

LES DIX GRANDES TENDANCES DU
Big Data
POUR 2017

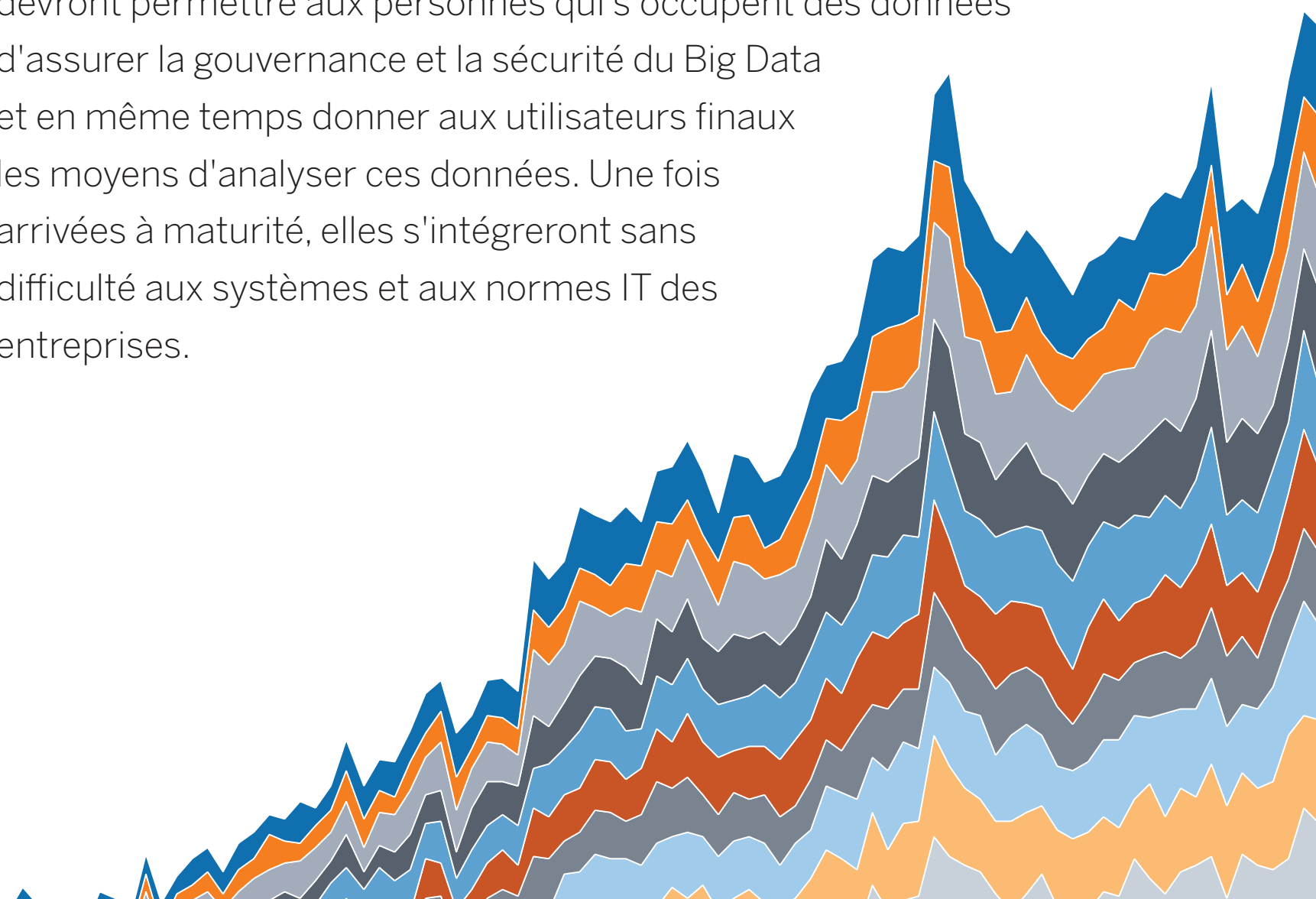




Chaque année, chez Tableau, nous lançons une discussion sur les avancées du secteur pour ensuite dresser la liste des grandes tendances du Big Data pour l'année suivante. Découvrez nos prévisions pour 2017.

Les 10 grandes tendances du Big Data pour 2017

L'année 2016 a été une année charnière pour le Big Data. En effet, les entreprises sont plus nombreuses à stocker, traiter et exploiter leurs données, quelles qu'en soient la forme et la taille. En 2017, les systèmes qui prennent en charge de grandes quantités de données, structurées et non structurées, continueront à se développer. Les plates-formes devront permettre aux personnes qui s'occupent des données d'assurer la gouvernance et la sécurité du Big Data et en même temps donner aux utilisateurs finaux les moyens d'analyser ces données. Une fois arrivées à maturité, elles s'intégreront sans difficulté aux systèmes et aux normes IT des entreprises.



BIG DATA

1

Le Big Data devient rapide et accessible : les options se multiplient pour accélérer Hadoop

Vous pouvez bien sûr pratiquer le machine learning et réaliser une analyse des sentiments sur Hadoop. Mais en général, les utilisateurs cherchent avant tout à savoir si le SQL interactif est rapide. Après tout, c'est le SQL que les utilisateurs métier utilisent lorsqu'ils veulent exploiter leurs données Hadoop afin d'obtenir des tableaux de bord plus rapides et reproductibles, et effectuer des analyses exploratoires.

Ce besoin de rapidité a favorisé l'adoption de bases de données plus rapides, comme [Exasol](#) et [MemSQL](#), de systèmes de stockage reposant sur Hadoop, comme [Kudu](#), et de technologies capables d'accélérer les requêtes. Grâce aux moteurs SQL-on-Hadoop, comme [Apache Impala](#), [Hive LLAP](#), [Presto](#), [Phoenix](#) et [Drill](#), et aux technologies OLAP-on-Hadoop comme [AtScale](#), [Jethro Data](#) et [Kyvos Insights](#), ces accélérateurs de requêtes gommant encore plus les différences entre les entrepôts classiques et l'univers du Big Data.

LISEZ AUSSI : [Le benchmark d'AtScale pour la BI sur Hadoop \(4e trimestre 2016, en anglais\)](#)

Le Big data ne se limite plus à Hadoop : les outils conçus sur mesure pour Hadoop deviennent obsolètes

Ces dernières années, avec l'essor du Big Data, plusieurs technologies se sont développées dans le but de répondre à la demande en matière d'analyse sur Hadoop. Toutefois, les entreprises avec des environnements complexes et hétérogènes ne souhaitent plus mettre en place un point d'accès BI cloisonné seulement pour une source de données (Hadoop). Les réponses à leurs questions sont enfouies dans une multitude de sources, qu'il s'agisse de systèmes SOR (System of Record), d'entrepôts dans le cloud, ou encore de données structurées et non structurées provenant aussi bien de sources Hadoop que de sources d'une autre nature. Notez que même les bases de données relationnelles commencent à s'adapter au Big Data. Par exemple, SQL Server 2016 prend désormais en charge les fichiers JSON.

En 2017, les clients voudront analyser toutes leurs données. Les plateformes capables de traiter tous les types de données et de sources prospéreront, tandis que celles qui ont été **conçues spécifiquement pour Hadoop** et qui ne pourront pas être déployées pour tous les cas d'utilisation tomberont dans l'oubli. La **sortie du marché de Platfora** constitue un premier signe de cette tendance.

LISEZ AUSSI : « **Uncommon sense: The big data warehouse** » (L'entrepôt Big Data hors du commun)



Les entreprises exploitent les lacs de données dès le départ afin d'en tirer le meilleur profit

Un lac de données est similaire à un réservoir artificiel. Vous devez d'abord construire un barrage à l'extrémité du lac (créer un cluster), puis le laisser se remplir d'eau (de données). Ensuite, vous commencez à utiliser l'eau (les données) à diverses fins, comme la production d'électricité ou l'alimentation en eau potable, mais aussi pour les activités plus ludiques, comme l'analyse prédictive, le machine learning, la cybersécurité, etc.

Jusqu'à présent, le remplissage du lac constituait une fin en soi. Ce ne sera plus le cas en 2017, car les entreprises choisiront de moins en moins Hadoop. Pour obtenir plus rapidement des réponses, elles voudront utiliser leur lac d'une manière plus agile et plus reproductible. Elles examineront attentivement leurs résultats avant d'investir dans le personnel, les données et l'infrastructure. Cela renforcera la collaboration entre **l'entreprise et l'IT**. Les plates-formes en libre-service s'imposeront davantage comme les outils à utiliser pour exploiter les ressources du Big Data.

LISEZ AUSSI : « [Maximizing data value with a data lake](#) » (Optimisation de la valeur des données avec un lac de données)

4

Des architectures arrivées à maturité permettent d'abandonner les structures universelles

Hadoop n'est plus seulement une plate-forme de traitement par lots pour les cas d'utilisation dans le domaine de la science des données. C'est devenu un véritable moteur polyvalent pour l'analyse ad hoc. Hadoop peut même servir quotidiennement au reporting opérationnel, tâche traditionnellement gérée par les entrepôts de données.

En 2017, les entreprises tiendront compte de cette diversité de besoins en favorisant une architecture adaptée au cas d'utilisation. Elles étudieront une multitude de facteurs, notamment les utilisateurs, les questions posées, les volumes, la fréquence des accès, la vitesse des données et le niveau d'agrégation avant d'adopter une stratégie pour leurs données. Ces nouvelles architectures de référence seront axées sur les besoins. Elles combineront les meilleurs outils de préparation de données en libre-service, Hadoop Core et autres plates-formes d'analyse pour les utilisateurs finaux de façon à permettre de reconfigurer la solution à mesure que les besoins évolueront. Au final, la flexibilité de ces architectures permettra de faire des choix sur le plan de la technologie.

LISEZ AUSSI : « [The cold/warm/hot framework and how it applies to your Hadoop strategy](#) » (La structure des différentes priorités des données et son application à votre stratégie Hadoop)



5

C'est la variété, et non le volume ou la vélocité, qui entraîne les investissements dans le Big Data

Gartner définit le Big Data en utilisant les « 3V » : Volume important, Vélocité élevée, grande Variété. Ces « 3V » progressent tous, mais c'est la variété qui s'impose comme le principal moteur des investissements dans le Big Data, comme en témoignent les résultats d'une **récente enquête** réalisée par New Vantage Partners. Cette tendance va s'accroître, puisque les entreprises cherchent à intégrer davantage de sources et à se concentrer sur le principe de la « **longue traîne** » du **Big Data**. Des fichiers JSON sans schéma aux types imbriqués dans d'autres bases de données (relationnelles et NoSQL) en passant par les données non plates (Avro, Parquet, XML), les formats de données se multiplient et les connecteurs deviennent indispensables. En 2017, les plates-formes d'analyse seront évaluées en fonction de leur capacité à fournir une connexion directe vers ces sources disparates.

LISEZ AUSSI : « **Variety, not volume, is driving big data initiatives** » (C'est la variété, et non le volume, qui déclenche les initiatives Big Data)

Spark et le machine learning illuminent le Big Data

[Apache Spark](#), qui était un composant de l'écosystème Hadoop, est en train de devenir la plate-forme Big Data préférée des entreprises. Dans une [enquête](#) menée auprès d'architectes de données, de responsables IT et d'analystes décisionnelles, près de 70 % des personnes interrogées préfèrent Spark à MapReduce, qui repose sur un traitement par lots et ne se prête pas aux applications interactives ni au traitement en temps réel des flux.

Ces grandes capacités de calcul pour le Big Data ont contribué à l'émergence de plates-formes offrant le machine learning avec calculs intensifs, l'intelligence artificielle et des algorithmes graphiques. Microsoft Azure ML, en particulier, doit son succès à sa facilité de prise en main et à son intégration aisée avec les plates-formes Microsoft existantes. La démocratisation du machine learning permettra de créer davantage de modèles et d'applications générant plusieurs pétaoctets de données. Tandis que les machines apprendront et que les systèmes deviendront de plus en plus intelligents, tous les regards seront tournés vers les fournisseurs de logiciels en libre-service, afin de découvrir comment ces derniers mettront ces données à la disposition de l'utilisateur final.

LISEZ AUSSI : « [Why you should use Spark for machine learning](#) » (Pourquoi utiliser Spark pour le machine learning)

La convergence de l'IoT, du cloud et du Big Data fait émerger de nouvelles opportunités pour l'analyse en libre-service

En 2017, de plus en plus d'objets du quotidien seront dotés de capteurs capables de transmettre des informations. L'IoT (the Internet of Things, autrement dit l'Internet des objets) génère d'énormes quantités de données structurées et non structurées, qui sont de plus en plus **déployées sur les services cloud**. Ce sont souvent des données hétérogènes, stockées sur plusieurs systèmes relationnels et non relationnels, que ce soit des clusters Hadoop ou encore des bases de données NoSQL. Même si les innovations en matière de stockage et de services gérés ont accéléré la capture, l'accès aux données et leur compréhension constituent toujours un défi de taille à l'autre bout de la chaîne. En conséquence, les utilisateurs sont de plus en plus nombreux à vouloir des outils d'analyse leur permettant de se connecter facilement à une grande variété de sources de données hébergées dans le cloud et de les combiner. De tels outils donnent aux entreprises les moyens d'explorer et de visualiser tous les types de données, où qu'elles soient stockées, ce qui les aide à déceler les opportunités cachées et à rentabiliser leur investissement dans l'IoT.

LISEZ AUSSI : « [Faire parler les données, un des défis liés à l'Internet des objets](#) »

La préparation des données en libre-service devient monnaie courante alors que les utilisateurs se mettent à structurer le Big Data

Permettre aux utilisateurs métier d'accéder aux données Hadoop constitue l'un des plus grands défis à l'heure actuelle. L'essor des plates-formes d'analyse en libre-service a été utile. Cependant, les utilisateurs métier souhaitent réduire encore davantage la durée et la complexité de la préparation des données à analyser, surtout s'ils sont confrontés à de nombreux types et formats de données.

Il existe des outils agiles qui permettent non seulement de préparer les données Hadoop au niveau de la source, mais également de fournir ces données sous forme d'instantanés pour accélérer et faciliter l'exploration. Des entreprises spécialisées dans la préparation des données Big Data par les utilisateurs finaux, telles que [Alteryx](#), [Trifacta](#) et [Paxata](#), sont à l'origine de nombreuses innovations. Ces outils aident les [sceptiques](#) et [ceux qui ont adopté Hadoop tardivement](#) à se lancer, et continueront à gagner du terrain en 2017.

LISEZ AUSSI : « [Why self-service prep is a killer app for big data](#) » (Pourquoi la préparation en libre-service est une application phare pour le Big Data)

Le Big Data grandit : Hadoop vient s'ajouter aux normes des entreprises

Hadoop occupe une place de plus en plus importante dans le paysage des technologies de l'information. En 2017, les entreprises investiront davantage dans les composants de sécurité et de gouvernance qui complètent leurs systèmes. Apache Sentry fournit un mécanisme granulaire et basé sur des rôles pour gérer les autorisations d'accès aux données et métadonnées stockées sur un cluster Hadoop. Le projet [Apache Atlas](#), né d'une initiative pour la gouvernance des données, permet d'appliquer la même classification à l'ensemble de l'écosystème de données. [Apache Ranger](#) fournit l'administration centralisée de la sécurité pour Hadoop.

Les clients commencent à demander ce type de fonctionnalités pour leurs plates-formes professionnelles de gestion des bases de données relationnelles. Ces fonctionnalités passent au premier plan des nouvelles technologies du Big Data, ce qui élimine encore un autre frein à l'adoption en entreprise.

LISEZ AUSSI : « [The phases of Hadoop maturity: Where exactly is it going?](#) »
([Les phases de maturité d'Hadoop : où en sommes-nous ?](#))

L'essor des catalogues de métadonnées permet de trouver des données à analyser dans le Big Data

Les entreprises ont longtemps supprimé des données parce qu'elles ne pouvaient pas toutes les traiter. Avec Hadoop, la capacité de traitement est importante, mais l'organisation des données ne permet généralement pas de les trouver facilement.

Les catalogues de métadonnées facilitent la découverte et la compréhension des données qui valent la peine d'être analysées à l'aide d'outils en libre-service. Des entreprises comme [Alation](#) et [Wamterline](#), qui utilisent le machine learning pour automatiser la recherche de données dans Hadoop, contribuent à répondre aux besoins des clients.

Elles classifient les fichiers en utilisant des balises, mettent en évidence les relations entre les éléments de données et fournissent même des suggestions de requêtes par l'intermédiaire de leur interface utilisateur dotée d'une fonction de recherche. Les utilisateurs et les gestionnaires peuvent tous réduire le temps nécessaire pour trouver et interroger précisément les données, et leur faire confiance. En 2017, la demande sera plus forte pour la découverte en libre-service, qui sera par ailleurs mieux connue. Celle-ci deviendra alors une extension naturelle de l'analyse en libre-service.

LISEZ AUSSI : [Data catalogs as a strategic requirement for data lakes \(Les catalogues de données : une condition stratégique pour les lacs de données\)](#)



À propos de Tableau

Intégrer la visualisation de données dans vos programmes et processus de vente au détail est plus facile qu'il n'y paraît.

Tableau Software aide les utilisateurs à voir et comprendre leurs données, peu importe le volume ou le nombre de systèmes dans lesquels elles sont stockées. Il vous permet de vous connecter à vos données, de les fusionner, de les visualiser et de les partager rapidement et facilement sous forme de tableaux de bord, aussi bien sur un ordinateur que sur une tablette. Créez et publiez des tableaux de bord avec des mises à jour de données automatiques, et partagez-les avec vos collègues, partenaires ou clients, même si vous n'avez aucune connaissance en programmation. Testez sans attendre la version d'évaluation gratuite.

[TABLEAU.COM/TRIAL](https://tableau.com/trial)