction . . . . .	45
3.2	Mécanisme de migration des machines virtuelles . . . . .	45
3.2.1	Quand doit-on démarrer le processus de migration ? . . . . .	46
3.2.2	Comment choisir la machine virtuelle à migrer ? . . . . .	46
3.2.3	Où migrer la machine virtuelle ? . . . . .	47
3.3	La taxonomie proposée . . . . .	47
3.3.1	Classification selon les caractéristiques du problème Etudié . . . . .	49
3.3.2	Classification selon les méthodes de résolution . . . . .	49
3.3.3	Classification selon les techniques d'évaluation . . . . .	50
3.4	Étude bibliographique . . . . .	51
3.4.1	Energy-aware resource allocation heuristics for efficient manage- ment of data centers for Cloud computing . . . . .	51
3.4.2	Energy Efficient Dynamic Integration of Thresholds for Migration at Cloud Data Centers . . . . .	51



3.4.3	Energy Conscious Dynamic Provisioning of Virtual Machines using Adaptive Migration Thresholds in Cloud Data Center . . . . .	52
3.4.4	BGM-BLA : a new algorithm for dynamic migration of virtual machines in cloud computing . . . . .	53
3.4.5	EAMA : Efficient Adaptive Migration Algorithm for Cloud Data Centers (CDCs) . . . . .	53
3.4.6	A modelling and Simulation via CloudSim for Live Migration in Virtual Machines . . . . .	54
3.4.7	Parallelizing Live Migration of Virtual Machines . . . . .	54
3.4.8	Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in Cloud data centers . . . . .	55
3.4.9	pMapper : Power and Migration Cost Aware Application Placement in Virtualized Systems . . . . .	56
3.5	Tableau comparatif . . . . .	57
3.5.1	disuccion du tableau comparatif . . . . .	57
3.6	Conclusion . . . . .	58
<b>4</b>	<b>Implémentation et résultats expérimentaux</b>	<b>59</b>
4.1	Introduction . . . . .	59
4.2	Langage et environnement de développement . . . . .	59
4.2.1	Le langage de programmation Java . . . . .	59
4.2.2	L'environnement de développement IntelliJ IDEA . . . . .	60
4.2.3	Le simulateur CloudSim . . . . .	60
4.2.3.1	Présentation du CloudSim . . . . .	60
4.2.3.2	Raisons de l'adoption de CloudSim pour la modélisation et la simulation . . . . .	61
4.3	Conception du système de migration de VMs . . . . .	61
4.3.1	Processus de migration de VMs . . . . .	61
4.3.1.1	Description du processus . . . . .	61
4.3.2	Algorithmes de migration . . . . .	61
4.3.2.1	Algorithme de sélection de VMs à migrer . . . . .	62
4.3.2.2	Algorithme de placement des VMs . . . . .	62
4.4	Stratégies de migration de VMs étudiées . . . . .	63
4.5	Les métriques utilisées . . . . .	64
4.5.1	Consommation d'énergie . . . . .	65
4.5.2	Utilisation moyenne du CPU . . . . .	65
4.5.3	Calcul des indicateurs de migration . . . . .	65
4.5.4	Le calcule de seuil supérieur et seuil inférieur . . . . .	66
4.5.4.1	Calcul du seuil supérieur $T_{upper}$ . . . . .	66
4.5.4.2	Calcul du seuil inférieur $T_{lower}$ . . . . .	66
4.6	Implémentation et Interprétations . . . . .	67
4.6.1	Configuration de l'environement de simulation . . . . .	67
4.6.2	Étude de l'approche à seuil unique statique (SSTP) . . . . .	67
4.6.2.1	Impact de $T_{upper}$ sur la consommation énergétique dans l'approche SSTP : . . . . .	68
4.6.2.2	Étude des paramètres coût de migration en fonction des différentes valeurs de $T_{upper}$ . . . . .	69

4.6.3	Étude de l'approche à seuil unique dynamique (DSTP) . . . . .	70
4.6.3.1	Évolution de l'utilisation du CPU pour SSTP, DSTP et WNMP . . . . .	70
4.6.3.2	Analyse de la dynamique du seuil supérieur ( $T_{upper}$ ) dans les politiques SSTP et DSTP . . . . .	71
4.6.4	Comparaison des approches sans consolidation SSTP et DSTP . . .	72
4.6.4.1	Comparaison du nombre de migrations entre SSTP et DSTP sans consolidation . . . . .	72
4.6.4.2	Comparaison de la consommation énergétique entre SSTP et DSTP sans consolidation . . . . .	73
4.6.5	Comparaison des approches avec consolidation SDTP, DDTP, HDTP . . . . .	74
4.6.5.1	Comparaison énergétique entre DDTP, HDTP, SDTP et WNMP . . . . .	74
4.6.5.2	Comparaison du nombre de migrations entre DDTP, HDTP et SDTP et WNMP . . . . .	75
4.6.5.3	Comparaison du Nombre d'hôtes actifs entre DDTP, HDTP et SDTP et WNMP . . . . .	76
4.7	Conclusion . . . . .	77
<b>Conclusion générale</b>		<b>79</b>

# Liste des figures

1.1	Historique du cloud computing . . . . .	17
1.2	Les modèles de service du Cloud Computing [1]. . . . .	18
1.3	Les modèles de déploiement du Cloud . . . . .	21
1.4	Architecture conceptuelle de référence du NIST . . . . .	23
1.5	Interactions entre les acteurs de l'informatique en nuage . . . . .	25
2.1	Architecture virtualisée vs architecture traditionnelle . . . . .	30
2.2	La machine virtuelle VM et l'hyperviseur VMM . . . . .	31
2.3	Architectures des hyperviseurs de type II . . . . .	32
2.4	Architectures des hyperviseurs de Type I . . . . .	33
2.5	Processus de la migration à froid. . . . .	38
2.6	Les techniques de migration à chaud [9] . . . . .	40
2.7	Aperçu de la migration (à chaud) hybride des VMs . . . . .	42
3.1	Le processus de migration des machines virtuelles . . . . .	46
3.2	Taxonomie des techniques de migration de machines virtuelles dans un environnement cloud . . . . .	48
4.1	Schéma hiérarchique des stratégies étudiées de migration basées sur les seuils	64
4.2	Impact de la valeur du seuil $T_{upper}$ sur l'approche Single Static Threshold Policy SSTP . . . . .	68
4.3	Comparaison graphique des consommations énergétiques pour différentes stratégies . . . . .	69
4.4	l'utilisation du CPU en fonction des politiques SSTP, DSTP et WNMP . .	70
4.5	Analyse de l'évolution du seuil supérieur en fonction des politiques SSTP et DSTP . . . . .	71
4.6	Comparaison du nombre de migrations entre SSTP et DSTP sans consolidation . . . . .	72
4.7	Comparaison énergétique entre SSTP et DSTP sans consolidation . . . . .	73
4.8	Comparaison de la consommation énergétique totale des approches WNMP, DDTP, HDTP et SDTP . . . . .	74
4.9	Comparaison du nombre de migrations entre DDTP, HDTP et SDTP et WNMP . . . . .	75
4.10	Comparaison du Nombre d'hôtes actifs entre DDTP, HDTP et SDTP et WNMP . . . . .	76

# Liste des tableaux

1.1	Les avantages et les inconvénients des modèles de services du Cloud selon le client . . . . .	20
1.2	Les avantages et les inconvénients des modèles de services du Cloud selon le fournisseur . . . . .	20
1.3	Comparaison des modèles de déploiement du cloud computing . . . . .	22
1.4	Parts du marché des principaux fournisseurs Cloud (2024) . . . . .	26
2.1	Tableau récapitulatif des techniques de virtualisation . . . . .	36
3.1	Synthèse des approches de migration des VMs dans le Cloud (classées par année) . . . . .	57

# Liste des algorithmes

1	Algorithme de sélection des machines virtuelles à migrer . . . . .	62
2	Algorithme de placement basé sur la consommation énergétique . . . . .	63

# Introduction générale

Au cours des dernières décennies, l'évolution rapide des technologies de l'information a profondément bouleversé la manière dont les services numériques sont consommés, déployés et administrés. Le Cloud Computing, en particulier, s'est imposé comme un paradigme incontournable en offrant des ressources informatiques à la demande, flexibles, évolutives et accessibles via Internet. Ce modèle a permis aux entreprises de réduire considérablement leurs coûts d'infrastructure tout en gagnant en agilité, en résilience et en performance.

Au cœur de cette révolution technologique se trouve la virtualisation, une technologie clé qui permet l'exécution de plusieurs machines virtuelles (VMs) sur un même hôte physique, optimisant ainsi l'utilisation des ressources matérielles. Toutefois, cette efficacité potentielle s'accompagne de défis majeurs, notamment en matière de consommation énergétique, de gestion de la charge, de continuité de service, et de respect des accords de niveau de service (SLA). En effet, les centres de données modernes, qui constituent l'infrastructure de base du cloud, sont énergivores et ont une empreinte environnementale significative.

Face à ces enjeux, la migration des machines virtuelles s'impose comme un mécanisme essentiel pour améliorer la performance globale du système tout en minimisant les coûts énergétiques. En réallouant dynamiquement les VMs d'un hôte à un autre, il est possible de regrouper les charges, d'éteindre les serveurs sous-utilisés et d'assurer un équilibrage de charge efficace. Cependant, la migration elle-même engendre un coût en termes de bande passante, de temps de traitement, de perturbation des services, et donc un compromis délicat entre efficacité énergétique et qualité de service.

Dans ce contexte, ce travail se concentre sur les différentes approches de migration de machines virtuelles (VMs) dans le cloud computing. Pour mieux structurer et analyser les travaux existants, nous avons proposé une taxonomie globale et compréhensive, articulée autour de trois axes principaux :

1. Les caractéristiques du problème étudié, à travers :
  - L'objectif principal de la migration ;
  - Le type de migration considéré ;
  - L'environnement cloud considéré, représenté par l'unité (migration d'une VM ou groupe de VMs) et le domaine de migration (intra-cloud ou inter-cloud).
2. Les méthodes de résolution proposées sont regroupées en deux sous-catégories : l'approche et l'algorithme de résolution, avec un accent particulier mis sur les approches basées sur des seuils, qui se caractérisent par
  - Un type : seuil unique ou double.
  - Une politique : statique, dynamique ou hybride.
3. L'évaluation des approches, en distinguant :

- Les méthodes d'évaluation : pendant ou après la migration ;
- Les métriques utilisées : consommation énergétique, nombre de migrations, temps de migration, etc.

L'objectif principal de ce mémoire est donc d'étudier, implémenter et évaluer différentes stratégies de migration de VMs dans un environnement cloud simulé, afin d'identifier les approches les plus efficaces pour réduire la consommation énergétique tout en maîtrisant les perturbations induites par la migration.

À travers une simulation menée avec CloudSim, nous avons proposé d'étudier l'impact de *la politique de seuil (statique, dynamique, hybride) et du type de seuil (unique ou double)*, en lien avec la présence ou l'absence de *techniques de consolidation*.

Ainsi, nous avons analysé et comparé les approches suivantes :

**1. Deux politiques à seuil unique sans consolidation :**

- **SSTP** (Static Single Threshold Policy).
- **DSTP** (Dynamic Single Threshold Policy).

**2. Trois politiques à seuil double avec consolidation :**

- **SDTP** (Static Double Threshold Policy).
- **DDTP** (Dynamic Double Threshold Policy).
- **HDTP** (Hybrid Double Threshold Policy).

Ces politiques sont comparées à une stratégie de référence, **WNMP** (With No Migration Policy), qui ne prévoit aucune migration. L'évaluation repose sur plusieurs indicateurs clés : le nombre de migrations, le temps de migration, la consommation énergétique et le nombre d'hôtes actifs.

Enfin, ce mémoire est structuré en quatre chapitres principaux :

- Le premier chapitre présente les fondements du cloud computing et ses composants essentiels.
- Le deuxième chapitre est consacré à la virtualisation et aux mécanismes de migration des machines virtuelles.
- Le troisième chapitre dresse un état de l'art des travaux récents, et introduit une taxonomie globale permettant de les classer de manière cohérente.
- Le quatrième chapitre est dédié à l'implémentation des stratégies étudiées, à la simulation dans CloudSim Plus, ainsi qu'à l'analyse des résultats obtenus.

# Chapitre 1

## Cloud Computing

### 1.1 Introduction

Le cloud computing est l'une des tendances qui évoluent le plus rapidement à l'ère des technologies modernes. Contrairement aux services informatiques traditionnels, il propose des services à la demande, un stockage accessible à distance, une élasticité rapide, une connectivité étendue, une sécurité renforcée, ainsi qu'un accès ubiquitaire, à tout moment et en tout lieu.

Ces caractéristiques incitent de plus en plus d'entreprises à migrer leur infrastructure depuis des environnements locaux vers le cloud, afin de réduire les charges administratives et de tirer parti des architectures virtualisées des centres de données, offrant ainsi aux clients une grande flexibilité et une haute disponibilité.

Dans ce chapitre, nous présentons une étude sur le cloud computing, en mettant en lumière ses concepts fondamentaux, ses principes architecturaux, ses technologies de mise en œuvre, ainsi que les principaux défis de recherche.

### 1.2 Définition, caractéristiques et évolution du cloud computing

#### 1.2.1 Définition du cloud computing

Le concept de cloud computing a fait l'objet de nombreuses définitions proposées par divers organismes, chercheurs et entreprises de Technologies de l'Information (TI).

Cependant, la définition la plus couramment adoptée est celle du NIST (National Institute of Standards and Technology), formulée en ces termes [15] : "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or provider interaction." C'est donc un modèle permettant un accès réseau omniprésent, pratique et à la demande à un ensemble de ressources informatiques configurables et partagées, pouvant être rapidement allouées et libérées avec un minimum d'effort de gestion ou d'interaction avec le fournisseur.



Par ailleurs, contrairement à d'autres termes techniques, l'informatique en nuage n'est pas une nouvelle technologie, mais plutôt un nouveau modèle opérationnel qui rassemble un ensemble de technologies existantes pour faire fonctionner les entreprises d'une manière différente. En effet, la plupart des technologies utilisées dans l'informatique en nuage, telles que la virtualisation ou la tarification basée sur l'usage, ne sont pas nouvelles. L'informatique en nuage exploite plutôt ces technologies existantes pour répondre aux exigences technologiques et économiques de la demande actuelle en matière de technologies de l'information [46].

### 1.2.2 Caractéristiques fondamentales du cloud computing

Le cloud computing se distingue par un ensemble de caractéristiques fondamentales qui définissent son fonctionnement et qui le différencient des modèles informatiques traditionnels.

- Service à la demande (On-demand self-service) : Un consommateur peut provisionner unilatéralement des ressources informatiques, telles que le temps processeur ou le stockage réseau, selon ses besoins, automatiquement, sans interaction humaine avec chaque fournisseur de services [15].
- Accès réseau large (Broad network access) : Les ressources sont disponibles sur le réseau et accessibles via des mécanismes standard qui favorisent leur utilisation par divers types de clients légers ou lourds (par exemple : téléphones mobiles, ordinateurs portables, assistants numériques personnels) [15].
- Mutualisation des ressources (Resource pooling) : Les ressources informatiques du fournisseur sont mutualisées pour servir plusieurs clients selon un modèle multi-locataires, avec un réajustement dynamique des ressources physiques ou virtuelles selon la demande. Les clients n'ont généralement ni contrôle ni connaissance précise de l'emplacement des ressources, mais peuvent parfois spécifier un niveau de localisation abstrait (par exemple : pays, région ou centre de données). Les ressources concernées incluent le stockage, le traitement, la mémoire, la bande passante réseau et les machines virtuelles [15].
- Élasticité rapide (Rapid elasticity) : Les ressources peuvent être provisionnées et libérées rapidement, parfois automatiquement, pour s'adapter à la demande. Pour le client, les capacités semblent souvent illimitées et disponibles à tout moment, en n'importe quelle quantité [15].
- Service mesuré (Measured service) : Les systèmes infonuagiques mesurent et optimisent automatiquement l'utilisation des ressources à un niveau d'abstraction adapté au type de service (par exemple : stockage, traitement, bande passante ou comptes d'utilisateurs actifs). L'utilisation peut être surveillée, contrôlée et rapportée, assurant ainsi une transparence à la fois pour le fournisseur et le client [15].

### 1.2.3 Évolution du Cloud Computing

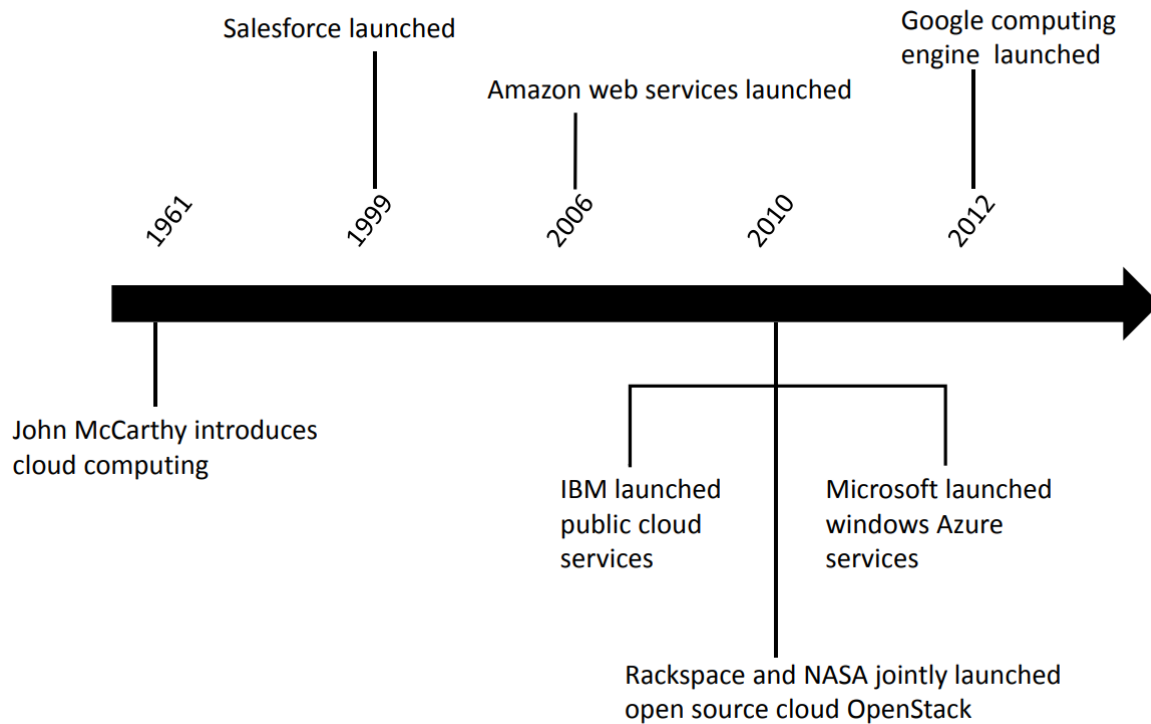


FIGURE 1.1 – Historique du cloud computing  
[39]

Comme l'illustre la figure 1.1, l'évolution du cloud computing peut être retracée à des systèmes plus anciens qui ont été utilisés en temps réel bien avant que l'informatique en nuage n'ait vu le jour. Le concept du cloud computing remonte à 1961 par John McCarthy au MIT : « Si les ordinateurs du genre que j'ai préconisé deviennent les ordinateurs du futur, alors l'informatique pourrait un jour être organisée comme une utilité publique tout comme le système téléphonique est une utilité publique [39].

*Salesforce* est l'une des premières entreprises à avoir exploité le concept de cloud computing à la fin des années 1990. Elle a commencé à fournir son logiciel en tant que service (Software as a Service, SaaS), notamment des outils de gestion de la relation client. Le modèle de Salesforce est un exemple typique du cloud computing, aux côtés du modèle de plateforme en tant que service (Platform as a Service, PaaS), qui fournit un environnement de développement aux clients, comme Microsoft Azure ou Google App Engine. Une autre forme est le modèle Infrastructure as a Service (IaaS), illustré par Amazon Elastic Compute Cloud (EC2), lancé en 2006 [39].

En juillet 2010, la NASA et Rackspace ont lancé un projet commun appelé OpenStack, avec la participation de plusieurs fournisseurs, dont AMD, Intel et Dell. Par la suite, de nombreuses autres organisations ont rejoint le projet. Une organisation à but non lucratif, la OpenStack Foundation, a été créée en septembre 2012 pour promouvoir la plateforme. Aujourd'hui, plus de 500 entreprises soutiennent OpenStack, et environ 6 800 entreprises l'utilisent pour déployer leurs services cloud [39].

En octobre 2011, l'initiative Trusted Cloud de la Cloud Security Alliance (CSA) a publié un livre blanc destiné à aider les fournisseurs de services informatiques à développer des solutions conformes aux normes de l'industrie. Ces services doivent être sécurisés, accessibles, contrôlables, interopérables et gérables [39].

## 1.3 Les modèles de service du Cloud Computing

Il existe trois principaux types de services du cloud computing (voir Figure 1.2), chacun offrant un certain degré de gestion : IaaS (Infrastructure-as-a-Service), PaaS (Platform-as-a-Service) et SaaS (Software-as-a-Service).

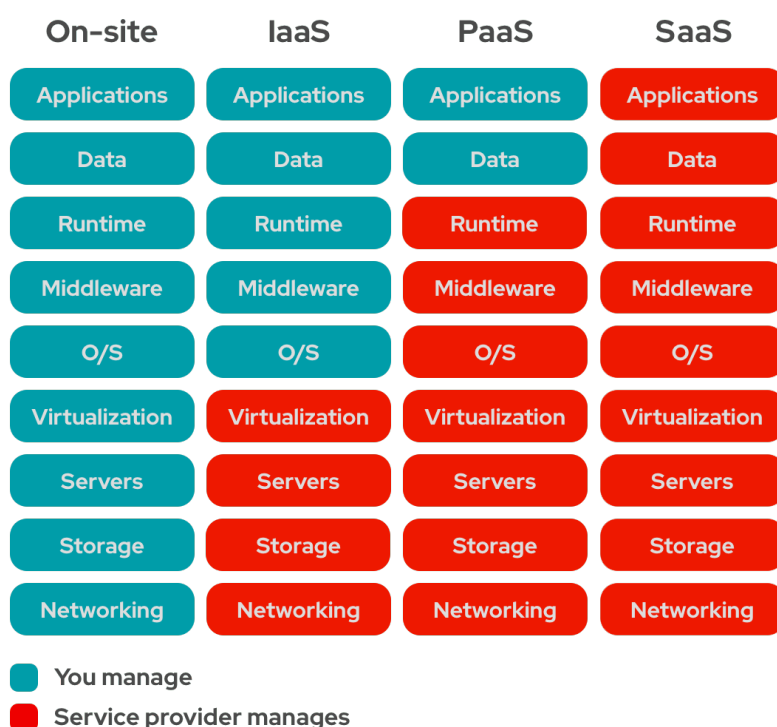


FIGURE 1.2 – Les modèles de service du Cloud Computing [1].

### 1.3.1 Infrastructure as a Service (IaaS)

Ce modèle de service permet la mise à disposition à la demande de ressources d'infrastructure, principalement hébergées dans des centres de données distants. Le modèle IaaS (Infrastructure as a Service) fournit aux administrateurs un accès aux machines virtuelles, au réseau, au stockage et aux configurations système. Le coût est généralement basé sur le principe du paiement à l'usage (pay-as-you-go), proportionnel à la consommation effective des ressources. Une analogie peut être faite avec le mode d'utilisation des industries des commodités (électricité, eau, gaz) ou des Télécommunications [21].

- **Avantages** : grande flexibilité, contrôle total des systèmes (administration à distance) qui permet d'installer tout type de logiciel métier.

- **Inconvénients** : besoin d'administrateurs système comme pour les solutions de serveurs classiques sur site.

Les cibles sont les responsables d'infrastructures informatiques. Amazon EC2 est le principal qui propose ce genre d'infrastructures. Eucalyptus est un exemple d'infrastructure[21].

### 1.3.2 Platform as a Service (PaaS)

Le PaaS permet non seulement de fournir des logiciels en mode SaaS, mais aussi de proposer des environnements de développement complets, incluant les langages, outils et modules nécessaires. L'avantage est que tout est hébergé chez un prestataire externe, ce qui évite à l'entreprise d'avoir sa propre infrastructure ou du personnel de maintenance, et lui permet de se concentrer uniquement sur le développement [21].

- **Avantages** : le déploiement est automatisé, pas de logiciel supplémentaire à acheter ou à installer.
- **Inconvénients** : limitation à une ou deux technologies (ex. : *Python* ou *Java* pour *Google AppEngine*, *.NET* pour *Microsoft Azure*, propriétaire pour *force.com*). Pas de contrôle des machines virtuelles sous-jacentes. Convient uniquement aux applications Web.

Les cibles sont les développeurs. *Google App Engine* est le principal acteur proposant ce genre d'infrastructures [21].

### 1.3.3 Software as a Service (SaaS)

Ce concept consiste à proposer un abonnement à un logiciel, au lieu d'en acheter la licence. Dans ce modèle, aucune application n'est installée sur l'ordinateur : tout se fait via un navigateur Web. L'utilisation est simple et transparente pour les utilisateurs, qui n'ont pas à se soucier de la plateforme ou du matériel, mutualisés entre plusieurs entreprises [21].

- **Avantages** : plus besoin d'installation ni de mise à jour (elles sont effectuées en continu par le fournisseur), plus de migration manuelle des données. Paiement à l'usage, possibilité de tester facilement de nouveaux logiciels.
- **Inconvénients** : limitation au logiciel proposé. Aucun contrôle direct sur le stockage ni sur la sécurisation des données associées.

Ce modèle vise principalement les utilisateurs finaux. Des entreprises comme *Salesforce* ou *Diva* en sont des exemples [21].

Les avantages et les inconvénients des modèles de services du cloud computing sont présentés dans les deux tableaux ci-dessous. La Table 1.1 portant sur le point de vue du client, et l'autre sur celui du fournisseur de services (Voir Table 1.2).

TABLE 1.1 – Les avantages et les inconvénients des modèles de services du Cloud selon le client

Modèle	Avantages	Inconvénients
<b>SaaS</b>	<ul style="list-style-type: none"><li>• Pas d'installation requise</li><li>• Maintenance gérée</li><li>• Accès rapide</li></ul>	<ul style="list-style-type: none"><li>• Sécurité des données</li><li>• Dépendance fournisseur</li><li>• Fonctionnalités limitées</li></ul>
<b>PaaS</b>	<ul style="list-style-type: none"><li>• Déploiement rapide</li><li>• Pas besoin d'infrastructure</li><li>• Environnement homogène</li></ul>	<ul style="list-style-type: none"><li>• Personnalisation limitée</li><li>• Compatibilité restreinte</li><li>• Risque de verrouillage technologique</li></ul>
<b>IaaS</b>	<ul style="list-style-type: none"><li>• Contrôle total</li><li>• Haute flexibilité</li><li>• Évolutivité</li></ul>	<ul style="list-style-type: none"><li>• Coût plus élevé</li><li>• Gestion technique nécessaire</li><li>• Sécurité</li></ul>

TABLE 1.2 – Les avantages et les inconvénients des modèles de services du Cloud selon le fournisseur

Modèle	Avantages	Inconvénients
<b>SaaS</b>	<ul style="list-style-type: none"><li>• Revenus récurrents</li><li>• Maintenance centralisée</li><li>• Facilité de mise à jour</li></ul>	<ul style="list-style-type: none"><li>• Forte exigence en infrastructure</li><li>• Support technique nécessaire</li><li>• Gestion des SLA</li></ul>
<b>PaaS</b>	<ul style="list-style-type: none"><li>• Attraction des développeurs</li><li>• Évolutivité simplifiée</li><li>• Offre clé en main</li></ul>	<ul style="list-style-type: none"><li>• Support multiple requis</li><li>• Coût de maintenance</li><li>• Sécurité et conformité</li></ul>
<b>IaaS</b>	<ul style="list-style-type: none"><li>• Vente d'infrastructure flexible</li><li>• Rentabilité élevée</li><li>• Marché en croissance</li></ul>	<ul style="list-style-type: none"><li>• Gestion technique complexe</li><li>• Risque de saturation des serveurs</li><li>• Coût énergétique</li></ul>

## 1.4 Les modèles de déploiement du Cloud

Le cloud computing peut être déployé sous différents modèles : Privé , public , hybride , communautaire (Voir figure 1.3) :

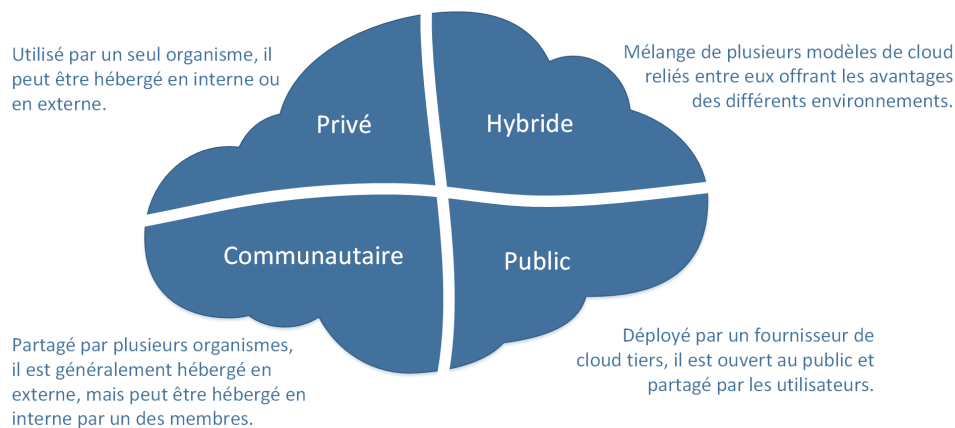


FIGURE 1.3 – Les modèles de déploiement du Cloud

### 1.4.1 Cloud Public

L'infrastructure du cloud est mise à la disposition du grand public. Il peut appartenir, être géré et exploité par une entreprise, un établissement d'enseignement, une organisation gouvernementale, ou une combinaison de ceux-ci. Il existe dans les locaux du fournisseur de cloud.[29]

### 1.4.2 Cloud Privé

L'infrastructure du cloud est mise en service pour un usage exclusif par une seule organisation comprenant plusieurs utilisateurs internes. Elle peut être détenue, gérée et exploitée par l'organisation elle-même, un prestataire tiers, ou une combinaison des deux. Cette infrastructure peut être localisée sur site ou hors site.[29]

### 1.4.3 Cloud Hybride

L'infrastructure du cloud est une composition d'au moins deux infrastructures infonuagiques distinctes (privées, communautaires ou publiques) qui restent des entités uniques, mais sont liées ensemble par une technologie standardisée ou propriétaire qui permet la portabilité des données et des applications .[29]

### 1.4.4 Cloud Communautaire

L'infrastructure du cloud est mise à la disposition exclusive d'une communauté spécifique de consommateurs provenant d'organisations qui ont des préoccupations communes (mission, exigences de sécurité, politique ou considérations de conformité). Elle peut être détenue, gérée et exploitée par une ou plusieurs organisations de la communauté, par une tierce partie, ou par une combinaison des deux. Elle peut être hébergée sur ou hors site.[29]

### 1.4.5 Comparaison des modèles de déploiement

Le tableau suivant présente les principales caractéristiques de chaque modèle de déploiement du cloud, et met en évidence la diversité de leurs approches et de leurs usages.

TABLE 1.3 – Comparaison des modèles de déploiement du cloud computing

Aspect	Public	Privé	Hybride	Communautaire
Contrôle	Limité	Total	Partiel	Partagé
Coût	Faible investissement initial	Investissement initial élevé	Variable	Partagé entre membres
Scalabilité	Très scalable	Moins scalable	Scalable	Dépend de la communauté
Sécurité	Moins sécurisé (partagé)	Très sécurisé	Sécurité ajustable	Sécurité partagée
Personnalisation	Limitée	Hautement personnalisable	Personnalisation partielle	Dépend des besoins communs
Exemples	AWS, Azure, GCP	VMware, OpenStack	Azure Hybrid, AWS Outposts	Clouds sectoriels (santé, finance)

## 1.5 Composants et architecture du Cloud Computing

### 1.5.1 Les composants du cloud computing

Le cloud computing repose sur une infrastructure distribuée qui permet d’offrir des services informatiques à la demande via Internet. Cette architecture s’appuie sur plusieurs composants essentiels qui travaillent ensemble pour assurer la disponibilité, la performance et la scalabilité des ressources. Ces composants sont : les ordinateurs clients, serveurs distribués, et les centres de données.

- **ordinateurs clients**

Ils sont des appareils comme les ordinateurs du bureau, les ordinateurs portables, les tablettes ou les téléphones intelligents que les utilisateurs utilisent pour interagir avec les services en nuage. Ces appareils accèdent au cloud via un navigateur web ou des applications spécialisées, ce qui permet aux utilisateurs de stocker, de récupérer et de gérer les données et les services hébergés dans le cloud. Les ordinateurs clients utilisent Internet pour se connecter à l’infrastructure du cloud, ce qui permet d’accéder aux ressources à distance à tout moment et de n’importe où.

- **Serveurs distribués**

Ils sont situés à des endroits géographiques différents, mais fonctionnent ensemble comme un système unifié. Cette distribution améliore les performances, la fiabilité et la disponibilité en permettant aux utilisateurs d’accéder aux données à partir du serveur le plus proche, ce qui réduit la latence. Même si ces serveurs sont répartis dans différents endroits, ils fonctionnent comme s’ils faisaient partie d’un environnement cloud unique et cohérent. Cette approche assure également une meilleure tolérance aux pannes et un meilleur équilibrage de la charge.

- **Centres de données**

Ils sont de grandes installations qui abritent un ensemble de serveurs, d’équipements de mise en réseau et de systèmes de stockage. Ces centres sont l’épine dorsale de

l'informatique en nuage, fournissant l'infrastructure physique nécessaire pour stocker, traiter et gérer de grandes quantités de données. Les fournisseurs de services infonuagiques gèrent plusieurs centres de données dans le monde entier pour assurer une haute disponibilité, la redondance et la prestation efficace des services. Ils sont conçus avec des systèmes avancés de refroidissement, de gestion de l'alimentation et de sécurité pour assurer la fiabilité et la sécurité des données stockées.

### 1.5.2 L'architecture référentielle du NIST

Le modèle conceptuel de référence du NIST (National Institute of Standards and Technology) constitue un cadre standardisé largement adopté pour décrire et comprendre les éléments clés du cloud computing (voir Figure 1.4). Il définit cinq acteurs principaux : le consommateur de services cloud, le fournisseur de services cloud, le courtier, le transporteur et l'auditeur. Ce modèle permet de clarifier les rôles, les interactions (Voir figure 1.5) et les responsabilités de chacun dans l'écosystème cloud [25].

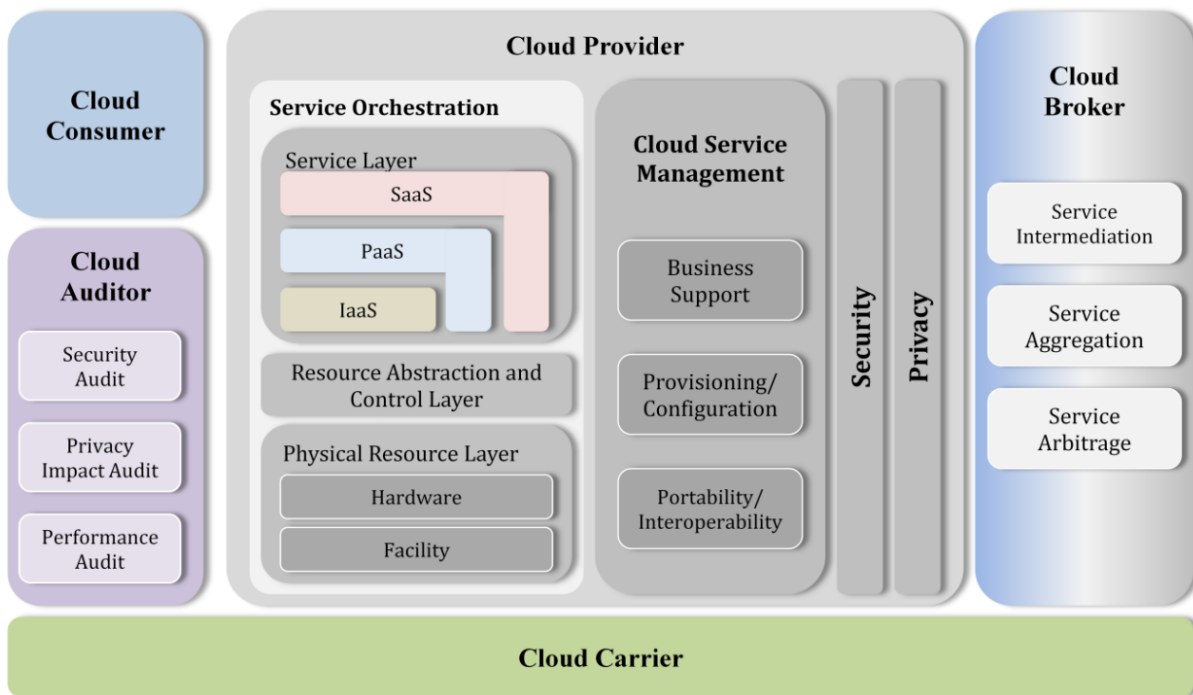


FIGURE 1.4 – Architecture conceptuelle de référence du NIST [25].

- **Cloud Consumer** : Le consommateur de l'infonuagique est la principale partie prenante du service d'infonuagique. Un consommateur de services infonuagiques représente une personne ou une organisation qui entretient une relation d'affaires avec un fournisseur de services infonuagiques et utilise le service. Un consommateur de services en nuage consulte le catalogue de services d'un fournisseur de services en nuage, demande le service approprié, établit des contrats de service avec le fournisseur de services en nuage et utilise le service. Le consommateur du cloud peut être facturé pour le service fourni et doit organiser les paiements en conséquence [25].
- **Cloud Provider** : Un fournisseur de services infonuagiques est une personne, une organisation ; il s'agit de l'entité responsable de la mise à disposition d'un service



aux parties intéressées. Un fournisseur de services infonuagiques acquiert et gère l'infrastructure informatique requise pour fournir les services, exécute le logiciel infonuagique qui fournit les services et prend des dispositions pour fournir les services infonuagiques aux consommateurs infonuagiques par l'accès au réseau [25].

- **Cloud Auditor** : Un vérificateur de services infonuagiques est une partie qui peut effectuer un examen indépendant des contrôles des services infonuagiques dans le but d'exprimer une opinion à leur sujet. Des vérifications sont effectuées pour vérifier la conformité aux normes par l'examen de preuves objectives. Un auditeur de services infonuagiques peut évaluer les services fournis par un fournisseur de services infonuagiques en ce qui concerne les contrôles de sécurité, l'impact sur la vie privée, le rendement, etc [25].
- **Cloud Broker** : Un cloud broker est une entité qui facilite la gestion, la performance et la livraison des services cloud pour les consommateurs en jouant un rôle d'intermédiaire entre eux et les fournisseurs cloud. Il propose trois types de services :
  - Service d'intermédiation** : amélioration des services en ajoutant des fonctionnalités comme la gestion des accès, l'identité, la sécurité et les performances.
  - Service d'agrégation** : combinaison et intégration de plusieurs services pour assurer une interaction sécurisée entre le consommateur et plusieurs fournisseurs.
  - Service d'arbitrage** : sélection flexible des services parmi plusieurs fournisseurs pour offrir les meilleures options disponibles [25].
- **Cloud Carrier** : Un fournisseur de services infonuagiques agit en tant qu'intermédiaire qui fournit la connectivité et le transport des services infonuagiques entre les consommateurs et les fournisseurs. Les fournisseurs de services infonuagiques fournissent un accès aux consommateurs par l'intermédiaire du réseau, des télécommunications et d'autres dispositifs d'accès. Par exemple, les consommateurs de services en nuage peuvent obtenir des services en nuage par l'intermédiaire d'appareils d'accès au réseau, comme des ordinateurs, des ordinateurs portatifs, des téléphones mobiles, des dispositifs Internet mobiles (MID), etc [25].

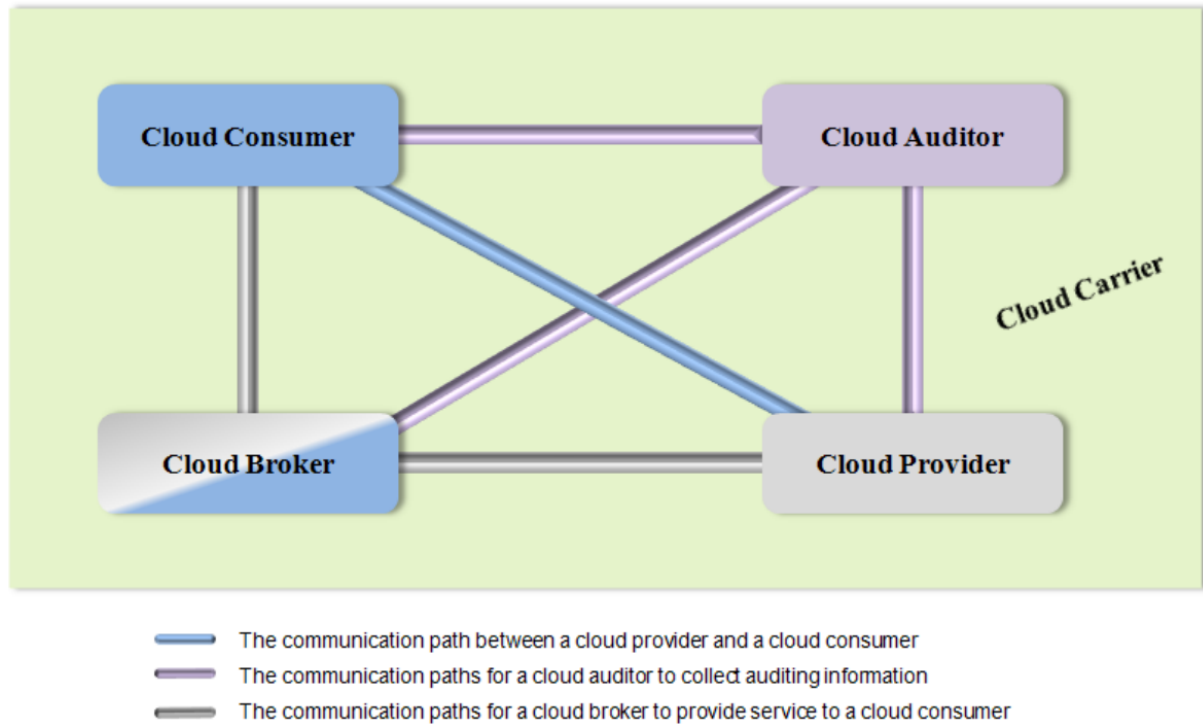


FIGURE 1.5 – Interactions entre les acteurs de l'informatique en nuage  
[25]

## 1.6 Le marché du cloud computing

### 1.6.1 Les fournisseurs majeurs du Cloud Computing

- **Amazon Web Services (AWS)** : Leader incontesté du marché, AWS propose une large gamme de services couvrant le calcul, le stockage, l'intelligence artificielle et les bases de données. Sa flexibilité et son écosystème robuste en font un choix privilégié par de nombreuses entreprises, des startups aux multinationales.
- **Microsoft Azure** : Principal concurrent d'AWS, Azure se distingue par son intégration transparente avec les produits Microsoft tels que Windows Server, Active Directory et Microsoft 365. Son approche hybride, combinant Cloud public et privé, est particulièrement appréciée par les entreprises recherchant une transition progressive vers le Cloud.
- **Google Cloud Platform (GCP)** : Spécialiste des services d'analyse de données et d'intelligence artificielle, GCP se démarque par son expertise en Big Data et en Machine Learning. Il est particulièrement prisé par les entreprises technologiques et les organisations ayant des besoins avancés en traitement de données.
- **IBM Cloud** : Fort de son expérience en solutions d'entreprise, IBM Cloud mise sur l'innovation en intelligence artificielle avec Watson et propose des services de Cloud hybride adaptés aux secteurs nécessitant une sécurité renforcée, tels que la finance et la santé.
- **Alibaba Cloud** : Acteur dominant en Asie, Alibaba Cloud connaît une expansion rapide à l'international. Il est particulièrement apprécié pour ses solutions d'hébergement, de stockage et ses services adaptés aux plateformes de commerce électronique.

Chacun de ces fournisseurs se distingue par des caractéristiques et des avantages spécifiques, permettant aux entreprises de choisir la solution la plus adaptée à leurs exigences en matière de sécurité, de performance, de scalabilité et de coûts. Le choix du prestataire dépend donc des besoins spécifiques de chaque organisation ainsi que des réglementations en vigueur dans son secteur d'activité.

### 1.6.2 Analyse du marché du cloud computing

Le marché mondial des services d'infrastructure cloud est dominé par quelques acteurs majeurs. Les données représentées dans la table 1.4 sont extraites de Statista[38] et représentent la répartition des parts de marché au niveau mondial pour l'année 2024.

TABLE 1.4 – Parts du marché des principaux fournisseurs Cloud (2024)

Fournisseur Cloud	Part de Marché (%)
Amazon Web Services (AWS)	31 %
Microsoft Azure	20 %
Google Cloud	11 %
Alibaba Cloud	5 %
IBM Cloud	4 %
Autres fournisseurs	29 %

Amazon Web Services (AWS) conserve une position dominante avec 31% du marché, suivi par Microsoft Azure (20%) et Google Cloud (11%). Ces trois principaux fournisseurs représentent ensemble 62% du marché mondial.

Alibaba Cloud et IBM Cloud, bien qu'importants sur certains marchés régionaux, détiennent respectivement 5% et 4% du marché global. Les "autres fournisseurs" représentent 29%, ce qui inclut des acteurs comme Oracle Cloud, OVHcloud, et Tencent Cloud.

En conclusion, le marché mondial est fortement concentré autour de quelques entreprises américaines, laissant peu de place aux nouveaux entrants pour concurrencer à grande échelle.

Cette domination se reflète également dans l'adoption des services de cloud computing par les entreprises européennes, comme le montre le graphique ci-dessus. Bien que l'usage du cloud progresse dans la majorité des pays de l'UE entre 2021 et 2023, on constate une forte disparité entre les États membres. Certains pays, comme la Finlande, la Suède ou Malte, affichent des taux d'adoption supérieurs à 65 %, tandis que d'autres, notamment en Europe de l'Est ou du Sud-Est, restent en retrait. Cette évolution montre une dépendance croissante aux services cloud, souvent fournis par les grands acteurs mondiaux, renforçant encore leur position dominante sur le marché.

## 1.7 Avantages et inconvénients du cloud computing

### 1.7.1 Avantages du cloud

*L'informatique en nuage* est une technologie émergente que presque toutes les entreprises adoptent à partir de leurs technologies sur site. Qu'il soit public, privé ou hybride, le cloud computing est devenu un facteur essentiel pour que les entreprises puissent se mesurer à la concurrence[20] :

- **Efficacité des coûts** : La principale raison pour laquelle les entreprises passent à l'informatique en nuage est qu'il faut coût considérablement moindre que toute technologie sur site. Désormais, les entreprises n'ont plus besoin de stocker des données dans disques plus comme le nuage offre l'espace de stockage énorme, économisant l'argent et les ressources.
- **Haute vitesse** : L'informatique en nuage nous permet de déployer le service rapidement en moins de clics. Ce déploiement nous permet d'obtenir les ressources nécessaires pour notre système en quelques minutes.
- **Sauvegarde et restauration des données** : une fois les données stockées dans le cloud, leur récupération devient plus simple et plus rapide, contrairement aux processus souvent longs et complexes des solutions sur site.
- **Facilité de gestion** : L'informatique en nuage élimine le besoin de mises à jour et d'entretien de l'infrastructure puisque le fournisseur de services assure la livraison en temps opportun, garantie et sans interruption et s'occupe également de l'ensemble de la maintenance et de la gestion de nos services informatiques conformément à l'entente sur les niveaux de service.
- **Traitement par lots sporadique** : L'informatique en nuage nous permet d'ajouter ou de retirer des ressources et des services en fonction de nos besoins. Ainsi, si la charge de travail n'est pas continue (24/7), nous n'avons pas à nous soucier du gaspillage de ressources, car nous ne restons pas avec des services inutilisés.
- **Avantage stratégique** : L'informatique en nuage offre à une entreprise un avantage concurrentiel sur ses concurrents, en lui permettant d'accéder rapidement aux applications les plus récentes et essentielles à sa mission, sans avoir à investir du temps et de l'argent dans leur installation[20].

### 1.7.2 Inconvénients du cloud

- **Vulnérabilité aux attaques** : Le stockage des données dans le nuage peut poser de graves problèmes d'information le vol, puisque dans le nuage toutes les données d'une entreprise sont en ligne. La violation de sécurité est quelque chose que même les meilleures organisations ont souffert et c'est un risque potentiel dans le nuage aussi. Bien que des mesures de sécurité avancées sont déployées dans le nuage, tout en conservant les données confidentielles dans le nuage peut être une affaire risquée[20].
- **Dépendance de la connectivité réseau** : L'informatique en nuage dépend entièrement d'Internet. This direct connection to the Internet signifies that an entreprise must have an affordable and cohérent Internet service, as well as a rapid and smooth connection and a bande passante allowing you to take advantage of the benefits of cloud computing.

- **Temps d'arrêt** : Le temps d'arrêt est considéré comme l'un des plus grands inconvénients potentiels de l'utilisation du cloud informatique. Les fournisseurs de cloud peuvent parfois faire face à des pannes techniques qui peuvent se produire en raison diverses raisons, telles que la perte de puissance, une faible connectivité Internet, des centres de données hors service pour la maintenance, etc. Cela peut entraîner une interruption temporaire du service cloud.
- **Verrouillage du fournisseur** : lorsqu'il est nécessaire de migrer d'une plate-forme cloud à une autre, une entreprise peut de faire face à de sérieux défis en raison des différences entre les plateformes. Hébergement et l'exécution des applications de la plateforme infonuagique actuelle sur une autre plate-forme peut entraîner un soutien problèmes, complexité de la configuration et dépenses supplémentaires. Les données de l'entreprise peuvent également être vulnérables aux attaques de sécurité dues à des compromis qui auraient pu être faits pendant les migrations.
- **Contrôle limité** : Les clients du cloud peuvent être confrontés à un contrôle limité sur leurs déploiements. Services cloud s'exécutent sur des serveurs distants qui sont entièrement détenus et gérés par des fournisseurs de services, ce qui il est difficile pour les entreprises d'avoir le niveau de contrôle qu'elles voudraient sur leur back-end infrastructure[20].

## 1.8 Défis du Cloud Computing

L'adoption du Cloud Computing soulève plusieurs enjeux majeurs, parmi lesquels

- **Sécurité et confidentialité des données** : les données étant hébergées à distance, leur protection contre les accès non autorisés, les pertes ou les fuites est cruciale.
- **Disponibilité et fiabilité des services** : les entreprises dépendent du bon fonctionnement continu du cloud, ce qui impose des exigences élevées en matière de disponibilité.
- **Souveraineté des données** : la localisation des centres de données peut poser des problèmes juridiques, notamment en matière de conformité aux réglementations locales.
- **Dépendance au fournisseur (vendor lock-in)** : il peut être difficile de migrer les services d'un fournisseur à un autre, limitant la flexibilité stratégique.
- **Gestion des coûts** : bien que le cloud permette de réduire les investissements initiaux, une mauvaise gestion des ressources peut entraîner des coûts élevés à long terme.
- **Interopérabilité et portabilité** : assurer la compatibilité entre différents services et plateformes cloud reste un défi technique important[17].

## 1.9 Conclusion

Dans ce chapitre, nous avons abordé les concepts de base du Cloud Computing, notamment ses modèles, ses avantages et ses enjeux. Le Cloud offre une solution flexible, évolutive et économique pour l'accès aux ressources informatiques. Son efficacité repose principalement sur la virtualisation et la migration des machines virtuelles, qui assurent une meilleure gestion et allocation des ressources. Le chapitre suivant se penchera sur ces deux technologies clés, indispensables au bon fonctionnement du cloud.

# Chapitre 2

## La virtualisation et la migration de VMs

### 2.1 Introduction

Récemment, la migration des machines virtuelles (VMs) est devenue une opération essentielle dans les environnements cloud. Elle permet de déplacer dynamiquement des VMs d'un serveur à un autre afin d'assurer la continuité de service, l'équilibrage de charge ou la maintenance des systèmes, tout en minimisant les interruptions.

Cette capacité repose sur la virtualisation qui est une technologie sous-jacente, incontournable, et permet de créer des machines virtuelles indépendantes du matériel physique.

Ainsi, nous présentons dans ce chapitre les concepts clés de la virtualisation, puis nous abordons les mécanismes, les techniques et les enjeux liés à la migration des VMs dans le cloud.

Ce chapitre constitue ainsi une base indispensable pour comprendre les mécanismes de virtualisation et de migration qui seront exploités et optimisés dans les chapitres suivants.

### 2.2 La virtualisation

#### 2.2.1 Définition et caractéristiques

C'est une technique permettant de mettre en œuvre une machine virtuelle (VM) sur un ordinateur physique (hôte). Elle repose sur le fait que le jeu d'instructions de l'ordinateur hôte soit identique (ou, plus précisément, en grande partie un sur-ensemble) à celui de la VM, et que le matériel dispose de certaines caractéristiques spécifiques permettant la virtualisation[33]. Ainsi, la virtualisation est une technologie qui permet de faire fonctionner plusieurs systèmes ou applications sur une seule machine physique, en créant des environnements virtuels séparés(Voir figure 2.1).

Cela permet d'utiliser les ressources de manière plus efficace, de tester des logiciels sans risque, ou de faire tourner plusieurs systèmes d'exploitation en même temps.

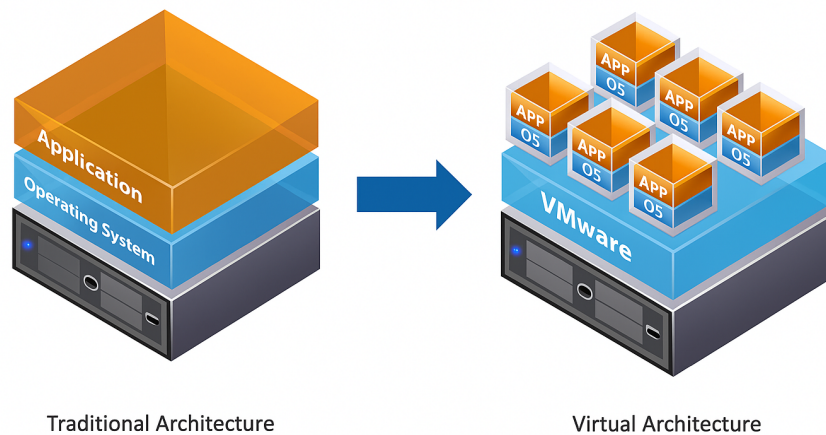


FIGURE 2.1 – Architecture virtualisée vs architecture traditionnelle

La virtualisation offre de nombreuses possibilités dans le domaine informatique, mais comme toute solution technologique, elle présente à la fois des points forts et des limites [33].

### Avantages

- Permet de tester et d'apprendre facilement.
- Moins d'investissement en matériel.
- Nécessite moins de maintenance.
- Amélioration des performances.
- Moindre consommation d'énergie, donc plus écologique.
- Facilite la migration vers le cloud.
- Hébergement facile d'une machine virtuelle invitée.
- Possibilité de générer un grand nombre de serveurs web

### Inconvénients

- Risque élevé de défaillances physiques.
- Nécessite des machines puissantes.
- Point unique de défaillance (single point of failure).
- Performances parfois faibles.
- Certaines applications ne peuvent pas être virtualisées.

Le concept de virtualisation repose sur plusieurs composants fondamentaux qui permettent sa mise en œuvre concrète. Parmi ces éléments, la machine virtuelle et l'hyperviseur jouent un rôle central. Leur interaction constitue le cœur de toute architecture virtualisée, en assurant l'isolation, la gestion et l'exécution des systèmes d'exploitation invités sur le matériel physique.

## 2.2.2 Machine virtuelle (VM Virtual Machine)

### 2.2.2.1 Concept de machine virtuelle

Une machine virtuelle (VM : Virtual Machine) est un environnement logiciel qui émule un ordinateur physique, avec son propre système d'exploitation (OS) et ses applications, fonctionnant sur un matériel réel (hôte) grâce à un hyperviseur (VMM - Virtual Machine Monitor).

De son côté, Goldberg considère une machine virtuelle comme une copie efficace et isolée d'une machine réelle, rendue possible grâce à un moniteur de machine virtuelle (VMM) [5] (voir Figure 2.2) .

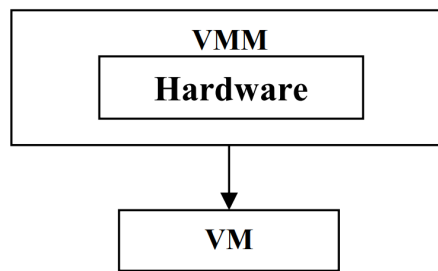


FIGURE 2.2 – La machine virtuelle VM et l'hyperviseur VMM  
[5]

Cependant, Kreuter [5] définit la machine virtuelle comme une couche d'abstraction ou un environnement situé entre les composants matériels et l'utilisateur final. Selon lui, les machines virtuelles permettent d'exécuter des systèmes d'exploitation et sont parfois appelées serveurs virtuels. Un système d'exploitation hôte peut exécuter plusieurs machines virtuelles et partager entre elles des ressources matérielles telles que le processeur, le disque, la mémoire, les contrôleurs et les entrées/sorties.

### 2.2.2.2 Applications des machines virtuelles

La machine virtuelle trouve son utilité dans de nombreux domaines, parmi lesquels [30] :

- Le déploiement et la création d'applications dans des environnements cloud .
- Le test de nouveaux systèmes d'exploitation, y compris les versions bêta .
- La mise en place d'environnements isolés pour faciliter et accélérer les processus de développement et de test par les développeurs .
- La sauvegarde complète d'un système d'exploitation en vue de sa restauration .
- L'exécution d'applications obsolètes ou l'analyse de fichiers potentiellement infectés par des virus, sans compromettre le système principal .
- L'utilisation de logiciels initialement non compatibles avec le système d'exploitation hôte.

## 2.2.3 L'hyperviseur (VMM Virtual Machine Monitor)

### 2.2.3.1 Définition

L'hyperviseur est un logiciel privilégié qui fonctionne soit près ou sous un système d'exploitation (OS Operating system). Il garantit l'indépendance et l'isolation entre les



différents OS invités et de ce fait fournit une meilleure sécurité pour les applications s'exécutant dans les VMs. L'hyperviseur a également pour rôle de partager et de coordonner les accès aux ressources de l'hôte physique sous-jacent entre les OS invités. Pour permettre l'exécution de plusieurs VMs, il fournit une abstraction d'une machine qui inclut un ensemble complet de ressources : processeur, mémoire, carte graphique, carte audio, disque dur, et interface réseau etc.

Les hyperviseurs les plus populaires sont KVM [22], Xen [6], VMware ESXi/ESX [44] et Hyper-V [42]. Il existe deux principaux types d'hyperviseur (Type I ou hyperviseur natif et le type II ou hyperviseur hébergé) [41].

### 2.2.3.2 Hyperviseur de type II

La première technologie de virtualisation est l'hyperviseur de type II. Cet hyperviseur s'installe sur un système d'exploitation comme n'importe quel autre programme. Il se charge de virtualiser les différents périphériques nécessaires aux machines virtuelles, et transmet les demandes de ressources au système d'exploitation hôte. Avec ce type d'hyperviseur, les applications dans les OS invités souffrent d'une baisse non négligeable des performances comparées aux performances sur un système natif. Cela est dû à la présence de deux couches logicielles intermédiaires : l'OS de l'hôte et l'hyperviseur. Comme exemple, nous pouvons citer : VMware Workstation, Oracle Virtualbox et Microsoft Virtual PC[41][11].

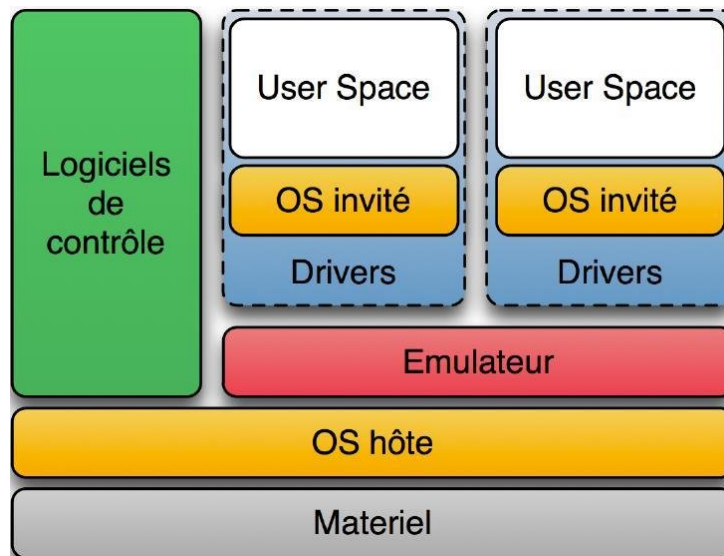


FIGURE 2.3 – Architectures des hyperviseurs de type II [28]

### 2.2.3.3 Hyperviseur de type I

, car il s'exécute directement au-dessus du matériel sous-jacent, comme illustré dans la figure 2. Dans ce cas, le VMM est un code léger dont la responsabilité est de planifier et d'allouer les ressources du système aux machines virtuelles, en l'absence de système d'exploitation en dessous [12]. Des exemples de VMM de type I incluent VMware ESX

et Xen. Ici, le VMM fournit des pilotes de périphériques que les systèmes d'exploitation invités utilisent pour accéder directement au matériel sous-jacent

Il est également connu sous le nom d'« hyperviseur natif » ou « bare metal », car il s'exécute directement sur le matériel comme une couche logicielle entre le matériel et les machines virtuelles. Ainsi, les OS invités s'exécutent directement au-dessus de l'hyperviseur. L'avantage avec cette approche vient de la faible baisse de performance due à la virtualisation, l'hyperviseur se tient entre les machines virtuelles et le matériel. Cependant, ce type d'hyperviseur est difficile à déployer. Comme exemple, nous pouvons citer l'hyperviseur Xen et VMware ESX/ESXi [41][11].

Un hyperviseur de type I est allégé de manière à se « concentrer » sur la gestion des OS invités. Ceci permet de libérer le plus de ressources possible pour les VMs. Toutefois, il est possible d'exécuter uniquement un hyperviseur à la fois sur une machine. Par opposition, l'hyperviseur de type II fournit moins de ressources disponible aux VMs, mais offre l'avantage de pouvoir exécuter plusieurs hyperviseurs simultanément [41].

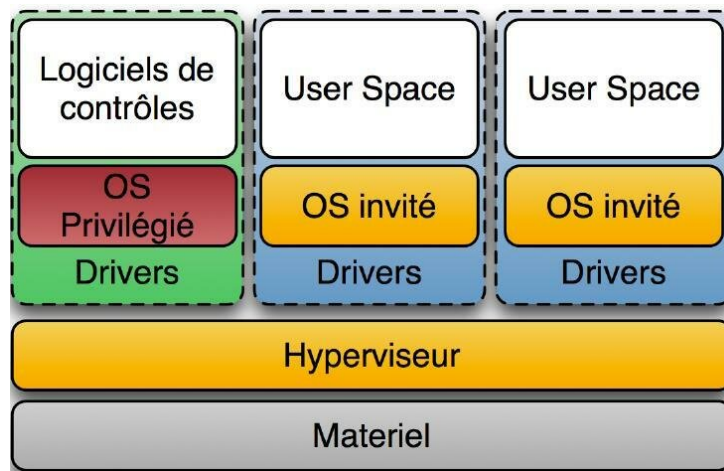


FIGURE 2.4 – Architectures des hyperviseurs de Type I [28]

## 2.2.4 Techniques de virtualisation

### 2.2.4.1 Émulation

L'émulation est une technique permettant de mettre en œuvre une machine virtuelle (VM) sur un ordinateur hôte dont le jeu d'instructions est différent de celui de la VM. Aucune des instructions exécutées par la VM n'est directement prise en charge par le processeur de l'ordinateur hôte ; elles sont toutes traduites par un logiciel de l'hôte, depuis le jeu d'instructions de la VM vers celui de l'hôte, avant d'être exécutées[33]. Les OS invités s'exécutent sans modification sur le matériel virtuel [23].

En règle générale, l'émulation est utilisée pour créer des systèmes d'exploitation nouveaux ou micro-codes pour de nouveaux matériels, avant que le matériel ne soit disponible physiquement [23].

#### 2.2.4.2 Virtualisation assistée par le matériel

Elle désigne l'utilisation d'un hyperviseur pour créer des machines virtuelles qui simulent un matériel complet (processeur, mémoire, périphériques...). Le support de la virtualisation peut être intégré au processeur ou assisté par celui-ci, le matériel se chargeant, par exemple, de virtualiser les accès mémoire ou de protéger le processeur physique des accès les plus bas niveau. Cela permet de simplifier la virtualisation logicielle et de réduire la dégradation de performances.[23]

Ainsi, la virtualisation assistée par matérielle repose sur des fonctionnalités intégrées directement dans le processeur (comme Intel VT-x ou AMD-V) pour aider l'hyperviseur à gérer les machines virtuelles plus efficacement. Ces extensions permettent :

- d'exécuter les systèmes d'exploitation non modifiés,
- de réduire la surcharge en évitant les techniques complexes comme la traduction binaire,
- et de mieux isoler les machines virtuelles.

#### 2.2.4.3 Virtualisation complète

La virtualisation complète (Full virtualization) du système fournit une réplique virtuelle du matériel du système, de sorte que les systèmes d'exploitation et les logiciels peuvent s'exécuter sur le matériel virtuel exactement comme ils le feraient sur le matériel d'origine[32].

Dans la virtualisation complète, le code source du système d'exploitation invité n'est pas modifié. Ainsi, lors de son exécution, le système invité n'a pas conscience d'être virtualisé[9].

La virtualisation complète diffère de l'émulation dans le fait que les systèmes d'exploitation et applications sont conçus pour fonctionner sur la même architecture que celle de la machine physique hôte. Ce qui permet à un système de virtualisation complète d'exécuter plusieurs instructions directement sur le matériel physique sous-jacent.

L'hyperviseur dans ce cas, donne l'accès au matériel physique et donne à chaque système d'exploitation invité l'illusion d'avoir sa propre copie. Il ne doit plus utiliser un logiciel pour simuler une architecture de base différente [23]. L'avantage de la virtualisation complète est que les systèmes d'exploitation invités peuvent s'exécuter sur un VMM sans être modifiés. Cependant, l'inconvénient est que le VMM doit utiliser des techniques particulières pour virtualiser le matériel pour chaque machine virtuelle, ce qui engendre une surcharge supplémentaire lors de l'accès au matériel [9].

#### 2.2.4.4 Para-virtualisation

Dans la paravirtualisation, le système d'exploitation invité requière quelques légères modifications pour fonctionner sur l'architecture émulée par l'hyperviseur, à l'aide d'*hyperappels*<sup>1</sup>. Les modifications effectuées visent à rendre le système émulé du fait qu'il s'exécute dans une machine virtuelle. De ce fait, il pourra collaborer plus étroitement avec le système hôte, en utilisant une interface spécifique et accéder au matériel virtuel via des couches d'abstraction[9]. La paravirtualisation est similaire à la virtualisation complète à la seule

---

1. Est une interface utilisée par un système d'exploitation invité pour demander directement des services à l'hyperviseur, notamment pour exécuter des instructions privilégiées qu'il ne peut pas exécuter lui-même dans un environnement virtualisé.

différence que les systèmes invités sont modifiés et ont conscience qu'il s'exécutent dans un environnement virtuel [23].

#### 2.2.4.5 Conteneurisation (virtualisation légère)

La virtualisation ne se limite pas aux machines virtuelles. Ces dernières années, les conteneurs sont devenus une alternative populaire grâce à leur légèreté et rapidité. Contrairement aux VM qui embarquent un système d'exploitation complet, les conteneurs partagent le noyau du système hôte tout en exécutant des applications de manière isolée. Cette approche réduit la consommation de ressources et accélère le déploiement des services[9].

Selon Pahl [31], ainsi que d'autres auteurs et éditeurs, les conteneurs complètent, mais ne remplacent pas totalement les machines virtuelles, notamment dans des environnements nécessitant une forte isolation ou des fonctionnalités avancées telles que la migration complète de l'état d'une machine .

La conteneurisation d'applications est facilitée par des solutions comme **Docker**, qui permettent de créer, déployer et exécuter des conteneurs de manière isolée. Pour gérer efficacement un grand nombre de conteneurs dans des environnements complexes, des orchestrateurs (ou des systèmes) tels que **Kubernetes** assurent leur coordination, leur répartition et leur montée en charge.

#### 2.2.4.6 Tableau récapitulatif des techniques de virtualisation

Nous récapitulons les techniques de virtualisation dans le tableau suivant qui montre que plus l'isolation est forte, comme dans l'émulation ou la virtualisation complète, plus la performance peut être affectée. À l'inverse, la virtualisation légère offre une excellente performance grâce à sa faible surcharge, mais au prix d'une isolation moindre. Le choix de la technique dépend donc du compromis recherché entre sécurité, performance et flexibilité.

TABLE 2.1 – Tableau récapitulatif des techniques de virtualisation

Type de virtualisation	Principe	Modification du système invité	Isolation	Légèreté	Exemple
Virtualisation complète	Exécute un OS non modifié ; l'hyperviseur intercepte les instructions sensibles.	Non	Forte	Moyenne	VMware Workstation, VirtualBox
Paravirtualisation	OS modifié pour communiquer avec l'hyperviseur via des hyperappels.	Oui	Moyenne à forte	Moyenne	Xen (avec OS modifié)
Émulation	Simulation complète du matériel, y compris un processeur différent.	Non	Forte	Faible	QEMU
Virtualisation matérielle	Le processeur assiste l'hyperviseur grâce à des extensions (VT-x, AMD-V).	Non	Forte	Bonne	KVM, VMware ESXi
Virtualisation légère (conteneurs)	Partage du noyau de l'hôte ; chaque conteneur exécute des processus isolés.	Non	Moyenne	Très bonne	Docker, LXC

## 2.2.5 Les types de virtualisation

### 2.2.5.1 Virtualisation de serveurs

L'objectif de la virtualisation de serveurs est de consolider au sein d'une infrastructure virtualisée plusieurs serveurs physiques. De cette manière, les ressources matérielles (CPU, RAM, disque, réseau) sont mutualisées et partagées en fonction du besoin entre plusieurs serveurs virtuels [35].

### 2.2.5.2 Virtualisation de réseau

C'est un ensemble de technologies qui permettent de penser autrement les infrastructures réseaux dans les entreprises et qui offrent la gestion simple et souple. On peut retrouver plusieurs éléments derrière la virtualisation du réseau, les Vlan et les Vswitch [35].

### 2.2.5.3 Virtualisation de stockage

C'est l'abstraction des ressources, elle consiste à agréger plusieurs formes de stockage différentes et de présenter à l'utilisateur un volume unique.

De cette manière, il est plus simple de stocker des données pour l'utilisateur, surtout plus simple de gérer les environnements par les administrateurs, ils peuvent aussi augmenter la taille du stockage, effectuer des opérations de maintenance, et déplacer des supports sans que cela interfère avec la production. [35]

### 2.2.5.4 Virtualisation de données

La virtualisation des données est une technologie qui permet d'accéder en temps réel à des données provenant de sources diverses (bases de données, fichiers, systèmes cloud, API, etc.) sans nécessiter leur déplacement ou duplication. Ainsi, un utilisateur peut accéder aux mêmes données depuis différents emplacements physiques.

Elle s'appuie sur une couche d'abstraction qui interagit directement avec les sources grâce à des connecteurs spécialisés, réalisant des opérations telles que des jointures ou des filtres, puis présente une vue unifiée aux utilisateurs ou applications. Elle permet de réduire les coûts de stockage et la complexité liée à la maintenance. Cette approche est particulièrement adaptée aux besoins d'analyse en temps réel et à l'intégration rapide de nouvelles sources de données.

### 2.2.5.5 Virtualisation de postes de travail

La virtualisation de postes de travail est une technologie qui consiste à héberger les environnements de bureau des utilisateurs sur des serveurs distants, plutôt que sur les machines locales. L'utilisateur peut ainsi accéder à son système d'exploitation, ses applications et ses données depuis n'importe quel terminal, via une connexion réseau.

L'enjeu de la virtualisation de postes de travail est de repenser la station de travail et de briser la dépendance entre un utilisateur et son matériel [35].

### 2.2.5.6 Virtualisation d'application

Elle se rapproche de la virtualisation de postes de travail ; son objectif est de simplifier le déploiement et la maintenance des applications. On distingue deux types de virtualisation d'applications : la virtualisation par présentation et le streaming d'application [35].

## 2.3 La migration des machines virtuelles

### 2.3.1 Définition de la migration de VMs

La **migration des machines virtuelles (VMs)** fait référence au processus de transfert d'une instance virtuelle d'un hôte physique vers un autre, tout en conservant son état complet (mémoire, CPU, stockage, E/S, etc.) et en minimisant, voire en évitant, l'interruption du service. Cette opération est rendue possible grâce aux hyperviseurs, qui assurent une abstraction entre le matériel physique et les environnements virtuels.[2] La

migration peut être effectuée à *chaud* (*live migration*), c'est-à-dire sans arrêter la machine virtuelle, ou à *froid* lorsque la VM est éteinte pendant le transfert. Dans les environnements cloud et les data centers modernes, la migration à chaud est particulièrement prisee pour assurer une haute disponibilité, répartir dynamiquement les charges de travail, et faciliter les opérations de maintenance sans impacter les utilisateurs finaux.

## 2.3.2 Technique de migration à froid «Stop and Copy Migration»

### 2.3.2.1 Définition de la migration à froid

La migration à froid, également appelée méthode d'arrêt et de copie (Stop and Copy), est une technique dans laquelle la machine virtuelle (VM) cesse complètement de fonctionner sur la machine source. Son exécution, ainsi que l'ensemble de sa mémoire, sont ensuite transférées vers la machine de destination. Durant tout ce processus, la VM reste à l'arrêt jusqu'à ce que toutes les pages mémoire aient été entièrement migrées. Ce n'est qu'à l'issue du transfert que la machine virtuelle est réactivée sur la machine de destination. Dans ce type de migration, le temps total de migration correspond au temps d'arrêt de la VM [12].

Pour décrire les étapes de la migration froid *Stop and Copy*, nous prenons en exemple la migration de la VM3 de la machine 1 vers la machine 2 (voir Figure 2.5) :

1. L'exécution de la VM3 est arrêtée sur la machine source.
2. L'état d'exécution ainsi que les pages mémoire de la VM3 sont transférés vers la machine 2.
3. La VM3 est ensuite activée sur la machine de destination.

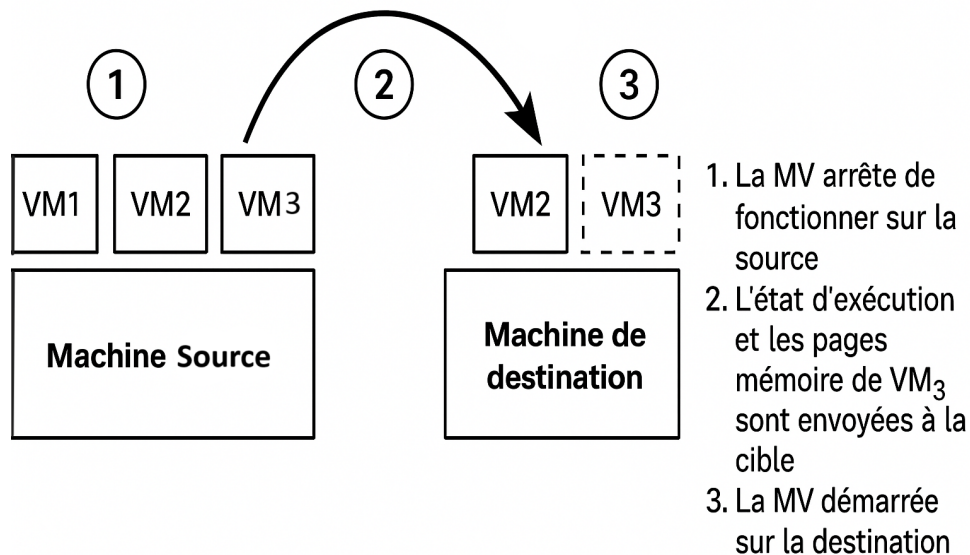


FIGURE 2.5 – Processus de la migration à froid.  
[26]

### 2.3.2.2 Avantages de la migration à froid

- **Simplicité d'implémentation** : La migration à froid est relativement simple à mettre en œuvre car elle ne nécessite pas de gestion complexe de l'état d'exécution de la machine virtuelle.
- **Consommation d'énergie réduite** : Puisque la VM est arrêtée pendant la migration, la charge sur le système source est réduite, ce qui peut diminuer la consommation d'énergie pendant le transfert.
- **Pas d'impact sur la performance de la VM** : La machine virtuelle étant arrêtée, il n'y a pas de consommation de ressources pendant la migration, et donc aucun impact sur les performances de la VM pendant ce processus.
- **Réduction du risque de corruption de données** : La migration se fait avec la VM arrêtée, ce qui élimine le risque de corruption des données, car il n'y a pas de processus en cours pendant le transfert.
- **Adapté aux environnements avec des exigences faibles en termes de disponibilité** : La migration à froid est bien adaptée aux systèmes où une brève période d'indisponibilité n'est pas un problème majeur.

### 2.3.2.3 Inconvénients de la migration à froid

- **Temps d'arrêt de la VM** : La VM doit être complètement arrêtée pendant la migration, ce qui peut entraîner un temps d'arrêt important, particulièrement gênant pour les services nécessitant une haute disponibilité.
- **Interruption du service** : Cette méthode n'est pas adaptée aux applications où l'indisponibilité temporaire peut affecter l'expérience utilisateur ou les services.
- **Impact sur l'expérience utilisateur** : L'arrêt de la machine virtuelle peut entraîner une mauvaise expérience pour les utilisateurs finaux, surtout pour les applications en ligne.
- **Efficacité réduite pour les charges de travail dynamiques** : Dans les environnements très dynamiques où les VMs sont constamment sollicitées, l'arrêt complet de la VM pour la migration peut ne pas être optimal.
- **Nécessité de gestion de l'état** : Le processus implique une gestion complète de l'état de la VM lors du redémarrage pour éviter des erreurs ou des incohérences.

## 2.3.3 Technique de migration à chaud «Live Migration»

### 2.3.3.1 Définition de la migration à chaud

La migration à chaud, également appelée Live Migration, est une technique qui permet de transférer une machine virtuelle (VM) d'un hôte physique à un autre sans interruption significative de service. Contrairement à la migration à froid, la VM continue de fonctionner pendant le processus de transfert[12].

Le mécanisme repose sur la copie progressive de la mémoire de la VM vers la machine de destination tout en maintenant son exécution sur la machine source. Les pages mémoire modifiées pendant le transfert initial sont recopiées en plusieurs passes successives. Une fois que le nombre de pages restantes est très faible, l'exécution de la VM est brièvement suspendue, les dernières pages sont transférées, puis la VM est immédiatement relancée



sur la machine de destination[12].

Cette technique est utile dans plusieurs scénarios, parmi lesquels :

- Une machine virtuelle (VM) hébergée sur un nœud physique défaillant peut être migrée vers un autre nœud fonctionnel afin d'assurer la continuité du service.
- Les VM en mode inactif sur une machine physique peuvent être déplacées vers d'autres hôtes dans le but d'optimiser l'utilisation des ressources disponibles.
- Les VM situées sur un nœud fortement sollicité peuvent être réparties sur plusieurs autres nœuds afin de garantir un meilleur équilibrage de charge.

Les techniques de migration à chaud se répartissent principalement en deux approches, différenciées par la manière dont la mémoire de la machine virtuelle est copiée : Pre-copy et Post-copy (Voir figure 2.7).

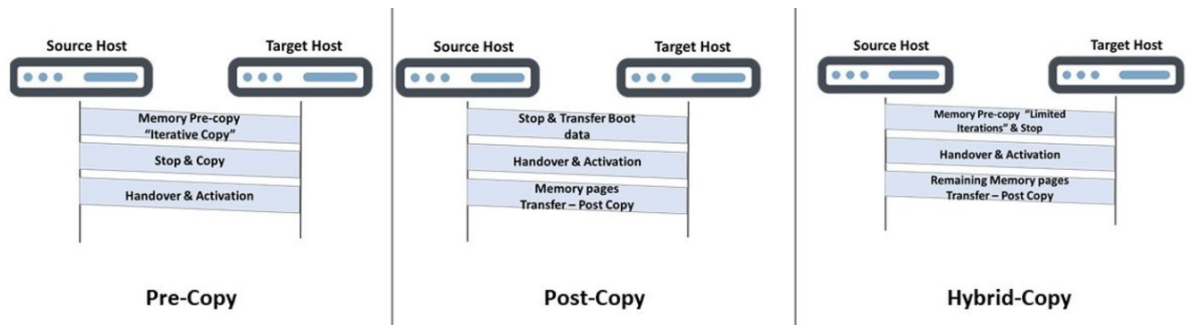


FIGURE 2.6 – Les techniques de migration à chaud [9]

### 2.3.3.2 Migration par pré-copie

La méthode de migration par *pré-copie* consiste à transférer la mémoire vers la machine de destination à travers une série d'itérations. À chaque itération, seules les pages nouvellement modifiées (dites «pages salies») sont transférées. Ce processus se répète jusqu'à ce que l'une des deux conditions soit remplie : soit le nombre maximal d'itérations est atteint, soit la quantité de données restantes est suffisamment faible pour ne pas entraîner de surcharge réseau[4].

- La mémoire et les VCPUs<sup>2</sup> sont réservés sur l'hôte de destination.
- Lors du déclenchement de la migration, un suivi des écritures en mémoire est lancé, et l'intégralité du contenu de la RAM est transférée vers l'hôte de destination. Cela constitue la première itération.
- Dans les itérations suivantes, seules les pages salies depuis la dernière itération sont transférées, jusqu'à atteindre la limite d'itérations ou jusqu'à ce que la quantité de mémoire modifiée soit négligeable.
- L'exécution de la machine virtuelle sur l'hôte source est alors arrêtée. L'état du processeur (CPU), les registres, l'état des périphériques virtuels et les dernières pages mémoire sont transférés à la destination.
- La machine virtuelle est ensuite relancée sur l'hôte de destination.

2. Un vCPU est l'unité de calcul allouée à une machine virtuelle (VM) par l'hyperviseur. Il représente une partie du ou des processeurs physiques (CPU) de la machine hôte.

La technique de pré-copie est bien adaptée aux charges de travail à prédominance en lecture. En revanche, dans le cas de machines virtuelles à forte écriture, le nombre important de pages salies peut provoquer une dégradation des performances [19]. Dans le pire des cas, la migration par pré-copie peut échouer si le taux d'écriture dépasse la bande passante réseau [34].

### 2.3.3.3 Migration par post-copie

Contrairement à la pré-copie, la migration par *post-copie* commence par transférer l'état des VCPUs et des périphériques vers la machine de destination, puis lance l'exécution de la machine virtuelle sur cette dernière [4].

1. La machine virtuelle est arrêtée sur l'hôte source.
2. Les registres du VCPU et l'état des périphériques sont copiés vers la VM de destination.
3. L'exécution est relancée sur la machine de destination.
4. Si la machine virtuelle tente d'accéder à une page mémoire qui n'a pas encore été transférée, une faute de page réseau se produit, entraînant le transfert immédiat de cette page vers l'hôte de destination (*paging à la demande*)[4].

Il s'agit là de la forme la plus simple de post-copie. Cependant, les fautes de page réseau entraînent une baisse de performance et une augmentation du temps total de migration. Plusieurs variantes de cette approche ont été développées pour améliorer les performances et réduire la durée de la migration.

En plus du *paging à la demande*, une copie en masse s'exécute en arrière-plan afin de transférer le maximum de mémoire aussi rapidement que possible. Une autre technique appelée *préchargement adaptatif* (*adaptive prepagging*) permet d'identifier intelligemment les pages susceptibles d'être utilisées prochainement par la VM et de les transférer à l'avance. Cela facilite le transfert rapide de l'ensemble de travail de la machine virtuelle [34].

### 2.3.3.4 Migration hybride

Dans la technique de migration hybride, il y a une combinaison des techniques de migration pré-copie et post-copie. Elle commence d'abord comme une migration en direct de type pré-copie, puis utilise ensuite le principe de la migration post-copie[14].

Un schéma de migration hybride des VM combine les caractéristiques des modèles de migration des VM avant et après la copie pour améliorer les performances de migration des VM. En comparaison avec le modèle de migration post-copie des VM, une méthode hybride exploite les tours pré-copie pour identifier et transférer l'ensemble de travail des VM afin de réduire le taux d'erreur des pages d'E/S. Après avoir terminé les tours de pré-copie, la méthode de migration de VM post-copie capture et transfère l'état minimal de la VM vers le serveur cible pour reprendre la VM. Le processus de migration des VM hybrides est basé sur cinq phases, incluant la préparation, les tours de pré-copie bornés, le transfert d'état des VM, la reprise des VM et la pagination à la demande, comme illustré dans la figure ???. Pendant la phase de préparation, les ressources système requises sont réservées au serveur cible. Les tours de pré-copie délimitées identifient et transfèrent le groupe de travail VM au serveur récepteur. En outre, après des cycles itératifs limités,

l'état minimal de la machine virtuelle est capturé et transféré au serveur récepteur pour reprendre la machine virtuelle sur le serveur cible. La phase de reprise des VMs lance les VMs à partir de l'état transféré sur le serveur récepteur. Dans la phase finale, en fonction des demandes de lecture/écriture de l'application, les pages de mémoire VM sont extraites du serveur source pour synchroniser les deux VMs afin de déconnecter la connexion au serveur source dès que possible.

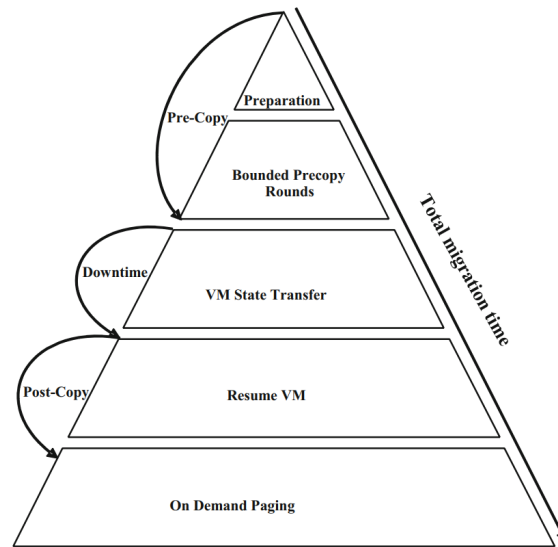


FIGURE 2.7 – Aperçu de la migration (à chaud) hybride des VMs

### 2.3.4 Ressources d'exécution de la migration des machines virtuelles

Après avoir consulté plusieurs articles traitant de la migration des machines virtuelles, nous remarquons que la migration des machines virtuelles (VM) s'appuie sur l'utilisation intensive de ressources systèmes, notamment le processeur (CPU), la mémoire, le réseau et parfois le stockage. L'efficacité de cette opération dépend largement de la disponibilité et de la qualité des ressources mobilisées au moment du transfert. Dans ce contexte, deux approches principales peuvent être adoptées : l'utilisation de ressources partagées ou de ressources dédiées à la migration. Chacune présente des avantages, des limites et des implications différentes en matière de performance, de coût et de qualité de service.

#### 2.3.4.1 Ressources partagées (Shared Resources)

Dans de nombreuses infrastructures virtualisées, la migration s'effectue en utilisant les ressources déjà mobilisées par d'autres charges de travail. Par exemple, le même lien réseau peut servir à la fois à la communication normale des applications et au transfert des données de migration. De même, les cœurs du processeur ou les accès disque peuvent être simultanément sollicités par la machine virtuelle et par le système d'exploitation hôte.

Ce choix, bien qu'économique et simple à mettre en place, présente plusieurs défis. En période de forte activité réseau ou CPU, les performances de la migration peuvent être fortement dégradées, entraînant des temps de transfert prolongés, une saturation des ressources et, dans certains cas, un impact négatif sur les VM en production. Ainsi, même

si cette approche convient à des environnements où les charges sont légères ou prévisibles, elle devient risquée dès lors que la migration s'effectue dans un contexte de production intensif ou en situation critique.

#### 2.3.4.2 Ressources dédiées (Dedicated Resources)

Face à ces limitations, certaines organisations choisissent de réserver des ressources exclusivement dédiées à la migration des VM. Cela peut se traduire par l'allocation d'une interface réseau spécifique, d'un canal de communication isolé, voire d'un nœud de calcul secondaire temporairement affecté à cette opération. Cette stratégie vise à assurer une isolation complète entre les flux de migration et les autres activités du système.

L'intérêt majeur de cette approche réside dans la prévisibilité des performances : la bande passante est disponible sans contention, le processeur n'est pas surchargé et les temps de migration sont mieux maîtrisés. En particulier, dans le cas des migrations à chaud, cette séparation permet de minimiser la fenêtre de bascule (*downtime*), un facteur clé pour maintenir la qualité de service perçue par l'utilisateur final.

Toutefois, cette performance a un coût. Mettre en place une infrastructure dédiée nécessite des investissements matériels supplémentaires, une complexité accrue dans la gestion des ressources, et une planification plus rigoureuse. Elle peut donc être difficile à justifier dans des environnements à faible budget ou à petite échelle.

#### 2.3.4.3 Vers une approche équilibrée

Dans la pratique, de nombreuses solutions hybrides émergent, combinant les deux stratégies selon les besoins et les contraintes. Par exemple, certaines plateformes de virtualisation permettent d'activer dynamiquement des canaux dédiés temporairement lors des pics de migration, tout en utilisant des ressources partagées le reste du temps. Ce compromis vise à concilier efficacité économique et exigences de performance, en s'appuyant souvent sur des algorithmes de planification intelligents ou des techniques d'analyse prédictive.

En définitive, le choix entre ressources partagées ou dédiées ne relève pas uniquement d'une logique technique. Il traduit également une vision stratégique de la gouvernance des infrastructures IT, où performance, résilience et maîtrise des coûts doivent en permanence être équilibrés.

## 2.4 Conclusion

Ce chapitre a présenté les concepts fondamentaux liés à la virtualisation, en mettant l'accent sur le rôle des machines virtuelles dans la modernisation des infrastructures informatiques. Nous avons également abordé les principes liés à la migration des VMs.

Dans le chapitre suivant, nous allons nous intéresser aux différentes approches de migration des machines virtuelles, en mettant en avant les critères de consommation

énergétique et de qualité de service.

# Chapitre 3

## État de l’art sur la migration des VMs dans le cloud computing

### 3.1 Introduction

La migration des machines virtuelles (VMs) est l’un des mécanismes les plus puissants et les plus utilisés dans les environnements virtualisés modernes, en particulier dans le contexte du *cloud computing*.

Ainsi, au fil du temps, plusieurs approches de migration ont vu le jour, chacune répondant à des objectifs spécifiques : équilibrage de la charge, gestion d’énergie, maintenance des serveurs, et tolérance aux fautes (pannes).

Dans ce chapitre, nous allons principalement nous focaliser sur l’approche énergétique, qui constitue aujourd’hui un axe de recherche prioritaire face aux enjeux mondiaux de réduction de l’empreinte carbone des infrastructures numériques.

Dans cette optique, nous proposons une taxonomie globale de la migration de VMs dans le cloud, structurée autour de trois axes majeurs : le problème, la résolution, et l’évaluation. Nous présentons ensuite un ensemble de travaux ayant pour objectif l’optimisation énergétique. Le chapitre se conclut par un tableau comparatif des approches étudiées, mettant en évidence les critères issus de la taxonomie proposée.

### 3.2 Mécanisme de migration des machines virtuelles

Pour initier le mécanisme de migration, trois questions fondamentales doivent être posées (Voir figure :3.1) : *quand* effectuer la migration, *quelles* machines virtuelles doivent être migrées, et *vers quel* hôte les transférer. Ainsi, le processus complet de migration de VMs comprend généralement quatre étapes principales, décrites ci-après.

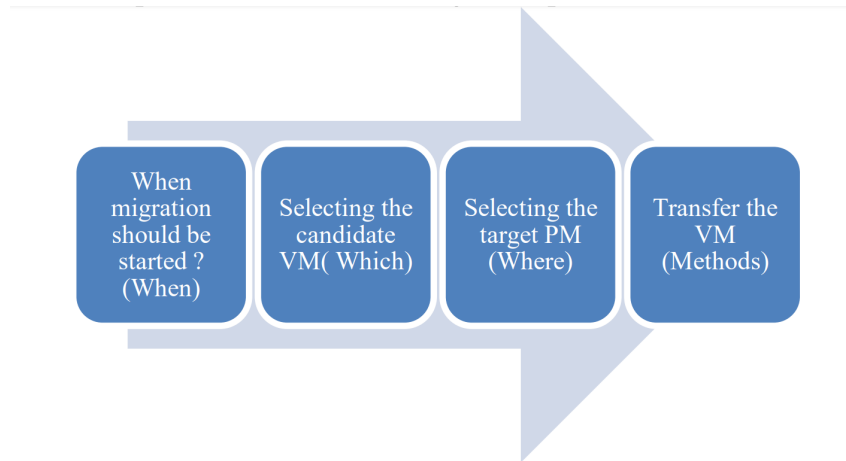


FIGURE 3.1 – Le processus de migration des machines virtuelles

### 3.2.1 Quand doit-on démarrer le processus de migration ?

En général, des seuils inférieur et supérieur sont utilisés pour déterminer quand lancer une migration de machines virtuelles (VM). Un système est considéré comme surchargé lorsque sa charge dépasse le seuil supérieur (dite  $T_{upper}$ ), et comme sous-utilisé lorsque sa charge est en dessous du seuil inférieur (dite  $T_{lower}$ ).

La migration des VM est généralement déclenchée dans trois situations principales :

1. **Équilibrage de charge** : des VM sont transférées depuis des machines physiques surchargées vers d'autres moins utilisées pour répartir la charge.
2. **Consolidation des serveurs** : toutes les VM hébergées sur une machine sous-utilisée sont déplacées vers d'autres hôtes. Cela permet d'éteindre la machine vide, augmentant ainsi l'efficacité énergétique.
3. **Atténuation des points chauds** : si une machine physique n'a pas suffisamment de ressources pour répondre aux besoins de ses VM, une ou plusieurs d'entre elles sont migrées vers d'autres hôtes.

### 3.2.2 Comment choisir la machine virtuelle à migrer ?

Le choix de la VM à migrer dépend du contexte : équilibrage de charge, consolidation des serveurs ou atténuation des points chauds.

- Dans le cas de la consolidation des serveurs, toutes les VM hébergées sur une machine peu utilisée doivent être déplacées.
- Dans les deux autres cas, le choix est plus complexe. Si on choisit une VM légère, plusieurs migrations peuvent être nécessaires pour libérer suffisamment de ressources. Si on choisit une VM plus lourde, cela peut allonger le temps de migration.

Différentes approches ont été proposées pour identifier la VM à migrer :

- selon [45] les auteurs proposent de choisir la VM ayant le meilleur compromis entre sa charge et sa taille mémoire.
- selon [36] les auteurs suggèrent de migrer la VM ayant la plus faible utilisation CPU, mais cela peut impliquer de nombreuses migrations.

- selon[8] les auteurs recommandent de choisir une VM dont la taille est suffisante pour réduire la surcharge de l'hôte en un seul déplacement. Si aucune VM ne répond à ce critère, alors plusieurs devront être migrées.

### 3.2.3 Où migrer la machine virtuelle ?

Le choix de la machine physique de destination, c'est-à-dire où migrer la VM, dépend de plusieurs facteurs, principalement la disponibilité des ressources (processeur, mémoire, bande passante, etc.) sur les autres hôtes du centre de données. L'objectif est de trouver une machine capable de recevoir la VM sans devenir elle-même surchargée. Plusieurs critères peuvent être pris en compte, tels que :

- La capacité restante en ressources,
- La consommation énergétique de la machine,
- La distance de communication entre la VM et ses ressources,
- La stabilité ou prévisibilité de la charge sur cette machine

Des algorithmes d'optimisation peuvent être utilisés pour sélectionner l'hôte le plus approprié. L'idée est d'assurer un bon équilibre entre performance et efficacité énergétique, tout en minimisant les risques de surcharge après la migration.

## 3.3 La taxonomie proposée

La migration de machines virtuelles (VMs) est un sujet largement abordé dans la littérature. Plusieurs classifications ont été proposées. À ce titre, Kapil et al. [10] ont traité, dans leur revue, *l'évaluation* des différentes techniques de migration de VMs. Ils ont identifié plusieurs indicateurs de performance utilisés à cette fin, notamment : le temps de préparation, le temps d'arrêt, le temps de reprise, le nombre de pages transférées, le temps total de migration et la dégradation des performances applicatives.

Par ailleurs, Ahmed et al. [3] ont classé les mécanismes de migration de VMs selon leurs *objectifs* (consolidation, optimisation de la bande passante, etc.). Plus précisément, ils ont catégorisé les travaux de consolidation de serveurs selon cinq caractéristiques communes : la politique d'allocation des ressources, l'architecture, les critères de co-localisation, le point de déclenchement de la migration et le modèle de migration.

Dans un autre travail, Ahmed et al. [3] ont proposé une taxonomie globale fournissant un cadre structuré pour comparer, analyser et guider les choix techniques en matière de migration dans les centres de données cloud. Cette taxonomie organise *les techniques de migration de VMs* selon plusieurs dimensions clés, telles que : le type de migration (pré-copie, post-copie, hybride, non-live), la fonction objective, la gestion de la bande passante, la distance du réseau (locale ou inter-centres), la granularité (page mémoire, VM entière, groupes de VMs), ainsi que le temps d'interruption (downtime).

De plus, Choudhary et al. [10] proposent une classification des approches de migration des VMs en trois catégories : selon le type de migration, la duplication, et le contexte (pages peu actives, pages modifiées, pages sensibles aux dépendances).



Enfin , L'auteur Le [24] présente une revue structurée des techniques de *migration à chaud* des machines virtuelles, en mettant en évidence leurs mécanismes, avantages, limites et cas d'usage.

D'autre part, nous proposons une taxonomie globale et compréhensive qui distingue trois axes principaux de la migration des VMs : selon les caractéristiques du problème étudié, sa résolution et son évaluation, où chaque axe correspond à un ensemble de critères spécifiques.

## Organigramme de classification

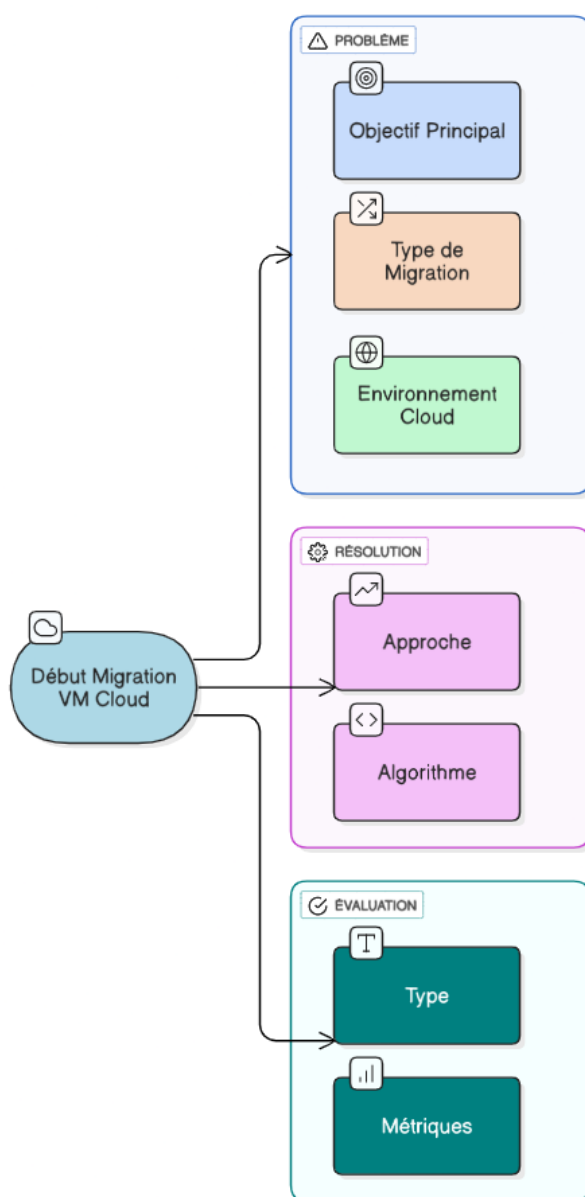


FIGURE 3.2 – Taxonomie des techniques de migration de machines virtuelles dans un environnement cloud

### 3.3.1 Classification selon les caractéristiques du problème Étudié

Cet axe concerne les caractéristiques du problème de migration de machines virtuelles à prendre en compte pour pouvoir le résoudre. On y distingue trois éléments principaux : le type de migration à traiter, l'objectif principal qui motive cette migration, et l'environnement cloud considéré.

1. **L'objectif principal** : la migration est mise en œuvre afin de réaliser l'un des objectifs suivants [3] :
  - *Tolérance aux pannes* : Un système tolérant aux pannes déclenche la migration de VM avant qu'une panne ne survienne. Si nécessaire, il remigre la VM vers son serveur d'origine après l'achèvement de la maintenance du système. Ce type de système améliore considérablement la disponibilité du système, renforçant ainsi la fiabilité du centre de données cloud.
  - *Maintenance du système* : La mise en œuvre d'une maintenance périodique ou dynamique permet de prolonger la durée de vie du système. La technologie de migration de VM permet de transférer les applications en cours d'exécution vers un autre hôte, assurant ainsi la continuité du service pendant la maintenance du système.
  - *Équilibrage de la charge* : L'équilibrage de la charge permet à l'opérateur cloud d'éviter les points de défaillance uniques en répartissant la charge de travail entre plusieurs hôtes physiques au sein d'un centre de données. Une charge de travail excessive sur un serveur dégrade les performances du système ; ainsi, les approches d'équilibrage de charge (basées sur la migration de VM) réduisent le risque de dégradation des performances applicatives en éliminant les points chauds dans le centre de données.
  - *Gestion de l'énergie* : Pour atteindre une efficacité énergétique au sein d'un centre de données cloud, le processus de migration de machines virtuelles (VM) transfère l'ensemble de la charge d'un serveur faiblement utilisé (par exemple lorsque l'utilisation des ressources est inférieure à un certain seuil) vers un autre serveur sous-utilisé, afin de pouvoir éteindre le premier. Les méthodes de migration de VM basées sur la consolidation de serveurs et le DVFS (Dynamic Voltage and Frequency Scaling) visent à améliorer l'efficacité énergétique du centre de données en regroupant les VMs sur un même hôte ou en réduisant la fréquence du processeur.
  - etc.
2. **Le type de migration** : migration à chaud (*Live*) ou à froid (*Non Live*) (voir les détails dans le chapitre précédent).
3. **L'environnement du cloud considéré** : y compris :
  - **Unité de migration** : une seule VM (*Single VM*) est migrée à la fois, ou plusieurs (*Cluster VM*).
  - **Domaine** : migration au sein du même cloud (*In the same cloud*), ou bien entre plusieurs clouds (*Inter-Clouds*)

### 3.3.2 Classification selon les méthodes de résolution

La résolution du problème de migration consiste à réaliser les étapes suivantes : la détection des hôtes surchargés ou sous-chargés, la sélection des machines virtuelles à migrer,

puis la décision de placement sur les hôtes cibles. Pour cela, différentes approches sont appliquées, chacune adoptant des algorithmes spécifiques pour la sélection et le placement des VMs.

1. **Approche de résolution** : des approches basées sur l'optimisation, l'apprentissage automatique, les heuristiques et les méthodes à base de seuils, qui correspondent à notre cadre d'étude. Nous avons également identifié deux paramètres caractérisant ces approches à base de seuil :
  - (a) **Le type de seuil** : la migration peut se faire sur la base de :
    - **Seuil unique** (*Single threshold*) : généralement un seuil supérieur (noté  $T_{upper}$ ), qui peut être statique ou dynamique.
    - **Seuils doubles** (*Double threshold*) : utilisation de deux seuils, un seuil supérieur ( $T_{upper}$ ) et un seuil inférieur ( $T_{lower}$ ), qui peuvent être statiques, dynamiques, ou hybrides.
  - (b) **La politique adoptée** : elle peut être :
    - **Statique** : la valeur du seuil déclenchant la migration reste *fixe*, quel que soit l'état ou les paramètres de l'environnement.
    - **Dynamique** : le seuil est *ajusté dynamiquement* en fonction de la charge moyenne, de la variance ou d'autres métriques (historique, tendance, etc.).
    - **Hybride** : concerne les approches à base de seuils doubles, où l'un des seuils est statique et l'autre dynamique.
2. **Algorithme de résolution** : ce sont les algorithmes utilisés pour les deux étapes de sélection et de placement, parmi lesquels :
  - *BFD Best Fit Decreasing* : permet de sélectionner les VMs par ordre décroissant de consommation de CPU, de mémoire (RAM) ou de stockage. Nous avons utilisé cet algorithme dans notre travail, en l'appliquant à la fois dans la sélection et le placement des VMs, sur la base de leur consommation CPU, dans le but de réduire la consommation énergétique globale.
  - *FFD First Fit Decreasing*,
  - *MFD Modified Fit Decreasing*,
  - *PAFD Power Aware Fit Decreasing* :
  - *PAFD (Power Aware Fit Decreasing)*,
  - *Round Robin* : il répartit les VMs de manière cyclique entre les hôtes disponibles, sans tenir compte de leur charge ou de leurs ressources.
  - etc.

### 3.3.3 Classification selon les techniques d'évaluation

Cet axe porte sur l'évaluation des performances liées à la migration des machines virtuelles (VMs), en s'appuyant sur deux critères principaux : le type d'évaluation et les métriques considérées.

1. **Type d'évaluation** : certains travaux analysent les performances au **cours** du processus de migration (During/While Migration), tandis que d'autres se focalisent sur l'évaluation **post**-migration (After Migration), ce qui constitue le cadre de notre étude.
2. **Métriques évaluées** : consommation énergétique, performances globales, respect des accords de niveau de service (SLA), entre autres.

## 3.4 Étude bibliographique

### 3.4.1 Energy-aware resource allocation heuristics for efficient management of data centers for Cloud computing

Dans cet article de référence, Beloglazov et al. [7] proposent plusieurs heuristiques visant à améliorer *la gestion énergétique des centres de données* dans les environnements cloud. Face à la consommation énergétique massive des datacenters, les auteurs soulignent l'urgence d'adopter des solutions de Green Cloud Computing, afin de réduire à la fois les coûts opérationnels et l'impact environnemental. Leur objectif principal est double : réduire significativement la consommation énergétique des centres de données, tout en garantissant le respect des accords de niveau de service (SLA) pour maintenir la qualité de service.

Pour cela, les auteurs s'appuient sur des *stratégies de consolidation* dynamique des machines virtuelles (VMs), en utilisant la migration à chaud (*live migration*) afin d'optimiser l'allocation des ressources. Leur méthodologie repose sur un modèle mathématique simple mais efficace, permettant d'estimer la consommation énergétique d'un serveur en fonction de son taux d'utilisation CPU. Ce modèle tient compte du fait qu'un serveur consomme une part significative de sa puissance maximale même lorsqu'il est inactif.

Ensuite, les auteurs développent plusieurs algorithmes d'allocation et de consolidation, notamment le *Modified Best Fit Decreasing* (MBFD), qui trie les machines virtuelles (VMs) selon leur consommation CPU, puis les place de manière à minimiser l'augmentation de la consommation énergétique. Ils proposent également trois politiques de sélection des VMs à migrer, dont la *Minimization of Migrations* (MM), qui vise à ne migrer que le nombre minimal de VMs nécessaire pour ramener un hôte en surcharge en dessous d'un seuil d'utilisation prédéfini.

Par ailleurs, les auteurs mettent en œuvre une politique de *seuils doubles* (double-threshold policy), selon laquelle chaque hôte doit maintenir son taux d'utilisation CPU entre un seuil inférieur et un seuil supérieur. Si l'utilisation est trop faible, les VMs sont migrées vers d'autres hôtes et le serveur est mis en veille afin d'économiser de l'énergie. En revanche, en cas de dépassement du seuil supérieur, certaines VMs sont migrées pour réduire la charge et éviter la saturation de l'hôte.

Les expérimentations réalisées démontrent la grande efficacité de la politique MM qui a permis de réduire la *consommation énergétique*, les *violations de SLA* et le *nombre de migrations*, limitant ainsi la surcharge réseau. Cependant, malgré ces avantages, l'étude présente certaines limites, notamment le fait que la politique de *seuils reste fixe* et non adaptative en fonction de la charge.

### 3.4.2 Energy Efficient Dynamic Integration of Thresholds for Migration at Cloud Data Centers

Dans le même cadre, R. Sinha et al. [36] ont traité le même objectif de migration de VMs, à savoir la gestion de la consommation énergétique. Ils ont proposé une stratégie de migration *dynamique* des machines virtuelles basée sur une politique de *seuils adaptatifs* (DyT) qui s'ajustent en fonction de la charge réelle des hôtes. Cette méthode repose sur

*deux seuils dynamiques :*

- Un seuil supérieur, propre à chaque hôte, déclenchant la migration partielle des VMs en cas de surcharge ;
- Un seuil inférieur, conduisant à la migration complète des VMs et à l'extinction de l'hôte pour économiser l'énergie.

Les seuils sont calculés à partir de l'utilisation du CPU, de la mémoire et de la bande passante, et sont associés à des algorithmes intelligents de sélection et de placement visant à *réduire le nombre de migrations* et à limiter leur impact sur la performance.

L'approche a été comparée à plusieurs politiques : DVFS (*Dynamic Voltage and Frequency Scaling* : ajuste dynamiquement la fréquence et la tension des processeurs selon la charge), NPA (*Non Power Aware* : sans optimisation énergétique), seuil statique (ST), et double seuil fixe (DoT). Les résultats montrent que l'approche proposée DyT réduit significativement *les violations de SLA* et *le nombre de migrations*, tout en maintenant une consommation énergétique comparable à celle du DVFS.

En conclusion, cette recherche met en avant une approche de consolidation dynamique, adaptée à la charge réelle des hôtes, permettant d'améliorer à la fois l'efficacité énergétique et la qualité de service. Toutefois, elle présente certaines limites, notamment l'absence de prise en compte de paramètres tels que le temps de migration et l'étude de la stratégie hybride basée sur des seuils doubles.

### 3.4.3 Energy Conscious Dynamic Provisioning of Virtual Machines using Adaptive Migration Thresholds in Cloud Data Center

Ce travail [27] aborde la problématique cruciale de *la consommation énergétique* dans les centres de données cloud, un enjeu majeur en raison des coûts élevés d'exploitation et des impacts environnementaux liés aux émissions de  $CO_2$ . Les auteurs [27] proposent une approche dynamique et économe en énergie basée sur la *consolidation* des machines virtuelles (VMs) en utilisant des seuils de migration *adaptatifs*. Cette méthode vise à *optimiser l'utilisation des ressources* physiques tout en maintenant un bon niveau de performance et en *réduisant les violations des accords de niveau de service (SLA)*.

Deux types de seuils sont introduits : *un seuil supérieur ( $T_{upper}$ )* qui déclenche la migration lorsqu'une surcharge est détectée en tenant compte de plusieurs ressources (CPU, RAM, bande passante), et *un seuil inférieur ( $T_{lower}$ )* qui identifie les serveurs sous-utilisés pour éteindre ceux qui n'hébergent plus de VMs.

La politique de migration minimale (*Minimum Migration Time - MMT*) permet de choisir les VMs à migrer en fonction du *temps de transfert le plus court*, réduisant ainsi les perturbations et la consommation d'énergie.

Les résultats de simulation montrent une réduction allant jusqu'à 50 % de la consommation énergétique, une diminution significative du nombre de migrations, ainsi qu'une meilleure stabilité et conformité aux SLA. Bien que cette méthode implique une certaine

complexité liée à la surveillance continue et aux calculs en temps réel, elle présente une grande flexibilité et s'adapte efficacement aux variations imprévisibles des charges de travail. Cette étude constitue donc une référence pertinente pour un mémoire portant sur l'optimisation énergétique, la gestion dynamique des VMs, ou encore l'équilibre entre performance, consommation énergétique et qualité de service dans les environnements cloud.

#### 3.4.4 BGM-BLA : a new algorithm for dynamic migration of virtual machines in cloud computing

Dans cette étude, Tao et al. [40] proposent une nouvelle approche pour optimiser la migration *dynamique (live)* des machines virtuelles (VM) dans les environnements de cloud computing. Ils introduisent une méthode BGM-BLA (Binary Graph Matching-based Bucket-code Learning Algorithm), conçue pour répondre à un modèle d'optimisation *multi-objectifs*. Ce modèle vise simultanément à *réduire la consommation énergétique des centres de données, minimiser les coûts liés à la migration des VM, et améliorer l'efficacité des communications entre les VM*.

Pour résoudre ce problème complexe, les auteurs le décomposent en deux phases : d'abord, le regroupement des VM fortement interconnectées, ensuite leur placement optimal sur les nœuds physiques. La force de BGM-BLA réside dans la combinaison de techniques avancées, notamment un mécanisme de codage compact (Bucket-Coding) pour restreindre l'espace de recherche, la correspondance de graphes binaires pour évaluer les solutions possibles, et des méthodes d'apprentissage automatique pour guider la sélection des migrations les plus pertinentes.

En réalité, l'algorithme BGM-BLA n'utilise aucun seuil (ni fixe ni dynamique) et ne réalise pas de consolidation explicite (absence d'extinction des hôtes). Il effectue une migration groupée de VMs (cluster de VMs) via une optimisation globale multi-objectif.

Les expérimentations menées montrent que BGM-BLA surpasse les algorithmes existants, tels que NSGA-II (Non-dominated Sorting Genetic Algorithm-II), aussi bien en termes de qualité des solutions de Pareto que de rapidité d'exécution, en particulier sur de grands jeux de données.

Parmi les contributions majeures de cette recherche figurent la proposition d'un cadre de résolution multi-objectif complet, l'introduction d'un schéma de codage efficace, ainsi qu'une approche hybride combinant méthodes déterministes et stochastiques.

#### 3.4.5 EAMA : Efficient Adaptive Migration Algorithm for Cloud Data Centers (CDCs)

Dans ce travail, les auteurs [16] ont abordé de *la consommation énergétique* élevée des centres de données cloud, qui engendre des coûts d'exploitation importants ainsi qu'un impact environnemental négatif dû aux émissions de  $CO_2$ . Face à ces enjeux, ils ont proposé l'algorithme EAMA, conçu pour optimiser l'utilisation des ressources tout en minimisant la consommation d'énergie grâce à l'extinction des machines physiques sous-utilisées (approche *avec consolidation*).

L'algorithme fonctionne par phases successives : il détecte d'abord les hôtes surchargés et sous-utilisés (*seuils doubles*) en analysant leur utilisation CPU, puis migre les VMs des hôtes surchargés vers ceux sous-utilisés, en tenant compte de la proximité géographique pour réduire la latence. Une phase finale vise à augmenter le nombre de serveurs éteints afin de maximiser les économies d'énergie.

Les résultats expérimentaux ont prouvé qu'EAMA réduit significativement *le nombre de migrations et les violations de SLA* par rapport à deux algorithmes concurrents (PACPA : *Predictive Anti-Correlated Placement Algorithm* et RUAAEE : *Resource Utilization-Aware Energy Efficient*), tout en améliorant l'utilisation des ressources et en diminuant la consommation énergétique.

Bien qu'encore limitée par l'usage de *seuils fixes*, cette approche contribue efficacement à concilier performance, économies d'énergie et respect des SLA dans les centres de données cloud.

### 3.4.6 A modelling and Simulation via CloudSim for Live Migration in Virtual Machines

Cette étude s'intéresse à l'optimisation de la migration *à chaud* des machines virtuelles (VMs) dans les environnements de cloud computing. Dans ce cadre, l'auteur Gupta [13] a proposé un modèle qui inclut un mécanisme de gestion proactive de la charge et introduit un nœud de suivi (tracker node) permettant de limiter les erreurs de page (page faults), ce qui améliore la *réduction du temps d'arrêt, l'équilibrage des charges et la continuité de service*. Deux méthodes majeures de migration sont détaillées :

- la technique *Pre-copy*, où les pages mémoire sont transférées progressivement avant l'arrêt de la machine, puis les données restantes sont copiées en une dernière étape ;
- la technique *Post-copy*, où la machine est arrêtée rapidement et redémarrée sur une autre hôte pendant que les pages mémoire sont transférées à la demande.

L'auteur Gupta [13] a comparé également les résultats obtenus avec ceux d'approches classiques telles que Virt-LM<sup>1</sup>, les optimisations de Pre-copy, la déduplication<sup>2</sup>, les méthodes hybrides<sup>3</sup> et les algorithmes Round Robin.

Les résultats expérimentaux montrent que le modèle offre de meilleures performances en termes de stabilité, de consommation énergétique et de capacité à maintenir la qualité de service, notamment dans des scénarios à forte variabilité de charge.

### 3.4.7 Parallelizing Live Migration of Virtual Machines

Cet article [37] propose une nouvelle approche pour optimiser la migration *à chaud* des machines virtuelles (VMs) en tirant parti du parallélisme offert par les architectures matérielles modernes. Le système introduit, appelé *PMigrate*, repose sur l'exploitation de deux types de parallélisme : le parallélisme de données (où différentes régions de mémoire ou de disque sont migrées simultanément) et le parallélisme en pipeline (où les étapes de

---

1. un système de gestion des migrations basé sur l'équilibrage de la charge entre hôtes

2. qui vise à éviter le transfert redondant de blocs mémoire déjà présents sur l'hôte cible

3. qui vise à éviter le transfert redondant de blocs mémoire déjà présents sur l'hôte cible

migration s'enchaînent de façon continue sans blocage).

Cette approche ne repose *ni sur des seuils* (fixes ou dynamiques) *ni sur une consolidation explicite* par extinction d'hôtes. L'objectif principal est de *réduire le temps total de migration* et, surtout, de *minimiser le downtime (interruption de service)*, une contrainte critique dans les environnements cloud. L'évaluation est effectuée *pendant la migration*.

Les résultats expérimentaux [37] ‘montrent que PMigrate permet une accélération significative du processus de migration et une réduction du temps d'arrêt, selon le type d'application concernée. En conclusion, cette recherche met en lumière l'intérêt d'intégrer le parallélisme au cœur du processus de migration live. PMigrate constitue ainsi une avancée importante pour la gestion efficace des machines virtuelles dans les centres de données cloud modernes, en conciliant performance, scalabilité et stabilité.

### 3.4.8 Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in Cloud data centers

Les algorithmes déterministes en ligne optimaux et les heuristiques adaptatives jouent un rôle crucial dans la consolidation dynamique des machines virtuelles (VMs) dans les centres de données cloud, permettant d'équilibrer efficacité énergétique et performances [8].

Ce travail [8] apporte une approche robuste pour la consolidation dynamique, combinant preuves théoriques et heuristiques efficaces, validées sur des traces réelles, pour optimiser la consommation énergétique des centres de données tout en maintenant la performance. Parmi les approches déterministes, l'algorithme Modified Best-Fit Decreasing (MBFD), une extension des méthodes de bin-packing, se distingue par sa capacité à optimiser le placement des VMs tout en réduisant la consommation d'énergie avec une complexité temporelle de  $O(n \log n)$ .

Les algorithmes intégrant la modulation de fréquence (DVFS) optimisent conjointement la consolidation et l'ajustement dynamique de la fréquence CPU, tandis que les méthodes basées sur des seuils ajustables permettent de détecter efficacement les surcharges ou sous-utilisations des serveurs.

Pour mieux s'adapter aux variations imprévisibles de charge, Beloglazov et Buyya [8] ont développé des heuristiques intelligentes, notamment des approches par apprentissage par renforcement (comme le Q-learning) qui *ajustent dynamiquement les seuils* de consolidation en fonction d'une fonction de récompense combinant énergie et QoS.

Les évaluations réalisées montrent des réductions de 25 à 35 % de la consommation énergétique par rapport à une allocation statique, avec moins de 2 % de violations de SLA et une diminution de 15 à 20 % du nombre de migrations. Ces méthodes maintiennent également une utilisation moyenne du CPU entre 70 et 80 % sans compromettre les performances. Les perspectives futures incluent l'intégration de la mémoire et du réseau dans les décisions de consolidation, l'amélioration des modèles de prédiction de charge, et l'exploration de l'apprentissage par renforcement profond pour des optimisations encore plus fines.



### 3.4.9 pMapper : Power and Migration Cost Aware Application Placement in Virtualized Systems

Dans leur article, Verma et al. [43] ont introduit *pMapper*, un framework conçu pour optimiser le placement d'applications virtualisées dans les centres de données, en tenant compte à la fois de la *consommation énergétique*, des *coûts de migration* et du *respect des accords de niveau de service (SLA)*. La solution proposée repose sur une architecture modulaire articulée autour d'un arbitrateur central coordonnant trois modules spécialisés : gestionnaire de performance, gestionnaire de puissance, et gestionnaire de migration.

Contrairement aux approches classiques, pMapper [43] *ne repose pas sur des politiques de seuil (fixes ou dynamiques)* pour déclencher la migration. Il *réalise une consolidation* basée sur une stratégie adaptative fondée sur une évaluation coût-bénéfice, intégrant un modèle global de coût combinant énergie consommée, coût de migration et pénalités liées aux SLA.

Ainsi, les migrations à chaud ne sont initiées que si le gain énergétique escompté compense les coûts associés, notamment le temps d'arrêt et la surcharge réseau.

Les résultats expérimentaux montrent une réduction d'environ 25% de la consommation énergétique par rapport aux méthodes statiques, tout en garantissant une qualité de service conforme aux exigences SLA.

### 3.5 Tableau comparatif

TABLE 3.1 – Synthèse des approches de migration des VMs dans le Cloud (classées par année)

Article	Environnement Cloud		Approche de résolution			Évaluation
	Unité	Domaine	Type de seuil	Politique	Consolidation	
pMapper [43]2008	VM seule	inter-cloud	simple	seuil unique	non	après migration
Energy Efficient Dynamic Integration of Thresholds [36]2011	VM seule	intra-cloud	dynamique	double seuil	oui	après migration
Energy-aware resource allocation heuristics [7]2012	VM seule	intra-cloud	simple	double seuil	oui	après migration
Optimal online deterministic algorithms [8]2012	VM seule	intra-cloud	dynamique	double seuil	oui	après migration
Energy Conscious Dynamic Provisioning [27]2013	VM seule	intra-cloud	dynamique	double seuil	oui	après migration
Parallelizing Live Migration [37]2013	VM seule	intra-cloud	simple	seuil unique	non	pendant migration
BGM-BLA [40]2015	VM seule	intra-cloud	dynamique	double seuil	oui	après migration
EAMA [16]2021	VM seule	intra-cloud	dynamique	double seuil	oui	après migration
Simulation via CloudSim [13]2021	VM seule	intra-cloud	simple	seuil unique	non	pendant migration

#### 3.5.1 disuccion du tableau comparatif

La majorité des approches étudiées utilisent une stratégie basée sur des seuils pour déclencher la migration des VMs. Les seuils dynamiques, comme dans les travaux de Sinha, Tao ou Ibrahim, s'adaptent à la charge du système et offrent de meilleures performances. À l'inverse, les seuils simples et statiques, présents dans des approches comme CloudSim ou pMapper, sont plus limités. La consolidation n'est réalisable que lorsque la politique repose sur un double seuil, ce qui permet d'identifier à la fois les hôtes surchargés et sous-utilisés. Ainsi, seules les méthodes avec double seuil, comme celles de Beloglazov ou Maurya, permettent une consolidation efficace. En revanche, les approches à seuil unique ne permettent pas une consolidation optimale. Tous les travaux analysés considèrent la migration d'unités individuelles (VM seule), sans prise en charge de groupes de VMs ou de clusters. Enfin, tous les scénarios évaluent des environnements au sein d'un même cloud,

à l'exception de pMapper, qui prend en compte un contexte inter-cloud.

## 3.6 Conclusion

Dans ce chapitre, nous avons présenté un état de l'art approfondi des techniques de migration des machines virtuelles (VMs) dans les environnements cloud. Une taxonomie complète a été proposée, articulée autour de trois axes principaux : le problème étudié, la solution proposée, et les modalités d'évaluation.

Sur la base de cette analyse, nous proposons, dans le chapitre suivant dédié à la simulation, d'étudier l'impact des *types de seuils et des politiques de déclenchement* sur les approches *avec ou sans consolidation*, dans le cadre de l'optimisation de *la consommation énergétique*.

# Chapitre 4

## Implémentation et résultats expérimentaux

### 4.1 Introduction

Dans ce dernier chapitre, nous allons étudier les approches de *migration live* de VMs basées sur la politique *de seuils*, dans le but de *réduire la consommation énergétique*.

À cet effet, nous avons utilisé le simulateur *CloudSim*, en adoptant un paradigme de programmation orienté objet avec le langage *Java*. Notre objectif est d'étudier l'impact du *type de seuil (statique, dynamique, hybride) et de la politique de seuil (unique ou double)*, en lien avec la présence ou l'absence de *techniques de consolidation*.

Ainsi, nous proposons d'analyser et de comparer les approches suivantes :

**1. Deux politiques à seuil unique sans consolidation :**

- **SSTP** (Static Single Threshold Policy).
- **DSTP** (Dynamic Single Threshold Policy).

**2. Trois politiques à seuil double avec consolidation :**

- **SDTP** (Static Double Threshold Policy)
- **DDTP** (Dynamic Double Threshold Policy)
- **HDTP** (Hybrid Double Threshold Policy)

### 4.2 Langage et environnement de développement

#### 4.2.1 Le langage de programmation Java

Java est un langage de programmation orienté objet, développé par Sun Microsystems en 1995, et actuellement maintenu par Oracle Corporation. Il a été conçu pour être portable, sécurisé et indépendant de la plateforme, ce qui en fait un choix privilégié pour le développement d'applications réparties, d'applications Web, mobiles (Android) et d'applications d'entreprise.

Java repose sur le principe "écrire une fois, exécuter partout" (Write Once, Run Anywhere – WORA), ce qui signifie qu'un programme Java compilé peut s'exécuter sur n'importe quelle machine disposant d'une Java Virtual Machine (JVM), sans modification du code source.

Aujourd'hui, Java est omniprésent dans le monde de la programmation. Les caractéristiques qui ont contribué à son succès sont nombreuses [18] parmi lesquelles :

- **Indépendance vis-à-vis de la plateforme** : le code Java peut être exécuté sur toute machine disposant d'une JVM, indépendamment du système d'exploitation.
- **Portabilité** : les applications Java peuvent être déployées sur différents systèmes sans nécessiter de recompilation.
- **Organisation modulaire** : le code est structuré en classes, chacune traitant un aspect spécifique du programme.
- **Gestion automatique de la mémoire** : grâce au *garbage collector*, Java gère la mémoire de manière efficace.
- **Multitâche** : Java prend en charge le *multithreading*, permettant l'exécution simultanée de plusieurs tâches.

## 4.2.2 L'environnement de développement IntelliJ IDEA

IntelliJ IDEA est un *environnement de développement intégré (IDE)* avancé conçu principalement pour le langage *Java*, développé par la société *JetBrains*. Réputé pour sa richesse fonctionnelle, il offre une prise en charge intelligente du code, avec notamment l'autocomplétion contextuelle, l'analyse statique, la détection d'erreurs en temps réel, ainsi que de puissants outils de refactorisation. IntelliJ IDEA propose une interface intuitive et intègre de nombreux outils de développement modernes, et les frameworks Java les plus utilisés (Spring, Hibernate, etc.). Grâce à sa compatibilité avec une grande variété de plugins, il s'adapte facilement aux besoins spécifiques des développeurs, ce qui en fait un IDE de référence dans le développement Java professionnel.

## 4.2.3 Le simulateur CloudSim

### 4.2.3.1 Présentation du CloudSim

CloudSim est un outil de simulation développé par le laboratoire CLOUDS de l'Université de Melbourne pour l'étude des environnements de cloud computing. Il permet aux utilisateurs d'évaluer des aspects spécifiques d'un système sans se soucier des détails de bas niveau liés aux infrastructures et services basés sur le cloud. Autrement dit, il prend en charge la modélisation, la simulation et la mise en œuvre fluide dans la conception de cadres de cloud computing.

CloudSim est une plateforme autonome qui prend en charge la simulation de clouds à grande échelle. Elle permet de modéliser les utilisateurs, les applications, les centres de données, les courtiers (brokers), ainsi que les politiques de planification (scheduling) et de provisionnement.

Il s'agit d'un logiciel open source qui offre aux utilisateurs la possibilité de tester des applications dans un environnement contrôlé et reproductible, d'identifier les goulets d'étranglement du système sans recourir à des infrastructures cloud réelles, et d'expérimenter différentes configurations afin de développer des techniques adaptatives de provisionnement.

#### 4.2.3.2 Raisons de l'adoption de CloudSim pour la modélisation et la simulation

- Simulation de différents centres de données dans un environnement cloud.
- Virtualisation des hôtes (serveurs), avec des politiques personnalisées pour le provisionnement des ressources aux machines virtuelles (VMs).
- Support des conteneurs logiciels à usage spécifique.
- Prise en charge des ressources informatiques sensibles à la consommation énergétique.
- Modélisation de différentes topologies de réseau de centres de données ainsi que des applications fondées sur l'échange de messages.
- Possibilité d'ajouter ou de retirer dynamiquement des composants de simulation.
- Simulation pouvant être arrêtée et reprise.
- Prise en charge des politiques définies par l'utilisateur pour l'allocation des hôtes aux VMs et des ressources d'hôtes aux VMs.

### 4.3 Conception du système de migration de VMs

#### 4.3.1 Processus de migration de VMs

##### 4.3.1.1 Description du processus

Dans ce travail, nous étudions les approches de migration live des machines virtuelles (VMs) basées sur des seuils d'utilisation des ressources, en particulier le CPU. L'objectif est de garantir l'efficacité énergétique des centres de données. Dans ce cadre, le système implémenté effectue une surveillance cyclique de l'état de chaque hôte actif. En fonction de l'utilisation détectée, trois situations peuvent se présenter, chacune entraînant un comportement spécifique du système :

- **Hôte surchargé** : la machine virtuelle (VM) présentant la plus faible utilisation du processeur (CPU) est sélectionnée (BFD CPU : Best Fit Decreasing CPU), dans le but de réduire la charge avec un impact minimal sur les performances globales. Cette VM est ensuite migrée vers un hôte actif choisi selon l'approche Best Fit Decreasing, minimisant ainsi la consommation énergétique.
- **Hôte sous-utilisé** : si la consolidation est activée, l'ensemble des VMs hébergées sont sélectionnées pour migration. Chacune est redéployée vers un hôte compatible, ce qui permet de mettre l'hôte initial hors service et de réduire la consommation énergétique.
- **Hôte dans un état normal** : aucune action de migration n'est effectuée.

Ce processus est répété régulièrement pour assurer une répartition optimale des charges et une réduction globale de la consommation énergétique dans le centre de données.

#### 4.3.2 Algorithmes de migration

La migration des machines virtuelles (VMs) basée sur les seuils s'effectue en deux étapes successives : la sélection des VMs à migrer, puis leur placement sur des hôtes compatibles, c'est-à-dire ceux qui satisfont leurs besoins en ressources (CPU, mémoire

vive et bande passante). Les algorithmes correspondant à chacune de ces étapes sont détaillés dans les sous-sections suivantes

#### 4.3.2.1 Algorithme de sélection de VMs à migrer

La sélection des machines virtuelles (VMs) repose sur l'état de charge de l'hôte physique ainsi que sur la stratégie adoptée, à seuil unique ou double.

- Dans la **stratégie à seuil unique**, une seule VM est migrée à partir d'un serveur surchargé (dont le taux d'utilisation est supérieur ou égal à  $T_{upper}$ ) ; il s'agit de la VM présentant la plus faible utilisation du processeur (CPU). Dans la **stratégie à seuil double**, une consolidation des serveurs est appliquée : toutes les VMs hébergées sur les serveurs sous-utilisés (dont le taux d'utilisation est inférieur ou égal à  $T_{lower}$ ) sont alors migrées.

L'algorithme suivant est utilisé pour faire cette sélection.

---

**Algorithm 1:** Algorithme de sélection des machines virtuelles à migrer

---

**Input:** `host` : hôte physique à analyser  
`estSurchargé` : booléen indiquant si `host` est surchargé (vrai) ou sous-utilisé (faux)  
`consolidation` : booléen indiquant si l'objectif est de consolider l'hôte (vrai) ou non (faux)  
**Output:** `listeVMsÀMigrer` : liste des machines virtuelles à migrer depuis `host`

```

1 begin
2   listeVMsÀMigrer  $\leftarrow$  NULL
3   vmList  $\leftarrow$  VMs de host
4   if estSurchargé == vrai then
5     if vmList  $\neq$  NULL then
6       vmMin  $\leftarrow$  VM avec la plus faible utilisation CPU dans vmList
7       listeVMsÀMigrer  $\leftarrow$  vmMin
8     else
9       if consolidation == vrai then
10        listeVMsÀMigrer  $\leftarrow$  toutes les VMs sur l'hôte sous-utilisé
11  return listeVMsÀMigrer

```

---

#### 4.3.2.2 Algorithme de placement des VMs

L'algorithme de placement des machines virtuelles (VMs) migrées repose sur une stratégie de type Best Fit Decreasing (BFD), visant à minimiser la consommation énergétique après migration.

Lorsqu'une VM doit être placée, l'algorithme parcourt l'ensemble des hôtes actifs du centre de données (c'est-à-dire ceux qui ne sont pas éteints) et évalue leur capacité à héberger la VM. Parmi les hôtes compatibles, il sélectionne celui pour lequel la consommation énergétique estimée après placement est la plus faible. Cette estimation est calculée à l'aide de l'équation suivante :

$$P_{\text{estimé}} = P_{\text{max}} \times (0,7 + 0,3 \times \text{CPU}_{\text{après placement}}) \quad (4.1)$$

---

**Algorithm 2:** Algorithme de placement basé sur la consommation énergétique

---

**Input:** `datacenter` : centre de données contenant la liste des hôtes`vm` : machine virtuelle à placer**Output:** `bestHost` : hôte cible optimal pour le placement de la VM

```
1 begin
2   hosts  $\leftarrow$  liste des hôtes actifs dans datacenter
3   Supprimer de hosts les hôtes incompatibles avec vm
4   Supprimer de hosts l'hôte d'origine de vm
5   foreach host dans hosts do
6     Estimer la consommation énergétique après placement :
7      $P \leftarrow$ 
8      $P_{\max} \times (0.7 + 0.3 \times (\text{utilisation CPU actuelle} + \text{utilisation CPU de } vm))$ 
9   bestHost  $\leftarrow$  l'hôte avec la plus faible valeur de  $P$ 
10  return bestHost
```

---

## 4.4 Stratégies de migration de VMs étudiées

Dans notre travail, nous avons exploré plusieurs politiques de migration basées sur l'utilisation de seuils. Chaque politique peut reposer sur un seuil unique ou double, et être de nature statique, dynamique (dans le cas d'un seuil unique) ou hybride (dans le cas de seuils doubles). Ainsi, nous avons proposer d'étudier cinq variantes de politiques de migration de machines virtuelles (voir Figure4.1) :

1. **SSTP** (Static Single Threshold Policy).
2. **DSTP** (Dynamic Single Threshold Policy).
3. **SDTP** (Static Double Threshold Policy)
4. **DDTP** (Dynamic Double Threshold Policy)
5. **HDTP** (Hybrid Double Threshold Policy)



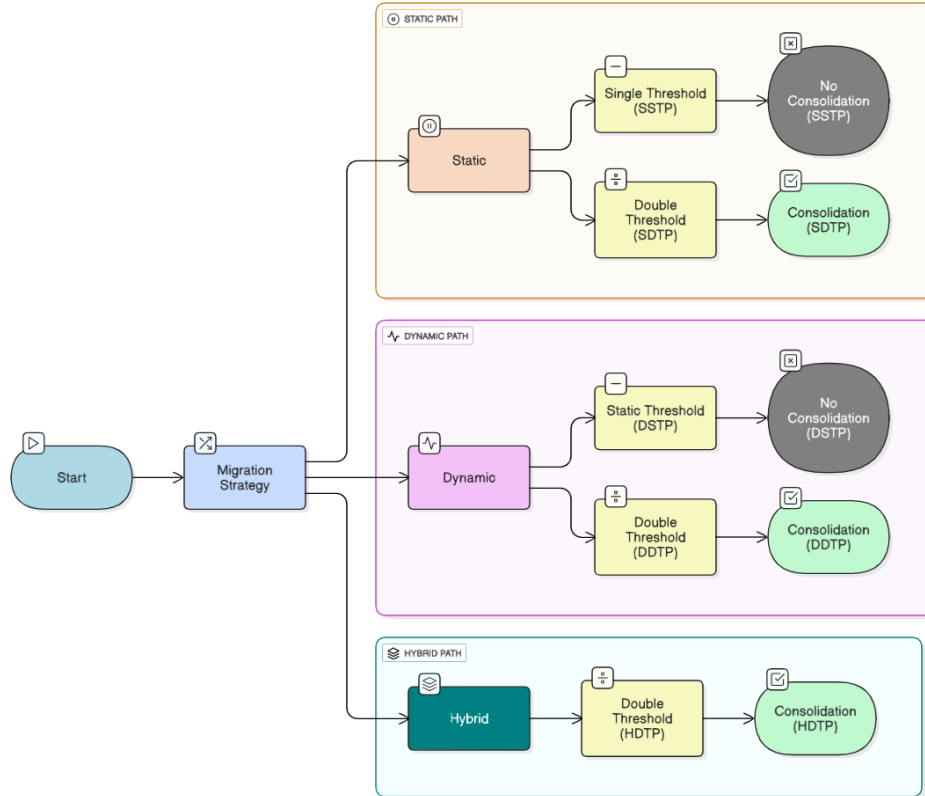


FIGURE 4.1 – Schéma hiérarchique des stratégies étudiées de migration basées sur les seuils

- **Single Threshold** : Utilise un seul seuil (souvent supérieur) pour détecter les hôtes surchargés. Si la charge dépasse ce seuil, une migration est déclenchée.
- **Double Threshold** : Définit deux seuils — supérieur et inférieur — pour éviter les migrations fréquentes et permettre un équilibre plus stable. Une migration est initiée si la charge dépasse le seuil supérieur ou tombe en dessous du seuil inférieur.
- **Hybride (Hybrid Double Threshold)** : Variante du Double Threshold qui combine un seuil statique et un seuil dynamique. Cette approche vise à tirer parti de la stabilité du seuil statique et de l'adaptabilité du seuil dynamique afin d'optimiser le processus de décision de migration.

Pour chaque variante, deux modes de gestion sont testés :

- **Sans consolidation** : Les migrations visent uniquement à équilibrer la charge entre les hôtes sans chercher à réduire le nombre d'hôtes actifs.
- **Avec consolidation** : En plus de l'équilibrage, cette approche tente de regrouper les charges sur un minimum d'hôtes afin de libérer certains hôtes et les éteindre, dans un but d'économie d'énergie.

## 4.5 Les métriques utilisées

Afin d'évaluer la performance des cinq politiques étudiées, il est nécessaire de définir des métriques pertinentes permettant de mettre en évidence leurs caractéristiques essentielles. Pour démontrer l'efficacité d'une approche par rapport à une autre, nous avons retenu

trois indicateurs de performance : la consommation d'énergie des centres de données, le nombre des hôtes actives, le nombre de migrations des machines virtuelles, ainsi que le temps total de migration.

### 4.5.1 Consommation d'énergie

Il existe plusieurs méthodes pour estimer la consommation d'énergie, parmi lesquelles nous avons retenu l'approche basée sur la "*puissance en fonction de l'utilisation*" [7]. Plusieurs travaux de recherche ont démontré qu'il existe une relation linéaire entre la consommation énergétique d'un serveur et le taux d'utilisation de son processeur. Ces études indiquent notamment que même à l'arrêt, un serveur consomme en moyenne 70 % de l'énergie qu'il utiliserait en pleine charge. Lorsque l'utilisation du processeur dépasse la limite inférieure reste fixée à 0,3. Ainsi, la consommation d'énergie  $P(u)$  a été définie à l'aide de la formule suivante :

$$\text{Power}_{\text{host}} = P_{\text{max}} \times (0.7 + 0.3 \times U_{\text{cpu}})$$

[7]

où :

- $P_{\text{max}} = 250 \text{ W}$  : puissance maximale de l'hôte.
- $U_{\text{cpu}} \in [0, 1]$  : taux d'utilisation du processeur (CPU).

Ce modèle suppose que la consommation au mode repos (*idle*) est de 70% de la puissance maximale, et qu'à charge maximale (*full load*), elle atteint 100% de  $P_{\text{max}}$ .

### 4.5.2 Utilisation moyenne du CPU

L'utilisation moyenne du CPU pour un ensemble de machines virtuelles (VMs) est donnée par la formule :

$$\text{CPU}_{\text{avg}} = \frac{1}{n} \sum_{i=1}^n U_i \times 100$$

[7]

où :

- $U_i$  : taux d'utilisation CPU (en fraction) de la VM  $i$
- $n$  : nombre total de VMs

Le résultat est multiplié par 100 pour exprimer l'utilisation moyenne en pourcentage.

### 4.5.3 Calcul des indicateurs de migration

Lors de chaque migration d'une machine virtuelle, plusieurs indicateurs sont calculés pour évaluer l'impact énergétique et temporel de la stratégie :

- **Temps de migration** : calculé par la formule

$$T_{\text{mig}} = \frac{RAM_{MB} \times 8}{BW_{Mbps}}$$

[7] où  $RAM_{MB}$  est la mémoire à transférer, et  $BW_{Mbps}$  la bande passante minimale entre hôte source et cible.

- **Énergie de migration** : Dans les travaux sur l'efficacité énergétique des centres de données, l'énergie consommée pendant la migration est donnée par :

$$E_{mig} = P_{max} \times \frac{T_{mig}}{3600}$$

[7] avec  $P_{max}$  la puissance maximale de l'hôte, exprimée en watts.

- **Nombre total de migrations** : il est compté dynamiquement dans la simulation à chaque migration effective.

#### 4.5.4 Le calcul de seuil supérieur et seuil inférieur

Dans les approches statiques, les seuils sont fixes et définis manuellement au début de la simulation. En revanche, dans les approches dynamiques, les seuils sont mis à jour instantanément selon les formules suivantes :

##### 4.5.4.1 Calcul du seuil supérieur $T_{upper}$

$$\begin{aligned} \text{temp} &= \text{Sum} + \left( \frac{\text{Bw}}{\text{Bw}(\text{host})} \right) + \left( \frac{\text{Ram}}{\text{Ram}(\text{host})} \right) \\ T_{upper} &= 1 - [(0,95 \cdot \text{temp} + \text{Sum}) - (0,90 \cdot \text{temp} + \text{Sum})] \end{aligned}$$

[7]

Où :

- Sum : somme des taux d'utilisation CPU de toutes les machines virtuelles (VMs) hébergées sur un hôte donné.
- Bw : bande passante totale allouée aux VMs sur cet hôte.
- Bw(host) : capacité totale de bande passante disponible sur l'hôte.
- Ram : mémoire RAM totale allouée aux VMs sur l'hôte.
- Ram(host) : capacité totale de RAM de l'hôte.

La variable temp représente une mesure globale de la charge actuelle de l'hôte, prenant en compte le CPU, la RAM et la bande passante. Ensuite,  $T_{upper}$  est calculé comme la différence entre deux bornes de sécurité (0,95 et 0,90) appliquées à cette charge, ce qui permet de détecter les situations de surcharge. Plus la valeur de temp est élevée, plus  $T_{upper}$  diminue, indiquant un risque de saturation.

##### 4.5.4.2 Calcul du seuil inférieur $T_{lower}$

$$\begin{aligned} n &= \text{nombre de VMs sur l'hôte} \\ \text{sum1} &= \frac{\text{sumCpu} + \text{allocatedBw} + \text{allocatedRam}}{n \cdot (\text{totalBw} + \text{totalRam})} \\ T_{lower} &= \min \left( 1.0, \max \left( 0.0, \sqrt{(\text{sumCpu} - \text{sum1})^2 - 0,3 \cdot \text{sum1}} \right) \right) \end{aligned}$$

[7]

Où :

- `sumCpu` : somme des pourcentages d'utilisation CPU des VMs sur l'hôte.
- `allocatedBw` : somme des bandes passantes allouées aux VMs.
- `allocatedRam` : somme des RAM allouées aux VMs.
- `totalBw` et `totalRam` : capacités totales respectives de bande passante et de RAM de l'hôte.

Cette approche permet de détecter les hôtes sous-utilisés en comparant la charge réelle à une moyenne relative normalisée, puis en appliquant un seuil ajusté par un facteur de tolérance (ici 30%).

## 4.6 Implémentation et Interprétations

Dans le cadre de ce travail, nous avons comparé les cinq approches basées sur la méthode *Threshold* décrites précédemment pour la migration live des machines virtuelles. L'objectif principal est d'évaluer l'impact de chaque approche sur la consommation énergétique, le coût de migration, ainsi que l'efficacité globale du système cloud.

Avant migration, nous surveillons continuellement les ressources (CPU) des hôtes afin de détecter les situations de surcharge ou de sous-utilisation. Lorsqu'un seuil est dépassé, une ou plusieurs machines virtuelles sont migrées vers des hôtes plus appropriés, selon une stratégie de placement définie.

Après migration, nous analysons les effets sur la répartition de la charge, l'économie d'énergie réalisée, et le temps nécessaire pour chaque opération.

Les stratégies comparées incluent l'utilisation de seuils statiques et dynamiques, avec ou sans consolidation des ressources.

Cette étude a été simulée avec *CloudSim Plus* et accompagnée d'une collecte de données telles que le nombre total de migrations, le temps cumulé de migration, et l'énergie consommée par les migrations, permettant ainsi une évaluation objective de chaque stratégie.

### 4.6.1 Configuration de l'environnement de simulation

Notre modèle se compose d'un centre de données unique constitué de 16 hôtes physiques. Chaque hôte peut héberger plusieurs machines virtuelles (VM) hétérogènes, avec un total de 64 VM déployées dans l'environnement. Ce système exécute un ensemble de 200 tâches (*Cloudlets*). Les hôtes physiques et les VM disposent de ressources variées telles que l'utilisation du processeur, la capacité de la mémoire vive (RAM), le nombre d'instructions par seconde (MIPS), les limites d'utilisation des ressources, la bande passante, entre autres.

### 4.6.2 Étude de l'approche à seuil unique statique (SSTP)

Nous avons débuté cette phase par l'évaluation de l'approche à seuil statique avec un seul seuil (*Single Threshold*). Pour ce faire, la valeur du seuil supérieur  $T_{upper}$  a été progressivement ajustée (0,9, 0,8, 0,7 et 0,6), en comparaison avec le scénario de référence sans politique de migration (*WNMP*). L'objectif est d'analyser l'impact de chaque configuration sur la consommation énergétique ainsi que sur le coût global de migration.

#### 4.6.2.1 Impact de $T_{upper}$ sur la consommation énergétique dans l'approche SSTP :

Le graphique ci-dessus illustre la consommation énergétique totale associée à différentes valeurs de seuil dans une stratégie de migration statique, comparée à une politique sans migration (WNMP – With No Migration Policy). :

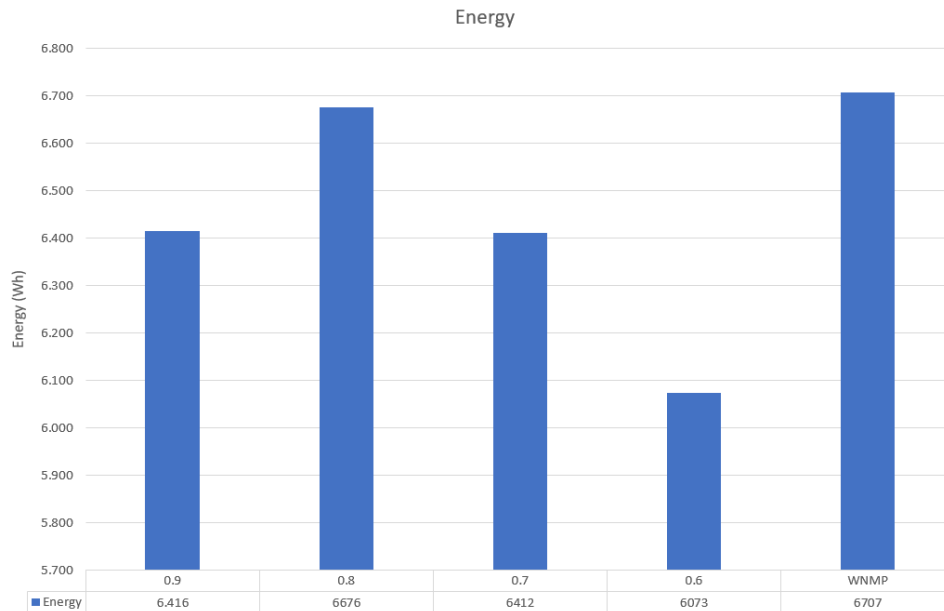


FIGURE 4.2 – Impact de la valeur du seuil  $T_{upper}$  sur l'approche Single Static Threshold Policy SSTP

Le graphique présenté dans la Figure 4.2 compare la consommation énergétique totale obtenue en appliquant la méthode **SSTP** (Single Static Threshold Policy), sans consolidation, pour différentes valeurs de seuil, avec celle obtenue en appliquant la politique **WNMP** (With No Migration Policy), qui ne prévoit aucun mécanisme de migration des machines virtuelles.

On observe que la stratégie **WNMP** entraîne la consommation énergétique la plus élevée, avec une valeur atteignant **6707 Wh**. Cette surconsommation s'explique par le fait que les hôtes restent actifs en permanence, même lorsqu'ils sont sous-utilisés, ce qui augmente considérablement la dépense énergétique.

En revanche, la stratégie **SSTP** avec un seuil fixé à **0.6** affiche la meilleure performance énergétique, avec une consommation réduite à **6073 Wh**. Ce résultat montre que, même sans consolidation, le fait d'activer la migration dès qu'un hôte dépasse une certaine utilisation permet de mieux répartir la charge, réduisant ainsi l'énergie consommée.

Les seuils plus élevés comme **0.7**, **0.8** ou **0.9** ne déclenchent la migration que rarement, ce qui rend leur comportement proche de **WNMP**. Cela se traduit par une consommation relativement élevée : **6412 Wh** pour un seuil de 0.7, **6676 Wh** pour 0.8, et **6416 Wh** pour 0.9. Cela confirme qu'un seuil trop élevé limite les gains énergétiques.

En conclusion, cette analyse montre que, même sans appliquer de consolidation, le choix judicieux d'un seuil fixe permet d'optimiser la consommation énergétique du centre de données par rapport à une politique qui n'inclut aucune migration (**WNMP**).

#### 4.6.2.2 Étude des paramètres coût de migration en fonction des différentes valeurs de $T_{upper}$

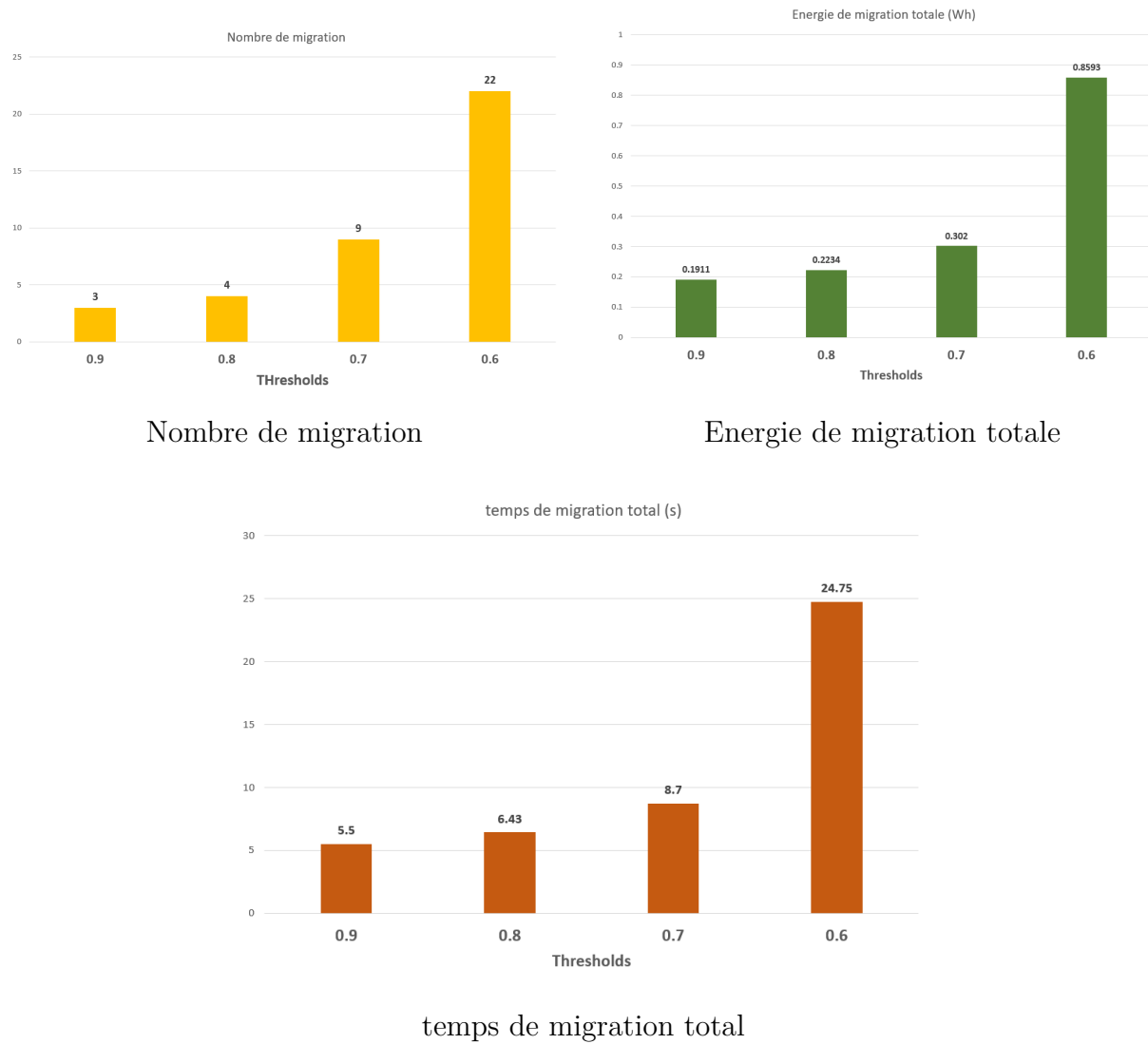


FIGURE 4.3 – Comparaison graphique des consommations énergétiques pour différentes stratégies

La figure ci-dessus met en évidence l'impact de la variation du *seuil statique* sur les coûts de migration au sein de l'environnement simulé. L'analyse révèle une tendance claire : plus la valeur du seuil diminue, plus le système devient sensible aux charges, ce qui se traduit par une augmentation progressive du **nombre de migrations**, du **temps total de migration** ainsi que de la **consommation énergétique**.

Avec un seuil élevé, tel que **0.9**, le système tolère une charge importante avant d'initier une migration. Cela limite considérablement le nombre d'opérations de migration (seulement 3), réduisant par conséquent le temps total (5.5 secondes) et l'énergie consommée (0.1911 Wh). À l'opposé, un seuil plus bas comme **0.6** rend le système beaucoup plus réactif : le nombre de migrations augmente fortement (22), entraînant un coût temporel (24.75 secondes) et énergétique (0.8593 Wh) significativement plus élevés.

Ces résultats permettent de tirer plusieurs enseignements :

- Un **seuil élevé** réduit le coût de migration, mais augmente le risque de surcharge des hôtes, ce qui peut compromettre la qualité de service.
- Un **seuil trop faible** déclenche des migrations excessives, parfois injustifiées, engendrant des surcoûts importants en temps et en énergie.
- Une **valeur intermédiaire**, telle que **0.7**, apparaît comme un compromis pertinent, permettant de concilier réactivité du système et maîtrise des ressources consommées.

### 4.6.3 Étude de l'approche à seuil unique dynamique (DSTP)

Dans cette phase, nous avons entamé l'évaluation de l'approche à seuil dynamique avec un seul seuil. L'analyse porte sur l'évolution de l'utilisation moyenne du CPU pour trois approches : SSTP (Static Single Treshold Policy), DSTP (Dynamic Single Treshold Policy) et sans politique de migration WNMP (With No Migration Policy). Cette étape inclut également une étude de l'évolution du seuil supérieur (*Tupper*) au cours de la simulation pour les stratégies SSTP et DSTP, afin de mieux comprendre leur comportement adaptatif face aux variations de charge.

#### 4.6.3.1 Évolution de l'utilisation du CPU pour SSTP, DSTP et WNMP

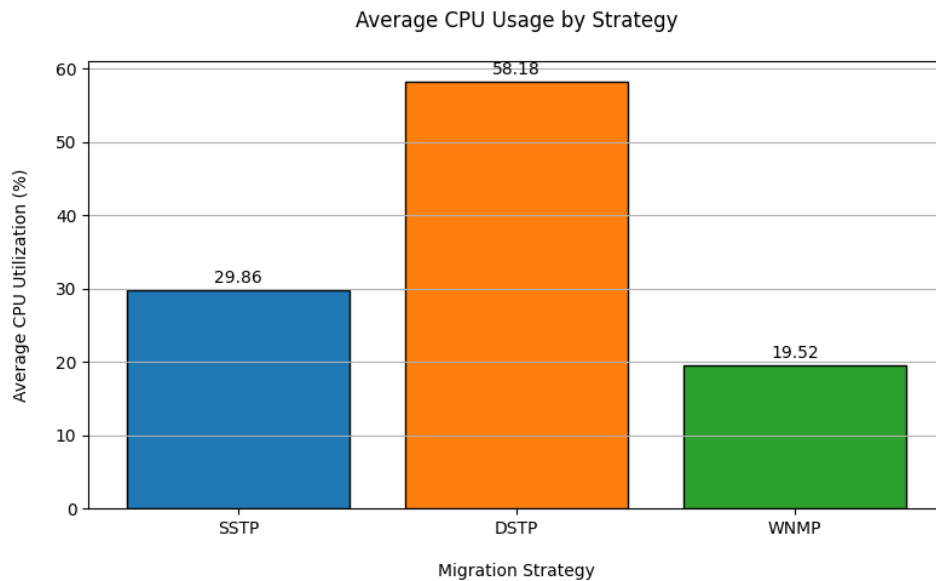


FIGURE 4.4 – l'utilisation du CPU en fonction des politiques SSTP, DSTP et WNMP

Les résultats obtenus concernant l'utilisation moyenne du processeur dans les différents scénarios de simulation mettent en évidence l'impact significatif des stratégies de migration sur l'efficacité de l'utilisation des ressources.

Dans le scénario WNMP, sans application de stratégie de migration, l'utilisation moyenne du CPU est relativement faible, avec une valeur de 19,52%. Ce faible taux d'occupation indique une répartition inefficace des charges entre les hôtes, entraînant une sous-utilisation des ressources disponibles et, potentiellement, une consommation énergétique inutilement élevée.

L'introduction de la stratégie SSTP (migration à seuil statique) permet d'améliorer cette efficacité, en atteignant une utilisation moyenne du CPU de 29,86%. Cette amélioration résulte d'une meilleure consolidation des machines virtuelles, bien que cette méthode reste peu flexible face aux variations dynamiques de la charge.

Enfin, la stratégie DSTP (migration à seuil dynamique) offre les meilleures performances, avec une utilisation moyenne du processeur atteignant 58,18%. Cette nette progression s'explique par l'adaptabilité de cette approche, qui ajuste les seuils de migration en fonction des conditions de charge, favorisant ainsi une consolidation plus efficace et une meilleure exploitation des ressources.

En résumé, plus la stratégie de migration est dynamique et réactive, meilleure est l'utilisation des ressources, permettant à la fois d'améliorer les performances globales du système et de réduire le gaspillage énergétique.

#### 4.6.3.2 Analyse de la dynamicité du seuil supérieur ( $T_{upper}$ ) dans les politiques SSTP et DSTP

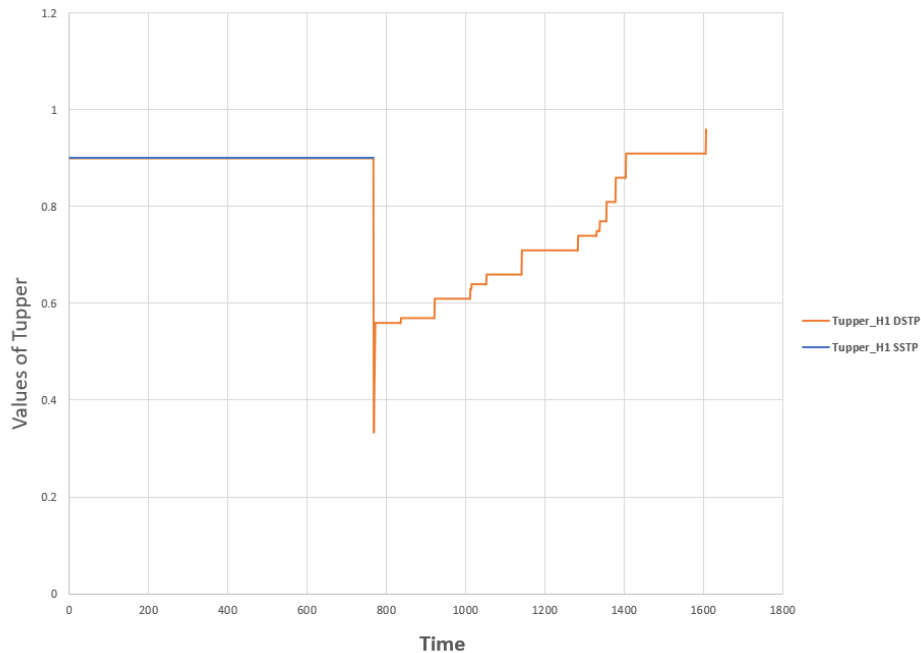


FIGURE 4.5 – Analyse de l'évolution du seuil supérieur en fonction des politiques SSTP et DSTP

- **SSTP (Seuil Statique)** : représentée en bleu, cette stratégie maintient une valeur constante de  $T_{upper}$  égale à **0.9** durant toute la simulation. Elle applique une politique fixe, indépendante de l'évolution réelle de la charge système.
- **DSTP (Seuil Dynamique)** : représentée en orange, cette stratégie ajuste dynamiquement la valeur de  $T_{upper}$  en fonction de l'utilisation du système. On remarque une baisse brutale de la valeur du seuil vers l'instant 750, suivie d'une remontée progressive, traduisant une adaptation continue à l'état du système.

**Interprétation :** La stratégie dynamique **DSTP** permet une plus grande sensibilité aux fluctuations de charge. Elle déclenche des migrations dès qu'un hôte approche du



seuil, même légèrement, ce qui renforce la réactivité du système face à la surcharge. En revanche, la stratégie statique **SSTP** reste moins réactive, car elle repose sur une valeur fixe, ne déclenchant des migrations que lorsque le seuil est strictement dépassé.

#### 4.6.4 Comparaison des approches sans consolidation SSTP et DSTP

##### 4.6.4.1 Comparaison du nombre de migrations entre SSTP et DSTP sans consolidation

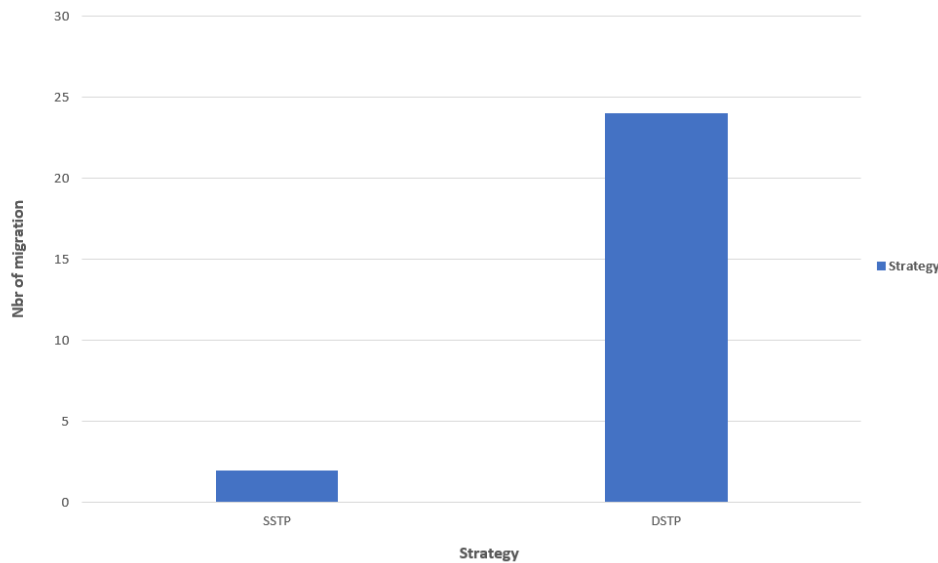


FIGURE 4.6 – Comparaison du nombre de migrations entre SSTP et DSTP sans consolidation

Le graphique présenté dans la Figure 4.6 compare le nombre total de migrations observé lors de l’application des stratégies SSTP (Single Static Threshold Policy) et DSTP (Dynamic Static Threshold Policy), sans recourir à la consolidation.

Les données montrent une différence marquée entre les deux approches :

- La méthode SSTP ne déclenche que 3 migrations sur l’ensemble de la simulation. Cette faible fréquence s’explique par l’utilisation d’un seuil statique élevé et constant, qui ne permet le déclenchement de la migration que dans des cas bien définis. Cela contribue à une stabilité du système, mais peut parfois limiter les possibilités d’optimisation plus fines.
- À l’inverse, la stratégie DSTP enregistre un total de 24 migrations, soit 8 fois plus que SSTP. Cette valeur élevée résulte de l’ajustement dynamique des seuils en fonction de la charge globale du système. Bien que cette approche soit plus réactive aux variations, elle entraîne une surcharge de migrations, ce qui peut nuire à la performance générale (augmentation du trafic réseau, surcharge des hôtes de destination, latence de service, etc.).

Ces résultats indiquent que, sans consolidation, la politique SSTP permet une gestion plus sobre et contrôlée des migrations, tandis que DSTP, bien qu’adaptative, génère un surcoût opérationnel important en raison de sa sensibilité aux fluctuations.

#### 4.6.4.2 Comparaison de la consommation énergétique entre SSTP et DSTP sans consolidation

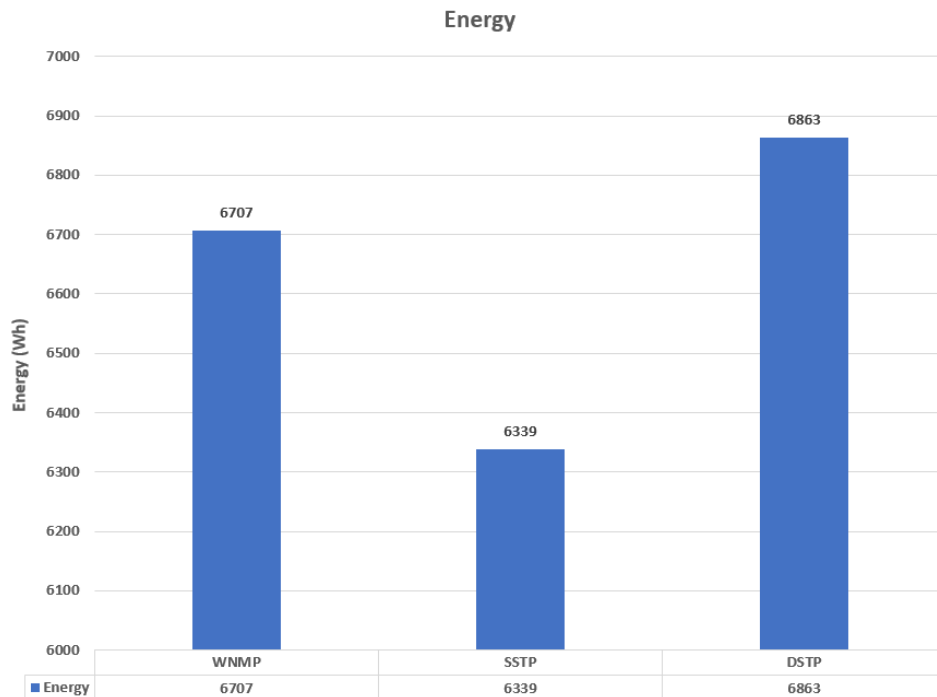


FIGURE 4.7 – Comparaison énergétique entre SSTP et DSTP sans consolidation

Le graphique ci-dessus illustre la consommation énergétique totale (exprimée en Wh) pour trois politiques de gestion des machines virtuelles : **WNMP** (With No Migration Policy), **SSTP** (Single Static Threshold Policy) et **DSTP** (Dynamic Static Threshold Policy), toutes appliquées sans mécanisme de consolidation.

Les résultats mettent en évidence les points suivants :

- La politique **SSTP** enregistre la meilleure performance énergétique avec une consommation totale de **6.339 Wh**. Malgré l'absence de consolidation, l'application d'un seuil statique bien calibré permet une répartition plus efficace des charges entre les hôtes, ce qui réduit le nombre de serveurs actifs simultanément, et donc la consommation globale.
- À l'inverse, la stratégie **DSTP** affiche une consommation plus élevée, atteignant **6.863 Wh**, dépassant même celle de la politique sans migration. Cette surconsommation peut s'expliquer par l'instabilité introduite par les seuils dynamiques, qui peuvent entraîner soit des migrations excessives et coûteuses, soit une mauvaise répartition de la charge.
- La politique **WNMP**, qui n'effectue aucune migration, présente une consommation intermédiaire de **6.707 Wh**. Ce résultat confirme que l'absence de migration empêche l'optimisation énergétique, en laissant tourner inutilement des hôtes peu utilisés.

Ces résultats démontrent que, même sans consolidation, la stratégie **SSTP** permet de réaliser des économies d'énergie significatives par rapport à **WNMP** et **DSTP**, à condition de bien choisir le seuil fixe de déclenchement des migrations.

## 4.6.5 Comparaison des approches avec consolidation SDTP, DDTP, HDTP

### 4.6.5.1 Comparaison énergétique entre DDTP, HDTP, SDTP et WNMP

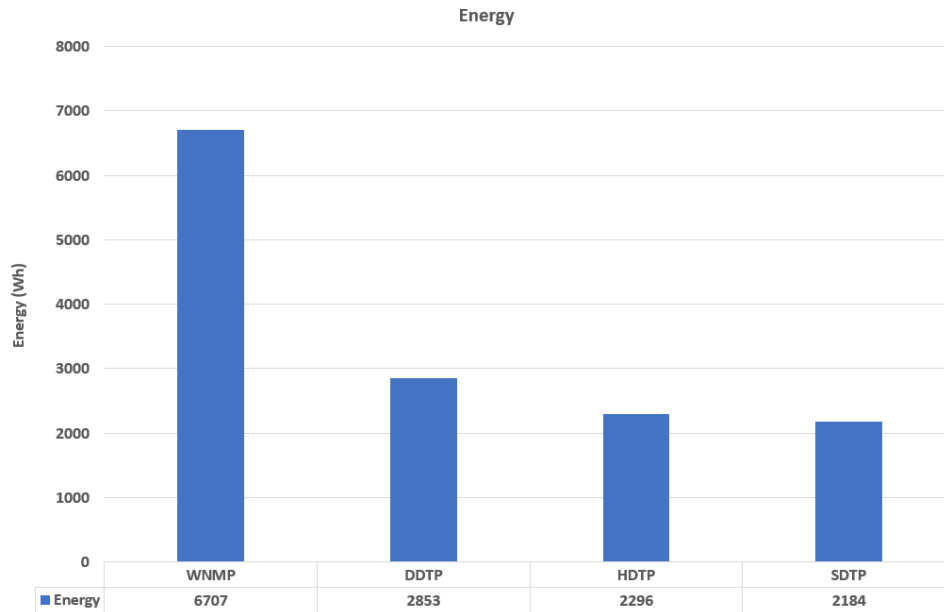


FIGURE 4.8 – Comparaison de la consommation énergétique totale des approches WNMP, DDTP, HDTP et SDTP

Dans cette expérience, nous comparons la consommation énergétique totale (en Wh) de quatre stratégies de gestion des machines virtuelles intégrant un mécanisme de consolidation. Le graphique à barres permet de visualiser de manière claire les différences de performance énergétique entre les approches étudiées.

Les stratégies analysées sont les suivantes :

- **WNMP** (With No Migration Policy),
- **DDTP** (Dynamic Double Threshold Policy),
- **HDTP** (Hybrid Double Threshold Policy),
- **SDTP** (Static Double Threshold Policy).

Les résultats obtenus sont les suivants :

- La stratégie **WNMP** enregistre la consommation énergétique la plus élevée avec **6 707 Wh**, en raison de l'absence de migration. Cela empêche toute consolidation, maintenant les hôtes actifs même lorsqu'ils sont sous-utilisés.
- La stratégie **DDTP** réduit considérablement cette consommation à **2 853 Wh**, grâce à l'utilisation de seuils dynamiques. Toutefois, une réactivité excessive peut engendrer des migrations non optimales.
- La stratégie **HDTP**, avec **2 296 Wh**, combine les avantages des seuils dynamiques et statiques. Elle assure un bon compromis entre adaptabilité et stabilité.
- Enfin, la stratégie **SDTP** offre la meilleure performance énergétique avec seulement **2 184 Wh**. Son approche statique et conservatrice permet une consolidation efficace tout en minimisant les migrations superflues.

**Conclusion** : Les résultats confirment que les stratégies intégrant une consolidation intelligente (notamment **SDTP** et **HDTP**) permettent de réduire la consommation énergétique de plus de **67 %** par rapport à une politique sans migration. Elles représentent des solutions viables et efficaces pour une gestion durable des ressources dans le cloud computing.

#### 4.6.5.2 Comparaison du nombre de migrations entre DDTP, HDTP et SDTP et WNMP

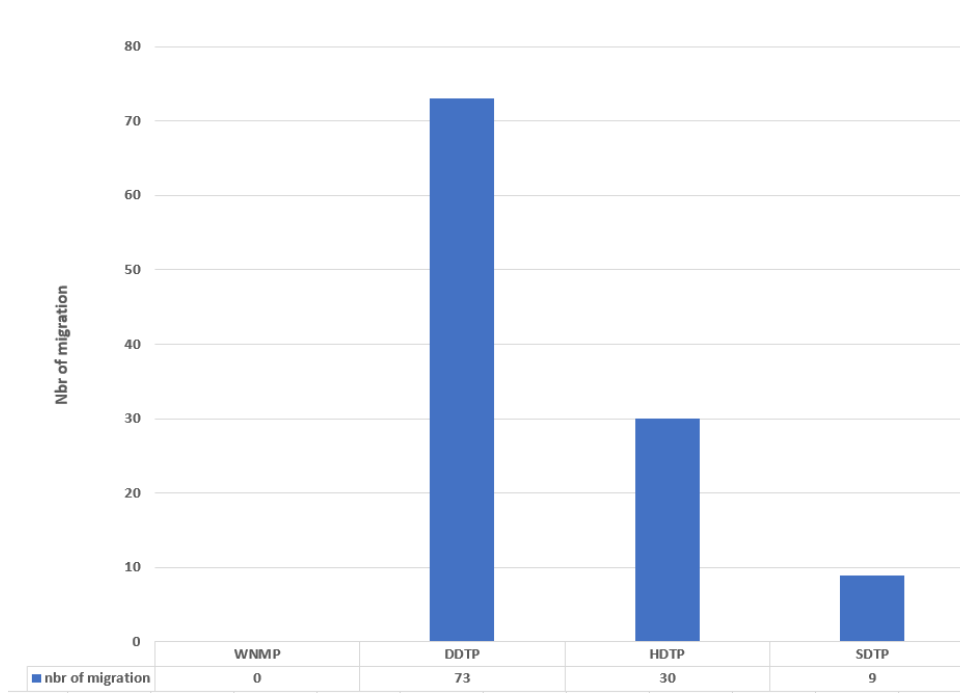


FIGURE 4.9 – Comparaison du nombre de migrations entre DDTP, HDTP et SDTP et WNMP

Dans cette partie, nous analysons le nombre total de migrations générées par différentes stratégies de gestion des machines virtuelles, en présence de consolidation. Les résultats sont les suivants :

- WNMP (With No Migration Policy) : 0 migration, ce qui est attendu, puisque cette stratégie n'intègre aucun mécanisme de réallocation. Cela se traduit par une stabilité totale mais au prix d'une consommation énergétique très élevée, comme vu précédemment.
- DDTP (Dynamic Double Threshold Policy) : 73 migrations. Cette stratégie, entièrement dynamique, ajuste continuellement ses seuils de décision en fonction de la charge du système. Cela la rend très réactive, mais aussi sujette à une fréquence de migration élevée, ce qui peut impacter les performances réseau et la stabilité globale.
- HDTP (Hybrid Double Threshold Policy) : 30 migrations. Elle constitue un compromis entre rigidité et adaptabilité. En intégrant à la fois des seuils statiques et dynamiques, elle permet une réduction significative du nombre de migrations par rapport à DDTP, tout en maintenant de bonnes performances énergétiques.

- SDTP (Static Double Threshold Policy) : 9 migrations seulement. C'est la stratégie qui génère le moins de migrations tout en atteignant une excellente efficacité énergétique. Son comportement plus conservateur minimise l'impact des migrations fréquentes sur les performances système.

Ces résultats soulignent que l'intégration de seuils bien calibrés, combinée à la consolidation, permet de limiter considérablement le nombre de migrations sans pour autant sacrifier l'efficacité énergétique. SDTP se distingue ici comme une solution équilibrée, offrant peu de perturbations tout en assurant une grande économie d'énergie.

#### 4.6.5.3 Comparaison du Nombre d'hôtes actifs entre DDTP, HDTP et SDTP et WNMP

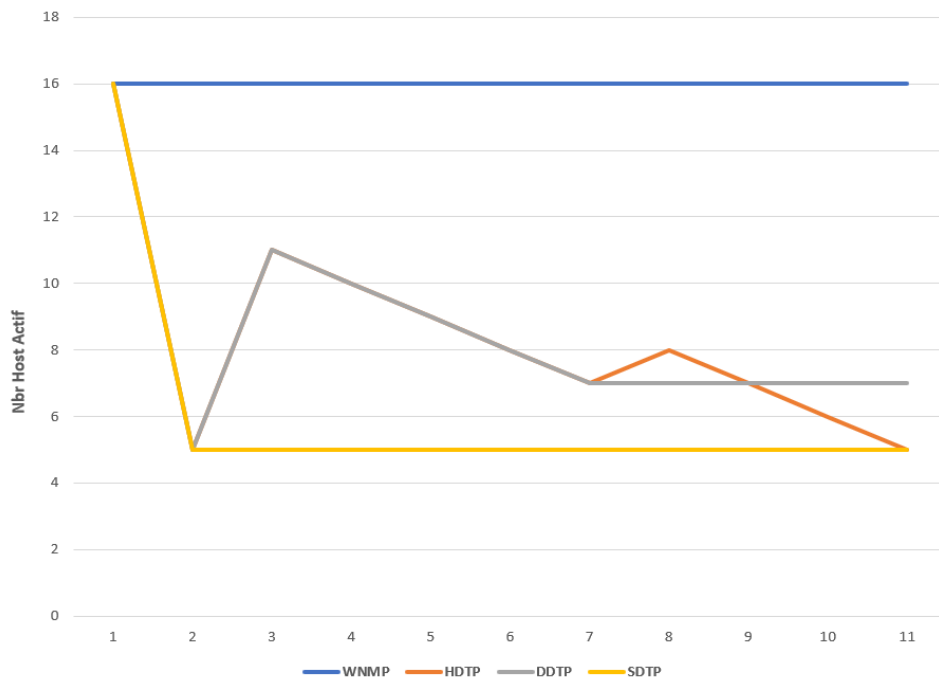


FIGURE 4.10 – Comparaison du Nombre d'hôtes actifs entre DDTP, HDTP et SDTP et WNMP

Dans le but de compléter l'évaluation énergétique et migration des différentes stratégies de migration des machines virtuelles, cette section analyse l'évolution du nombre d'hôtes physiques actifs au cours de la simulation, avec l'activation du mécanisme de consolidation. Le graphique correspondant présente cette évolution pour les stratégies suivantes : WNMP, DDTP, HDTP et SDTP.

- La stratégie WNMP (With No Migration Policy) maintient 16 hôtes actifs en permanence, représentant l'intégralité de l'infrastructure disponible. Ce comportement s'explique par l'absence totale de migration et de consolidation, entraînant une sous-utilisation des ressources et une consommation énergétique maximale, comme démontré précédemment.
- La stratégie SDTP (Static Double Threshold Policy) réduit immédiatement le nombre d'hôtes actifs à 5, niveau qu'elle conserve de manière stable tout au long de la simulation. Ce résultat témoigne de l'efficacité d'une politique basée sur des seuils fixes

rigoureux, permettant la désactivation rapide des hôtes inutilisés, et assurant ainsi une optimisation maximale des ressources.

- En comparaison, la stratégie DDTP (Dynamic Double Threshold Policy) affiche un comportement fluctuant, avec un nombre d'hôtes variant entre 5 et 11. Cette instabilité découle de l'adaptation dynamique des seuils selon la charge du système. Si cette stratégie permet une réaction rapide aux variations, elle induit également une volatilité du nombre d'hôtes actifs, susceptible d'affecter la stabilité opérationnelle.
- Quant à la stratégie HDTP (Hybrid Double Threshold Policy), elle se caractérise par une réduction progressive du nombre d'hôtes actifs, oscillant entre 5 et 8. Grâce à la combinaison de seuils statiques et dynamiques, cette approche hybride parvient à trouver un équilibre entre performance énergétique et stabilité, avec des variations modérées dans le parc actif.

Les résultats obtenus démontrent que les stratégies intégrant la consolidation, notamment SDTP et HDTP, permettent de réduire significativement le nombre d'hôtes physiques actifs tout en garantissant une gestion efficace des ressources. Cela contribue directement à une consommation énergétique optimisée, tout en limitant les perturbations et en assurant une stabilité satisfaisante du système.

## 4.7 Conclusion

Ce chapitre a permis d'implémenter, simuler et analyser plusieurs stratégies de migration de machines virtuelles basées sur des seuils, avec ou sans consolidation, en utilisant l'outil CloudSim Plus. L'objectif principal était d'évaluer leur impact sur **la consommation énergétique, le nombre de migrations, le temps de migration, et le nombre d'hôtes actifs**, dans un environnement cloud.

Les résultats peuvent être résumés comme suit :

### 1. Pour les approches sans consolidation SSTP et DSTP :

la stratégie **SSTP (Single Static Threshold Policy)** s'est montrée plus efficace que **DSTP (Dynamic Single Threshold Policy)** en matière de consommation d'énergie (6.3 kWh contre 6.86 kWh), grâce à un nombre de migrations réduit (3 contre 24), ce qui limite les coûts supplémentaires. En revanche, DSTP a permis une meilleure utilisation du CPU (58 % contre 29%), grâce à son adaptabilité dynamique.

### 2. Pour les approches sans consolidation SDTP, DDTP, et HDTP

- **SDTP (Static Double Threshold Policy)** : s'est révélée la plus performante, avec la plus faible consommation énergétique (2.184.714 Wh), le plus petit nombre de migrations (9), et le plus faible nombre d'hôtes actifs (5). Cela confirme l'efficacité d'une stratégie rigide mais économe.
- **HDTP (Hybrid Double Threshold Policy)** : a offert un bon compromis, combinant adaptabilité et stabilité. Elle a atteint une consommation très proche de SDTP (2.296.530 Wh), avec un nombre modéré de migrations (30), et un comportement stable du nombre d'hôtes actifs (entre 5 et 8).
- **DDTP (Dynamic Double Threshold Policy)** , bien qu'adaptative, a généré le plus grand nombre de migrations (73) et une consommation plus élevée (2.853.496 Wh) que les deux autres. Son comportement instable s'est traduit par une variation importante du nombre d'hôtes actifs (5 à 11), réduisant ainsi son efficacité énergétique.

Ainsi, au terme de cette étude comparative, nous pouvons conclure que la stratégie SDTP (Static Double Threshold Policy avec consolidation) est la plus efficace pour optimiser la consommation énergétique, limiter les migrations et stabiliser l'utilisation des ressources. Toutefois, HDTP représente une bonne alternative hybride pour les environnements nécessitant un compromis entre performance adaptative et stabilité, tandis que DDTP, malgré sa réactivité, reste moins performant en raison de sa volatilité.

Ces résultats confirment que l'intégration d'une politique de seuils statiques bien calibrés, combinée à la consolidation, constitue la meilleure solution pour une gestion durable des ressources dans les datacenters cloud.

# Conclusion générale

Ce mémoire a permis d'explorer en profondeur le domaine du Cloud Computing, en mettant l'accent sur la gestion intelligente des ressources à travers la migration des machines virtuelles, dans le but de réduire la consommation énergétique tout en maintenant une qualité de service optimale.

Dans un premier temps, nous avons présenté les concepts fondamentaux du cloud, de la virtualisation et de la migration des VMs. Une étude approfondie de la littérature a permis de recenser les principales stratégies existantes, qu'elles soient basées sur des seuils statiques, dynamiques ou hybrides, avec ou sans mécanisme de consolidation. Cette analyse a mis en lumière la diversité des approches proposées, tout en révélant plusieurs défis ouverts, notamment en matière de coût de migration, de maintien des performances et de maîtrise énergétique.

Sur le plan expérimental, des simulations ont été menées à l'aide de l'environnement CloudSim, à travers l'implémentation de plusieurs politiques de migration telles que WNMP, SSTP, DSTP, DDTP, HDTP et SDTP. Les résultats obtenus ont montré que les politiques intégrant la consolidation, en particulier SDTP et HDTP, permettent de réduire efficacement le nombre d'hôtes actifs et la consommation d'énergie, tout en limitant le nombre de migrations. À l'inverse, certaines politiques purement dynamiques, comme DSTP, peuvent engendrer un nombre excessif de migrations, compromettant ainsi la stabilité du système.

Ces résultats démontrent qu'un équilibre optimal entre performance, efficacité énergétique et qualité de service peut être atteint à condition de choisir des politiques de migration bien adaptées. Les approches hybrides, combinant des seuils fixes à des ajustements dynamiques, s'avèrent particulièrement efficaces dans cette optique.

Enfin, ce travail ouvre la voie à plusieurs perspectives de recherche, notamment l'intégration de l'intelligence artificielle pour améliorer la prise de décision, l'utilisation d'approches multi-objectifs, ou encore le recours à des techniques comme la programmation dynamique pour résoudre de manière optimale les problèmes complexes liés à la planification des migrations. Une validation expérimentale à plus grande échelle représenterait également une étape cruciale pour confirmer la robustesse des solutions proposées dans des environnements réels.



# Bibliographie

- [1] Red hat, 2025. Consulté le 22 février 2025.
- [2] KELLA Abdelaziz. *VERS UNE GESTION DE MIGRATION DES MACHINES VIRTUELLES POUR AMÉLIORER LA QUALITÉ DE SERVICE ET RÉDUIRE LA CONSOMMATION D'ÉNERGIE DANS LES CLOUD COMPUTING*. PhD thesis, Université d'Oran, 2015.
- [3] Raja Wasim Ahmad, Abdullah Gani, Siti Hafizah Ab. Hamid, Muhammad Shiraz, Feng Xia, and Sajjad A Madani. Virtual machine migration in cloud data centers : a review, taxonomy, and open research issues. *The Journal of Supercomputing*, 71 :2473–2515, 2015.
- [4] Gupta Ambika. A modelling & simulation via cloudsim for live migration in virtual machines [j]. In *IOP Conference Series : Materials Science and Engineering*, volume 1116, 2021.
- [5] Radhwan Y. Ameen and Asmaa Y. Hamo. Survey of server virtualization. *International Journal of Computer Science and Information Security (IJCSIS)*, 11(3), 2013. <https://sites.google.com/site/ijcsis/>.
- [6] Paul Barham, Boris Dragovic, Keir Fraser, Steven Hand, Tim Harris, Alex Ho, Rolf Neugebauer, Ian Pratt, and Andrew Warfield. Xen and the art of virtualization. In *Proceedings of the Nineteenth ACM Symposium on Operating Systems Principles (SOSP '03)*, pages 164–177, New York, NY, USA, 2003. ACM.
- [7] Anton Beloglazov, Jemal Abawajy, and Rajkumar Buyya. Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing. *Future generation computer systems*, 28(5) :755–768, 2012.
- [8] Anton Beloglazov and Rajkumar Buyya. Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers. *Concurrency and Computation : Practice and Experience*, 24(13) :1397–1420, 2012.
- [9] Lyes Bouali, Emad Abd-Elrahman, Hossam Afifi, Samia Bouzefrane, and Mehammed Daoui. Virtualization techniques : Challenges and opportunities. In *Mobile, Secure, and Programmable Networking : Second International Conference, MSPN 2016, Paris, France, June 1-3, 2016, Revised Selected Papers 2*, pages 49–62. Springer, 2016.
- [10] Anita Choudhary, Mahesh Chandra Govil, Girdhari Singh, Lalit K Awasthi, Emmanuel S Pilli, and Divya Kapil. A critical survey of live virtual machine migration techniques. *Journal of Cloud Computing*, 6 :1–41, 2017.

- [11] Ankita Desai, Rachana Oza, Pratik Sharma, and Bhautik Patel. Hypervisor : A survey on concepts and taxonomy. *International Journal of Innovative Technology and Exploring Engineering*, 2(3) :222–225, 2013.
- [12] Dad Djouhra. *Optimisation des performances des Datacenter des Cloud sous Contrainte d'énergie consommée*. Thèse de doctorat, 2016.
- [13] Ambika Gupta. A modelling & simulation via cloudsim for live migration in virtual machines. In *IOP conference series : materials science and engineering*, volume 1116, page 012138. IOP Publishing, 2021.
- [14] Ambika Gupta and Suyel Namasudra. A novel technique for accelerating live migration in cloud computing. *Automated Software Engineering*, 29(1) :34, 2022.
- [15] Michael Hogan, Fang Liu, Annie Sokol, and Jin Tong. Nist cloud computing standards roadmap - version 1.0. Special Publication 500-291, National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA, July 2011.
- [16] Muhammad Ibrahim, Muhammad Imran, Faisal Jamil, Yun-Jung Lee, and Do-Hyeun Kim. Eama : Efficient adaptive migration algorithm for cloud data centers (cdcs). *Symmetry*, 13(4) :690, 2021.
- [17] Rafia Islam, Vardhan Patamsetti, Aparna Gadhi, Ragha Madhavi Gondu, Chinna Manikanta Bandaru, Sai Chaitanya Kesani, and Olatunde Abiona. The future of cloud computing : benefits and challenges. *International Journal of Communications, Network and System Sciences*, 16(4) :53–65, 2023.
- [18] John Ivan. *Introduction to Java Programming*. Tech Books Publishing, New York, 2010.
- [19] Nancy Jain and Sakshi Choudhary. Overview of virtualization in cloud computing. In *2016 Symposium on Colossal Data Analysis and Networking (CDAN)*, pages 1–4. IEEE, 2016.
- [20] Tarandeep Kaur. *Cloud Computing*. Lovely Professional University, Jalandhar Delhi GT Road, Phagwara - 144411, n.d.
- [21] Vincent Kherbache, Mohamed Moussalih, Yannick, and Allan Lefort. Cloud computing. Rapport de licence professionnelle, IUT Nancy-Charlemagne, 2010. Tuteur : M. Lucas Nussbaum.
- [22] Avi Kivity, Yaniv Kamay, Dor Laor, Uri Lublin, and Anthony Liguori. KVM : The linux virtual machine monitor. In *Proceedings of the 2007 Ottawa Linux Symposium (OLS'07)*, 2007.
- [23] Dan Kusnetzky. *Virtualization : A manager's guide*. " O'Reilly Media, Inc.", 2011.
- [24] Tuan Le. A survey of live virtual machine migration techniques. *Computer Science Review*, 38 :100304, 2020.
- [25] Fang Liu, Jin Tong, Jian Mao, Robert Bohn, John Messina, Lee Badger, and Dawn Leaf. Nist cloud computing reference architecture. Special Publication 500-292, National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA, September 2011.

- [26] L. M. and E. R. Live migration of virtual machines in cloud environment : A survey. *Indian Journal of Science and Technology*, 8 :1–7, 2015.
- [27] Khushbu Maurya and Richa Sinha. Energy conscious dynamic provisioning of virtual machines using adaptive migration thresholds in cloud data center. *International Journal of Computer Science and Mobile Computing*, 2(3) :74–82, 2013.
- [28] Ishak H.A Meddah. *Virtualisation et Cloud Computing*. 10 2021.
- [29] Peter Mell, Tim Grance, et al. The nist definition of cloud computing. 2011.
- [30] Oo2 Formations & Consulting. Comprendre la technologie virtualisation et la machine virtuelle. <https://www.oo2.fr/actualites/comprendre-la-technologie-virtualisation-et-la-machine-virtuelle>, mai 2021. [Consulté en décembre 2021].
- [31] Claus Pahl, Antonio Brogi, Jacopo Soldani, and Pooyan Jamshidi. Cloud container technologies : a state-of-the-art review. *IEEE Transactions on Cloud Computing*, 7(3) :677–692, 2017.
- [32] Robert Rose. Survey of system virtualization techniques. 2004.
- [33] David SH Rosenthal. Emulation & virtualization as preservation strategies. *Andrew W. Mellon Foundation*, 2015.
- [34] Shashank Sahni and Vasudeva Varma. A hybrid approach to live migration of virtual machines. In *2012 IEEE international conference on cloud computing in emerging markets (CCEM)*, pages 1–5. IEEE, 2012.
- [35] Benhacine Samira, Cherikh Zouina, and Amroune Kamal. Etude de migration de l’infrastructure hyper convergente. Master’s thesis, Université Abderrahmane Mira - Bejaia, 2020.
- [36] Richa Sinha, Nidhi Purohit, and Hiteishi Diwanji. Energy efficient dynamic integration of thresholds for migration at cloud data centers. *IJCA Special Issue on Communication and Networks*, 1(44-49) :57, 2011.
- [37] Xiang Song, Jicheng Shi, Ran Liu, Jian Yang, and Haibo Chen. Parallelizing live migration of virtual machines. In *Proceedings of the 9th ACM SIGPLAN/SIGOPS international conference on Virtual execution environments*, pages 85–96, 2013.
- [38] Statista. Worldwide market share of leading cloud infrastructure service providers in q3 2024, 2024. Accédé le 18 juin 2025.
- [39] Jayachander Surbiryala and Chunming Rong. Cloud computing : History and overview. In *2019 IEEE Cloud Summit*, pages 1–7. IEEE, 2019.
- [40] Fei Tao, Chen Li, T Warren Liao, and Yuanjun Laili. Bgm-bla : a new algorithm for dynamic migration of virtual machines in cloud computing. *IEEE Transactions on Services Computing*, 9(6) :910–925, 2015.
- [41] Boris Teabe. *Performance et qualité de service de l’ordonnanceur dans un environnement virtualisé*. PhD thesis, Institut National Polytechnique de Toulouse-INPT, 2017.

- [42] Anthony T. Velte and Toby J. Velte. *Microsoft Virtualization with Hyper-V*. McGraw-Hill, Inc., New York, NY, USA, 1 edition, 2010.
- [43] Akshat Verma, Puneet Ahuja, and Anindya Neogi. pmapper : power and migration cost aware application placement in virtualized systems. In *ACM/IFIP/USENIX international conference on distributed systems platforms and open distributed processing*, pages 243–264. Springer, 2008.
- [44] VMware. *vSphere Virtual Machine Administration*. VMware, Inc., 2015. Disponible sur : <https://www.vmware.com>.
- [45] Timothy Wood, Prashant J Shenoy, Arun Venkataramani, Mazin S Yousif, et al. Black-box and gray-box strategies for virtual machine migration. In *NSDI*, volume 7, pages 17–17, 2007.
- [46] Qi Zhang, Lu Cheng, and Raouf Boutaba. Cloud computing : state-of-the-art and research challenges. *Journal of Internet Services and Applications*, 1(1) :7–18, 2010.