

# Les principales tendances en 2016 du Big data



**La nouvelle année verra le Big Data prendre de l'ampleur et de la vitesse. L'évolution des usages des données va progresser. Les entreprises ne pourront pas ignorer la généralisation des analyses en libre-service et l'adoption à grand échelle du cloud et de Hadoop, ainsi que les nouvelles technologies venant compléter ce framework, entraînant de nombreux changements.**

## **1. Montée en puissance du NoSQL**

Dans nos prévisions des tendances du Big Data de l'année dernière, nous avons noté une accentuation de l'adoption des technologies NoSQL, qui sont généralement associées aux données non structurées. Dorénavant, les bases de données NoSQL commenceront à occuper une place centrale dans le paysage IT des entreprises, tandis que les avantages des bases de données sans schéma seront de plus en plus notables. Pour s'en convaincre, il suffit de consulter le Magic Quadrant de Gartner consacré aux systèmes de gestion de bases de données opérationnelles. Par le passé, Oracle, IBM, Microsoft et SAP dominaient le classement. Aujourd'hui, ce Magic Quadrant fait la part belle aux prestataires de solutions NoSQL, comme MongoDB, DataStax, Redis Labs, MarkLogic et Amazon Web Services (avec DynamoDB), qui viennent supplanter les fournisseurs traditionnels dans la catégorie des leaders.

## **2. Apache Spark révolutionne le Big Data**

De simple composant de l'écosystème Hadoop, Apache Spark est devenu une référence en matière de plate-forme Big Data pour de nombreuses entreprises. Spark offre une rapidité de traitement bien supérieure à Hadoop et constitue désormais le plus important projet Big Data open source, selon son créateur Matei Zaharia, également cofondateur de Databricks. Les exemples convaincants de mises en œuvre de Spark en milieu professionnel sont de plus en plus nombreux, à l'instar de Goldman Sachs qui en a fait sa solution de choix pour ses analyses du Big Data.

## **3. Les projets Hadoop arrivent à maturité et les entreprises passent à l'environnement de production**

Selon une enquête récente menée auprès de 2 200 clients Hadoop, seuls 3 % d'entre eux envisagent de moins s'appuyer sur ce framework au cours des 12 prochains mois. 76 % des sondés prévoient de recourir davantage à Hadoop au cours des 3 prochains mois, et enfin près de 50 % des entreprises qui n'ont pas déployé Hadoop déclarent avoir l'intention de le faire dans les 12 prochains mois. Cette enquête révèle également que Tableau est le principal outil d'aide à la décision pour les entreprises utilisant Hadoop ou prévoyant de l'utiliser, ainsi que pour celles qui ont déjà parfaitement intégré ce framework à leurs projets.

## **4. Le Big Data prend de l'ampleur : Hadoop s'ajoute aux normes de l'entreprise**

Hadoop occupe une place de plus en plus importante dans le paysage de l'IT. L'augmentation des investissements dans les éléments gravitant autour des systèmes professionnels, comme la sécurité, viendra corroborer cette tendance. Apache Sentry procure un système granulaire et basé sur des rôles pour gérer les autorisations d'accès aux données et métadonnées stockées sur un cluster Hadoop. Il s'agit là d'exemples de ce que les clients attendent de leur plate-forme de gestion de bases de données relationnelles. Ces fonctionnalités se retrouvent désormais à l'avant-garde des nouvelles technologies en matière de Big Data, ce qui simplifie d'autant plus l'adoption de telles solutions en entreprise.

## **5. Le Big Data prend de la vitesse : Hadoop aussi**

À mesure que Hadoop gagne en importance dans les entreprises, les utilisateurs attendent de plus en plus des fonctions d'exploration de données aussi rapides que celles proposées par les entrepôts de données traditionnels. En réponse à cette demande grandissante, nous assistons à une adoption croissante de technologies telles que Cloudera Impala, AtScale, Actian Vector et Jethro Data, qui favorisent la compatibilité de Hadoop avec les cubes OLAP et simplifient d'autant plus le rapprochement entre solutions traditionnelles d'aide à la décision et Big Data.

## **6. Le nombre croissant d'outils de préparation des données favorise la découverte d'informations**

Les outils de préparation de données en libre-service gagnent en popularité. Cette explosion est due en partie à l'adoption d'outils de découverte de données générées par les utilisateurs métiers, tels que Tableau, qui permettent d'accélérer les analyses. Désormais, ces utilisateurs souhaitent également réduire le temps nécessaire à la préparation des données et la complexité d'une telle opération, ce qui revêt une importance toute particulière pour le Big Data qui implique une multiplicité de types et de formats de données. Nous avons pu assister à de nombreuses innovations en matière de préparation de données pour le Big Data de la part de prestataires comme Alteryx, Trifacta, Paxata ou Lavastorm. Les leaders traditionnels en matière de solutions ETL, comme Informatica, qui développe Rev, ne sont pas en reste et réalisent d'importants investissements en la matière.

## **7. Les entrepôts de données à traitement MPP se tournent vers le cloud**

Même si la croissance du segment des entrepôts de données ralentit, ces solutions ne sont pas pour autant en passe de disparaître. Cette technologie opère actuellement sa transition vers le cloud, avec en tête Amazon Redshift, un entrepôt de données à la demande dans le cloud. Redshift est le service AWS qui a connu la croissance la plus rapide, mais il doit désormais faire face à la concurrence de Google et sa solution BigQuery, d'autres acteurs bien établis sur le marché comme Microsoft (avec Azure SQL Data Warehouse) ou Teradata, ou encore de nouveaux prestataires tels que Snowflake, lauréat du Strata + Hadoop World 2015 Startup Showcase, qui connaissent un succès grandissant dans le secteur. Les analystes estiment que 90 % des entreprises qui ont adopté Hadoop conserveront également leurs entrepôts de données. De plus, grâce aux nouvelles offres cloud, ces entreprises peuvent augmenter ou diminuer la capacité de stockage et la puissance de calcul de leurs entrepôts en fonction du volume d'informations stockées dans leur lac de données Hadoop.

## **8. Convergence de l'IoT, du cloud et du Big Data**

Même si l'IoT n'en est encore qu'à ses balbutiements, les pétaoctets de données générées par les différents objets connectés vont favoriser l'explosion des solutions cloud. Dans cette optique, les leaders en la matière, comme Google, Amazon Web Services et Microsoft proposent désormais des services d'IoT pour transférer en toute transparence ces données vers leurs moteurs analytiques dans le cloud.

Si ces tendances et ces évolutions peuvent sembler disparates, elles sont toutes liées par une même nécessité : exploiter les données rapidement et confortablement. Alors que le Big Data continue d'évoluer et que de nouvelles manières d'exploiter ces données voient le jour, une seule chose ne change pas : l'analyse de données est désormais à la portée de tous, et nous avons hâte de nous y mettre.



Source : *Big data : les principales tendances de 2016*