

# Vertigineux "big data"

LE MONDE | 28.12.2012 Par Sarah Belouezzane et Cécile Ducourtieux

Photos, vidéos, sons, chiffres, textes... Depuis la démocratisation de l'Internet haut débit, ce sont des milliards de données qui sont déversées par les particuliers, les entreprises et même l'Etat sur la Toile. Les spécialistes appellent ça le "big data", littéralement "grosse donnée"... Ce vocable est un peu "marketing" : depuis dix-huit mois, pas une semaine ne passe sans un séminaire ou une nouvelle offre big data sur la planète high-tech. Mais il recouvre une réalité : la quantité de données envoyées, reçues, stockées toutes les secondes est devenue tellement importante qu'il est quasiment impossible de la traiter avec les outils d'analyse habituels.

Tous les jours, ce sont 118 milliards de mails qui sont envoyés à travers le monde, 2,45 milliards de contenus différents qui sont postés sur Facebook. L'opérateur de télécommunications américain AT & T transfère chaque jour 240 000 milliards d'octets de données ! Des chiffres qui donnent le vertige et qui vont encore exploser. A l'aube de l'an 2020, il y aura 10,4 zettaoctets, soit 10 400 milliards de gigaoctets de données déversés tous les mois sur Internet. 160 000 millions de milliards la capacité du plus cher des iPhone !

Les humains ne sont pas les seuls responsables de ce déluge d'informations. De plus en plus "connectées", les machines contribuent aussi à faire exploser les compteurs. Les stations pétrolières, les voitures et les compteurs électriques peuvent être dotés de capteurs ou de cartes SIM. Pour communiquer des informations sur leur environnement, la température... Certaines entreprises proposent même d'équiper les réfrigérateurs et les machines à café de cartes SIM pour les connecter aux réseaux mobiles. Histoire d'être informé à temps si le lait ou les dosettes de café venaient à manquer. Il y a aussi les données "libérées" par les pouvoirs publics : horaires de transports en commun, statistiques sur les populations...

## "ON EXPLOITE DÉSORMAIS TOUTES LES DONNÉES SUR UN CLIENT"

"Cela fait quelque temps déjà que le nombre de données produites et stockées est important, note Patrice Poiraud, directeur business et analyse chez IBM. Le phénomène a pris une ampleur sans

précéder ces deux dernières années." Cette accélération est principalement due à un changement dans nos habitudes. L'apparition des réseaux sociaux mais aussi la démocratisation des smartphones avec toutes les perspectives de création et de partage de contenu qu'ils offrent ont changé la donne.

Ces données, de plus en plus d'entreprises cherchent à les exploiter. Car "elles représentent une mine d'informations potentiellement incroyable, à condition de pouvoir les analyser", souligne Gilbert Grenié, analyste au cabinet de conseil PricewaterhouseCoopers (PWC). Or "les outils classiques ne suffisent plus", explique son collègue de PWC, Zouheir Guedri. Habités aux données dites "structurées" - textes, chiffres -, nos ordinateurs n'étaient, jusqu'à il y a peu de temps, pas aptes à analyser les vidéos, enregistrements de voix, sons, bruits et autres images que nous nous sommes mis à stocker de façon effrénée.

Du coup, des éditeurs de logiciels informatiques ont mis au point des techniques empruntées au calcul scientifique pour "mouliner toutes ces données" de manière exhaustive. C'est notamment le cas en marketing : "On exploite désormais toutes les données dont on dispose sur un client, en les mettant toutes au même niveau, sans les hiérarchiser : leur âge, leur profession, les boutiques autour de chez eux, cela nous permet d'établir des profils de clients totalement nouveaux", explique Marc Atallah, directeur chez Deloitte France. Le but : "Savoir quel client viser, quel client garder, quel client est le plus rentable."

## RESPECT DE LA VIE PRIVÉE

Yann Chevalier, PDG de la société Intersec, travaille pour les opérateurs de téléphonie mobile, dont SFR. "Il dispose d'une base de données clients de 20 millions d'abonnés. Toutes les données de localisation sont collectées et stockées. Elles ne sont exploitées que si le client le souhaite", explique-t-il. On peut, grâce au big data, lui envoyer des offres de "couponing" sur son smartphone quand il passe à côté d'une galerie commerciale. Le fait que le prix de la capacité de calcul ait beaucoup baissé compte énormément : "Certaines de ces techniques d'exploitation des données, on les possède depuis un certain temps.

95 Mais jusqu'à présent, les opérateurs télécom ne voyaient pas l'intérêt de dépenser des millions d'euros pour envoyer des publicités ciblées", précise Yann Chevalier.

De fait, le "cloud computing", l'informatique dans les nuages, a fait beaucoup pour le big data. Aujourd'hui, il est possible de louer d'énormes capacités de calcul, parfois pendant quelques heures seulement. A moindre coût : ces serveurs en batterie sont fournis par des prestataires comme IBM ou Amazon. Le big data "permet presque de prédire l'avenir", affirme Gilbert Grenié. En moulinant les milliards de données disponibles sur le climat, la géologie, la géothermie d'un lieu, on peut, par exemple, savoir dans quelle direction ira le vent et orienter son éolienne en fonction. D'autres exemples existent dans l'urbanisation : grâce au croisement des données de cartographie et de géolocalisation des individus, on peut prévoir où seront les embouteillages.

115 Mais la manipulation à grande échelle de ces données pose aussi le problème de la vie privée. Trouver l'équilibre entre le respect de son intimité et les bénéfices tirés du big data n'est pas simple. "Le problème s'est présenté aux Etats-Unis, raconte Roxane Edjlali du cabinet Gartner, grâce aux sites Internet visités et aux achats effectués par des jeunes femmes, une entreprise américaine a su qu'elles attendaient un bébé. Elle a donc envoyé un coupon aux domiciles des concernées et a appris à un père que sa fille adolescente était enceinte..."

Sarah Belouezzane et Cécile Ducourtieux