

Big Data : il est temps de grandir... ou de mourir

Une nouvelle étape. Après des éditions précédentes marquées par le déploiement du socle technique du Big Data – en particulier les datalake Hadoop -, le cru 2017 du salon Big Data Paris, qui se tient en ce moment au Palais des congrès de la porte Maillot, semble tout entier voué à la question de la valeur de la donnée. Et aux prérequis indispensables pour exploiter les applications Big Data au quotidien. Car, trop concentrées à justifier la construction d'une plate-forme technique, certaines entreprises ont développé des prototypes incapables de passer en production. Soit parce qu'ils demandent trop d'interventions manuelles. Soit parce que les algorithmes qu'ils emploient ne passent pas à l'échelle quand il s'agit d'injecter la totalité des données de production. Résultat : fréquemment six mois à un an consacrés à industrialiser les plates-formes, autrement dit à reconstruire ce qui était, aux yeux des métiers, censé être terminé.

Autre question souvent laissée dans l'ombre dans ces phases de prototype : celle de la sécurité. Une question largement abordée lors de la première journée du salon Big Data, notamment par Guillaume Poupard, le directeur général de l'Anssi, en vedette lors de la plénière. *« Il est essentiel que ces questions de cybersécurité sortent du champ trop étroit où elles ont été maintenues trop longtemps »,* a expliqué Guillaume Poupard, appelant les praticiens du Big Data à intégrer la sécurité dès le design de l'application. *« Car miser sur des patch qui vont sécuriser les systèmes à posteriori coûte cher et a toutes les chances de s'avérer inefficace, reprend le directeur général de l'Anssi. Investir dans la sécurité dès la conception s'avère rentable à moyen et long terme. Ceux qui vont survivre sont ceux qui seront capables d'assurer la sécurité de leurs systèmes ».*

Reconstruire les processus analytiques

Et d'encaisser la GDPR, le règlement européen sur la protection des données applicable en 2018, serait-on tenté d'ajouter. L'arrivée de ce texte soulève, lui aussi, pas mal d'incertitudes pour certaines initiatives Big Data existantes, où de grandes masses de données ont été amassées en vue d'exploitations futures hypothétiques. *« La GDPR va tout changer car aujourd'hui, les obligations des entreprises restent ambiguës, explique Florian Douetteau, le Pdg de l'éditeur français Dataiku. Avec la GDPR, la notion de finalité sera étendue et il faudra préciser l'impact de tout profilage aux personnes concernées. Ce qui passe par une reconstruction des processus analytiques de bout en bout. »* Selon lui, ce travail est déjà en cours aux Etats-Unis, sous la pression d'autres réglementations poussant elles aussi à l'auditabilité des processus analytiques.

Pour Laurent Guiraud, responsable du développement de CSC sur le Big Data et le Cloud, l'arrivée de cette législation peut être l'occasion d'un *« retour aux fondamentaux »* sur les usages de la donnée dans l'entreprise, après les premières années un peu débridées du Big Data où des datalakes ont été mis sur pied sans forcément maîtriser les données qui y étaient injectées, leur origine, leur validité... Mais cette transition pourrait bien s'avérer douloureuse dans certaines organisations où des dirigeants ont la tentation de tirer un trait sur les initiatives du passé pour repartir d'une feuille blanche.

Sans Big Data, pas d'IA

La question est d'autant plus brûlante que la maîtrise de la donnée sert de socle à deux autres révolutions technologiques en cours. L'IoT et la robotique ainsi que l'informatique cognitive (ou IA), des segments que le cabinet d'études IDC voit progresser de près de 20 % par an d'ici à 2020. Là où la génération de technologies actuelles (Big Data, Cloud, Mobile et social) devra se contenter d'une croissance de 6,6 % à l'année en moyenne. « *Le Big Data n'est pas grand chose sans intelligence artificielle. Inversement, l'IA n'est rien sans Big Data* », résume Emmanuel Mogenet, directeur de Google Research Europe. Une affirmation qu'appuie Jean-Michel Cambot, le développeur à l'origine de Business Objects aujourd'hui directeur de la stratégie et fondateur de l'éditeur TellMePlus : « *tout le monde a compris que c'est l'IA qui permettra d'extraire la valeur des données* ». Même si, aujourd'hui, les projets en production sont encore rares et que les entreprises lancent avant tout des prototypes en la matière. En essayant de ne pas reproduire les erreurs d'hier.

A lire aussi :

[Projets Hadoop : un échec dans 70 % des cas](#)

[Un cours sur le « bullshit » à l'ère du Big Data passionne](#)

[Transavia : la donnée client pour nouveau carburant](#)