

Puissance à la demande : promesses et défis. (Paru dans LuxBox Magazine Mai 2004)

C'est sous ce nom que se cache la stratégie transversale qui devrait conduire à court terme à une modification complète de l'offre des plus grands acteurs du marché. « On Demand Computing » chez IBM, « Adaptive Enterprise » chez HP, « N1 » chez SUN,...se basent sur le constat que la majorité des infrastructures informatiques sont actuellement sous utilisées. Le but initial est de permettre aux entreprises de ne payer que les ressources réellement utilisées. Les technologies porteuses sont les grappes de serveurs, l'architecture multiprocesseur et les grilles de calcul.

Un cas concret :

La société Lego a vu les coûts de gestion de son infrastructure informatique diminuer d'un tiers suite à la mise en place d'un contrat « On Demand » avec IBM. L'activité cyclique de Lego en fait une cible idéale, auparavant cette société aurait dû tailler son infrastructure de manière à supporter les demandes maximales quitte à ce que la majorité de ces ressources soit sous-utilisée durant les autres périodes. La solution implémentée permet à Lego d'ajuster ses ressources informatiques aux besoins réels en fonction de la demande.

Les clefs de facturation :

Avant toute mise en production, une analyse détaillée des besoins permet l'établissement de l'indice 100 servant de base au calcul de l'indemnité mensuelle et à l'établissement de paliers de facturation. Les données relatives à la consommation réelle, collectées par un logiciel spécifique sont, dans la plupart des cas, accessibles en temps réel et déterminent le palier de facturation relatif à la période couverte.

Les constructeurs sont également capables d'activer ou désactiver à distance des ressources (processeurs, disques,...) entraînant une facturation selon un barème établi lors de la signature du contrat.

Quelles sont les difficultés liées à ces solutions :

Selon les analystes du MetaGroup on peut relever plusieurs inconvénients majeurs.

1. La mensualité minimale est souvent assez élevée et la négociation demeure difficile.
2. Les fabricants n'offrent que rarement la possibilité d'inclure l'infrastructure préexistante.
3. Cette solution recueille difficilement l'appui des responsables informatiques car perçue comme une solution d'externalisation conduisant à une précarité de leur emploi.

Une question reste ouverte : « Comment les créateurs de logiciels vont-ils adapter leur politique de licence à cette nouvelle architecture ? » Affaire à suivre...

Le SMP ou « Symmetric multiprocessing » et les grappes ou « cluster » :

L'ajout systématique de serveurs n'est pas la solution idéale pour faire face à des demandes de ressources informatiques.

L'utilisation de plusieurs processeurs au sein d'un même serveur ou la création de grappes de serveurs permet d'augmenter les capacités d'une infrastructure de calcul. La combinaison de ces deux technologies a permis à un cluster d'atteindre 35,9 Teraflops soit un trillion d'opérations/sec ou 35 fois la performance d'un processeur Pentium IV - 3 GHz.

Les serveurs SMP tournent une seule instance du système d'exploitation et partagent le même IO système et la même mémoire de manière parallèle. Il va de soit que pour profiter pleinement de cette architecture il est indispensable que les applications soient optimisées pour le parallélisme et conçues sous la forme de module de traitement élémentaires (threads). Cette solution est assez coûteuse car elle exige une grande quantité de mémoire vive et nécessite le déploiement initial de serveurs « haut de gamme » capables de suivre l'évolution de la solution.

Le « clustering » est une technologie moins coûteuse consistant à ajouter des serveurs à la grappe en fonction des besoins, de plus elle offre des services de redondance/reprise de pannes.

Supportée par le monde Open Source, elle permet de créer des grappes sur base d'ordinateurs conventionnels moins coûteux.

Le SMP est destiné aux applications supportant le partage de tâches et ayant besoin de communiquer entre elles de façon très rapide. Dans ce cas, seule la capacité à traiter les données sur un seul et même système permet d'optimiser les performances.

Les grilles de calcul ou « Grid Computing » :

On appelle grille de calcul, un ensemble de machines mutualisant leurs ressources au travers d'un réseau. L'acronyme GRID est utilisé en français pour signifier « Globalisation de Ressources Informatiques Distribuées ». Dassault Aviation, et Arcelor utilisent cette technique de mutualisation de leurs ressources de calcul.

Cette architecture logique permet à un gestionnaire d'allouer des ressources à une application ou un groupe d'utilisateurs en fonction de critères tels que les créneaux horaires, les consommations de ressources locales,... Le logiciel de gestion de la grille permet le partage de ressources hétérogènes (pc, serveurs,...) au travers de groupes d'utilisateurs ou d'applications tout en garantissant la sécurité, l'intégrité et la confidentialité des données.

Afin d'assurer un bon rendement à cette solution il est indispensable de s'assurer que le réseau d'interconnexion offre un taux de latence aussi limité que possible.

Conclusion :

Le « Computing On Demand » est la réponse du marché à un besoin actuel d'optimisation de l'usage des ressources informatiques. Que cela passe par l'offre d'externalisation initialisée par les fabricants ou par le déploiement interne des technologies décrites ci-dessus, il semble inévitable de recadrer l'infrastructure informatique dans son rôle d'outil de production au service du métier.

DELBRASSINE Charles
IT WORKS SA Luxembourg
www.itworks.lu