

La gestion de capacité des infrastructures virtuelles, un enjeu clé (avis d'expert)

L'avènement de la virtualisation a permis de s'affranchir des problématiques liées à la gestion du matériel informatique ; tout du moins, en apparence. Les demandes de provisionnement de serveurs virtuels explosent, les délais de livraison chutent drastiquement et la qualité des solutions de virtualisation semble pouvoir permettre une mutualisation infinie des ressources. Or la charge de l'infrastructure est complexe à contrôler. La constatation d'une surcharge est bien souvent le signe d'une dégradation générale avancée. La charge globale de l'infrastructure virtuelle se reporte alors inexorablement, et bien souvent de manière inégale, sur les applications qu'elle héberge. L'infrastructure virtualisée est régulièrement considérée comme une entité indépendante des services qu'elle fournit. Cependant, socle de virtualisation et services applicatifs fournis sont bien plus liés qu'il n'y paraît.

Partir sur de bonnes bases

Les bases d'une infrastructure virtualisée saine sont établies dès sa conception. Une des premières étapes assurant la future maîtrise de l'infrastructure est la définition de sa place au sein du système d'information et sa puissance globale. Il conviendra tout d'abord d'identifier les ressources nécessaires à l'exécution de l'ensemble des applications qu'elle hébergera. L'infrastructure virtualisée étant par essence une entité hautement mutualisée, la seconde étape de sa conception est la planification de sa mise en œuvre. Il s'agit d'assurer une bonne répartition de la charge sur l'ensemble des hyperviseurs et de pouvoir supporter la perte d'un, ou plusieurs, de ses équipements sans impacter de manière non acceptable l'ensemble du service rendu. On préférera donc utiliser des équipements plus modestes en termes de puissance brute mais disposant de nombreux cœurs de calcul et périphériques de communication. Enfin, il ne faut pas oublier qu'une infrastructure virtuelle ne s'auto-suffit pas. Elle est, dans la grande majorité des cas, couplée à un système de stockage mutualisé (SAN, NAS...) et connectée au réseau d'entreprise. Il faudra donc rester cohérent dans le dimensionnement de ces éléments afin que l'un ne diminue pas les performances globales des autres.

La gestion de la capacité pour assurer la maîtrise de son infrastructure

L'importance de l'activité de surveillance ne doit pas être minimisée. Comme nous l'avons développé dans notre [second article](#), il s'agit d'identifier les métriques à surveiller, de déterminer les seuils définissant un comportement acceptable ou non et de mettre en place différents niveaux d'alertes permettant de suivre l'évolution de l'infrastructure dans son ensemble. D'une manière générale, il faut suivre les évolutions, appréhender les changements et ne surtout pas subir la montée en charge de l'infrastructure. La production régulière de rapports et tableaux de bord permet de justifier de la bonne santé de l'infrastructure. Le cas échéant, ces mêmes rapports

permettront de légitimer la nécessité d'investir dans une mise à jour de l'infrastructure.

La maîtrise du placement des ressources applicatives est le second axe lié à la gestion de capacité. Il faut alors prendre un peu de recul et comprendre le fonctionnement des applications ; en identifier les impacts sur l'infrastructure, indépendamment et en tant qu'ensemble. Cette compréhension permettra, en mutualisant les profils applicatifs complémentaires, de limiter l'impact de la charge d'une application vis-à-vis d'une autre. De même, la connaissance du fonctionnement technique des applications permet de s'assurer de latences d'accès minimales entre les différentes entités d'une application ; par exemple, en gardant « proche » une application et ses données. Il conviendra cependant de ne pas sacrifier à la recherche du temps de latence minimal la capacité d'une application à supporter la perte d'un élément de l'infrastructure qui la supporte. On devra, par exemple, préférer la localisation d'éléments d'un même cluster applicatif sur plusieurs des nœuds de virtualisation déployés ; le risque d'arrêt d'une application complète en cas de la perte d'un hyperviseur est alors fortement réduit.

Définir des classes de service

Une erreur fréquente est de penser qu'il suffit de déployer ses machines virtuelles et que le système de gestion de l'infrastructure fera automatiquement en sorte que tout fonctionne de manière optimale. D'une manière générale, et dans les limites basses de l'utilisation des ressources disponibles, cette assertion est vraie. Cependant, les fonctionnalités de distribution automatique des ressources et de placement de VM peuvent être aussi pratique que destructrices en terme de qualité de service si elles ne sont pas bien configurées. Pour cela, il convient de définir des classes de service ; plus ou moins complexes selon la taille de l'infrastructure. Il ne s'agit cependant pas de réserver des ressources figées de l'infrastructure pour des applications que l'on juge critiques. Sauf dans quelques cas bien précis, cela pose plus de problèmes que cela n'en résout. Il conviendra plutôt de définir des niveaux de priorité entre différents groupes d'applications. Le moteur d'ordonnancement de la solution virtuelle sera alors en mesure de rapidement faire les « bons » choix pour assurer que les machines virtuelles prioritaires ont toujours un accès privilégié aux ressources de l'hyperviseur sur lequel elles s'exécutent.

L'apparente facilité de gestion d'une infrastructure virtuelle ne doit pas faire oublier que c'est un élément du SI d'autant plus critique qu'il en occupe une place centrale. La bonne santé des applications qui s'y exécutent n'est assurée que par une maîtrise de bout en bout de l'infrastructure et une surveillance permanente de celle-ci. Une infrastructure virtuelle maîtrisée saura se faire oublier au profit des services qu'elle fournit. C'est d'ailleurs l'une des fonctionnalités plébiscitées du Cloud Computing : être un fournisseur de services informatiques ; d'applications hautement disponibles, performantes, sécurisées et maîtrisées.

En complément :

[Avis d'expert : Pour un meilleur dimensionnement des applications et des infrastructures](#)