

Focus sur 11 technologies Big Data en forte croissance

Après notre article sur les technologies [Big Data prometteuses](#), voici celles qui ont fait leurs preuves et contribuent chaque jour l'essor du Big data. 10 technologies sont en "Phase de Croissance" dans le rapport "TechRadar : Big Data, Q1 2016" de Forrester : bon niveau d'adoption, diversité et résistance. Il demeure cependant une petite incertitude sur leur devenir.

Enfin, une seule d'entre elles vient clore l'étude avec le qualificatif de "technologie en phase d'équilibre".

Naviguer en toute confiance en comprenant la situation

Les informations favorisant la prise de décision sont forcément (au moins en partie) stratégiques. D'où la nécessité de protéger leur accès, et plus encore sur le cloud ou en milieu hybride. C'est justement ce que se proposent de réaliser **le chiffrement et le masquage de données Big Data**. Reposant généralement sur les politiques de sécurité et de gouvernance de l'entreprise, ces outils automatisent et rendent transparentes les fonctions sécurisation, anonymisation, respect des règles de conformité, etc. Forrester prévoit encore une forte croissance de ces solutions.

Associant monitoring de l'infrastructure sous-jacente et optimisation des tâches, les fonctions de **supervision et gestion Big Data** ont sans aucun doute favorisé l'émergence de ces projets. En dépit de ses propres prédictions en 2014, le cabinet d'études reconnaît une évolution et une adoption très rapide de ces solutions dont l'essor va se poursuivre.

L'intégration Big Data agit sur les données via des traitements et une infrastructure massivement parallèles et évolutifs. Un processus qui permet d'ingérer des informations et d'envoyer des résultats dans de multiples cas d'usage comme la détection de fraude, la recherche en temps réel, la personnalisation, etc. le succès des infrastructures distribuées pour les données appelle d'autant plus cette intégration Big Data. Une forte valeur ajoutée qui assure la croissance de ces outils pendant des années.

Simplifier l'infrastructure à tout prix

Puiser des informations depuis diverses sources (SGBD, datawarehouse, Hadoop, GED, progiciels...) permet de restituer une information consolidée et très pertinente aux métiers. Ces opérations sont amplement simplifiées (ou simplement rendues possibles) grâce à **la virtualisation des données**. Et les sources variées (l'un des trois V du Big Data) peuvent résider sur site ou dans le cloud. Le succès de cette virtualisation devrait s'amplifier avec l'émergence des plateformes en self-service à destination des utilisateurs métier.

Les systèmes de stockage de fichiers distribués (Distributed file stores) via des grappes de

serveurs (clusters) peuvent opérer et ouvrir l'accès rapide à de très gros volumes de données, via des infrastructures matérielles évolutives. Qualifié de *"technologie la plus mature de la jeune histoire de Hadoop"* par les auteurs du rapport, ce stockage HDFS combiné au traitement distribué MapReduce est à l'origine du Big Data. Le stockage de fichiers distribué est appelé à une forte croissance pendant encore plusieurs années, selon Forrester.

Parce que l'infrastructure distribuée ne suffit pas toujours au Big Data, **la fabrique de données in-Memory** joue le rôle d'accélérateur en proposant le stockage, le traitement et l'accès à d'énormes volumes de données en mémoire, sur SSD ou Flash. Analyse en temps réel, algorithmes prédictifs, streaming... autant de technologies qui bénéficient des performances et de la faible latence du In-Memory pour répondre rapidement à l'utilisateur. Comment un tel succès pourrait-il s'arrêter?

Éprouvées ou renouvelées, presque installées.

Totalement innovantes ou revisitées, plusieurs technologies apportent un nouveau souffle au Big Data et à l'informatique au sens large. Quatre d'entre elles sont également considérées en "phase de croissance".

Profitant pleinement des infrastructures en cluster et des traitements de données distribués, **les bases de données NOSQL** sont fortement évolutives, stockent tout type de format, et proposent un schéma de gestion flexible, ainsi que des processus de requêtes variés. Idéales pour les moteurs de recommandation, les analyses de logs, l'analyse de comportements de clients, l'analyse de modèles, les applications Internet à grande échelle... mais aussi pour les applications prédictives. Bref, la solution pour les applications de troisième génération.

L'analyse prédictive est déjà une discipline installée, comme le prouvent les rachats [de SPSS par IBM](#) ou [de KXen par SAP](#). Sans oublier des acteurs majeurs comme SAS, entre autres. Tous les acteurs de cette discipline affichent aujourd'hui le support du Big data. Enrichies de machine learning, d'outils de data science et d'interfaces graphiques simplifiant l'accès à plus d'utilisateurs, ces solutions à très forte valeur ajoutée devraient continuer à croître fortement selon Forrester, mais sans atteindre la phase d'équilibre avant 10 ans. Malgré ces simplifications, la discipline reste complexe.

Les outils de recherche et de découverte de l'information ont fortement évolué avec le Big Data afin de permettre à tous d'accéder à ces grands volumes de données très variées. Souvent conçus autour de projets open source, ils se sont adaptés à ces infrastructures et gros volumes pour apporter des réponses très rapides et les plus pertinentes. Des outils qui devraient encore progresser et gagner des parts de marché pendant 3 à 5 ans.

Les solutions proposant de **l'analyse de flux** (stream analytics) enregistrent une demande croissante dans un écosystème robuste, essentiellement constitué de très grandes entreprises. Cependant, elles se démocratisent un peu plus. Selon Forrester, les éditeurs devront se positionner clairement face à l'arrivée de projets open source d'ingestion de flux (sous Hadoop notamment) en trouvant le meilleur équilibre pour combiner leur code propriétaire avec ces projets.

La seule techno Big Data à l'équilibre ?

Le cabinet d'études estime une technologie en "phase d'équilibre" lorsqu'elle est éprouvée via plusieurs cas d'usages et bénéficie d'une large base installée de clients.

Le datawarehouse massivement parallèle (ou MPP datawarehouse) stocke et traite de très grands volumes de données structurées à très grande vitesse. Orienté lignes et/ou colonnes, il est conçu pour les hautes performances, l'ingestion parallélisée, et organisé pour supporter les applications de BI. Processus optimisés, données compressées, et intégrant des traitements analytiques évolués... ces solutions peuvent répondre à une grande partie des besoins, y compris en analyse prédictive, par exemple. Bien installée, cette technologie croîtra donc modestement, mais sûrement dans les années à venir.

Et après la phase d'équilibre? Les auteurs estiment que cette phase *«va perdurer encore des années voire des décades et que ces technologies Big Data ne passeront pas en phase de déclin de sitôt, essentiellement parce qu'elles pourront s'intégrer avec de nouvelles technologies qui enrichiront la valeur ajoutée pour l'entreprise.»*

A lire aussi :

[Tour d'horizon des 11 technologies prometteuses du Big Data](#)

[BluData : comment Auchan bâtit son bras armé Big Data](#)

Crédit Photo: Solarseven-Shutterstock