

Une mission à chaque étage

L'affichage

Le premier niveau est composé d'une batterie de serveurs Web légers, de répartiteurs de charge, de serveurs de cache. Ils ont pour fonction d'assurer l'affichage des pages HTML tout en l'accéléralant.

La construction

Au deuxième niveau, les serveurs d'applications se chargent de générer dynamiquement les pages Web après consultation des bases de données et en fonction des requêtes reçues par les serveurs Web.

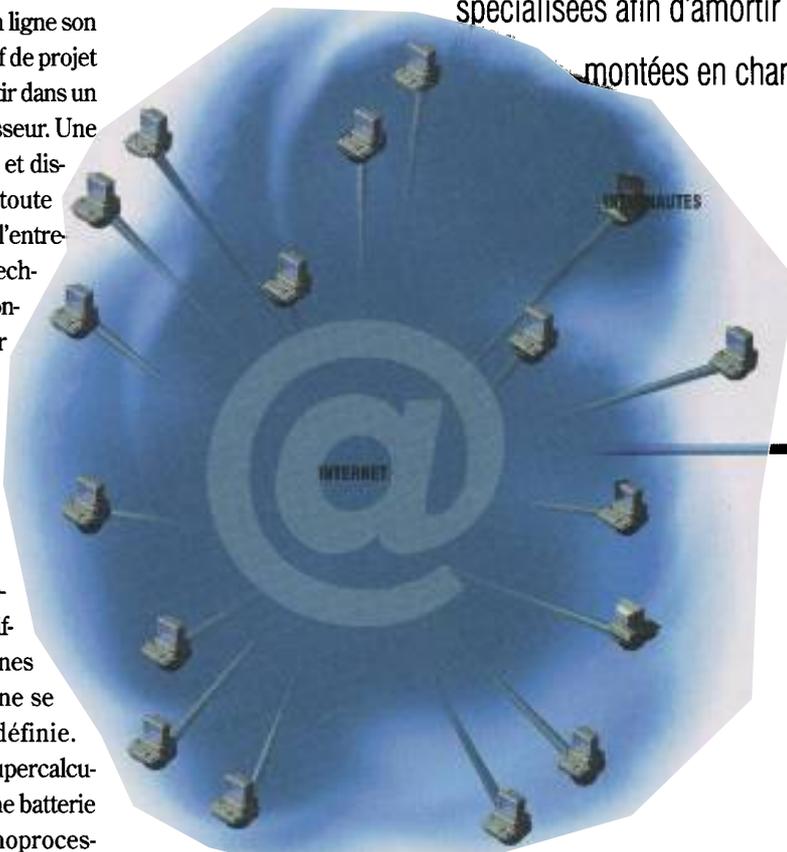
Le stockage

Le dernier niveau de cette infrastructure Internet héberge les SGBD (tous ceux du marché). Le stockage est laissé à des serveurs dédiés.

Trois de serveurs se

En informatique comme ailleurs, l'union fait la force. Les serveurs se multiplient sur différentes couches spécialisées afin d'amortir les montées en charge.

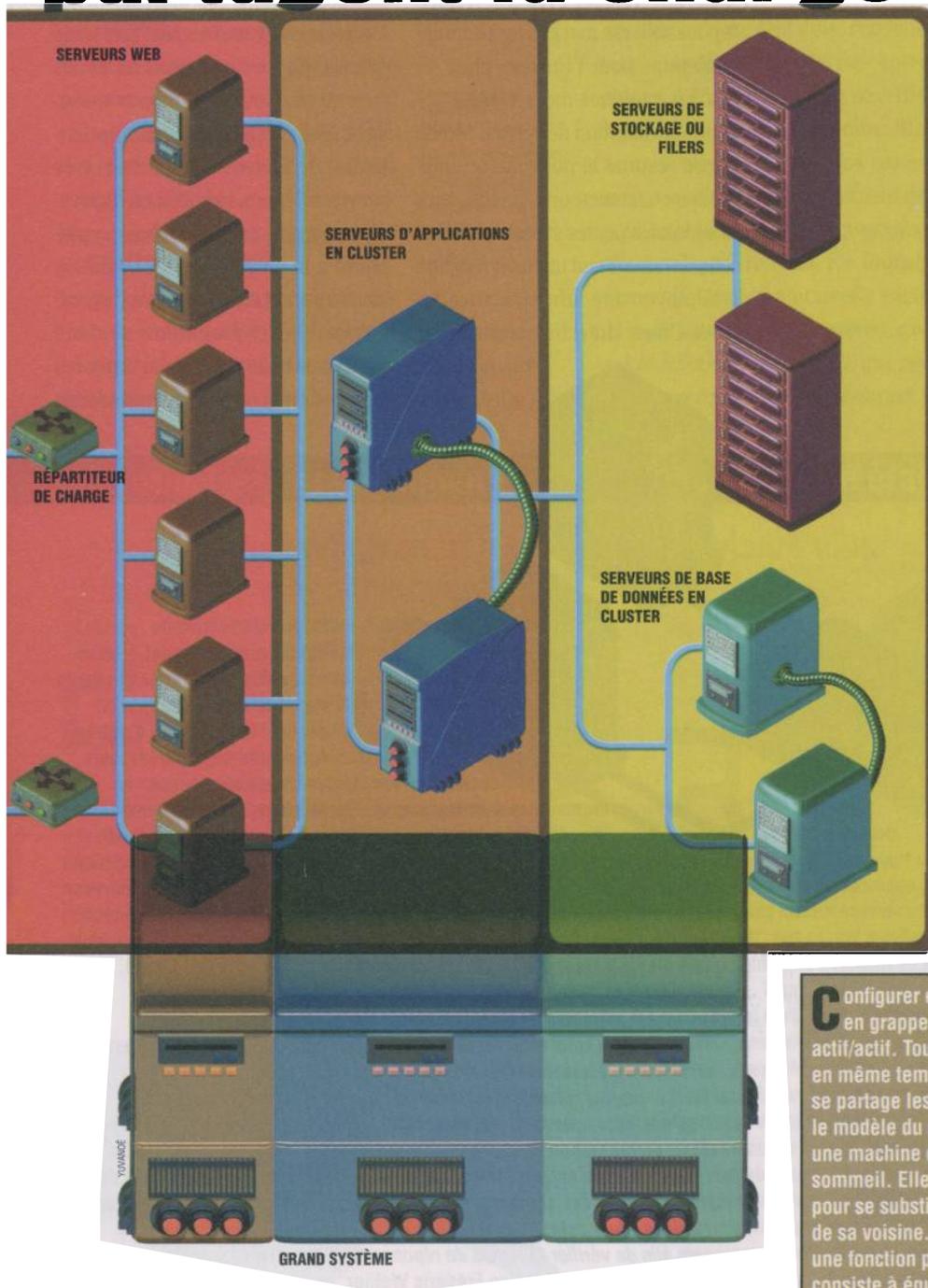
À l'heure de mettre en ligne son site Internet, le chef de projet peut choisir d'investir dans un super-serveur multiprocesseur. Une machine alliant puissance et disponibilité et concentrant toute l'architecture Internet de l'entreprise. Une merveille de technologie qui, une fois partitionnée, accueille le serveur Web, la base de données et le serveur d'applications. Il peut aussi opter, et c'est la tendance du jour, pour une infrastructure se déployant sur trois étages de serveurs. L'idée étant de répartir la charge entre les différents niveaux de machines complémentaires, chacune se chargeant d'une tâche définie. Ainsi, en frontal, pas de supercalculateurs dernier cri, mais une batterie de petites machines monoprocesseurs faisant office exclusivement de serveurs Web. Plus bas dans l'infrastructure, deux niveaux hébergent des applications plus critiques exigeant le déploiement de machines puissantes, « des machines multiprocesseurs, plus traditionnelles que celles déployées en frontal. Implémentées en cluster, elles offrent une puissance de calcul capable d'accueillir les serveurs



Trois étages de serveurs au lieu d'un super-serveur multiprocesseur

En frontal, plusieurs serveurs légers HTTP sont déployés pour traiter exclusivement l'envoi et l'affichage des pages Web. Ce modèle est rendu possible par l'ajout de répartiteurs de charge qui équilibrent les flux IP de données entre les différentes machines et veillent ainsi à ce qu'aucun des serveurs ne soit en surrégime. Au final, la ferme de serveurs Web se comporte comme une machine unique fragmentée. Cependant, si un serveur HTTP s'arrête, les éléments restants continuent à répondre aux requêtes des internautes. Quant aux serveurs de cache, ces machines dédiées allègent la charge du frontal Web en stockant en mémoire les pages les plus demandées. Les serveurs d'applications (voir IFM n°112, p.51) fonctionnent en parallèle pour se répartir la construction des pages dynamiques avant de les envoyer aux serveurs Web. Autre nouveauté dans les architectures Internet, l'arrivée des serveurs dédiés exclusivement au stockage. Ils permettent d'alléger la tâche des serveurs de base de données et proposent des dispositifs de sécurité.

étages partagent la charge



d'applications», explique Christophe Lauer, consultant responsable R&D chez Techmetrix Research. Enfin, à la base du système se trouvent les serveurs de bases de données. Configurés en grappes, ils se différencient par des capacités de mémoire étendues pour répondre aux contraintes SGBD du marché. En revanche, par rapport aux serveurs de données classiques, elles n'accueillent plus de systèmes de stockage. Ces derniers sont autonomes, prêts à monter en puissance en cas de besoins accrus en archivage, ils font office d'auxiliaires aux SGBD.

Chaque pièce du puzzle est préconfigurée et dédiée exclusivement à une tâche. En cas de panne, les clones de serveurs se partagent la tâche du serveur défaillant, tandis que d'autres se tiennent prêts à prendre le relais en cas de montée en charge. On évite ainsi la gabegie de puissance en sous employant un grand système coûteux. La tendance est donc à investir moins et mieux en réalisant une architecture modulaire et réactive dimensionnée au plus près des besoins de monter en charge d'un site. Mais attention, même bien armée, cette infrastructure ne supportera pas à coup sûr des pics extrêmes, il faut donc connaître ses limites.

Cluster et répartition de charge

Configurer en cluster consiste à mettre des serveurs en grappe. On parle aujourd'hui de cluster actif/actif. Tous les serveurs travaillent en parallèle, en même temps, pour construire les pages Web. On se partage les ressources des machines. À l'inverse, le modèle du cluster actif/passif, plus ancien, déploie une machine de secours identique, qui reste en sommeil. Elle ne se réveillera qu'en cas de panne pour se substituer physiquement aux défaillances de sa voisine. Quant à la répartition de charge, c'est une fonction propre aux commutateurs IP. Leur fonction consiste à équilibrer les flux IP, afin que les serveurs Web se partagent le flux des requêtes des internautes.