Combell
Your host on the internet

Réduisez votre temps d'indisponibilité à un minimum avec un concept multi-centre de données

Livre blanc

Confiez vos activités critiques à un expert

S'il est crucial pour vos activités commerciales que vos serveurs soient disponibles en continu, vous devez demander à votre hébergeur de vous fournir une solution haute disponibilité (high availability). Si vous exploitez p. ex. une boutique en ligne et que vous perdez des clients ou que vous ratez des commandes à chaque minute que votre serveur web ne fonctionne pas, ou que vous proposez une plateforme de commerce électronique pour laquelle chaque minute compte, vous avez tout intérêt à opter pour un concept multi-centre de données : deux centres de données qui fonctionnent tel un seul et unique réseau logique et qui optimisent ainsi votre disponibilité : Grâce à cette répartition entre deux endroits, vous pouvez garantir la disponibilité de votre service Internet. Lorsqu'une panne se produit dans un des centres de données, vos serveurs continuent à tourner dans le deuxième centre de données. Dans ce document technique, nous allons analyser la technologie qui rend cela possible et nous allons également expliquer à quoi vous devez faire attention lorsque vous optez pour une solution multi-centre de données.

Un concept multi-centre de données traditionnel basé sur le DNS

D'habitude, un concept multi-centre de données est basé le DNS (Domain Name System), le protocole de réseau qui est utilisé pour convertir les noms des ordinateurs (noms de domaines) en adresses IP. On redirige donc p. ex. le nom de domaine www.example.com vers l'adresse IP 1 d'un serveur web dans le premier centre de données. Si ce centre de données est confronté à un problème technique, on fait en sorte que le même nom de domaine soit redirigé vers l'adresse IP 2 d'un serveur web identique dans le deuxième centre de données. À partir de là, toutes les requêtes des clients sont redirigées vers le serveur web qui se trouve dans le deuxième centre de données.

Du fait que les paramètres DNS ne changent généralement pas très souvent, le DNS utilise cependant la mise en cache (caching) : un ordinateur qui demande l'adresse IP derrière un nom de domaine obtient aussi une valeur TTL (time to live) qui indique le temps durant lequel cette adresse reste valable. Cette durée peut être comprise entre quelques minutes et quelques jours. Si durant ce temps l'ordinateur a besoin de retrouver cette adresse qui fait partie de ce nom de domaine, il la récupèrera dans un cache local. Ainsi, on évite que les serveurs DNS soient constamment surchargés avec des requêtes superflues.

Si vous voulez que www.example.com soit redirigé vers le serveur web qui se trouve dans le deuxième centre de données au moment où le premier centre de données n'est plus disponible, vous devez paramétrer le TTL de www.example.com au préalable sur p. ex. cinq minutes. Les visiteurs de votre site web redemanderont dans ce cas l'adresse IP de votre serveur web toutes les cinq minutes. À l'instant où le premier centre de données n'est plus disponible, faites en sorte que votre nom de domaine soit redirigé vers l'adresse IP 2 de votre serveur web identique dans le deuxième centre de données. En théorie, après cinq minutes, chaque serveur DNS dans le monde entier aura attribué cette nouvelle adresse IP au nom de domaine www.example.com, ce qui fait que les visiteurs arriveront sur le deuxième serveur web.

Le DNS n'offre pas de garanties suffisantes

Un concept multi-centre de données qui utilise le DNS fonctionne dans la plupart des cas, mais il a quelques inconvénients majeurs :

- **L'approche n'est pas fiable**

En théorie, après cinq minutes, votre serveur web est à nouveau disponible via le deuxième centre de données, mais dans la pratique, certains fournisseurs Internet ne tiennent pas compte d'un TTL inférieur à quelques heures. De ce fait, certains visiteurs pourront déjà atteindre votre site web après cinq minutes via le deuxième centre de données mais peut-être pas tous. Il ne s'agit donc pas d'une bonne solution multi-centre de données.

- **Vous surchargez l'infrastructure DNS**

Avec un TTL aussi court, vous abusez en fait du protocole DNS, car ce dernier n'a pas été conçu pour de telles situations. Avec un TTL de 5 minutes, les serveurs DNS reçoivent chaque jour d'énormes quantités de requêtes superflues, ce qui explique aussi pourquoi certains fournisseurs Internet voient cela d'un mauvais œil : car, ce faisant, vous surchargez en effet l'infrastructure DNS.

- **Une période d'indisponibilité de cinq minutes est trop longue**

Un TTL inférieur à cinq minutes n'est pas réaliste pour toute une série de raisons. Dans le cas d'un concept multi-centre de données avec DNS, vous devez donc – dans le meilleur des cas – compter sur un temps d'indisponibilité de cinq minutes, ce qui reste trop long pour un grand nombre d'applications.

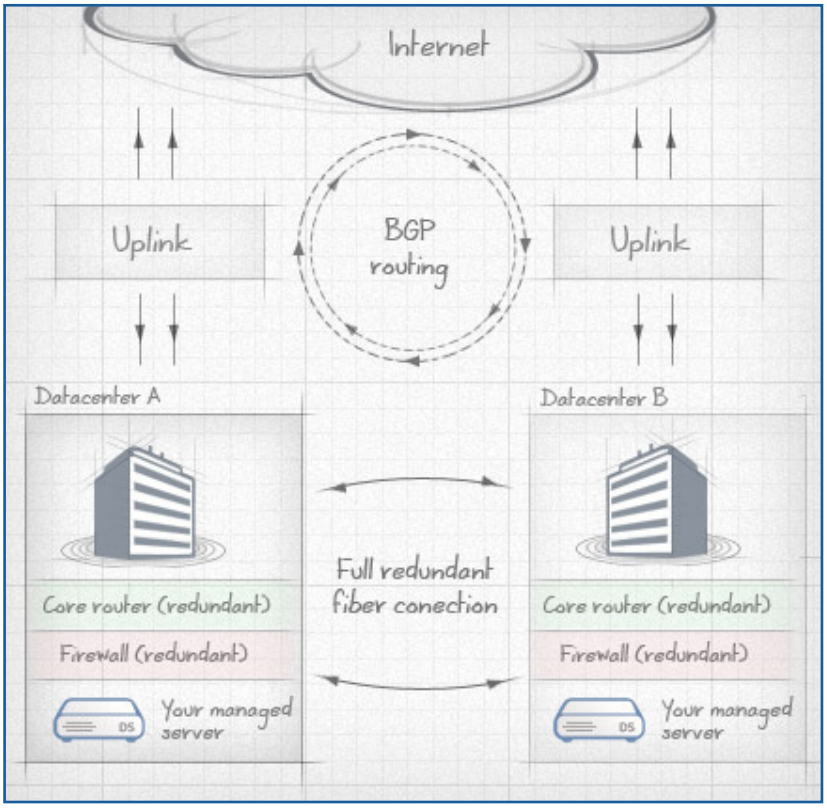
Bien qu'un concept multi-centre de données basé sur le DNS soit toujours utilisé par maints hébergeurs, Combell Solutions a cessé de proposer cette solution en 2009. Au lieu de cela, nous avons opté pour une technologie qui permet de passer plus rapidement d'un centre de données à un autre.

Un concept multi-centre de données basé sur BGP : un taux de disponibilité atteignant presque 100 %

BGP (Border Gateway Protocol) est le protocole qui est utilisé pour le routage du trafic réseau entre plusieurs fournisseurs : les routeurs BGP communiquent entre eux afin de connaître les adresses IP disponibles. Ainsi, il est possible de déterminer (automatiquement, à tout moment, et pour chaque visiteur de votre site web) le chemin le plus rapide pour accéder au centre de données.

Combiné avec un répartiteur de charge, le protocole BGP est une manière plus logique que le DNS pour passer d'un centre de données à un autre. Dans chaque centre de données, vous devez donc faire tourner un serveur web identique derrière un répartiteur de charge (load balancer). Si les visiteurs veulent atteindre votre site web via l'adresse IP de votre serveur web, ils doivent introduire l'adresse IP publique du répartiteur de charge. Celui-ci vérifie constamment lequel des deux serveurs web configurés de manière redondante est accessible et transmet les paquets réseau au(x) serveur(s) disponible(s).

En configurant correctement le répartiteur de charge et le routeur BGP, il y aura toujours un chemin permettant d'atteindre le serveur. Si un des centres de données est confronté à une panne, le routeur remarque que le répartiteur de charge n'est plus disponible via un des chemins et emprunte donc le chemin encore disponible. En même temps, le répartiteur de charge cesse de rediriger le trafic réseau vers le serveur web qui ne répond plus vers le serveur web sous-jacent. Une fois la panne solutionnée, le répartiteur de charge recommence à rediriger le trafic réseau vers le serveur web. Tout cela se fait de manière complètement transparente pour les internautes, qui ne peuvent d'aucune manière se douter qu'une panne est survenue dans un des centres de données.



Pour eux, c'est comme si le serveur web avait toujours fonctionné. Votre site web peut ainsi rester accessible au public à presque 100 % du temps : l'indisponibilité n'existe pratiquement plus.

Quelles garanties de disponibilité sont possibles sur le plan technologique?

Grâce à la technologie multi-centre de données basée sur BGP, vos serveurs peuvent atteindre le taux de disponibilité le plus élevé qui soit. Quant aux garanties envisageables, tout dépend par contre de ce qui se passe dans les centres de données et de la distance qui les sépare.

L'avantage de Combell Solutions est que, pour un surcoût relativement faible, vous pouvez effectuer la transition de votre infrastructure de serveurs prévue pour un seul centre de données vers notre infrastructure multi-centre de données. Nous nous chargeons de la gestion pour vous, afin que vous puissiez vous concentrer sur vos applications sans devoir vous soucier de la disponibilité de l'infrastructure.

- **Quelle redondance le centre de données offre-t-il ?**

Un centre de données peut offrir plusieurs degrés de disponibilité, en fonction du degré de redondance qui s'y trouve. Si vous n'utilisez qu'un seul centre de données sans prendre de mesures spéciales, votre serveur tombera en panne ou sera indisponible en cas de problème technique. Il est cependant possible d'éviter cela en utilisant une architecture intelligente (multi-room) à l'intérieur du centre de données, qui peut être composée de deux salles fonctionnant de manière totalement indépendante. Ainsi, si un incendie devait p. ex. se produire dans une des salles, l'autre salle continuerait à fonctionner et un serveur pourrait donc prendre le relais de l'autre dans la deuxième salle. Dans le cas d'une catastrophe à plus grande échelle qui toucherait l'entièreté du centre de données, un deuxième centre de données situé à un autre endroit serait cependant nécessaire (une solution multi-centre de données). Naturellement, même dans un concept multi-centre de données, chaque centre de données peut aussi utiliser une architecture multi-room.

- **L'écriture miroir des données doit-elle se faire de manière synchrone ou asynchrone ?**

Dans le cas d'une catastrophe survenant dans un centre de données, il est non seulement impératif que vos serveurs restent disponibles, mais aussi que vous perdiez le moins de données possibles. Voilà pourquoi vous avez, dans de nombreux cas, intérêt à synchroniser le stockage dans chaque centre de données (mirroring). À chaque fois que votre serveur web enregistre des données dans l'espace de stockage d'un des centres de données, ces mêmes données doivent aussi être enregistrées dans l'autre centre de données. Cela peut se faire de manière synchrone ou asynchrone.

Dans le cas de l'écriture miroir synchrone (synchrone mirroring), les données sont enregistrées simultanément dans le premier centre de données et transmises vers le deuxième centre de données. L'application attend jusqu'à ce que les données soient enregistrées dans les deux centres de données. La connexion réseau vers le deuxième centre de données entraîne un ralentissement supplémentaire de l'application, mais la perte de données est exclue. Dans le cas de l'écriture miroir asynchrone (asynchrone mirroring), les données sont également transmises vers le deuxième centre de données, mais l'application n'attend pas que l'opération soit terminée. L'application fonctionne donc plus rapidement, mais il existe un risque de perdre des données dans le cas où une catastrophe devrait avoir lieu dans le premier centre de données.

Le choix entre les deux technologies dépend en grande partie du type d'application que l'on souhaite répartir sur les 2 centres de données (à quel point l'impact du ralentissement causé par la réplication synchrone est notable) et des exigences professionnelles : combien de données peut-on se permettre de perdre et à quelle vitesse le 2ème centre de données doit-il être disponible en cas de panne ?

Combien de temps d'indisponibilité pouvez-vous vous permettre ?

Avec une solution multi-centre de données basée sur BGP, vos serveurs peuvent atteindre un taux de disponibilité de presque 100 % : si une catastrophe survient dans un des centres de données, vos serveurs se mettent en route dans l'autre centre de données en à peine quelques secondes. Ce qui est possible sur le plan technologique ne l'est cependant pas toujours sur le plan économique. Commencez de ce fait toujours par effectuer une analyse de l'impact sur votre business, dans laquelle vous effectuez une analyse des risques de tout ce qui peut (mal) se passer et des conséquences sur votre entreprise. Vous pourrez ensuite déterminer les paramètres suivants auxquels la solution de votre hébergeur devra répondre :

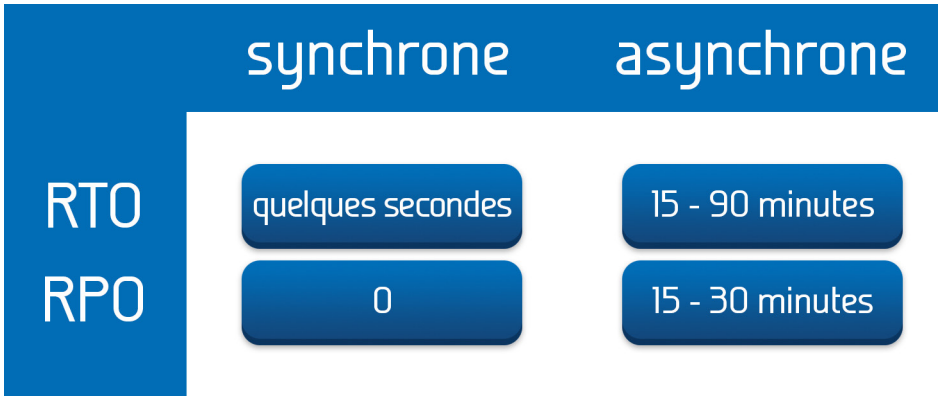
- **Combien de temps votre application peut-elle être indisponible ?**

La réponse à cette question est exprimée dans le RTO (recovery time objective), via lequel vous indiquez le délai nécessaire pour que l'application soit à nouveau fonctionnelle sans compromettre la continuité de vos activités commerciales. Attention : veillez à ce que votre fournisseur et vous-mêmes compreniez le terme RTO de la même manière. Imaginez p. ex. qu'un hébergeur avec une infrastructure multi-centre de données basée sur le DNS définisse le RTO comme le temps nécessaire pour démarrer un serveur de remplacement dans le deuxième centre de données. Si ce serveur n'est pas disponible dans ce même délai en raison du caractère peu fiable du DNS, pour vous, tout ne sera pas rentré dans l'ordre : vous n'êtes en effet pas d'accord avec le fait que le serveur tourne si les visiteurs de votre site web ne peuvent pas y accéder.

- **Quelle quantité de données pouvez-vous perdre si un centre de données tombe en panne ?**

La réponse à cette question est exprimée dans le RPO (recovery point objective), via lequel vous indiquez le délai maximum (avant la panne) durant lequel vous pouvez perdre des données sans compromettre la continuité de vos activités commerciales. Veillez à ce que les accords avec votre hébergeur soient clairs au sujet du moment où le temps commence à s'écouler après une catastrophe.

Tant le RTO que le RPO sont exprimés en unités de temps. Quant à savoir quels RTO et RPO sont envisageables, cela dépend de la technologie que vous utilisez. Le stockage est un des principaux facteurs déterminants dans ce contexte. Les RTO et RPO suivants peuvent généralement être atteints avec l'écriture miroir synchrone et asynchrone :



Il s'agit là de valeurs théoriques. Dans la pratique, vous devez faire certains choix technologiques pour pouvoir déterminer cela. Vous devez ensuite effectuer des comparaisons entre les frais encourus pour atteindre les RTO et RPO susmentionnés avec ces choix et les pertes que votre entreprise pourrait subir durant une telle période d'indisponibilité ou en perdant des données durant ce temps. En fonction de votre application, cette perte peut être facile à calculer. Une boutique en ligne dispose p. ex. de chiffres concernant les recettes moyennes par minute et on peut donc en déduire un niveau d'indisponibilité acceptable.

À propos de Combell Solutions

En tant que leader absolu du marché des services d'hébergement pour entreprises, intégrateurs de TI et développeurs de logiciels, Combell est devenu, depuis sa fondation en 1999, un des partenaires uniques les plus fiables pour l'hébergement d'environ n'importe quelle infrastructure de TI ou application ou n'importe quel site web. Combell Solutions est une division de Combell qui se charge de la réalisation de projets d'hébergement avancés sur mesure.