Conception de Centres de Données Internet à Haute Disponibilité

by *Gregory J. Graham*
Telecom Marketing Associates, Kanata, ON

(Traduit par Isabel Deslauriers)

Abstract

As services based on the Internet have become part of everyday life for Canadians, the infrastructure that provides these services must scale upwards in robustness and capacity. Lessons learnt from the public switched telephone network are being applied, with new twists, to Internet infrastructure. The importance of providing secure, high-availability infrastructure has become even more obvious following the recent events of September 11, 2001.

This article describes leading approaches to designing high-availability data centres used for web-hosting, e-commerce, storage and other IP applications.

Sommaire

Comme les services basés sur l'Internet font maintenant partie de la vie quotidienne des canadiens, l'infrastructure qui assure ces services doit être “scale upwards” en robustesse et en capacité. Les leçons apprises des réseaux téléphoniques publics à interrupteurs sont appliqués à l'infrastructure de l'Internet avec des variantes. Suite aux évènements du 11 septembre 2001, il est devenu encore plus important d'offrir une infrastructure sécuritaire et a disponibilité élevée. Cet article décrit les approches d'avant-garde dans la conception de centres de données à haute disponibilité utilisés pour hôtes de pages web, commerce électronique, entreposage de données et autres applications utilisant le protocole Internet.

1.0 Introduction

Avec l'usage grandissant d'Internet, les cédules 24 heures sur 24, 7 jours semaine, et la disponibilité globale permise par le web, les sites corporatifs et le commerce électronique ne peuvent simplement pas se permettre de repos. Toute interruption a clairement un coût apparaissant par le biais d'une perte de clients, de revenu ou de productivité.

Les Centres de données Internet (CDI) sont des installations créées avec soin, qui combinent la fidélité d'une centrale téléphonique avec des concepts de centre de traitement des données ainsi qu'une sécurité physique à la fine pointe de la technologie. Le résultat? Une infrastructure extrêmement robuste. En fait, les meilleurs CDIs sont basés sur deux objectifs:

- Desservir de façon continue, en minimisant le risque d'interruption;
- Si une interruption se produit, y remédier le plus rapidement possible.

Cependant, une telle infrastructure a son coût. Les centres de données Internet requièrent un emplacement et un édifice choisis avec soin, un système de sécurité physique et informatique, des systèmes de contrôles environnementaux, une source d'énergie de secours et une redondance complète. Le partage de ces coûts parmi un grand nombre d'utilisateurs permet de réduire le coût individuel, ce qui est le concept fondamental supportant le modèle de centre de données Internet commercial.

Cet article examine les concepts supportant les centres de données Internet, et démontre comment on minimise le risque dès l'étape initiale de la conception.

2.0 Quelle est la composition d'un centre de données?

Ce qui compose un centre de données dépend du modèle de commerce souhaité. En général, les CDIs dérivent leurs profits de plusieurs sources:

- La location d'espace, d'énergie et de connectivité,
- L'opération de services informatiques sous-contractés et
- La mise en disponibilité de services professionnels optionnels.

2.1 Accueil de services

Un centre de données peut être aussi simple que la mise en disponibilité d'un espace aux fournisseurs de communications qui y installeront eux-même leur équipement et réseau. Ces fournisseurs louent en fait des espaces à température contrôlée, ainsi que l'énergie et la connectivité offertes.

La situation physique de l'espace est le point clé, et doit tenir en ligne de compte la proximité des clients ainsi que des réseaux optiques redondants. Cet aspect devient particulièrement important en milieu urbain dense, où l'accessibilité à ces items peut devenir difficile.

Il faut noter que dans ce cas-ci, les fournisseurs d'accès sont responsables de l'installation et du maintien de leurs opérations, et deviennent en quelque sorte sous-locataires face au centre de données.

2.2 Colocation

Le déploiement et l'opération d'équipement dans un espace appartenant à un autre propriétaire s'appelle `colocation`. Un centre de données peut générer un revenu de colocation en fournissant de l'espace et de la connectivité à des fournisseurs de services locaux compétitifs et à des fournisseurs d'accès Internet. Ces locataires obtiennent en échange une connexion au réseau téléphonique local sans devoir débourser pour développer leurs propres installations.

Les éléments tels que commutateurs, passerelles, et multiplexeurs de disponibilité sont situés dans la centrale, et se connectent directement à la boucle locale. Des lignes de jonction redondantes garantissent une haute disponibilité sous presque n'importe quelles conditions. Les centrales téléphoniques ont donc été les premiers centres de données. La majeure différence est que les CDIs abritent des serveurs, alors que les centrales téléphoniques abritent des commutateurs et l'équipement relié à la disponibilité.

2.3 Accueil avec gestion

Contrairement à la colocation, l'accueil avec gestion peut être qualifié de service clé en main. Dans ce cas, les clients sous-contractent leur commerce électronique à l'opérateur du centre de données Internet. En retour, les opérateurs de CDI fournissent une qualité garantie via un contrat de niveau de service (SLA), c'est- à- dire un contrat définissant les prix, les termes, la performance informatique et les pénalités qui peuvent y être associées. L'équipement peut être la propriété soit du client, soit de l'opérateur du CDI, dépendamment de l'arrangement commercial.

Une variante sur ce thème est l'hébergement dédié, une approche où l'on dédie des serveurs distincts à chaque client. Ceci élimine les conflits de logiciels possibles et les difficultés amenées par l'établissement de cédules de maintien lorsque les équipements sont partagés.

2.4 Services à valeur additionnelle

Les centres de données Internet ont rajouté à ces modèles de base une panoplie de services à la carte. Les services professionnels couramment offerts incluent les coupe-feu (`firewall`), la surveillance de réseau, l'équilibrage de charge, les tests de performance, les audits de sécurité, des réseaux de données privés et des services de banques de données. Ces services requièrent tous une certaine expertise en réseautique qui peut ne pas être disponible de façon interne pour les clients.

3.0 De la Centrale téléphonique au CDI

Les centrales téléphoniques (CT) sont depuis longtemps équipées de systèmes de chauffage, climatisation, énergie de secours et lignes de jonction redondantes, ayant pour but d'assurer un espace conditionné, de l'énergie, de la sécurité et de la connectivité.

Parce que les centrales électroniques terminent les boucles téléphoniques locales par des commutateurs, les ports de commutateurs sont l'unité de base des opérations de CT. Une centrale Nord-Américaine typique peut abriter 10 000 boucles locales, et donc nécessiter 10 000 ports.

D'un autre côté, les entreprises requièrent un établissement capable d'abriter les serveurs qui livrent leur contenu web, commerce électronique, et services de gestion des clients. Ces systèmes opèrent couramment sur 110 VCA, contrairement aux 48 VCC typiques des centrales téléphoniques. Il s'agit d'une des différences clés entre une CT et un CDI.

L'espace intérieur d'un CDI peut se trouver divisé en espaces accommodant les supports à équipement, l'énergie disponible, et la connectivité. Dépendamment des demandes des clients et de la philosophie de gestion appliquée, les supports à équipement peuvent être individuels, groupés par rangées, dans des cages verrouillées ou même dans des voûtes verrouillées séparément.

Sans égard à la manière dont l'espace est loué, l'équipement des clients se retrouve habituellement monté sur des supports de groupe. Ces supports sont les unités de base du CDI car l'utilisation de l'espace - contrairement aux ports de commutateurs - est le principal critère générateur de revenus et de dépenses. Les dimensions de la base peuvent se trouver entre 20 et 30 pieds carrés (1.85 à 2.79 mètres carrés); leur réduction augmente la densité de supports et conséquemment les revenus possibles.

La densité de supports peut également être augmentée en choisissant des supports de plus haute taille. Les manufacturiers réduisent continuellement le volume des serveurs. Alors que les serveurs présents peuvent avoir une taille comparable à des boîtes de pizza et une épaisseur aussi petite que 1.75 pouces (4.45 cm, ou une unité de support; un support standard de 78 pouces de hauteur abrite 45 unités), les serveurs varient énormément en ce qui a trait à leur volume et à leur puissance de calcul. Les très gros serveurs sont équipés de leurs propres supports, alors que quarante serveurs de grosseur boîte à pizza peuvent partager un même support.

L'objectif visant à maximiser la densité de serveurs dans un espace minimal crée un problème de disponibilité d'électricité et de climatisation assez sévère, et qui n'existait pas pour les centrales téléphoniques, les CT du passé ayant une puissance de calcul limitée et une grande quantité de ports entrée/sortie consumant un espace important. En contraste, les centres de données ont une puissance de calcul éblouissante qui cause la dissipation de grandes quantités de chaleur.

4.0 Élimination de facteurs à risque

4.1 Situation

L'élimination de facteurs à risque débute avec le choix d'emplacement pour le CDI. Le choix du site peut minimiser le risque d'un désastre naturel en évitant les zones d'inondation et les zones souffrant de tornades fréquentes, etc. Les centres de données devraient aussi se situer à une distance sécuritaire de toute autoroute, chemin de fer et ligne de vol.

La disponibilité de communication par fibre optique est une autre considération. Idéalement, un CDI devrait être situé de façon à ce qu'il soit possible de se connecter à des réseaux optiques redondants dépendants de plus d'un opérateur.

4.2 Construction

De plus, le bâtiment abritant le centre de données devrait être construit de façon à supporter les désastres naturels ainsi que toute attaque. Même si ces constructions devraient être habituelles, elles se font plus recherchées dans cette réalité suivant les attaques du onze septembre. Les CDI ne se contentent pas seulement de murs en béton épais, un minimum de fenêtres, du vitrage et des murs à l'épreuve des balles incluant un renforcement d'acier inséré sous les cloisons sèches, et une entrée par sas habituellement verrouillée, mais demandent aussi la possibilité d'une structure renforcée à l'épreuve des tremblements de terre et des détonations de bombes (Figure 1).

Comme a été noté ci-dessus, une plus grande densité de serveurs peut augmenter le revenu de CDI. Cependant, ce poids additionnel peut nécessiter un plancher plus épais. Le sol sous le CDI doit aussi être capable de supporter une charge exceptionnellement lourde, et toute nouvelle construction doit conséquemment être spécifiée en accord avec ces critères. Un édifice existant sous considération pour établissement de CDI doit aussi être inspecté par des experts en structure afin de déterminer s'il pourra supporter un tel poids.

4.3 Sécurité Physique

Un haut niveau de sécurité physique et informatique est fondamental au concept de CDI. Les mesures de sécurité physique incluent les clôtures de périmètre, détecteurs de mouvement, alarmes, gardes, surveillance par vidéo et l'accès contrôlé aux équipements.

En particulier, le personnel de sécurité et les mesures prises doivent

Figure 1 (gauche): Le vestibule d'un centre de données intelligent de TELUS possède des murs et du vitrage anti-balles, un sas verrouillé, des détecteurs biométriques, et des gardes de sécurité à tout moment.

Figure 2 (droite): Les méthodes d'identification biométriques vérifient qu'une personne est bien celle qu'elle prétend être. Nous voyons ici un détecteur d'empreinte digitales à un centre de données intelligent TELUS.

prévenir tout accès par un client à l'équipement d'un autre client, spécialement lorsqu'il s'agit de compétiteurs directs.

Par exemple, dans les centres de données TELUS, toute personne désireuse d'accès doit se soumettre à plusieurs niveaux de sécurité. Premièrement, les clients doivent fournir une liste du personnel auquel ils cherchent à autoriser l'accès des équipements en question. Tous ces clients doivent se soumettre à un examen de dossier criminel par le biais des autorités policières locales. Ensuite, on émet à chacun une carte d'identité avec photo et bande magnétique. Finalement, un échantillon d'empreinte digitale est soutiré.

Quelqu'un pénétrant à l'intérieur d'un centre de données doit apparaître sur la liste fournie, présenter sa carte d'identité au garde qui la vérifie à l'aide de la bande magnétique et faire reconnaître son empreinte digitale par un détecteur approprié. Seulement après toutes ces étapes lui sera-t-il permis de traverser le sas normalement verrouillé (Figures 2 et 3).

Une fois à l'intérieur, le visiteur n'est pas libre de se promener à sa guise. L'accès aux différentes zones du CDI est contrôlé par des lecteurs de bande magnétique et d'empreintes digitales additionnels. Tout équipement est de plus verrouillé de façon individuelle. Certains centres de données emploient des escortes de sécurité, cependant ce système est moins sécuritaire. Il y a un petit risque que l'escorte soit momentanément distraite ou requiert une pause biologique, ce qui laisse un saboteur potentiel sans surveillance.

Sous l'éventualité d'une situation d'urgence, les centres de donnés sont connectés aux stations de pompiers, de police et autres autorités.

Cette approche compréhensive permet aux centres de données Internet de mettre en disponibilité le maximum en terme de sécurité - lecteurs biométriques, gardes en tout temps, vitrage anti-balles, détection d'intrusion de réseau avancée, et multiples coupe-feu.

4.4 Sécurité de réseau

La base de la sécurité de réseau est le coupe-feu, un dispositif qui inspecte le trafic présent sur le réseau. Les coupe-feu filtrent le trafic en se basant sur l'origine et la destination des paquets, les numéros de ports ainsi que d'autres paramètres afin de bloquer tout trafic non autorisé à entrer le réseau du centre de données.

Une stratégie d'architecture de réseau assez commune consiste à connecter les réseaux de colocation et d'accueil avec gestion sur des sous-réseaux séparés. Une seconde série de coupe-feu est établie afin d'isoler les serveurs d'accueil avec gestion des serveurs de colocation. À l'occasion, les coupe-feu choisis sont d'un différent fabricant, afin de présenter une deuxième difficulté aux pirates qui peuvent avoir une expertise spécialisée et conséquemment être familiers avec les faiblesses d'un genre de coupe-feu en particulier.

Les clients du centre de données peuvent aussi établir une troisième protection en implémentant leur propre protocole de sécurité par le biais de leur propre coupe-feu dédié. L'installation, l'opération et le maintien de coupe-feu est un des services professionnels les plus populaires parmi ceux offerts par les centres de données.

La prochaine mesure de sécurité vitale aux technologies de l'informatique est le système de détection d'intrusion, ou IDS. Un IDS surveille le trafic du réseau afin de détecter les trains d'attaque potentiels.

Plusieurs attaques pirates exploitent les vulnérabilités présentées par les protocoles de communication afin de créer délibérément des erreurs ou d'autres conditions sous lesquelles le mal peu être commis. La différence entre le coupe-feu et le système de détection d'intrusion est que le IDS maintien l'état du protocole et l'information contextuelle en temps réel. Ceci permet la détection du trafic malicieux ayant pénétré le coupe-feu grâce à l'usage d'adresses IP et de numéros de ports autorisés. Le IDS peut subséquemment déployer des alertes, terminer des sessions d'utilisateurs, et même reconfigurer les coupe-feu afin de bloquer le trafic.

Dans le cas d'un centre de données Internet opéré par une compagnie de communications, ces systèmes de sécurités peuvent être intégrés et opérés de façon indépendante afin d'offrir de très hauts niveaux de sécurité. Par exemple, la plupart des opérateurs de dorsale IP ont déjà des coupe-feu et des IDS en place.

Figure 3: Un unique point d'entrée et de sortie contrôle l'autorisation d'accès, prévient le prélèvement d'équipement et nuie à une éventuelle attaque physique.

4.5 Redondance

Une très haute disponibilité est presque toujours atteinte par le biais de la redondance. En théorie, les centres de données Internet ne devraient avoir aucun point de coupure simple. Ceci implique que tous les systèmes doivent avoir un secours automatique disponible - même lorsque certains systèmes sont mis hors de service pour cause de maintien.

Au minimum, de multiples lignes de jonction doivent être utilisées, idéalement sous le contrôle de différentes compagnies et pénétrant le CDI à différents endroits.

La disponibilité à long terme d'électricité est une considération importante. On n'a qu'à considérer les récentes pannes et mises en veilleuse de la Californie, le plus important centre géographique de centres de données.

Les centres de données Internet utilisent typiquement de 85 à 100 watts par pied carré afin d'alimenter leurs équipements et de maintenir l'air climatisé tel que requis. Ceci représente plus de vingt fois l'usage commercial moyen. Lorsque possible, le CDI devrait avoir accès à une grille d'énergie redondante. Les sources d'énergie continues ou générateurs diesel sont aussi communément utilisés afin de garantir une source d'électricité continue.

Étant donné la grande densité de puissance dans les CDIs, une panne de système de refroidissement causerait une hausse de température rapide. Même si une catastrophe thermique totale n'a jamais été rapportée publiquement, il s'agit d'une possibilité réelle et donc une redondance de systèmes de refroidissement est essentielle. Il ne faut pas par contre s'attendre à voir deux immenses climatiseurs dans un centre de données. Cette redondance est parfois atteinte en sur-dimensionnant intentionnellement une unité composée de climatiseurs modulaires redondants.

Un concept emprunté directement des centres de traitement des données est l'utilisation de planchers surélevés. En plus de simplifier le câblage, ils créent un corridor de ventilation efficace. Certains CDI ont pour cette raison des planchers surélevés de plus de deux pieds.

La redondance ultime consiste en un second centre de données. Certains opérateurs de CDI planifient la construction de réseaux reliant leurs centres de données, afin de favoriser la sauvegarde d'information mutuelle.

Les centres de données TELUS ont une solution élégante à ce problème, qui utilise la technologie d'équilibrage de charge de Cisco Systems, Inc. afin de distribuer la charge de serveurs géographiquement parmi les centres de données. Par exemple, une entreprise pourrait avoir ses opérations web accueillies simultanément à Calgary et à Toronto. La technologie d'équilibrage de charge détecte la situation géographique de l'utilisateur et distribue la charge de façon à maximiser la performance et minimiser le temps de réponse. Si l'inconcevable devait arriver et l'un des centres de données tombait en panne, l'équilibreur de charge transfèrerait automatiquement le trafic vers le deuxième site.

4.6 Gestion

La disponibilité de compétence en technologie de l'information est un autre critère favorisant la demande envers les CDIs. Les centres de données sont typiquement opérés par une tierce personne, un arrangement qui peut engendrer plusieurs bénéfices.

Premièrement, la délégation des opérations à une tierce personne peut être un objectif stratégique, surtout lorsque ces compétences ne sont pas trouvées à l'intérieur de la compagnie en question.

Deuxièmement, une tierce personne externe peut être tenue responsable de rencontrer les contrats de niveau de service. Ceci peut être difficile lorsqu'il s'agit d'une fonction interne à la compagnie.

Et troisièmement, une tierce personne neutre peut accumuler la demande provenant de plusieurs clients qui ne collaboreraient pas ordinairement. Ceci peut causer une baisse de coûts d'opérations pour tous ceux impliqués.

L'expertise technique et gestionnaire forme une importante partie de la valeur proposée par les centres de données.

5.0 Conclusion

Les centres de données Internet offrent un environnement à haute disponibilité, donc les services vitaux qu'ils abritent peuvent être offerts avec un risque d'interruption minimal.

Les hauts coûts du IDC sont habituellement difficiles à assumer pour une entreprise individuelle. Cependant, lorsque partagés par plusieurs, l'investissement de plusieurs milliers de dollars peu être réduit à un coût d'opération plus acceptable pour chacun. Ce partage des coûts permet aux clients d'IDC de prendre avantage des technologies réduisant les risques qui ne seraient pas normalement accessible pour eux, et ils obtiennent finalement une plus grande fiabilité à un moindre coût.

6.0 Remerciements

Les contributions de Craig Richardson, directeur à TELUS Business Solutions (hosting solutions), Calgary lors de la préparation de cette publication sont appréciées.

À propos de l'auteur

Gregory Graham détient un Baccalauréat en Génie Électrique de l'Université Concordia à Montréal, et un MBA de l'université d'Alberta à Edmonton.

Il représente près de 20 années d'expérience dans l'industrie de la télécommunication et de l'informatique. Son dossier inclus des positions à responsabilité grandissante en marketing et en planification stratégique à Hewlett-Packard, TELUS Advanced Communications, et Nortel Networks. Il fut un membre de l'équipe de gestion qui transforma TELUS PLAnet en le plus important service d'Internet de l'Ouest du Canada, et qui introduisit le premier service ADSL résidentiel en Amérique du Nord. Il a aussi créé le premier service de fin d'appels VoIP offert par une entreprise de télécommunications. Il peut être rejoint par courriel à greg_graham@compuserve.com.