

Définition et présentation

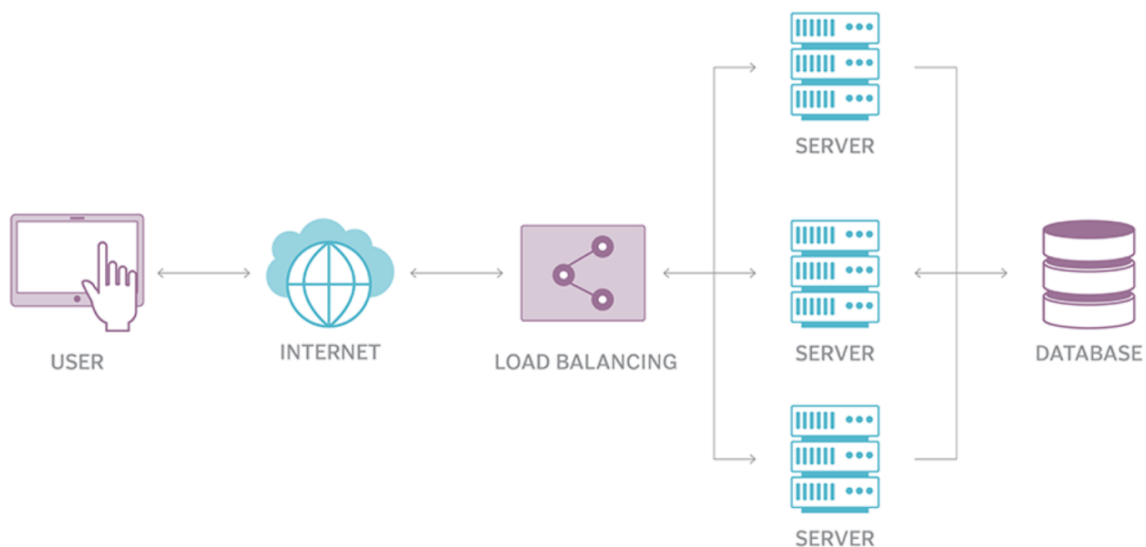
Date : 20-02-2023

Auteur: Eloïd DOPPEL

Load Balancing ou équilibreur de charge:

Le **load balancing** est une technique qui permet de répartir équitablement la charge de travail entre plusieurs serveurs. Lorsque vous avez un site Web ou une application qui reçoit un grand nombre de requêtes simultanées, il peut être difficile pour un seul serveur de traiter toutes ces demandes efficacement et rapidement. C'est là que le load balancing entre en jeu.

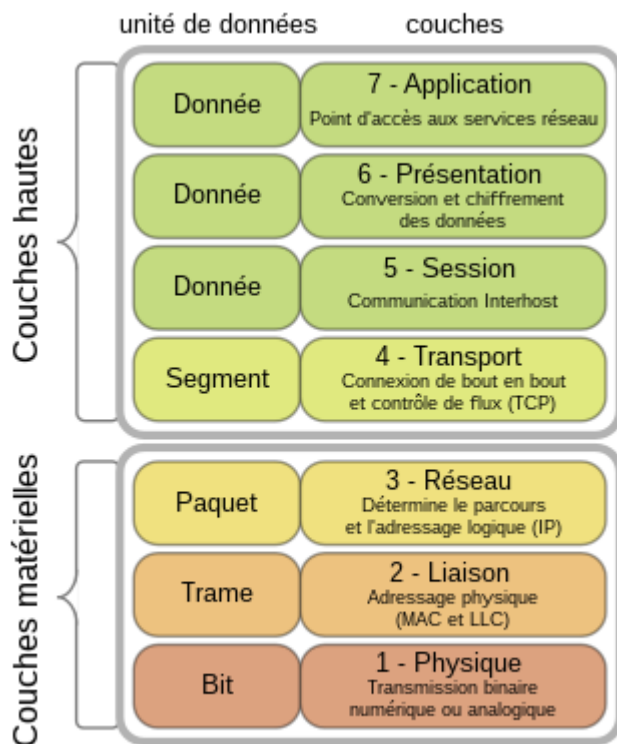
FONCTIONNEMENT DU RÉPARTITEUR DE CHARGE



Le load balancing peut être réalisé de différentes manières, mais le principe de base est le même. Il s'agit de placer plusieurs serveurs derrière un équilibreur de charge (load balancer en anglais) qui se charge de répartir équitablement la charge de travail entre ces différents serveurs. Les requêtes sont acheminées vers le serveur le moins occupé à ce moment-là, de sorte que chaque serveur est utilisé de manière égale.

Il existe plusieurs types de load balancing, notamment le load balancing de niveau réseau, le load balancing de niveau application et le load balancing de DNS. Le load balancing de niveau réseau

est effectué au niveau de la couche réseau, tandis que le load balancing de niveau application est effectué au niveau de la couche applicative. Le load balancing de DNS utilise des enregistrements DNS pour répartir la charge de travail entre plusieurs serveurs.



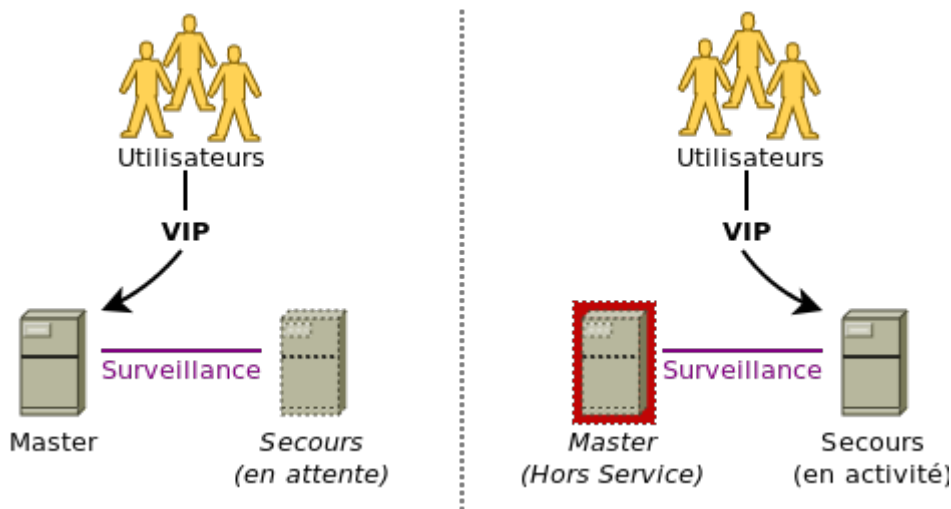
La haute disponibilité:

La haute disponibilité, quant à elle, se réfère à la capacité d'un système à rester disponible et à fonctionner correctement même en cas de panne d'un ou plusieurs composants. Pour assurer la haute disponibilité d'un système, il est nécessaire de mettre en place des mécanismes de redondance et de reprise sur panne.

Il existe plusieurs techniques pour assurer la haute disponibilité, notamment la mise en place de serveurs redondants, la réplication de données et la sauvegarde régulière des données. L'objectif est de minimiser les temps d'arrêt en assurant la disponibilité continue du système, même en cas de panne.

Haute Disponibilité - mode Actif / Passif

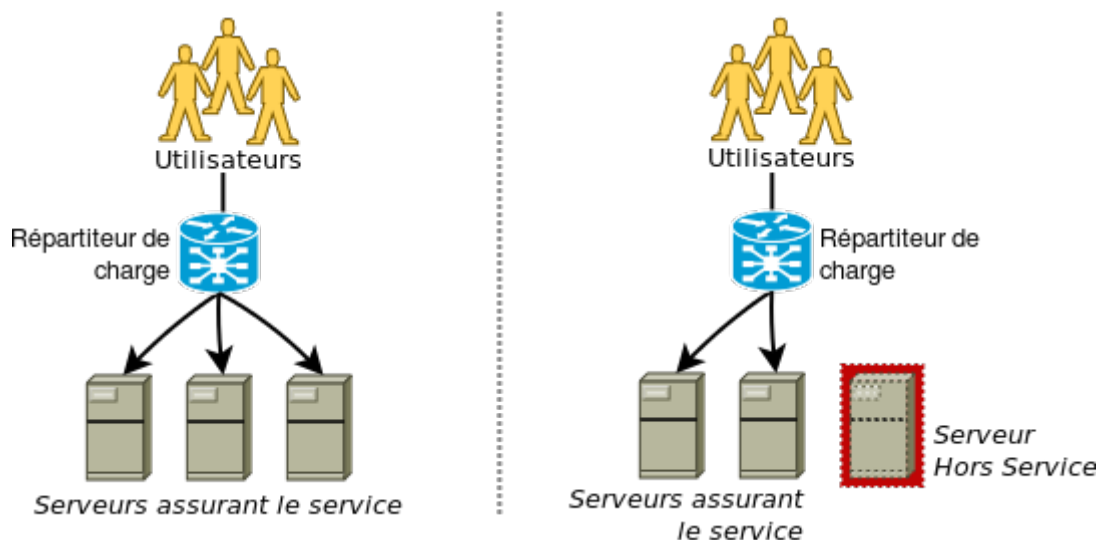
Pour passer en mode Haute Disponibilité Redondé afin d'accéder à un SLA de 99,9% garantie à l'année, nous proposons une redondance de votre serveur (et de vos services) sur un mode Actif / Passif.



Après étude, nous nous occupons de la mise en place de votre architecture et nous effectuons la recette des tests de bascule et des différents indicateurs de supervision avec vous.

Haute Disponibilité - mode Répartition de Charge

Pour vos installations nécessitant à la fois de la Haute Disponibilité Redondée et de Hautes Performances pour répondre à d'importantes sollicitations, en pique ou en continu, nous vous proposons des architectures avec Répartition de Charge.



Un équipement Levelsys (en mode mutualisé ou dédié) situé en amont de vos serveurs, attribut les différentes tâches / connexions à vos serveurs suivant une politique prédéfinie.

Si un de vos serveurs n'est plus en mesure d'assurer le service, notre équipement le détecte, ne lui envoie plus de tâche et répartit les tâches qui lui étaient précédemment attribuées sur l'ensemble des serveurs restant.

Cette solution vous apporte une grande souplesse :

Vous pouvez simplement monter en charge et accompagner ainsi la croissance de votre activité et intégrer les futurs coûts de votre infrastructure SI dans votre projet. Vous pouvez piloter

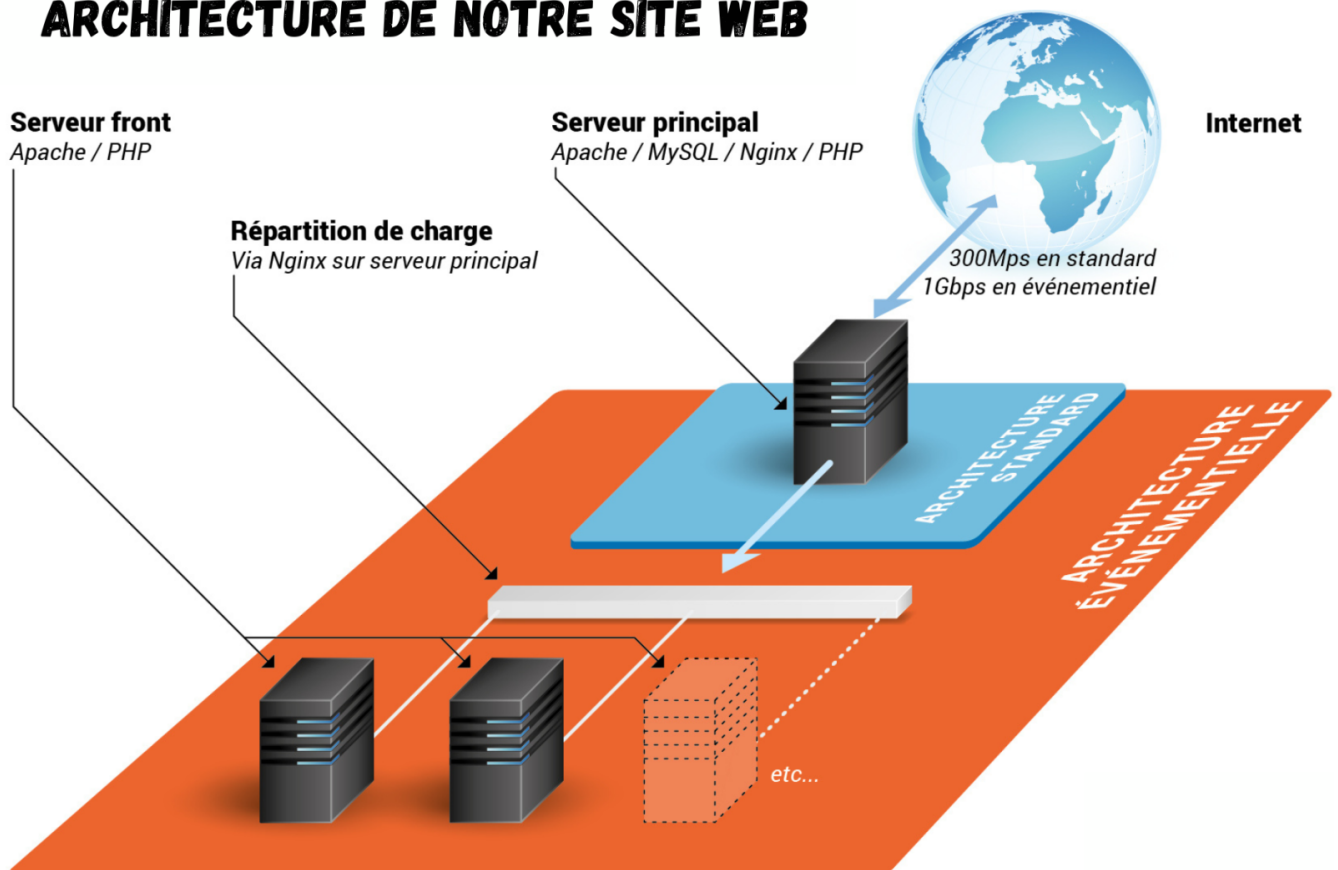
intuitivement et adapter l'architecture de votre plateforme pour absorber une charge passagère due à votre activité grâce à notre plateforme IaaS avec autodéploiement intégré automatiquement dans la répartition de charge).

Cas pratique

Pour mieux comprendre les différences entre le load balancing et la haute disponibilité, prenons l'exemple d'un site Web de commerce électronique. Supposons que ce site Web reçoive un grand nombre de requêtes simultanées pendant les périodes de pointe, comme les soldes ou les vacances de fin d'année. Si un seul serveur était utilisé pour traiter toutes ces requêtes, le site pourrait devenir indisponible ou ralentir considérablement en raison de la surcharge.

Avec le load balancing, plusieurs serveurs sont utilisés pour répartir équitablement la charge de travail. Ainsi, chaque serveur est utilisé de manière égale, ce qui permet de traiter un plus grand nombre de requêtes simultanées. Le site reste donc disponible même pendant les périodes de pointe.

ARCHITECTURE DE NOTRE SITE WEB



Cependant, même avec le load balancing, il est possible que l'un des serveurs tombe en panne ou rencontre un problème. Dans ce cas, la haute disponibilité entre en jeu. Si le système est configuré pour être hautement disponible, les utilisateurs peuvent être redirigés automatiquement vers un serveur de secours en cas de panne. Ainsi, le site reste disponible même en cas de panne d'un serveur.

En résumé, le load balancing et la haute disponibilité sont deux techniques utilisées pour garantir la disponibilité des services.

Revision #3

Created 20 February 2023 13:06:40 by edoppel

Updated 20 February 2023 14:34:55 by edoppel