

La haute disponibilité des services

Les solutions de haute disponibilité

Les solutions sont nombreuses et dépendent de l'application cliente ou serveur :

- utilisation d'un protocole applicatif garantissant la qualité des échanges de données ;
- la réplication des données ;
- la redondance des services ;
- la redondance des serveurs ;
- la redondance de serveurs partageant la même information mais situés sur des réseaux IP différents ;
- la redondance de serveurs par équilibrage de la charge réseau ou la mise en cluster NLB (Network Load Balancing) ;
- le cluster failover ou cluster à tolérance de pannes ;
- l'utilisation de serveurs en attente (Standby ou log shipping) ;
- l'utilisation de miroirs ;
- l'utilisation des solutions de virtualisation.

La réplication des données

Le principe est de mettre en place une solution qui garantit que les mêmes données sont répliquées entre des serveurs qui utilisent le même service. Chaque serveur dispose de sa propre copie des données et communique aux autres les modifications qui le concernent de telle sorte que ces modifications soient répercutées sur l'ensemble des **réplicats**.

La redondance de serveurs

Le principe de la redondance de serveur est de permettre à un service de pouvoir fonctionner sur plusieurs serveurs. Chaque serveur délivre le service indépendamment des autres et peut pallier les dysfonctionnements d'un autre serveur. Par exemple le service DHCP peut être configuré en plaçant un serveur DHCP dans chaque sous-réseau afin de délivrer des adresses IP pour le sous-réseau concerné mais peut aussi contenir des étendues pour les autres sous-réseaux. En cas d'indisponibilité, d'un serveur DHCP dans un sous-réseau, c'est un autre serveur DHCP qui attribuera les adresses IP pour l'étendue concernée. On garantit de cette manière la continuité du service.

La redondance de serveurs partageant la même information mais situés sur des réseaux IP différents

Les services utilisant la réplication gère souvent aussi la redondance de serveur afin que les clients continuent à disposer du service lorsque le serveur qu'ils sollicitent habituellement est indisponible. Ces serveurs sont situés sur le même réseau ou sur des réseaux IP différents. Cela garantit la continuité du service mais également la répartition de la charge. Pour reprendre l'exemple précédent, les clients Active Directory sont capables de contacter l'un après l'autre les contrôleurs de domaine jusqu'à ce qu'ils puissent valider leurs ouverture de session. De même, un client DNS sollicitera le serveur DNS primaire puis le DNS secondaire si nécessaire.

La redondance de serveurs par équilibrage de la charge réseau ou la mise en cluster NLB

Un cluster NLB est un système logique composé d'un ensemble de serveurs physiques appelés nœuds ou hôte. Ce système redirige la requête du client vers le nœud disponible de manière transparente pour le client. Le cluster NLB dispose de sa propre adresse IP et sa propre adresse MAC et garantit la continuité de service en gérant la répartition de la charge sur l'ensemble des nœuds.

Utilisant sur du matériel spécifique ou des solutions logicielles comme Vcenter de VMware ou les solutions Serveur de Microsoft, le cluster NLB peut aussi s'adapter, sous certaines conditions, à des services qui n'ont pas été conçus pour cela.

Peuvent fonctionner sur des clusters NLB des applications Web, FTP, VPN, etc.

Le cluster failover ou cluster à tolérance de pannes (cluster de basculement)

Le principe est proche de celui du cluster NLB. Le cluster **failover** est constitué d'un ensemble de nœuds mais la différence est que la requête du client est redirigée de manière transparente vers le nœud actif. Cela signifie que les nœuds composant le cluster sont soit actifs soit passifs. Le nœud passif est en attente d'une indisponibilité du nœud actif. Cela nécessite que les nœuds doivent partager les mêmes

données qui sont alors situées sur un volume réseau distinct des nœuds, en général un espace de stockage à haut débit comme un SAN.

- Le terme **heartbeat** est employé et signifie que le nœud passif envoie à intervalles réguliers une demande pour savoir si le nœud actif est toujours disponible.
- Le terme **failover** désigne le processus par lequel le nœud passif prend le contrôle du cluster après avoir déterminé l'indisponibilité du nœud actif.
 - Le terme **failback** désigne le processus par lequel le nœud indisponible redevient disponible et fait basculer les services afin que le cluster soit de nouveau dans l'état d'origine.

L'utilisation de serveurs en attente

Il s'agit d'avoir un serveur en attente, en général pour un service spécifique. Il sera manuellement mis en œuvre en cas de défaillance du serveur principal. C'est une solution moins onéreuse que le cluster mais moins efficace car il y a un temps de latence dans la mise du basculement et donc un risque de perte d'activité durant cette période.

L'utilisation des solutions de virtualisation

La virtualisation offre de nouvelle solution de haute disponibilité en permettant d'avoir des sys-tèmes plus réactifs. La machine virtuelle qui héberge le service est aisément déployable d'un hôte à l'autre car il s'agit d'un déplacement de fichier. Cela est facilité si les données sont hébergées sur un volume partagé de type SAN, SAS ou iSCSI. Le temps d'indisponibilité peut être très court, le temps de provisionner la VM avec une copie de sauvegarde du fichier ou la restauration d'un snapshot.

Il est bien sur possible de gérer des machines virtuelles en cluster failover.

From:

/ - **Les cours du BTS SIO**

Permanent link:

</doku.php/sisr3/c1>

Last update: **2014/10/09 15:42**

