

Origines et enjeux du big data

Le développement spectaculaire d'internet, des réseaux sociaux, de la technologie mobile et des objets connectés provoquent une croissance exponentielle des données, structurées ou non, à laquelle les entreprises sont confrontées : c'est le phénomène du « big data » traduit le plus souvent par le terme « mégadonnées ».

Les enjeux sont de taille, car le big data n'est pas qu'une question technique de volumétrie et de stockage. Il constitue au contraire l'opportunité de valoriser le contenu de ces nouvelles sources et d'en tirer profit. Encore récemment, l'analyse de l'information se faisait *a posteriori* selon un cycle en trois temps : collecte, traitement, analyse. Grâce aux nouvelles solutions du big data, les événements peuvent être traités au fil de l'eau. Mieux, les solutions d'analyse prédictive permettent désormais d'anticiper certains événements ce qui procure aux entreprises un avantage concurrentiel évident. Des modèles mathématiques appliqués aux mégadonnées sont ainsi utilisés pour connaître les attentes des clients, faire des prévisions météorologiques fiables, estimer les prix potentiels des actifs financiers, ou prévenir les épidémies. Et ce ne sont là que quelques exemples, car les domaines d'application sont vastes et variés.

Les métiers et les secteurs concernés

Si le marketing est certainement le domaine qui a le plus d'attentes, le big data s'adresse aussi à la plupart des autres fonctions de l'entreprise. Les algorithmes de plus en plus pointus permettent d'améliorer l'efficacité des fonctions production, logistique, achats, maintenance qui manipulent de grandes quantités de données. Le big data est aussi en train de transformer les pratiques RH au sein des organisations. Les possibilités de l'analyse prédictive trouvent en effet des applications concrètes dans les domaines du recrutement, de la gestion des risques et des performances des collaborateurs.

La plupart des secteurs économiques sont concernés. L'industrie, l'assurance, les services, la finance et la distribution font face à des masses de données toujours plus importantes. Dans le secteur de l'énergie, la mise en place de compteurs intelligents capables d'enregistrer des mesures tous les quarts d'heure permet ainsi d'optimiser la production, la distribution et la consommation.

En dehors du monde de l'entreprise, les techniques du big data apportent aussi des solutions à la recherche, à la médecine (prévention des épidémies, médecine personnalisée à distance...) et d'une manière plus générale à l'humanité.

Il est intéressant de souligner que les états ont pris la mesure du phénomène, puisqu'aux États-Unis comme en France et dans bien d'autres pays, des budgets ont été débloqués pour soutenir des projets de R&D portant sur les technologies liées à l'exploitation de très grands volumes de données.

Les technologies du big data

Si l'exploitation de gros volumes de données est à présent possible, c'est grâce à la technologie qui a évolué ces dix dernières années sous l'impulsion de sociétés issues du *web*. Les GAFAs (Google, Amazon, Facebook, Apple) ont clairement joué un rôle moteur dans la recherche de solutions pour traiter les mégadonnées, car ces pionniers ont été les premiers à être confrontés à de gigantesques volumes de données *via* leurs moteurs de recherche ou leurs sites internet.

Sans ces sociétés, les innovations techniques telles que Hadoop ou MapReduce n'auraient certainement pas vu le jour. Le fait d'avoir partagé ces innovations grâce à l'*open source* a grandement facilité l'expansion de ces technologies. Aujourd'hui, la plupart des grands noms du logiciel proposent des solutions big data reposant sur ces briques *open source*. Les améliorations sont continues aussi bien dans les plateformes permettant la parallélisation des traitements, que dans les bases de données (en mémoire, NoSQL...) ou encore dans les outils d'analyse et de visualisation des données.

Il est impossible de citer toutes les technologies du big data. Toutefois, l'écosystème Hadoop demeure très présent au sein des entreprises utilisatrices. On note aussi depuis quelques années la percée de la solution *open source* Spark qui semble très prometteuse.

Spark est reconnu pour :

- sa simplicité d'utilisation : Spark permet d'écrire rapidement des applications en Java, Scala ou Python. Outre les opérations de Map et Reduce, Spark supporte les requêtes SQL et propose des fonctionnalités de *machine learning* et de traitements orientés graphe ;
- ses performances : la plateforme peut faire tourner des requêtes 100 fois plus vite qu'avec MapReduce avec les données directement en mémoire, ou 10 fois plus vite si les données restent sur disque.

Le big data au cœur de l'expérience client

Avec les nouvelles possibilités du numérique, de nombreuses entreprises remettent au goût du jour le thème de l'expérience client, que les anglo-saxons intitulent *customer experience* (CX) ou *digital customer experience* (DCX). Cela consiste à développer une démarche « client d'abord », en prenant en charge la dimension émotionnelle et humaine qui relie les entreprises à leurs clients.

À l'ère du digital et de la *data*, l'expérience client se voit réinventée car le consommateur possède de plus en plus de supports numériques tels que des smartphones, des tablettes ou des objets connectés, ce qui a des répercussions sur le processus d'achat. De nombreuses entreprises, notamment B to C, développent des stratégies e-commerce pour intégrer les nouvelles possibilités technologiques dans leur *business model*. Le traitement des données collectées permet de mieux connaître chaque client et de personnaliser les offres. C'est ainsi que les algorithmes se basent sur les données issues de l'enregistrement des parcours client sur le web (clics, sites visités, etc.) pour proposer des produits ou services adaptés.

L'objectif n'est cependant pas de basculer toutes les activités physiques classiques vers le digital, mais de se repositionner face aux nouvelles tendances de parcours client. Ainsi, le point de vente physique demeure présent dans la relation client et le parcours d'achat. Il est intéressant de noter qu'Amazon fait le chemin inverse, puisque cet acteur purement digital à l'origine a ouvert sa première librairie physique à Seattle en novembre 2015. La stratégie digitale devient très nettement omnicanal afin de favoriser la synergie entre les multiples canaux (physique, e-shopping, réseaux sociaux) et d'apporter un maximum de confort au consommateur lors de son processus d'achat.

L'essor du *click and collect* est un très bon exemple de stratégie mêlant digital et physique. Ce mode de vente permet au client de réserver en ligne un ou plusieurs produits disponibles en magasin. Le service fonctionne à partir d'une page produit, sur laquelle le client sélectionne les produits qu'il souhaite réserver. Une fois son choix effectué, l'internaute est informé des magasins de proximité qui possèdent le produit dans leur stock. Le client indique alors le magasin dans lequel il souhaite se rendre pour retirer son produit. Sur ce principe, la grande distribution a développé le modèle *click and drive* permettant de faire ses courses sur internet et de récupérer ses produits dans un espace dédié ayant des pistes de stationnement. Fin 2015, 36 % des français ont déjà fait leurs courses via un *drive*.

Un peu partout des initiatives *web-to-store* facilitant le parcours d'achat en intensifiant la relation client voient le jour. Des enseignes comme Burberry, Mango, Darty ou Sephora installent ainsi des boutiques connectées. Lorsqu'ils ne trouvent pas leur bonheur en magasin, les clients sont invités à se connecter aux bornes disponibles dans les rayons pour passer leur commande et se faire livrer à domicile ou en boutique.

Et ça marche ! Le cabinet Forrester et la société Watermark Consulting ont mené une étude très intéressante montrant l'impact de la prise en compte de l'expérience client sur la performance boursière. Le résultat est sans appel : les organisations qui se concentrent sur l'expérience client surperforment (+ 35,2 %), tandis que celles qui sont en retard sous-performent (- 44,7 %).

Les compétences et les pré-requis

Un frein aux possibilités du big data est cependant présent : les ressources humaines. Il y a en effet besoin de former des architectes et des développeurs aux technologies de traitement des mégadonnées tels que Hadoop, Spark, NoSQL, SAP HANA, R, etc. Ce point devrait être relativement simple à régler, car des formations existent, et l'histoire montre que le monde de l'informatique s'est toujours adapté à l'évolution des technologies.

Des tensions relativement importantes se font aussi sentir pour les scientifiques de données, les fameux *data scientists* aux compétences mathématiques et statistiques pointues. Ces experts sont capables de donner du sens aux données grâce à des modèles mathématiques élaborés, notamment les modèles prédictifs. Des solutions pour pallier ce déficit sont dès à présent apportées, et on assiste depuis peu à une multiplication des formations dans ce domaine, notamment au sein des Grandes Ecoles. On peut aussi parier sur le fait que les outils qui seront proposés à l'avenir seront de plus en plus exploitables par des non-statisticiens.

Un peu partout, des projets big data sont engagés. Mais la maîtrise des méthodes et des outils ne suffit pas à en assurer le succès. L'entreprise doit avant tout entrer dans l'ère de l'information et placer les données au cœur de sa stratégie. La mise en place d'une gouvernance des données, avec notamment une politique affirmée de qualité des données, s'avère indispensable à la réussite de projets big data. Comment tirer profit des données massives provenant de nouvelles sources si l'on peine aujourd'hui à gérer les données existantes ? Comment intégrer de nouvelles données dans un paysage existant si celui-ci n'est pas réellement maîtrisé ? Enfin, à quoi bon faire des analyses prédictives si les données massives que l'on traite sont de mauvaise qualité ? Ces questions sont de plus en plus prises au sérieux, car les chances de sortir vainqueur de la compétition économique reposent sur la maîtrise de l'information. La transformation digitale des entreprises qui est en marche un peu partout est un signe encourageant. Dans la majorité des cas, un volet *data* est intégré à ces initiatives. Les données constituent sans aucun doute le matériau de base de la révolution numérique.

Conclusion : en route vers les smart data

Demain, la gestion des données se trouvera naturellement au centre de nombreux projets. Le volume et la diversité des données ne seront plus considérés comme un phénomène nouveau, et il y a fort à parier que le terme « big data » va disparaître.

C'est en partie déjà le cas, puisque l'on parle depuis quelques temps de *smart data*, c'est-à-dire de données intelligentes. Ce qui est important, c'est la valeur que l'on peut tirer des données, peu importe leur volume. Peut-être reviendrons-nous tout simplement au terme *data*, dès lors que la volumétrie croissante des données sera banalisée.

Aussi, nous nous dirigeons vers un monde de plus en plus connecté, avec des échanges de données qui s'intensifient. L'analyse de ces données progresse également avec le perfectionnement des algorithmes

et du *machine learning*. La frontière entre les capacités humaines et celles des machines se réduit, et les applications du *cognitive computing*, permettant aux machines d'interagir avec les humains, s'annoncent prometteuses. Certains futurologues vont plus loin, notamment avec le mouvement transhumaniste visant à doter les êtres humains de capacités surhumaines grâce à l'intégration des biotechnologies et des nanotechnologies.

Raymond Kurzweil, chef de file de ce mouvement et directeur de l'ingénierie chez Google, a ainsi annoncé lors de la conférence TED de mars 2014 à Vancouver que l'homme pourra télécharger son cerveau dans un ordinateur en 2030. Cela paraît fou, et quoiqu'il en soit nous devons réfléchir à la place de l'être humain vis-à-vis de l'intelligence artificielle, et d'une manière plus générale à l'équilibre à trouver entre les bienfaits de la technologie et les dérives potentielles.

Enjeux et usages du big data (2^{ème} édition)

de Christophe Brasseur

Editeur : Lavoisier / Hermes

Collection : Information numérique - Traitement, interprétation, communication



Cette nouvelle édition de ***Enjeux et usages du big data*** prend en compte les principales évolutions du big data et approfondit les nouveaux enjeux numériques parmi lesquels figurent l'analytique, l'expérience client, le cloud, les réseaux sociaux et les objets connectés. Les exemples et les cas ont été entièrement renouvelés afin d'être au plus près des préoccupations actuelles des entreprises. L'ouvrage met également l'accent sur les méthodes, les techniques et les ressources nécessaires pour permettre aux entreprises d'entrer avec succès dans l'ère de l'information à grande échelle.

Cet ouvrage concis et didactique s'adresse aux dirigeants d'entreprise, aux managers opérationnels et aux professionnels des systèmes d'information, ainsi qu'aux étudiants des écoles de commerce, d'ingénieurs et des universités, soucieux de comprendre les problématiques et les applications du big data.

Lien : [Enjeux et usages du big data](#)