



Stratégies de migration de centres de données
en nuage informatique

par

TUDOR ANTOHI

Essai présenté au CeFTI

en vue de l'obtention du grade de maître en technologies de l'information
(Maîtrise en génie logiciel incluant un cheminement de type court en technologies de
l'information)

FACULTÉ DES SCIENCES
UNIVERSITÉ DE SHERBROOKE

Longueuil, Québec, Canada, mai 2017

Sommaire

Le but de cet essai est de vérifier si l'ordre de séquence d'une migration d'un centre de données vers l'infonuagique a un impact significatif sur le coût de migration. L'essai finit par prouver un patron d'analyse d'un système logique à partir duquel il est possible de calculer et optimiser le coût de la migration de ce système vers le nuage.

Le nuage informatique est un nouveau paradigme qui change dans les dernières années le paysage informatique des centres de données. La plus populaire solution infonuagique est actuellement IaaS. Les deux autres solutions PaaS et SaaS sont au-delà de la portée de cet essai. Sur le marché, il est possible de choisir plusieurs solutions pour adapter un centre de données traditionnel dans des solutions infonuagiques.

La problématique est de trouver les conséquences de cette solution. Plusieurs enjeux sont possibles et le coût de migration est parmi eux. Comprendre et convertir les systèmes logiques qui s'exécutent dans les centres de données traditionnels dans des modèles de coûts viables est un défi actuellement. Cet essai propose d'identifier une approche pertinente et générique pour optimiser le coût de migration.

La méthodologie utilisée est pragmatique. L'architecture générale d'un système Web exigeant une haute disponibilité est choisie et testée pour confirmer l'impact sur le coût de la séquence de migration des composants. Premièrement, on dérive le modèle d'architecture abstrait de ce système logique dans les éléments qui influencent le coût. Cette partie utilise une méthode qualitative. La deuxième partie trouve les formules et les variables et propose un algorithme d'optimisation de coût selon une méthodologie quantitative. La méthodologie est aussi confirmée par une analyse quantitative et qualitative. Finalement, les résultats sont vérifiés avec des statistiques et confirmés par un panel de spécialistes.

Les résultats confirment la stratégie utilisée dans l'algorithme d'optimisation de coût. Les mesures manuelles concordent à 100 % avec les calculs de l'algorithme. Mais c'est absolument sûr que l'échantillon possible n'est pas très large, attribuable à la complexité des architectures informatiques. Donc, dans cet essai, encore une fois, les opinions des experts

servent à confirmer l'approche et sa généralisation. Un résumé de résultats confirme la validité de la méthodologie et répond à la question de recherche.

La plus importante conclusion est que la méthodologie est correcte et la réponse est oui, l'ordre de migration a un effet important sur le coût de migration. Elle ouvre aussi la possibilité pour d'autres analyses qui dépassent les limitations forcées par la portée de l'étude. Une de recommandations est de développer à partir de cette méthodologie une solution intégralement automatisée. La deuxième recommandation est de mettre en pratique le modèle dans un contexte de nuage informatique PaaS.

Remerciements

Je tiens tout d'abord à remercier chaleureusement mon directeur de recherche, le professeur Pierre Martin Tardif, pour son expertise, ses conseils et son soutien. Je remercie aussi Lynn Legault et Vincent Echelard pour m'avoir guider dans les cours INF787 et INF788, pour leurs conseils académiques.

Surtout à Lynn Legault car sans elle, je n'aurais jamais pu finir cet essai.

Je tiens à remercier tous les membres du jury de m'avoir fait l'honneur d'accepter de juger mon mémoire.

Mes remerciements s'adressent également à tous les membres du comité Iteraplan de Loto-Québec qui ont contribué à la modélisation de l'architecture TI. C'est un travail d'excellente qualité, qui de base à une autre itération utilisée dans cet essai, porteuse de nombreuses améliorations.

Laurent Lamiaux, Avigaël Lévy, Chau Thien, Tony Cung, Félix Langelier, Benoit Desjardins, Rafal Ulatovski et Claude Lamy ont aidé de façon importante au développement de cet essai avec leurs observations et leur expérience. Un grand merci !

Simplifier et modéliser les connaissances diversifiées, expérimenter les modèles sont des techniques qui peuvent paraître simples. Je dois aussi une bonne partie du mon travail aux gens qui ne sont plus vivants. Capablanca et Karpov, grands joueurs d'échecs, qui m'ont appris comment extraire la simplicité dans les positions compliquées. Richard Feynman et Carl Sagan, professeurs capables de rédiger des cours clairs et concis de mécanique quantique et d'astronomie.

Je remercie aussi à ma famille, la vieille et la nouvelle. Mes deux filles ont continué à me soutenir pendant cette période de ma vie.

Table des matières

| | |
|--|-----|
| Sommaire | ii |
| Remerciements..... | iv |
| Table des matières | v |
| Liste des tableaux..... | vii |
| Liste des figures..... | ix |
| Liste des sigles, des symboles et des acronymes..... | xi |
| Introduction..... | 12 |
| Chapitre 1 Mise en contexte | 15 |
| 1.1 Centre de données en nuage informatique..... | 15 |
| 1.2 Modélisation d'un système informatique | 19 |
| 1.3 Migration en nuage informatique | 25 |
| Chapitre 2 Revue de la littérature..... | 29 |
| 2.1 Méthodologie de recherche | 29 |
| 2.1.1 Mots-clés | 29 |
| 2.1.2 Sources bibliographiques | 31 |
| 2.2 Coûts de migration vers le nuage | 31 |
| 2.3 Fournisseurs de solutions infonuagiques..... | 35 |
| 2.3.1 Amazon AWS | 35 |
| 2.3.2 Microsoft Azure..... | 38 |
| Chapitre 3 Problématique | 40 |
| 3.1 Description | 40 |
| 3.1.1 Question de recherche | 41 |
| 3.1.2 Objectifs et hypothèses | 43 |
| 3.1.3 Limites et contraintes..... | 45 |
| 3.2 Méthodologie proposée | 46 |
| Chapitre 4 Approche proposée | 47 |

| | |
|--|-----|
| 4.1 L'approche | 47 |
| 4.2 Choisir l'architecture | 49 |
| 4.2.1 Sélection d'un modèle d'architecture | 49 |
| 4.2.2 Sélection d'un système pour l'expérimentation | 50 |
| 4.3 Modéliser l'architecture..... | 50 |
| 4.4 Mapper l'architecture analysée aux blocs d'architecture des fournisseurs de service nuage..... | 56 |
| 4.5 Calculer et valider les coûts..... | 58 |
| 4.6 Identifier le coût optimal de migration selon la séquence de migration | 61 |
| 4.7 Analyser et valider les résultats | 63 |
| Chapitre 5 Analyse des résultats | 65 |
| 5.1 Mesures de la modélisation..... | 65 |
| 5.1.1 Composants technologiques..... | 65 |
| 5.1.2 Consommation de ressources | 66 |
| 5.1.3 Bande passante réseau..... | 66 |
| 5.1.4 Mappage infonuagique | 67 |
| 5.2 Analyse quantitative de résultats | 68 |
| 5.2.1 Résultats obtenus..... | 68 |
| 5.2.2 Statistiques descriptives | 70 |
| 5.2.3 Statistiques inférentielles | 72 |
| 5.3 Analyse qualitative de résultats | 74 |
| 5.4 Développements futurs..... | 76 |
| Conclusion..... | 77 |
| Références | 79 |
| Bibliographie..... | 84 |
| Annexe I Mappage..... | 85 |
| Annexe II Analyse de coût | 88 |
| Annexe III Méthode de points de migration dans le nuage..... | 98 |
| Annexe IV Vendeurs IaaS et PaaS | 101 |
| Annexe V Documents d'évaluation de spécialistes | 103 |

Liste des tableaux

| | |
|---|----|
| Tableau 1.1 Tableau de systèmes logiques génériques..... | 23 |
| Tableau 1.2 Éléments d'infrastructure d'un centre de données..... | 24 |
| Tableau 2.1 Tableau de ressources payables dans le nuage informatique | 32 |
| Tableau 4.1 Estimé des composants technologiques du système à migrer..... | 52 |
| Tableau 4.2 Exemple de compilation possible des estimations d'experts..... | 53 |
| Tableau 4.3 Consommation de bande passante..... | 55 |
| Tableau 4.4 Ressources payables dans le nuage informatique | 57 |
| Tableau 4.5 Étapes de migration dans le nuage informatique..... | 58 |
| Tableau 4.6 Tableau de résultats attendus et obtenus..... | 62 |
| Tableau 4.7 Formulaire d'évaluation de la méthode et de résultats..... | 63 |
| Tableau 5.1 Estimé des composants technologiques du système à migrer..... | 65 |
| Tableau 5.2 Tableau de machines..... | 66 |
| Tableau 5.3 Tableau de consommation de la bande passante | 67 |
| Tableau 5.4 Tableau de blocs payables dans le nuage informatique | 67 |
| Tableau 5.5 Tableau de résultats attendus et obtenus..... | 68 |
| Tableau 5.6 Tableau de résultats optimaux | 69 |
| Tableau 5.7 Tableau statistiques descriptives..... | 70 |
| Tableau 5.8 Tableau avec la position d'un EI dans la séquence et le coût..... | 73 |
| Tableau 5.9 Tableau corrélation de paires *position d'un EI dans la séquence et le coût) | 73 |
| Tableau 5.10 Formulaire d'évaluation de la méthode et de résultats..... | 74 |
| Tableau I.1 Tableau de blocs payables dans le nuage informatique | 86 |

| | |
|--|-----|
| Tableau I.2 Tableau de blocs payables dans le nuage informatique Azure | 86 |
| Tableau IV.1 Mappage de couche logique dans le nuage d'Amazon | 101 |

Liste des figures

| | |
|--|----|
| Figure 1-1 NIST sur les types de nuages informatiques..... | 16 |
| Figure 1-2 Étude Gartner Group sur les tendances dans les technologies de centres de données | 17 |
| Figure 1-3 Virtualisation vers le plateau de productivité | 18 |
| Figure 1-4 Architecture multiniveau avec redondance par centre et centres de données redondants | 20 |
| Figure 1-5 Architecture multiniveau sans redondance par centre et avec centres de données redondants | 21 |
| Figure 1-6 Architecture MASA | 22 |
| Figure 2-1 CMPs selon [20] | 34 |
| Figure 4-1 Méthodologie | 48 |
| Figure 4-2 Architecture multiniveau source | 52 |
| Figure 4-3 Architecture multiniveau en cours de migration..... | 54 |
| Figure 4-4 Architecture multiniveau cible migrée en nuage | 55 |
| Figure 4-5 Algorithme d'optimisation | 62 |
| Figure 5-1 Histogrammes Pareto et cumulatives..... | 71 |
| Figure I-1 Recommandation Amazon pour les machines I3 | 85 |
| Figure I-2 Prix Amazon I3 | 85 |
| Figure I-3 Recommandations Azure..... | 86 |
| Figure II-1 Aucun élément migré | 88 |
| Figure II-2 Élément E2 migré | 89 |
| Figure II-3 Élément E1 migré | 89 |

| | |
|--|-----|
| Figure II-4 Élément E4 migré | 90 |
| Figure II-5 Élément E3 migré | 90 |
| Figure II-6 Élément E6 migré | 91 |
| Figure II-7 Élément E5 migré | 91 |
| Figure II-8 Élément E7 migré | 92 |
| Figure II-9 Fin de migration et coût total..... | 92 |
| Figure II-10 Sortie de script Python..... | 92 |
| Figure III-1 Points de migration en infonuagique (pondération) | 100 |
| Figure IV-1 Gartner Group et les solutions IaaS en 2016..... | 101 |

Liste des sigles, des symboles et des acronymes

| | |
|---|----|
| AMI Amazon Machine Image – Image machine Amazon | 36 |
| C2P Cloud to Physical – Infonuagique vers physique | 25 |
| CAPEX Capital expenditure – Dépense de capital..... | 15 |
| CMP Cloud Migration Point – Point de migration nuage..... | 33 |
| EAM Enterprise Architecture Management (Iteratec) – Gestion d’architecture d’entreprise... 12 | |
| EBS Elastic Block Store – Block stockage Amazon | 37 |
| EC2 Elastic Compute – Serveur générique Amazon..... | 36 |
| IaaS Infrastructure as a Service – Infrastructure comme service | 41 |
| IaC Infrastructure as Code – Infrastructure comme code | 19 |
| MASA Meshed and Service Architecture – Architecture de services enfiletés..... | 19 |
| NIST National Institute of Standards and technology | 16 |
| OPEX Operational expenses – dépenses opérationnelles | 15 |
| P2C Physical to Cloud – Physique vers infonuagique..... | 25 |
| P2V Physical to Virtual – Physique vers Virtuel..... | 25 |
| PaaS Platform as a Service – Plateforme comme service | 16 |
| pl/sql Langage de programmation Oracle | 23 |
| SaaS Software as a Service – Logiciel comme service..... | 16 |
| SDDC Software Defined Data Center – Centre de données défini logiciel | 19 |
| SDN Software Defined Network – Réseau définie logiciel..... | 19 |
| Togaf The Open Group Architecture Framework | 12 |
| V2C Virtual to Cloud – Virtuel vers infonuagique..... | 25 |
| V2P Virtual to Physical – Virtuel vers physique..... | 25 |

Introduction

Cet essai étudie les stratégies de migration des centres de données vers l'infonuagique, pour trouver l'impact de l'ordre de migration des éléments d'infrastructure sur le coût de migration. Plus précisément, il étudie les méthodes d'évaluation des coûts de migration en utilisant une approche basée sur les points de fonction.

L'analyse utilise la méthodologie de développement architectural Togaf et EAM qui classifie le sujet dans le domaine technologique de l'informatique.

Le premier chapitre est une mise en contexte qui introduit les concepts clés de la recherche. Les centres de données, le nuage informatique et les techniques de migration sont présentés brièvement. La modélisation de l'architecture d'infrastructure d'un centre de données utilise les concepts EAM Iteraplan. Chaque système logique est décomposé et quantifié au niveau de ses éléments d'infrastructure, de ses composants technologiques et de ses interrelations.

Les techniques de migration en nuage diffèrent au niveau des coûts, de la performance, de la disponibilité et subissent des impacts par la virtualisation. Chaque vendeur de solutions infonuagiques offre sa technologie à des prix spécifiques. Le chapitre expose quelques-unes de ces offres et la problématique qui en découle. Puisque la migration requiert des dispositions particulières afin de conserver la disponibilité des systèmes, l'opération peut engendrer des coûts supplémentaires qui pourraient être minimisés. Le but de l'essai est d'estimer la meilleure stratégie de migration d'un grand centre de données vers le nuage informatique en utilisant la méthode de points de fonction, dans un environnement de type infrastructure comme service (IaaS) pour trouver l'ordre de migration optimal. Deux contraintes sont présentes, soit le besoin d'une disponibilité de 100 % des applications ainsi qu'un risque à zéro.

Le chapitre 2 présente l'analyse d'articles scientifiques et de documents trouvés sur les sites scholar.google.com, Université de Sherbrooke et d'autres sites pour obtenir un état de la connaissance actuelle du sujet. Les compagnies Amazon, Microsoft sont actuellement les entreprises les plus populaires des offres de produits et services infonuagiques de pointe. Ce

chapitre effectue une revue de la littérature scientifique orientée sur les migrations de centres de données et les mesures de succès.

Le chapitre 3 présente la problématique qui consiste à vérifier s'il est possible, à partir d'une fonction mathématique qui définit la complexité d'une migration d'un centre de données, de réduire les coûts estimés par des fournisseurs de solution infonuagiques en trouvant une séquence de migration optimale. Les méthodes de migration sont spécifiques à chaque fournisseur et le défi est de trouver un modèle général, utilisant les points de fonctions, capable d'évaluer les coûts de migration vers le nuage d'un centre de données. Certaines limitations s'appliquent et elles sont expliquées et documentées.

Le chapitre 4 présente l'approche de recherche sur la problématique définie au chapitre précédent. La démarche consiste à trouver un modèle et les interrelations d'un système logique, représentatif pour les entreprises qui désirent migrer leurs infrastructures vers le nuage et qui ont des contraintes de disponibilité, et où le risque est un enjeu. Une analyse de haut niveau présente les éléments d'infrastructure et les composantes technologiques d'un système logique et fournit la topologie du système logique. Le chapitre explique comment mapper les éléments de cette architecture source vers la solution infonuagique choisie, considérée architecture cible. Le but est de calculer le coût à partir d'une formule mathématique et de trouver l'ordre de migration qui offre le coût le plus bas.

La formule utilisée prend en compte la complexité des interconnexions réseau et la complexité des procédures d'installation et de configuration des éléments d'infrastructure ainsi que la complexité du transfert des données.

À partir de cette fonction, les impacts de la séquence des opérations sur les coûts d'une migration sont analysés. Ce chapitre présente aussi les risques et les faiblesses d'un modèle universel de coût et énumère des contraintes supplémentaires pour réduire la complexité de cette analyse.

Un système Web classique est choisi pour l'expérimentation. Les technologies de migration de Microsoft et Amazon sont analysées pour cette simulation de migration en nuage. Togaf impose une définition de l'architecture initiale et cible pour dériver la méthode de migration. Pour simuler la migration, l'essai utilise un diagramme d'états intermédiaires. Il existe une architecture source, plusieurs architectures intermédiaires qui changent avec chaque élément

d'infrastructure migré, et une architecture cible. La fonction mathématique qui calcule le coût s'applique sur chaque étape de migration et les résultats finaux sont comparés pour identifier le coût minimal.

Des statistiques sont appliquées pour valider les résultats et mesurer l'importance de la séquence opérationnelle. Cette validation est basée sur les statistiques descriptives et inférentielles. Une seconde validation consiste à consulter l'opinion de plusieurs spécialistes expérimentés. Ils répondent à des questions qui visent à confirmer la conformité de la méthodologie.

Le chapitre 5 fait l'analyse des résultats. C'est un chapitre pragmatique qui présente les tableaux de résultats et applique les méthodologies présentées au chapitre 4 pour confirmer l'hypothèse. Les résultats des analyses sont présentés avec des statistiques quantitatives. Enfin, la méthodologie est notée et approuvée par les spécialistes participant au sondage.

Chapitre 1

Mise en contexte

Le centre de données représente l'infrastructure physique et virtuelle des systèmes informatiques qui répondent aux besoins d'une entreprise. Il évolue actuellement à un rythme rapide, sous l'impulsion de l'arrivée de mégacentres de données mis en place par Amazon, Facebook, Google et Microsoft. Pour une entreprise, il est souhaitable de connaître les nouvelles technologies informatiques pour avoir un avantage face à la concurrence. Par exemple, les systèmes analytiques, l'intelligence artificielle, l'apprentissage automatique et les statistiques sont des technologies qui exigent plus de ressources informatiques. Les investissements internes sont coûteux et les entreprises cherchent des modèles alternatifs, dont les solutions infonuagiques font partie.

1.1 Centre de données en nuage informatique

Les architectes considèrent le modèle CAPEX traditionnel, qui consiste à acheter du matériel dédié et l'amortir sur une période de temps, ainsi que le modèle OPEX, qui consiste à utiliser une infrastructure de nuage partagé et payer en fonction de l'utilisation, et ce pour des raisons de flexibilité et de coût. [1]. Le nuage informatique est un modèle d'organisation qui permet un accès réseau fiable et sur demande à un groupe partagé de ressources informatiques configurables (réseaux, serveurs, stockage, applications et services) pouvant être mis en et hors service rapidement, avec un effort de gestion minimal [2]. Le nuage informatique joue un rôle important dans la façon dont les centres de données d'entreprise sont maintenant construits. Alors que les besoins en matière de centres de données augmentent, ils deviennent également de plus en plus complexes. Cette complexité se matérialise par un nombre croissant de dispositifs et d'applications, voir par exemple la prochaine révolution nommée l'Internet des objets (IoT).

Selon le National Institute of Standards and Technology (NIST), le nuage informatique a cinq caractéristiques, trois modèles de services et quatre modèles de déploiement. Les cinq caractéristiques sont : l'accès libre-service sur demande, l'accès large au réseau, la mise en commun des ressources, l'élasticité rapide et la résilience.

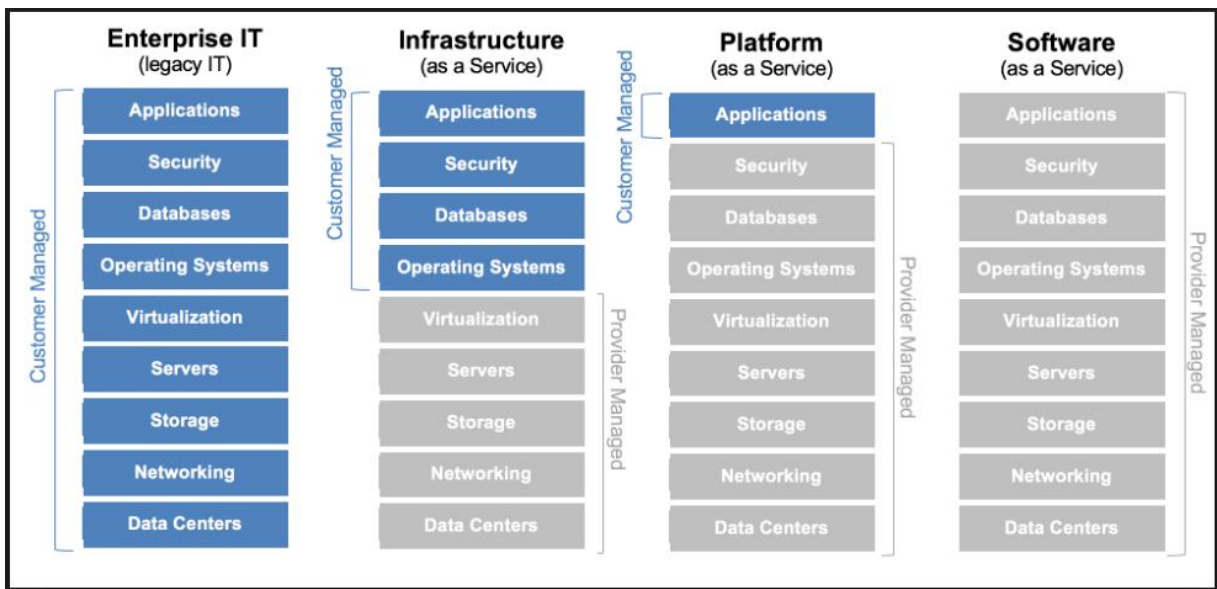


Figure 1-1 NIST sur les types de nuages informatiques [2]

Les trois modèles de services dans le nuage sont : l'infrastructure comme service (IaaS), la plate-forme comme service (PaaS) et les logiciels comme service (SaaS). Les modèles de déploiement dans le nuage sont : le nuage privé, le nuage communautaire, le nuage public et le nuage hybride. L'essai se concentre sur les migrations IaaS dans un nuage privé. Il présente l'analyse et le choix technologique des nuages Amazon et Microsoft, les plus populaires actuellement sur le marché, selon le graphique présenté.

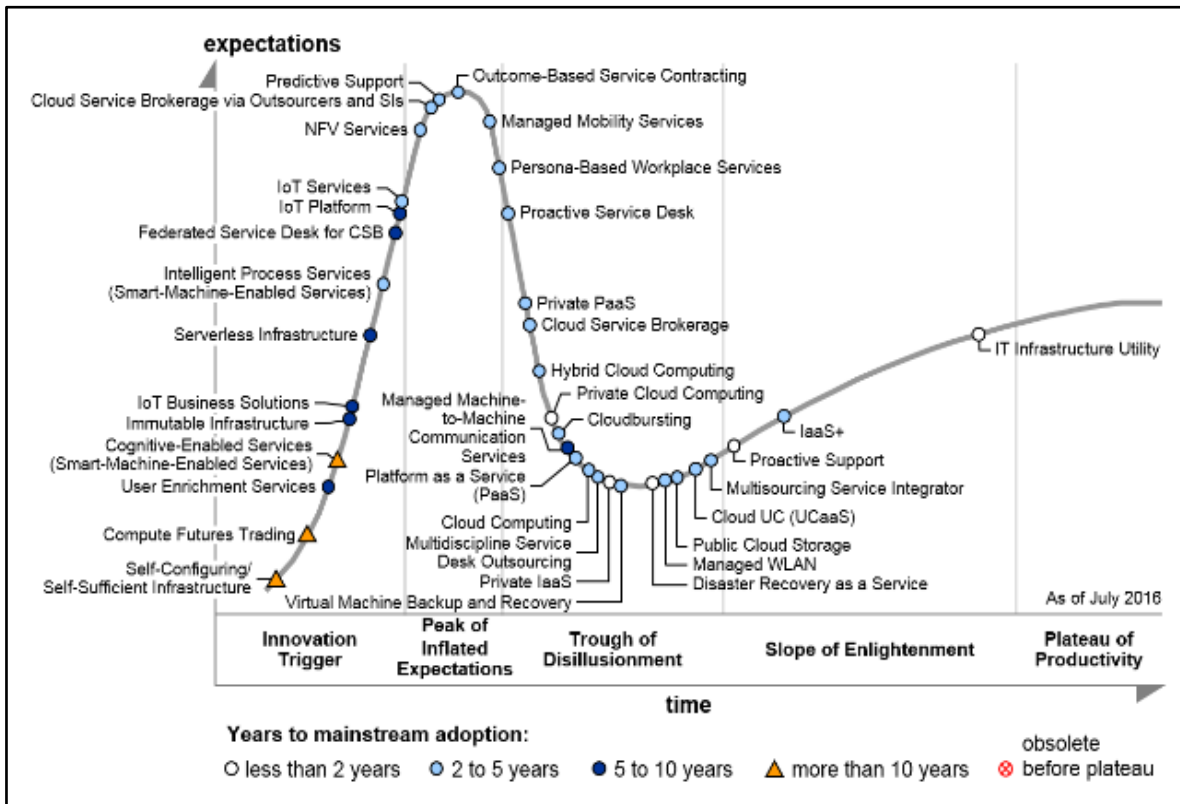


Figure 1-2 Étude Gartner Group sur les tendances dans les technologies de centres de données [3]

La Figure 1-2 démontre, à partir de l'étude de Gartner, que la transformation vers IaaS a bien dépassé la phase d'innovation et est une technologie mature avec un temps d'adoption de moins de deux ans. Elle se situe très proche de l'impartition (*IT Infrastructure utility*). Selon la même étude, les trois facteurs qui déclenchent la migration en nuage sont :

- L'économie des coûts opérationnels (48 %)
- L'extensibilité applicative (26 %)
- L'utilisation efficace de ressources (21 %)

La virtualisation est la création d'une version virtuelle d'un élément d'infrastructure, comme un serveur, un périphérique de stockage, un réseau ou un composant technologique comme un système d'exploitation qui peut être divisé dans un ou plusieurs environnements d'exécution. L'économie au niveau des solutions informatiques s'acquiert par le partage de

ressources à partir d'un modèle d'abstraction de ressources qui évite la dépendance sur les ressources physiques. La ressource physique est multiplexée pour plusieurs usagers et l'architecture d'infrastructure TI devient une architecture abstraite, définie comme logiciel SDDC. Cette couche d'abstraction est offerte par un système d'opération dédié nommé hyperviseur. L'hyperviseur est exécuté directement sur la machine physique appelée machine hôte et les machines virtuelles sont souvent appelées visiteur.

La virtualisation permet d'augmenter la densité des ordinateurs qui s'exécutent sur la même ressource physique. Une plus grande densité implique normalement une plus rapide migration des éléments d'infrastructure.

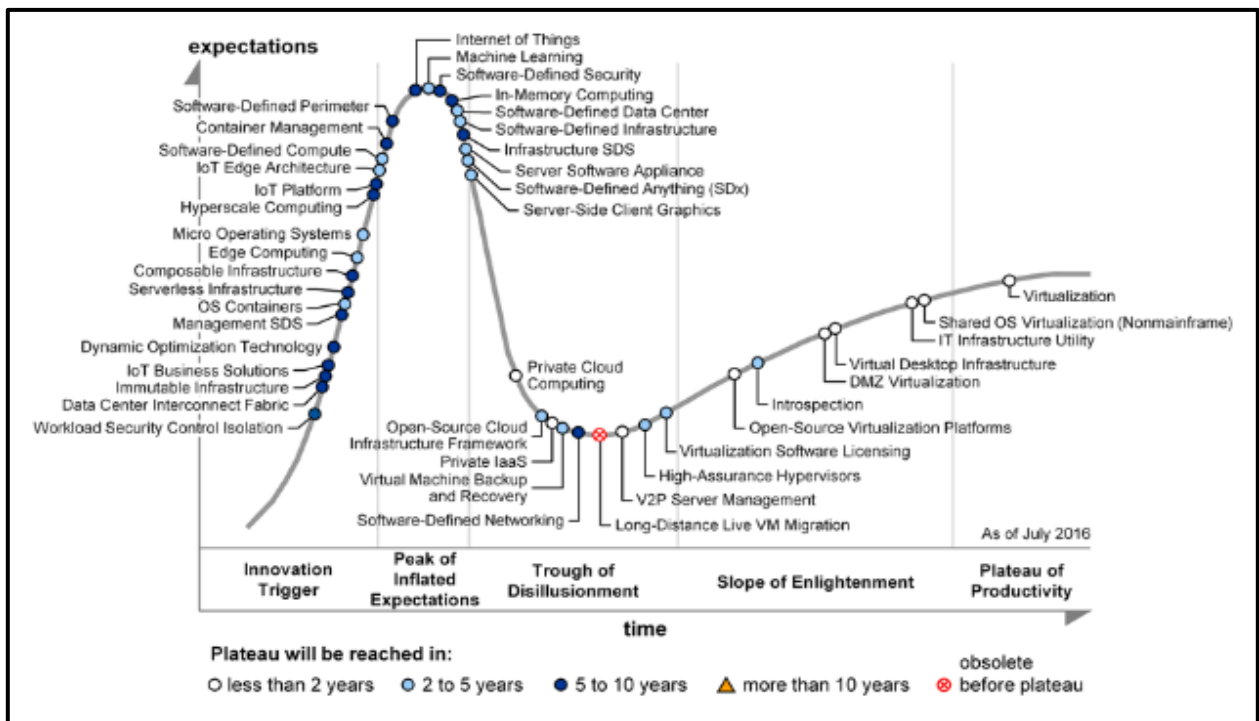


Figure 1-3 Virtualisation vers le plateau de productivité [4]

Il existe quatre types de virtualisation : native, assistée matériel, paravirtualisation et virtualisation de système d'opération. La virtualisation native simule le matériel pour permettre à un système d'opération visiteur de s'exécuter de façon isolée sans aucune modification de code. Dans la virtualisation assistée matériel, la machine a sa propre partition

de matériel qui permet l'isolation du système d'opération. La paravirtualisation est une technique dans laquelle l'hyperviseur fournit un API et le système d'exploitation de la machine virtuelle hôte appelle cet API. Cette technique impose des modifications dans le système d'exploitation visiteur. Le résultat offre une meilleure performance, mais avec le coût de ces modifications. La dernière technique est la conteneurisation, la virtualisation d'un serveur physique au niveau du système d'opération, qui donne la possibilité d'exécuter de multiples machines virtuelles isolées et sécurisées partageant la même machine physique et le même noyau de système d'opération. La virtualisation prend en charge la gestion de ressources physiques pour les systèmes d'opération(OS) et unifie leur utilisation dans une couche logicielle standard. La technique permet d'offrir des grappes de haute disponibilité, des aiguilleurs de charge, le déploiement en temps réel sans interruption, surtout grâce à cette unification qui fait que les équipements communiquent avec un seul langage. Le système d'opération peut être Windows ou Linux, avec des versions différentes, mais contrôlées par une seule version d'hyperviseur, capable d'unifier la gestion des ressources physiques. Les migrations de matériel et le système d'opération peuvent effectuer des mises à niveau progressivement. En anglais, le terme est *rolling upgrade*.

Les prochaines étapes franchies par la virtualisation sont les concepts SDN et le SDDC. Le SDN est une approche de mise en réseau informatique qui permet une abstraction et une unification de gestion des ressources réseau. SDDC arrive à gérer un parc informatique complet dans cette couche d'abstraction. La grande majorité des nuages informatiques sont gérés avec les concepts SDDC et IaC (Infrastructure as Code – Infrastructure comme code).

Tout élément d'infrastructure peut être virtualisé, et l'impact sur la migration se mesure sur des critères. Le séquençement des opérations établi lors du plan de migration a possiblement un impact sur les coûts de migration vers le nuage, dans un contexte de temps réel

1.2 Modélisation d'un système informatique

Actuellement, deux tendances d'architectures d'infrastructure sont populaires: multiniveau et MASA [3]. MASA est plus récent, mais le type d'infrastructure choisi est multiniveau, qui reste le plus populaire.

Une architecture typique multiniveau consiste en :

- a) Un ou plusieurs aiguilleurs de charge ;
- b) Deux ou plusieurs centres de données redondants ;
- c) Deux ou plusieurs serveurs web redondants, minimalement lweb1 et lweb2 situés dans des centres de données différents ;
- d) Deux serveurs applicatifs, minimalement disons ltom1 et ltom2 situés dans des centres de données différents ;
- e) Deux serveurs de bases de données, minimalement disons lora1 et lora2 situés dans des centres de données différents.

Des exemples des architectures multiniveaux sont présentés dans les figures suivantes:

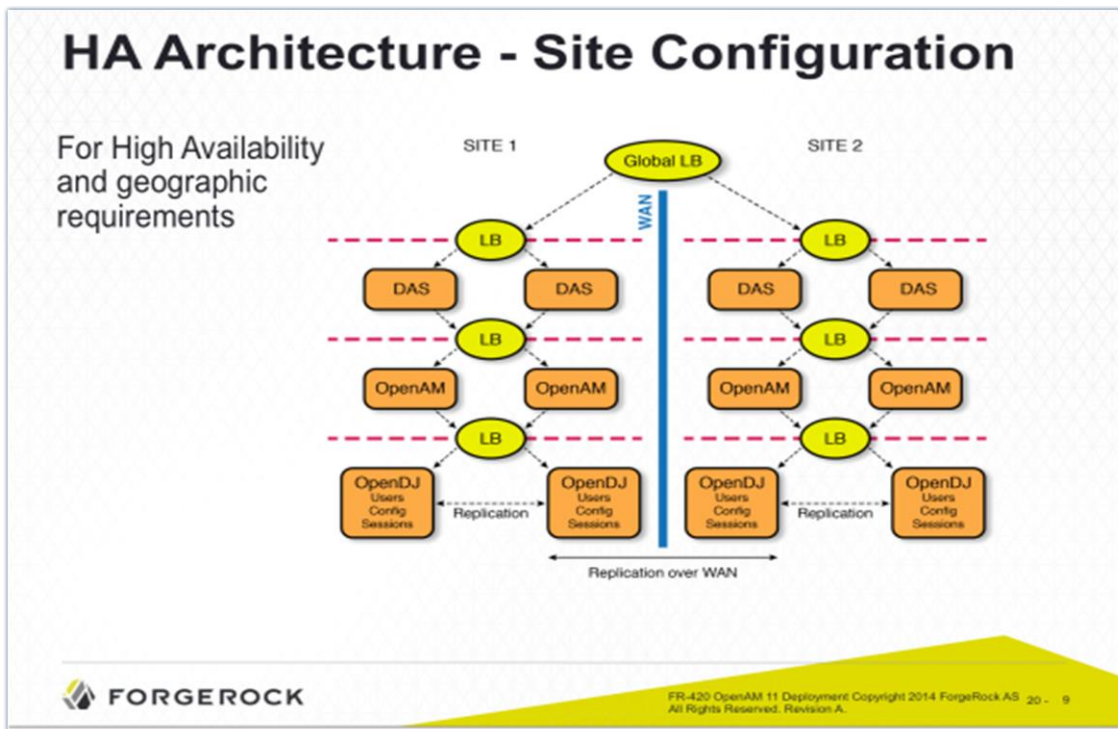


Figure 1-4 Architecture multiniveau avec redondance par centre et centres de données redondants [6]

Et aussi :

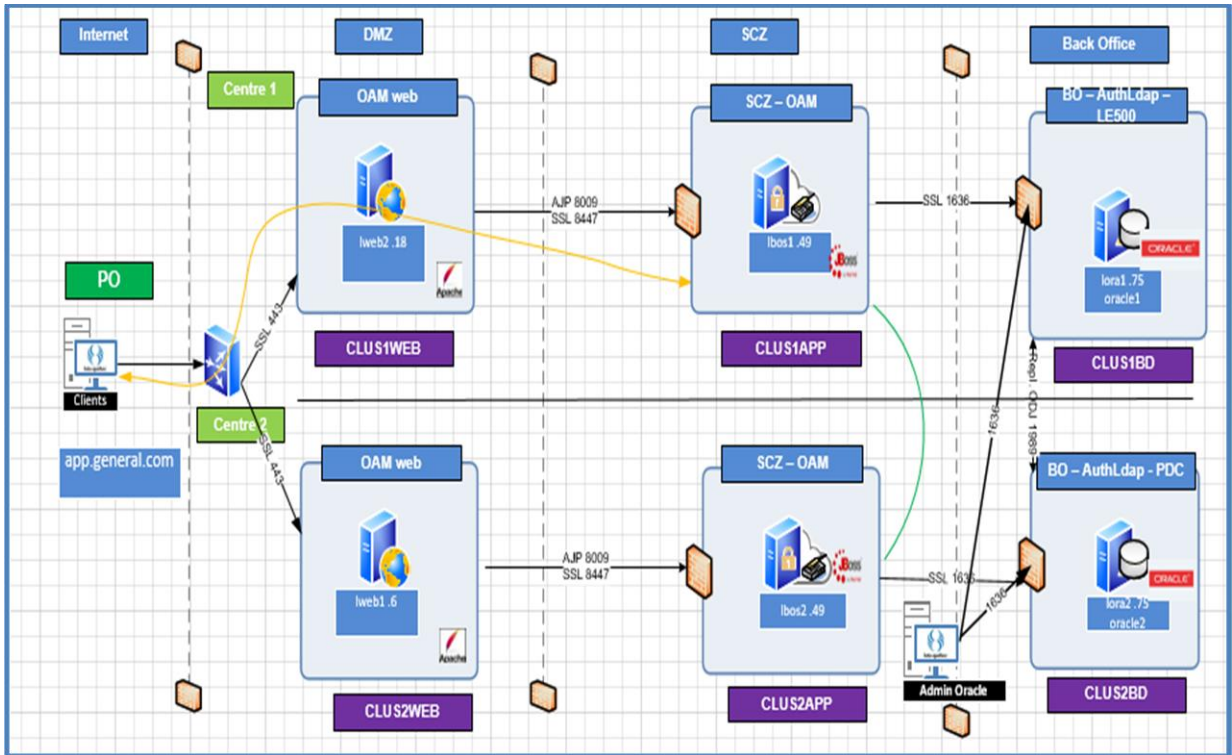


Figure 1-5 Architecture mult niveau sans redondance par centre et avec centres de données redondants [6]

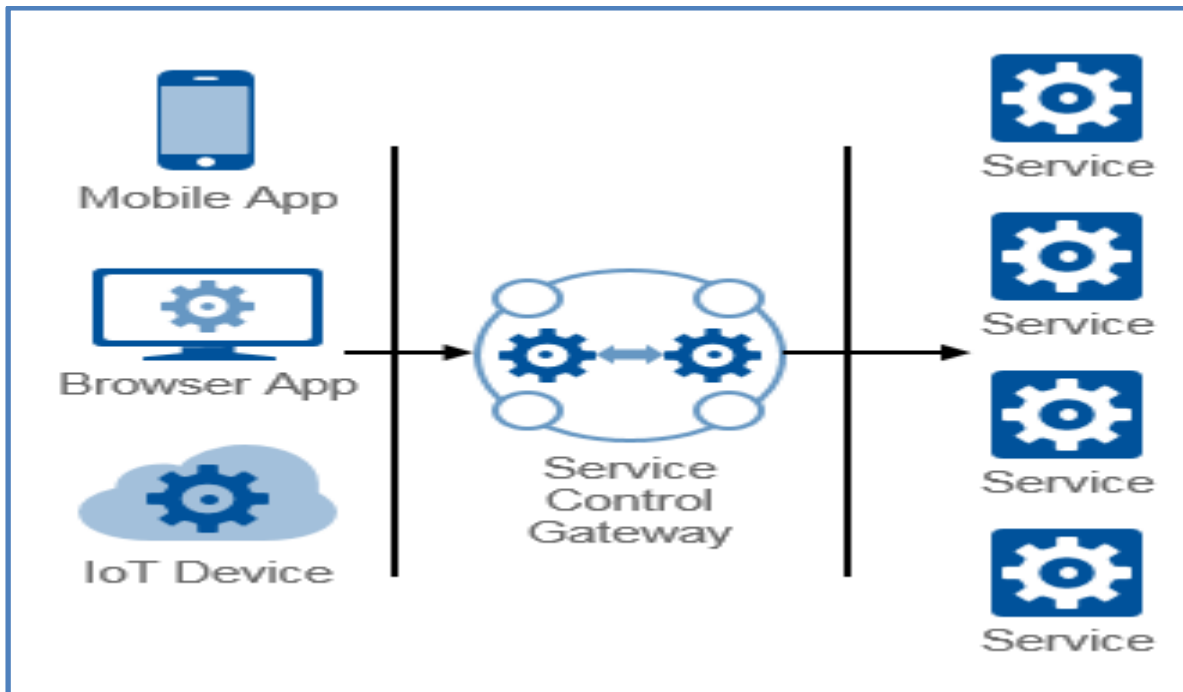


Figure 1-6 Architecture MASA [7]

L'architecture MASA est considérée selon Gartner le futur des architectures mobiles. Elle est constituée par des services qui sont situés sur des serveurs physiquement différents. Pour le cas de cette architecture, il y a seulement les composants technologiques liés aux systèmes logiques de type service. La modélisation utilisée dans cet essai a un grand niveau de généralité et couvre aussi cette architecture selon le même sondage effectué.

Pour estimer la maintenance d'un centre de données ou un système informatique, il faut d'abord identifier les éléments qui le composent. Dans le cadre du travail effectué pour cet essai, une partie importante concerne la modélisation d'un centre de données utilisant la méthode EAM.

La méthode d'architecture EAM considère que les entreprises et leurs systèmes informatiques sont organisés dans trois couches, une couche affaires, une couche logique et une couche physique. Chaque couche contient ses propres blocs architecturaux. La migration vers des solutions de type IaaS concerne la couche logique et la couche physique.

Voici, dans le tableau 1.1, les éléments composants de la couche logique :

Tableau 1.1 Tableau de systèmes logiques génériques

| # | Système logique | Description |
|----|--|--|
| 1 | Système logique | Système informatique qui effectue un ensemble de fonctions. Le plus haut niveau de définition. Ex. : système de ventes en ligne. |
| 2 | Dépôt de données | Sous-système générique qui gère de données persistantes. Ex. : les applications pl/sql et les tables d'une base de données Oracle. |
| 3 | Application | Sous-système applicatif générique qui gère de données non persistantes. Ex. : applications Java intergiciel |
| 4 | Équilibreur de charge (<i>Load Balancer</i>) | Sous-système qui gère l'équilibre de charge entre plusieurs systèmes |
| 5 | Mandataire (<i>Proxy</i>) | Sous-système qui filtre toutes les connexions vers l'extérieur de façon qu'elles semblent provenir de la même machine afin de protéger le réseau interne sous-jacent contre les intrus |
| 6 | CDN ou Réseau de diffusion de continu (<i>Content Delivery Network</i>) | Sous-système qui consiste à reproduire, sur un serveur situé à proximité des clients, les contenus qui comportent un fort volume d'informations, de telle sorte que ceux-ci leur sont acheminés plus rapidement, lorsqu'ils en font la demande |
| 7 | DNS | Sous-système qui gère un les domaines de système |
| 8 | Gestion des identités, accès et autorisation (<i>Identity access and authorization management</i>) | Sous-système destiné à gérer les identités, les accès et l'autorisation d'un système logique |
| 9 | File de messages (<i>Message Queue</i>) | Sous-système qui gère une file de messages |
| 10 | Pare-feu (<i>Firewall</i>) | Sous-système de pare-feu applicatif ou réseau |
| 11 | Stockage (<i>Storage</i>) | Sous-système de stockage de données |
| 12 | Messagerie | Système de courriel |
| 13 | Cache | Système destiné pour garder dans la mémoire volatile les plus fréquents objets utilisés |

Les systèmes logiques du Tableau 1.1 sont suffisants pour modéliser l'architecture logique d'une entreprise. Les interrelations sont de type « utilise » et « est utilisé ». Les systèmes logiques ont des composants technologiques utilisés pour remplir les fonctions. Un dépôt de données utilise le composant technologique logiciel pl/sql à une certaine version, par exemple oracle12c. Normalement, les composants technologiques des systèmes logiques sont soit hors de la boîte soit des logiciels maison.

Les blocs de type interfaces transportent des données persistantes entre deux systèmes logiques. Ils utilisent des composants technologiques, mais pour de raisons de clarté, ils sont présentés dans les modèles logiques.

La couche physique représente l'infrastructure d'un centre de données. Les centres de données sont composés d'éléments d'infrastructure tels que des serveurs, des serveurs applicatifs, des bases de données, des équipements de stockage et de réseau et d'autres éléments tels que présentés dans le Tableau 1.2.

Tableau 1.2 Éléments d'infrastructure d'un centre de données

| # | Élément Infrastructure | Type |
|----|--------------------------|-------------|
| 1 | Hyperviseur | Système |
| 2 | Machine physique | Système |
| 3 | Machine virtuelle | Système |
| 4 | Cluster OS | Système |
| 5 | Cluster Hyperviseur | Système |
| 6 | Conteneur | Système |
| 7 | Cluster Conteneur | Système |
| 8 | Instance applicative | Intergiciel |
| 9 | Cluster applicatif | Intergiciel |
| 10 | Instance Base de données | BD |
| 11 | Cluster BD | BD |
| 12 | BD Physique | BD |
| 13 | LUN | Stockage |
| 14 | Cluster Lun | Stockage |
| 15 | Baie de stockage | Stockage |
| 16 | Contrôleur Stockage | Stockage |
| 17 | Équipement Réseau | Réseau |

Le tableau présente qu'une partie des éléments d'infrastructures. Une liste plus exhaustive est présentée en Annexe I.

La liaison entre les éléments d'infrastructure et les systèmes logiques est effectuée avec des interactions de type : « s'exécute sur ». Un dépôt de données s'exécute sur un élément d'infrastructure de type instance de base de données et sur la base de données physique. Les éléments d'infrastructure sont aussi connectés avec des relations de type « utilise » ou « est utilisé ». Une instance de base de données utilise une base de données physique qui utilise un élément de stockage. Les relations « utilise » et « est utilisé » se présentent selon

le principe suivant : bloc #1 utilise le bloc #2 si le bloc #1 cesse de fonctionner alors bloc #2 cesse de fonctionner, mais pas l'inverse. Cette approche permet d'effectuer une analyse d'impacts et une décomposition des systèmes logiques pour trouver les éléments d'infrastructure qui migrent vers le nuage.

1.3 Migration en nuage informatique

Les systèmes logiques sont classifiés par Gartner en systèmes de type Mode 1 et Mode 2. Mode 1 est traditionnel et séquentiel, mettant l'accent sur la sécurité et la précision. Mode 2 est exploratoire et non linéaire, mettant l'accent sur l'agilité et la vitesse et sont les applications les plus propices à migrer dans le nuage. Chaque migration est mesurée sur trois variables qui doivent être minimisées : le temps de migration, le temps d'arrêt du système et le coût.

Dans certains cas, il faut minimiser l'interruption de service pendant la migration. Il faut aussi réduire la surcharge supplémentaire causée par la période de migration quand la bande réseau est intensivement utilisée que certains éléments d'infrastructure sont doublés et que des données supplémentaires sont migrées. Finalement, il faut payer des frais supplémentaires pendant la migration parce que pour cette période les éléments d'infrastructure sont doublés (une instance dans le centre de données et une instance dans le nuage). Il faut s'assurer que le temps d'amortissement est correct alors que le nuage offre aussi la performance et l'élasticité envisagée. Il est donc avantageux de minimiser les coûts autant que possible.

Les acronymes P2V (physique vers virtuel), P2C (physique vers nuage), V2C (virtuel vers nuage) et les acronymes inverses (une migration de nuage vers un centre de données local est absolument possible), V2P, C2P, C2V sont des concepts à considérer lors du processus de migration. La différence donnée par le degré de virtualisation est importante dans les mesures de migration parce que la densité supérieure de ressources dans un environnement virtuel et les procédures de contrôle de consommation de ressources assurent, dans situations particulières, plus de rapidité dans une migration. Il est possible de contrôler les variables qui déterminent l'efficacité d'une migration, c'est-à-dire la consommation CPU,

mémoire, disque et réseau. Et finalement, il est possible, dans certains cas, que le coût de P2V plus le coût V2C soit plus petit que le coût de P2C. Autrement dit, dans certains cas, c'est plus efficace de prévirtualiser un centre de données avant de migrer dans le nuage.

La littérature [2] mentionne 4 types de migration : Type I (remplacer le tout) implique que les systèmes sont remplacés avec des solutions nuage informatique SaaS. C'est le type de migration la moins invasive. Les systèmes logiques sont intégralement migrés vers le nuage. Ce type de migration nécessite une série de configurations pour régler les incompatibilités à utiliser les fonctionnalités de la couche en charge. Le type II (migration partielle) implique la migration de certains des logiciels et composants vers le nuage. À titre d'exemple, la migration de la fonctionnalité d'audit d'un système dans le nuage peut être classée comme un type II de migration. Le type III (migrer l'ensemble de la pile d'applications) est le type de migration le plus facile, car toute l'application est monolithique et encapsulée dans une ou plusieurs machines virtuelles en cours d'exécution sur le nuage IaaS. Ce type de migration ne nécessite aucune adaptation en supposant que la pile d'applications peut être portée telle quelle dans une machine virtuelle. Enfin, le type IV (nuage natif) représente la migration complète, l'application est convertie dans un système en nuage entièrement recomposé des blocs d'architectures natifs vendeur dans une solution PaaS.

Il est possible de migrer dans le nuage en utilisant trois techniques. La première technique implique des arrêts. La deuxième technique implique des répliques itératives, sans arrêt. Une troisième technique hybride se situe entre les deux premières, impliquant des temps d'arrêt presque nuls.

Dans la première technique, le système cesse de servir les mises à jour, déplace une image persistante et volatile dans le nuage via le réseau et redémarre le système à la destination. Cette technique entraîne une interruption de service. Le transfert s'effectue sur le réseau en utilisant un matériel qui est envoyé avec la copie des données au fournisseur de service infonuagique. Le système démarre dans le nuage quand la copie est terminée. L'avantage de cette technique est sa simplicité.

La réplique itérative est plus complexe. Une copie initiale des données persistantes est effectuée sans arrêter les systèmes. Une fois copiée dans le nuage, la synchronisation démarre. Cependant le système source continue à journaliser les changements qui sont

envoyés dans le nuage par la synchronisation. Une fois la synchronisation terminée, les applications sont déplacées dans le nuage avec un temps d'arrêt presque nul.

Pour trouver la meilleure méthode de migration, Togaf est utilisé. Togaf est un ensemble de concepts et un standard industriel couvrant le domaine des architectures informatiques d'entreprise. Le Cycle ADM définit les bonnes pratiques pour développer l'architecture d'entreprise d'une organisation. Il est constitué d'une phase préliminaire suivie de huit phases, permettant de construire, entre autres, l'architecture technique, de planifier le déploiement, de la mettre en œuvre et finalement, de gérer les changements.

- 1) Vision de l'architecture
- 2) Architecture affaires
- 3) Architecture des systèmes d'information
- 4) Architecture technologique
- 5) Opportunités et solutions
- 6) Planification de migration
- 7) Gestion de l'implémentation
- 8) Gestion du changement d'architecture

Selon J. Lin [4], une migration en nuage qui utilise Togaf contient six phases :

- 1) Identification de l'architecture source
- 2) Identification de l'architecture cible
- 3) Identification des vendeurs de solutions infonuagiques
- 4) Sélection des vendeurs
- 5) Plan de migration
- 6) Migration et test

Il y a beaucoup de documentation pour prédire les coûts d'hébergement d'un système dans le nuage après avoir migré les systèmes logiques, mais très peu d'études ciblent le coût de la migration. Il manque de modèles, pour aider à quantifier et optimiser le coût d'une migration de façon détaillée. Un modèle est nécessaire pour estimer les coûts de migration. Ce modèle doit être validé par des experts à l'aide d'une méthode formelle. La méthode vise à organiser

la consultation des experts sur un sujet choisi. Les experts sont impliqués dans chaque étape de la recherche, lisent le contenu écrit, répondent aux questionnaires, participent aux réunions, pour établir un consensus sur les zones d'incertitude et converger vers une opinion commune.

Cette méthode qualitative a deux caractéristiques communes: la rétroaction et la réponse statistique. L'anonymat assure l'objectivité, la rétroaction implique l'expert dans chaque étape de la recherche et la décision est prise statistiquement. Un facilitateur, normalement l'auteur de l'essai, coordonne le processus et modère les opinions de participants.

Chapitre 2

Revue de la littérature

Ce chapitre vise à présenter les principales notions théoriques nécessaires à la compréhension du problème de recherche. En effet, il s'articule autour de trois sections. La première section présente la méthodologie de recherche utilisée, les sources bibliographiques et les mots-clés utilisés. La seconde section expose le vécu de l'industrie de déploiement en nuage informatique en termes de coûts de migration. La dernière section fait appel aux livres blancs existants sur les sites de vendeurs connus. Cela permet de connaître leurs stratégies de coût ainsi que leurs solutions infonuagiques.

2.1 Méthodologie de recherche

La recherche cible des articles scientifiques et des livres de référence sur les méthodes d'estimation des coûts de migration vers le nuage. Les mots-clés et les sources bibliographiques sont présentés et analysés ainsi que la valeur scientifique de chaque source.

2.1.1 Mots-clés

Les mots clés utilisés sélectionnés sont groupés dans quatre catégories principales selon le plan énoncé dans la mise en contexte : centre de données en nuage informatique, virtualisation, migration en nuage et coût de migration en nuage. La grande majorité des documents sont en anglais. Les mots clés pour la catégorie de nuage informatique sont :

- a) Blocs d'architecture d'un centre de données (*Data Center Building Blocks*) [5] [6]
- b) Méthodologie EAM Iteraplan (*EAM Iteraplan*)
- c) Patrons de nuage (*Cloud Patterns*) [7] [8]
- d) Architecture nuage (*Cloud architecture*) [5] [9] [10] [11]
- e) Blocs d'architectures dans le nuage (*Cloud Building Blocks*) [3] [12]

- f) Architecture de nuage sans serveur (*Cloud serverless architecture*) [13]
- g) Centre de données défini logiciel (*Software defined data center*) [14]
- h) Réseau défini par logiciel (*Software defined network*) [15] [8]
- i) Consommation de ressources en nuage (*Cloud Ressource Consumption*) [5]
- j) Centre de données avec haute disponibilité (*High Availability Data Centers*) [14]

La virtualisation a été analysée en 1.2 Les termes clés de recherche sont :

- a) Types de virtualisation (*Virtualisation types, Virtualisation categories*) [16]
- b) Virtualisation réseau (*Network Virtualisation*) [14]
- c) Coût de la virtualisation (*Virtualisation Cost*) [5]
- d) Conteneurs (*Containers, Light Virtualisation*) [11]
- e) Grappes virtuelles (*Virtual Clusters*) [14]
- f) Consommation et gestion de ressources dans les machines virtuelles (*Resource management virtual machines*) [11]
- g) Solutions virtuelles de haute disponibilité (*High Availability Virtual solutions*) [17]
- h) Migration en temps réel des éléments d'infrastructures virtuels. (*Live migration virtual machines*) [9] [18]

Pour la migration en nuage, les termes clés sont :

- a) Migration en nuage (*Cloud Migration*) [18]
- b) Méthodologie Togaf appliquée en nuage (*Togaf Cloud*) [18] [4]
- c) Migration en temps réel en nuage (*Live migration in cloud*) [18]
- d) P2V, V2C, C2V, V2P, P2C, C2P [18] [12]

Finalement le coût, la plus compliquée partie de cet essai, est couvert par cette séquence de mots clés :

- a) Coût de nuage informatique (*Cloud Cost*) [17] [5] [19] [20]
- b) Gestion de nuage informatique (*Cloud Management*) [5]
- c) Calculatrice de coût de nuage (*Cloud Cost Calculator*) [21]
- d) Analyses de coûts (*Cost Analysis*) [5]
- e) Analyses de coûts et avantages de nuage (*Cost benefit analysis in cloud*) [5]

- f) Points de fonction dans le nuage informatique (*Cloud Migration Points, Function Points Method*) [5]
- g) Mesures de migration en nuage (*Cloud Migration Measures or Measurements*) [5]
- h) Risques de migration en nuage (*Cloud Risk Migration*) [17]

2.1.2 Sources bibliographiques

L'environnement infonuagique change et s'améliore chaque année. La technologie évolue offrant de meilleures performances à des coûts moindres. Afin de procéder à la revue de littérature, les sources bibliographiques principales utilisées sont :

- articles scientifiques provenant de site <http://scholar.google.com> [22] [23] [19] [24] [20] [25] [26] [27] [4]
- cours en ligne offerts par Edx, Udacity et Coursera pour une mise à jour permanente de connaissances sur le sujet de nuage informatique, virtualisation et migration [9] [10] [11] [14]
- la méthodologie Togaf pour modéliser la migration d'un centre de données [28] [4]
- la méthodologie EAM de Iteraplan pour la modélisation d'un centre de données
- une sélection de livres et documents académiques [5] [12] [6]
- livres blancs de sites Amazon et Azure [29] [3]
- articles sur le site Gartner Group [30]

2.2 Coûts de migration vers le nuage

Les prochains paragraphes définissent les modèles et les analyses de coûts de migrations vers le nuage. Les références principales étudiées sont : [5], [20] et [24].

Le choix de la référence [5] est basé sur l'approche théorique et pratique proposée. Les auteurs adoptent une approche théorique et pratique pour calculer le coût d'une migration dans le nuage. L'ouvrage établit la liaison entre la technologie, la gestion de coûts et le modèle IaaS. Le chapitre 16 du livre est particulièrement pertinent. Il s'intitule Rétrofacturation (en anglais *Chargeback*) et il explique comment calculer le coût opérationnel

par utilisation de ressources ou abonnement ainsi que les modèles de coût du nuage informatique.

Le coût opérationnel pour l'utilisation de ressources est calculé sur une approche strictement basée sur l'infrastructure informatique de consommation. En règle générale, une série de paramètres est mesurée sur les éléments d'infrastructure pour déterminer le coût d'un environnement informatique migré en nuage. Un utilisateur peut utiliser une machine virtuelle pour héberger un serveur web, qui consomme très peu des ressources, tandis qu'un autre utilisateur peut utiliser une machine virtuelle d'un serveur de bases de données qui consomme presque toutes les ressources disponibles. Le fournisseur de services infonuagiques peut différencier ces deux cas de consommation de ressources particulières avec différentes charges.

C'est une méthode statique ou préconfigurée d'allocation de ressources. Si un consommateur ne peut pas prévoir ou estimer la charge de travail, un des moyens est d'utiliser une méthode d'essai et d'erreur risquée. Par contre, un avantage de la répartition des coûts basée sur l'utilisation des ressources est qu'il peut contrôler les demandes excessives. Il peut également gérer le comportement utilisateur dans la façon dont il consomme des ressources informatiques d'infrastructure.

La référence [5] donne aussi une liste de ressources payables d'infrastructure, prix possibles et unités de mesure :

Tableau 2.1 Tableau de ressources payables dans le nuage informatique [5]

| # | Ressource | Unité de mesure | Prix possible (\$) |
|---|--------------------|-----------------|--------------------|
| 1 | vCPU | GHz/heure | 0,045 |
| 2 | Memoire | Gb/heure | 0,1 |
| 3 | Entrée disk | Heure | 0,06 |
| 4 | Sortie disk | Heure | 0,06 |
| 5 | Entrée/Sortie disk | Heure | 0,1 |
| 6 | Entrée Réseau | Gb/heure | 0,1 |
| 7 | Sortie Réseau | Gb/heure | 0,1 |

| | | | |
|---|-------------------------|----------|-------|
| 8 | Entrée/Sortie reseau | Gb/heure | 0,1 |
| 9 | Stockage | Gb/heure | 0,150 |

Le coût par souscription est habituellement mensuel. Les frais uniques sont les frais d'installation. Le coût total est composé par un coût de souscription, un coût par ressource et les frais uniques (frais d'installation et configuration).

La référence [20] trouvée est la plus importante pour cet essai. La méthode de points de migration au nuage essaie d'abstraire l'effort de migration avec une méthode appelée CMP (méthode de points de fonction nuage).

Selon [20] les facteurs qui déterminent, le coût d'une migration en nuage sont :

- 1) Installation et configuration - Lors d'une migration vers un nuage IaaS l'effort est nécessaire pour installer et configurer les éléments d'infrastructure du système nécessaires (serveurs de base de données et intergiciel et les environnements).
- 2) Les modifications de bases de données – Une migration d'une base de données vers le nuage peut entraîner des changements de schéma de base de données et les changements de la requête en raison de différences de versions, variantes (MySQL et MSSQL), ou des types de bases de données (relationnelles ou NoSQL).
- 3) Les changements de code - Dans certains cas, le changement de code est nécessaire pour s'adapter au nouveau modèle de programmation dans les nuages.
- 4) Changement de connexions - Au sein d'un système S avant la migration, la connexion entre deux éléments A et B est une connexion LAN. Si seulement B est migré vers le nuage et est A est conservé dans un centre de données locales, la connexion LAN entre A et B devient une connexion WAN. Si A et B sont migrés vers le nuage, la connexion LAN entre A et B devient une connexion LAN (réseau infonuagique dans les conditions peuvent être différentes dans les nuages).

Les auteurs proposent un tableau de points de migration avec pondérations selon le degré de complexité et la pondération de chaque facteur mentionné.

| CMP | Type | Complexity Level | | | Weight |
|---------------------|--------------------|------------------|---------|------|--------|
| | | Low | Average | High | |
| CMP _{conn} | LAN-to-LAN | 1 | 3 | 4 | 3 |
| | LAN-to-WAN | 1 | 6 | 9 | |
| | WAN-to-LAN | 1 | 6 | 9 | |
| CMP _{code} | PDT | 3 | 6 | 10 | 5 |
| | HIT | 4 | 7 | 12 | |
| | DMT | 5 | 8 | 13 | |
| | TMT | 4 | 6 | 9 | |
| CMP _{ic} | Application | 1 | 2 | 7 | 2 |
| | Infrastructure | 1 | 3 | 9 | |
| CMP _{db} | Query Modification | 1 | 3 | 8 | 1 |
| | Data Population | 3 | 4 | 10 | |

Figure 2-1 CMPs selon [20]

La formule proposée pour calculer l'effort de la migration est la somme de tous ces facteurs :

$$CMP = \sum_{i=1}^4 CMP_i \times W_i \quad (2.1)$$

$$= CMP_{conn} \times W_{conn} + CMP_{code} \times W_{code} + CMP_{ic} \times W_{ic} \\ + CMP_{db} \times W_{db}$$

La référence [24] est un répertoire utile de plusieurs analyses académiques effectuées sur les migrations dans le nuage informatique. Les catégories analysées sont les processus d'une migration en nuage :

- 1) Processus de planification de migration telle que l'étude de faisabilité, l'analyse des besoins de migration, ainsi que certaines prises de décisions concernant le fournisseur à choisir, les sous-systèmes à migrer, les services infonuagiques à utiliser, et l'élaboration d'une stratégie de migration. L'artéfact livré est le plan de migration.

- 2) L'exécution de la migration. Dans ce deuxième processus, la migration effective des tâches telles que l'extraction de données, l'architecture de recouvrement ainsi que des modifications de code à exécuter.
- 3) Évaluation de migration. Dans ce troisième processus, le système migré est prêt à l'emploi et à la validation. Les tâches de test, de validation et de déploiement des applications sont effectuées.
- 4) Préoccupations transversales. Certaines tâches transversales comme la gouvernance, l'analyse de la sécurité, la formation, l'effort d'estimation, le changement organisationnel, l'analyse de l'élasticité ainsi que les activités de coordination.

La référence [24] est utile surtout pour les références vers d'autres études de qualité qui analysent ces processus.

2.3 Fournisseurs de solutions infonuagiques

Les paragraphes suivants analysent les principaux vendeurs sur le marché, leurs équipements typiques d'infrastructure, les blocs d'architecture physiques et logiciels ainsi que les patrons de migration dans le nuage informatique. La littérature est principalement composée des livres et d'articles dédiés et le but est d'extraire les paradigmes, les modèles courants de pensées qui s'appliquent sur le sujet de cet essai.

2.3.1 Amazon AWS

AWS (Amazon Web Services) offre une architecture orientée service composée de blocs d'architecture type PaaS, IaaS, SaaS [13]. Le modèle PaaS est mappé dans l'annexe IV dédiée à un début de modélisation de coûts Amazon PaaS.

Les éléments d'infrastructure modélisés IaaS en 1.3 sont mappables en :

- 1) Amazon EC2 (machines, conteneurs, instances applicatives, instances BD et BD physiques)
- 2) Amazon RDS, Redshift (certaines instances et bases de données relationnelles)

- 3) Amazon Dynamo DB, Simple DB, MapReduce (instances et bases de données non relationnelles)
- 4) Amazon S3 pour les éléments de stockage
- 5) Les éléments d'infrastructure réseau migrent surtout en Amazon EC2

EC2 (*Elastic Compute Cloud*) [31] est un service Web avec une interface simple pour lancer les instances d'une application sous plusieurs systèmes d'exploitation, tels que plusieurs distributions Linux, Windows, OpenSolaris, FreeBSD, et NetBSD. Une instance est créée soit à partir d'une image prédéfinie (AMI) signée numériquement et stockée dans S3 (élément de stockage) ou d'une image définie par l'utilisateur. L'image comprend le système d'exploitation, l'environnement d'exécution, les bibliothèques API et l'application souhaitée par l'utilisateur. Un utilisateur peut : (a) lancer une instance d'un AMI existant et mettre fin à une instance ; (b) démarrer et arrêter une instance ; (c) créer une nouvelle image ; (d) ajouter des balises pour identifier une image ; et (e) redémarrer une instance.

EC2 est basée sur la stratégie de virtualisation Xen. Chaque machine virtuelle ou instance fonctionne comme un serveur privé virtuel. Un utilisateur peut interagir avec EC2 à l'aide d'un ensemble de messages SOAP. Les instances peuvent être placées à différents endroits dans différentes régions et zones de disponibilité.

EC2 permet l'importation des images de machines virtuelles à partir de l'environnement utilisateur pour une instance et distribuer automatiquement le trafic d'application entrant entre plusieurs instances à l'aide de l'équilibrage de charge élastique. EC2 associe une adresse IP fixe à un compte ; ce mécanisme permet de masquer la défaillance d'une instance par une relocalisation dynamique d'une adresse IP publique à toute instance du compte, sans interagir avec l'équipe de support logiciel.

Le système de stockage simple (S3) [32] est un service de stockage conçu pour stocker des objets. Il prend en charge un ensemble minimal de fonctions : écrire, lire et supprimer. S3 permet à une application de gérer un nombre illimité d'objets allant de la taille d'un octet à téraoctets.

Amazon S3 SLA garantit la fiabilité et utilise des interfaces REST et SOAP et le protocole de téléchargement par défaut est HTTP.

Un autre élément de stockage, EBS fournit des services permanents au niveau bloc, les volumes de stockage pour une utilisation avec les instances Amazon EC2. Un volume semble à une application comme la matière non formatée et fiable d'un disque physique. La taille du volume de stockage varie de 1 Go à 1 To. Les volumes sont regroupés en zones de disponibilité et sont automatiquement répliqués dans chaque zone. Une instance EC2 peut monter plusieurs volumes, mais un volume ne peut pas être partagé entre plusieurs instances. EBS prend en charge la création d'instantanés (*snapshots*) des volumes associés à une instance, puis les utilise pour redémarrer une instance. La stratégie de stockage fourni par EBS est adaptée pour les applications de bases de données, systèmes de fichiers et les applications qui roulent à l'aide des données brutes.

En tant qu'laaS, Amazon offre les régions assimilables avec les centres de données et les zones de disponibilités assimilables avec les zones physiquement séparées. Le reste, comme présenté, est la machine de calcul, le stockage, la base de données et le réseau.

Le document [13] fournit les repères principaux d'une migration dans le nuage. AWS Storage Gateway est un élément de stockage utilisé pour les sauvegardes. AWS Snowball est un dispositif adapté pour les migrations de grandes quantités de données de type *fork and lift*. AWS Direct Connect est la solution pour une migration de données utilisant le réseau WAN. Amazon S3 XA est destiné aux migrations rapides de données stockage. Finalement Amazon Kinesis capture les changements et les envoie dans le nuage.

La migration en nuage sans temps d'arrêt est modélisable facilement avec ces solutions. Le modèle de prix d'Amazon est basé sur l'élément EC2 et il est calculé face cette formule :

Prix = Prix EC2 (grandeur de la machine) + Prix Stockage (persistant ou volatile) + Prix réseau (gratuit à l'intérieur de la même zone de disponibilité et gratuite pour les données entrantes) + Prix logiciel installé (Oracle, MySQL) + Prix des options de services additionnelles (surveillance ou équilibrage de charges).

Les éléments d'infrastructure des centres de données sont facilement retrouvables dans des solutions Amazon laaS.

2.3.2 Microsoft Azure

Microsoft Azure [33] offre aussi une architecture orientée service ou on peut distinguer PaaS, IaaS, SaaS. Azure IaaS est constitué de trois composants clés : calcul (*Compute*), qui fournit un environnement de calcul ; le stockage pour un stockage évolutif ; le contrôleur (*Fabric Controller*), qui déploie, gère et surveille les applications ; en reliant des serveurs avec des connexions à haute vitesse et des commutateurs. Le réseau de diffusion de contenu (CDN) conserve des copies de données de cache pour accélérer les calculs. Le sous-système Connect prend en charge les connexions IP entre les utilisateurs et leurs applications s'exécutant sur Windows Azure

Le mappage IaaS est assez simple, une combinaison machine virtuelle (calcul), stockage et réseau similaire à Amazon. Les extractions de données de Microsoft Azure sont payées, pendant que les transferts de données ne sont pas payés. La facturation est mensuelle.

En ce qui concerne la migration Microsoft offre un accélérateur MA qui possède les caractéristiques suivantes :

- 1) Hétérogénéité : possibilité de migrer les charges d'exécution d'un large éventail de plates-formes telles que VMware, Microsoft Hyper-V, Amazon Web Services et/ou serveurs physiques.
- 2) Automatisation : permission de découvrir automatiquement les charges de l'entreprise à distance, à partir de nuage. MA permet de tester la charge de travail dans le nuage sans impact sur la charge de travail de production.
- 3) Migration d'applications multiniveaux : MA possède la capacité de migrer le système de production multiniveaux avec cohérence.
- 4) Réplication continue pour moins de temps de transfert : MA pour Azure fournit une réplication du système y compris l'exploitation et l'application des données. Cette réplication permanente et le suivi des modifications en mémoire réduisent le temps de transfert et mettent au minimum l'impact sur les charges de travail.

Autant Microsoft qu'Amazon ne donnent pas dans cet outil MA la possibilité de simuler et calculer le prix de la migration dans les conditions flexibles désirées par un client.

Chapitre 3

Problématique

3.1 Description

La stratégie est définie comme un « ensemble d'actions coordonnées en vue d'atteindre un objectif précis » [34]. L'objectif de cet essai est de valider la stratégie de migration d'un centre de données dans un environnement infonuagique et de répondre si l'identification d'une meilleure stratégie permet des économies importantes.

Le but précis est de vérifier si l'ordre de migration a une influence sur le coût. La revue de littérature présentée au Chapitre 2 fait l'état des connaissances sur les stratégies de migrations possibles basées sur l'ordre de migration et la configuration des éléments d'infrastructure migrés dans le nuage. Il existe plusieurs autres éléments à considérer, dont les compétences et l'acceptation des spécialistes impliqués, la complexité des systèmes, le degré de virtualisation, mais qui ne font pas l'objet de cet essai.

Plusieurs entreprises optent pour des solutions infonuagiques dues aux avantages que celles-ci apportent. Toutefois, l'opération de migration engendre des coûts non négligeables, surtout lorsque la disponibilité est un enjeu. La stratégie de migration a aussi un impact sur les coûts de celle-ci. Le fait de migrer un serveur avant l'autre influence les coûts en fonction du flux de données entre les serveurs. Les données qui circulent entre le nuage et l'extérieur de ce dernier sont facturables. Ainsi, la stratégie utilisée influence les coûts. Le but de cette recherche est d'identifier les facteurs à considérer pour calculer ces coûts et trouver le meilleur ordre de séquençement de migration, dans un contexte de système à haute disponibilité.

Cet essai s'adresse aux spécialistes de migration dans des environnements infonuagiques ainsi qu'au personnel responsable des infrastructures TI. Un modèle permettant d'estimer le

coût de migration d'un système logique aide à la prise de décisions et sauve du temps et argent.

Les concepts clés recherchés sont le nuage, les systèmes logiques, les centres de données, les éléments d'infrastructure, le coût de migration en nuage et finalement l'optimisation des coûts. La modélisation de tous ces concepts sur un cas réel permet d'analyser des scénarios et vérifier si cette séquence permet de réduire les coûts de migration et ainsi obtenir une méthode pour identifier la meilleure séquence.

Une de difficultés rencontrées par les entreprises est le manque d'information concernant les coûts de migration vers le nuage informatique. Il existe énormément d'information pour les frais d'exploitation, mais très peu sur la période de migration. C'est pourquoi la solution proposée est d'estimer une formule à partir de [20], et d'identifier un modèle générique de migration selon un algorithme qui minimise le coût.

3.1.1 Question de recherche

La question se formule ainsi : est-ce qu'un ordre de séquençement de migration des infrastructures réduit les coûts de migration en environnement infonuagique, dans un contexte de haute disponibilité où les systèmes sont complexes? Pour répondre, il faut trouver le modèle formel qui permet d'estimer et de minimiser le coût de migration d'un centre de données, dans un contexte de haute disponibilité. Le modèle doit résister aux tests et recevoir la validation des experts.

L'hypothèse est qu'un ordre de séquençement de migration des infrastructures réduit les coûts de migration en environnement infonuagique, dans un contexte de haute disponibilité. Dans un contexte de haute disponibilité, les infrastructures sont migrées une à la suite des autres pour assurer la continuité du service. Puisque les coûts sont attribuables aux transferts de données entre l'environnement interne et externe au nuage, ils sont influencés par l'ordre de migration des infrastructures dans le nuage. Dans des environnements complexes, l'ordre risque d'influencer de façon plus importante les coûts de migration. Le présent essai se limite à trouver un modèle qui permet d'identifier le meilleur ordre de séquençement pour réduire les coûts.

Le contexte de la recherche est une migration en nuage informatique de type IaaS. La migration doit maintenir au maximum la disponibilité des applications (si possible migration en temps réel). Les solutions informatiques analysées sont celles des vendeurs Amazon et Azure. Aussi, afin de diminuer le risque au niveau de la disponibilité, la stratégie retenue est de migrer un élément d'infrastructure à la fois.

Le système se migre vers le nuage en respectant une séquence des opérations à partir de son architecture source vers son architecture cible. Pendant la migration l'infrastructure de système migré passe par des architectures intermédiaires.

La variable indépendante est la séquence de migration des éléments d'infrastructure. La variable dépendante est donc le coût de la migration, c'est cette variable qui est mesurée. En procédant à l'expérimentation, la recherche vise à vérifier si le coût de la migration diffère en fonction de l'ordre de séquençage du transfert des éléments du système lors d'une migration vers le nuage et à identifier la méthode de calcul à utiliser.

C'est pourquoi, il faut considérer la spécificité de chaque élément d'infrastructure, c'est-à-dire chaque composant technologique ainsi que la consommation des ressources. Par exemple, les serveurs Web frontaux, les bases de données, les équipements réseau, les éléments de stockage consomment différemment en terme de CPU, mémoire, bande réseau et lecture/écriture de données. Chaque élément migré a un impact différent sur le coût de migration.

Les paragraphes suivants présentent le modèle formel de la question de cette recherche.

Considérons un centre de données composé de N éléments d'infrastructure E_i , avec i variable de 1 à N . M composants techniques C_k , avec k variable de 1 à M . Chaque système logique S fonctionne sur un sous-ensemble d'éléments d'infrastructure et des composants techniques liés à des éléments.

Les composants techniques sont affectés aux éléments d'infrastructure. Exemple : C_k est assigné à E_i pour le fonctionnement de E_i .

Par exemple, $S1 = \{E_1, E_2, E_3\}$ indique que le système fonctionne sur les éléments d'infrastructure E_1 , E_2 et E_3 .

Un élément E_i contient des composants techniques C_{i1} et C_{i2} .

Il est possible de décrire le système logique S1 utilisant les conventions de langage Python [35] S1 comme un dictionnaire de listes :

$$S1 = \{E1[C1, C2], E2[C1, C4], \dots, EN[C1, C2, CM]\} \quad (3.1)$$

Les éléments d'infrastructure sont interconnectés. Exemple : E_i utilise E_j , car E_i dépend de E_j . Les interconnexions entre les éléments E_i et E_j sont modélisées avec une matrice bidimensionnelle :

| | | | | | | | |
|------|------|------|-----|------|----|------|--|
| | $E1$ | $E2$ | ... | Ei | .. | En | |
| $E1$ | 0 | 1 | | 1 | | 0 | |
| $E2$ | 1 | 0 | | | | 0 | |
| ... | | | 0 | | | | |
| Ej | | 1 | | 0 | | 1 | |
| | 1 | | | | 0 | | |
| En | 1 | 0 | | | | 0 | |

(3.2)

Si $E(i,j) = 1$, E_i est connecté réseau avec E_j (échange de données avec E_j directement).

Le système migre en nuage élément par élément pour diminuer le risque.

La question de cet essai est de trouver l'ordre de migration qui donne le plus bas coût et analyser l'impact sur plusieurs scénarios possibles.

3.1.2 Objectifs et hypothèses

L'objectif est de démontrer l'impact de l'ordre de migration des éléments qui composent un système dans un environnement infonuagique et de partager l'algorithme permettant d'effectuer les calculs pour ainsi déterminer le coût. À partir de l'équation qui calcule le coût de migration, un algorithme est produit pour automatiser le calcul de plusieurs scénarios en changeant la séquence des éléments à migrer. La question de recherche est : « Est-ce que

l'ordre de séquençement du transfert des éléments qui composent un système influence le coût de migration vers le nuage, dans un contexte où la disponibilité du système doit être 100 % et sans risque ? ».

L'hypothèse est que oui, l'ordre de séquençement du transfert des éléments d'un système influence le coût de migration vers le nuage, dans un contexte où la disponibilité du système doit être 100 % et sans risque. Puisque la migration en nuage passe par des architectures intermédiaires et que les données qui circulent de l'intérieur du nuage à l'extérieur sont facturables, il y a une possibilité que l'ordre de migration amène des coûts différents selon le scénario adopté.

Il existe une façon de calculer l'effort d'une migration d'un système dans le nuage. Il faut adapter cette équation pour obtenir le coût de la migration d'un système S composé de N éléments d'infrastructure et M composants technologiques.

La migration en nuage de système S passe par des architectures intermédiaires. Pendant la période de migration il faut ajouter le nouveau coût de chaque élément migré dans le nuage. La migration est terminée quand tous les éléments de système S sont dans le nuage.

Selon la littérature présentée, il faut ajouter à la formule le coût dans le nuage en tant que CMP_4 . La formule se change pour dériver le coût à partir d'une formule qui calcule l'effort.

$$CMP = \sum_{i=1}^2 CMP_i \times W_i \quad (3.3)$$

CMP_1 est la mesure de transferts WAN et LAN pendant la migration en nombre de jours. W_1 est le prix unitaire des gigaoctets transférés par jour à l'extérieur de nuage.

CMP_2 est la mesure d'installation, de configuration et de tests des éléments E_i en nombre total de jours. Ce coût dépend des composants technologiques C_{ij} associés à chaque élément d'infrastructure. W_2 est le prix unitaire de la machine réservée dans le nuage.

L'hypothèse est qu'il y a une directe corrélation entre cette formule et le coût de la migration et aussi qu'il est possible de trouver le meilleur ordre des opérations de migrations à partir de cette formule.

La migration est de type IaaS ce qui implique qu'aucun code applicatif ne change, et qu'on a aucune transformation logique de la BD.

3.1.3 Limites et contraintes

L'essai concerne les environnements dont la haute disponibilité est un enjeu. Dans un autre contexte, cette recherche n'a pas le même impact puisqu'il est alors possible de migrer l'environnement dans une seule opération, les coûts de transfert de données entre l'environnement externe et interne du nuage n'est pas en jeu.

L'essai n'a pas l'ambition de déterminer un modèle universel, mais bien de trouver les caractéristiques qui permettent de déterminer l'impact de l'ordre de séquençement de migration.

Une autre limite est sur le nombre de fournisseurs de services en nuage analysés. La technologie évolue à grande vitesse et les offres de nuage informatique aussi. Les compagnies retenues dans cet essai pour effectuer les analyses de coût, sont les deux plus populaires Amazon et Microsoft (Azure). Plusieurs autres fournisseurs importants existent sur le marché, mais l'essai ne les retient pas: Google, VMware, Rackspace, HP, IBM et Oracle.

De plus, certains facteurs comme la formation du personnel, la résistance au changement et le coût des ressources humaines externes ne sont pas considérés parce qu'ils font partie de coûts d'installation et de configuration.

Le fait que le système migre un élément d'infrastructure à la fois est une autre contrainte. Il est plus complexe de considérer plusieurs éléments migrés à la fois et le travail de cet essai ouvre porte vers une solution générique.

Enfin, les coûts du centre de données sources ne sont pas considérés, car ils sont négligeables et le coût de migration est calculé seulement sur les étapes qui impliquent l'environnement infonuagique.

3.2 Méthodologie proposée

La recherche est quantitative de type expérimental. La méthodologie proposée consiste à vérifier l'influence qui existe entre l'ordre de séquençement d'une migration d'un système et le coût de la migration à l'aide d'un modèle supervisé par des experts. Une première partie de l'étude consiste à modéliser l'architecture cible pour déterminer l'ensemble des éléments à migrer. Ensuite, le coût est calculé selon plusieurs scénarios et les résultats sont comparés afin de trouver le meilleur scénario. L'approche est définie en détail au prochain chapitre.

Chapitre 4

Approche proposée

L'approche proposée pour vérifier l'hypothèse est présentée dans ce chapitre. Une démarche en six étapes permet de vérifier si l'ordre de séquençement du transport des éléments d'un système influence le coût de migration vers le nuage, dans un contexte où la disponibilité du système doit être 100% et sans risque.

4.1 L'approche

L'approche méthodologique consiste en six étapes :

- 1) Choisir l'architecture
- 2) Modéliser l'architecture dans les blocs selon la méthodologie Iteraplan
- 3) Mapper l'architecture analysée aux blocs d'architecture des fournisseurs de solution infonuagique
- 4) Calculer et valider les coûts (concevoir une formule générale pour le calcul de coût de migration)
- 5) Identifier le coût de migration le plus bas selon la séquence de migration (utilisation d'un programme pour automatiser les calculs)
- 6) Analyser et valider les résultats

La Figure 4-1 utilise un diagramme de flux fonctionnel croisé pour présenter la démarche méthodologique de chacune des étapes.

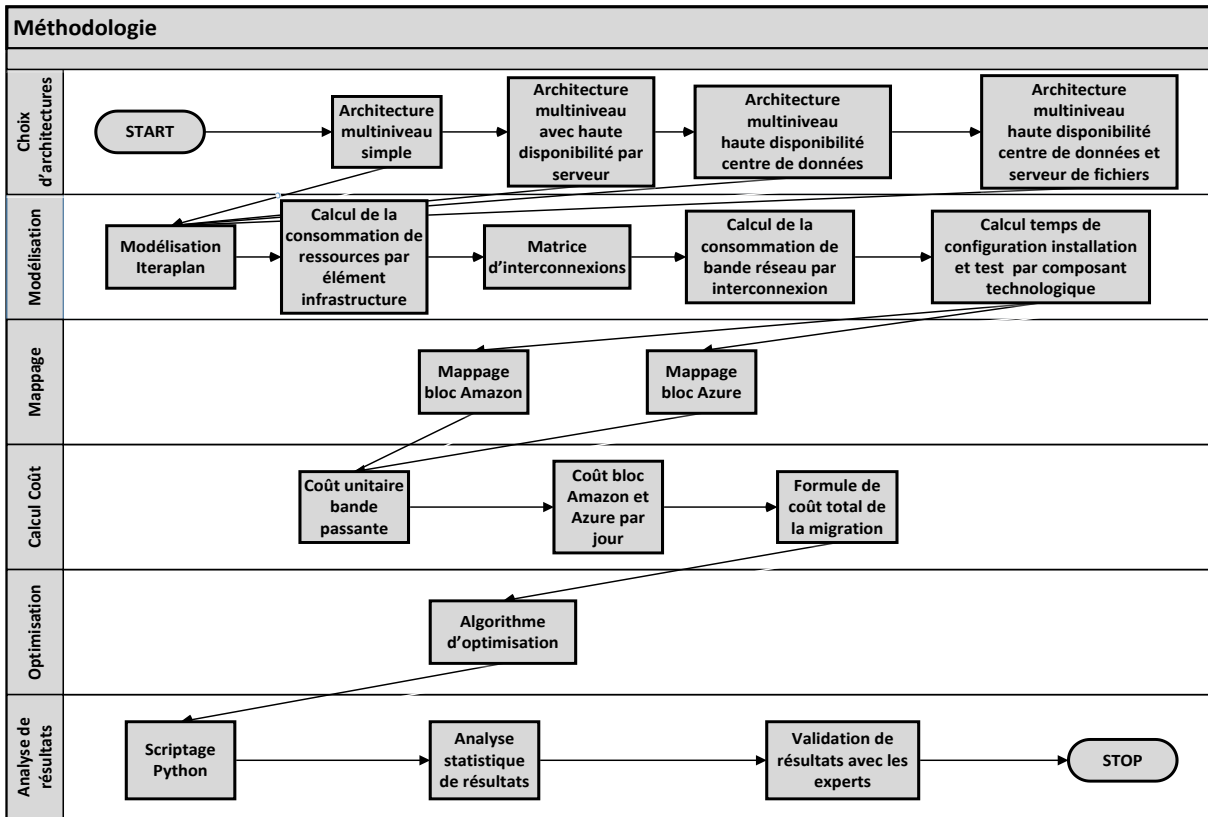


Figure 4-1 Méthodologie

L'expérimentation utilise cette méthode pour procéder aux six étapes. Environ sept spécialistes experts sont impliqués dans l'approche, à chacune des étapes. Ils participent au choix de l'architecture et du système utilisé pour l'expérimentation. Ils procèdent à l'estimation du nombre de jours requis pour migrer les éléments d'infrastructure, approuvent le résultat de l'opération de mappage, effectuent le calcul manuel de scénarios de migration et approuvent le modèle mathématique conçu. Enfin, ils valident les résultats.

Les paragraphes qui suivent décrivent chaque étape en détail.

4.2 Choisir l'architecture

Le contexte de cette étude est particulier. Il vise les entreprises qui ne peuvent se permettre d'arrêter leurs systèmes lors de la migration des infrastructures dans un environnement infonuagique. Dans ce contexte particulier, des coûts supplémentaires sont en jeu. Ce genre de besoin touche, entre autres, des firmes qui gèrent des environnements complexes.

Pour établir le choix de l'architecture, sept spécialistes participent à élaborer un environnement représentatif de la population visée. Il s'agit d'un échantillonnage probabiliste par réseau, car ils travaillent dans la même entreprise et acceptent de participer à cette étude. Ces spécialistes ont entre 10 et 20 ans d'expérience dans le domaine des architectures Web et d'infrastructure.

4.2.1 Sélection d'un modèle d'architecture

Le groupe d'experts est contacté afin de déterminer l'architecture la plus représentative de la réalité de leur entreprise. La façon de procéder pour communiquer afin d'obtenir le modèle à retenir se fait à l'aide du courriel. Ils reçoivent trois modèles de base sélectionnés à partir de la littérature [3], [7], [10].

Figure 1-4, Figure 1-5 et Figure 1-6 présentent ces modèles.

Les spécialistes cotent chacun des modèles en fonction du besoin de l'entreprise. Ensuite, les cotes sont compilées et le modèle obtenant la plus haute cote est retenu pour l'expérimentation. Ainsi, les experts sont impliqués dès le début de l'expérimentation, tel que la méthode qualitative utilisée le suggère.

C'est le modèle d'architecture multiniveau, avec redondance par centre de données, mais sans centre de données redondants qui est retenu. Le choix de l'architecture pour l'étude est présenté à la Figure 1-5.

L'architecture cible est obtenue à partir d'une opération de mappage, entre les éléments d'infrastructure identifiés dans l'architecture source et les blocs d'architecture obtenus auprès des fournisseurs de solutions infonuagiques. Cette opération est décrite dans la section 4.4

L'expérimentation s'effectue avec deux fournisseurs connus des spécialistes qui participent à l'expérimentation, soit Amazon AWS et Microsoft Azure. Un tableau de Gartner Group 2016 confirme la popularité de ces deux fournisseurs. Le tableau indique que les plus importants sont, en 2016, Amazon avec 31% des parts du marché et Microsoft Azure avec 9% des parts de marché. De plus, ces deux entreprises possèdent des centres de données infonuagiques au Québec. Le choix convient donc pour l'expérimentation.

4.2.2 Sélection d'un système pour l'expérimentation

Le choix du système pour l'expérimentation utilise le CMS Magnolia 5.4. Il contient des aiguilleurs de charge A10, des instances applicatives Web Apache, des instances applicatives J2EE Tomcat et des bases de données Oracle.

Le système choisi pour l'expérimentation est un système Web, qui est jugé complexe à cause du nombre d'éléments d'infrastructure qui le composent, des flux de données et divers types d'éléments. Le système contient les composants technologiques les plus populaires en 2016.

Togaf demande la création d'un document de définition d'architecture [28] qui contient, entre autres, l'architecture source et l'architecture cible. L'architecture source du système retenue est décomposée en éléments d'infrastructure et composants technologiques pour analyser sa migration vers le nuage.

4.3 Modéliser l'architecture

Tous les éléments d'infrastructure sont interconnectés au réseau. Sur chaque interconnexion, un nombre de giga-octets est transporté. Seul le flux de données sortant du nuage est facturé.

La modélisation combine chaque élément d'infrastructure avec les composants technologiques liés. Les composants technologiques sont ensuite analysés pour estimer le

nombre de jours d'installation, de configuration et de test à partir d'une expérience confirmée par les experts qui participent à l'étude.

Chaque élément d'infrastructure est modélisable selon la méthodologie Iteraplan et est réduit à un degré d'abstraction permettant le calcul du coût de l'ensemble du système. Le but est d'extraire les éléments d'infrastructure et les composants technologiques qui vont migrer dans le nuage et de dériver les mesures de coût.

Le mappage des éléments d'infrastructure avec les blocs d'architecture dans le nuage est effectué à partir des caractéristiques de chaque élément, machine physique ou virtuelle, selon le nombre de processeurs, la mémoire et le stockage. Cette opération donne le prix unitaire par jour d'utilisation de chaque élément en nuage selon les caractéristiques de l'élément d'infrastructure nuage correspondant. La figure 4.2 présente le schéma de l'architecture du système.

Un nombre de sept éléments d'infrastructure et huit composants technologiques sont identifiés. Chaque élément d'infrastructure réside sur une machine (serveur).

L'aiguilleur de charge (*load balancer*) joue le rôle d'interface avec l'Internet. Les utilisateurs, nommés acteurs selon Togaf, frappent l'aiguilleur avec des requêtes qui sont filtrées et envoyées aux deux serveurs web redondants qui hébergent des instances applicatives Apache. Les instances Apache jouent le rôle de serveurs proxy qui enlèvent la couche de sécurité (SSL) avant passer les paquets vers les instances applicatives. Les instances applicatives utilisent les composants Tomcat et hébergent les applications Magnolia écrites en Java. Les bases de données Oracle jouent le rôle de dépositaires de métadonnées et informations présentées sur Internet.

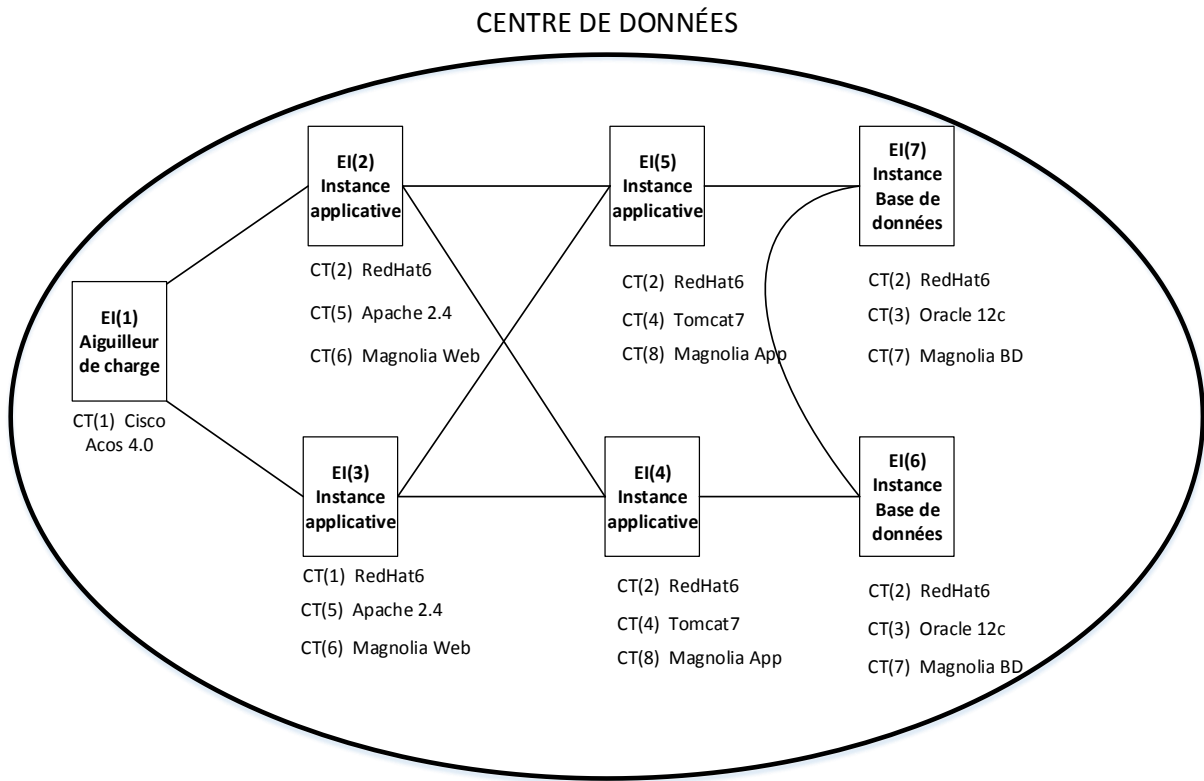


Figure 4-2 Architecture multiniveau source

Les machines utilisées sont Linux version RedHat 6.

Chaque expert analyse les huit composants technologiques du système choisi afin d'estimer le nombre de jours de test, d'installation et de configuration requis pour migrer chacun des composants. Le tableau 4.1 présente le canevas envoyés aux experts.

Tableau 4.1 Estimé des composants technologiques du système à migrer

| Expert X | | | | | |
|----------|-------------------------|-----------------|---------------|------|-------|
| | Composant technologique | Nombre de Jours | | | |
| | | Installation | Configuration | Test | Total |
| 1 | A10 Acos 4.0 | | | | |
| 2 | RedHat 6 | | | | |

| Expert X | | | | | |
|----------|-------------------------|-----------------|---------------|------|-------|
| | Composant technologique | Nombre de Jours | | | |
| | | Installation | Configuration | Test | Total |
| 3 | Oracle 12c | | | | |
| 4 | Tomcat 7 | | | | |
| 5 | Apache 2.4 | | | | |
| 6 | Magnolia Web | | | | |
| 7 | Magnolia BD | | | | |
| 8 | Magnolia Applicatif | | | | |

Le tableau contient les huit composants technologiques. Pour chacun des composants, chacun des experts doit fournir l'estimation du nombre de jours pour installer, configurer et tester le composant. Les données sont ensuite compilées par composant et un tableau global avec les estimations des experts est inscrit. L'estimation la plus basse et la plus haute sont retirées et la moyenne permet d'obtenir le nombre de jours à considérer pour le calcul. Le tableau 4.2 présente un exemple de tableau de compilation des estimations possibles provenues des experts dans la gestion de la base de données Oracle 12c.

Tableau 4.2 Exemple de compilation possible des estimations d'experts

| Oracle 12c | | | | | |
|------------|----------|-----------------|---------------|------|-------|
| | Expert | Nombre de Jours | | | |
| | | Installation | Configuration | Test | Total |
| 1 | Expert 1 | 4 | 1 | 2 | 7 |
| 2 | Expert 2 | 3 | 1 | 2 | 6 |
| 3 | Expert 3 | 4 | 1 | 2 | 7 |
| 4 | Expert 4 | 3 | 1 | 2 | 6 |
| 5 | Moyenne | 3,5 | 1 | 2 | 6,5 |

L'architecture évolue avec chaque migration en nuage en passant par des étapes intermédiaires.

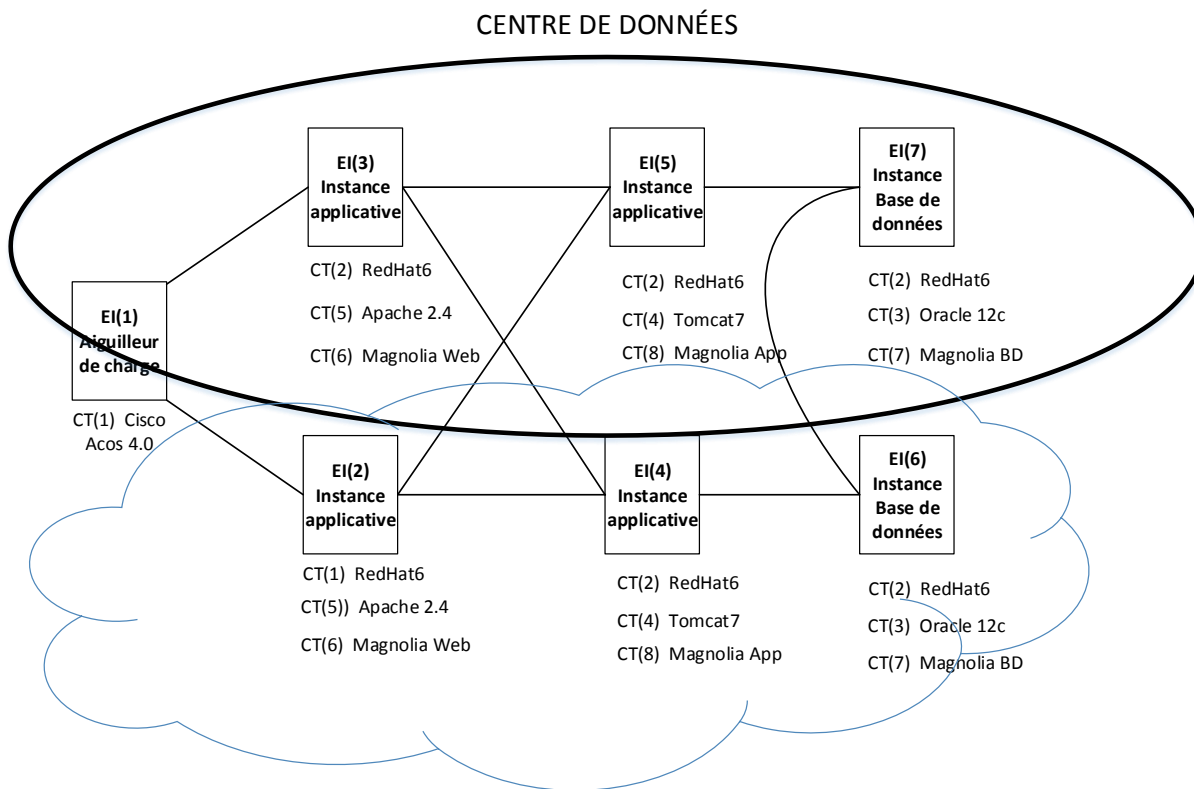


Figure 4-3 Architecture multiniveau en cours de migration

À la fin, l'infrastructure doit être complètement migrée dans l'environnement infonuagique.

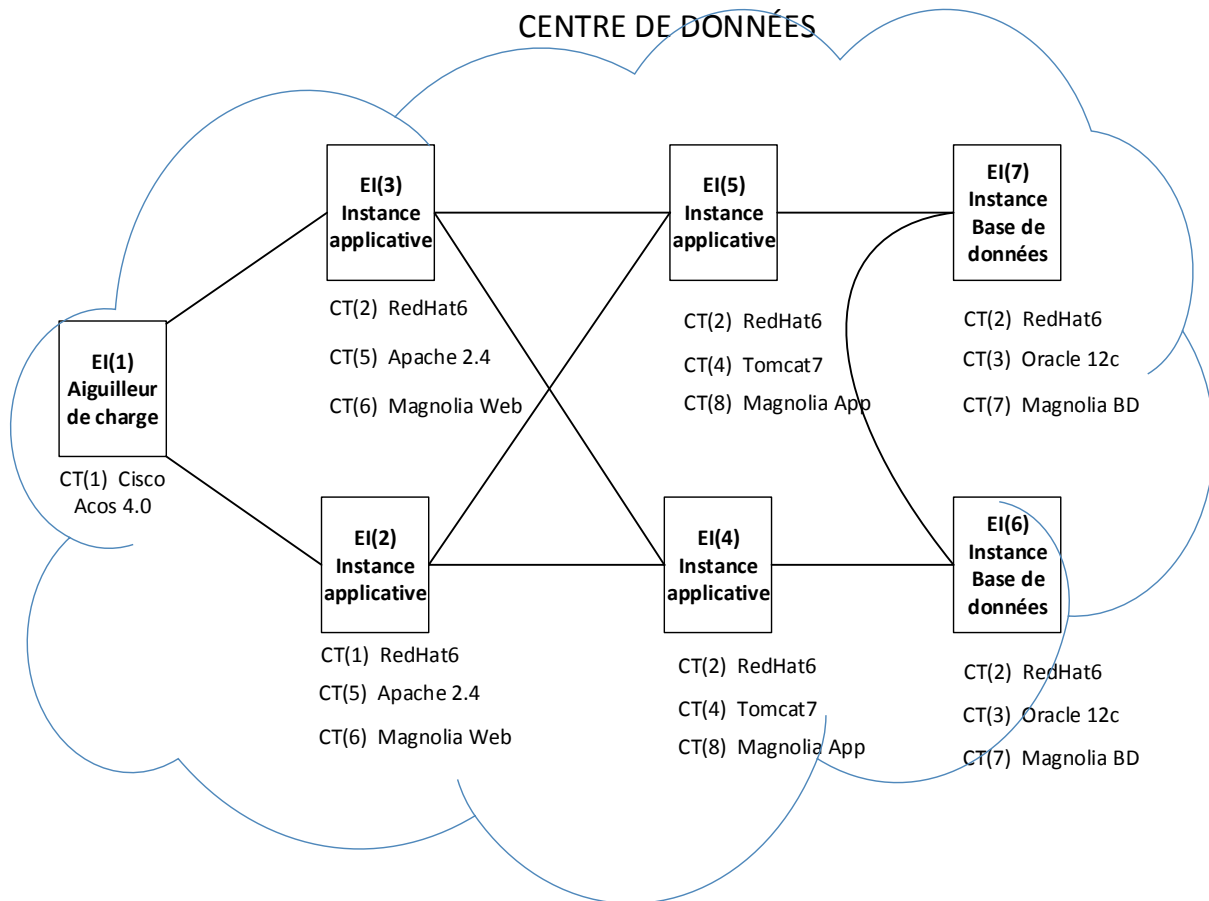


Figure 4-4 Architecture multiniveau cible migrée en nuage

Chaque étape de migration est déterminante pour le coût. L'emplacement des systèmes change, les interconnexions aussi. Les prix varient par bloc utilisé dans le nuage et pour la bande passante sortante du nuage utilisé. Le tableau suivant présente la matrice des interconnexions et le nombre de Giga-octets sortants par heure pour les éléments interconnectés. Tenant compte qu'il s'agit d'un système Web, l'acteur utilisateur applicatif Internet est aussi représenté.

Le calcul de bande passante est effectué avec les logiciels « Tcpcdump », « Wireshark » et « Grafana » capable de tracer la consommation entre deux adresses et ports TCP/IP différents. Cette opération est effectuée par les administrateurs réseau et permet de remplir le tableau 4.3

Tableau 4.3 Consommation de bande passante

| EI(i,j) => I vers J | | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---------------------|------------|----------|------------|------|------|------|------|-----|-----|
| | | Internet | Aiguilleur | Web1 | Web2 | App1 | App2 | BD1 | BD2 |
| EI(0) | Internet | X | | X | X | X | X | X | X |
| EI(1) | Aiguilleur | | X | | | X | X | X | X |
| EI(2) | Web1 | X | | X | X | | | X | X |
| EI(3) | Web2 | X | | X | X | | | X | X |
| EI(4) | App1 | X | X | | | X | X | | |
| EI(5) | App2 | X | X | | | X | X | | |
| EI(6) | BD1 | X | X | X | X | | | X | |
| EI(7) | BD2 | X | X | X | X | | | | X |

Pour l'élément EI(i,j), une valeur N abstraite dans la cellule du tableau signifie que EI(i) envoie N giga-octets de bande passante vers un élément EI(j). Comme mentionné, seulement la bande réseau sortante du nuage est payante. Il est à noter que c'est le cas en 2017, mais il est possible que les fournisseurs changent leur modèle de coût. La méthode de calcul est indépendante du fournisseur. Le modèle de coût doit être identique même si les prix diffèrent.

4.4 Associer l'architecture analysée aux blocs d'architecture des fournisseurs de service nuage

Les éléments d'infrastructure sont associés avec les éléments correspondants dans les nuages informatiques Amazon et Microsoft. Comme règle, le « mappage » effectue la conversion d'une machine dans des blocs d'architecture équivalents vendeur nuage. Les blocs mappés doivent avoir des ressources physiques supérieures ou égales à ceux d'origine. Les blocs infonuagique sont donc choisis à partir du nombre de cœurs CPU, giga-octets de mémoire RAM et giga-octets de stockage persistant.

L'utilisation des outils des sites Amazon et Microsoft termine l'opération pour déterminer les coûts par heure de chaque machine et ensuite, pour calculer le coût pour l'ensemble des éléments d'infrastructure. Le travail est effectué par l'auteur, mais les résultats sont validés

par les experts. Le document d'évaluation se trouve sous Annexe V Documents d'évaluation de spécialistes.

Tableau 4.4 Ressources payables dans le nuage informatique

| | Éléments d'infrastructure | Bloc Amazon | Prix Amazon / heure | Bloc Microsoft | Prix Azure / heure |
|---|--------------------------------------|-------------|---------------------|----------------|--------------------|
| 1 | EI(0) Aiguilleur de charge | | | | |
| 2 | EI(1) Instance Applicative Apache | | | | |
| 3 | EI(2) Instance Applicative Apache | | | | |
| 4 | EI(3) Instance Applicative Tomcat | | | | |
| 5 | EI(4) Instance Applicative Tomcat | | | | |
| 6 | EI(5) Instance BD Oracle | | | | |
| 7 | EI(6) Instance BD Oracle | | | | |

Ce tableau est rempli avec des blocs infonuagiques qui dépassent minimalement ou sont égaux aux caractéristiques des éléments d'origine. Le coût de la bande passante réseau est ajouté pour déterminer le coût total d'un système en nuage.

Une observation s'impose. Les prix sont variables, selon les contrats et la consommation de ressources par mois et selon le fournisseur. Le but de cet essai n'est pas de choisir le meilleur prix au niveau des fournisseurs, mais de choisir le meilleur scénario de migration selon la séquence de migration

L'expérience avec la conception en nuage est déterminante, il faut tenir compte d'une possible croissance, il faut penser à l'élasticité des ressources consommées pour les périodes d'achalandage.

4.5 Calculer et valider les coûts

La formule de coût de migration est calculée dans deux étapes. La première étape fait un calcul du coût de la migration d'une seule machine dans le nuage informatique. Il faut calculer le prix de la machine par jour et le prix de la bande passante.

Donc :

- 1) CMP1 est le nombre de jours d'installation, configuration et test par machine migrée
- 2) W1 est le prix unitaire de la machine par jour
- 3) CMP2 est le nombre de Go sortants de réseau/jour pendant la période de test et d'activité
- 4) W2 est le prix unitaire d'un Gigaoctet sortant de nuage par jour
- 5) CMP3 est le nombre de jours d'installation et configuration par machine migrée
- 6) W3 est le prix unitaire d'un Gigaoctet sortant de nuage par jour (égal à W2)

La formule est :

$$Coût = \sum_{i=1}^3 CMP_i * W_i \quad (4.1)$$

Chaque migration d'un équipement d'infrastructure ajoute un coût supplémentaire. À la fin, toutes les machines sont testées et considérées migrées dans le nuage. Le tableau suivant présente les étapes et les calculs.

Tableau 4.5 Étapes de migration dans le nuage informatique

| | Étape de migration |
|------|---|
| 1(a) | EI(1) migre. Configuration réseau change dans un nouvel état. |
| 1(b) | Les composants technologiques de EI(1) sont installés, configurés et testés. Pendant cette étape le prix de la machine EI(1) dans le nuage est payé. Pendant cette étape le prix de la bande passante sortante de la machine EI(1) est payé pendant la période de test. |
| 2(a) | EI(2) migre. Configuration réseau change dans un nouvel état. |

| | Étape de migration |
|------|---|
| 2(b) | Les composants technologiques de EI(2) sont installés, configurés et testés. Pendant cette étape le prix de machines EI(1), EI(2) dans le nuage est payé. Pendant cette étape le prix de la bande passante sortante de machines EI(1), EI(2) est payé de cette façon : a) le prix de la bande passante de EI(1) pour la période d'installation et configuration de EI(2) dans la vieille configuration réseau. b) le prix de la bande passante de EI(2) et EI(1) pour la période de tests de EI(2) dans la nouvelle configuration réseau |
| -- | ---- |
| J(a) | EI(j) migre. Configuration réseau change dans un nouvel état. |
| J(b) | Les composants technologiques de EI(j) sont installés, configurés et testés. Pendant cette étape le prix de machines EI(1), EI(2),...,EI(j) dans le nuage est payé. Pendant cette étape le prix de la bande passante sortante de machines EI(1), EI(2),...,EI(j) est payé de cette façon : a) le prix de la bande passante de EI(1), EI(2),...,EI(j-1) pour la période d'installation et configuration de EI(j) dans la vieille configuration réseau. b) le prix de la bande passante de EI(1), EI(2),...,EI(j) pour la période de test de EI(j) dans la nouvelle configuration réseau |
| -- | --- |
| N(a) | Dernier EI(N) migre. Configuration réseau change dans un nouvel état. |
| N(b) | Les composants technologiques de EI(N) sont installés, configurés et testés. Pendant cette étape le prix de machines EI(1), EI(2),...,EI(N) dans le nuage est payé. Pendant cette étape le prix de la bande passante sortante de machines EI(1), EI(2),...,EI(N) est payé Pendant cette étape le prix de la bande passante sortante de machines EI(1), EI(2),...,EI(N) est payé de cette façon : a) le prix de la bande passante de EI(1), EI(2),...,EI(N-1) pour la période d'installation et configuration de EI(j) dans la vieille configuration réseau. b) le prix de la bande passante de EI(1), EI(2),...,EI(N) pour la période de test de EI(N) dans la nouvelle configuration réseau |
| Fin | Fin de migration |

La formule qui calcule le coût cumulatif d'installation, configuration et test des machines migrées en nuage selon leurs prix d'utilisation par jour est :

$$\sum_{j=1}^N (CMPict(j) * (\sum_{k=1}^j W(k))) \quad (4.2)$$

Les valeurs CMPict(j) restent constantes pour chaque élément J migré. Ce sont les nombres totaux de jours d'installation, configuration et test pour l'élément J.

$$CMPict(j) = NombreJOURSict(j) \quad (4.3)$$

Par contre, les valeurs de CMP changent par chaque élément migré qui va créer une autre topologie réseau et va influencer le coût. La formule qui calcule le coût cumulatif de la bande passante de machines migrées en nuage pendant la période de test et activité en nuage suivant le teste, est la suivante :

$$\sum_{j=1}^N (\sum_{k=1}^j (CMPtest(j, k) * W2)) \quad (4.4)$$

CMPtest(j,k) est défini pour élément K dans la topologie réseau après la migration d'élément J :

$$CMPtest(j, k) = JRSest(j) * ConsBandeSort(k \text{ après élém. } j \text{ migré}) \quad (4.5)$$

La formule qui calcule le coût cumulatif de la bande passante de machines migrées en nuage pendant la période d'installation et configuration est la suivante :

$$\sum_{j=1}^N (\sum_{k=1}^j (CMPic(j - 1, k) * W2)) \quad (4.6)$$

Dans la période d'installation et configuration d'un élément K, la consommation réseau est spécifique à la topologie réseau après la migration d'élément J-1, mais avant la migration d'élément J. Élément J change la topologie réseau seulement quand le test commence. CMPic(j,k) est défini:

$$CMPic(j, k) = JRSic(j) * ConsBandeSort(k \text{ après élém. } (j - 1) \text{ migré}) \quad (4.7)$$

Le calcul de coûts pendant la migration sera confirmée par le panel des experts sélectionnés qui comptent au moins cinq ans d'expérience du travail avec l'infonuagique.

4.6 Identifier le coût optimal de migration selon la séquence de migration

Pour calculer la séquence de migration au meilleur coût, tous les scénarios de migration doivent être estimés. Cette étape exige d'effectuer les calculs pour chaque séquence possible.

Les séquences possibles sont multiples, ce qui engendre une multitude de scénarios, avec seulement sept éléments à migrer. Un programme est conçu afin de calculer le coût de migration pour toutes les combinaisons de séquences possibles et choisir la séquence optimale. La méthode est générale et s'applique pour tous les systèmes logiques qui migrent dans le nuage. Les résultats sont présentés au Chapitre 5.

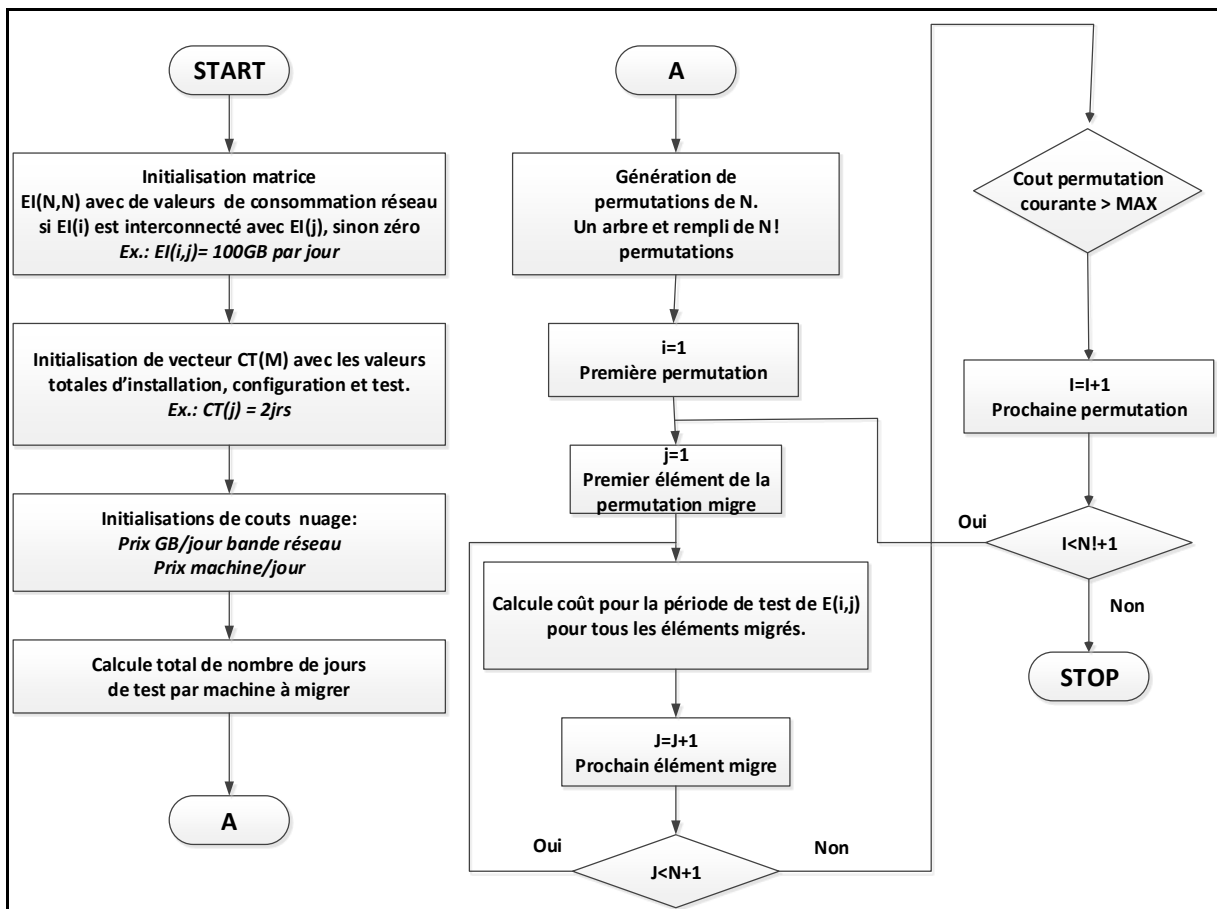


Figure 4-5 Algorithme d'optimisation

L'algorithme utilise une technique récursive. Il génère les permutations possibles de migrations (la variable séquence) et calcule le coût pour chaque étape de migration par séquence. Le langage d'implantation utilisé est Python, spécialisé en ensembles de données. La validation d'algorithme est vérifiée et approuvée par deux programmeurs d'expérience.

La partie quantitative de la méthodologie est validée dans deux étapes :

- a) Calcul manuel de coût de migration
- b) Comparaison de résultats avec la méthode automatique originale de cet essai

Les résultats sont compilés dans un tableau afin de les comparer aux résultats de l'algorithme programmé. Un échantillon de 7 séquences aléatoires de migration est calculé manuellement et comparé avec les résultats obtenus par la programmation.

Tableau 4.6 Tableau de résultats attendus et obtenus

| | Séquence de migration | Coût manuel Amazon | Coût Amazon | Coût manuel Azure | Coût Azure |
|---|------------------------------|---------------------------|--------------------|--------------------------|-------------------|
| 1 | (1, 2, 3, 4, 6, 5, 7) | | | | |
| 2 | (7, 6, 5, 4, 3, 2, 1) | | | | |
| 3 | (1, 2, 4, 3, 5, 7, 6) | | | | |
| 4 | (7, 5, 6, 4, 3, 1, 2) | | | | |
| 5 | (2, 1, 4, 3, 6, 5, 7) | | | | |
| 6 | (2, 4, 1, 3, 5, 7, 6) | | | | |
| 7 | (1, 5, 2, 3, 4, 6, 7) | | | | |

L'algorithme doit aussi identifier la meilleure séquence pour chacun des fournisseurs de services.

Chaque expert est appelé à vérifier un scénario manuellement. Les données sont comparées avec celles calculées par le programme automatique. À partir de cette vérification statistique, qui démontre une corrélation entre l'ordre de migration et le coût, le panel des experts confirme ou non la méthode utilisée et les résultats. Les questions et les réponses doivent être pertinentes au cas concret présenté dans cet essai, voir 4.7.

4.7 Analyser et valider les résultats

Un formulaire est distribué aux sept experts afin d'évaluer la méthode utilisée. Les questions sont présentées dans le tableau 4.7. Les questions concernent l'approche utilisée et permettent de vérifier que la méthodologie est effectuée de façon adéquate et selon les règles qui régissent le travail. Les spécialistes peuvent inclure dans le courriel de réponse des commentaires et des recommandations pour améliorer la cotation.

Tableau 4.7 Formulaire d'évaluation de la méthode et de résultats

| | Question | 1 Pauvre | 2 Passe | 3 Bien | 4 Très bien | 5 Excellent |
|---|---|-------------|------------|-----------|-------------------|----------------|
| 1 | La modélisation proposée est-elle correcte? | | | | | |
| 2 | Le « mappage » dans le nuage est-il bien conçu? | | | | | |
| 3 | Les contraintes sont-elles respectées? | | | | | |
| 4 | La méthodologie de migration suit-elle les règles du procédé? | | | | | |
| 5 | Les résultats confirment-ils l'hypothèse? | | | | | |

Le formulaire comportant ces cinq questions est utilisé afin de valider l'exécution des différentes étapes et pour confirmer la validité des résultats.

La question 1 permet d'évaluer la modélisation de l'architecture cible proposée par le chercheur. Selon la réponse, des corrections seront appliquées jusqu'au moment quand un consensus existe.

La question 2 confirme la qualité de « mappage ». L'offre nuage est vaste et plusieurs options sont possibles. L'expérience pratique avec les technologies infonuagiques est déterminante pour un mappage qui offre au minimum la même performance, disponibilité et fonctionnalité.

La question 3 s'assure que la méthode respecte les contraintes énumérées dans la section 3.1.3 qui stipule que l'essai se concentre seulement sur les vendeurs Amazon et Azure, que

le système migre avec seulement un élément d'infrastructure à la fois et que le facteur humain n'est pas pris en considération.

La question 4 valide la méthodologie utilisée pour répondre à la question de recherche. Les commentaires et les recommandations des experts ainsi qu'une cotation positive sont critiques pour le succès de cet essai.

La dernière question numéro 5 confirme l'arrimage de la méthodologie avec les résultats et la réponse à la question de recherche. C'est la question la plus importante.

Les questions sont notées sur une échelle de 1 à 5. 1 est pauvre, 2 passe, 3 est bien, 4 est très bien et 5 est excellent. Ceci permet d'obtenir un degré quantifié de crédibilité.

Les valeurs moyennes de notations seront calculées et commentées. Pour les questions qui reçoivent des notes en dessous de la moyenne, des discussions permettent de comprendre et d'ajuster la méthodologie. Puisque les spécialistes sont impliqués tout au long du processus, il est plus facile de reprendre une partie de la méthodologie si cette dernière est jugée non faible.

Le prochain chapitre présente les résultats obtenus et répond à la question de recherche.

Chapitre 5

Analyse des résultats

5.1 Mesures de la modélisation

Les mesures de normalisation concernent le nombre de jours estimés pour la migration des composants technologiques multiplié par la consommation des ressources au niveau des serveurs et de la bande passante pour obtenir le coût selon le taux de chaque fournisseur.

5.1.1 Composants technologiques

Suite à des échanges courriels avec les 3 experts en administration de systèmes, serveurs applicatifs et bases de données, le Tableau 5.1 a été complété avec des chiffres estimés d'effort, en nombre de jours, pour mettre en place les composants dans l'infonuagique.

Tableau 5.1 Estimé des composants technologiques du système à migrer

| | Composant technologique | Nombre de Jours | | | |
|---|-------------------------|-----------------|---------------|------|-------|
| | | Installation | Configuration | Test | Total |
| 1 | A10 Acos 4.0 | 0,5 | 0,5 | 1 | 2 |
| 2 | RedHat 6.0 | 0,5 | 0,25 | 0,25 | 1 |
| 3 | Apache 2.4 | 0,1 | 0,1 | 0,2 | 0,4 |
| 4 | Tomcat 7 | 0,2 | 0,2 | 0,4 | 0,8 |
| 5 | Oracle 12c | 0,25 | 0,25 | 0,5 | 1 |
| 6 | Magnolia Web | 0,2 | 0,3 | 1 | 1,5 |
| 7 | Magnolia BD | 0 | 0 | 0,2 | 0,2 |
| 8 | Magnolia Applicatif | 0,2 | 0,3 | 5 | 5,5 |

5.1.2 Consommation de ressources

Chaque serveur est analysé sur quatre caractéristiques de base :

- 1) Le nombre de cœurs CPU et la performance mesurée en nombre de cœurs et fréquence
- 2) La grandeur de la mémoire volatile mesurée en GB
- 3) L'espace physique nécessaire mesuré en GB
- 4) Les contraintes de performance E/S mesurées en IOPS

Cette analyse permet l'association dans les blocs d'infrastructure nuage. Les mesures sont effectuées par le logiciel Grafana installé sur chaque serveur ainsi qu'en prenant les caractéristiques de chaque serveur migré.

Tableau 5.2 Tableau de machines

| Machine | CPU | | Mémoire | E/S | Espace physique |
|---------|-------|--|---------|------|-----------------|
| | #vCPU | Fréquence (GHz) | GB | IOPS | GB |
| 1 Web1 | 2 | INTEL® CORE™ i5-7600T PROCESSOR 3.7GHZ | 8 | 100 | 16 |
| 2 Web2 | 2 | INTEL® CORE™ i5-7600T PROCESSOR 3.7GHZ | 8 | 100 | 16 |
| 3 App1 | 4 | INTEL® CORE™ i5-7600T PROCESSOR 3.7GHZ | 8 | 300 | 32 |
| 4 App2 | 4 | INTEL® CORE™ i5-7600T PROCESSOR 3.7GHZ | 8 | 300 | 32 |
| 5 BD1 | 8 | INTEL® CORE™ i7-6950X PROCESSOR EXTREME EDITION 3.5GHZ | 32 | 1200 | 64 |
| 6 BD2 | 8 | INTEL® CORE™ i7-6950X PROCESSOR EXTREME EDITION | 32 | 1200 | 64 |

5.1.3 Bande passante réseau

Pour calculer le coût de la migration, il faut estimer la consommation de la bande passante entre les serveurs qui hébergent le système.

Les mesures sont effectuées par trois administrateurs de systèmes avec les logiciels Wireshark et TCPdump capables de calculer la bande passante entre les machines pendant une journée complète, soit 24 heures, surchargée d'activités.

Tableau 5.3 Tableau de consommation de la bande passante

| EI(i,j) => I vers J (Gio/heure) | | J=0 | J=1 | J=2 | J=3 | J=4 | J=5 | J=6 | J=7 |
|------------------------------------|------------|----------|------------|------|------|------|------|------|------|
| | | Internet | Aiguilleur | Web1 | Web2 | App1 | App2 | BD1 | BD2 |
| I=0 | Internet | X | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| I=1 | Aiguilleur | 1,1 | X | 0,6 | 0,6 | 0 | 0 | 0 | 0 |
| I=2 | Web1 | 0 | 0,5 | X | 0 | 0,5 | 0,5 | 0 | 0 |
| I=3 | Web2 | 0 | 0,5 | 0 | X | 0,5 | 0,5 | 0 | 0 |
| I=4 | App1 | 0 | 0 | 0,45 | 0,45 | X | 0 | 0,75 | 0,75 |
| I=5 | App2 | 0 | 0 | 0,45 | 0,45 | 0 | X | 0,75 | 0,75 |
| I=6 | BD1 | 0 | 0 | 0 | 0 | 1 | 1 | X | 2 |
| I=7 | BD2 | 0 | 0 | 0 | 0 | 1 | 1 | 2 | X |

5.1.4 Mappage infonuagique

L'association au nuage est effectuée avec le groupe d'experts comptant de cinq à dix ans d'expérience.

Cette solution est validée et signée sur le document papier présenté en 4.7 à l'Annexe V.

L'aiguilleur de charge est remplacé par une solution PaaS. Les autres blocs d'architectures sont des machines virtuelles avec des systèmes d'opération RedHat6 préinstallés. Le Tableau 5.4 présente le mappage Azure et Amazon.

Tableau 5.4 Tableau de blocs payables dans le nuage informatique

| | Éléments d'infrastructure | Bloc Amazon | Prix Amazon \$/ heure | Bloc Microsoft | Prix Azure \$/ heure |
|---|-----------------------------------|-----------------------|-----------------------|----------------------|----------------------|
| 1 | EI(1) Aiguilleur de charge | Elastic Load Balancer | 0,025 | Public Load Balancer | 0 |
| 2 | EI(2) Instance Applicative Web | I3.large | 0,156 | F4 | 0,724 |

| | Éléments d'infrastructure | Bloc Amazon | Prix Amazon \$/ heure | Bloc Microsoft | Prix Azure \$/ heure |
|---|-----------------------------------|-------------|-----------------------|----------------|----------------------|
| 3 | EI(3) Instance Applicative Web | I3.large | 0,156 | F4 | 0,724 |
| 4 | EI(4) Instance Applicative | I3.xlarge | 0,312 | F4 | 0,724 |
| 5 | EI(5) Instance Applicative | I3.xlarge | 0,312 | F4 | 0,724 |
| 6 | EI(6) Instance BD | I3.2xlarge | 0,624 | L8 | 2,621 |
| 7 | EI(7) Instance BD | I3.2xlarge | 0,624 | L8 | 2,621 |

Les sites Azure ne permettent pas facilement de trouver des instances réservées.

5.2 Analyse quantitative de résultats

5.2.1 Résultats obtenus

Le chemin optimal est calculé avec l'algorithme présenté au chapitre 4, dans la section 4.6 mais les résultats d'un échantillon de sept séquences sont aussi calculés manuellement pour valider les données obtenues avec l'algorithme programmé. Les résultats calculés manuellement sont présentés dans la section Analyse de coût.

Tableau 5.5 Tableau de résultats attendus et obtenus

| | Séquence de migration | Coût manuel Amazon (\$) | Coût Amazon(\$) (algorithme) | Coût manuel Azure(\$) | Coût Azure(\$) (algorithme) |
|---|-----------------------|-------------------------|---------------------------------|-----------------------|--------------------------------|
| 1 | (1, 2, 3, 4, 6, 5, 7) | 734,59 \$ | 734,59 \$ | 3821,45 \$ | 3821,45 \$ |
| 2 | (7, 6, 5, 4, 3, 2, 1) | 1242,74 \$ | 1242,74 \$ | 26768,10 \$ | 26768,10 \$ |
| 3 | (1, 2, 4, 3, 5, 7, 6) | 741,64 \$ | 741,64 \$ | 5610,05 \$ | 5610,05 \$ |
| 4 | (7, 5, 6, 4, 3, 1, 2) | 1177,45 \$ | 1177,45 \$ | 26352,28 \$ | 26352,28 \$ |

| | Séquence de migration | Coût manuel Amazon (\$) | Coût Amazon(\$) (algorithme) | Coût manuel Azure(\$) | Coût Azure(\$) (algorithme) |
|---|-----------------------|-------------------------|---------------------------------|-----------------------|--------------------------------|
| 5 | (2, 1, 4, 3, 6, 5, 7) | 848,61 | 848,61 | 4215,15 | 4215,15 |
| 6 | (2, 4, 1, 3, 5, 7, 6) | 736,3 | 736,3 | 5657,87 | 5657,87 |
| 7 | (1, 5, 2, 3, 4, 6, 7) | 755,29 | 755,29 | 3700,47 | 3700,47 |

Tableau 5.6 Tableau de résultats optimaux

| | Séquence de migration avec meilleur et pire coût | Coût Amazon(\$) | Coût Azure(\$) |
|---|--|-----------------|----------------|
| 1 | (4, 2, 1, 3, 5, 6, 7) | 730,68 \$ | |
| 2 | (6, 7, 2, 3, 1, 4, 5) | 1353,69 \$ | |
| 3 | (1, 4, 5, 2, 3, 6, 7) | | 3612,78 \$ |
| 4 | (7, 6, 2, 3, 4, 5, 1) | | 27151,29 \$ |

La nécessité d'une méthode optimale est rapidement observable. Les coûts maximaux sont entre deux à huit fois plus grands que les coûts minimaux obtenus donc une migration à l'aveugle peut avoir des impacts négatifs sur le coût.

Ce qui confirme l'hypothèse selon laquelle, la séquence de migration, dans un contexte de très haute disponibilité et de tolérance aux risques nulle, permet de réduire le coût d'une migration vers un environnement infonuagique.

Les paragraphes suivants détaillent les analyses statistiques utilisant Excel avec le module d'analyses Microsoft ainsi que le module complémentaire *Real Statistics Analysys Pack* pour les statistiques inférentielles.

5.2.2 Statistiques descriptives

Les statistiques descriptives décrivent en résumé les données disponibles. Les variables utilisées sont quantitatives. La séquence de migration est une variable discrète et la variable de type coût est continue.

Tableau 5.7 Tableau statistiques descriptives

| # | Statistiques Descriptives | Valeur |
|----|---|------------|
| 1 | Moyenne | 1050,47 |
| 2 | Erreur-type | 2,20 |
| 3 | Médiane | 1061,39 |
| 4 | Mode | 972,81 |
| 5 | Écart-type | 156,39 |
| 6 | Variance de l'échantillon | 24457,37 |
| 7 | Kurstosis (Coefficient d'aplatissement) | -0,90 |
| 8 | Coefficient d'asymétrie | -0,14 |
| 9 | Plage | 623,01 |
| 10 | Minimum | 730,68 |
| 11 | Maximum | 1353,69 |
| 12 | Somme | 5294346,72 |
| 13 | Nombre d'échantillons | 5040,00 |

La moyenne et la médiane sont utilisées pour les analyses de tendance centrale qui se concentrent sur les valeurs typiques. Les deux valeurs sont proches, la moyenne est 1050,47 et la médiane est 1061,39. Les deux valeurs sont très proches et confirment que les cas particuliers sont presque inexistants. Une différence plus grande pourrait impliquer la présence de ces cas (*outliers*), mais ce n'est pas le cas. Le mode est la valeur la plus fréquente (972,81), seize variables sont associées à cette catégorie. C'est une plus grande probabilité de se trouver dans cette zone de coût si la bonne méthode n'est pas utilisée.

Les statistiques descriptives contiennent aussi des mesures de variabilité capables d'analyser la distribution des coûts. La rangée est la valeur maximale moins la valeur minimale (423,01). La variance mesure la dispersion. La déviation standard est une mesure

de degré d'éloignement de la moyenne. Sa valeur de 156,39 confirme que les valeurs sont proches de la moyenne et l'erreur standard aux alentours de 1.5 %.

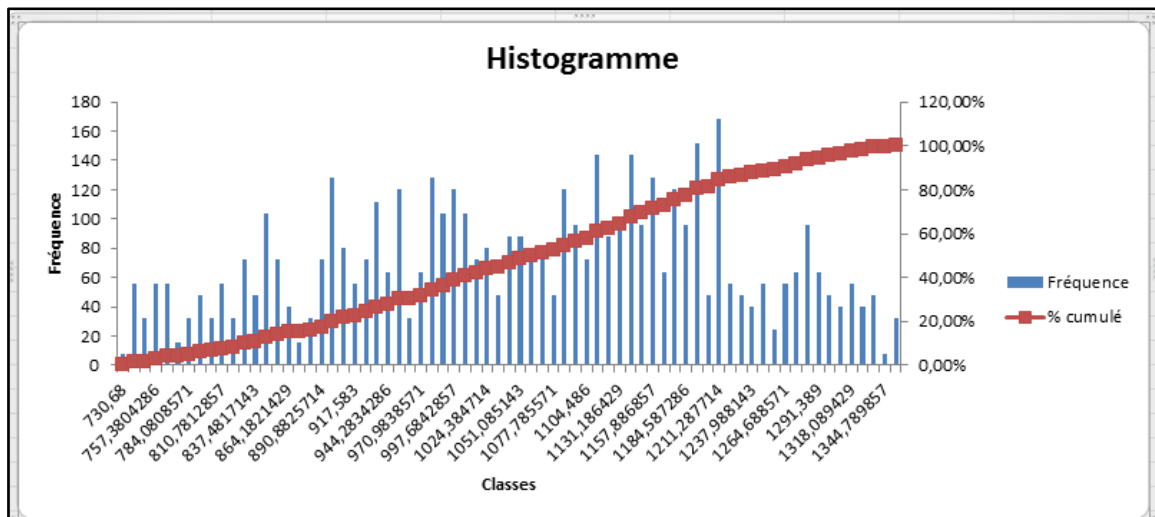
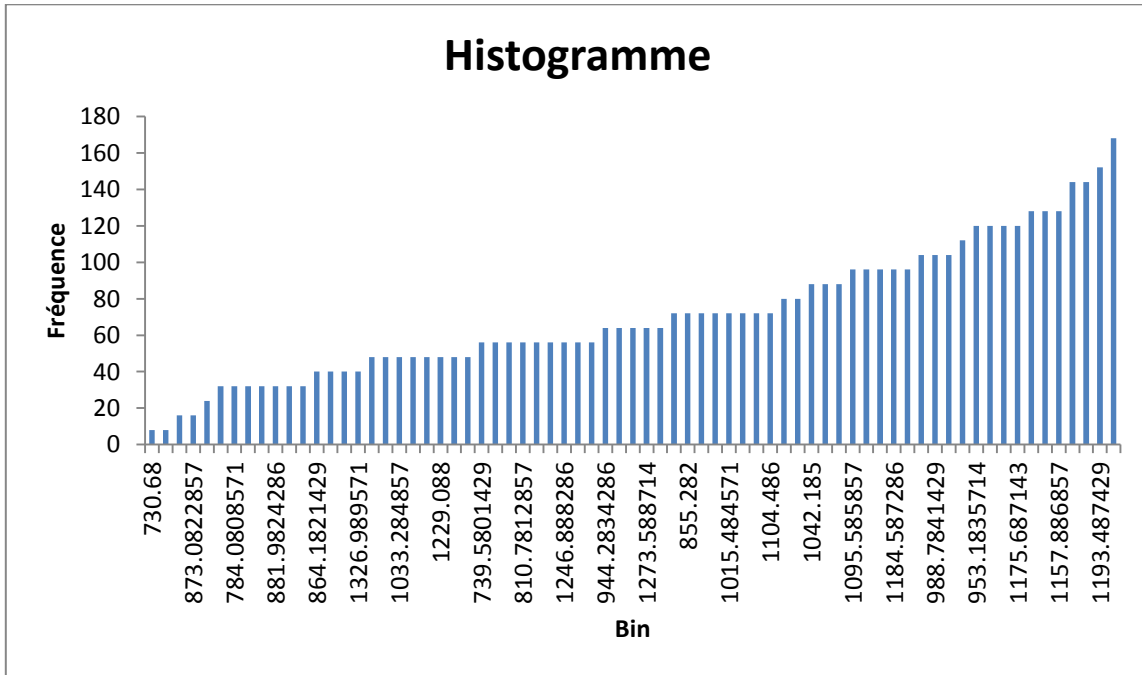


Figure 5-1 Histogrammes Pareto et cumulatives

Les histogrammes démontrent que les grandes valeurs sont surtout à droite de la moyenne (kurtosis est à une valeur négative), mais la dissymétrie est à 0.

Les chances de tomber sur des valeurs bonnes de coûts sont petites si les bonnes méthodes ne sont pas appliquées. En fait la probabilité est de huit (nombre des apparences de plus petit coût) sur 5040 (nombre d'échantillons) environ 0,15 %.

5.2.3 Statistiques inférentielles

Le défi est de trouver la corrélation entre les deux variables, séquence et coût et prouver l'effet de manière scientifique. Les corrélations mesurent la force de la relation, ou d'association, entre deux variables, mais elles n'impliquent pas nécessairement la causalité. Le coefficient de corrélation de l'échantillon est noté r tandis que la corrélation de population est notée par ρ ou R .

Le signe et la valeur absolue d'une corrélation décrivent la direction et l'ampleur de la relation entre deux variables. Pour interpréter les corrélations:

- la valeur d'un coefficient de corrélation est comprise entre -1 et 1.
- plus la valeur absolue est égale à 1, plus la relation linéaire est forte.
- plus proche de 0, plus faible est la relation.

Une corrélation près de 0 ne signifie pas nécessairement qu'il n'y a pas de relation. Cela signifie seulement que la relation n'est pas linéaire (elle pourrait être curviligne ou logarithmique, par exemple).

Le signe de la corrélation indique la direction de la relation. Les corrélations positives signifient que les variables se déplacent dans la même direction. Les corrélations négatives signifient que les variables se déplacent dans des directions opposées.

Par défaut, Excel possède un outil limité à l'analyse d'une seule variable d'entrée sans prendre en considération l'impact d'un ensemble de variables. L'analyse est basée sur la meilleure position d'un élément d'infrastructure E_i et son impact sur le meilleur coût. Un échantillon de toutes les permutations calculées est présenté dans le tableau 5.8.

Tableau 5.8 Tableau avec la position d'un EI dans la séquence et le coût

| Séquence | E1 | E2 | E3 | E4 | E5 | E6 | E7 | Coût(\$) |
|-----------------------|----|----|----|----|----|----|----|-----------|
| (4, 2, 1, 3, 5, 6, 7) | 2 | 1 | 3 | 0 | 4 | 5 | 6 | 730,68 \$ |
| (4, 2, 1, 3, 5, 7, 6) | 2 | 1 | 3 | 0 | 4 | 6 | 5 | 730,68 \$ |
| (4, 3, 1, 2, 5, 6, 7) | 2 | 3 | 1 | 0 | 4 | 5 | 6 | 730,68 \$ |
| (4, 3, 1, 2, 5, 7, 6) | 2 | 3 | 1 | 0 | 4 | 6 | 5 | 730,68 \$ |
| (5, 2, 1, 3, 4, 6, 7) | 2 | 1 | 3 | 4 | 0 | 5 | 6 | 730,68 \$ |
| (5, 2, 1, 3, 4, 7, 6) | 2 | 1 | 3 | 4 | 0 | 6 | 5 | 730,68 \$ |
| (5, 3, 1, 2, 4, 6, 7) | 2 | 3 | 1 | 4 | 0 | 5 | 6 | 730,68 \$ |
| (5, 3, 1, 2, 4, 7, 6) | 2 | 3 | 1 | 4 | 0 | 6 | 5 | 730,68 \$ |
| (4, 2, 3, 1, 5, 6, 7) | 3 | 1 | 2 | 0 | 4 | 5 | 6 | 731,90 \$ |
| (4, 2, 3, 1, 5, 7, 6) | 3 | 1 | 2 | 0 | 4 | 6 | 5 | 731,90 \$ |
| (4, 3, 2, 1, 5, 6, 7) | 3 | 2 | 1 | 0 | 4 | 5 | 6 | 731,90 \$ |
| (4, 3, 2, 1, 5, 7, 6) | 3 | 2 | 1 | 0 | 4 | 6 | 5 | 731,90 \$ |
| (5, 2, 3, 1, 4, 6, 7) | 3 | 1 | 2 | 4 | 0 | 5 | 6 | 731,90 \$ |
| (5, 2, 3, 1, 4, 7, 6) | 3 | 1 | 2 | 4 | 0 | 6 | 5 | 731,90 \$ |
| Etc... | | | | | | | | |

L'expérimentation a permis d'analyser, dans une première phase, l'impact de la position de chaque élément d'infrastructure sur le coût de migration, voir le tableau 5.9.

Tableau 5.9 Tableau corrélation de paires *position d'un EI dans la séquence et le coût)

| | Pos E1 | Pos E2 | Pos E3 | Pos E4 | Pos E5 | Pos E6 | Pos E7 |
|------|------------|------------|------------|------------|------------|-------------|-------------|
| Coût | 0.03847242 | 0.14960698 | 0.14960698 | 0.41700794 | 0.41700794 | -0.58585113 | -0.58585113 |

Les éléments avec le plus grand impact sont les éléments E6 et E7 qui sont les machines qui migrent avec des bases de données, plus lourdes en trafic réseau et coût d'infrastructure. Plus E6 et E7 migrent à la fin, plus l'impact est positif sur le coût de migration. Si E4, E5 migrent au début, le coût est plus bas.

L'outil Real Statistics a permis une analyse d'impact de toutes les variables sur le coût de migration. La fonction RSquare sert à ce calcul qui donne une valeur de 0,99. Les valeurs de positionnement sont calculées pour l'ensemble des éléments E0, E7. La valeur totale détermine le coefficient de régression cumulatif défini comme $RSquare(=RSquare(C1:I5040,J1:J5040))$.

Cette forte corrélation, même si elle ne signifie pas la causalité, appuie l'affirmation de notre hypothèse dans cet essai.

5.3 Analyse qualitative de résultats

L'analyse qualitative de résultats est effectuée avec une méthodologie qui ressemble à la méthode Delphi. Elle comporte au minimum trois tours d'avis et parfois plus, autant qu'il en faut pour aboutir à un maximum de consensus au sein d'un groupe. Chaque participant donne son avis (1), est informé des avis exprimés par les autres ainsi que des réactions par rapport à son propre avis (2) pour lui permettre de réagir en tentant de se rapprocher de la réponse consensuelle (3).

Les spécialistes choisis sont des experts en infrastructure, infonuagique et programmation. Leur expérience est très diversifiée et ils sont capables d'analyser chaque étape de cette méthodologie.

Tableau 5.10 Formulaire d'évaluation de la méthode et de résultats

| | Question | 1 Pauvre | 2 Passe | 3 Bien | 4 Très bien | 5 Excellent |
|---|--|-------------|------------|-----------|-------------------|----------------|
| 1 | La modélisation proposée est-elle correcte ? | | | | 4,7 | |
| 2 | Le mappage dans le nuage est-il bien conçu ? | | | 3,9 | | |
| 3 | Les contraintes sont-elles respectées ? | | | | 4,2 | |
| 4 | La méthodologie de migration suit-elle les règles du procédé ? | | | | 4,1 | |
| 5 | Les résultats confirment-ils l'hypothèse ? | | | | 4,7 | |

Quelques idées intéressantes ont contribué à enrichir cet essai :

- 1) Pendant la période de migration, l'existence d'un réseau virtuel privé (VPN) est absolument nécessaire. Ce réseau est payé à l'heure. Dans une migration 1 par 1 le coût de ce réseau est constant parce que le temps total de migration est toujours

constant. Par contre, dans les migrations de plusieurs éléments au même moment, ce prix varie et il doit être pris en compte.

- 2) Une confusion persiste par rapport aux prix de l'infonuagique. Présentement, c'est seulement la bande passante qui est facturée dans les cas Amazon et Azure.
- 3) L'algorithme initial prenait en compte le changement de la topologie réseau à partir de l'installation et la configuration d'un élément d'infrastructure. Une bonne observation a aidé à voir que le changement de la topologie se passe seulement quand on active le test. La formule a changé.
- 4) Une idée proposée est de trouver le modèle de coût pour plusieurs migrations au même moment et de le démontrer par le raisonnement par récursivité.
- 5) La suggestion d'explorer plusieurs outils infonuagiques a été proposée mais les outils connus (Cloudamize, Oracle, Microsoft) ne calculent pas le coût de la migration.

La moyenne d'évaluation obtenue est de 4,32, voir le tableau 5.10, ce qui est excellent car elle valide l'hypothèse et la réponse à la question de recherche et la qualité de cet essai.

5.4 Développements futurs

Certaines améliorations sont possibles avec des étapes intermédiaires qui éliminent la nécessité de calculer le coût de toutes les séquences avec l'algorithme de type tri en bulle. Les algorithmes de type « diviser pour régner » sont connus plus performants. Le nombre d'éléments d'infrastructure augmente la complexité des analyses et l'algorithme programmé fait une différence.

La période totale de migration est plus courte s'il est possible de migrer plusieurs éléments d'infrastructure en parallèle. Cette version d'algorithme est quantifiable et modélisable pour le meilleur coût dans n'importe quelle situation, mais pour des raisons de risque, cet essai limite la migration à une machine à la fois.

La recherche se limite au service IaaS, même s'il est possible de migrer dans un nuage PaaS. Cependant, la phase de mappage est différente, car les blocs d'architecture et les composants technologiques sont plus complexes.

La seule variable considérée dans l'étude, pour optimiser le coût, est la séquence de migration. Une autre variable possible est le degré de virtualisation. Le degré de virtualisation change la vitesse d'installation, de configuration et de test de chaque élément d'infrastructure. Il y a aussi, la consommation de la bande passante et la consommation des ressources qui sont d'autres variables à étudier.

Une variable possible est la complexité cyclomatique d'un système logique et son influence sur le coût de migration. Il est possible que le sujet soit beaucoup plus complexe, mais il serait intéressant d'étudier son effet sur le coût de la migration, surtout de prédire s'il existe une liaison entre le coût de migration et cette variable. Le but serait de quantifier le coût avant la migration.

Enfin, toutes ces variables peuvent faire l'objet de recherches futures.

Conclusion

L'infonuagique représente un concept qui transcende la manière de livrer des services aux différents clients que ce soit pour les entreprises ou le grand public. Cette idée permet de réaliser des économies substantielles tout en éliminant des tâches associées à la maintenance ainsi que la gestion de l'infrastructure informatique exigeante et parfois complexe pour le client.

Cet essai aborde un problème relié à la migration des applications dans le nuage informatique, afin de calculer et de réduire le coût de la migration.

L'essai étudie les principales technologies et bonnes pratiques qui favorisent une migration haute-disponibilité vers l'infonuagique. Une modélisation abstraite est utilisée pour séparer les éléments qui influencent le coût de migration. Puis, le développement d'une stratégie permet de trouver la meilleure séquence de migration selon les contraintes.

Le modèle proposé, est conçu pour assurer la liberté du choix de l'environnement de destination, d'un environnement traditionnel à un environnement infonuagique quelconque et vice-versa. Cela garantit l'interopérabilité avec toutes les plateformes de nuage et la portabilité des applications à la fois empêchant l'enfermement propriétaire sur notre plateforme. Les outils utilisés sont simples et les interactions automatisées accélèrent l'identification de la séquence optimale de migration. La solution proposée comprend un algorithme qui trouve le meilleur coût.

Deux cas d'études concrets ont été réalisés pour tester et valider le modèle, sur Amazon et Azure, sur une architecture de système très populaire sur le marché. Les résultats obtenus sont concluants et démontrent l'efficacité de ce modèle pour analyser la migration de différents types d'applications pour toutes les architectures possibles.

Cet essai a quelques contributions importantes. Premièrement, la modélisation proposée d'un système logique est universelle et permet la séparation des éléments qui influencent le coût d'une migration en nuage. En fait, le modèle réduit chaque système logique dans un graphe avec plusieurs nœuds. Le coût, développé selon les éléments de ce modèle, est

confirmé avec de résultats pratiques. La stratégie pour obtenir le meilleur coût utilise la théorie de graphes et le chemin critique, définie comme une liste ordonnée des opérations nécessaires pour obtenir un résultat voulu. Dans notre cas, le résultat voulu est le coût minimal.

Donc, pour répondre à la question de la recherche, oui, il est possible de trouver une stratégie qui calcule le meilleur coût pour la migration en nuage et la stratégie trouvée optimise le processus de migration en nuage.

Les prochains paragraphes adressent les travaux futurs possibles qui sont au-delà de la portée de cet essai.

La complexité des systèmes devient de plus en plus grande et aujourd'hui la modélisation peut être effectuée avec des moyens automatisés. La simulation de coûts est possible, mais les solutions sur le marché considèrent surtout les ressources consommées pour simuler le coût en nuage. Il est possible de développer une solution intégralement automatisée pour la stratégie présentée dans cet essai.

L'essai limite la migration à un élément à la fois. Mathématiquement, il est possible de changer les contraintes et développer les formules pour les cas de migrations multiples d'éléments d'infrastructure et les démontrer avec le raisonnement par récursivité.

Une autre direction possible est de calculer la complexité d'un système logique et prédire le coût de la bande passante. C'est fortement probable qu'il y ait une liaison entre la complexité cyclomatique et le coût de migration d'un système assumant une consommation de ressources données.

Finalement, il est possible d'analyser aussi une migration PaaS utilisant le même patron, mais c'est plus complexe. Le mappage des éléments est plus difficile parce qu'il remplace les composants technologiques et pas seulement les éléments d'infrastructure.

Références

- [1] R. B. Caesar Wu, Cloud Data Centers and Cost Modeling: A Complete Guide To Planning, Designing and Building a Cloud Data Center, Morgan Kaufman, 2014, p. 848 p..
- [2] NIST Cloud Computing Security Reference Architecture.
- [3] G. Group, «Hype Cycle for Infrastructure Strategies, 2016,» 2016.
- [4] Gartner Group, «Hype Cycle for Compute Infrastructure, 2016».
- [5] E. a. Microsoft, Architecting Microsoft Azure Solutions, 2016.
- [6] ForgeRock, «Architecture of OpenAM and OpenDJ,» 2015.
- [7] Gartner Group, «Top 10 Strategic Technology Trends for 2017: Mesh App and Service Architecture,» 2017.
- [8] J. Lin, Migrating Applications to Clouds with BPM and TOGAF Framework, 2015.
- [9] R. B. Caesar Wu, Cloud Data Centers and Cost Modeling: A Complete Guide To Planning, Designing and Building a Cloud Data Center, Morgan Kaufman, 2014,

p. 848 p..

[10] H. Geng, Data Center Handbook, 2015.

[11] EDX Microsoft, Building Cloud Apps with Microsoft Azure, 2016.

[12] Microsoft , Cloud patterns, 2016.

[13] C. a. U. o. Illinois, Cloud Networking, 2016.

[14] C. a. U. o. Illinois, Cloud Computing Applications, 2016.

[15] E. a. LinuxFoundationX, LFS151x: Introduction to Cloud Infrastructure Technologies.

[16] D. C. Marinescu, Cloud Computing, Morgan Kaufmann, 2013, p. 503.

[17] Amazon, AWS Database Migration Service Best Practices, 2015.

[18] Nick Feamster, Princeton University, Software Defined Networking.

[19] M. Washam, Automating Microsoft Azure Infrastructure Services, 2015.

[20] S. Laan, IT Infrastructure Architecture, 2013.

[21] Carlos Becker Westphall, Federal University of California, CLOUD COMPUTING 2015 -
The Sixth International Conference on Cloud Computing, GRIDs, and

Virtualization, 2015, p. 169.

[22] A. J. Elmore, Zephyr: Live Migration in Shared Nothing Databases for Elastic Cloud Platforms, 2014.

[23] Planning for Beneficial Migration of Enterprise Applications to the Cloud, 2014.

[24] K. L. A. F. A. L. a. J. K. Van T. K. Tran, Size Estimation of Cloud Migration Projects with Cloud Migration Point (CMP), School of Information Technologies, The University of Sydney, Australia.

[25] Amazon, Amazon pricing calculator, 2016.

[26] S. D. D. A. A. E. A. Aaron J. Elmore, «Zephyr: Live Migration in Shared Nothing Databases for,» *SIGMOD'11, June 12–16, 2011, Athens, Greece..*

[27] V. S. S. L. F. Andrikopoulos, Decision Support for Application Migration to the Cloud : Challenges and Vision, 2014.

[28] A. A. a. C. P. M. I. Pooyan Jamshidi, Cloud Migration Research: A Systematic Review, 2015.

[29] M. a. Ranjan, Migrating multi-tier application to IaaS, 2012.

[30] W. M. Zhao Yangyang, «Performance Modeling of Virtual Machine Live Migration,» *School of Computing and Information Sciences Florida, 2012.*

[31] C. L. B. R. Zhang Qi, Cloud computing : state-of-the-art and research challenges, 2015.

[32] The Open Group, Togaf 9.1, 2011.

[33] Microsoft, Azure Moving Apps to Cloud, 2014.

[34] B. C. Tak, To move or not to move: the economics of cloud computing.

[35] Amazon EC2, <https://aws.amazon.com/ec2/>.

[36] Amazon S3, <https://aws.amazon.com/s3/>.

[37] Microsoft Azure, <https://azure.microsoft.com/en-us/pricing/calculator/>, 2016.

[38] Wikipedia, <https://www.wikipedia.org>.

[39] M. Lutz, Programming Python, 2016.

[40] Gartner Group, «Magic Quadrant for Cloud Infrastructure as a Service, Worldwide,»
2016.

[41] M. a. Ranjan, CloudGenius: Decision Support for WebServer Cloud Migration, 2012.

[42] L. Leong, Three Journeys define migrating a data center to Cloud IaaS, 2016.

[43] U. a. G. T. University, Health Informatics in the Cloud.

[44] Loto-Québec, «Architecture of external authentication infrastructure,» 2016.

Bibliographie

- [1] R. B. Caesar Wu, Cloud Data Centers and Cost Modeling: A Complete Guide To Planning, Designing and Building a Cloud Data Center, Morgan Kaufman, 2014, p. 848
- [2] The Open Group, Togaf 9.1, 2011.
- [3] K. L. A. F. A. L. a. J. K. Van T. K. Tran, Size Estimation of Cloud Migration Projects with Cloud Migration Point (CMP), School of Information Technologies, The University of Sydney, Australia.
- [4] H. Geng, Data Center Handbook, 2015.

Annexe I Mappage

Les sites Internet Amazon utilisés pour le mappage par ressources sont <https://aws.amazon.com/fr/ec2/instance-types/>. Le site suivant <https://calculator.s3.amazonaws.com/index.html> est utilisé pour le calcul de prix.

Storage Optimized I3 – High I/O Instances

This family includes the High Storage Instances that provide Non-Volatile Memory Express (NVMe) SSD backed instance storage optimized for low latency, very high random I/O performance, high sequential read throughput and provide high IOPS at a low cost.

Features:

- High Frequency Intel Xeon E5-2686 v4 (Broadwell) Processors with base frequency of 2.3 GHz
- NVMe SSD Storage
- Support for TRIM
- Support for Enhanced Networking
- High Random I/O performance and High Sequential Read throughput

Figure I-1 Recommandation Amazon pour les machines I3

| | | | | | | | |
|------------|---|-------|-------------------|------------|-----|---------|---------------|
| i3.large | 2 | 15.25 | 1 x 475 NVMe SSD | 10 Gigabit | Yes | \$0.216 | \$0.126 (42%) |
| i3.xlarge | 4 | 30.5 | 1 x 950 NVMe SSD | 10 Gigabit | Yes | \$0.372 | \$0.192 (48%) |
| i3.2xlarge | 8 | 61.0 | 1 x 1900 NVMe SSD | 10 Gigabit | Yes | \$0.754 | \$0.394 (48%) |

Figure I-2 Prix Amazon I3

À partir de 5.1.1, 5.1.2 et 5.1.3., sur le site d'Amazon les choix confirmés par les architectes pour le mappage sont :

- a) Le choix pour instances applicatives Web est pour les instances EC2 **i3.large**. C'est un produit qui nécessite de la consommation de la bande passante grande et présente un excellent fit avec les caractéristiques de 5.1.2 et 5.1.3.

- b) Le choix pour l'instance applicative Tomcat est une machine EC2 **i3.xlarge** pour les mêmes raisons.
- c) Le choix pour les instances BD Oracle est une machine EC2 **i3.2xlarge**

Tableau I.1 Tableau de blocs payables dans le nuage informatique

| Modèle | vCPU | Mém. (Gio) | Stockage (Gio) | Prix \$/jour |
|-------------------|----------|------------|----------------|--------------|
| I3.large | 2 | 15 | 20 | 0,126 |
| I3.xlarge | 4 | 30 | 40 | 0,192 |
| I3.2xlarge | 8 | 61 | 80 | 0,394 |

Les recommandations Azure sont :

| Type | Sizes | Description |
|--------------------------|-----------------------------|--|
| General purpose | DSv2, Dv2, DS, D, Av2, A0-7 | Balanced CPU-to-memory ratio. Ideal for testing and development, small to medium databases, and low to medium traffic web servers. |
| Compute optimized | Fs, F | High CPU-to-memory ratio. Good for medium traffic web servers, network appliances, batch processes, and application servers. |
| Memory optimized | GS, G, DSv2, DS | High memory-to-core ratio. Great for relational database servers, medium to large caches, and in-memory analytics. |
| Storage optimized | Ls | High disk throughput and IO. Ideal for Big Data, SQL, and NoSQL databases. |
| GPU | NV, NC | Specialized virtual machines targeted for heavy graphic rendering and video editing. Available with single or multiple GPUs. |
| High performance compute | H, A8-11 | Our fastest and most powerful CPU virtual machines with optional high-throughput network interfaces (RDMA). |

Figure I-3 Recommandations Azure

Le site <https://azure.microsoft.com/en-ca/pricing/calculator/> est utilisé pour le mappage en Azure.

Tableau I.2 Tableau de blocs payables dans le nuage informatique Azure

| Modèle | vCPU | Mém. (Gio) | Stockage (Gio) | Prix \$/jour |
|--------|------|------------|----------------|--------------|
|--------|------|------------|----------------|--------------|

| | | | | |
|-----------|----------|-----------|-----------|--------------|
| F4 | 4 | 8 | 20 | 0,724 |
| F4 | 4 | 8 | 40 | 0,724 |
| L8 | 8 | 32 | 80 | 2,621 |

Plusieurs preuves de tests montrent un prix de 0,1\$/heure par 1Gio pour la bande réseau pour les deux vendeurs. Les choix ne sont pas pour faire une comparaison entre les deux vendeurs.

Annexe II Analyse de coût

Pour commencer la vérification, l'algorithme est testé dans une première phase manuellement. La séquence utilisée pour cette calcul est : (2, 1, 4, 3, 6, 5, 7). La procédure manuelle est donnée dans les pages suivantes pour le cas d'Amazon.

Les figures suivantes expliquent phase par phase l'application de l'algorithme.

| Coût initial 0 | |
|-----------------------------------|---------------------|
| Aucun élément migré dans le nuage | |
| | |
| Centre de données | Nuage |
| E1 | Aiguilleur |
| E2 | Machine pour IA web |
| E3 | Machine pour IAweb |
| E4 | Machine pour IA |
| E5 | Machine pour IA |
| E6 | Machine pour BD |
| E7 | Machine pour BD |
| | |
| | |
| | |

Figure II-1 Aucun élément migré

| Centre de données | Nuage | Prix/heure EC2 E1(2) Web | Prix/jour | |
|---------------------|-------------|--------------------------|----------------------------|--------------|
| E1 | | | | |
| | E2 | 0,156 | 3,744 | |
| E3 | | | | |
| E4 | | | | |
| E5 | | | | |
| E6 | | | | |
| E7 | | | | |
| | | | | |
| #JRs | RH6.0 | Apache 2.4 | Magnolia Web | |
| Installation E2 | 0,5 | 0,1 | 0,2 | |
| Configuration E2 | 0,25 | 0,1 | 0,3 | |
| Test E2 | 0,25 | 0,2 | 1 | |
| | | | | |
| Réseau pendant test | Nuage | Centre de Données | Ei vers Ej (Gio à l'heure) | Gio par jour |
| | E2 | E1 | 0,5 | 12 |
| | E2 | E4 | 0,5 | 12 |
| | E2 | E5 | 0,5 | 12 |
| | | | | 36 |
| | | | | |
| Coût | Coût Réseau | Coût Utilisation EC2 | | |
| 16,0776 | 5,22 | 10,8576 | | |

Figure II-2 Élément E2 migré

| Centre de données | Nuage | Prix/heure EC2 E1(2) Web | Prix/jour | |
|---------------------|-------------------|--------------------------|------------------------------|---------------------------|
| | E1 | 0,025 | 0,6 | |
| | E2 | 0,156 | 3,744 | |
| E3 | | | | |
| E4 | | | | |
| E5 | | | | |
| E6 | | | | |
| E7 | | | | |
| | | | | |
| #Jours | Acos 4.0 | | | |
| Installation E1 | 0,5 | | | |
| Configuration E1 | 0,5 | | | |
| Test E1 | 1 | | | |
| | | | | |
| Réseau pendant test | Nuage | Centre de Données | Ei vers Ej (Gio à l'heure) | Ei vers Ej (Gio par jour) |
| | E1 | E0 | 1,1 | 26,4 |
| | E1 | E3 | 0,6 | 14,4 |
| | E2 | E4 | 0,5 | 12 |
| | E2 | E5 | 0,5 | 12 |
| | | | | 64,8 |
| | | | | |
| Coût | Coût Réseau 2 (T) | Coût Utilisation EC2 | Coût réseau résiduel 1 (I+C) | |
| 18,768 | 6,48 | 8,688 | 3,6 | |

Figure II-3 Élément E1 migré

| Centre de données | Nuage | Prix/heure EC2 El(2) Web | Prix/jour | |
|---------------------|-------------------|--------------------------|------------------------------|---------------------------|
| | E1 | 0,025 | 0,6 | |
| | E2 | 0,156 | 3,744 | |
| E3 | | | | |
| | E4 | 0,312 | 7,488 | |
| E5 | | | | |
| E6 | | | | |
| E7 | | | | |
| #Jours | Rh6 | Tomcat7 | Magnolia App | |
| Installation E4 | 0,5 | 0,2 | 0,2 | |
| Configuration E4 | 0,25 | 0,2 | 0,3 | |
| Test E4 | 0,25 | 0,4 | 5 | |
| Réseau pendant test | Nuage | Centre de Données | Ei vers Ej (Gio à l'heure) | Ei vers Ej (Gio par jour) |
| | E1 | E0 | 1,1 | 26,4 |
| | E1 | E3 | 0,6 | 14,4 |
| | E2 | E5 | 0,5 | 12 |
| | E4 | E6 | 0,75 | 18 |
| | E4 | E3 | 0,45 | 10,8 |
| | E4 | E7 | 0,75 | 18 |
| | | | | 99,6 |
| Coût | Coût Réseau 2 (T) | Coût Utilisation EC2 | Coût réseau résiduel 1 (I+C) | |
| | 153,3396 | 56,274 | 86,3736 | 10,692 |

Figure II-4 Élément E4 migré

| Centre de données | Nuage | Prix/heure EC2 El(2) Web | Prix/jour | |
|---------------------|---------------|--------------------------|------------------------------|---------------------------|
| | E1 | 0,025 | 0,6 | |
| | E2 | 0,156 | 3,744 | |
| | E3 | 0,156 | 3,744 | |
| | E4 | 0,312 | 7,488 | |
| E5 | | | | |
| E6 | | | | |
| E7 | | | | |
| #JRs | RH6 | Apache 2.4 | Magnolia Web | |
| Installation E3 | 0,5 | 0,1 | 0,2 | |
| Configuration E3 | 0,25 | 0,1 | 0,3 | |
| Test E3 | 0,25 | 0,2 | 1 | |
| Réseau pendant test | Nuage | Centre de Données | Ei vers Ej (Gio à l'heure) | Ei vers Ej (Gio par jour) |
| | E1 | E0 | 1,1 | 26,4 |
| | E2 | E5 | 0,5 | 12 |
| | E3 | E5 | 0,5 | 12 |
| | E4 | E6 | 0,75 | 18 |
| | E4 | E7 | 0,75 | 18 |
| | | | | 86,4 |
| Coût | Coût Réseau 2 | Coût Utilisation EC2 | Coût réseau résiduel 1 (I+C) | |
| | 72,1404 | 12,528 | 45,1704 | 14,442 |

Figure II-5 Élément E3 migré

| Centre de données | Nuage | Prix/heure EC2 E1(2) Web | Prix/jour | |
|---------------------|-------------|--------------------------|------------------------------|---------------------------|
| | E1 | 0,025 | 0,6 | |
| | E2 | 0,156 | 3,744 | |
| | E3 | 0,156 | 3,744 | |
| | E4 | 0,312 | 7,488 | |
| E5 | | | 0 | |
| | E6 | 0,624 | 14,976 | |
| E7 | | | 0 | |
| #JRs | RH6 | Oracle 12c | Magnolia BD | |
| Installation E6 | 0,5 | 0,5 | 0 | |
| Configuration E6 | 0,25 | 0,25 | 0 | |
| Test E6 | 0,25 | 0,25 | 0,2 | |
| Réseau pendant test | Nuage | Centre de Données | Ei vers Ej (Gio à l'heure) | Ei vers Ej (Gio par jour) |
| | E1 | E0 | 1,1 | 26,4 |
| | E2 | E5 | 0,5 | 12 |
| | E3 | E5 | 0,5 | 12 |
| | E6 | E5 | 1 | 24 |
| | E4 | E7 | 0,75 | 18 |
| | E6 | E7 | 2 | 48 |
| | | | | 140,4 |
| Coût | Coût Réseau | Coût Utilisation EC2 | Coût réseau résiduel 1 (I+C) | |
| | 90,0024 | 9,828 | 67,2144 | 12,96 |

Figure II-6 Élément E6 migré

| Centre de données | Nuage | Prix/heure EC2 E1(2) We | Prix/jour | |
|---------------------|---------------|-------------------------|------------------------------|---------------------------|
| | E1 | 0,025 | 0,6 | |
| | E2 | 0,156 | 3,744 | |
| | E3 | 0,156 | 3,744 | |
| | E4 | 0,312 | 7,488 | |
| | E5 | 0,312 | 7,488 | |
| | E6 | 0,624 | 14,976 | |
| E7 | | | 0 | |
| #Jours | Rh6 | Tomcat7 | Magnolia App | |
| Installation E5 | 0,5 | 0,2 | 0,2 | |
| Configuration E5 | 0,25 | 0,2 | 0,3 | |
| Test E5 | 0,25 | 0,4 | 5 | |
| Réseau pendant test | Nuage | Centre de Données | Ei vers Ej (Gio à l'heure) | Ei vers Ej (Gio par jour) |
| | E1 | E0 | 1,1 | 26,4 |
| | E5 | E7 | 0,75 | 18 |
| | E4 | E7 | 0,75 | 18 |
| | E6 | E7 | 2 | 48 |
| | | | | 110,4 |
| Coût | Coût Réseau 2 | Coût Utilisation EC2 | Coût réseau résiduel 1 (I+C) | |
| | 363,234 | 62,376 | 277,692 | 23,166 |

Figure II-7 Élément E5 migré

| Centre de données | Nuage | Prix/heure EC2 E1(2) Web | Prix/jour | |
|---------------------|-------------------|--------------------------|------------------------------|---------------------------|
| | E1 | 0,025 | 0,6 | |
| | E2 | 0,156 | 3,744 | |
| | E3 | 0,156 | 3,744 | |
| | E4 | 0,312 | 7,488 | |
| | E5 | 0,312 | 7,488 | |
| | E6 | 0,624 | 14,976 | |
| | E7 | 0,624 | 14,976 | |
| #JRs | RH6 | Oracle 12c | Magnolia BD | |
| Installation E7 | 0,5 | 0,5 | 0 | |
| Configuration E7 | 0,25 | 0,25 | 0 | |
| Test E7 | 0,25 | 0,25 | 0,2 | |
| Réseau pendant test | Nuage | Centre de Données | Ei vers Ej (Gio à l'heure) | Ei vers Ej (Gio par jour) |
| | E1 | E0 | 1,1 | 26,4 |
| | | | | 26,4 |
| Coût | Coût Réseau 2 (T) | Coût Utilisation EC2 | Coût réseau résiduel 1 (I+C) | |
| | 135,0432 | 1,848 | 116,6352 | 16,56 |

Figure II-8 Élément E7 migré

| Cout final | E2 | E1 | E4 | E3 | E6 | E5 | E7 |
|------------|-------|-------|--------|-------|-------|--------|--------|
| 848,61 | 16,08 | 18,77 | 153,34 | 72,14 | 90,00 | 363,23 | 135,04 |

Figure II-9 Fin de migration et coût total

Le script Python est utilisé pour les analyses statistiques et applique les algorithmes de 4.5 et 4.6. Le test effectué sur (2, 1, 4, 3, 6, 5, 7) et les autres séquences du tableau confirment la valeur de 848.61 calculée manuellement et aussi les autres.

| | | |
|-----------------------|------------------|----------|
| (2, 1, 4, 3, 5, 7, 6) | :::1:0:3:2:4:6:5 | : 739.76 |
| (2, 1, 4, 3, 6, 5, 7) | :::1:0:3:2:5:4:6 | : 848.61 |
| (2, 1, 4, 3, 6, 7, 5) | :::1:0:3:2:6:4:5 | : 896.61 |

Figure II-10 Sortie de script Python

Le script produit une sortie texte analysée dans Excel.

```

# exécute moi python Master_final.py > sortie.txt
# conda config --set ssl_verify false

#Importations
import numpy
from numpy import matrix
import itertools
#####
# Première partie initialisation variables
#####
# Composants technologiques ET JOURS DE CONGIGURATION ET TEST
# CT(0) = A10 Acos 4.0
# CT(1) = RedHat 6.0
# CT(2) = Apache 2.4
# CT(3) = Tomcat7
# CT(4) = Oracle 12c
# CT(5) = Magnolia Web
# CT(6) = Magnolia App
# CT(7) = Magnolia BD
M=7 #components techno 0..7
CT_TEMPS_IC= [1,0.75,0.2,0.4,0.75,0.5,0.5,0]
CT_TEMPS_T= [1,0.25,0.2,0.4,0.25,1,5,0.2]

# Éléments infrastructure
# N=8 éléments
# EI(0) = Internet
# EI(1) = Load Balancer utilise Cisco
# EI(2) = Application Web utilise RedHa6 et Apache et EBS Web
# EI(3) = Application Web utilise RedHat6 et Apache et EBS Web
# EI(4) = Instance applicative utilise RedHat6 et Jboss et EBS App
# EI(5) = Instance applicative utilise RedHat6 et Jboss et EBS App
# EI(6) = Base de Données utilise Oracle
# EI(7) = Base de Données utilise Oracle
N=7 #éléments infra, zero included

# Initialisation coût per EI par jour en nuage
# Les blocs nuage sont mappés aux blocs de centres de données
EI_COUT_CLOUD_MACHINE= [0,0.6,3.744,3.744,7.488,7.488,14.976,14.976]
#Amazon
#EI_COUT_CLOUD_MACHINE= [0,0,5.26,5.26,8.76,8.76,32.26,32.26] #Microsoft

# Generate EI
EI= [ i for i in range(N+1) ]

# Définitions des utilisations de CTs en EIs
# EI_CT(i,j)=1 si EI(i) utilise CT[j]
EI_CT= [ [ 0,0,0,0,0,0,0,0],[ 1,0,0,0,0,0,0,0],[ 0,1,1,0,0,1,0,0],[
0,1,1,0,0,1,0,0],[ 0,1,0,1,0,0,1,0],[ 0,1,0,1,0,0,1,0],[
0,1,0,0,1,0,0,1],[ 0,1,0,0,1,0,0,1] ]

```



```

j=0
if len(list_temp) > 0:
    j=list_temp[-1]
#print(list_temp)
if len(list_temp) > 1 :
    for k in list_temp[:-1]:
        #cout residuel
        if len(list_temp) > 1 :
            for l in range(N+1):
                if l not in list_temp[:-1]:

EI_COUT_CLOUD[k]=EI_COUT_CLOUD[k]+EI_TEMPS_IC[j]*24*EI_EI_RESEAU_CONS[k]
[l]*EI_COUT_CLOUD_UNIT_RESEAU
        #print('after loop reseau residuel ic=',i,EI_COUT_CLOUD)

        #somme étape par élément migré
        #somme totale SUM
    for k in list_temp:
        #fill the costs by time for the max_permutation
        #based on time multiplied with cost per element mapped
        #print(EI_COUT_CLOUD_MACHINE[k])

EI_COUT_CLOUD[k]=EI_COUT_CLOUD[k]+(EI_TEMPS_T[j]+EI_TEMPS_IC[j])*EI_COUT
_CLOUD_MACHINE[k]
        #calculer la bande passante de chaque EI dans le nuage
pendant la periode de test
        #parse EI_EI réseau
        #print('after loop i=',i,'k=',k,EI_COUT_CLOUD)

        if len(list_temp)>0:
            for l in range(N+1):
                if l not in list_temp:

EI_COUT_CLOUD[k]=EI_COUT_CLOUD[k]+EI_TEMPS_T[j]*24*EI_EI_RESEAU_CONS[k] [
l]*EI_COUT_CLOUD_UNIT_RESEAU
        #print('after loop reseau 1 t=',i,'k=',k,EI_COUT_CLOUD)
        #print EI_COUT_CLOUD[k]
        #somme étape par élément migré
        #somme totale SUM
#print EI_COUT_CLOUD
SUM=round(sum(EI_COUT_CLOUD),2)
#print (SUM)
#concatenation=''.join(map(str,lpermutation))
stringrank='';
for u in range(N+1):
    if u>0 :

stringrank=stringrank+';'+str(lpermutation.index(u))
print (lpermutation,';',stringrank,';',round(SUM,2))
#stopbreak=stopbreak+1

```

```

#if stopbreak>1:
#   stopbreak=stopbreak+1
#   break
if SUM > MIN_COST:
    MIN_COST=SUM
    MIN_PERMUTATION=lpermutation
if SUM < MAX_COST:
    MAX_COST=SUM
    MAX_PERMUTATION=lpermutation
# Initialisation
EI_COUT_CLOUD = [0,0,0,0,0,0,0,0]
#break

print (MAX_COST)
print (MAX_PERMUTATION)
print (MIN_COST)
print (MIN_PERMUTATION)

```

Le fichier **result.txt** est importé dans Excel et affiche le coût pour chaque séquence. De plus, le fichier contient les colonnes qui donnent l'ordre de chaque élément migré.

| A | B | C | D | E | F | G | H | I |
|-----------------------|----|----|----|----|----|----|----|---------|
| Séquence | E1 | E2 | E3 | E4 | E5 | E6 | E7 | Coût |
| (1, 2, 3, 4, 5, 6, 7) | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 734.59 |
| (1, 2, 3, 4, 5, 7, 6) | 0 | 1 | 2 | 3 | 4 | 6 | 5 | 734.59 |
| (1, 2, 3, 4, 6, 5, 7) | 0 | 1 | 2 | 3 | 5 | 4 | 6 | 843.43 |
| (1, 2, 3, 4, 6, 7, 5) | 0 | 1 | 2 | 3 | 6 | 4 | 5 | 891.43 |
| (1, 2, 3, 4, 7, 5, 6) | 0 | 1 | 2 | 3 | 5 | 6 | 4 | 843.43 |
| (1, 2, 3, 4, 7, 6, 5) | 0 | 1 | 2 | 3 | 6 | 5 | 4 | 891.43 |
| (1, 2, 3, 5, 4, 6, 7) | 0 | 1 | 2 | 4 | 3 | 5 | 6 | 734.59 |
| (1, 2, 3, 5, 4, 7, 6) | 0 | 1 | 2 | 4 | 3 | 6 | 5 | 734.59 |
| (1, 2, 3, 5, 6, 4, 7) | 0 | 1 | 2 | 5 | 3 | 4 | 6 | 843.43 |
| (1, 2, 3, 5, 6, 7, 4) | 0 | 1 | 2 | 6 | 3 | 4 | 5 | 891.43 |
| (1, 2, 3, 5, 7, 4, 6) | 0 | 1 | 2 | 5 | 3 | 6 | 4 | 843.43 |
| (1, 2, 3, 5, 7, 6, 4) | 0 | 1 | 2 | 6 | 3 | 5 | 4 | 891.43 |
| (1, 2, 3, 6, 4, 5, 7) | 0 | 1 | 2 | 4 | 5 | 3 | 6 | 982.93 |
| (1, 2, 3, 6, 4, 7, 5) | 0 | 1 | 2 | 4 | 6 | 3 | 5 | 1030.94 |
| (1, 2, 3, 6, 5, 4, 7) | 0 | 1 | 2 | 5 | 4 | 3 | 6 | 982.93 |
| (1, 2, 3, 6, 5, 7, 4) | 0 | 1 | 2 | 6 | 4 | 3 | 5 | 1030.94 |
| (1, 2, 3, 6, 7, 4, 5) | 0 | 1 | 2 | 5 | 6 | 3 | 4 | 1109.6 |
| (1, 2, 3, 6, 7, 5, 4) | 0 | 1 | 2 | 6 | 5 | 3 | 4 | 1109.6 |
| (1, 2, 3, 7, 4, 5, 6) | 0 | 1 | 2 | 4 | 5 | 6 | 3 | 982.93 |
| (1, 2, 3, 7, 4, 6, 5) | 0 | 1 | 2 | 4 | 6 | 5 | 3 | 1030.94 |
| (1, 2, 3, 7, 5, 4, 6) | 0 | 1 | 2 | 5 | 4 | 6 | 3 | 982.93 |
| (1, 2, 3, 7, 5, 6, 4) | 0 | 1 | 2 | 6 | 4 | 5 | 3 | 1030.94 |
| (1, 2, 3, 7, 6, 4, 5) | 0 | 1 | 2 | 5 | 6 | 4 | 3 | 1109.6 |
| (1, 2, 3, 7, 6, 5, 4) | 0 | 1 | 2 | 6 | 5 | 4 | 3 | 1109.6 |
| (1, 2, 4, 3, 5, 6, 7) | 0 | 1 | 3 | 2 | 4 | 5 | 6 | 741.64 |
| (1, 2, 4, 3, 5, 7, 6) | 0 | 1 | 3 | 2 | 4 | 6 | 5 | 741.64 |
| (1, 2, 4, 3, 6, 5, 7) | 0 | 1 | 3 | 2 | 5 | 4 | 6 | 850.48 |
| (1, 2, 4, 3, 6, 7, 5) | 0 | 1 | 3 | 2 | 6 | 4 | 5 | 891.43 |

Sous-chapitres 5.2.2 et 5.2.3 sont dédiées aux analyses statistiques effectuées dans Excel. Le paquetage supplémentaire utilisé provient de site <http://www.real-statistics.com/>.

Une présentation plus ample est donnée sur le site <http://www.real-statistics.com/correlation/multiple-correlation/>.

Si y est une variable dépendante (variable de réponse) et x_1, \dots, x_k sont des variables indépendantes (variables prédictives), alors le modèle de régression multiple fournit une prédiction de y à partir du x_i de la forme :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (\text{II.1})$$

Le paquetage de ressources de statistiques réelles fournit la fonction de tableau RSquare (R1, R2) qui calcule cette valeur où R1 est un réseau $n \times k$ contenant les données d'échantillon X et R2 est un réseau $n \times 1$ contenant les données d'échantillons Y. Le coefficient de corrélation peut être calculé par la formule = SQRT (RSquare (R1, R2)).

La fonction RSQUARE permet la corrélation multiple entre un ensemble de variables indépendantes (numéros dans la séquence de migration d'un élément EI) et la variable dépendante Coût.

Annexe III Méthode de points de migration dans le nuage

Cet essai a été possible grâce à des idées provenant de [5] et [20]. Voilà un résumé de [20].

Comme l'effort est nécessaire pour migrer vers un nuage et la quantité d'efforts requis est diversifiée, l'estimation de l'effort pour un projet de migration vers une plate-forme infonuagique est essentielle pour la gestion de projet, en particulier la planification des projets et la planification budgétaire. Étant donné que l'infonuagique est nouveau, il n'y a pas suffisamment de données accessibles au public en ce qui concerne les efforts de migration en nuage. Par conséquent, au mieux de notre connaissance, aucune estimation d'effort n'a été proposée spécifiquement conçue pour les projets de migration en nuage, alors que les approches d'estimation d'effort traditionnel existantes pour le développement de logiciels ne sont pas applicables dans ce contexte.

Les auteurs ont développé une métrique spécifique aux points de fonction et aux infonuagiques, appelé Cloud Migration Point (CMP), pour mesurer la taille d'un projet de migration de nuage, qui sert alors de base à l'estimation de l'effort de migration en nuage. Dans cet article, les auteurs présentent l'approche pour développer la CMP et sa méthode de comptage, ainsi que des validations empiriques théoriques et initiales. Les facteurs de coût suivants ont été trouvés :

1) Installation et configuration (CMP_{ic}) - Lors de la migration vers un nuage IaaS tel que Amazon EC2, un effort est nécessaire pour installer le logiciel système nécessaire, les serveurs de base de données ou les intergiciels; les variables d'environnement et les paramètres doivent également être configurés. Lors de la migration vers un nuage PaaS tel que Microsoft Azure, l'installation et l'effort de configuration résident dans la couche d'application, comme les bibliothèques ou les plug-ins.

2) Changements de base de données (CMP_{bd}) - La migration d'une base de données vers le nuage peut entraîner des modifications de schéma de base de données et des changements

de requêtes en raison de différences dans les versions, les variantes (MySQL vs MSSQL) ou les types de bases de données (Relationnel vs. NoSQL).

3) Changements de code (CMPcode) - Dans certains cas de migration, la modélisation du code est nécessaire pour s'adapter au nouveau modèle de programmation en nuage, ou la couche d'accès à la base de données doit être modifiée pour fonctionner en toute transparence avec différentes bases de données en nuage. Les types de changements de code sont groupés en :

- Type de domaine de problème (PDT): classes qui représentent des entités du monde réel dans le domaine d'application système.
- Type d'interaction humaine (HIT): classes conçues pour la visualisation de l'information et l'interaction homme-ordinateur.
- Type de gestion de données (DMT): classes qui assurent le stockage et la récupération des données.
- Type de gestion des tâches (TMT): classes qui sont responsables de la définition et du contrôle des tâches.

4) Changements de connexion (CMPconn) - Dans un système S avant migration, la connexion entre deux composants A et B est une connexion LAN. Si seulement B est migré vers un nuage et que A est conservé dans un centre de données local, la connexion LAN entre A et B devient une connexion WAN. Si A et B sont migrés vers le nuage, la connexion LAN entre A et B devient une connexion LAN en nuage (les conditions du réseau peuvent être différentes en nuage).

Table VII
EVALUATING CMP'S COMPONENTS

| CMP | Type | Complexity Level | | | Weight |
|---------------------|--------------------|------------------|---------|------|--------|
| | | Low | Average | High | |
| CMP _{conn} | LAN-to-LAN | 1 | 3 | 4 | 3 |
| | LAN-to-WAN | 1 | 6 | 9 | |
| | WAN-to-LAN | 1 | 6 | 9 | |
| CMP _{code} | PDT | 3 | 6 | 10 | 5 |
| | HIT | 4 | 7 | 12 | |
| | DMT | 5 | 8 | 13 | |
| | TMT | 4 | 6 | 9 | |
| CMP _{ic} | Application | 1 | 2 | 7 | 2 |
| | Infrastructure | 1 | 3 | 9 | |
| CMP _{db} | Query Modification | 1 | 3 | 8 | 1 |
| | Data Population | 3 | 4 | 10 | |

Figure III-1 Points de migration en infonuagique (pondération)

La somme des toutes ces pondérations donne une estimation d'effort. Elle ressemble, mais elle n'est pas la même dans cet essai.

$$CMP = \sum_{i=0}^3 CMP_i \times w_i \quad (III.1)$$

Les auteurs continuent avec une validation théorique et pratique de cette estimation. L'essai considère seulement CMP_{ic} et CMP_{conn} et s'en va dans la direction d'un calcul de coût selon l'arbre de séquences de migration.

Annexe IV Vendeurs IaaS et PaaS

Le choix d'Amazon et Microsoft pour cette recherche est expliqué par ce rapport qui provient aussi de Gartner Group. Le meilleur cadran est occupé par ces deux vendeurs de IaaS.

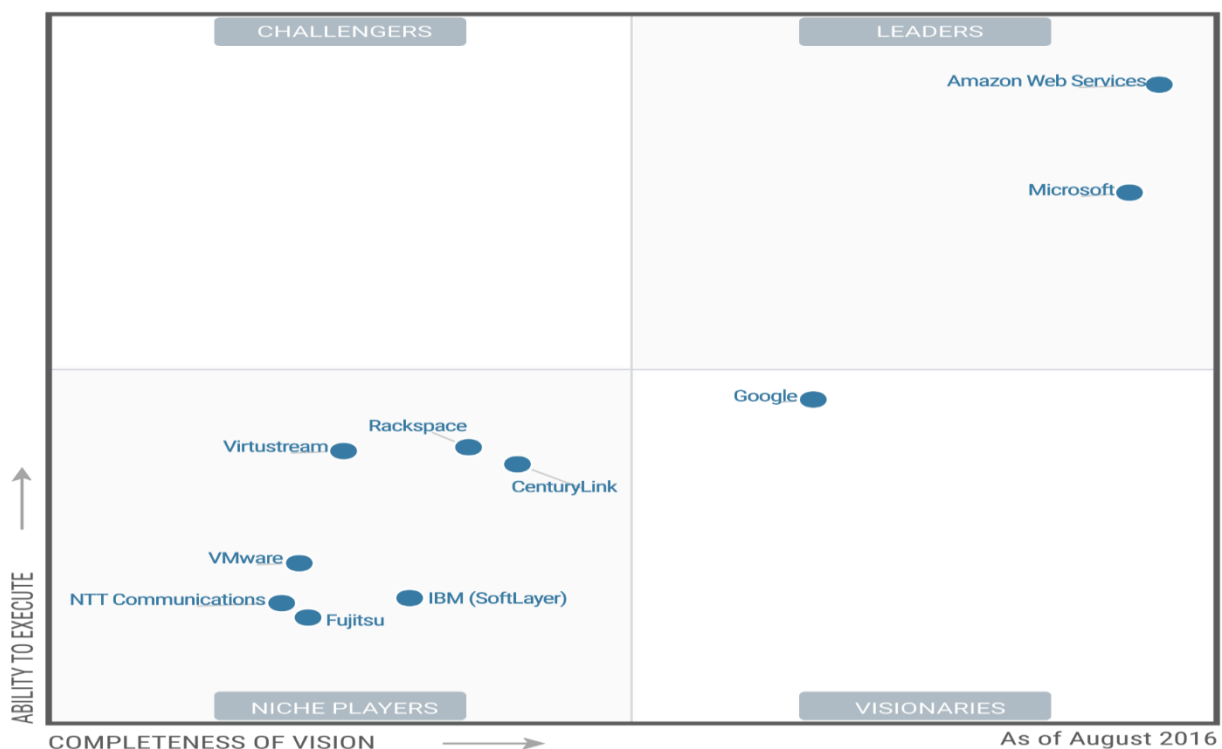


Figure IV-1 Gartner Group et les solutions IaaS en 2016 [40]

Selon Gartner, le marché infonuagique s'en va vers une hyperconvergence IaaS et PaaS. Le tableau suivant explique le mappage possible de l'architecture logique Iteraplan (plutôt mappable dans l'univers PaaS) vers le nuage d'Amazon. Cette explication est donnée pour garder la méthodologie viable pour encore 3-5 années.

Tableau IV.1 Mappage de couche logique dans le nuage d'Amazon

| # | Système logique | Amazon |
|---|-----------------|--------|
| | | |

| # | Système logique | Amazon |
|----|--|---|
| 1 | Système et Sous-système logique (SaaS) | Sans traduction. N'importe quelle solution SaaS est un système logique |
| 2 | Dépôt de Données | <p>Les dépôts de données sont groupés en relationnels et non-relationnels.</p> <p>Amazon offre dans le relationnel :</p> <ul style="list-style-type: none"> • Amazon SimpleDB, une base de données simple, capable de requêtes et indexations. • Amazon RDS, un service capable de fournir mySQL, Oracle et SQL Server. <p>Dans le non-relationnel :</p> <ul style="list-style-type: none"> • Amazon Dynamo DB • Amazon Redshift (un entrepôt de données) • Amazon Elastic Map Reduce (pour le traitement type HDFS de grandes quantités de données) |
| 3 | Application | Configuration et déploiement en EC2 |
| 4 | Équilibreur de charge (<i>Load Balancer</i>) | Elastic Load Balancing |
| 5 | Mandataire (<i>Proxy</i>) | Configuration et déploiement en EC2 |
| 6 | CDN ou réseau de diffusion de contenu (<i>Content Delivery Network</i>) | Amazon Cloud Front et le fournisseur typique d'Amazon |
| 7 | DNS | Amazon Route 53 |
| 8 | Gestion des identités, accès et autorisation (<i>Identity access and authorization management</i>) | AWS Directory Service |
| 9 | File de messages (<i>Message Queue</i>) | Amazon Simple Queue Service (Amazon SQS) |
| 10 | Pare-feu (<i>Firewall</i>) | Amazon WAF Groupes de sécurité Amazon EC2 |
| 11 | Stockage (<i>Storage</i>) | Amazon S3 pour stockage normal Amazon Glacier pour archivage |
| 12 | Messagerie courriel | Amazon Simple Email Service (Amazon SES) |
| 13 | Cache | Amazon ElastiCache |
| 14 | Gestionnaire de flux de travail | Amazon Simple Workflow Service (Amazon SWF) |

Annexe V Documents d'évaluation de spécialistes

Le temps de configuration, installation et test a été établi suite à des échanges de courriels avec de spécialistes en nuage. Le courriel suivant est un exemple des échanges.

Les valeurs dans ton tableau sont représentatives.

Olivier
120-8437

De : Antohi Tudor
Envoyé : 22 mars 2017 14:22
À : Gagnon Marchand Olivier
Cc : Langelier Félix; Cung Dinh Le Son Tony
Objet : RE: Question

Effectivement Merci beaucoup.

Olivier, pour configurer un elastic load balancer sur amazon à partir de ce que tu sais en A10, combien du temps pour être sur avant migrer?

Tudor

| | Composant technologique | Nombre de Jours | | | |
|---|-------------------------|-----------------|---------------|------|-------|
| | | Installation | Configuration | Test | Total |
| 1 | A10 Acos 4.0 | 0,5 ? | 0,5 ? | 1 ? | 2 |

De : Langelier Félix
Envoyé : mercredi 22 mars 2017 14:20
À : Antohi Tudor; Cung Dinh Le Son Tony
Cc : Gagnon Marchand Olivier
Objet : RE: Question

Salut Tudor,
Voici mon estimé!

| | Composant technologique | Nombre de Jours | | | |
|---|-------------------------|-----------------|---------------|------|-------|
| | | Installation | Configuration | Test | Total |
| | Tomcat 7 | 0,2 | 0,2 | 0,4 | 0,8 |
| 5 | Apache 2.4 | 0,1 | 0,1 | 0,2 | 0,4 |
| 6 | Magnolia code Web | 0,2 | 0,3 | 1 | 1,5 |
| 7 | Magnolia code BD | 0,0 | 0,0 | 0,2 | 0,2 |
| 8 | Magnolia Applicatif | 0,2 | 0,3 | 5 | 5,5 |

De : Antohi Tudor
Envoyé : 22 mars 2017 11:59
À : Cung Dinh Le Son Tony

Laurent Lamiaux compte 20 ans d'expérience en architecture de bases de données chez Loto-Québec. Son expérience dans les analyses de coût a été essentielle et dernièrement il a passé beaucoup du temps avec les analyses de consolidations de bases de données.

De : Lamiaux Laurent Date : lun.10/04/2011
 À : Antohi Tudor
 Cc :
 Objet : RE: Je vais t'expliquer point par point

Tableau 4.7 Formulaire d'évaluation de la méthode et de résultats

| | Question | 1 pauvre | 2 passe | 3 bien | 4 Très bien | 5 excellent |
|---|--|-------------|------------|-----------|----------------|----------------|
| 1 | La modélisation proposée est-elle correcte ? | | | | | X |
| 2 | Le mappage dans le nuage est-il bien conçu ? | | | | X | |
| 3 | Les contraintes sont-elles respectées ? | | | | | X |
| 4 | La méthodologie de migration suit-elle les règles du procédé ? | | | | X | |
| 5 | Les résultats confirment-ils l'hypothèse ? | | | | | X |

Claude Lamy compte 20 ans d'expérience en architecture d'entreprise chez Loto-Québec. Son expérience avec la méthodologie Iteraplan a été utilisée pendant le développement de cet essai.

Bonjour Tudor,

Voici le document avec le chapitre 4 révisé.

C'est pour moi de l'excellent travail ! Bravo !

Mes amis me disent que tu es un homme riche !!! Et qu'ils peuvent t'aider ... ;-)

J'espère répondre le mieux possible (en tout cas, au meilleur de mes connaissances) et que cela puisse t'aider.

À bientôt.

Merci.

Claude

De : Antohi Tudor

Envoyé : 5 avril 2017 15:15

À : Chau Thuy-Tien; Cung Dinh Le Son Tony; Ulatowski Rafal; Lamiaux Laurent; Lamy Claude; Desjardins Benoit

Objet : Last one

| | Question | 1 médiocre | 2 passable | 3 bon | 4 très bon | 5 excellent |
|---|---|---------------|---------------|----------|------------------|----------------|
| 1 | La modélisation proposée est-elle correcte? | | | | | X |
| 2 | Le mappage dans le nuage est-il bien conçu? | | | | X | |
| 3 | Les contraintes sont-elles respectées? | | | | | X |
| 4 | La méthodologie de migration suit-elle les règles du procédé? | | | | | X |
| 5 | Les résultats confirment-ils l'hypothèse? | | | | | X |

De : Cung Dinh Le Son Tony
Envoyé : lundi 10 avril 2017 12:56
À : Antohi Tudor
Objet : RE: Maitre Tony can u help me ?

Salut Tudor,

Travail très professionnel et les éléments abordés sont pertinents.

Voici mon formulaire rempli :

| | Question | 1 médiocre | 2 passable | 3 bon | 4 très bon | 5 excellent |
|---|---|---------------|---------------|----------|------------------|----------------|
| 1 | La modélisation proposée est-elle correcte? | | | | | X |
| 2 | Le mappage dans le nuage est-il bien conçu? | | | | X | |
| 3 | Les contraintes sont-elles respectées? | | | | | X |
| 4 | La méthodologie de migration suit-elle les règles du procédé? | | | | X | |
| 5 | Les résultats confirment-ils l'hypothèse? | | | | | X |

Désolé encore pour le retard.

Bonne journée.

Félix Langelier est architecte intergiciel et spécialiste infonuagique avec plus de 10 ans d'expérience. Ses observations ont été sur les prix de conversions vers blocs infonuagiques et il a mentionné le prix de la connexion virtualisée qui doit exister d'une façon permanente entre le centre de données et le vendeur de solution en nuage. Ce prix n'a pas été pris en considération étant donné que c'est une constante et dans les migrations 1 par 1, le temps total est toujours le même.

Benoit Desjardins M.Sc.A., architecte concepteur a une expérience de plus de 20 ans en technologies de l'information. Avigal Levy, Benoit Desjardins et l'auteur de cet essai ont

développé la méthodologie d'architecture d'une entreprise basée sur ITERAPLAN et appliquée chez Loto-Québec.

De : Desjardins Benoit
Envoyé : vendredi 7 avril 2017 20:30
À : Antohi Tudor
Objet : Évaluation document maîtrise
 Bonjour Tudor

C'est un document très intéressant. J'ai lu l'ensemble.

Voici ma notation et le document joint avec quelques commentaires, des petites corrections pour ma contribution lorsque je suis nommé à la fin.
 Il y a quelques éléments orthographiques remarqués, mais j'ai manqué de temps pour aller à fond à ce niveau.
 Pour ce qui est de l'algorithme j'ai compris l'idée générale, j'ai eu quelques doutes dans la notation des formules, mais cela me semble très bien.

| | Question | 1 médiocre | 2 passable | 3 bon | 4 très bon | 5 excellent |
|---|---|---------------|---------------|----------|---------------|----------------|
| 1 | La modélisation proposée est-elle correcte? | | | | X | |
| 2 | Le mappage dans le nuage est-il bien conçu? | | | | X | |
| 3 | Les contraintes sont-elles respectées? | | | | X | |
| 4 | La méthodologie de migration suit-elle les règles du procédé? | | | | X | |
| 5 | Les résultats confirment-ils l'hypothèse? | | | | X | |

Merci

Tony Cung est administrateur de système depuis plus de 5 ans et compte des études universitaires. Sa spécialité est Linux, Apache, l'infonuagique Amazon et aussi avec les configurations automatisées par Puppet. Il a aidé avec les estimations des efforts pour l'installation, configuration et test de machines et a testé les algorithmes.

Chau Thuy-Thien est architecte infrastructure avec une carrière de plus de 20 ans avec Desjardins et Loto-Québec. Elle a participé dans plusieurs projets d'analyse réseau et

informatique. Sa suggestion d'étudier les outils qui déjà existe sur le marché a été déjà pris en compte. Aucun outil ne donne pas directement le coût de migration prêt-à-l'emploi.

De : Chau Thuy-Tien
Envoyé : mercredi 5 avril 2017 17:11
À : Antohi Tudor
Objet : RE: Last one

Tel que demandé.

De : Antohi Tudor
Envoyé : 5 avril 2017 15:15
À : Chau Thuy-Tien; Cung Dinh Le Son Tony; Ulatowski Rafal; Lamiaux Laurent; Lamy Claude; Desjardins Benoit
Objet : Last one

| | Question | 1 médiocre | 2 passable | 3 bon | 4 très bon | 5 excellent |
|---|---|---------------|---------------|----------|---------------|----------------|
| 1 | La modélisation proposée est-elle correcte? | | | | X | |
| 2 | Le mappage dans le nuage est-il bien conçu? | | | | X | |
| 3 | Les contraintes sont-elles respectées? | | | | X | |
| 4 | La méthodologie de migration suit-elle les règles du procédé? | | | | X | |
| 5 | Les résultats confirment-ils l'hypothèse? | | | | X | |

Rafal Ulatovski compte plus de 10 ans d'expérience en programmation et architecture applicative. Son expérience avec Python, les statistiques et la mathématique a été utilisée pour la validation des algorithmes. Son idée a été surtout un défi. Il a demandé une

généralisation de la méthodologie utilisant le raisonnement par récurrence connu aussi comme raisonnement par induction.

Tableau 4.7 Formulaire d'évaluation de la méthode et de résultats

| | Question | 1 pauvre | 2 passe | 3 bien | 4 Très bien | 5 excellent |
|---|--|-------------|------------|-----------|-------------------|----------------|
| 1 | La modélisation proposée est-elle correcte ? | | | | x | |
| 2 | Le mappage dans le nuage est-il bien conçu ? | | | x | | |
| 3 | Les contraintes sont-elles respectées ? | | | x | | |
| 4 | La méthodologie de migration suit-elle les règles du procédé ? | | | | x | |
| 5 | Les résultats confirment-ils l'hypothèse ? | | | | x | |

Avigaël Lévy est une experte en programmation et bases de données avec une solide expérience d'architecture. À partir de ses observations une erreur dans l'algorithme final a été corrigée. Il a fallu séparer la période d'installation et configuration de la période de test quand le système migre entre l'architecture antérieure et l'architecture qui succède.

Bonjour Tudor,

Voici mes réponses.

Cordialement,

Avigaël Lévy
DevOps et Intégration
Loto-Québec
T. interne 120-4261
T. externe (514) 285-2929 poste 4261
avigael.levy@loto-quebec.com

De : Antohi Tudor

Envoyé : 29 mars 2017 15:15

À : Chau Thuy-Tien; Sicard Jean-Christophe; Huynh Can; Cung Dinh Le Son Tony; Lévy Avigael

Objet : Lire chapitre 4 et remplir un formulaire

Bonjour à vous,

En fait sont 60 pages après avoir changé le format mais avec beaucoup de figures.

Merci de votre patience.

Tudor

| | Question | 1 médiocre | 2 passable | 3 bon | 4 très bon | 5 excellent |
|---|---|---------------|---------------|----------|------------------|----------------|
| 1 | La modélisation proposée est-elle correcte? | | | | | X |
| 2 | Le mappage dans le nuage est-il bien conçu? | | | | X | |
| 3 | Les contraintes sont-elles respectées? | | | | | X |
| 4 | La méthodologie de migration suit-elle les règles du procédé? | | | | X | |
| 5 | Les résultats confirment-ils l'hypothèse? | | | | | X |