

Big Data : une affaire de traitement et non de volume (tribune)

Le concept de Big Data renvoie à l'origine à la masse de données complexes et polystructurées, des informations en constante inflation et de toutes natures, auxquelles les entreprises sont confrontées depuis plusieurs années. Leur impressionnant volume a conduit fort justement les puristes de la langue française à créer le néologisme de « mégadonnées », équivalent exact de l'anglais Big Data (1). Toutefois, par glissement métonymique, le Big Data en est venu peu à peu à désigner non plus les données elles-mêmes, mais un mode d'exploitation, autrement dit l'ensemble des outils décisionnels et des techniques analytiques conçus pour tirer au mieux parti de ces data.

La polysémie du terme ne cessant de s'enrichir, on voit fleurir de toutes parts aujourd'hui de nouvelles déclinaisons du Big Data. Nous avons ainsi identifié plusieurs termes évoquant moins une nouvelle génération de données que de nouvelles façons de les traiter dans des contextes d'usage très divers :

- **Smart Data** ou « *données intelligentes* » : pour les entreprises du BtoC (grande distribution par exemple), les Smart Data sont les données résultant d'un processus transformant les données brutes récoltées (émanant de diverses sources, sites web, mails, réseaux sociaux...) en données ultraqualifiées sur chaque consommateur. Là encore, le terme a tendance à s'appliquer moins aux données qu'au processus de leur monétisation, très prisé des services marketing et des gestionnaires de l'expérience client.

- **Open Data** : le traitement de ces données « *ouvertes* », à forte volumétrie (données statistiques, géographiques...) intéresse aujourd'hui en particulier les administrations et collectivités territoriales souhaitant donner aux citoyens un libre accès aux données de l'Etat pour améliorer la relation entre les services publics et leurs usagers.

- **Small Data** : par opposition aux volumineuses mégadonnées, il s'agit ici de données à faible volumétrie, mais de structure complexe et fortement évolutive. De source locale (capteurs, bases limitées), il s'agit de données contenant des caractères spécifiques, permettant de déterminer un état ou une condition à un moment précis, comme une température, un rythme cardiaque ou encore une localisation. Les Small Data sont appelées à jouer un grand rôle dans le développement de l'Internet des objets.

- **Long Data** : les « *Long Data* » (intraduisible en français) font référence à une approche intégrant la durée. Il s'agit de données courant sur une longue période, le temps constituant un critère clé de leur analyse. Les utilisateurs en sont des historiens, des scientifiques, des chercheurs en géologie, astronomie, biologie, etc.

- **Slow Data** : par opposition aux « *Fast Data* » (à consommation immédiate, comme le *fast food*), les Slow Data désigneraient des informations relativement statiques, qui évolueraient peu mais de façon imprévisible. Utilisées pour faire, par exemple, de l'analyse comportementale.

Clean, dark ou obsolètes ?

Quelles que soient les évolutions de ces nouvelles Big Data, leur exploitation fructueuse reste soumise à une condition: leur pertinence. Un récent rapport signé Veritas et intitulé « **Databerg 2015** » (2) classe les données stockées par les entreprises en trois catégories :

- les **Clean Data**, ou données « *propres* » : il s'agit des données stratégiques et reconnues comme telles, et qui donc doivent être sécurisées de manière proactive et gérées en temps réel ;
- les « **ROT data** » (ROT pour « *Redundant, Obsolete, Trivial* ») : données identifiées comme inutiles, périmées ou redondantes, qu'il est nécessaire de limiter et de supprimer régulièrement ;
- les « **Dark Data** », ou données « obscures » dont la pertinence n'a pu être identifiée.

Selon ce rapport, l'entreprise française typique posséderait 57% de données obscures, 21% de données inutiles et seulement 22% de données « *propres* ». Faute d'une meilleure gestion de leurs mégadonnées, l'absence de pertinence des informations traitées pourrait coûter aux entreprises la bagatelle de 784 milliards d'euros d'ici 2020. D'où encore la nécessité de mettre l'accent sur la façon de les traiter. On n'a décidément pas fini de parler des (Big) Data et des outils nécessaires à leur valorisation.



Par Claire Leroy, Chief Editor au CXP Group, premier cabinet européen indépendant d'analyse et de conseil dans le domaine des logiciels, des services informatiques et de la transformation numérique.

(1) comme l'indique le [grand dictionnaire terminologique](#).

(2) « Databerg 2015 », enquête commandité par Veritas Technologies et réalisée par Vanson Bourne de juillet à septembre 2015 auprès de 1475 personnes dans 14 pays de la zone EMEA (dont 200 en France).

A lire aussi : les précédentes analyses du CXP

[La cybersécurité, fil d'Ariane de la transformation numérique](#)

[Big Data : ne sous-estimez pas le potentiel économique de vos données](#)

Crédit photo : agsandrew / Shutterstock