



Accueil » Terre » Tribunes

Le Big Data est-il polluant ?

LAURENT LEFÈVRE LABORATOIRE D'INFORMATIQUE DU PARALLÉLISME ET JEAN-MARC PIERSON INSTITUT DE RECHERCHE EN INFORMATIQUE DE TOULOUSE 2 AVRIL 2015 À 18:36



Serveurs de Google, Etats-Unis, 2012. (Photo : Google/ZUMA/REA)

TRIBUNE Du commerce à l'énergie, en passant par la finance, la santé, le transport, la culture... et même la science, les données numériques représentent pour beaucoup le pétrole du XXI^e siècle. Chaque jour, l'industrie du Big Data se fait plus grosse, représentant déjà plus de 4 millions d'emplois directs dans le monde ! Mais il y a une donnée que nous avons peut-être oubliée en route : tout a un coût. Et celui du Big Data est écologique. Derrière cette informatique virtualisée, distribuée et distante, se trouvent, en effet, des infrastructures bien réelles qui ont une consommation énergétique et un impact carbone forts. N'ayons pas peur des mots : oui, le Big Data pollue.



Tous les premiers vendredis du mois, Libération publie en partenariat avec le magazine en ligne de l'organisme (<https://lejournal.cnrs.fr>), une analyse scientifique originale.

Mais reprenons du début. Le Big Data (ou « mégadonnées » en français) désigne tout à la fois la faculté de produire ou de collecter des données numériques, de les stocker, de les analyser et de les visualiser. Il est très souvent défini par ses caractéristiques liées aux « 3V » (volume, variabilité, vitesse) : les données arrivent en masse - notamment avec l'arrivée conjointe de l'Internet des objets - sont de nature plus variée que par le passé, et à une vitesse sans précédent. En 2015, le patrimoine mondial de données atteindra 8 zettaoctets (10²¹ octets). Des infrastructures pharaoniques sont donc déjà nécessaires pour stocker cette avalanche de données, mais il faut y ajouter celles pour les traiter. L'analyse de données, provenant d'observations de l'environnement, d'expériences scientifiques ou encore de données marketing, nécessite des moyens de calcul très importants, concentrés dans des grands centres et des supercalculateurs.

Prenons l'exemple de Google, qui indexe des millions de documents pour faciliter et accélérer la recherche. Ces données sont aussi analysées pour pouvoir proposer aux utilisateurs des contenus publicitaires : c'est sur cette approche que repose son modèle économique. Or, pour réaliser ces traitements, on estime que Google possède plus d'un million de serveurs, comme les trois autres géants du Web que sont Amazon, Microsoft et Facebook. Qu'elles possèdent leurs propres infrastructures ou qu'elles les louent, les entreprises et les institutions plus modestes font souvent appel à des milliers d'équipements interconnectés, depuis la capture de données jusqu'à leur analyse.

Des travaux, comme ceux du groupement de service Ecoinfo, créé par le CNRS, sont venus rappeler à quel point les technologies de l'information sont consommatrices d'énergie, et produisent des gaz à effet de serre dans toutes les étapes de leur cycle de vie : conception et transport des équipements, usage et fin de vie. En phase d'utilisation, les éléments essentiels du Big Data peuvent se répartir en trois catégories : les équipements terminaux, les réseaux et les centres de données qui consomment chacune une puissance électrique comparable, de l'ordre de 40 gigawatts en 2013 soit une quarantaine de tranches de centrales nucléaires (1). Un chiffre qui a forcément des répercussions sur le climat, même si cet impact carbone dépend du mix énergétique du pays d'usage (34 grammes de CO₂ par kWh en France en février). Alors, en arriverons-nous un jour au scénario (provocateur) de la société Cisco, dans lequel seules certaines machines auront le droit de communiquer (celles qui ont une adresse IP paire par exemple), à l'instar de la circulation alternée utilisant les plaques minéralogiques des voitures ? Fort heureusement, une multitude de réponses et d'alternatives se mettent en place. Des codes de conduites sont proposés aux industriels et aux hébergeurs pour améliorer leurs infrastructures de calcul, de communication et de stockage à grande échelle. Des innovations sont appliquées sur le design des équipements et leur potentialité de recyclage. Des recherches sur l'efficacité énergétique proposent de nouvelles voies pour limiter ces impacts. Autant d'initiatives indispensables si l'on veut pouvoir contenir la demande énergétique croissante de ces infrastructures tout en garantissant une bonne qualité de service aux utilisateurs.

Si le Big Data, en lui-même, est polluant, permet-il, par ailleurs, de diminuer d'autres sources de pollution ? Plusieurs exemples semblent en témoigner. Ainsi, l'analyse de données massives peut permettre d'optimiser des processus industriels, et donc de réduire les émissions polluantes associées. De même, les fermiers peuvent recevoir en temps réel des informations sur leurs plantations, issues de capteurs et d'images satellites, afin de ne diffuser que la quantité d'eau nécessaire, et le juste dosage de pesticides.

La simple collecte des données et sa mise à disposition peut également amener des changements de comportements. A Portland, une communauté de citoyens déploie des capteurs mesurant la qualité de l'air dans leur voisinage. Les données, de meilleure précision et qualité que celles mesurées par l'Agence de protection de l'environnement, sont analysées et visualisées sur un site web. La présence de ces données ouvertes amènera de nouveaux usages et comportements, moins centrés sur les voitures en cas de pic de pollution. Enfin, un dernier exemple vient de Chine, où la ville de Pékin s'est lancée, avec IBM, sur un programme de diminution de la pollution en ville (dont on connaît l'ampleur). Pour cela, des données sont récoltées par une multitude de capteurs, couplés avec des données satellites de météorologie, et analysées par des systèmes d'intelligence artificielle à grande échelle. Le but : produire une carte de prédiction de la pollution soixante-douze heures en avance.

Le Big Data est la ruée vers l'or des temps modernes. Comme sa glorieuse aînée, elle draine beaucoup d'espoirs - fondés ou non - et pose d'importants problèmes tout en permettant le développement de nouveaux territoires. Il nous appartient de veiller à ce que le coût environnemental de ces technologies soit contrebalancé - au moins en partie - par des progrès dans la lutte contre le réchauffement climatique et la pollution.

(1) Une tranche est une unité de production électrique au sein d'une centrale.

Laurent LEFÈVRE **Laboratoire d'informatique du parallélisme** et

Jean-Marc PIERSON **Institut de recherche en informatique de Toulouse**

0 COMMENTAIRES

Identifiez-vous pour commenter

45 suivent la conversation 

[Plus récents](#) | [Plus anciens](#) | [Top commentaires](#)