

Le Big data, ça n'existe pas !

JEAN-PAUL CRENN / FONDATEUR DU CABINET CONSEIL EN E-COMMERCE ET TRANSFORMATION DIGITALE WEBCOLIBRI | LE 29/10 À 08:31



Le Big data, ça n'existe pas !

L'appellation "Big data" est à la mode, mais elle sert souvent à camoufler des réalités et des technologies très différentes.

Comment faire le tri ?

"Les Big data, c'est comme le bug de l'an 2000, ça n'existe pas ! Ou plutôt, c'est une appellation marketing plaquée sur un phénomène dont l'existence remonte aux premiers pas de l'informatique et pour lequel un certain nombre d'acteurs, SS2I, ou boîtes de Business Intelligence en tête, ont inventé un sobriquet, à des fins commerciales". En tout cas, c'est ce qu'affirme Julien Zakoian, ancien Directeur Marketing & Communication du site vente-privee.com (source : livre blanc de l'EBG consacré aux Big data, 2014).

Directions générales et marketing, informaticiens et tous ceux qui travaillent autour des technologies ont l'habitude des modes et des mythes qui s'évanouissent après quelques années de promesses démesurées. Rappelons-nous par exemple les promesses de Second Life il y a 10 ans. Le Big data est-il juste un habillage marketing, comme nous l'affirme Julien Zakoian ?

Tout d'abord, ce Big data mériterait bien une définition... "*Collecte, exploration et analyse de grandes masses de données*" semble être celle qui est la plus généralement admise. Et l'on voit bien que cette description est loin d'être suffisamment précise.

Rêves et fantasmes

Prenons tout d'abord le terme "*grande masse de données*". Pour les géants du Web tels que Google ou Facebook, nous imaginons bien que cela représente des millions de téraoctets (To), mais beaucoup de petits malins maquillent en "big" ce qui est en réalité du "small". Du coup, ils vendent doublement du rêve : d'une part celui du fantasme lié à l'espionnage du type NSA et d'autre part celui d'une exploitation enfin aisée, complète et peu onéreuse de leurs données.

Le "Big data" est donc bien "Big". Mais ce "big" débute à partir de quel seuil ? La "norme" la plus commune est celle de 100 To de données. En deçà, vous restez dans le "small". Quoi qu'en dise votre SSII ou votre consultant favori – votre ego dût-il en souffrir.

Mais le Big data n'est pas qu'une question de taille. Car quand on creuse la première définition, il s'avère que le Big data consiste en fait à traiter un ensemble de données plus ou moins structurées qui deviennent tellement volumineuses qu'elles sont difficiles à travailler avec des outils classiques de gestion de base de données. "Big data" est alors une appellation attrape-tout pour désigner les données qui ne vont pas dans les cases habituelles des systèmes d'information.

Il se réfère aux données qui sont trop volumineuses pour tenir sur un seul serveur, trop peu structurées pour tenir dans une base de données traditionnelle, générées en un flot continu tellement massif qu'elles ne peuvent tenir dans un entrepôt de données classique. Typiquement des données provenant de réseaux sociaux, de capteurs liés à des objets connectés, de site Web...

Du coup, cela se complexifie. D'autant plus que d'avoir beaucoup de données c'est bien, mais en faire quelque chose c'est mieux.

La promesse de l'abduction

Et c'est là que nous arrivons à la promesse la plus révolutionnaire du Big data : l'abduction. Les données vont "parler" d'elles-mêmes, sans qu'il y ait besoin d'avoir de modèles préconçus pour les analyser. Adieu les notions d'induction et de déduction ! Le Big data promet donc un nouveau paradigme : celui de la fin du cartésianisme et de la pensée scientifique – et donc de la pensée marketing – telle que nous la connaissons. Rien que cela ! Mais, bien sûr, rien n'est moins certain et les débats font rage. Notamment parce que les outils utilisés intègrent, de par leur construction, leurs propres biais.

Alors, "Big data" est-elle une simple appellation marketing, conçue pour optimiser la valorisation financière des GAFA sur laquelle se sont précipités SSII, éditeurs de Business Intelligence et autres auteurs d'ouvrages en manque de droits d'auteur ?

Rendez-vous en 2025

Pour la plus grande partie du discours sur les "Big data", cela est clairement le cas. Il est en effet tellement aisé pour l'offreur de service de maquiller en "big" ce qui est "small", tellement valorisant pour le client de se sentir faisant partie des "grands" de ce monde. Il ne faudrait cependant pas s'arrêter là car nous risquerions de jeter le bébé avec l'eau du bain. La masse des informations liées au numérique, leur flot continu, leur hétérogénéité, nécessitent clairement de nouvelles approches pour pouvoir être traitées et surtout analysées afin d'aider à l'action et à la prise de décision – ces dernières étant tellement nombreuses qu'elles doivent être en bonne partie automatisées, grâce à des agents intelligents.

Selon le "Hype Cycle" de Gartner, nous pouvons prendre rendez-vous en 2020 ou en 2025 pour voir si "Big data" aura tenu toutes ses promesses. À la lecture d'un article de Joël Rosney de 1995 consacré à l'imminence de l'arrivée d'agents intelligents dans les cyberespaces, nous pouvons penser que 2035 sera sans doute la bonne date. ●