

Avis d'expert : les origines du big data

Cet avis d'expert est signé **Mike Lynch**, fondateur et président-directeur général d'Autonomy et vice-président exécutif de la division *information management* de HP

Le *big data* est devenu sans conteste le sujet du moment : nous sommes incroyablement inondés de vastes flux d'information. Nous produisons de plus en plus de données chaque jour, à des débits de plus en plus rapides. Collectivement, nous produisons chaque jour 2,5 quintillions d'octets de données et le taux de croissance est si élevé que 90 % de l'information jamais créée jusqu'alors a été produite au cours des deux dernières années. Nous constatons également une plus grande variété d'information qu'auparavant sachant qu'un ensemble de données peut se composer à lui seul de vidéo, audio, texte, données de capteur...

Alors que les ordinateurs sont devenus des outils utiles aux entreprises, le premier défi était pour elles de pouvoir stocker et catégoriser leurs données. Et en raison des limites technologiques, la base de données était née. L'étape suivante allait être de commencer à réellement exploiter ces données. Les bases de données consistant en des rangées et des colonnes, les ordinateurs étaient en mesure d'interagir avec elles et de les consulter. Et c'est parce que l'ordinateur savait que la colonne 3, ligne 6 correspondait au nombre d'ours en peluche stockés dans l'entrepôt, qu'il savait qu'il allait devoir passer une nouvelle commande dès lors que le stock allait descendre à zéro. C'est ici que les entreprises allaient pouvoir créer de la valeur ajoutée dans notre monde réel de l'information : en insérant ces informations dans une base de données, les ordinateurs allaient pouvoir commencer à les automatiser et les analyser.

Notre dernier défi en date : celui de gérer le volume, la quantité et la vitesse de l'information que nous produisons aujourd'hui. Avec une quantité accrue de données, nous sommes face à une plus-value potentielle, ce qui nous a conduits à attaquer le problème en trouvant une solution avec une base de données 'intelligente' capable de faire face à cette explosion des données. Ces bases de données existent depuis si longtemps que nous avons finalement oublié le problème d'origine. Car même avec la création de la base de données dite intelligente capable de traiter d'innombrables quantités d'informations à grande vitesse, il était impossible de résoudre le problème dans son ensemble.

Nous cherchons au mauvais endroit

On se réfère souvent au *big data* pour parler des données non structurées, ce qui a finalement abouti à créer la confusion entre les deux termes. Les données non structurées peuvent être volumineuses en tant que telles (une image est bien plus lourde que des chiffres ou des mots mémorisés dans une base de données typique). Les entreprises sont désormais face à un besoin croissant de récupération et d'utilisation de données qui ne sont pas simplement composées de texte ou de chiffres et elles doivent par ailleurs répondre de plus en plus à des demandes d'analyses en temps réel de ces données.

Traditionnellement, on considérait qu'en fonction du volume et de la vitesse des données vous étiez en mesure de les analyser. Aujourd'hui, ce n'est plus pertinent. Il est très important de garder à l'esprit que les entreprises qui sauront intégrer et exploiter les données non structurées ou « informations humaines » (vidéo, image, *email*, SMS, contenus provenant des réseaux sociaux...)

pourront en tirer un réel avantage concurrentiel. C'est le problème le plus important à résoudre et il est bien plus important que celui de savoir comment gérer le *big data*.

Vous pouvez pallier des problèmes d'évolutivité liés au *big data* en investissant toujours plus dans des solutions logicielles et matérielles. Mais l'information humaine représente aujourd'hui 90 % de l'ensemble des données d'une entreprise, ce qui signifie que si vous les ignorez et êtes uniquement à la recherche de « la base de données parfaite », vous n'allez traiter que 10 % de l'information. Plutôt que de se focaliser uniquement là-dessus, il est important d'avoir une vue d'ensemble.

De quelle manière le big data peut-il apporter de la valeur ajoutée ?

Pour mieux comprendre à quel point l'information humaine peut permettre de tirer de la valeur ajoutée, mettons-nous dans la peau d'un responsable de site de commerce électronique. Lors de sa création, le *e-commerce* était tout simplement un point de vente complémentaire pour les commerçants traditionnels ; puis c'est devenu un canal de distribution à part entière incitant les consommateurs à y acheter produits et services. Désormais, la disponibilité de grandes quantités de données sur des prospects ou clients offrent à l'entreprise un énorme potentiel dès lors qu'ils peuvent recueillir et analyser correctement ces informations. La navigation du client et ses habitudes d'achat engendrent une importante traçabilité des données et, couplés à l'intégration d'informations issues des réseaux sociaux, permet un ciblage efficace du contenu.

Mais le défi ici est de garder à l'esprit qu'avoir des dizaines, des centaines ou des milliers de clients potentiellement connectés les uns aux autres par des milliers de points de données engendre la collecte d'une masse d'information très importante. Une partie pourra être structurée en données facilement catégorisables (âge, sexe, zone géographique) mais la majorité sera constituée de données non structurées : ce sont des informations humaines. La teneur d'un avis consommateur et son ressenti sur un produit par exemple n'est pas quelque chose qui peut être aisément classé dans une base de données standard ; pourtant, la possibilité de tirer parti de ce genre d'information en l'analysant est vitale pour toute entreprise qui veut augmenter ses parts de marché.

Le meaning est la solution

Pendant qu'on s'accrochait tous à nos bases de données, le monde a changé. Le « problème » base de données est devenu un faux prétexte détournant l'attention sur la question essentielle : notre capacité à pouvoir gérer l'information humaine. Les bases de données ont été créées il y a un demi-siècle comme la solution pour pallier les limites des ordinateurs. Alors, à quoi bon attendre d'elles qu'elles nous fournissent la solution miracle dans un monde aujourd'hui complètement différent ?

La clé est de bien comprendre qu'en les laissant là où elles se trouvent, les données constituent une puissante plateforme d'analyse infiniment évolutive pour l'entreprise. Grâce au *meaning* (c'est-à-dire l'analyse de l'information contextuelle et conceptuelle par les ordinateurs) nous pouvons « trancher le nœud gordien » et aller au cœur du sujet. Nous serons alors capables de traiter le sens de 100 % de l'information, structurée et non structurée et de générer ainsi une réelle valeur ajoutée

dans l'entreprise.

En se donnant les capacités de comprendre ce qu'est le *meaning*, les entreprises peuvent tirer parti de 100 % de leurs données et non pas seulement de 10 % d'entre elles que sont les données structurées. Ce dont nous sommes capables aujourd'hui n'a jamais été possible jusqu'alors parce que nous avons oublié les principes fondamentaux liés à l'industrie informatique. C'est désormais aux ordinateurs de s'adapter à l'homme et non l'inverse. Nous avons besoin d'examiner le problème de plus près pour enfin réaliser que nous avons déjà la réponse.

Crédit photo : © Autonomy