**TABLE DES MATIÈRES**

Page

INTRODUCTION .....1

CHAPITRE 1 REVUE DE LITTÉRATURE .....11
1.1 Introduction.....11
1.2 Introduction aux centres de données.....12
1.2.1 L'architecture physique des centres de données..... 12
1.2.2 Les niveaux de redondance des centres de données ..... 17
1.2.3 Les services des centres de données ..... 18
1.3 Les problèmes de localisation des centres de données .....19
1.3.1 Les critères de sélection des localisations..... 19
1.3.1.1 Exemples de quelques entreprises ..... 19
1.3.1.2 Les critères de sélection des localisations : Généralisation ..... 24
1.3.1.3 Le coût d'énergie ..... 27
1.3.1.4 Le climat ..... 27
1.3.1.5 Les catastrophes naturelles ..... 27
1.3.1.6 La pollution..... 28
1.3.1.7 L'interférence électromagnétique ..... 29
1.3.1.8 L'énergie renouvelable ..... 29
1.3.1.9 Le transport ..... 30
1.3.2 Les contraintes de transfert des données..... 30
1.3.2.1 Les contraintes réglementaires sur l'exportation des données... 30
1.3.2.2 Transfert des données en dehors du Canada ..... 31
1.3.3 Modélisation des problèmes de localisation ..... 33
1.3.3.1 Texte introductif..... 33
1.3.3.2 Localisation des centres de données ..... 33
1.4 Conclusion .....39

CHAPITRE 2 MODÉLISATION MATHÉMATIQUE .....41
2.1 Introduction.....41
2.2 Les hypothèses .....42
2.3 Approche d'optimisation proposée.....43
2.3.1 Description du modèle mathématique ..... 43
2.3.2 Les ensembles et les indices ..... 44
2.3.3 Les paramètres ..... 45
2.3.4 Les variables de décision ..... 48
2.3.5 Modèle d'optimisation..... 49
2.4 Conclusion .....53

CHAPITRE 3 EXPÉRIMENTATION, ANALYSE ET DISCUSSION .....55
3.1 Introduction.....55
3.2 Caractérisation des localisations.....55

3.2.1 Les critères de localisation ..... 55
3.3 Étude de cas ..... 56
3.3.1 Environnement de programmation ..... 57
3.3.2 Localisation des centres de données ..... 57
3.3.3 Temps d'exécution des méthodes de résolution ..... 59
3.4 Minimiser les coûts ..... 60
3.5 Les serveurs de secours "backup" ..... 64
3.6 Indicateur d'efficacité énergétique ..... 66
3.7 Temps de réponse des services Internet ..... 67
3.8 Évaluation de l'approche proposée ..... 69
3.9 Conclusion ..... 70

CONCLUSION ..... 71

RECOMMANDATIONS ..... 73

ANNEXE I DATACENTER LOCATION FOR INTERNET SERVICES IN THE PROVINCE OF QUEBEC ..... 75

ANNEXE II LES CATÉGORIES DE L'EFFICACITÉ ÉNERGÉTIQUE (PUE) ..... 89

ANNEXE III COMPARAISON DES PROFILS DE L'ÉNERGIE D'UN CENTRE DE DONNÉES ET D'UN BÂTIMENT TYPIQUE ..... 91

LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES ..... 93

## LISTE DES TABLEAUX

Page

Tableau 1.1 Classification des centres de données .....18

Tableau 1.2 Résumé des critères de sélection des localisations .....23

Tableau 3.1 Les tailles des centres de données.....59

Tableau 3.2 Temps d'exécution de la solution proposée.....59

Tableau 3.3 La quantité d'énergie pour le refroidissement.....60

Tableau 3.4 Délai de latence en secondes.....65

Tableau 3.5 Nombre de serveurs pour chaque centre de données localisé.....70

## LISTE DES FIGURES


Figure 1.1 Méthodologie de recherche .....8

Figure 1.1 Architecture physique de centres de données.....13

Figure 1.2 Conception en couche de l'infrastructure.....15

Figure 1.3 Architecture physique d'un centre de données .....16

Figure 1.4 Les proportions pour les facteurs .....26

Figure 1.5 Exemple de service d'infonuagique déployé.....34

Figure 1.6 Les proportions d'utilisation d'énergie .....36

Figure 2.1 Schéma descriptif du modèle mathématique.....44

Figure 3.1 Résultat de localisation des centres de données .....57

Figure 3.2 Les zones de marché desservies par les centres de données .....58

Figure 3.3 Les coûts fixes des centres de données.....61

Figure 3.4 Les coûts variables des centres de données.....61

Figure 3.5 Les coûts d'énergie des centres de données.....62

Figure 3.6 Répartition des coûts des centres de données localisés.....63

Figure 3.7 Le nombre de serveurs pour chaque zone de marché.....64

Figure 3.8 Nombre total des serveurs de secours " backup .....65

Figure 3.9 Valeurs des PUE pour chaque centre de données .....67

Figure 3.10 Temps de réponse par rapport au nombre de clients .....68

Figure 3.11 Variation de temps de réponse en fonction de service Internet.....68

## LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

| | |
|---|---|
| CDN | Content delivery network |
| PUE | Power Usage Effectiveness |
| GRDA | Grand River Dam Authority |
| UPS | Uninterruptible Power Supply |
| PATRIOT | Providing Appropriate Tools Required to Intercept and Obstruct Terrorism |
| SOS | Special Ordered Sets |
| QoS | Quality of service |
| CNIL | Commission nationale de l'informatique et des libertés |

Rapport-Gratuit.com

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

# INTRODUCTION

## Contexte général

Aujourd'hui, les compagnies Internet les plus populaires telles que Google©, Yahoo© et Microsoft© offrent une gamme de services à des millions d'utilisateurs chaque jour. Ces services sont hébergés dans des centres de données.

Également, le besoin d'une grande mémoire de stockage, une infrastructure et une plus grande puissance de traitement des données est une nécessité. En effet, les centres de données présentent un point axial sur lequel toutes les données et les applications des utilisateurs sont déployées. De même, ils offrent différents services.

Les centres de données informatiques présentent un investissement énorme et la plupart des entreprises ont besoin de mettre en place leurs propres centres de données pour exécuter des opérations commerciales, stocker des données ou offrir des services à des clients. En effet, la mise en place des nouveaux centres de données nécessite un effort de longue haleine, des millions de dollars et des mois de planification. Dans le même contexte, Greenberg *et al.* (2009) indiquent que les centres de données sont très coûteux et le coût peut éteindre 12 millions par mois pour un centre de données de 100,000 serveurs. Également, le coût total des centres de données varie selon les secteurs d'industrie (communications, commerce électronique, défense, technologies & logiciels, éducation*, etc.*). Selon LLC (2013), le coût total d'un centre de données varie entre un coût minimum de 74 223 $ et un coût maximum de 1,734,433 $. En outre, les centres de données sont "énergivores". À cet égard,Pakbaznia et Pedram (2009) indiquent que l'énergie consommée par les centres de données est énorme. De plus, ils affirment, selon l'agence de protection de l'environnement des États-Unis, que les centres de données consomment environ 61 milliards de kilowattheures (kWh) en 2006, soit environ 1,5 pour cent de la consommation totale d'électricité des États-Unis, pour un coût total d'environ 4.5 milliards de dollars. De même, Pakbaznia et Pedram (2009) ont déclaré que la consommation d'énergie des serveurs et des centres de données a doublé au cours des cinq dernières années et devrait quadrupler au cours des cinq prochaines années pour plus de

100 milliards de kWh à un coût annuel d'environ 7,4 milliards de $. Également, Pakbaznia et Pedram (2009) prouvent que le coût de refroidissement est l'un des principaux contributeurs de la facture totale d'électricité des centres de données. Pour ces raisons, la localisation des centres de données est très importante et présente un axe de recherche active.

Ainsi, les entreprises ne peuvent se permettre de prendre une mauvaise décision lors de la conception et la mise en place d'un nouveau centre de données, car la planification est pour environ 20 ans. Pour cette raison, la définition des critères de sélection des centres de données est la chose la plus basique avant de mettre en place et choisir un site pour un nouveau centre de données.

Également, les compagnies veulent bâtir des centres de données dans des localisations qui ne sont pas sujettes aux catastrophes naturelles, aux terrorismes et également offrent des ressources moins chères comme l'électricité, réseau et transport.

C'est pourquoi que nous choisissons la province de Québec pour appliquer notre projet de recherche. En effet, la province de Québec bénéficie de la tarification énergétique la plus basse en Amérique du Nord, un risque minimal des catastrophes naturelles, un climat frais, la confidentialité des données (la loi américaine *Patriot Act* ne s'applique pas), l'accès aux marchés (accord de libre-échange avec les États-Unis et plusieurs pays d'Amérique latine) et l'utilisation des technologies avancées de refroidissement qui est aussi gratuite puisqu'il utilise l'air frais hivernal de la province de Québec pour rafraîchir les serveurs. Les températures moyennes sont significativement plus basses qu'aux États-Unis, avec des températures moyennes sous le point de congélation cinq mois par an et par conséquent un coût réduit pour le refroidissement.

En plus, une variété de services tels que l'hébergement Web, les services de base de données, les services de télécommunications et services bancaires et les services Internet. Ces services dépendent de l'infrastructure des centres de données pour offrir des services aux clients. Avec l'avènement du commerce électronique, des téléphones intelligents, des réseaux sociaux

et des communications mobiles, les exigences des entreprises continuent d'augmenter. Pour notre projet, nous admettons que les centres de données à mettre en place offrent des services Internet aux utilisateurs. Ces exigences sont stockées dans ces centres de données répartis géographiquement.

Les compagnies Internet veulent bâtir des centres de données dans des localisations qui ne sont pas sujettes aux catastrophes naturelles, aux terrorismes et également offrent des ressources moins chères comme l'électricité et le réseau de transport. En même temps, elles ont besoin d'une installation robuste qui peut fonctionner 365/7/24 sans perturbation et offre une puissance de calcul évolutive.

Parce que les utilisateurs des services Internet demandent une haute disponibilité de service et un temps de réponse minime, chaque service est dupliqué dans des centres de données distribués géographiquement. Chaque centre de données est censé de servir les requêtes des utilisateurs qui sont les plus proches (en terme de latence réseau). Dans le cas où les centres de données deviennent inaccessibles, les requêtes des utilisateurs seront transmises à des miroirs des centres de données.

Dans ce cadre, nous présentons un modèle mathématique tout en prenant en considération les critères mentionnés précédemment pour choisir la localisation potentielle des centres de données. Ce choix vise à optimiser les coûts d'implémentation des centres de données ainsi que la disponibilité du service Internet. Également, les compagnies Web doivent prendre en considération les coûts énormes d'approvisionnement et d'exploitation tel un réseau de centres de données. De même, l'aspect intéressant de temps de réponse, la disponibilité et les coûts dépendent fortement à la localisation spécifique des centres de données. Par exemple, le coût des terrains, électricité aussi bien que le temps de latence dépend directement de la localisation. Encore plus intéressante, la relation coût et localisation n'est pas toujours valide. Par exemple, le coût de terrain dans une zone désertique n'est pas cher par contre le refroidissement est coûteux.

**Définitions des centres de données**

Stryer (2010) a défini les centres de données en tant qu'un entrepôt centralisé pour le stockage, la gestion et la diffusion des données et des informations. Également, l'association industrielle de télécommunication (TIA) a produit une nouvelle spécification nommée « TIA-942 : un nouveau standard pour définir les infrastructures de télécommunication des centres de données. » En effet, ce standard englobe toutes les parties de conception des centres de données y compris le câblage, installation et conception de réseaux.

D'après Daim, Bhatla et Mansour (2012) un centre de données est une installation des systèmes informatiques et des composants associés comme les systèmes de stockage et de télécommunications. Les centres de données comprennent les alimentations d'énergie de secours, des données redondantes, un environnement de contrôle (climatisation d'air) et des dispositifs de sécurité. Également, un centre de traitement de données est un site physique sur lequel se trouvent regroupés des équipements constituants du système d'information de l'entreprise (ordinateurs centraux, serveurs, baies de stockage, équipements réseaux et de télécommunications, etc.). Il comprend en général un contrôle sur l'environnement (climatisation, système de prévention contre l'incendie), une alimentation d'urgence et redondante, ainsi qu'une sécurité physique élevée. Il sert surtout à traiter les informations nécessaires aux activités d'une entreprise.

**Problématique**

Comme nous avons prouvé que les centres de données présentent un point axial et critique pour les processus d'affaires des entreprises et des clients des services Internet. Il est donc important de définir un algorithme robuste pour assurer la sélection des localisations potentielles des centres de données. Dans ce cas, il existe plusieurs problèmes de localisation, parmi lesquels nous pouvons citer :

**Problème I** – localisation des centres de données tout en minimisant leur coûts d'installation.

**Problème II** – Localisation des centres de données tout en minimisant l'énergie consommée par les équipements de technologie d'informations (TI).
**Problème III**– Localisation des centres de données pour des services Internet tout en assurant une haute disponibilité pour les clients.

Ces problèmes relèvent un défi pour les chercheurs afin de développer un algorithme pour la sélection des localisations pour les nouveaux centres de données. En effet, la combinaison de ces problèmes à la fois pour définir un *Framework* de localisation des centres de données présente à son tour un défi. Dans ce contexte, plusieurs algorithmes sont définis tel que proposé par Goiri *et al.* (2011). Ces algorithmes visent à sélectionner les localisations des centres de données tout en minimisant les coûts de mise en place. Également, plusieurs autres approches de sélection ont été proposées dans des recherches antérieures. Cependant, les approches de Gandhi *et al.* (2011) et Lei *et al.* (2010) visent à minimiser respectivement l'énergie consommée par les équipements TI et le coût d'énergie électrique alimentant les centres des données. De plus, aucune de ces approches ne combine à la fois les problèmes précédemment cités. Nous nous intéressons dans ce mémoire par la combinaison à la fois des deux premiers problèmes de localisation cités antérieurement. Notre approche pourra pallier aux lacunes des différentes approches définies dans les recherches antérieures. Avec plus de détail, notre approche visera à définir un algorithme qui assure la sélection des localisations des centres de données offrant des services Internet tout en minimisant, d'une part, les coûts de mise en place des centres de données d'autre part, la consommation énergétique des équipements informatiques et par conséquent, une bonne efficacité énergétique des centres de données.

**Objectifs de recherche**

Le but principal de cette recherche est de choisir la bonne localisation et la bonne dimension pour mettre en place des centres de données des régions de la province de Québec afin de desservir des zones de marché. Les centres de données, qui vont être mis en place, sont des centres de données fournissant des services Internet.

Pour atteindre ce but, les objectifs / sous-objectifs sont les suivants :

- Définir les contraintes et les critères de sélection des sites pour mettre en place des centres de données pour des services Internet.
- Définir un algorithme de localisation des centres de données en prenant en considération les contraintes déjà définis.
- Évaluation et validation de l'approche proposée à travers une étude de cas.

Notre recherche est fondée sur les questions suivantes :
QR1 : Quels sont les critères de sélection des localisations pour placer un centre de données pour des services Internet ?
QR2 : Quelles sont les localisations potentielles pour installer un centre de données pour des services Internet dans la province de Québec ?
QR3 : Quel est le nombre et la taille des centre de données à mettre en place dans chaque localisation de la province de Québec ?

**Les hypothèses**

Tout au long de notre projet, nous admettons deux hypothèses.

1) La première hypothèse indique que notre projet de recherche est appliqué à la province de Québec. De plus, le choix de la province de Québec est basé sur le fait que nous avons plus de données sur les centres de données déjà existants. Également, les visites guidées des centres de données et la communication avec des experts des centres de données à Montréal nous ont permis de collecter beaucoup de données pour le cas de la province de Québec.
2) La deuxième hypothèse est portée sur le type de service de centre de données à mettre en place. En effet, nous admettons de mettre en place des centres de données offrant des services Internet aux utilisateurs.

La validation de ces hypothèses est effectuée quant à la conception et la modélisation du modèle mathématique pour la sélection des localisations potentielles des centres de données offrant des services Internet.

**Méthodologie de recherche**

Cette recherche suit une approche itérative et incrémentale afin de répondre aux objectifs de recherche déjà mentionnés. En effet, cette approche est formée de deux principales itérations :

- Définir les critères de sélection des localisations :

Afin de définir les critères de sélection des localisations, nous avons fait une analyse exhaustive des travaux de recherche actuels sur la sélection des critères de localisation; cette itération est incrémentale tout en faisant recours chaque fois à des revues de littératures. En effet, la revue de la littérature a comme objectif principal de contextualiser les lecteurs par rapport aux éléments présents dans la problématique déjà déterminée. De plus, cette itération permet de répondre à la question de recherche QR1.

- Modélisation du problème de localisation :

Cette itération permet de répondre aux questions de recherche QR2 et QR3. En effet, nous définissons, tout d'abord, un modèle mathématique pour résoudre le problème de localisation des centres de données. Par la suite, nous faisons recours à une partie d'expérimentation illustrée par une étude de cas pour valider ce travail.

Rapport-Gratuit.com

La figure 1.1 illustre la méthodologie de recherche suivie.

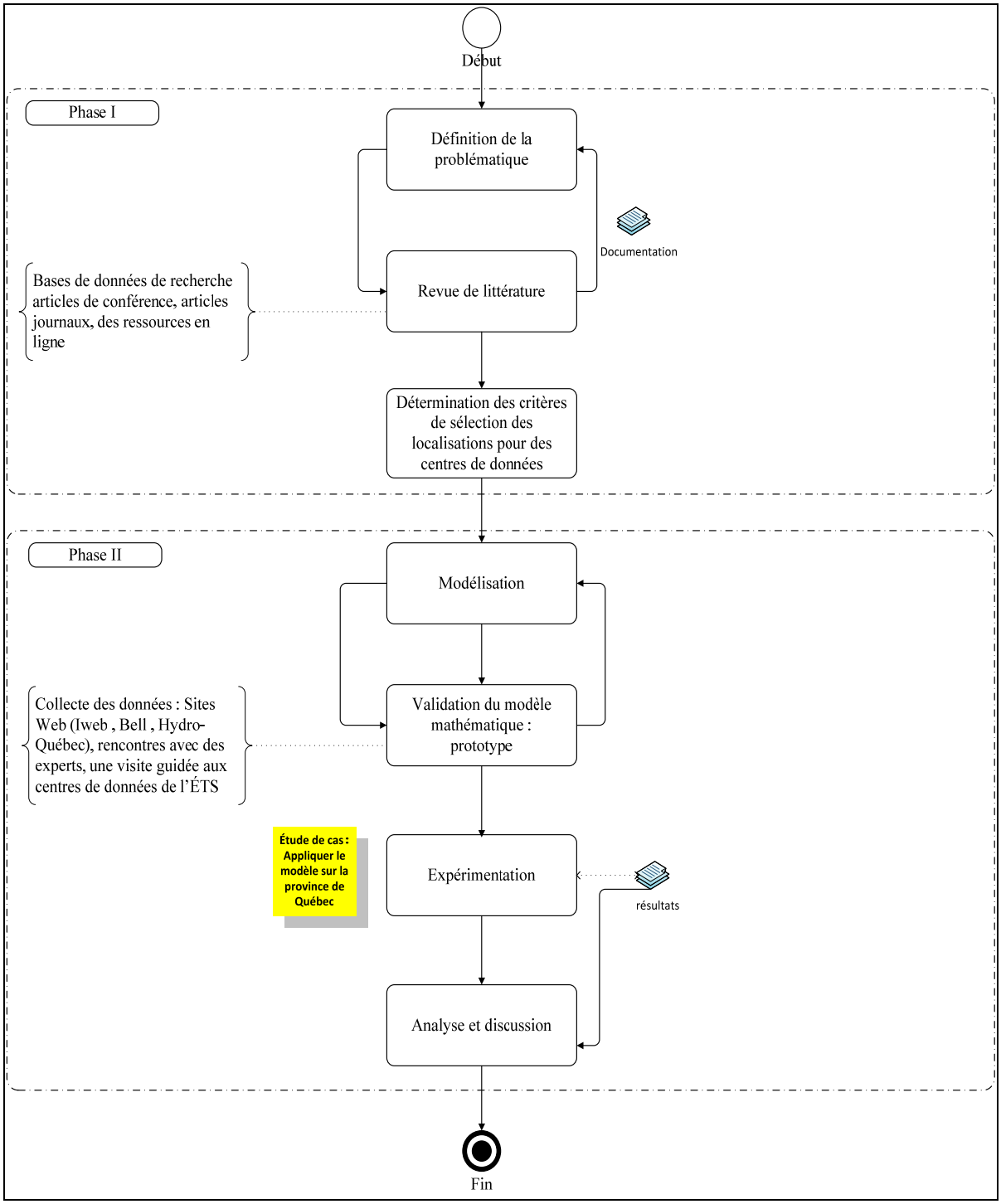


Figure 1.1 Méthodologie de recherche

**Organisation du rapport**

Le reste du rapport est organisé comme suit. Le chapitre 1 est une revue de littérature dont nous présenterons dans une première partie un état de l'art sur les centres de données. Dans une deuxième partie, nous présenterons des recherches connexes sur les problèmes de localisation des centres de données, le projet a été mis dans son contexte. Dans le chapitre 2, nous décrirons en détail le modèle mathématique et la solution proposée. Le chapitre 3 présente la partie expérimentation où nous présenterons et interpréterons les résultats obtenus. La conclusion et nos recommandations sont présentées à la fin du mémoire.

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

# CHAPITRE 1

# REVUE DE LITTÉRATURE

## 1.1 Introduction

Le besoin d'une grande mémoire de stockage, une infrastructure et une plus grande puissance de traitement des données sur la machine d'un utilisateur n'est plus une nécessité. En effet, les centres de données présentent un point axial sur lequel toutes les données et les applications des clients sont déployées. La centralisation de ces données dans un même endroit rend la gestion et la sécurité des données plus facile. Dans ce chapitre, nous présentons un état de l'art sur les centres de données pour arriver ultérieurement à présenter l'importance de sélection des localisations pour mettre en place ces centres de données.

Plusieurs théories sont proposées afin d'expliquer le concept de localisation des centres de données. Bien que la revue de littérature couvre une grande variété de ces théories, cette revue de littérature va se concentrer sur quatre thèmes majeurs qui émergent à plusieurs reprises tout au long de la revue de littérature. Ces thèmes sont : la minimisation des coûts des centres de données, la minimisation de l'énergie consommée par les centres de données, la minimisation de la bande passante et l'efficacité énergétique des centres de données. Bien que la revue de littérature présente ces thèmes dans des contextes variés, tout au long de ce chapitre, nous allons nous concentrer principalement sur les concepts des centres de données qui offrent les services Internet.

Également, ce chapitre est formé de deux sections principales. La première section est une introduction aux centres de données dont nous allons introduire des notions basiques sur les et exactement l'architecture physique et les équipements de technologie d'information (TI) qui nous intéressent principalement. La deuxième section se concentre sur les critères de sélection des localisations pour les centres de données ainsi qu'une étude de modélisation et les solutions proposées pour la localisation des centres de données que nous les décomposons en quatre thèmes, comme mentionné précédemment.

## 1.2 Introduction aux centres de données

Les centres de données ont commencé à fleurir dans les années 1990 avec la révolution Internet et l'inflation de la demande en connexion haut débit ininterrompue, aussi bien pour le grand public que les entreprises dont les moyens informatiques en interne ne suffisaient plus.

C'est tout d'abord un centre qui centralise des données informatiques. Ce terme s'utilise dans le cas de l'hébergement d'un site Web, car c'est généralement l'endroit où les serveurs sont placés. Ces centres de données regroupent une multitude de machines. Les données mises en place sur les serveurs doivent être accessible à tout moment et protégées des dégâts extérieurs, par conséquent les centres de données ont pour avantage d'être conçus de façon à être protégés des principaux risques de coupure ou d'intrusion. Il y a ainsi des protections contre les coupures électriques, les risques d'incendie, l'accès de personnes malveillantes aux serveurs. Les serveurs sont généralement rangés dans des baies (ressemble à une grosse armoire) et les connexions sont prévues pour faciliter la rapidité d'accès aux machines. Les premiers centres de données avaient pour objectif d'offrir un centre d'hébergement et de traitement des données hypersensibles à de grandes entreprises internationales, d'un niveau de sécurité très élevé. Au fil du temps et de la démocratisation des composantes ainsi que la virtualisation des serveurs, les centres de données se sont positionnés comme étant une solution économique pour les PME et petits indépendants.

### 1.2.1 L'architecture physique des centres de données

Un centre de données est généralement organisé dans des rangées « baies » où chaque rangée contient des actifs modulaires tels que les serveurs, les commutateurs, des unités de stockage « briques », ou des appareils spécialisés comme représentés sur la Figure 2.2 une baie standard de hauteur 78 po, de largeur de 23 à 25 po et de 26 à 30 po de profondeur. Typiquement, chaque baie prend un certain nombre de « baies modulaires » actives insérées horizontalement dans les baies. L'épaisseur de l'actif est mesurée en utilisant une unité appelée " U " qui est de 45 mm (ou environ 1,8 po).

La majorité des serveurs possèdent des processeurs simple ou double et peuvent s'adapter à la taille de 1U, mais les plus grands (par exemple, multiprocesseurs) peuvent exiger 2U ou une taille de plus. Un rack standard peut prendre un total de 42 actifs à 1U lorsqu'il est entièrement rempli.

Figure 1.1 Architecture physique de centres de données
« centre de données de McGill » localisé à l'ÉTS

Il est aujourd'hui plus intéressant d'utiliser les services ainsi que la qualité d'un centre de données plutôt que d'investir dans du matériel propre. Le plus souvent, les armoires sont placées sur des planchers en hauteur permettant d'y passer les câbles réseau, l'alimentation électrique, mais également une circulation de l'air par le sol. La température d'un centre de données doit être constante à 20 °C. Elle est assurée par une climatisation classique ainsi que des systèmes de refroidissement par eau dans les villes plus chaudes. Lorsque le local est situé dans une ville où la température moyenne est proche ou en dessous des 20 °C souhaités, un système de récupération d'air froid extérieur pourra être utilisé. Cela permet des économies d'énergie, devenu un nouvel objectif d'évolution (*Green Computing*). La disposition des serveurs en allées a également été pensée de manière à répartir la chaleur qu'ils dégagent. L'alimentation électrique peut être assurée par plusieurs circuits différents et complétée par des systèmes de batteries de secours (UPS) ou de générateurs. L'objectif est d'alimenter les systèmes selon le principe de zéro coupure (*No-Break*). Si malgré tout un début d'incendie devait se déclarer, il serait immédiatement neutralisé à l'aide de gaz inerte

ou de micros particules d'eau azotée. La technique du gaz inerte est également utilisée dans les musées, car elle n'endommage pas les alentours et peut être utilisée tout en maintenant le circuit électrique actif.

En outre, Foh et al. (2012) ont montré que la réduction de la consommation d'énergie peut être obtenue en optimisant les usages des serveurs, du stockage et des réseaux. Pour cette finalité, plusieurs techniques peuvent être envisagées pour optimiser les usages de serveurs. Des techniques telles que l'utilisation des serveurs et la virtualisation sont quelques exemples qui peuvent conduire à des gaspillages faibles dans l'utilisation des ressources du serveur et de partage élevé de l'infrastructure commune. De plus, Foh et al. (2012) ont prouvé que les périphériques de stockage sont un autre facteur principal de la consommation d'énergie. Les dispositifs de stockage génèrent aussi une quantité importante de chaleur qui doit être examinée pour l'énergie de refroidissement. La virtualisation du stockage et de placement optimale des données dans des dispositifs de stockage hétérogènes seront des solutions possibles pour améliorer les performances d'accès au stockage et à accroître l'utilisation du stockage.

Les centres de données sont formés par des serveurs, des routeurs, des commutateurs, *etc*. En outre, Zhang, Cheng et Boutaba (2010) ont prouvé que l'approche en couches est la base pour la conception de l'architecture réseau qui est testée sur plusieurs centres de données déployés. Cette approche est illustrée à la figure 1.2

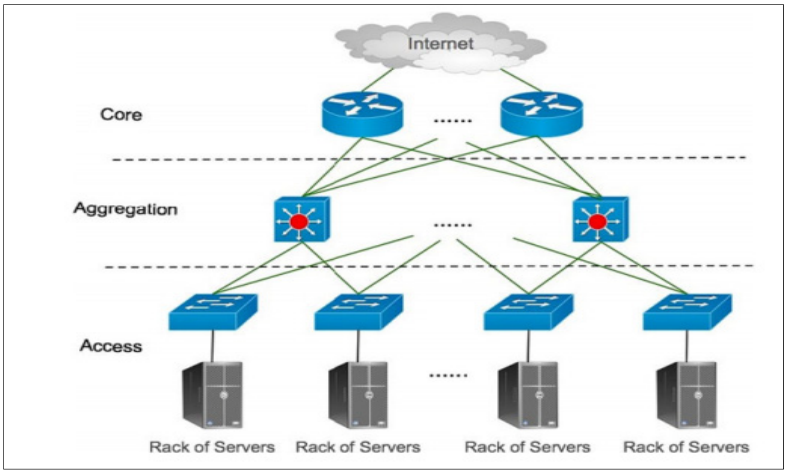


Figure 1.2 Conception en couche de l'infrastructure
de réseau de centre de données
Tirée de Zhang *et al.* (2010)

La couche d'accès est au niveau de laquelle les serveurs sont connectés au réseau. Alors que la couche d'agrégation fournit des fonctionnalités importantes telles que le service de domaine, service de location, service d'équilibrage de charge *(load balancing*). Finalement, la couche de noyau qui fournit une connectivité aux plusieurs commutateurs d'agrégation. Les routeurs de la couche noyau gèrent le trafic entrant et sortant de centre de données. Un projet de virtualisation concerne aussi les centres de données, puisque différents aspects de dimensionnement tels que la place au sol, la consommation électrique (et donc la climatisation) vont être différents par rapport à l'hébergement de serveurs physiques dédiés. Également, on peut ajouter que des demandes spécifiques peuvent accompagner la mise en œuvre de serveurs de virtualisation, notamment en ce qui concerne l'architecture réseau (avec des réseaux dédiés pour les clusters ou les communications avec les SAN / NAS) où le positionnement dans la salle machine des serveurs et baies de stockage (armoires de stockage).

❖ **Exemple des centres de données de Bell**

Bell (2005) a défini dans son article un plan tactique pour implémenter un centre de données. En effet, ce plan contient cinq critères de mise en place :

- La source d'énergie : Distribution de l'énergie y compris les UPS (alimentation de secours), les unités de distribution de l'énergie, les unités de distribution intermédiaires.
- Les systèmes de ventilation et de refroidissement.
- Les systèmes de protection de feux.
- Les systèmes de sécurité.
- Les systèmes des planchers surélevés :
  - ✓ 12 pouces (30,5 cm) pour 1000 pieds carrés.
  - ✓ 12 à 18 pouces (31- 46 cm) pour 1000 à 5000 pieds.
  - ✓ 18 à 24 pouces pour 5000 à 10000 pieds carrés.
  - ✓ >= 24 pouces pour 10000 pieds carrés ou plus.

La figure ci-dessous présente l'architecture d'un centre de données :

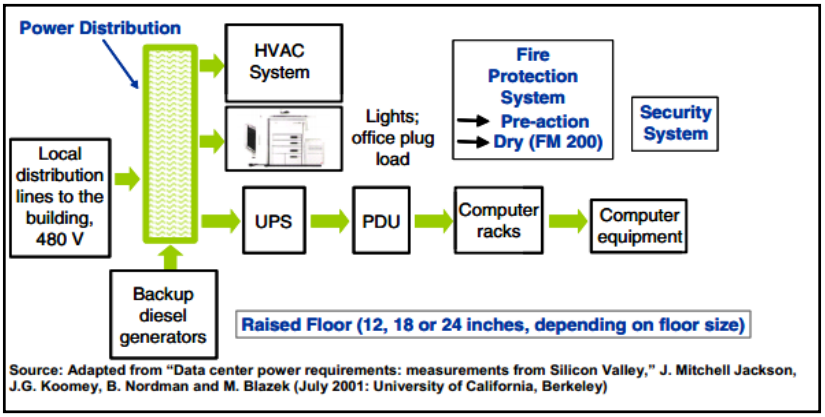


Figure 1.3 Architecture physique d'un centre de données
Tirée de Bell (2005)

Les centres de données sont essentiellement de très grands appareils qui consomment de l'énergie électrique et de la chaleur. Le centre de données est formé par un système de refroidissement (comme bien illustré dans la figure 1.3). Ce système de refroidissement supprime l'énergie supplémentaire consommant de la chaleur dans le processus, dont la chaleur doit être retirée aussi.

Il n'est pas surprenant, alors, que la majeure partie des coûts de construction d'un centre de données soient proportionnels à la quantité de puissance délivrée et la quantité de chaleur à

enlever; en d'autres termes, la grande partie de l'argent est dépensé, soit sur, le conditionnement de puissance et de distribution ou sur les systèmes de refroidissement. Les coûts de construction typiques pour un grand centre de données sont dans la gamme de 10 à 20 $ / Watt, mais varie considérablement selon la taille, l'emplacement et la conception.

### 1.2.2 Les niveaux de redondance des centres de données

La conception globale d'un centre de données est souvent classée en quatre niveaux (*Tier I-II, III et IV*), en fonction de résilience et disponibilité de certains systèmes. La valeur N décrit le type de secours disponible dans l'éventualité de la défaillance du composant. Par exemple, la redondance de type N+1 signifie que les composants (N) ont au moins un composant de secours indépendant (+1).

**Le niveau1 (*Tier I*) :** Les centres de données ont un seul chemin pour la distribution d'énergie et le de refroidissement, sans composants redondants.
**Le niveau 2 (*Tier II*) :** il ajoute des composants redondants à cette conception (N + 1), ce qui améliore la disponibilité.
**Le niveau 3 (*Tier III*) :** Les centres de données possèdent multiples chemins de distribution de puissance et de refroidissement, mais un seul chemin actif. Également, ils ont aussi des composants redondants et qui sont maintenable en même temps. Ce qu'ils assurent la redondance, même pendant la maintenance, généralement avec une configuration N + 2.
**Le niveau 4 (*Tier IV*) :** Les centres de données possèdent deux chemins actifs pour la distribution de puissance et de refroidissement. Il existe des composants redondants dans chaque chemin. Les équipements redondants sont sensés tolérer toute défaillance des équipements sans influencer sur la charge.

Les niveaux de redondance précédemment décrits sont détaillés au tableau 1.1.

Rapport-Gratuit.com

Tableau 1.1 Classification des centres de données

| Niveau | Génératrice | Système d'alimentation sans coupure | Services publics | Refroidissement | Disponibilité estimée |
|---|---|---|---|---|---|
| **Niveau 1** | Génératrice de type N | Système d'alimentation sans coupure de type N | Source d'alimentation unique | Système de refroidissement dédié de type N | 99,671 % Temps d'arrêt d'environ 29 h par année |
| **Niveau 2** | Génératrice de type N | Système d'alimentation sans coupure de type N+1 avec redondance des modules | Source d'alimentation unique | Système de refroidissement de type N+1 avec éléments redondants | 99,750 % Temps d'arrêt d'environ 22 h par année |
| **Niveau 3** Maintenabilité concurrente | Système de génératrices de type N+1 | Système d'alimentation sans coupure de type N+1 avec redondance des modules | Deux sources d'alimentation, 1 active et 1 passive | Système de refroidissement de type N+1 avec éléments redondants | 99,980 % Temps d'arrêt d'environ 105 min par année |
| **Niveau 4** Tolérance aux pannes | Système de génératrices de type 2N | Systèmes d'alimentation sans coupure de type 2N | Deux sources d'alimentations actives | Systèmes de refroidissement | de type 2N 99,990 % Temps d'arrêt d'environ 52 min par année |

Source : Uptime Institute

### 1.2.3 Les services des centres de données

Barroso et Hölzle (2009) indiquent que les centres de données offrent différents types de services. En effet, ces services sont répartis en fonction des besoins de l'entreprise et en fonction des demandes des utilisateurs. Souvent, ces services sont répartis dans plusieurs ordinateurs, chacun exécutant une partie de la tâche. Les composantes les plus habituelles sont des systèmes de gestion de base de données, des serveurs de fichiers et des serveurs d'applications. Également, des services de colocalisation, service d'hébergement et les services Internet.

En ce qui concerne les services Internet, nous distinguons les services de base d'Internet :

- Services internes au réseau
  - ✓ essentiellement les protocoles de routage,
- Services inter-machines
  - ✓ synchronisation du temps (NTP),
- Services transparents à l'utilisation des services utilisateurs
  - ✓ résolution de noms (DNS),
- Services traditionnels offerts aux utilisateurs
  - ✓ courrier électronique (email),
  - ✓ session distante (telnet),
  - ✓ transfert de fichiers (ftp),
  - ✓ accès aux news (nntp),
  - ✓ hypertexte (http),
- Services de recherche d'information
  - ✓ moteurs de recherche à travers le Web,
  - ✓ annuaires,
- Services interactifs
  - ✓ communication textuelle (talk, irc),
- Services temps réel
  - ✓ audio, vidéo, tableau blanc, slides,
  - ✓ téléphone Internet (VOIP).

## 1.3 Les problèmes de localisation des centres de données

### 1.3.1 Les critères de sélection des localisations

#### 1.3.1.1 Exemples de quelques entreprises

Plusieurs grandes entreprises dans le monde telles que Google©, IBM©, Microsoft©, DELL© et Amazon© définissent des critères en commun pour mettre en place un centre de données. Avant de mettre en place un centre de données, il faut penser tout d'abord à satisfaire les

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

besoins des clients tout en assurant la haute disponibilité des services, des données sécurisées, un temps de réponse minimal et un transfert rapide des données. Ces finalités seront obtenues tout en définissant dès le départ une bonne localisation pour un centre de données. De même, la définition des critères pour le choix de meilleures localisations pour les centres des données diffère selon les priorités des entreprises.

- **Google**

Google est l'une des grandes entreprises dans le monde et possède un grand nombre d'utilisateurs et ce nombre augmente jour après jour avec l'apparition de ses nouvelles applications. Afin de satisfaire ces besoins, Google essaye d'installer des centres de données partout dans le monde. En effet, Google installe un centre de traitement de données en Belgique. Ce site belge a été choisi en raison de critères environnementaux. Également, Dumout (2007) prouve que Google consacre 250 millions d'euros à la construction d'un centre de traitement de données informatique en Belgique, à St Ghislain. En effet, le choix de la Belgique s'est fait pour des raisons essentiellement d'aménagement du territoire. Le terrain est proche d'un point d'eau, car Google utilise un système de refroidissement par évaporation pour les installations informatiques. Ce système est plus écologique que les systèmes classiques. Dans le cas de St Ghislain, le lieu choisi se situe donc tout près d'un canal; les autres critères reposaient sur la proximité d'une source d'énergie et sur le bon raccordement de la zone en fibre optique.

En plus, d'après (Google, 2012a), la construction d'un premier centre de données en Asie du Sud-est représente une étape importante pour l'entreprise. En effet, le taux quotidien de nouveaux utilisateurs d'Internet y est le plus élevé au monde. En outre, Google (2012a) prouve que le choix de Singapour pour mettre en place le centre de données est en raison des critères suivants : une infrastructure fiable, une main-d'œuvre expérimentée, un engagement envers des réglementations transparentes et favorables aux entreprises. De même, Singapour possède une économie Internet dynamique et la ville se situe au centre de l'un des marchés Internet connaissant la croissance la plus rapide au monde. Également, Google choisit les localisations où se trouve l'énergie renouvelable pour la mieux exploiter et par conséquent

diminuer les coûts d'énergie consommée. Dans ce cas, Google a recours à l'énergie renouvelable qui est le critère le plus important ces dernières années. En effet, d'après Google (2012b), l'entreprise a signé un accord avec 'Grand River Dam Authority' (GRDA), pour fournir au centre de données de Google situé dans le comté de Mayes (Oklahoma), environ 48 MW d'énergie éolienne.

- **Microsoft**

L'utilisation des technologies Microsoft de virtualisation et d'administration du centre de données et une transition bien planifiée vers l'infonuagique aide les départements informatiques à fournir des services fiables, à améliorer l'efficacité opérationnelle et à permettre un meilleur alignement avec l'objectif métier de l'entreprise, pour transformer les centres de données en actifs stratégiques. En effet, Fei et Li Xuan (2011) présentent que Microsoft prévoie de mettre en place un centre de données en Sibérie pour deux raisons essentielles : le climat dont la température est à 50 ° au-dessous de zéro chaque hiver, par conséquent il peuvent utiliser le refroidissement naturel pour les centres de données. Également, le coût d'électricité est stable et réduit.

- **IWEB**

  a. **Caractéristiques**

IWEB est une entreprise d'hébergement des sites et des applications Web. Cette entreprise a lancé récemment un des plus grands centres de données à Montréal. En effet, ce centre de données est localisé à Montréal et possède une capacité de 33,000 serveurs. Également, ce centre est placé sur un terrain de taille 28,000 pieds carrés et il est géré par plus 220 employés spécialistes. En plus, trois niveaux d'alimentation électrique préviennent les pannes. Des systèmes d'alimentation sans coupures UPS (*Uninterruptible Power Supply*) non interruptibles garantissent l'alimentation en cas de panne et trois génératrices au diésel dans chacun des centres de données fournissent de l'électricité pour une plus longue période.

### b. Les critères de sélection

Iweb (2013) déclare que ses centres de données « verts » sont situés à Montréal, au Canada. Ces centres sont placés à 72 kilomètres des États-Unis tout en offrant des connexions rapides à des réseaux de fibres optiques. L'hébergement vers Montréal pour mettre en place le nouveau centre de données suite à un nombre de critères définis par Iweb :

- Le coût d'électricité à Montréal est très bas par rapport aux autres régions.
- La pollution.
- Montréal bénéficie de la tarification la plus basse en Amérique de nord pour l'énergie renouvelable. En effet, 100 % de l'électricité d'Iweb est générée au Canada, il s'agit d'énergie renouvelable.
- Risque minimal de catastrophes naturelles.
- Un climat frais : en effet, pour les données, Iweb utilise des technologies avancées de refroidissement qui sont aussi gratuites puisqu'il utilise l'air frais hivernal du Canada pour rafraîchir les serveurs. Les températures moyennes à Montréal sont significativement plus basses qu'aux États-Unis, avec des températures moyennes sous le point de congélation cinq moins par an.

**Remarque** *:* Amazon, Apple et Microsoft n'accordent pas suffisamment d'attention à la provenance de l'électricité qu'elles consomment et continuent d'avoir largement recours aux énergies non renouvelable pour alimenter leur infonuagique alors que leurs besoins sont exponentiels. Par contre, Yahoo et Google continuent de montrer l'exemple en faisant de l'accès aux énergies renouvelables une priorité pour la croissance de leur infonuagique ; ces deux entreprises soutiennent de plus en plus activement les politiques favorables à l'augmentation des investissements dans les énergies propres.

### ❖ Technoparc de Montréal

D'après monsieur Charles Lambert, le directeur de TIC au Technoparc de Montréal, le Technoparc offre un terrain pour placer les centres de données tout en faisant aux critères mentionnés dans le tableau suivant également, le Technoparc offre trois solutions de centres de données qui sont comme suit : Technoparc Montréal propose trois solutions

d'hébergement de données dans son Campus Saint-Laurent pour répondre aux besoins spécifiques tout en réduisant considérablement l'empreinte environnementale. En outre, l'infrastructure physique des centres de données varie selon la sécurité de l'endroit ou un centre de données est placé. Ainsi, les entreprises suivent L'Uptime Institute (*The global Datacenter Authority*) qui est l'autorité mondiale des centres de données. De même, pour le transfert des données, il existe une contrainte de bande passante. À Montréal, on a "Data communication" comme un fournisseur de bande passante. Comme on a déjà mentionné précédemment que Technoparc Montréal offre un terrain favorable pour mettre en place des centres de données tout en ayant recours à des critères propres à lui et presque considérés comme un standard. Ces critères sont définis dans le tableau 1.2

Tableau 1.2 Résumé des critères de sélection des localisations à Technoparc, Montréal

| Catégorie | Facteur | Position |
|---|---|---|
| **Énergie** | Coût | Parmi les taux d'électricité les plus compétitifs en Amérique du Nord : 4.83 ¢/KWh (6.0 MW de puissance) |
| | Fiabilité de la source d'énergie | • Le Canada se classe premier au monde pour la sécurité d'approvisionnement énergétique<br>• L'électricité du Québec est à l'abri des fluctuations mondiales du prix du pétrole<br>• Réseau électrique indépendant vous protégeant contre les pannes de courant(réf. 2003)<br>• Réseau électrique en boucle de 735 KV qui fait le tour de Montréal |
| | Distribution | • Connecté à un réseau d'énergie souterrain<br>• 2 sources d'alimentation d'énergie à tension moyenne (25 KV) |
| | Énergie renouvelable | 97 % de l'électricité de sources renouvelables |
| **Bande passante** | Connexion | De multiples connections à des transporteurs et fournisseurs de premier niveau (Tier 1) |
| | Réseau | 3 lignes différentes (n + 2) |
| | Performance | Performance pour réseau d'affaires fixe (vitesse de téléchargement et débit montant): le Canada est premier en Amérique du Nord (Cisco 2011) |
| **Localisation** | Accessibilité | • À proximité des autoroutes 40 et 13<br>• Près de l'Aéroport international de Montréal et du centre-ville de Montréal |
| | Qualité | Parc technologique et scientifique de classe mondiale situé sur un site prestigieux |
| | Expansion | Capacité d'expansion des infrastructures pour appuyer vos opportunités de croissance |
| | Environnement | Environnement de travail exceptionnel bordé de magnifiques espaces verts |
| | Sécurité | Situé dans une zone délimitée et patrouillée 7 jours/semaine |

| | | |
|---|---|---|
| **Main d'œuvre qualifiée** | R&D | • Fonds pour la recherche universitaire : Montréal se classe première au Canada<br>• 8 universités and 2 écoles d'ingénierie avec des départements TIC<br>• Plus de 40 centres de recherche en TIC |
| | Finissants universitaires | Plus de 2 000 gradués de programmes reliés au secteur TIC par année |
| | Disponibilité | 120 000 employés en TIC hautement qualifiés dans la région métropolitaine de Montréal |
| **Environnement propice aux affaires** | Lois sur la propriété intellectuelle | Lois efficaces pour protéger la propriété intellectuelle |
| | Accès au marché | Accord de libre-échange avec les États-Unis et plusieurs pays d'Amérique latine |
| | Coût d'opération | La région métropolitaine de Montréal se classe au premier rang en Amérique du Nord pour la compétitivité des coûts totaux d'opération des compagnies en TIC (KPMG 2012) |
| | Crédits de taxes | • Crédit d'impôts pour le développement des affaires électroniques et pour les dépenses de R&D<br>• Crédit de taxe sur la propriété pendant 5 ans |
| | Confidentialité | La loi américaine *Patriot Act* ne s'applique pas |

Source : Directeur de TIC au Technoparc de Montréal

### 1.3.1.2 Les critères de sélection des localisations : Généralisation

Covas, Silva et Dias (2013) présentent une analyse décisionnelle multicritères pour identifier la bonne localisation des centres de données en prenant en considération les dimensions économiques, sociales, techniques et environnementales. Ces dimensions seront évaluées en tant que problème de tri à multicritères. Ces problèmes seront analysés en utilisant la méthode de surclassement « ELECTRE TRI ». En outre, ils déclarent que les centres de données sont caractérisés par une grande consommation d'énergie. En effet, lorsqu'une entreprise décide d'installer un centre de données, la première variable de décision est le coût d'énergie. De même, la sécurité des données ainsi que la fiabilité des centres de données sont reliées à la localisation des centres de données. Également Covas, Silva et Dias (2013) ont défini les critères de sélection des sites pour installer un centre de données durable. Tout d'abord le coût d'énergie consommée par le centre de données. L'évaluation des critères de sélection pour placer des centres de données se diffère d'une dimension à une autre : en effet, au niveau de l'évaluation de l'environnement, l'économie de l'énergie est mesurée en termes de quantité de dioxyde de carbone ($CO_2$) émise alors que pour l'évaluation économique, l'économie d'énergie est mesurée en unités monétaires.

Covas, Silva et Dias (2013) ont classifié les critères de sélection des emplacements des centres de données en 4 catégories d'évaluation :

1) Critère d'évaluation des risques.
2) Critère d'évaluation économique.
3) Critère d'évaluation environnemental.
4) Critère d'évaluation social.

Dans une autre partie, Daim, Bhatla et Mansour (2012) définissent quatre facteurs pour la sélection des locaux des centres de données : facteur géographique, facteur financier, facteur politique et facteur social. En effet, le facteur géographique prend en compte l'absence de catastrophes naturelles, la disponibilité de transport, disponibilité de l'énergie et la disponibilité des ressources d'eau. Alors que le facteur financier prend en considération le coût de terrain, le coût de construction et la variabilité de coût d'électricité et les taxes. En plus, le facteur politique prend en considération essentiellement la création d'emploi. Finalement, le facteur social qui se base sur la sécurité contre les crimes et les lois et validité des lois sur l'urbanisme.

De plus, Fei et Li Xuan (2011) indiquent que les deux buts essentiels d'un centre de données est de maintenir les données stockées en sécurité et d'assurer la continuité des activités de l'entreprise. En effet, le choix des locaux est très primordial et une priorité pour les entreprises. De même, ils déclarent que la sécurité de centres de données se manifeste en trois points : tout d'abord, la sécurité physique de centre de données envers les catastrophes naturelles et humaines. Ensuite, la sécurité des informations stockées par l'utilisation des systèmes de prévention contre les attaques et les virus. Finalement, la continuité des activités des centres de données tout en planifiant des centres de récupération après les catastrophes. En outre Fei et Li Xuan (2011) utilisent la méthode Delphi pour le choix des localisations des centres de données. Cette méthode est utilisée souvent aux États-Unis pour une analyse quantitative pour un emplacement intelligent des centres de données. La figure 1.4 illustre les proportions des différents facteurs déterminés pour choisir une localisation d'un centre de données Fei et Li Xuan (2011).

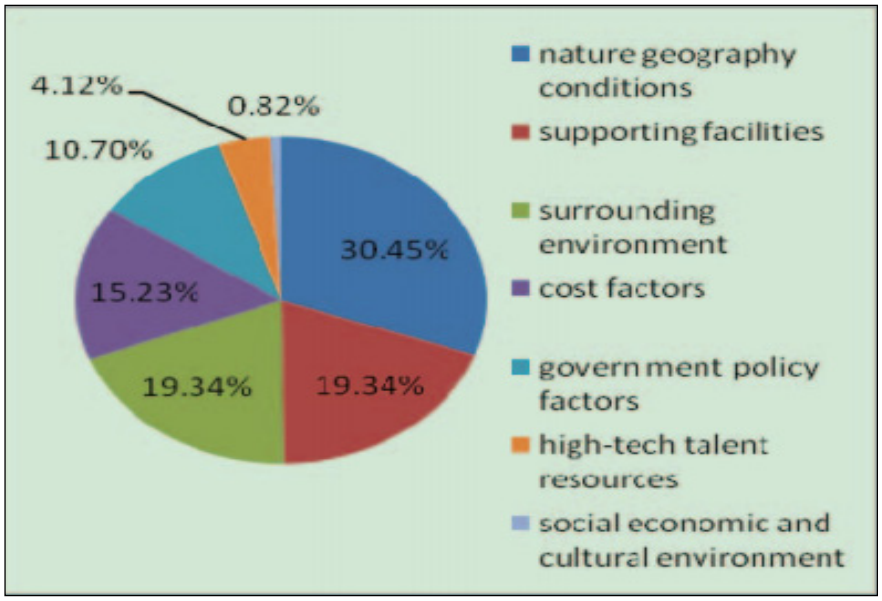


Figure 1.4 Les proportions pour les facteurs
de localisation des centres de données
Tirée de Fei and Li Xuan (2011)

Comme on a déjà mentionné que les critères de sélection se classifient selon les priorités et les besoins des entreprises concernées. En effet, les critères de sélection des sites pour les centres de données :

- Les coûts de l'énergie;
- Le climat;
- Les catastrophes naturelles;
- La situation politique de la région;
- La main d'œuvre;
- Les incitations fiscales;
- Le climat d'affaires;
- Le prix de terrain;
- La réglementation gouvernementale;
- Le taux des taxes (les impôts fonciers, les impôts sur les sociétés);
- Le transport.

#### 1.3.1.3 Le coût d'énergie

Pakbaznia et Pedram (2009) déclarent que la consommation d'énergie des serveurs et des centres de données a doublé au cours des cinq dernières années et devrait quadrupler au cours des cinq prochaines années pour plus de 100 milliards de kWh à un coût annuel d'environ 7,4 milliards de $. Également, Pakbaznia et Pedram (2009) indiquent que l'énergie représente aujourd'hui environ les deux tiers du coût total du centre de données; même supérieurs aux coûts totaux des équipements. Cette énergie est principalement consommée dans les systèmes auxiliaires tels que le système de refroidissement. Au niveau réglementaire, cette croissance massive de la consommation d'énergie et les émissions de dioxyde de carbone associées à la gestion de centres de données est susceptible d'attirer un examen par les législateurs qui cherchent des façons de réduire les émissions de dioxyde de carbone. Avec la croissance des demandes des services des centres de données, la quantité d'énergie consommée augmente. Ce qui engendre une augmentation des coûts énergétiques des centres de données. Par conséquent, ce critère représente le premier critère de sélection des localisations pour les centres de données.

#### 1.3.1.4 Le climat

Le climat est très important pour choisir une bonne localisation des centres de données et surtout un climat frais pour réduire l'empreinte écologique et par conséquent une réduction des coûts.

#### 1.3.1.5 Les catastrophes naturelles

Les risques des catastrophes naturelles influencent beaucoup sur les décisions prises pour choisir les terrais des centres de données. En effet, Montréal par exemple, se caractérise par un risque minimal des catastrophes naturelles. Également, si le site de centre de données est dans une région connue pour les séismes dans ce cas le bâtiment doit être construit de façon à réduire l'impact des tremblements de terre.

Rapport-Gratuit.com

Dans ce même contexte, Covas, Silva et Dias (2013) et Rath (2012) définissent les catastrophes naturelles possibles à rencontrer telles que :

- Les séismes;
- Les tempêtes de verglas : si le centre de donnée est localisé dans une région susceptible aux tempêtes de verglas, il faut prendre en considération que le centre de données doit s'alimenter via une alimentation de secours;
- Les ouragans;
- Les tornades;
- Les inondations;
- Les glissements de terrain : prendre les mêmes directives de sécurité que pour les inondations tout en ajoutant des murs extérieurs étanches et solides;
- Les incendies : Ils engendrent des dommages remarquables. Il faut installer des murs et des portes résistants aux feux et les systèmes d'extinction d'incendie.

**Remarque 1** : d'après Daim, Bhatla et Mansour (2012), le but de prendre en considération les catastrophes et risques naturels n'est pas d'éliminer ces risques, mais plutôt de les atténuer et de réduire les coûts et les charges pour faire fonctionner les centres de données.
**Remarque 2** : Le choix des emplacements des sites des centres de données doit prendre en considération, d'après Sun Microsystems, les risques naturels (les orages, les séismes), les risques humains (les pollutions industrielles, les vibrations) et la disponibilité des services publics.

#### 1.3.1.6 La pollution

Si le centre de données se trouve dans une région ou zone polluée ou il y a des traces de pollution dans l'air, il faut dans ce cas protéger les centres de données en limitant l'air entrant

dans les centres de données. En effet, la pollution peut endommager les serveurs et par conséquent l'infrastructure de centre de données.

#### 1.3.1.7 L'interférence électromagnétique

Si les centres de données sont localisés près des sources d'interférences électromagnétiques, ils reçoivent uniquement ¼ des radiations alors que si les centres de données sont 10 fois loin, ils reçoivent 1/100 des radiations électromagnétiques. Dans ce cas, il faut placer les centres de données loin des sources de radiations telles que les réseaux cellulaires.

#### 1.3.1.8 L'énergie renouvelable

L'énergie renouvelable ou appelée parfois l'énergie « verte » est l'un des critères importants les dernières années surtout avec la convergence des entreprises vers l'infonuagique. De même, Google a annoncé le 26 septembre 2014 avoir signé un accord avec 'Grand River Dam Authority' (GRDA), pour fournir à son centre de données situé dans le comté de Mayes (Oklahoma), environ 48 MW d'énergie éolienne. En effet, Lemay et al. (2012) indiquent que les problèmes du climat est causé par le gaz à effet de serre dont le dioxyde de carbone est le facteur clé. Ce problème climatique influence beaucoup sur la durabilité et le bon fonctionnement des centres de données. De même, les grandes entreprises de service TIC comme Microsoft et Google ont placé leurs centres de données près des sources d'énergies renouvelables.

Les centres de données de l'ÉTS et de l'UQAM utilisent l'énergie hydraulique. Étant donné que le Québec possède une grande capacité de production hydroélectrique. Pour cette raison, l'interruption des services dans le réseau est peu probable. En effet, Lemay *et al*. (2012) visent à utiliser l'énergie solaire et l'énergie éolienne à la place d'hydroélectricité, car ils pensent que cette dernière n'est pas obligatoirement renouvelable.

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

#### 1.3.1.9 Le transport

C'est un facteur primordial pour assurer la disponibilité et l'accès rapide aux centres de données à n'importe quel moment. En effet, Daim, Bhatla et Mansour (2012) indiquent que les centres de données devraient être toujours facilement accessibles par tout moyen de transport. Par conséquent, les fournisseurs de matériel informatique et les équipes de construction doivent être en mesure d'accéder facilement au site de centre.

### 1.3.2 Les contraintes de transfert des données

#### 1.3.2.1 Les contraintes réglementaires sur l'exportation des données

Haas (2013) indique que les récentes révélations de l'existence du programme PRISM qui est un programme de surveillance mis en place aux États-Unis et permettant l'accès par les autorités américaines aux données stockées dans l'infonuagique ont inquiété pouvoirs publics et citoyens. Pourtant, les entreprises doivent également considérer la localisation des données traitées dans l'infonuagique comme l'un des points essentiels qui doivent être contrôlés lorsqu'elles souhaitent bénéficier de ce type de services pour externaliser le traitement de leurs données. En effet, d'une part, il existe des contraintes règlementaires encadrant l'exportation de données dans l'infonuagique et, d'autre part, la localisation des données peut affecter le cadre légal applicable à leur traitement, y compris pour un éventuel accès à ces données par les autorités gouvernementales.

Dans le même contexte, Haas (2013) indique que le secteur bancaire est ainsi soumis aux dispositions du règlement n°97-02 du 21 février 1997 relatif au contrôle interne des établissements de crédit et des entreprises d'investissement. Il impose notamment des contrôles et formalités supplémentaires aux prestataires de services d'investissement qui ont recours à un prestataire situé dans un État non membre de la Communauté européenne.

Dans le secteur de la santé, par exemple, l'hébergement de données de santé à caractère personnel sur support informatique est également strictement encadré. Cette activité est

soumise à l'obtention d'un agrément, délivré en contrepartie de garanties en termes d'intégrité, de sécurité et de confidentialité des données traitées. Le dossier d'agrément doit notamment préciser le lieu où l'hébergement sera effectué, et toute modification de ce lieu devra faire l'objet d'une notification auprès du ministre chargé de la santé.

En outre, les transferts de données personnelles hors de l'Union européenne requièrent :

- Un contrat de transfert conforme aux clauses contractuelles types adoptées par la Commission européenne, ou la mise en place de règles internes d'entreprise (*binding corporate rules*).
- Une autorisation préalable par la CNIL (Commission nationale de l'informatique et des libertés)**,** sauf lorsque le transfert est à destination de pays reconnu par la Commission européenne comme assurant un niveau de protection suffisant des données, ou de destinataires situés aux États-Unis et ayant adhéré aux principes du Sphère de sécurité.

**Remarque 1** : Les *Binding Corporate Rules* (BCR) constituent un code de conduite, définissant la politique d'une entreprise en matière de transferts de données. Les BCR permettent d'offrir une protection adéquate aux données transférées depuis l'Union européenne vers des pays tiers à l'Union européenne au sein d'une même entreprise ou d'un même groupe.

**Remarque 2** : Les principes de la Sphère de sécurité (*SafeHarbor*) permettent à une entreprise américaine de certifier qu'elle respecte la législation de l'Espace Économique Européen (EEE) afin d'obtenir l'autorisation de transférer des données personnelles de l'EEE vers les États-Unis et Canada.

#### 1.3.2.2 Transfert des données en dehors du Canada

Pour les contraintes de transfert des données, en dehors du Canada, nous avons recours à Industrie Canada pour déterminer les lois de transfert des données en dehors du pays. De même, la commissaire à la protection de la vie privée du Canada déclare sur son site officiel :

(www.priv.gc.ca) deux lois pour le transfert des données en dehors du Canada. En effet, ces deux lois sont nommées comme suit :

- **La Loi sur la protection des renseignements personnels**

La loi sur la protection des renseignements personnels, qui est entrée en vigueur le 1er juillet 1983, impose des obligations à quelque 250 ministères et organismes fédéraux en matière de respect du droit à la vie privée en limitant la collecte, l'utilisation et la communication de renseignements personnels. Elle confère aux personnes le droit d'avoir accès aux renseignements personnels les concernant qui sont détenus par des organismes fédéraux et à demander qu'ils soient corrigés Canada (2013).

- **La Loi sur la protection des renseignements personnels et les documents électroniques (LPRPDE)**

La loi sur la protection des renseignements personnels et les documents électroniques (LPRPDE) définit les règles applicables à la façon dont les organisations du secteur privé peuvent recueillir, utiliser et communiquer des renseignements personnels dans le cadre d'activités commerciales. La LPRPDE s'applique également aux entreprises fédérales pour ce qui est des renseignements personnels des employés. En vertu de cette loi, les personnes peuvent avoir accès aux renseignements personnels que les organisations ont recueillis à leur sujet et demander que des corrections y soient apportées. En général, la LPRPDE s'applique aux activités commerciales des organisations dans toutes les provinces, à l'exception des organisations qui recueillent, utilisent ou communiquent des renseignements personnels exclusivement dans les provinces qui ont leurs propres lois en matière de protection des renseignements personnels qui ont été déclarées essentiellement similaires à la loi fédérale. Dans ces cas, c'est la loi provinciale essentiellement similaire qui s'appliquera au lieu de la LPRPDE, même si la LPRPDE continue de s'appliquer aux transferts interprovinciaux ou internationaux de renseignements personnels Canada (2013).

**Exemple** : pour le secteur de télécommunication et de transport, la loi fédérale est appliquée et non pas la loi provinciale.

### 1.3.3 Modélisation des problèmes de localisation

#### 1.3.3.1 Texte introductif

Dans cette section, nous présentons quelques problèmes de localisation des centres de données en se basant sur certains critères : minimisation des coûts, minimisation de l'énergie consommée et minimisation de bande passante et l'efficacité énergétique. Comme récapitulatif, notre projet de recherche porte sur le problème de localisation des centres de données offrant des services Internet. Alors qu'il existe quelques revues qui parlent de problème de localisation et uniquement une seule revue qui s'intéresse et coïncide avec notre projet de recherche, mais avec quelques limites. Pour ces raisons, nous allons présenter les solutions proposées par les revues de littérature une fois que les centres de données sont localisés (problèmes à l'intérieur des centres de données). Par la suite, nous allons présenter les problèmes de localisation décrits dans la revue. Finalement, nous allons critiquer la solution proposée qui porte sur la localisation des centres de données pour des services Internet.

#### 1.3.3.2 Localisation des centres de données

Lorsqu'on parle de l'énergie, on dit l'énergie électrique, mais essentiellement l'énergie renouvelable qui sera la solution pour les grandes entreprises comme Google, IBM, Microsoft. En effet, le coût de l'énergie est l'un des principaux constituants du coût d'exploitation d'une installation. En effet, Raghavendra *et al*. (2008) indique que le coût d'énergie qui alimente le centre de données et l'énergie pour le refroidissement des centres de données est plus de 30 millions de dollars en 2008. De plus, Van Heddeghem *et al.* (2014) indiquent que la consommation d'énergie électrique totale dans le monde augmente de 3.9 % en 2007 à 4.6 % en 2012.

Dans ce contexte, Hong et Baochun (2012) développent un système d'optimisation qui prend en considération l'utilisation de la bande passante et le coût de l'électricité pour résoudre les

problèmes de sélection des centres de données pour les services de l'infonuagique. Les services d'infonuagique sont répartis dans des infrastructures géographiquement distribuées. Également, les centres de données sont localisés dans différentes régions pour fournir des services plus performants. Pour cette raison, les fournisseurs de services ont besoin d'une méthode pour rediriger les clients vers les centres de données appropriés. En effet, Hong et Baochun (2012) ont montré que la méthode traditionnelle pour la sélection des centres de données utilise les serveurs DNS comme illustré à la figure 1.5.

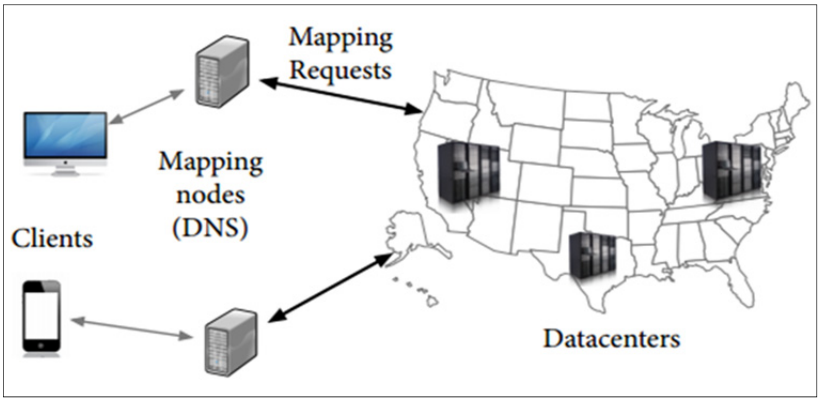


Figure 1.5 Exemple de service d'infonuagique déployé
sur une infrastructure géographiquement distribuée
Tirée de Hong and Baochun (2012)

Chaque zone géographique possède un coût d'électricité différent. De même, Hong et Baochun (2012) indiquent que la distribution géographique des centres de données présente un avantage. En effet, cette distribution est liée à une diversité des prix d'électricité. Dans ce cas, les demandes de clients seront redirigées vers les centres de données dont le coût d'électricité est réduit. Les auteurs indiquent qu'il existe autre que le coût d'électricité, un autre contributeur significatif de coût qui est la bande passante de réseau étendu (à grande échelle). La méthode utilisée par Hong et Baochun (2012) est le multiplexage de la bande passante dans le but de réduire la consommation de la bande passante et le coût. Dans ce cadre, Hong et Baochun (2012) ont développé un système d'optimisation des coûts de sélection des centres de données. Ce système prend en considération le coût d'électricité et de la bande passante.

En outre, ce système est à la base un algorithme d'optimisation qui vise à sélectionner l'affectation optimale des nœuds de centre de données. Cette affectation optimale est basée sur la méthode de sous-gradient « *subgradient* ».
**Note :** La méthode sous-gradient **«** *subgradient* **»** est une méthode itérative pour la résolution des problèmes de minimisation convexe.

Comme nous l'avons mentionné, le système d'optimisation développé par Hong et Baochun (2012) vise à minimiser les coûts de consommation de l'électricité et de la bande passante. Dans ce cadre, ils ont donné deux exemples pour le coût d'électricité dans le cas où les centres de données se trouvent soit dans une région règlementée ou non réglementée. Par exemple, pour certains états aux États-Unis (US) en tant que région règlementée, le prix d'électricité est fixe. Alors que pour les états de la Californie et du Texas en tant que deux régions non-règlementées, le prix d'électricité est variable. Par ailleurs, Hong et Baochun (2012) n'ont pas pris en considération le prix de la bande passante : ce prix est presque fixe pour une grande majorité des régions.

Afin de résoudre le problème d'optimisation défini sous forme d'un programme linéaire Hong et Baochun (2012) utilisent la relaxation lagrangienne. En effet, ils prouvent que la fonction d'optimisation des centres de données « DC-OPT » peut être décomposée en $n$ problèmes de maximisation par centre de données.

**Remarque** : si le 95 % des bandes passantes de tous les centres de données dépasse le maximal de la bande passante, dans ce cas l'infonuagique augmente son prix de bande passante pour supprimer le trafic en excès.

Les centres de données sont choisis aléatoirement et en lui affectant un jeton avec un budget total $B$. En effet, le centre de données marque le jeton et l'envoie au centre de données suivant qui est à son tour met à jour le budget qui reste. De même, la fonction d'optimisation DC-OPT réduit l'utilisation de la bande passante de la plupart des centres de données de 15 à 20 %.

Comme on a déjà mentionné que le coût d'électricité varie selon les régions et pour certaines régions, ce coût est plus élevé par rapport à d'autres régions. Dans ce cas, l'énergie renouvelable sera une bonne solution pour réduire les coûts des centres de données. Afin de présenter l'importance de minimisation des coûts d'énergie des centres de données, Covas, Silva et Dias (2012) présentent les différentes proportions d'utilisation de l'énergie par un centre de données à la figure 1.6.

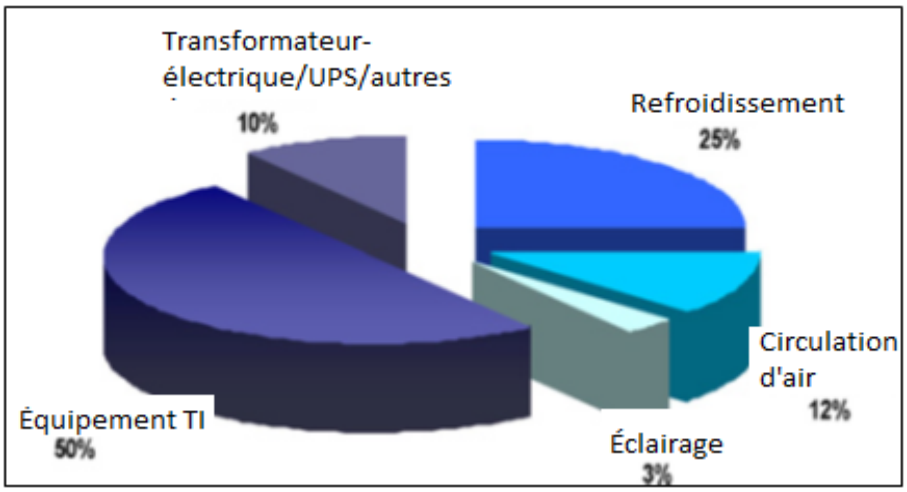


Figure 1.6 Les proportions d'utilisation d'énergie
dans un centre de données
Tirée de Covas *et al.* (2012)

Également, Lei et al. (2010) décrivent le problème de minimisation de coût d'électricité pour des centres de données offrant des services Internet. En effet, le problème de minimisation porte essentiellement sur un environnement des marchés à électricités multiples. La notion de multi-électricités signifie que le coût d'électricité varie en fonction de lieu et de l'heure. Ce qui engendre un problème des coûts pour les centres de données qui fonctionnent 356/7/24. Alors, l'objectif de Lei *et al.* (2010) est de minimiser les coûts des centres de données tout en garantissant la qualité de service adapté à la diversité des prix de l'électricité en fonction de diversité du lieu et l'heure de fonctionnement des centres de données. Le modèle proposé est un programme linéaire en nombres mixtes.

En effet, le coût total de localisation de $N$ centres de données est donné par la fonction suivante :

$$T_{total} = \sum_{i=1}^{N} m_i Pr_i(t) Po_i \tag{1.1}$$

Cette fonction est le produit entre le nombre de serveurs allumés $m_i$ par l'énergie nécessaire pour le fonctionnement d'un serveur $Po_i$ par le prix de l'électricité $Pr_i(t)$ dans la localisation $i$.

Yuanxiong *et al*. (2011) présentent le même problème de minimisation des coûts d'électricité proposés par Lei et al. (2010) mais leur résolution de ce problème est faite à travers la technique d'optimisation de Lyapunov. En effet, l'algorithme proposé permet de résoudre le problème de la répartition optimale du trafic et le chargement / déchargement de la batterie énergétique pour des centres de données offrant des services Internet.

Dans le même contexte, Jie et al. (2012) et Yao et al. (2012) décrivent le problème de minimisation des coûts d'énergie électrique des centres de données tout en proposant une modélisation mathématique d'un programme linéaire en nombres mixtes. Cette solution assure une gestion optimale de la demande d'énergie électrique des centres de données. De plus Yao et al. (2012) présentent la gestion des demandes d'énergie par un bon lissage des demandes énergétiques (une répartition des demandes en fonction des besoins).

Autre que les problèmes liés directement aux centres de données une fois localisés tels que la minimisation des coûts d'énergie électrique et les coûts de bande passante, nous nous intéressons aussi au problème de sélection de localisation tout en répondant aux demandes des clients.

Chang, Patel et Withers (2007) présentent un modèle d'optimisation pour déterminer les localisations des centres de données dans le domaine militaire.

Rapport-Gratuit.com

Afin de choisir la bonne localisation, les auteurs définissent quatre facteurs de décision : 1) les capacités associées aux sites d'installation, 2) la capacité et la latence de réseau WAN, 3) la proximité à la demande et au volume de demande et 4) la performance des applications (qualité de service (Qos) des applications).

Dans un autre contexte, Goiri et al. (2011) choisissent la localisation des centres de données pour des services Internet en se basant sur les critères fondamentaux suivants: le temps de réponse, le délai de cohérence, la disponibilité des serveurs et l'émission de $CO_2$. En outre, le modèle d'optimisation présenté par Goiri et al. (2011) vise à minimiser le coût total du réseau de centre de données en respectant la latence du réseau, le délai de cohérence et/ou la disponibilité.

Mais, il existe quelques lacunes de cette solution. Tout d'abord, Goiri et al. (2011) n'ont pas spécifié la nature des services Internet. En effet, pour les services Internet nous pouvons distinguer les services Web, les services courriel, les services de communication en direct, les services de traitement informatique, les services de messagerie, *etc*. Chaque service Internet possède un temps de réponse différent en fonction des demandes des clients. Finalement, la solution proposée par es auteurs ne prend pas en considération le calcul de la métrique de l'efficacité énergétique (PUE).

Également, quelques articles électroniques ont parlé de localisation des centres de données, mais en terme de coût opérationnel ou de la probabilité des catastrophes naturelles telles que les solutions proposées par Oley (2010), Alger (2005) et Stansberr (2006). Cependant, les travaux mentionnés précédemment n’ont pas pris en considération l’optimisation de tous les coûts ou des contraintes pertinentes. Par exemple, aucun d’entre eux ne considère l’efficacité énergétique d’un centre de données et la détermination de l’énergie consommée par les équipements TI.

Malgré que les travaux antérieurs n’ont pas étudié le processus de localisation et emplacement des centres de données d’une manière complète ou détaillée, certaines études

rigoureuses ont étudié la mise en place d'autres ressources informatiques distribuées. Par exemple, Wierzbicki (2002) utilise des techniques d'optimisation pour placer des caches dans les *content delivery network* (CDN).

Notre travail diffère de ces efforts en ce qui concerne notre focalisation sur les centres de données qui offrent des services Internet. Également, nous travaillons sur les coûts physiques, l'efficacité énergétique et les caractéristiques géographiques des localisations. La combinaison de ces problématiques dans seul modèle d'optimisation rend notre travail différent des travaux antérieurs.

## 1.4 Conclusion

Les centres de données présentent un élément stratégique pour les entreprises. La localisation des centres de données a un impact direct sur le temps de réponse des services, les coûts opérationnels, la qualité des services, *etc*. Également, l'apparition et la croissante utilisation d'infonuagique influent beaucoup la localisation et surtout le dimensionnement des centres de données. Pour cette fin, nous définissons les critères de sélection des sites pour placer un centre de données ainsi que les contraintes liées au transfert des données.

Évidemment, les grandes entreprises d'Internet, telles que Microsoft et Google ont leurs propres processus de sélection des emplacements pour leurs centres de données. Malheureusement, leurs processus sont très confidentiels pour des fins commerciales. L'un des objectifs de notre travail est de mettre l'accent sur ce problème et permettre à la communauté de recherche de traiter cette problématique dans le domaine public.

L'étape suivante sera la modélisation du problème de localisation des centres de données. En effet, nous allons utiliser ces contraintes et ces critères pour définir un modèle mathématique qui a comme objectif la localisation des centres de données.

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

# CHAPITRE 2

# MODÉLISATION MATHÉMATIQUE

## 2.1 Introduction

Les centres de données et la sélection de leur localisation sont un axe de recherche très actif. En effet, la revue de la littérature présentée, dans le chapitre précédent ont permis de recenser des méthodes de résolutions des problèmes de localisations des centres de données ainsi que le problème de minimisation des coûts de la mise en place. De plus, nous avons pris en considération les problèmes de localisation indiqués dans le premier chapitre. Par conséquent, nous formulons un problème de localisation qui combine à la fois le problème de minimisation des coûts totaux des centres de données et le problème de leur efficacité énergétique. Également, le problème de localisation est appliqué sur des centres de données qui offrent des services Internet.

En outre, les services Internet populaires sont hébergés dans des centres de données distribuées géographiquement. La localisation des centres de données a un impact direct sur le temps de réponse des services et leurs coûts (fixes et variables). La localisation prend en considération plusieurs critères y compris la proximité aux zones de marché, les sources d'énergie dans la région, les coûts des terrains, électricité, la main d'œuvre et la température moyenne au site retenu. Dans ce cadre, nous présentons un modèle mathématique tout en prenant en considération ces critères pour sélectionner la localisation potentielle des centres de données. Ce choix vise d'une part, à optimiser le coût total d'implémentation des centres de données et d'autre part, nous déterminons l'efficacité énergétique des centres de données à installer avec une bonne gestion des équipements TI et par conséquent une meilleure gestion des tailles des centres de données.

## 2.2 Les hypothèses

Tout au long de cette partie, nous admettons cinq hypothèses afin de tester le modèle mathématique précédemment présenté.

**Hypothèse 1** : Nous choisissons de mettre en place des centres de données qui offrent trois types de services Internet différents :

- Le service Web (serveurs Web),
- Le service de traitement information (Serveur FTP),
- Le service de communication en direct (Serveur de vidéoconférence).

**Remarque** : La disponibilité diffère d'un service à un autre. Par exemple, le service de vidéoconférence nécessite une haute disponibilité pour assurer une meilleure qualité de service (QoS) alors que le service de recherche nécessite moins de disponibilité et moins de ressources pour livrer des résultats aux clients.

**Hypothèse 2** : Nous admettons que le niveau de redondance des centres de données à mettre en place est le niveau 1 (Level I).
En effet, c'est le niveau de redondance le plus basique. Et c'est le moins coûteux pour le mettre en place (une seule source d'alimentation; les systèmes de refroidissement et les génératrices sont de types N[1]).

**Hypothèse 3 :** Nous sélectionnons de travailler sur 5 régions de la province de Québec pour mettre en place des centres de données. Ainsi, nous choisissons que les centres de données desservent 13 zones de marché réparties dans de la province.

---

[1]La valeur N décrit le type de secours disponible dans l'éventualité de la défaillance du composant. Par exemple, la redondance de type N+1 signifie que les composants (N) ont au moins un composant de secours indépendant (+1).

**Hypothèse 4 :** Nous affectons un commutateur à 10 ports pour interconnecter les serveurs, dont 8 ports entrants pour les serveurs et 2 ports pour la connexion Ethernet.

**Hypothèse 5 :** Les serveurs des centres de données sont tous de même ressource $r \in R$. En effet, ils ont les mêmes caractéristiques physiques (même capacité de stockage, même taille de RAM, même type de carte réseau et carte de fibre optique).

## 2.3 Approche d'optimisation proposée

Afin de résoudre le problème de localisation des centres de données, nous proposons une modèle mathématique comme solution.

### 2.3.1 Description du modèle mathématique

Dans cette section, nous présentons notre modèle dans le cadre de la sélection des localisations des centres de données pour des services Internet. En effet, notre modèle est un programme linéaire en nombres mixtes qui assure la localisation des centres de données tout en minimisant le coût total de mise en place des centres de données. Par ailleurs, la sélection des localisations prend en considération les contraintes suivantes : le temps de réponse de service Internet, le délai de latence et l'efficacité énergétique de centres de données à installer.

Notre modèle mathématique est décrit à la figure 2.1

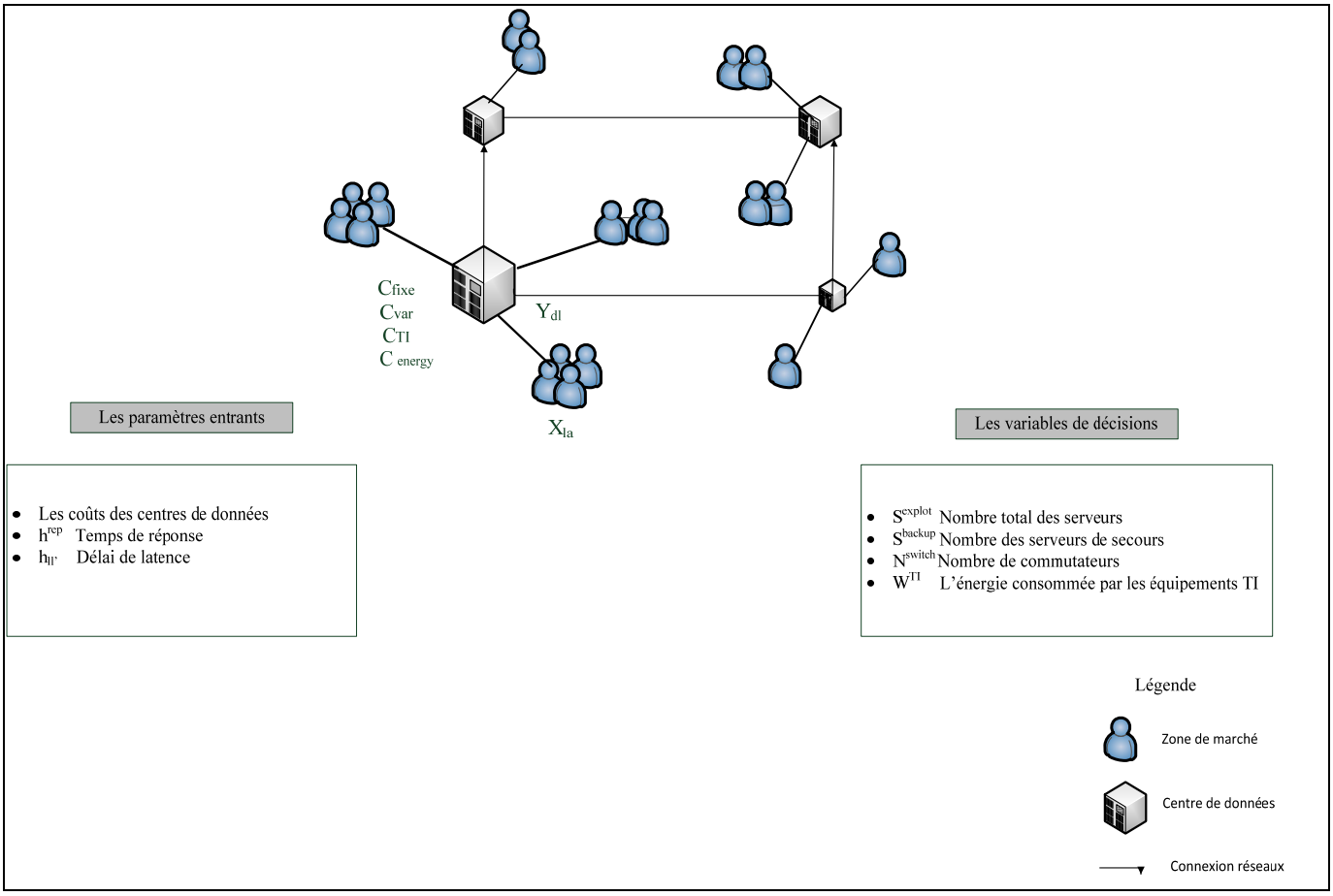


Figure 2.1 Schéma descriptif du modèle mathématique

### 2.3.2 Les ensembles et les indices

$l \in L$ Localisation des centres de données.

$a \in A$ Zone de marché.

$d \in D$ Taille d'un centre de données.

$s \in S$ Les services Internet à fournir. (serveur WEB, serveur de traitement d'information, serveur de communication en direct).

$r \in R$ Ensemble de ressources que chaque serveur dispose, qui comprennent : 1=CPU, 2=Carte réseau, 3=RAM, 4=Stockage, 5=carte Fibre optique.

### 2.3.3 Les paramètres

Afin de définir notre modélisation, nous avons besoin de définir tout d'abord, les paramètres qui vont être utiles pour la formulation mathématique. En effet, ces paramètres sont divisés en deux parties. Les coûts de mise en place des centres de données et d'autres paramètres qui sont requis et utiles pour notre formulation.

Les coûts des centres de données sont présentés par les paramètres suivants

| | |
|---|---|
| $c_{dl}^{maint}$ | Coût annuel de maintenance d'un centre de données localisé dans $l \in L$ et de taille $d \in D$. |
| $c_{r}^{server}$ | Coût net de serveur de ressources $r \in R$. |
| $c_{d}^{admin}$ | Coût d'administration d'un centre de données de taille $d \in D$ ($/administrateur). |
| $c_{d}^{erergy}$ | Coût d'énergie d'un centre de données de taille $d \in D$. |
| $c_{dl}^{land}$ | Coût de terrain d'un centre de données localisé dans $l \in L$ et de taille $d \in D$. |
| $c_{dl}^{cooling}$ | Coût de refroidissement d'un centre de données localisé dans $l \in L$ et de taille $d \in D$. |
| $c^{BW}$ | Coût de bande passante. |
| $c^{switch}$ | Coût net de commutateur ($/commutateur). |
| $c^{FiberOptic}$ | Coût de fibre optique ($/ km). |

**Les coûts** : Les coûts totaux d'un réseau des centres de données sont décortiqués en coûts fixes ($C^{fixe}$) et coûts variables ($C^{\text{var}}$). Mais pour mettre l'accent sur d'autres coûts, nous décortiquons en plus les coûts variables en coûts des équipements TI ($C^{TI}$), coûts d'énergie ($C^{energy}$) et coûts de bande passante ($C^{BW}$). En effet, les coûts fixes sont formés par les coûts des terrains et les coûts de refroidissement. De plus, le coût de terrain varie selon la

localisation, tout comme pour le coût de refroidissement qui dépend de la localisation et par conséquent, il dépend de la température ambiante de la localisation. Les coûts variables sont formés par les coûts d'administration et les coûts de maintenance des centres de données. Ces coûts dépendent principalement des tailles des centres de données. Également, ces coûts sont dominés par la rémunération du personnel. Ensuite, les coûts d'énergie dépendent du nombre des serveurs hébergés et de la taille du centre de données. Également, ces coûts varient selon la localisation. De même, les coûts des équipements TI et les coûts de la bande passante dépendent du nombre des serveurs hébergés et des ressources qui les caractérisent. Tous ces coûts dépendent du niveau de redondance qui sera construit dans chaque centre de données. En effet, les niveaux de redondances sont caractérisés par une redondance des éléments de secours comme : le générateur d'énergie, les serveurs, les équipements réseau, *etc*.

Les paramètres suivants sont aussi requis :

| | |
|---|---|
| $d_{la}^{location}$ | Distance entre la localisation $l \in L$ et la zone de marché $a \in A$. |
| $n_d$ | Le nombre d'administrateurs requis pour l'administration d'un centre de données de taille $d \in D$. |
| $b_s$ | La bande passante externe requise pour chaque serveur de service $s \in S$ (la vitesse de connexion des serveurs à Internet). |
| $w_d$ | L'énergie entrante nécessaire pour alimenter un centre de données de taille $d \in D$. |
| $h_{ll'}$ | Délai de latence entre deux localisations potentielles $l \in L$ et $l' \in L$. |
| $h_{sa}^{rep}$ | Temps de réponse de service $s \in S$ pour une zone de marché $a \in A$. |
| $f_d^{\max}$ | La capacité maximale de centre de données en nombre des équipements TI (serveurs, commutateurs) pour un centre de données de taille $d \in D$. |
| $\delta_a$ | Nombre des serveurs hébergés pour desservir une zone de marché $a \in A$. |
| $k_{sa}$ | Temps de réponse maximal d'un service $s \in S$ pour une zone de marché $a \in A$. |
| $p^{switch}$ | Puissance entrante pour alimenter un commutateur. |

Les constantes du modèle :

| | |
|---|---|
| $m$ | Délai de latence maximale. |
| $\beta$ | Facteur de pondération de l'énergie hydraulique. |
| $\lambda$ | La valeur maximale de l'efficacité énergétique. |
| $\alpha$ | Facteur de pondération de l'énergie électrique. |
| $M$ | L'énergie maximale qui peut être consommée par les équipements TI. |

Afin de bien éclaircir notre contribution, nous décrivons en détail les principaux paramètres qui forment notre modèle mathématique**.**

**Le temps de réponse :** Les demandes des services Internet sont augmentées rapidement. Afin de maintenir la haute disponibilité des services, il faut prendre en considération, dans notre modélisation, le temps de réponse des services. Dans ce cas, nous définissons le paramètre $h_{sa}^{rep}$ qui définit le temps de réponse de service $s \in S$ pour une zone de marché $a \in A$. Notamment, chaque type de service Internet possède un temps de réponse différent. Par exemple, le service Web exige moins de temps de réponse que le service de communication en direct.

**Le délai de latence :** un des objectifs des centres de données qui offrent des services Internet est le maintien de la disponibilité des services offerts pour les utilisateurs potentiels. Dans ce cadre, nous pensons au délai de latence qui est calculé une fois qu'un centre de données sera inaccessible ou indisponible. Également, parce que les services doivent être toujours disponibles même si les centres de données deviennent inaccessibles ou indisponibles, les centres de données qui fournissent les services Internet doivent être des miroirs les uns des autres. Ça signifie que le changement d'état d'un centre de données doit être propagé aux centres de données de secours (miroirs). Le délai de latence réfère au temps requis pour atteindre le centre de données de secours. Dans ce cas, le paramètre $h_{ll'}$ est calculé en

Rapport-Gratuit.com

fonction de la distance qui sépare un centre de données actif et son centre de secours. Les valeurs des délais de latence doivent être inférieures ou égales à une valeur maximale $m$ qu'on va définir dans le prochain chapitre.

**L'indicateur d'efficacité énergétique (PUE) :** les centres de données sont caractérisés par une grande consommation d'énergie. Également, avec l'augmentation des demandes des utilisateurs pour les services, le nombre des serveurs hébergés va augmenter et par conséquent l'énergie consommée par ces équipements. Dans ce cas, il est important de déterminer la valeur de PUE. Dans notre modélisation mathématique, nous définissons PUE en fonction de la variable de l'énergie consommée par les équipements TI. Cet indicateur permet de calculer la bonne dimension des centres de données à mettre en place et d'affecter par la suite aux zones de marché le nombre adéquat des serveurs hébergés et de secours. La valeur de PUE est calculée comme suit :

$$\frac{w_d \times \alpha + w_{dl}^{cooling} \times \beta}{w_{dl}^{TI} \times \alpha} \tag{2.1}$$

Cette valeur est le rapport de l'énergie total d'un centre de données et l'énergie nécessaire pour le refroidissement par l'énergie totale consommée par les équipements TI (serveurs et commutateurs).

### 2.3.4 Les variables de décision

Pour sélectionner la localisation potentielle des centres de données dans la province de Québec, les variables de décisions ci-dessous sont requises. En effet, les variables de décision de notre modèle sont composées des variables binaires et des variables continues pour la demande des utilisateurs.

$Y_l^{center}$ Variable binaire égale à 1 si le centre de données localisé dans la localisation $l \in L$ est de taille $d \in D$, 0 sinon.

$X_{la}^{backup}$ Variable binaire égale à 1 si le centre de données de secours est localisé dans $l \in L$

pour desservir une zone de marché $a \in A$, 0 sinon.

$Y_{dl}^{size}$ Variable binaire égale à 1 si le centre de données localisé dans $l \in L$ est de taille $d \in D$, 0 sinon.

$X_{la}^{center}$ Variable binaire égale à 1 si le centre de données est localisé dans $l \in L$ pour desservir une zone de marché $a \in A$, 0 sinon.

$S_{la}^{backup}$ Le nombre total des serveurs de secours pour un centre de données localisé dans $l \in L$ et dessert une zone de marché $a \in A$.

$W_{dl}^{TI}$ L'énergie consommée par les équipements TI pour un centre de données localisé dans $l \in L$ et de taille $d \in D$.

$S_{la}^{explot}$ Le nombre total de serveurs pour chaque centre de données placé dans la localisation $l \in L$ pour desservir une zone de marché $a \in A$.

$N_{la}^{switch}$ Le nombre total des commutateurs pour interconnecter les serveurs d'un centre de données localisé dans $l \in L$ qui dessert une zone de marché $a \in A$.

### 2.3.5 Modèle d'optimisation

En utilisant les notations décrites précédemment, un modèle de programmation en nombres mixtes est formulé. En effet, la configuration du problème consiste en une compagnie de service Internet qui vise à sélectionner des emplacements pour un ensemble de centres de données dans un ensemble des localisations potentielles ($l \in L$). Cette compagnie cherche à offrir ses services à un ensemble de zones de marché ($a \in A$). Par conséquent, chaque centre de données possède une taille $d \in D$. L'objectif d'optimisation est de minimiser le coût total d'un réseau de centres de données tout en respectant les contraintes de temps de réponse des services Internet, le délai de latence et l'efficacité énergétique.

Le coût total de mise en place des centres de données est décomposé en six coûts : 1) le coût fixe (2.2) qui est formé par le coût de terrain et le coût de refroidissement. 2) Un coût variable (2.3) qui est formé par le coût d'administration et le coût de maintenance des centres de données. 3) Le coût des équipements TI (2.4) qui est constitué par les coûts des serveurs y

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

compris les serveurs de secours et les serveurs hébergés et les coûts des commutateurs. 4) Le coût d'énergie (2.5) est formé par l'énergie totale entrante pour alimenter un centre de données de taille $d \in D$. 5) Le coût de la bande passante (2.6) et 6) le coût de fibre optique (2.7) en fonction de la distance séparant la localisation et la zone de marché à desservir.

$$C^{fixe} = \sum_{d \in D} \sum_{l \in L} \left( c_{dl}^{land} + c_{dl}^{cooling} \right) Y_{dl}^{size} \tag{2.2}$$

$$C^{\text{var}} = \sum_{d \in D} \sum_{l \in L} \left( n_d c_d^{admin} + c_d^{maint} \right) Y_{dl}^{size} \tag{2.3}$$

$$C^{TI} = \sum_{l \in L} \sum_{a \in A} c_1^{server} (S_{la}^{\exp lot} + S_{la}^{backup}) + c^{switch} N_{la}^{switch} \tag{2.4}$$

$$C^{energy} = \sum_{l \in L} \sum_{d \in D} w_d \, c_d^{energy} Y_{dl}^{size} \tag{2.5}$$

$$C^{BW} = \sum_{s \in S} \sum_{l \in L} \sum_{a \in A} b_s \, c^{BW} S_{la}^{\exp lot} \tag{2.6}$$

$$C^{distance} = c^{fiberOptic} \sum_{l \in L} \sum_{a \in A} d_{la}^{location} X_{la}^{center} \tag{2.7}$$

La fonction-objectif (2.8) vise à minimiser les coûts précédemment mentionnés pour sélectionner la localisation potentielle des centres de données.

$$Min\ F1 = C^{energy} + C^{fixe} + C^{\text{var}} + C^{TI} + C^{BW} + C^{distance} \tag{2.8}$$

Les contraintes sont :

Un centre de données peut desservir plusieurs zones de marchés à la fois.

$$\sum_{l \in L} X_{la}^{center} \geq 1, \forall a \in A \tag{2.9}$$

Déterminer un centre de donnée : ouvert ou non avec une seule dimension.

$$\sum_{d \in D} Y_{dl}^{size} \leq 1, \forall l \in L \tag{2.10}$$

Le centre de données doit exister (ouvert) pour desservir une zone de marché

$$X_{la}^{center} \leq \sum_{d \in D} Y_{dl}^{size}, \forall l \in L, a \in A \tag{2.11}$$

Chaque centre de données a au moins un centre de secours

$$\sum_{l \in L} X_{la}^{backup} \geq 1, \forall a \in A \tag{2.12}$$

Contrainte informative pour déterminer l'état d'un centre de données

$$\sum_{d \in D} Y_{dl}^{size} = Y_{l}^{center}, \forall l \in L \tag{2.13}$$

Un centre de données de secours doit exister pour desservir une zone de marché.

$$\begin{aligned} &\sum_{d \in D} Y_{dl}^{size} \geq X_{la}^{bakup}, \forall l \in L, a \in A \\ &\sum_{d \in D} Y_{dl}^{size} \geq X_{la}^{bakup}, \forall l \in L, a \in A \end{aligned} \tag{2.14}$$

Le centre de données et son centre de secours ne sont pas placés dans la même localisation.

$$X_{la}^{center} + X_{la}^{backup} \leq 1, \forall l \in L, a \in A \tag{2.15}$$

Contrainte de demande qui détermine le nombre total des serveurs

$$\delta_a X_{la}^{center} = S_{la}^{\exp lot}, \forall l \in L, a \in A \tag{2.17}$$

Le centre de données de secours répond aux besoins maximaux des zones de marché en nombre de serveurs

$$\delta_a X_{la}^{backup} \leq S_{la}^{backup}, \forall l \in L, a \in A \tag{2.18}$$

La capacité maximale de centre de données en nombres de commutateurs et serveurs

$$\sum_{a \in A} (S_{la}^{\exp lot} + S_{la}^{backup} + N_{la}^{switch}) \leq \sum_{d \in D} f_d^{\max} Y_{dl}^{size}, \forall l \in L, a \in A \tag{2.19}$$

Assure la proportionnalité entre le nombre de serveurs et commutateurs

$$\begin{aligned} &\mu N_{la}^{switch} - 1 < S_{la}^{\exp lot}, \forall l \in L, a \in A \\ &\mu N_{la}^{switch} \geq S_{la}^{\exp lot} + S_{la}^{backup}, \forall l \in L, a \in A \end{aligned} \tag{2.20}$$

Délai de latence entre deux localisations potentielles

$$h_{ll'}(X_{la}^{center} + X_{la}^{backup} - 1) \leq m, \forall l \in L, l' \in L, a \in A, l \neq l' \tag{2.21}$$

L'énergie consommée par les équipements TI

$$\sum_{a\in A}(p^{server}S_{la}^{explot}+p^{switch}N_{la}^{switch})=\sum_{d\in D}W_{dl}^{TI},\forall l\in L \tag{2.22}$$

Assure que la variable de décision $W_{dl}^{TI}$ va indiquer la taille sélectionnée

$$\sum_{a\in A}(p^{server}S_{la}^{explot}+p^{switch}N_{la}^{switch})=\sum_{d\in D}W_{dl}^{TI},\forall l\in L \tag{2.23}$$

$$W_{dl}^{TI}\leq M\,Y_{dl}^{size},\forall d\in D,l\in L$$

Détermine la valeur de l'efficacité énergétique

$$\alpha w_d Y_{dl}^{size}\leq\lambda\alpha w_{dl}^{TI},\forall d\in D,l\in L \tag{2.24}$$

Temps de réponse des services Internet

$$h_{sa}^{rep}X_{la}^{center}\leq k_{sa},\forall s\in S,l\in L,a\in A \tag{2.25}$$

Intégrité et non-négativité

$$Y_{dl}^{size}\in\{0,1\}\ \forall d\in D,l\in L \tag{2.26}$$

$$Y_{l}^{center}\in\{0,1\}\ \forall l\in L \tag{2.27}$$

$$X_{la}^{center}\in\{0,1\}\ \forall a\in A,l\in L \tag{2.28}$$

$$X_{la}^{backup}\in\{0,1\}\ \forall a\in A,l\in L \tag{2.29}$$

Le nombre des équipements TI (serveurs et commutateurs) doit être un entier. Dans ce cas, nous le considérons comme entier. Par contre, nous allons les relâcher dans le prochain chapitre, parce qu'avec des milliers de serveurs, les fractions deviennent négligeables.

$$S_{la}^{explot}\geq 0 \text{ et entier } \forall l\in L,a\in A \tag{2.30}$$

$$N_{la}^{switch}\geq 0 \text{ et entier } \forall l\in L,a\in A \tag{2.31}$$

$$S_{la}^{backup}\geq 0 \text{ et entier } \forall l\in L,a\in A \tag{2.32}$$

Comme nous l'avons déjà mentionné la formule pour calculer l'efficacité énergétique est présentée à (2.1). Cette contrainte est non-linéaire. Pour assurer que notre modèle mathématique est linéaire, nous rendons cette contrainte linéaire tel que présenté à (2.24).

## 2.4 Conclusion

Tout au long de ce chapitre, nous avons présenté la modélisation du problème de localisation des centres de données pour des services Internet.

Notre contribution a deux objectifs majeurs : 1) minimiser les coûts de mise en place des centres de données et 2) déterminer la taille des centres de données afin de répondre aux contraintes (2.23) et (2.24). Ces deux dernières contraintes visent, comme nous avons déjà mentionné, à déterminer l'énergie consommée par les équipements TI des centres de données de taille $d \in D$.

Également, le problème de localisation est un problème théorique bien connu. Cependant, nous avons mis l'accent sur les aspects algorithmiques et implications théoriques, dans l'absence de données de localisation réelles.

En outre, dans la phase de planification et dimensionnement des centres de données, il faut penser à la durée pour le renouvellement d'un centre de données et les serveurs

- 12 ans pour le renouvellement des centres de données.
- 4 ans pour les serveurs : il faut toujours penser au développement technologique.

Dans la suite, nous allons nous intéresser à la résolution de modèle proposé à l'aide d'une étude de cas.

# CHAPITRE 3

# EXPÉRIMENTATION, ANALYSE ET DISCUSSION

## 3.1 Introduction

Dans cette section, nous allons valider la solution proposée et décrite dans le précédent chapitre au moyen d'un prototype et une étude de cas. En effet, afin de résoudre le problème de localisation, nous l'avons approximé par une programmation linéaire. De plus, nous avons converti le problème en un problème de minimisation des coûts.

Le prototype que nous allons présenter est appliqué dans le contexte de la province de Québec. En effet, nous avons pris un échantillon formé par cinq régions de la province. En outre, le choix de cet échantillon est basé sur le nombre de population à desservir ainsi que les caractéristiques réseaux (à proximité des sources de fibre optique) et énergétique (à proximité des fournisseurs d'énergie).

## 3.2 Caractérisation des localisations

Afin de démontrer le lien direct entre les coûts des centres de données et les localisations, nous sélectionnons cinq régions de la province de Québec : Montréal, Québec, Laval, Gatineau et Trois-Rivière. Ces localisations répondent aux critères de sélection de localisations pour des centres de données.

### 3.2.1 Les critères de localisation

Au niveau de chapitre de la revue de littérature, nous avons présenté des critères de sélection des localisations. Parmi ces critères, nous sélectionnons quelques un qui correspondent le plus à notre problème de localisation.

De plus, les critères que nous sélectionnons sont :

- Le coût d'énergie : la province de Québec est caractérisée par un coût d'énergie à prix compétitif. En effet, selon Hydro-Québec, les centres de données suivent les tarifs de classe "L". Le détail de ce tarif est donné comme suit :

  **Structure du tarif** L [2]

  Prix de la puissance :

  - ☑ Puissance à facturer : 12,63 $/kW,
  - ☑ Par jour où il y a dépassement : 7,38 $/kW,
  - ☑ Maximum mensuel : 22,14 $/kW.

  Prix de l'énergie: 3,17 ¢/kWh.
- La proximité aux zones de marché.
- La température moyenne de la région : ce facteur a un impact direct sur le processus de refroidissement des centres de données et par conséquent, le coût de refroidissement.
- Le prix de fibre optique : un prix compétitif des fibres optiques avec le fournisseur « Fibre Noire » : ce coût est très important surtout que les centres de données que nous voulons localiser, offrent des services Internet.
- La proximité aux points d'accès de la fibre optique : la disponibilité des services Internet dépend directement de ce critère, surtout lors de transfert des données d'un centre à un autre.
- La population des régions.

## 3.3 Étude de cas

Une compagnie d'Internet "X" souhaite mettre en place un réseau des centres de données dans la province de Québec afin de desservir des zones de marché. Cette compagnie veut sélectionner des localisations parmi les régions de la province de Québec. La sélection répond à la fois au besoin de la compagnie et aux clients à desservir.

---

[2] Tarifs en vigueur le 1er avril 2014.
Source : Hydro-Québec

En effet, le besoin de l'entreprise est basé sur la minimisation des coûts totaux d'installation et mise en place des centres de données. Les clients des services Internet nécessitent une haute disponibilité de service et un temps de réponse minime.

La solution proposée est un programme linéaire en nombres mixtes qui permet de minimiser les coûts de mise en place d'un réseau de centre de données. Les centres de données localisés servent 13 zones de marché de la province de Québec.

### 3.3.1 Environnement de programmation

Afin de résoudre le problème proposé pour la localisation des centres de données pour des services Internet, nous faisons recours à *Lingo* version 12.0. En effet, *Lingo* est un outil (basé sur un langage script spécifique) permettant d'optimiser rapidement des modèles mathématiques de tous genres, développée par Lindo systems. *Lingo* est en mesure d'importer et d'exporter des tableaux de données sur des feuilles de calcul inhérentes à *Microsoft Excel*.

### 3.3.2 Localisation des centres de données

L'exécution du modèle mathématique donne des résultats illustrés par la figure 3.1

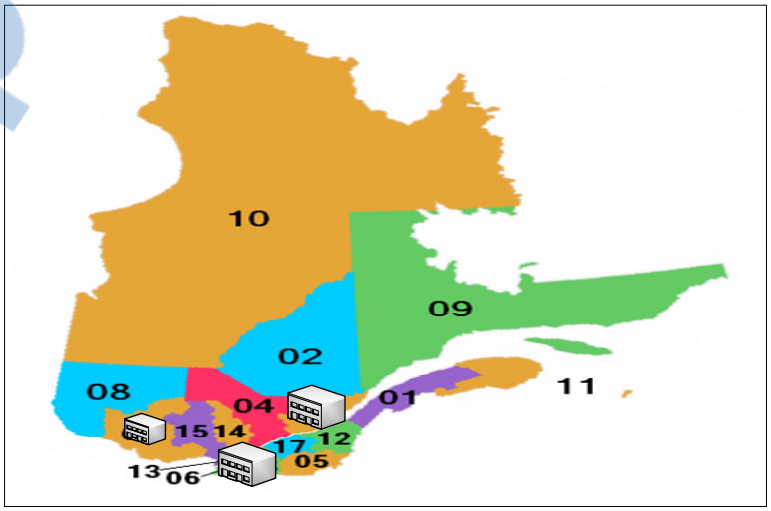


Figure 3.1 Résultat de localisation des centres de données

Rapport-Gratuit.com

La figure 3.1 présente la localisation des centres de données respectivement dans Montréal, Québec, Gatineau, Trois-Rivière et Terrebonne. En effet, les centres de données mis en place possèdent des tailles différentes. Les tailles sont respectivement d3, d4 et d5. La répartition des tailles est faite en fonction de la population (zone de marché) à desservir. Et par conséquent, en fonction de la demande (nombre de serveurs dans chaque centre de données).

Ces centres de données servent des zones de marchés de la province de Québec présenté dans la figure 3.2.

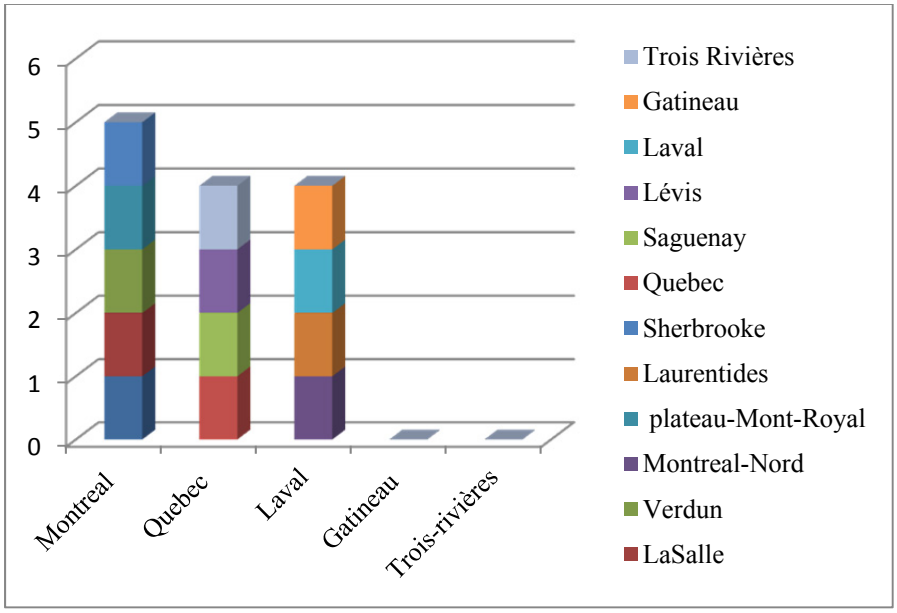


Figure 3.2 Les zones de marché desservies par les centres de données

La figure 3.2 indique que les centres de données localisés desservent les zones de marchés les plus proches afin de bien répondre aux besoins des clients.

Les dimensions des centres de données que nous fixons dès le départ sont présentées au tableau 3.1

Tableau 3.1 Les tailles des centres de données

| | d1 | d2 | d3 | d4 | d5 |
|---|---|---|---|---|---|
| Dimension (m²) | 2 600 | 2 200 | 1 700 | 1 200 | 800 |

### 3.3.3 Temps d'exécution des méthodes de résolution

Nous observons que le temps de réalisation de notre modèle à l'aide le *Lingo* est de 90 secondes. En effet, *Lingo* supporte des ensembles des commandes spéciaux (*Special Ordered Sets* SOS). Dans notre cas, nous utilisons les deux fonctions SOS1 et SOS3 respectivement pour les contraintes (2.9) et (2.10). Il définit un ensemble de variables binaires et impose des restrictions sur leurs valeurs collectives. Pour SOS1, au plus, une seule variable appartenant à un ensemble SOS1 sera supérieure à 0. Pour SOS3, exactement une variable d'un ensemble SOS3 donnée sera égale à 1.

Suite à l'ajout de ces commandes spéciales SOS1 et SOS3, le temps de résolution a baissé de 90 secondes à 11 secondes. Tous les détails sont présentés dans le tableau 3.2.

Tableau 3.2 Temps d'exécution de la solution proposée

| | mémoire (k) | Temps d'exécution (s) |
|---|---|---|
| MILP | 3537 | 90 |
| MILP<br>Avec SOS | 241 | 11 |

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

## 3.4 Minimiser les coûts

Le coût total minimisé de mise en place des centres de données dans trois localisations potentielles est 122 897 216 $. En effet, ce coût total est décortiqué en coût fixe, coût variable, coût d'énergie, coût des équipements TI et coût de la bande passante.
Mais avant de déterminer les valeurs de ces coûts, il faut mentionner que nous suivons cette décomposition afin d'insister sur l'importance de chaque coût. En effet, les coûts fixes de mise en place des centres de données dans 3 régions de la province de Québec (Montréal, Québec, Laval). Le coût fixe est la somme des coûts de terrain et de refroidissement, sachant que le coût de refroidissement est calculé comme suit : coût de refroidissement = coût d'énergie (selon Tarif L)$\times$quantité d'énergie entrante pour le refroidissement.

La quantité d'énergie pour le refroidissement dépend de la température moyenne de la localisation. Ainsi, elle dépend de la taille du centre de données. Le tableau 3.2 présente les valeurs d'énergie pour le refroidissement.

Tableau 3.3 La quantité d'énergie pour le refroidissement

| | **Montréal** | **Québec** | **Laval** | **Gatineau** | **Trois-Rivière** |
|---|---|---|---|---|---|
| **d1** | 18 000 | 17 280 | 16 560 | 15 480 | 15 840 |
| **d2** | 16 560 | 16 560 | 15 840 | 15 120 | 15 120 |
| **d3** | 15 480 | 16 200 | 15 120 | 14 400 | 14 760 |
| **d4** | 14 760 | 14 436 | 14 400 | 12 960 | 13 320 |
| **d5** | 14 400 | 14 040 | 13 680 | 12 240 | 12 960 |

Les valeurs des coûts fixes sont réparties comme présentée à la figure 3.3

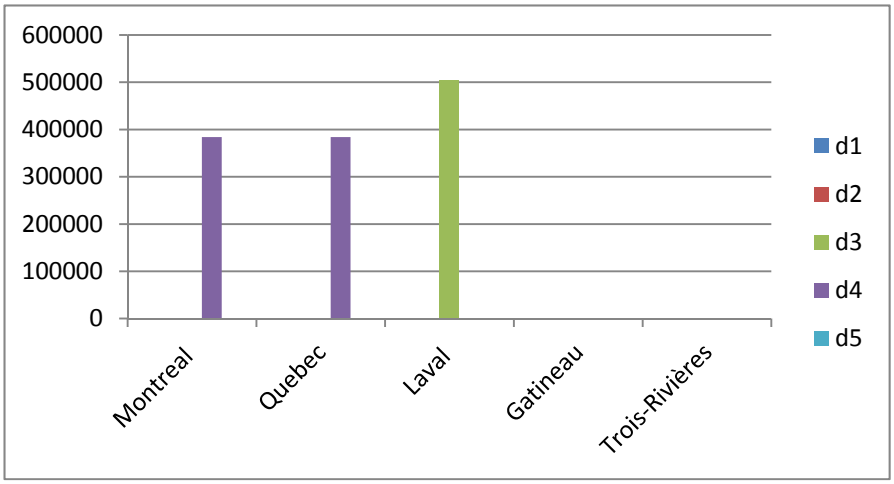


Figure 3.3 Les coûts fixes des centres de données

Les coûts variables de mise en place des centres de données dans cinq régions de la province de Québec sont la somme des coûts de maintenance et les coûts d'administration des centres de données. Les résultats d'exécution de modèle mathématique, présenté dans la section II, sont présentés à la figure 3.4

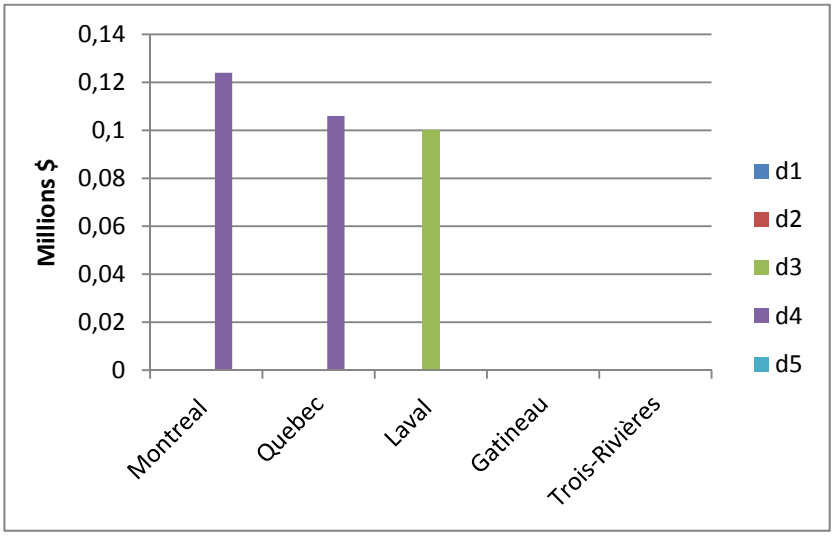


Figure 3.4 Les coûts variables des centres de données

Les centres de données sont caractérisés par une grande consommation d'énergie. En effet, lorsqu'une entreprise décide d'installer un centre de données, la première variable de décision est le coût d'énergie. Dans notre cas, la position stratégique de la province de Québec offre des prix compétitifs d'énergie. Par conséquent, une diminution des coûts d'installation et mise en place des centres de données.

La figure 3.5 illustre les coûts d'énergie pour alimenter les centres de données :

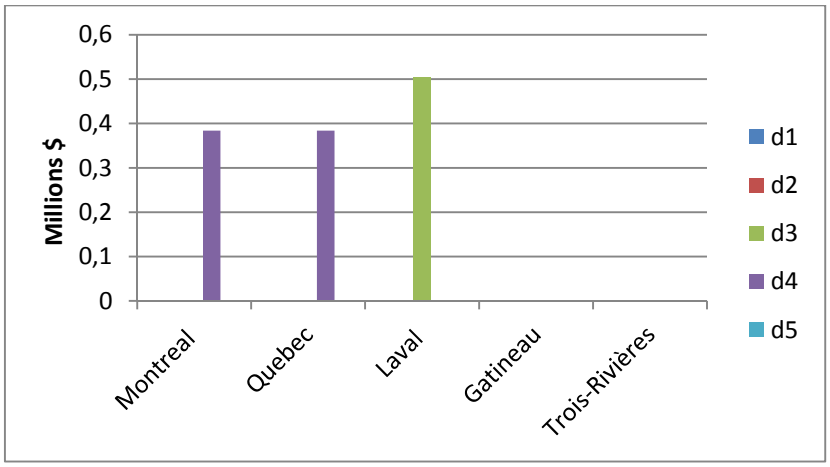


Figure 3.5 Les coûts d'énergie des centres de données

Selon Hydro-Québec, le tarif L s'applique au contrat annuel d'un client dont la puissance à facturer minimale est de 5 000 kilowatts ou plus et qui est lié principalement à une activité industrielle.

De plus, l'énergie totale consommée par un centre de donnée est répartie comme suit :

- Énergie entrante pour alimenter un centre de données.
- L'énergie consommée par les équipements TI.
- L'énergie pour l'éclairage.
- Énergie pour le refroidissement (district de l'eau glacée : pour éliminer l'excès de chaleur).

La répartition des coûts de  mise en place des centres de données est donnée par la figure 3.6

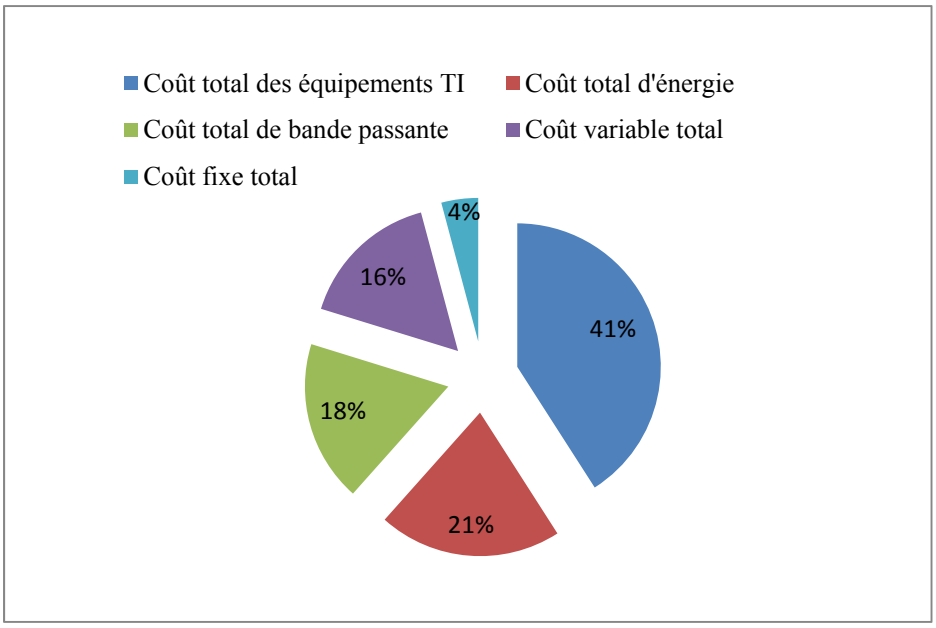


Figure 3.6  Répartition des coûts des centres de données localisés

Le coût d'énergie présente 21 % de coût annuel de centre de données. Ce coût présente la deuxième proportion des coûts après le coût des équipements TI. En effet, les équipements TI de nos centres de données localisés sont formés par les serveurs et les commutateurs. En outre, le nombre de serveurs satisfait la demande des clients afin de fournir les services Internet. À travers cette valeur, nous pouvons déterminer la demande pour chaque zone de marché.

D’après la figure 3.6, le coût des équipements TI présente 41 % de coût total d'un centre de données. Ce coût dépend directement du nombre de serveurs et commutateurs dans chaque centre de données. La figure 3.7 présente le nombre de serveurs d'exploitation affecté pour chaque zone de marché.

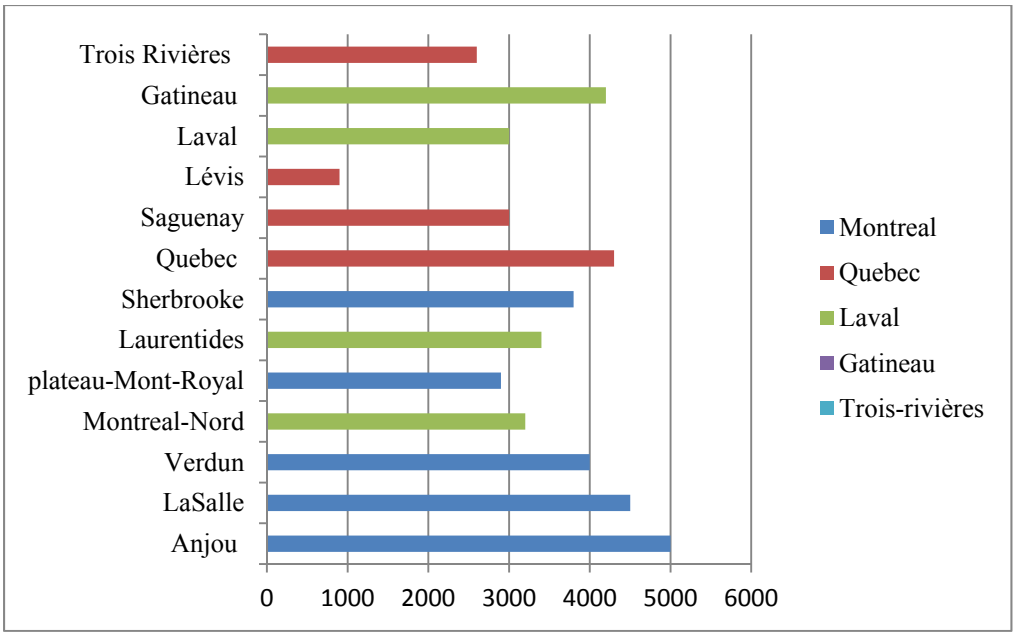


Figure 3.7 Le nombre de serveurs pour chaque zone de marché

Nous voulons mentionner que la contrainte d'intégrité de nombre de serveurs est relâchée. Tout simplement, parce que nous travaillons sur un grand nombre de serveurs (des milliers de serveurs). Alors, le nombre de serveurs et la même chose pour les commutateurs (qui forme les équipements TI de nos centres de données) sont des variables continues.

## 3.5 Les serveurs de secours "backup"

Nous indiquons précédemment que chaque centre de données présente un miroir de l'autre centre de données. Afin d'assurer la continuité et la disponibilité des services Internet à offrir par les centres de données localisés, nous admettons une solution de secours. En effet, cette solution indique que chaque centre de données possède des serveurs de secours afin d'assurer la continuité en cas de problème. De plus, nous admettons dès le départ que tous les serveurs des centres de données ont les mêmes caractéristiques afin d'éviter les problèmes de compatibilité lors de la migration des données d'un centre vers un autre. Dans le chapitre 2, section 2.4.3, nous décrivons la notion de délai de latence. En effet, ce délai sera calculé en cas de discontinuité des services d'un centre de données. Le délai de latence réfère au temps requis pour atteindre le centre de données de secours.

Dans ce cas, le paramètre $h_{ll'}$ est calculé en fonction de la distance qui sépare un centre de données actif et son centre de secours.

Le tableau 3.4 présente les valeurs des délais de latence pour chaque localisation possible.

Tableau 3.4 Délai de latence en secondes

| | **Montréal** | **Québec** | Laval | **Gatineau** | **Trois-Rivières** |
|---|---|---|---|---|---|
| **Montréal** | 0 | 20 | 0,5 | 15 | 10 |
| **Québec** | 20 | 0 | 25 | 40 | 6,5 |
| **Laval** | 0,5 | 25 | 0 | 12 | 10,5 |
| **Gatineau** | 15 | 40 | 12 | 0 | 30 |
| **Trois-Rivières** | 10 | 6,5 | 10,5 | 30 | 0 |

La figure 3.8 indique le nombre total des serveurs de secours "backup" qui dessert chaque zone de marché.

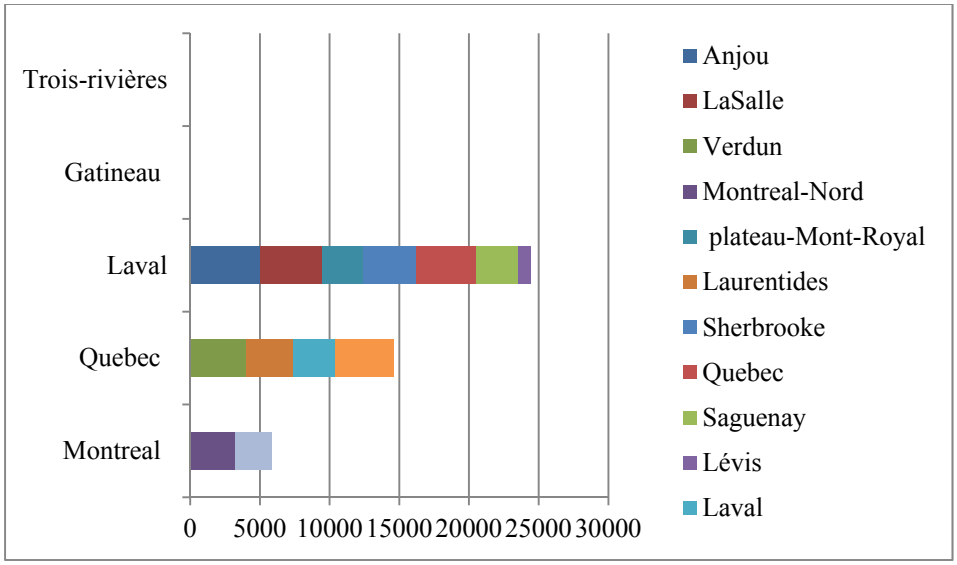


Figure 3.8 Nombre total des serveurs de secours " backup

## 3.6 Indicateur d'efficacité énergétique

Nous calculons ces valeurs afin de déterminer la valeur de PUE (voir équation 2.1)

Également, le facteur de pondération varie pour chaque type d'énergie.

- Facteur de pondération (Énergie électrique)=1.0
- Facteur de pondération (Énergie hydraulique)= 0.31

En plus, pour calculer le PUE, il faut choisir une catégorie parmi quatre catégories. Pour plus de détail, voir annexe II.

Dans notre cas, nous choisissions la catégorie 3 avec :

- ☑ L'énergie consommée par les équipements TI dépend du nombre des équipements TI (serveurs et commutateurs).

Le problème que nous définissons est un problème combinatoire. En effet, il permet de sélectionner une taille pour un centre de données et à la fois il faut prendre en considération le nombre de serveurs afin de bien gérer l'énergie consommée. De plus, l'ouverture d'un nouveau centre de données est directement liée à la demande de nos clients, ce qui rend le modèle plus complexe.

Également, la bonne gestion de taille de centre de donnée est liée à la consommation d'énergie des équipements TI (serveurs et commutateurs), ce qui nous amène à calculer l'efficacité énergétique de chaque centre de données ouvert. La contrainte (2.24) permet de déterminer les valeurs de PUE pour chaque centre localisé en fonction de taille et localisation.

Selon Uptime Institute (2012), la valeur moyenne de PUE est 1.7. Donc, nous prenons cette valeur comme valeur maximale de l'efficacité énergétique.

Les résultats d'exécution sont présentés à la figure 3.9. En comparant ces valeurs avec ceux proposés par Uptime Institute (2012), nous trouvons que nos centres de données répondent

aux exigences énergétiques. En effet, la localisation des centres de données dans la province de Québec montre un effet direct sur les valeurs d'efficacité énergétique.

Les valeurs de PUE pour chaque centre de données localisé sont bien illustrées dans la figure 3.9. Cette valeur dépend de la température moyenne de localisation et de nombre total des serveurs en cours de fonctionnement. À noter que les serveurs de secours ne sont pas pris en considération lors du calcul de PUE. Les centres de données localisés possèdent différentes tailles et par conséquent un nombre différent de serveurs. Alors, les valeurs des PUE sont respectivement, 1.3, 1.3 et 1.34.

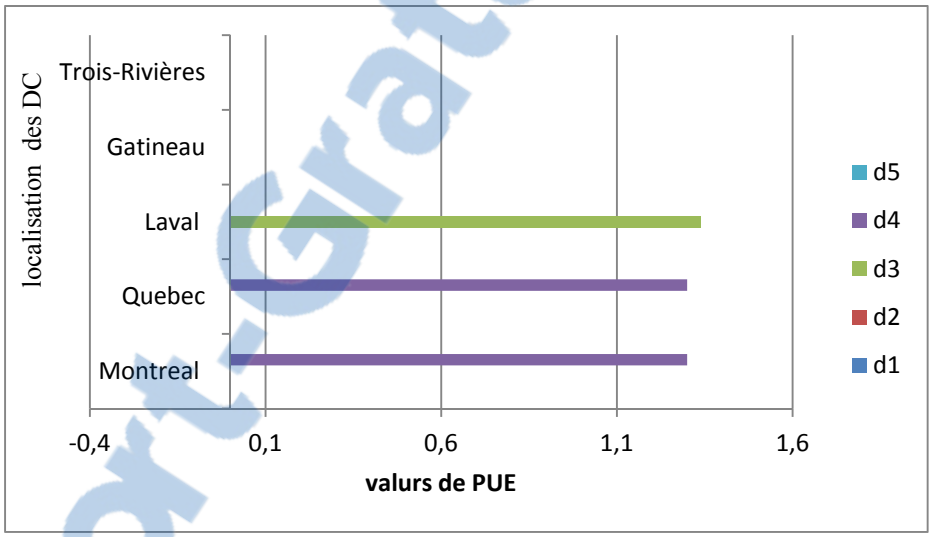


Figure 3.9 Valeurs des PUE pour chaque centre de données

Ces valeurs indiquent que les serveurs hébergés sont distribués en fonction de taille des centres de données et la population à desservir. En nous basant sur ces valeurs, nous constatons que les centres de données consomment la bonne quantité d'énergie. De plus, la contrainte (2.22) assure la non-ouverture des centres de données qui ne seront plus utiles même ils sont près des zones de marché.

## 3.7 Temps de réponse des services Internet

Nous choisissons de mettre en place des centres de données qui offrent trois types de services Internet : Service Web, service de traitement informatique et service de communication.

Rapport-Gratuit.com

En effet, le temps de réponse varie d'un service à un autre. Également, le paramètre de temps de réponse est calculé en fonction de la demande de population et en fonction de nature de service à offrir. En effet, le temps de réponse augmente si la population augmente.
Voici un exemple de cette variation en fonction de nombre de clients qui peuvent utiliser le service demandé. Un exemple de variation est illustré par la figure 3.10

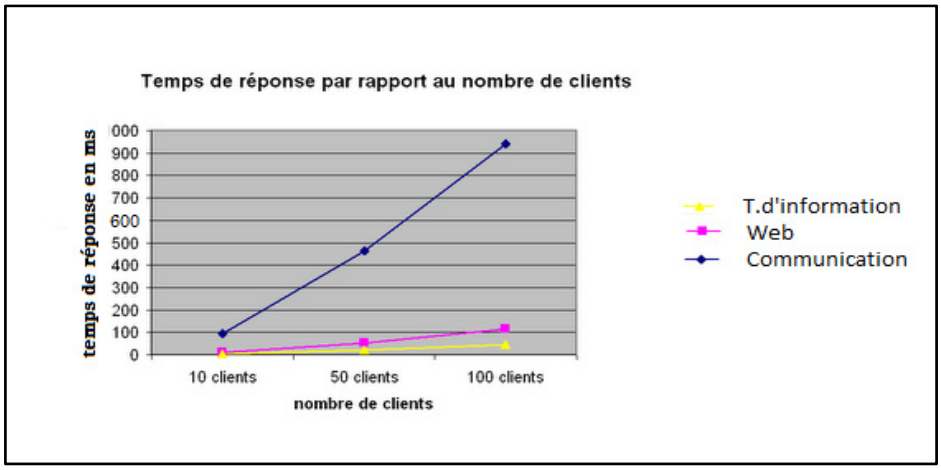


Figure 3.10 Temps de réponse par rapport au nombre de clients

La figure 3.11 illustre cette variation pour chaque localisation possible d'un centre de données

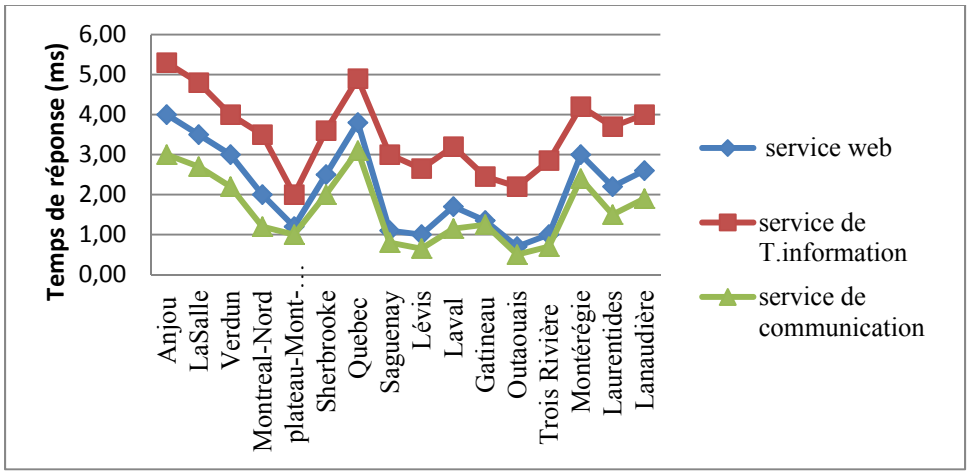


Figure 3.11 Variation de temps de réponse en fonction de service Internet

## 3.8 Évaluation de l'approche proposée

Notre contribution se base sur la mise en place des centres de données dans 8 régions de la province de Québec. Le modèle mathématique que nous proposons permet de sélectionner les localisations qui répondent aux contraintes déterminées. Dans ce contexte, BELL (2012) indique que les responsables des centres de données doivent tenir compte des critères basiques tels que le coût d'énergie, le climat, les taxes, etc.

En outre, pour assurer un accès ininterrompu aux données, BELL (2012) indique qu'il faut investir dans la redondance et la disponibilité. En effet, notre approche vise à respecter la contrainte de disponibilité des services tout en localisant des centres de données qui prennent en considération le temps de réponse des services Internet pour chaque zone de marché à desservir. Également, nous admettons une contrainte liée au délai de latence en cas d'inaccessibilité ou non-fonctionnement d'un centre de données. Dans ce cas, chaque centre de données représente un miroir pour l'autre centre de données. Mais, la redondance n'est pas appliquée pour notre approche, puisque nous admettons le niveau 1 de redondance dont la disponibilité selon Uptime Institute (2012) est de 99.671 % et dont le temps d'arrêt est d'environ 29 h par année.

Dans le même domaine de recherche, nous trouvons le projet de Goiri et al. (2011) qui présente une solution de localisation des centres de données pour des services Internet. Cette solution indique que le coût total de mise en place de quatre centres de données est de valeur de 12.5 millions de dollars. Les centres de données sont caractérisés par une taille de 1421 m² et 16 000 serveurs.

Alors que notre solution, permet de mettre en place trois centres de données de taille d3 = 1 700 m² et d4 = 1 200 m². Le nombre de serveurs pour chaque centre de données localisé est indiqué dans le tableau 3.5

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

Tableau 3.5 Nombre de serveurs pour chaque centre de données localisé

| | |
|---|---|
| **Montréal** | **20200** |
| **Québec** | **10800** |
| **Laval** | **13800** |
| **Gatineau** | **0** |
| **Trois-rivières** | **0** |

Le coût total de mise en place de trois centres de données avec ces caractéristiques est 10.41 millions de dollars par mois.

Tout en comparant nos résultats avec ceux présentés par Goiri et al. (2011), nous constatons que la mise en place des centres de données dans la province de Québec est moins couteuse qu'aux États-Unis.

## 3.9 Conclusion

Tout au long de ce chapitre, nous avons présenté en détail les résultats de l'approche proposée. En effet, la sélection de la province de Québec comme une localisation des centres de données est un défi. En effet, les prix compétitifs et chers de l'énergie électrique engendrent une minimisation des coûts surtout que ce dernier représente 23 % de coût total de mise en place d'un centre de données.

# CONCLUSION

Les centres de données jouent un rôle-clé dans la gestion, la qualité, le stockage et le partage des données. Afin d'assurer convenablement ces fonctionnalités, il faut toujours penser à une meilleure sélection des emplacements pour mettre en place les centres de données. Surtout que la localisation a un impact direct sur les coûts fixes, les coûts variables y compris particulièrement le coût d'énergie.

Évidemment, les grandes entreprises d'Internet, telles que Microsoft® et Google® ont leurs propres processus de sélection des emplacements pour leurs centres de données. Malheureusement, leurs processus sont hautement confidentiels pour des fins commerciales. L'un des objectifs de notre travail est de mettre l'accent sur ce problème et permettre à la communauté de recherche de traiter cette problématique dans le domaine public et rendre le processus de sélection des localisations plus actif.

Ce mémoire présente dans une première partie les critères de sélection de localisation. À ce niveau, le projet suscite une question : la province de Québec, avec ses températures froides, sa proximité aux marchés des États-Unis, la présence des liaisons de télécommunication à fibre optique (fibre noire), la disponibilité d'une main-d'œuvre qualifiée et la proximité des centrales d'hydroélectriques, est-elle convenable et propice à la mise en place des centres de données. Tout en appliquant notre projet pour la province de Québec, nous constatons que les résultats obtenus sont proches à la réalité et aux résultats discutés dans la revue de littérature.

Dans une deuxième partie, nous modélisons le problème de sélection des localisations tout en définissant des contraintes pour réponde aux besoins des clients des services Internet. Ces contraintes sont : le temps de réponse des services pour des zones de marchés à desservir, le délai de latence lors de transfert des données d'un centre vers un autre et l'efficacité énergétique (PUE) qui permet d'assurer une meilleure gestion des équipements TI.

## RECOMMANDATIONS

Dans les travaux futurs, quelques recommandations peuvent être apportées à l'approche proposée dans ce mémoire dans le cadre d'amélioration :

- L'application de la présente approche sur un cas d'étude réel et sur des données de localisation réelles. En effet, la qualité des résultats retournés par la présente approche dépend principalement de la qualité et l'exactitude des données en question. De plus, l'emplacement des installations est un problème théorique bien connu. Cependant, les traitements de ce problème ont porté sur ses aspects algorithmiques et implications théoriques, dans l'absence de données de localisation réelles.

- L'ajout du critère de sélection de niveaux de redondance adéquats pour chaque centre de données. En effet, le niveau de redondance joue un rôle très important afin d'assurer la haute disponibilité et la continuité des centres de données. L'importance de ce critère varie en fonction des domaines d'application des centres des données. Pour le domaine de technologie d'information et de communication (TIC), tel que note cas, le choix d'un niveau de redondance est basique.

- L'utilisation de l'énergie renouvelable et surtout l'énergie hydraulique pour le refroidissement afin de réduire le coût total d'énergie.

- L'utilisation des serveurs « verts » afin de réduire l'énergie consommée par les équipements TI et par conséquent augmenter l'efficacité énergétique de centre de données.

- L'application de notion de virtualisation des serveurs afin de réduire les coûts des équipements TI. Également, avec la virtualisation des serveurs, nous pouvons réduire la taille de centre de données et par conséquent une réduction au niveau de coût des terrains.

# ANNEXE I

# DATACENTER LOCATION FOR INTERNET SERVICES IN THE PROVINCE OF QUEBEC

Hibat Allah OUNIFI[1, a], Mustapha OUHIMMOU[1, b,] Marc PAQUET[1, c]

And Julio MONTECINOS[1, d]

[1]École de technologie supérieure, 1100 Rue Notre-Dame Ouest, Montréal, QC H3C 1K3

[a] hibat-allah.ounifi.1@ens.etsmtl.ca, [b]Mustapha.ouhimmou@etsmtl.ca, [c]Marc.paquet@etsmtl.ca, [d]Julio.Montecinos@etsmtl.ca

**Keywords:** data centers, service response time, power usage effectiveness, latency delay, cost optimization.

**Abstract.** This paper introduces an efficient algorithm based on mixed integer linear programming (MILP) for the datacenter localization problem. The proposed model is used to locate datacenters under Internet services demands and, at the same time, ensure a low energy consumption of information technology (IT) equipment. Model validation and results are proved via experimentation example and numerical study.

## Introduction

In datacenter studies, there are some points that are important to evaluate. Those points are: datacenter topology, network connectivity, latency time, costs per megawatts (MW) and the use of clean energy. Daim, Bhatla et Mansour (2012) discussed two essential purposes of a datacenter: 1) ensure data security and 2) guarantee continuity of company's operations. Similarly, they claim that the security of the datacenter is manifested in two ways: 1) physical security of the datacenter from natural and human disasters and 2) business continuity of datacenter while planning centers recovery after disasters. In the latter case, we can speak about datacenter of recovery (backup).

Many companies such as IBM®, Amazon® and Microsoft® offer many services like database services and Web services to millions of users. These services are stored in datacenters that are distributed geographically. The main needs of datacenter user are high availability and low response time. To maintain these needs, we assume that each datacenter has as a backup to another datacenter at a distinct location. In addition, Covas, Silva et Dias (2013) indicate that companies want to build datacenters in locations that are not prone to natural disasters, terrorism and also offer less expensive resources such as electricity, network and easy access way to transport. At the same time, they need a robust system that can operate 24/7/365 without interference and provides scalable computing power. Similarly,

the response time, the availability and the costs depends strongly on the specific location of datacenters.

Today, site selection is also a key element in datacentre's network design problem. The location of datacenter presents the strategic setting to determine the total cost of ownership and, to a some extent, life span of a datacenter. In this paper, we study the Internet datacenter location in the province of Quebec. First of all, we propose a mixed integer linear programming (MILP) model to select datacenter locations and sizes. This model defines costs minimization objective with service response time, latency delay and power usage effectiveness (PUE) as main operational constraints.

This paper is organized as follows. Section 2 presents some related work and gives some examples about datacenter budgets and costs. Based on this literature review, we give an overview of datacenter location criteria and highlight the advantages of the province of Quebec as a potential location to setup datacenters. Section 3 explains in detail the mathematical model. Section 4 gives the numerical results. Finally, section 5 draws our conclusion and research perspectives.

## Literature Review

Several large companies in the world such as Google®, IBM®, Microsoft®, Dell® and Amazon® define common criteria for establishing a datacenter.

Before setting up a datacenter, all user needs should be met while ensuring service high availability, data security, good response time and a fast data transfer. These goals will be achieved while defining from the outset for the right location and size of datacenter. Similarly, the definition of criteria for the selection of good datacenter location differs according to business priorities. Covas, Silva et Dias (2013) present a multi-criteria decision analysis to identify the best location of datacenters, taking into account economic, social, technical and environmental dimensions. These dimensions will be evaluated as a problem of multi-criteria sorting. In another paper, Daim, Bhatla et Mansour (2012) define four factors to select datacenter locations: geographical factor, cost factor, political factor and social factor. Indeed, the choice of site location should consider, according to Sun Microsystems, natural hazards (storms, earthquakes), human risks (industrial pollution, vibration) and availability of public services. In addition, Rath (2007) presents many other criteria to choose the potential location of datacenters such as workforce qualification, land cost, green power, insurance... Also, Covas, Silva et Dias (2013); Lei et al. (2010) indicate that datacenters are characterized by high energy consumption. When a company decides to open a datacenter, the first decision variable is the energy cost. Similarly, data security and reliability of datacenters are connected to the location of datacenters. Thus, according to Covas, Silva et Dias (2013), the criteria of evaluation for the datacenter site selection have different units . In fact, in the environmental evaluation, energy saving is measured in terms of quantity of CO2 while in the economic evaluation, energy savings will be measured in financial units. In our case, we choose the geographical factor and cost factor by selecting the province of Quebec as a strategic position to setup datacenters. Thus, the province of Quebec has energy at competitive prices, renewable energy, and suitable climate for cooling, qualified workforce, robust telecommunications infrastructure and proximity to the US market.

Our aim is to define, first of all, site location criteria then, minimize the total costs (including fixed and variable costs). In this context, Goiri et al. (2011) present an optimization framework to choose the potential location of datacenters by respecting response time and service availability constraints. But, there are some gaps: first, they do not specify the nature of Internet services. For Internet services, there are many services such as Web service, mailing service, computer processing service, Video conference service... For each service, there are different response times. In addition, this response time depends on the demands of users. In addition, they do not calculate the power usage effectiveness (PUE) metric.

Indeed, energy consumption is one of the most important criteria when companies plan to build datacenter. In this case, Raghavendra et al. (2008); Yao et al. (2012) Lei et al. (2011); Yuanxiong et al. (2011) present a power management solutions in order to decrease the energy consumption of server and electricity costs. In this context, Uptime Institute (2012) mentions that the PUE is a metric used to define energy efficiency of datacenters. Therefore, Abbasi, Varsamopoulos et Gupta (2012) show that calculating the power usage effectiveness (PUE) is important to define the energy used by IT equipment of datacenter.

In addition, Hong et Baochun (2012) develop an optimization system that takes into consideration the use of bandwidth and the cost of electricity to solve the problems of datacenters selection for cloud computing services. Indeed, the method used in Hong et Baochun (2012) is the multiplexing of bandwidth in order to reduce the consumption of bandwidth and cost. Based on these related works, we follow an approach which combines two approaches: 1) minimize the total costs of datacenter setup and 2) minimize the energy consumption. For the latter, especially the IT equipment energy such as servers by selecting the right size of datacenter and assigning a specific number of hosted servers for each area zone (population center). With the increase demand of Internet services in recent years, the energy consumed by datacenters, which is related directly to the number of hosted servers, experiencing remarkable increase. In the same context, Raghavendra et al. (2008) indicates that the total datacenter power and cooling energy costs is over $30 billion in year 2008, in the world. In another context, Van Heddeghem et al. (2014) indicate that the total worldwide electricity consumption has increased from 3.9% in 2007 to 4.6% in 2012.

To achieve the objectives stated previously, we should respect the Internet services response time and latency delay constraints. Our second contribution is the calculation of the energy efficiency (PUE) for each datacenter and in each potential location which makes our problem a complex combinatory problem.

## Model Formulation

**Model Description.** In this section, we summarize the model to capture the behaviour of datacenter location for Internet services in the province of Quebec. It describes the mixed integer linear programming (MILP) model that considers the critical aspects for the design and selection of datacenter locations and sizes while minimizing the total costs of setup, and determines the energy efficiency (PUE) of each datacenter. Indeed, companies take into account the value of the PUE while selecting locations of datacenters. In this project, we chose, in advance, five regions on the province of Quebec to be the possible location to setup

Rapport-Gratuit.com

datacenters. These regions are Montréal, Québec, Laval, Trois-Rivières and Gatineau. Also, we selected the area zones to be served later by located datacenters: Anjou, LaSalle, Verdun, Montréal-Nord, Plateau-Mont-Royal, Les Laurentides, Sherbrooke, Québec, Gatineau, Laval, Saguenay, Lévis and Trois-Rivièrse. Those area zones are characterized by high population density (the largest cities in the province of Quebec.

*Service response time*: As previously mentioned, Internet services demands are increased quickly. To maintain the high availability of services, we should take into account, in the mathematical model, the service response time. In this case, we define $h_{sa}^{rep}$ parameter which presents the response time of a service $s$ to an area zone (population center) $a$. In more detail, for each Internet service, it has a different response time. Indeed, this parameter depends on the type of service (e.g., Web service needs less response time than the communication service).

*Latency delay:* One of the most important objectives of Internet datacenter is to maintain the service availability for potential user. In this case, we think to the latency delay parameter that is defined when a datacenter becomes unreachable or unavailable. That means that the active datacenter become a center of backup for unavailable datacenters. The $h_{ll'}$ parameter is calculated as a function of distance from the unavailable or unreachable datacenter and a backup datacenter. This parameter should be less or equal to the maximum latency delay value.

*Power usage effectiveness (PUE):* As mentioned above, the Internet datacenters have a high consumption of energy. With the increase of service demands, the numbers of hosted services will increase and consequently the energy consumed by servers. Thus, in selecting location, it is important to calculate the PUE value. In our mathematical model, we define the PUE as constraint to help us to choose the good size of datacenter and affect later, for each population center the required number of hosted servers and backup servers too. The value of PUE is calculated as follow:

$$\frac{electricity \times weighting\ factor + DistrictChilledWater \times weighting\ factor}{IT\ Energy}$$

**Sets and Indices.** In this mathematical model, the following sets and indices are used:

$l \in L$ Datacenter location.

$a \in A$ Population center.

$d \in D$ Size of datacenter.

$s \in S$ Provided Internet services.

$r \in R$ Set of resources that each server has, which include 1 = CPU, 2 =

Network card, 3=RAM, 4 = Storage,

5 = Fiber Optic card.

**Parameters.** To formulate the model, the following costs are required:

$c_{dl}^{maint}$ Maintenance cost of datacenter has a size $d \in D$.

$c_{r}^{server}$ Server cost with resources $r \in R$.

$c_{d}^{admin}$ Administration cost of a datacenter located in and has a size $d \in D$ ($/administrator).

$c_{d}^{erergy}$ Energy cost for each datacenter has a size $d \in D$.

$c_{dl}^{land}$ Land cost for a datacenter in the location $l \in L$ and has a size $d \in D$.

$c_{dl}^{cooling}$ Cooling cost for a datacenter in the location $l \in L$ and has a size $d \in D$.

$c^{BW}$ Cost of required bandwidth for each server providing a service $s \in S$.

$c^{switch}$ Switch cost ($/switch).

$c^{FiberOptic}$ Fiber-optic cost ($/ Km).

The following data are also required:

$d_{la}^{location}$ Distance between location $l \in L$ and population center $a \in A$.

$n_d$ The number of administrators required to manage datacenter has a size $d \in D$.

$b_s$ Bandwidth required for each server has a service $s \in S$.

$w_d$ The amount of energy required to power a datacenter has a size $d \in D$.

$h_{ll'}$ Latency between two potential locations $l \in L$ and $l' \in L$.

$h_{sa}^{rep}$ Response time of service $s \in S$ for a population center $a \in A$.

$f_{d}^{\max}$ Maximum capacity of a datacenter in number of IT equipment (servers, Switches) for a datacenter, which has a size $d \in D$.

$\delta_a$ Number of hosted servers to serve a population center $a \in A$.

$k_{sa}$ Maximum response time of a service $s \in S$ for a population center $a \in A$.

$p^{switch}$ Incoming power to serve a switch.

$p^{server}$ Incoming power to serve a server.

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

$m$ Maximum latency delay.

$\beta$ Weighting factor for water energy.

$\lambda$ Maximum power usage effectiveness (PUE).

$\alpha$ Weighting factor for electric energy.

$\mu$ Switch incoming ports for servers.

**Decision variables.** In order to find the optimal datacenter location in the province the Quebec, the following decision variables are required:

$Y_{l}^{center}$ Binary variable equals to 1 if a datacenter is located in $l \in L$; 0 otherwise.

$X_{la}^{backup}$ Binary variable equals to 1 if a backup datacenter is located in $l \in L$ to serve the population center $a \in A$; 0 otherwise.

$S_{la}^{backup}$ The total number of servers of a backup datacenter located in $l \in L$ and serves a population center $a \in A$.

$W_{dl}^{TI}$ The IT energy consumption for a datacenter of size $d \in D$ located in $l \in L$.

$Y_{dl}^{size}$ Binary variable equals to 1 if datacenter located in $l \in L$ has as size $d \in D$; 0 otherwise.

$X_{la}^{center}$ Binary variable equals to 1 if datacenter located in $l \in L$ serves the population center $a \in A$; 0 otherwise.

$S_{la}^{explot}$ The total number of servers for a datacenter located in $l \in L$ that serves the population center $a \in A$.

$N_{la}^{switch}$ The number of switches to connect servers for a datacenter located in $l \in L$ which serves the population center $a \in A$.

**Optimization model.** Using these notations, a mixed integer linear programming (MILP) model can be formulated to select Internet datacenter locations and sizes. The total datacenter costs are segmented in six costs. Fixed costs (3.1) are formed by land cost and cooling cost. Hence, variable costs (3.2) are formed by administration cost and maintenance cost. Information technology (IT) costs (3.3) that contain the server and the

switch costs. The energy cost (3.4) that depends on the total energy needed to power a datacenter which has as size $d$.In addition, the bandwidth cost (3.5) depends on the total number of servers in each datacenter location. Finally, the distance cost (3.6) is calculated in function of fiber-optic cost. The sum of the costs mentioned above constitutes the objective function of our problem.

The objective function to be minimized is given by:

$$C^{fixe} = \sum_{d \in D} \sum_{l \in L} \left( c_{dl}^{land} + c_{dl}^{cooling} \right) Y_{dl}^{size} \tag{3.1}$$

$$C^{\text{var}} = \sum_{l \in L} \sum_{d \in D} \left( n_d c_d^{admin} + c_d^{maint} \right) Y_{dl}^{size} \tag{3.2}$$

$$C^{TI} = \sum_{a \in A} \sum_{l \in L} c_1^{server} (S_{la}^{\exp lot} + S_{la}^{backup}) + c^{switch} N_{la}^{switch} \tag{3.3}$$

$$C^{energy} = \sum_{l \in L} \sum_{d \in D} w_d \, c_d^{energy} Y_{dl}^{size} \tag{3.4}$$

$$C^{BW} = \sum_{a \in A} \sum_{s \in S} \sum_{l \in L} b_s \, c^{BW} S_{la}^{\exp lot} \tag{3.5}$$

$$C^{distance} = c^{fiberOptic} \sum_{a \in A} \sum_{l \in L} d_{la}^{location} \, X_{la}^{center} \tag{3.6}$$

The objective-function (3.7) aims to minimize the costs mentioned above to select the potential location of a data center:

$$Min\ F1 = C^{energy} + C^{fixe} + C^{\text{var}} + C^{TI} + C^{BW} + C^{distance} \tag{3.7}$$

The objective-function must meet the following constraints

$$\sum_{l \in L} X_{la}^{center} \geq 1 \, , \forall a \in A \tag{3.8}$$

$$\sum_{d \in D} Y_{dl}^{size} \leq 1 \ \forall l \in L \tag{3.9}$$

$$X_{la}^{center} \leq \sum_{d \in D} Y_{dl}^{size} \, , \forall a \in A, l \in L \tag{3.10}$$

$$\sum_{l \in L} X_{la}^{backup} \geq 1, \forall a \in A \tag{3.11}$$

$$\sum_{d \in D} Y_{dl}^{size} = Y_{l}^{center} \, , \forall l \in L \tag{3.12}$$

$$\sum_{d\in D} Y_{dl}^{size} \geq X_{la}^{backup}, \forall a \in A, l \in L$$
$$Y_{l}^{center} \geq X_{la}^{backup}, \forall a \in A, l \in L \tag{3.13}$$

$$X_{la}^{center} + X_{la}^{backup} \leq 1, \forall a \in A, l \in L \tag{3.14}$$

$$\delta_a X_{la}^{center} = S_{la}^{explot}, \forall l \in L, a \in A \tag{3.15}$$

$$\delta_a X_{la}^{backup} \leq S_{la}^{backup}, \forall l \in L, a \in A \tag{3.17}$$

$$\sum_{a\in A} (S_{la}^{explot} + S_{la}^{backup} + N_{la}^{switch}) \leq \sum_{d\in D} f_d^{\max} Y_{dl}^{size}, \forall l \in L, a \in A \tag{3.18}$$

$$\mu * N_{la}^{switch} - 1 < S_{la}^{explot} + S_{la}^{backup}, \forall l \in L, a \in A \tag{3.19}$$
$$\mu * N_{la}^{switch} \geq S_{la}^{explot} + S_{la}^{backup}, \forall l \in L, a \in A$$

$$h_{ll'}(X_{la}^{center} + X_{la}^{backup} - 1) \leq m \ \ \forall l \in L, l' \in L, a \in A \tag{3.20}$$

$$\sum_{a\in A} (p^{server} S_{la}^{explot} + p^{switch} N_{la}^{switch}) = \sum_{d\in D} W_{dl}^{TI}, \forall l \in L \tag{3.21}$$

$$W_{dl}^{TI} \leq M Y_{dl}^{size}, \forall l \in L, d \in D \tag{3.22}$$

$$(\alpha w_d + \beta w_{dl}^{cooling}) Y_{dl}^{size} \leq \lambda(\alpha W_{dl}^{TI}), \forall d \in D, l \in L \tag{3.23}$$

$$h_{sa}^{rep} X_{la}^{center} \leq k_{sa} \ \forall a \in A, l \in L, s \in S \tag{3.24}$$

Binary variables:

$Y_{dl}^{size} \in \{0,1\} \ \forall d \in D, l \in L$

$Y_{l}^{center} \in \{0,1\} \ \forall l \in L$

$X_{la}^{center} \in \{0,1\} \ \forall a \in A, l \in L$

$X_{la}^{backup} \in \{0,1\} \ \forall a \in A, l \in L$

Integer and positive variables:

$S_{la}^{explot} \geq 0$ and Integer $\forall l \in L, a \in A$

$N_{la}^{switch} \geq 0$ and Integer $\forall l \in L, a \in A$

$S_{la}^{backup} \geq 0$ and Integer $\forall l \in L, a \in A$

Constraint (3.8) ensures that each datacenter serves at least one population center. Constraint (3.9) defines if datacenter is opened or not. Constraint (3.10) ensures that a datacenter must exist to serve a population center. Constraint (3.11) ensures that each population center is deserved by a backup datacenter. Constraint (3.12) is just an informative constraint that contains an information variable $Y_l^{center}$. Constraint (3.13) ensures that a backup datacenter exists to serve a population center; this backup datacenter has an identical characteristic than a datacenter. Constraint (3.14) indicates that the datacenter and its backup are not in the same location. Constraint (3.15) is the demand constraint that defines the total number of servers that can be defined to serve a population center. In addition, the constraint (3.16) ensures that the backup datacenter is equal to the biggest needs. Constraint (3.17) defines the total number of switches needed to interconnect servers. Then, constraint (3.18) is added to ensure proportionality between the number of servers and switches. Constraint (3.19) defines the maximum capacity of a datacenter in number of IT equipment (servers and switches). Constraint (3.20) defines the datacenter latency between two potential locations in order to determine the datacenter of backup. Constraint (3.21) calculates the energy consumed by IT equipment in order to calculate in the next step, the power usage effectiveness. In addition, the backup servers could be included, but they are in stand-by and no energy consumed. Constraint (3.22) ensures that $W_{dl}^{TI}$ will indicate the selected size, with $M$ as the maximum energy that can be consumed by IT equipment. Constraint (3.23) defines the power usage effectiveness, this constraint is a way to impose a minimum use of resources over the size, so it increases artificially the number of IT equipment. Finally, constraint (3.24) defines the service response time that should be less than a maximum response time.

**Experimentation and Analysis**

We present a solution method for the problem described in section 3. In order to solve the problem, we approximate the problem by a linear programming formulation. Then, we convert the problem to a minimization cost problem. The objective of this section is to validate the model and determine the numerical results. The integer programming model was tested with a real scenario and real data. In addition, throughout our project, we assume two assumptions. First, we choose to setup datacenters that offer three different types of Internet services: Web service (Web server), processing information service (FTP Server) and communication service (Server of videoconferencing). Indeed, service availability varies from one service to another. Thus, in the second hypothesis, we assign a 10-ports switch to interconnect servers with eight incoming ports for servers and 2-ports for Ethernet connection. The described model in section 3 is solved by using Lingo optimizer version 12.0. The optimal total cost is 122 897 216 M$. This cost is the result of the setup of three datacenters with different sizes: the first one is located in Montréal, the second one in Québec and the third one in Laval. Each datacenter serves area zones (population centers). The solution is detailed in Fig.1.

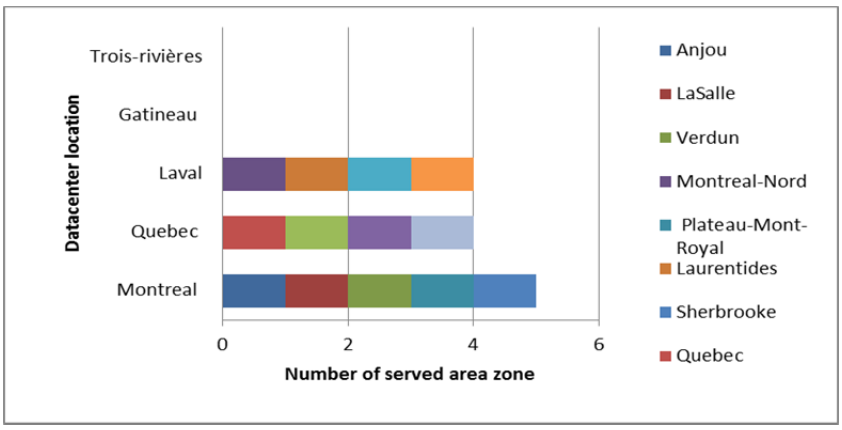


Figure-A I- 1 Datacenter location and assigned area zones

This figure indicates that three opened datacenters serve 13 populations centers (e.g., Montréal serves Anjou, LaSalle, Verdun, Plateau-Mont-Royal and Sherbrooke). The percentages of the total costs of three datacenters setup are distributed as follows and presented in Fig. 2.

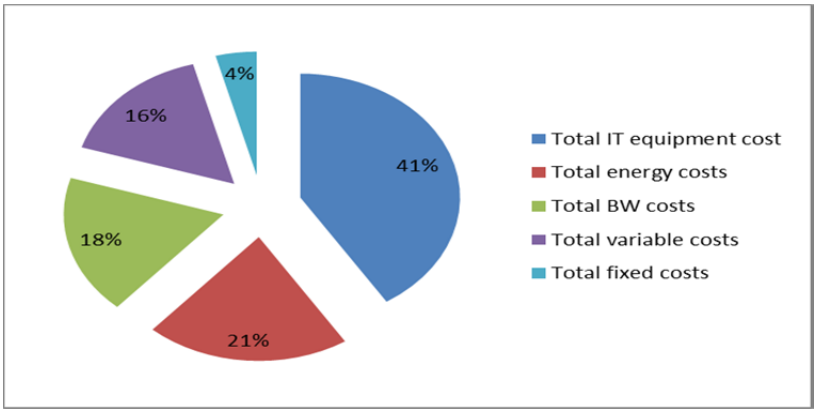


Figure-A I- 2 Datacenter costs for the optimal solution to F1

Indeed, all costs assume a 12-year lifetime for datacenters and a 4-year lifetime for servers.

As indicated in the above figure that the total energy cost has 21% of the total costs. In fact, when a company decides to locate a datacenter, the first decision variable is the cost of energy. In our case, the strategic position of the province of Quebec offers competitive energy prices. Therefore, a lower cost of installation and implementation. The energy cost is depending on the total number of IT equipment and especially the number of servers affected

for each area zone. In this case, we present in Fig.3 the total number of servers assigned for each area zone.

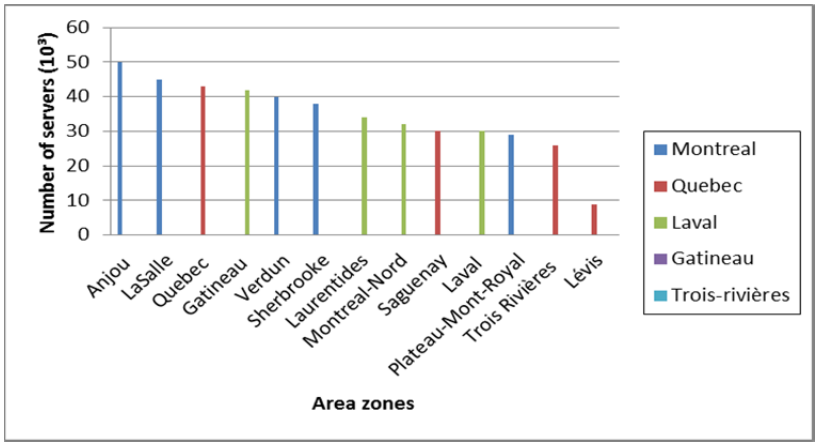


Figure-A I- 3 Total servers numbers affected for each zone

Throughout this project, we assume that the redundancy level of our datacenter is Level I does not require that redundancy in number of IT equipments. To resolve the problem of breakdown of datacenter, we define a latency constraint that is detailed in constraint (20). If such problem occurs, a mirror datacenter takes over the service in order to maintain high availability of services. Also, the backup capacity is shared between datacenters. In addition, constraints (21) and (22) define the total energy consumed by IT equipment, thus, we calculate the power usage effectiveness (PUE) of located datacenters. The PUE value requires a minimum number of servers in use. So, we don't take into account the total number of servers of backup. In fact, these constraints force the opening of the datacenters only if the datacenters have sufficient equipment, while some centers take over hardware required for the operation and backup. This excess is contained in *variable* $S_{la}^{backup}$ . Fig. 4 presents in details the total number of server in datacenters of backup and the repartition for population center.

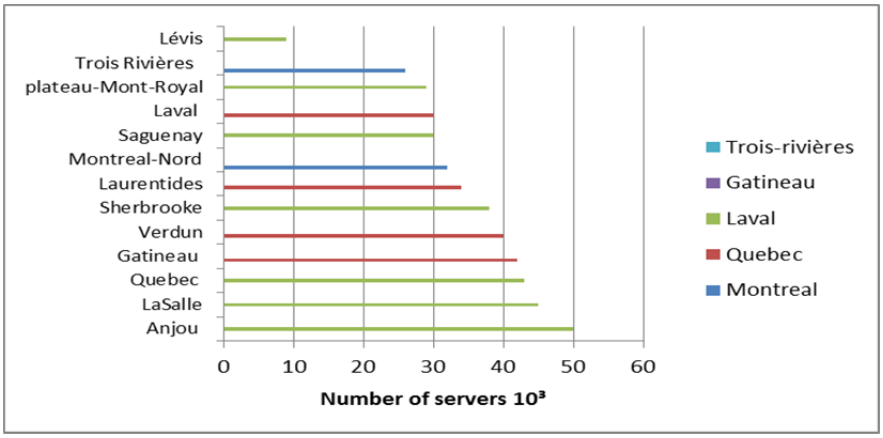


Figure-A I- 4 Total servers' numbers affected for each area zone in datacenters of backup

As it is calculated in constraint (22), the PUE values are given in the Fig.5. This value depends on location temperature and the total number of functional servers in each datacenter. As the located datacenters have different sizes (size d4= 1200 m² and size d3= 1700 m²), we have as it is indicated on Fig.3 and Fig.4, different number of servers in datacenters and those used as backup. Therefore, the power usage effectiveness (PUE) values are respectively 1.3, 1.3 and 1.34.

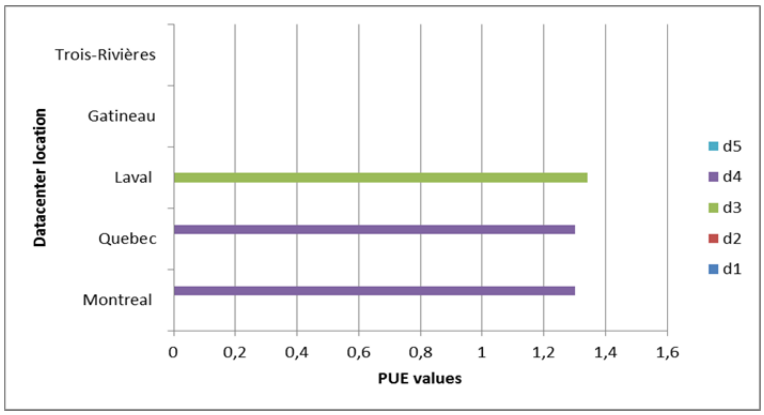


Figure-A I- 5 Power usage effectiveness (PUE) values

These PUE values indicate that the numbers of hosted server are distributed in function of datacenter sizes and the population center to serve. Based on these values, we can state that the opened datacenters consume the right amount of energy. Thus, the constraint (22) indicated that we cannot open a datacenter if it we do not need it. In addition, this constraint ensures to select just one datacenter size. In addition, when our model selects datacenter

sizes, it should take into account the total number of servers for not using more energy. In this case, the model is highly combinatorial.

Finally, we observe that the solution runtime take 90 seconds. Lingo supports Special Ordered Sets (SOS). In this case, we use the SOS1 and SOS3 functions respectively in constraints (9) and (10). It defines a set of binary variables and places restrictions on their collective values. For SOS1, at most, only one variable belonging to an SOS1 set will be greater than 0. For SOS3, exactly one variable from a given SOS3 set will be equal to 1.

As result, the elapsed runtime of our solution is just 11 seconds. All details are presented in table 1.

Table1. Running time of solutions approach

| | Generator memory (k) | Runtime (s) |
|---|---|---|
| Pure MILP | 3537 | 90 |
| MILP with SOS sets | 241 | 11 |

**Conclusions and Perspectives**

Throughout this paper, we presented a location model in a MILP framework. This model aims to minimize the datacenter costs including variable costs, fixed costs and energy costs. This objective is obtained by respecting three constraints: service response time, datacenter latency time and the power usage effectiveness metric. By calculating the energy consumption by IT equipments, we determined the PUE values for each location in order to verify the energetic efficiency of located datacenters.

As future work, we will apply our mathematical model for a real life size example. Also, we will consider minimizing and increasing energy efficiency by integrating “green” servers.

**References**

[1] T. U. Daim, A. Bhatla, and M. Mansour, "Site selection for a data centre - a multi-criteria decision-making model," *International Journal of Sustainable Engineering,* vol. 6, pp. 10-22, 2012.

[2] M. T. Covas, C. A. Silva, and L. C. Dias, "Multicriteria decision analysis for sustainable data centers location," *International Transactions in Operational Research,* vol. 20, pp. 269-299, 2013.

Rapport-Gratuit.com

[3] J. Rath, "Data center site selection," p. 37, 2007.

[4] R. Lei, L. Xue, X. Le, and L. Wenyu, "Minimizing Electricity Cost: Optimization of Distributed Internet Data Centers in a Multi-Electricity-Market Environment," in *INFOCOM, 2010 Proceedings IEEE*, 2010, pp. 1-9.

[5] I. Goiri, L. Kien, J. Guitart, J. Torres, and R. Bianchini, "Intelligent Placement of Datacenters for Internet Services," in *Distributed Computing Systems (ICDCS), 2011 31st International Conference on*, 2011, pp. 131-142.

[6] R. Raghavendra, P. Ranganathan, V. Talwar, Z. Wang, and X. Zhu, "No "power" struggles: coordinated multi-level power management for the data center," *SIGPLAN Not.,* vol. 43, pp. 48-59, 2008.

[7] J. Yao, L. Xue, H. Wenbo, and A. Rahman, "Dynamic Control of Electricity Cost with Power Demand Smoothing and Peak Shaving for Distributed Internet Data Centers," in *Distributed Computing Systems (ICDCS), 2012 IEEE 32nd International Conference on*, 2012, pp. 416-424.

[8] R. Lei, L. Xue, X. Le, and P. Zhan, "Hedging Against Uncertainty: A Tale of Internet Data Center Operations Under Smart Grid Environment," *Smart Grid, IEEE Transactions on,* vol. 2, pp. 555-563, 2011.

[9] G. Yuanxiong, D. Zongrui, F. Yuguang, and W. Dapeng, "Cutting Down Electricity Cost in Internet Data Centers by Using Energy Storage," in *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, 2011, pp. 1-5.

[10] L. Uptime Institute (2012, Data center site infrastructure Tier Standard Available: http:// uptimeinstitute.com

[11] Z. Abbasi, G. Varsamopoulos, and S. K. S. Gupta, "TACOMA: Server and workload management in internet data centers considering cooling-computing power trade-off and energy proportionality," *ACM Trans. Archit. Code Optim.,* vol. 9, pp. 1-37, 2012.

[12] X. Hong and L. Baochun, "Cost efficient datacenter selection for cloud services," in *Communications in China (ICCC), 2012 1st IEEE International Conference on*, 2012, pp. 51-56.

[13] W. Van Heddeghem, S. Lambert, B. Lannoo, D. Colle, M. Pickavet, and P. Demeester, "Trends in worldwide ICT electricity consumption from 2007 to 2012," *Computer Communications,* vol. 50, pp. 64-76, 9/1/ 2014.

**ANNEXE II**

**LES CATÉGORIES DE L'EFFICACITÉ ÉNERGÉTIQUE (PUE)**

PUE(s)

| | PUE Category 0 | PUE Category 1 | PUE Category 2 | PUE Category 3 |
|---|---|---|---|---|
| Point de mesure IT | Sortie des UPS | Sortie des UPS | Sortie des PDU | Entrée des serveurs |
| Définition de l'énergie IT | Demande électrique ponctuelle (IT) (kW) | Energie IT annuelle | Energie IT annuelle | Energie IT annuelle |
| Définition de l'énergie totale | Demande électrique totale ponctuelle (kW) | Energie totale annuelle | Energie totale annuelle | Energie totale annuelle |

Source : Energy Star Data Center Energy Efficiency Initiatives

Figure-A II- 1 Les catégories de l'efficacité énergétique (PUE)

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

## ANNEXE III

## COMPARAISON DES PROFILS DE L'ÉNERGIE D'UN CENTRE DE DONNÉES ET D'UN BÂTIMENT TYPIQUE

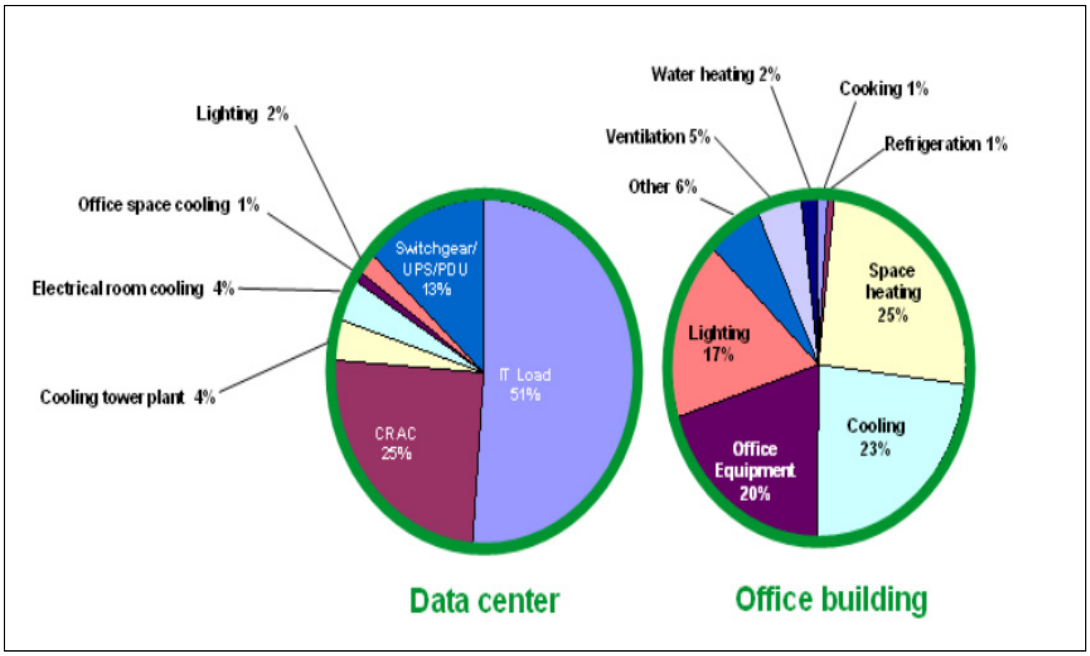


Figure-A III- 1 Comparaison des profils de l'énergie d'un centre de données et d'un bâtiment typique
Tirée de Bouley (2010)

## LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES


Abbasi, Zahra, Georgios Varsamopoulos et Sandeep K. S. Gupta. 2012. « TACOMA: Server and workload management in internet data centers considering cooling-computing power trade-off and energy proportionality ». *ACM Trans. Archit. Code Optim.,* vol. 9, n° 2, p. 1-37.

Alger, D. 2005. « Choosing an Optimal Location for Your Data Center ».

Barroso, Luiz André, et Urs Hölzle. 2009. « The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines ». *Synthesis Lectures on Computer Architecture,* vol. 4, n° 1, p. 1-108.

BELL. 2012. *Le coût réel de la gestion de données*. 10 p.

Bell, Michael. 2005. *"Use Best Practices to Design Data Center Facilities"*. Coll. « Gartner Research ». Consulté le April 22,.

Canada, la commissaire à la protection de la vie privée du. 2013. « Coin juridique ». < http://www.priv.gc.ca/leg_c/index_f.asp >.

Chang, Shin-Jyh Frank, Susmit Harihar Patel et James Marc Withers. 2007. « An Optimization Model to Determine Data Center Locations for the Army Enterprise ». In *Military Communications Conference, 2007. MILCOM 2007. IEEE*. (29-31 Oct. 2007), p. 1-8.

Covas, Miguel T., Carlos A. Silva et Luis C. Dias. 2012. « On locating sustainable Data Centers in Portugal: Problem structuring and GIS-based analysis ». *Sustainable Computing: Informatics and Systems,* vol. 3, n° 1, p. 27-35.

Covas, Miguel T., Carlos A. Silva et Luis C. Dias. 2013. « Multicriteria decision analysis for sustainable data centers location ». *International Transactions in Operational Research,* vol. 20, n° 3, p. 269-299.

Daim, Tugrul U., Ashok Bhatla et Mohammad Mansour. 2012. « Site selection for a data centre - a multi-criteria decision-making model ». *International Journal of Sustainable Engineering,* vol. 6, n° 1, p. 10-22.

Dumout, Estelle. 2007. « Google installe un centre de traitement des données en Belgique ». < http://www.zdnet.fr/actualites/google-installe-un-centre-de-traitement-des-donnees-en-belgique-39369079.htm >. Consulté le April 27th.

Fei, Yang, et Ye Li Xuan. 2011. « Method of Locating Data Center Based on Delphi ». In *Innovations in Bio-inspired Computing and Applications (IBICA), 2011 Second International Conference on*. (16-18 Dec. 2011), p. 299-302.

Foh, Chuan Heng, Yonggang Wen, Joel Rodrigues et Debojyoti Dutta. 2012. « SIG on Green Data Center and Cloud Computing ».< http://www3.ntu.edu.sg/home/aschfoh/sig/sig-on-green-data-center-and-cloud-computing/ >.

Gandhi, A., Chen Yuan, D. Gmach, M. Arlitt et M. Marwah. 2011. « Minimizing data center SLA violations and power consumption via hybrid resource provisioning ». In *Green Computing Conference and Workshops (IGCC), 2011 International*. (25-28 July 2011), p. 1-8.

Goiri, I., Le Kien, J. Guitart, J. Torres et R. Bianchini. 2011. « Intelligent Placement of Datacenters for Internet Services ». In *Distributed Computing Systems (ICDCS), 2011 31st International Conference on*. (20-24 June 2011), p. 131-142.

Google. 2012a. « Centre de données de Singapour». <http://www.google.com/intl/fr/about/datacenters/locations/singapore/opportunities-contacts.html >.

Google. 2012b. « Google Centres de données ». < http://www.google.com/about/datacenters/ >. Consulté le 2012.

Greenberg, Albert, James R. Hamilton, Navendu Jain, Srikanth Kandula, Changhoon Kim, Parantap Lahiri, David A. Maltz, Parveen Patel et Sudipta Sengupta. 2009. « VL2: a scalable and flexible data center network ». In *Proceedings of the ACM SIGCOMM 2009 conference on Data communication*. (Barcelona, Spain), p. 51-62. 1592576: ACM.

Haas, Olivier. 2013. « Cloud Computing et localisation - « Ciel, où sont mes données ?! » ». <http://www.journaldunet.com/solutions/expert/54839/cloud-computing-et-localisation-----ciel--ou-sont-mes-donnees.shtml >. Consulté le 17 juillet

Hong, Xu, et Li Baochun. 2012. « Cost efficient datacenter selection for cloud services ». In *Communications in China (ICCC), 2012 1st IEEE International Conference on*. (15-17 Aug. 2012), p. 51-56.

Iweb. 2013. « Centre de données «verts» : Près des États-Unis, les avantages du Canada». < http://iweb.com/fr/centre-de-donnees >.

Jie, Li, Li Zuyi, Ren Kui et Liu Xue. 2012. « Towards Optimal Electric Demand Management for Internet Data Centers ». *Smart Grid, IEEE Transactions on,* vol. 3, n° 1, p. 183-192.

Lei, Rao, Liu Xue, Xie Le et Liu Wenyu. 2010. « Minimizing Electricity Cost: Optimization of Distributed Internet Data Centers in a Multi-Electricity-Market Environment ». In *INFOCOM, 2010 Proceedings IEEE*. (14-19 March 2010), p. 1-9.

Lei, Rao, Liu Xue, Xie Le et Pang Zhan. 2011. « Hedging Against Uncertainty: A Tale of Internet Data Center Operations Under Smart Grid Environment ». *Smart Grid, IEEE Transactions on,* vol. 2, n° 3, p. 555-563.

Lemay, M., Nguyen Kim-Khoa, B. St. Arnaud et M. Cheriet. 2012. « Toward a Zero-Carbon Network: Converging Cloud Computing and Network Virtualization ». *Internet Computing, IEEE,* vol. 16, n° 6, p. 51-59.

LLC, Ponemon Institute. 2013. *2013 Cost of Data Center Outages*.

Oley, B. 2010. « Where's the Best Place to Build a Data Center? ».

Pakbaznia, Ehsan, et Massoud Pedram. 2009. « Minimizing data center cooling and server power costs ». In *Proceedings of the 2009 ACM/IEEE international symposium on Low power electronics and design*. (San Fancisco, CA, USA), p. 145-150. 1594268: ACM.

Raghavendra, Ramya, Parthasarathy Ranganathan, Vanish Talwar, Zhikui Wang et Xiaoyun Zhu. 2008. « No "power" struggles: coordinated multi-level power management for the data center ». *SIGPLAN Not.,* vol. 43, n° 3, p. 48-59.

Rath, John. 2007. « Data center site selection ». p. 37.

Rath, John. 2012. « Data center site selection ».<http://rath-family.com/rc/DC_Site_Selection.pdf >.

Stansberr, M. 2006. « Data Center Locations Ranked by Operating Cost, ».

Uptime Institute , LLC. 2012. « Data center site infrastructure Tier Standard ». < http:// uptimeinstitute.com >.

Van Heddeghem, Ward, Sofie Lambert, Bart Lannoo, Didier Colle, Mario Pickavet et Piet Demeester. 2014. « Trends in worldwide ICT electricity consumption from 2007 to 2012 ». *Computer Communications,* vol. 50, n° 0, p. 64-76.

Wierzbicki, Adam. 2002. « Internet Cache Location and Design of Content Delivery Networks ». In *Revised Papers from the NETWORKING 2002 Workshops on Web Engineering and Peer-to-Peer Computing*. p. 69-82. 714918: Springer-Verlag.

Yao, Jianguo, Liu Xue, He Wenbo et Ashikur Rahman. 2012. « Dynamic Control of Electricity Cost with Power Demand Smoothing and Peak Shaving for Distributed

Internet Data Centers ». In *Distributed Computing Systems (ICDCS), 2012 IEEE 32nd International Conference on*. (18-21 June 2012), p. 416-424.

Yuanxiong, Guo, Ding Zongrui, Fang Yuguang et Wu Dapeng. 2011. « Cutting Down Electricity Cost in Internet Data Centers by Using Energy Storage ». In *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*. (5-9 Dec. 2011), p. 1-5.

Zhang, Qi, Lu Cheng et Raouf Boutaba. 2010. « Cloud computing: state-of-the-art and research challenges ». *Journal of Internet Services and Applications,* vol. 1, n° 1, p. 7-18.