

What Is Big Data?

Michael Chen | Content Strategist | September 23, 2024



Définition du big data

Qu'est-ce **exactement** que le Big Data ?

La définition du Big Data est la suivante : des données plus variées, arrivant dans des volumes croissants et à une vitesse plus élevée. C'est ce que l'on appelle les trois « V ».

En d'autres termes, le Big Data est composé de jeux de données complexes, provenant essentiellement de nouvelles sources. Ces ensembles de données sont si volumineux qu'un logiciel de traitement de données traditionnel ne peut tout simplement pas les gérer. Mais ces énormes volumes de données peuvent être utilisés pour résoudre des problèmes que vous n'auriez jamais pu résoudre auparavant.

Télécharger l'e-book sur l'évolution du big data et des data lakehouses (PDF)

Les trois « V » du big data

- Volume**
La quantité de données est importante. Avec le Big Data, vous devrez traiter de gros volumes de données non structurées et à faible densité. Il peut s'agir de données de valeur inconnue, comme des flux de données X (anciennement Twitter), des flux de clics sur une page Internet ou une application mobile ou d'un appareil équipé d'un capteur. Pour certaines entreprises, cela peut correspondre à des dizaines de téraoctets de données. Pour d'autres, il peut s'agir de centaines de pétaoctets.
- Vélocité**
La vitesse à laquelle les données sont reçues et, éventuellement, traitées. Normalement, les données haute vitesse sont transmises directement à la mémoire, plutôt que d'être écrites sur le disque. Certains produits intelligents accessibles via Internet opèrent en temps réel ou quasi réel et nécessitent une évaluation et une action en temps réel.
- Variété**
La variété désigne les types de données disponibles. Les types de données traditionnels ont été structurés et trouvent naturellement leur place dans une [base de données relationnelle](#). Avec l'augmentation du Big Data, les données ne sont pas nécessairement structurées. Les types de données non structurés et semi-structurés, tels que le texte, l'audio et la vidéo, nécessitent un prétraitement supplémentaire pour en déduire le sens et prendre en charge les métadonnées.

La valeur et la vérité du Big Data

Deux autres « V » ont vu le jour ces dernières années : **valeur** et **véracité**. Les données possèdent une valeur intrinsèque. Mais cela ne sert à rien tant que cette valeur n'est pas découverte. Tout aussi important : dans quelle mesure vos données sont-elles véridiques - et dans quelle mesure pouvez-vous vous y fier ?

À l'heure actuelle, le Big Data revêt une importance capitale. Pensez à certaines des plus grandes entreprises de technologie du monde. Une grande partie de la valeur qu'ils offrent provient de leurs données qu'ils analysent en permanence pour accroître leur efficacité et développer de nouveaux produits.

Les avancées technologiques récentes ont réduit de manière exponentielle le coût de stockage et de calcul des données, ce qui facilite plus que jamais leur stockage. Un plus grand volume de Big Data étant maintenant plus économique et accessible, vous êtes en mesure de prendre des décisions commerciales plus précises.

Trouver de la valeur dans le Big Data ne consiste pas seulement à l'analyser (ce qui est un avantage supplémentaire). C'est un processus de découverte complet qui nécessite des analystes perspicaces, des utilisateurs professionnels et des dirigeants qui posent les bonnes questions, reconnaissent les tendances, fassent des suppositions éclairées et prédisent les comportements.

Mais comment en sommes-nous arrivés à cet état de fait ?

Histoire du Big Data

Bien que le concept de big data soit relativement récent, l'apparition de grands jeux de données remontent aux années 60 et 70. À cette époque, les tout premiers centre de données sortaient de terre et que les bases de données relationnelle apparaissaient lentement.

En 2005, on assista à une prise de conscience de la quantité de données que les utilisateurs généraient sur Facebook, YouTube et autres services en ligne. Hadoop (une infrastructure open source créée spécifiquement pour stocker et analyser de grands jeux de données) fut développé cette même année. NoSQL commença également à être de plus en plus utilisé à cette époque.

Le développement d'infrastructures open source telles que Hadoop (et, plus récemment, Spark) a été primordial pour la croissance du big data, car celles-ci facilitent l'utilisation du big data et réduisent les coûts de stockage. Depuis, le volume du Big Data a explosé. Les utilisateurs génèrent toujours d'énormes quantités de données, mais ce ne sont pas seulement les humains qui les utilisent.

Avec l'avènement de l'Internet of Things (IoT), de plus en plus d'objets et de terminaux sont connectés à Internet, collectant des données sur les habitudes d'utilisation des clients et les performances des produits. L'émergence du [machine learning](#) a produit encore plus de données.

Alors que le Big Data a fait beaucoup de chemin, son utilité commence à peine à se faire sentir. Le Cloud computing a encore décuplé ses possibilités. Le cloud offre une évolutivité considérable, les développeurs peuvent simplement faire fonctionner rapidement des clusters dédiés pour tester un sous-ensemble de données. En outre, les [bases de données graphiques](#) deviennent de plus en plus importantes, avec leur capacité à afficher d'énormes quantités de données de manière à rendre les analyses rapides et complètes.



Découvrez les informations dans vos données

- Qui sont les criminels qui blanchissent de l'argent sale et qui commettent des fraudes au fisc ?
- Qui a été en contact avec une personne infectée et doit se mettre en quarantaine ?
- Comment simplifier et optimiser l'ingénierie des fonctionnalités dédiées à la data science ?

Cliquez ci-dessous pour consulter l'e-book « 17 cas d'utilisation pour les bases de données orientées graphe et l'analyse de graphe ».

Télécharger l'e-book

Avantages du big data

- Le Big Data permet d'obtenir des réponses plus complètes, car le volume d'informations est plus important.
- Des réponses plus complètes signifient plus de confiance dans les données, ce qui signifie une approche complètement différente de la résolution des problèmes.

Cas d'utilisation du Big Data

Le big data peut vous aider à réaliser diverses activités commerciales, de l'expérience client aux analyses. En voici quelques-unes.

1. Développement de produit

Des entreprises telles que Netflix et Procter & Gamble utilisent le big data pour anticiper la demande des clients. Elles créent des modèles prédictifs pour de nouveaux produits et services, en classant les principaux attributs de produits ou services passés et présents et en modélisant la relation entre ces attributs et le succès commercial de leurs offres. De plus, P&G utilise les données et analyses émanant de groupes cibles, réseaux sociaux, marchés test et présentations en avant-première pour prévoir, produire et lancer de nouveaux produits.

2. Maintenance prédictive

Les facteurs permettant de prédire les défaillances mécaniques peuvent se cacher profondément dans les données structurées, telles que l'année, la marque et le type d'équipement, ainsi que dans des données non structurées couvrant des millions d'entrées de journal, de données de capteur, de messages d'erreur et de température. En analysant ces indications de problèmes potentiels avant que ceux-ci surgissent, les entreprises sont à même de déployer leur maintenance de manière plus rapide et d'optimiser le temps de fonctionnement de leurs pièces et équipements.

 | 

Contacter l'équipe commerciale

3. Expérience client

La course aux clients est lancée. Il est désormais possible d'avoir une meilleure vue d'ensemble de l'expérience client qu'auparavant. Le Big Data vous permet de rassembler des données provenant de médias sociaux, de visites Web, de journaux d'appels et d'autres sources pour améliorer l'expérience d'interaction et maximiser la valeur fournie. Commencez à proposer des offres personnalisées, à réduire la perte de clients et à traiter les problèmes de manière proactive.

4. Fraude et conformité

Votre sécurité n'est pas seulement mise à mal par une poignée de pirates informatiques ; vous faites face à des équipes entières d'experts. Les paysages de la sécurité et les exigences de conformité sont en évolution constante. Le Big Data vous aide à identifier des modèles dans les données qui indiquent une fraude et à agréger de grands volumes d'informations permettant d'accélérer le reporting réglementaire.

5. Machine learning

Le machine learning est un sujet d'actualité. Et les données, en particulier le Big Data, en sont l'une des raisons. Nous sommes désormais capables d'enseigner aux machines, plutôt que de simplement les programmer. La disponibilité du Big Data pour former des modèles de machine learning rend cela possible.

6. Efficacité opérationnelle

L'efficacité opérationnelle ne se trouve pas sous le feu des projecteurs, mais c'est dans ce domaine que l'impact du big data est le plus important. Grâce au Big Data, vous pouvez analyser et évaluer la production, les commentaires et retours des clients, ainsi que d'autres facteurs, afin de réduire les pannes et d'anticiper les demandes à venir. Le Big Data permet également d'améliorer la prise de décision, en adéquation avec la demande du marché.

7. Stimuler l'innovation

Le big data peut vous aider à innover en étudiant les interdépendances entre les êtres humains, les institutions, les entités et les processus, puis en déterminant de nouvelles manières d'utiliser ces informations. Exploiter les informations pour améliorer les décisions dans les domaines financiers et de planification. Examiner les tendances et les souhaits des clients pour offrir de nouveaux produits et services. Mettre en place une tarification dynamique. Les possibilités sont infinies.

22 cas d'usage du big data à connaître

Téléchargez votre e-book gratuit pour en savoir plus sur :

- De nouvelles façons d'utiliser vos données
- Comment la concurrence pourrait innover
- Avantages et défis des différents cas d'usage

Consultez l'e-book

Les défis liés au Big Data

Si le Big Data ouvre des perspectives intéressantes, il n'en présente pas moins certains écueils.

Premièrement, le Big Data est... volumineux. Même si de nouvelles technologies ont été mises au point pour le stockage des données, les volumes de données doublent environ tous les deux ans. Les entreprises éprouvent toujours des difficultés à maîtriser leur croissance et à trouver des moyens de les stocker efficacement.

Mais il ne suffit pas de stocker les données. Pour être utiles, celles-ci doivent être exploitées et, en amont, organisées. Des données propres, ou des données pertinentes pour le client et organisées de manière à permettre une analyse significative, nécessitent beaucoup de travail. Les spécialistes des données passent 50 à 80 % de leur temps à organiser et à préparer les données avant leur utilisation.

Enfin, la technologie du Big Data évolue rapidement. Il y a quelques années, Apache Hadoop était la technologie la plus utilisée pour traiter le Big Data. Puis, Apache Spark fit son apparition en 2014. Actuellement, l'association des deux infrastructures semble constituer la meilleure approche. Maîtriser la technologie du Big Data est un enjeu continu.

Découvrez d'autres ressources concernant le Big Data :

En savoir plus sur le Big Data chez Oracle

Fonctionnement du Big Data

Le Big Data offre de nouvelles perspectives, qui ouvrent de nouvelles opportunités et favorisent de nouveaux business models. Son adoption implique trois actions principales :

1. Intégrer

Le Big Data rassemble des données provenant de sources et d'applications disparates. Les mécanismes d'intégration des données classiques, comme ETL (extraire, transformer et charger) ne sont généralement pas à la hauteur. Pour analyser des jeux de Big Data à l'échelle de téraoctets, voire de pétaoctets, il est nécessaire d'adopter de nouvelles stratégies et technologies.

Lors de l'intégration, vous devez importer les données, les traiter et vous assurer qu'elles sont formatées et disponibles sous une forme accessible à vos analystes.

2. Gérer

Le Big Data nécessite du stockage. Votre solution de stockage peut se trouver dans le cloud, sur site, ou les deux à la fois. Vous pouvez stocker vos données sous la forme de votre choix et imposer à ces jeux de données vos exigences de traitement, ainsi que les moteurs de traitement nécessaires, à la demande. Nombreux sont ceux qui choisissent leur solution de stockage en fonction de l'endroit où sont hébergées leurs données. Le cloud est de plus en plus adopté, car il prend en charge vos besoins informatiques actuels et laisse la possibilité d'augmenter les ressources en fonction des besoins.

3. Analyser

Votre investissement dans le Big Data est rentable lorsque vous analysez et agissez sur vos données. Forgez-vous un nouveau point de vue grâce à une analyse visuelle de vos divers jeux de données. Explorez davantage les données afin de faire de nouvelles découvertes. Partagez vos conclusions avec d'autres utilisateurs. Créez des modèles de données avec le machine learning et l'intelligence artificielle. Exploitez vos données.

Meilleures pratiques en matière de Big Data

Pour vous aider dans votre aventure Big Data, nous avons rassemblé quelques bonnes pratiques clés à garder à l'esprit. Voici nos conseils pour établir un socle solide pour le Big Data.

1. Alignez le big data avec des objectifs commerciaux spécifiques

Des jeux de données plus complets vous permettent de faire de nouvelles découvertes. Afin d'y parvenir, il est important d'ancrer les nouveaux investissements dans des compétences, une organisation, ou une infrastructure avec un contexte axé sur l'entreprise, afin de garantir la continuité des investissements et du financement du projet. Pour déterminer si vous êtes sur la bonne voie, interrogez-vous sur la façon dont le Big Data étaye et renforce vos priorités commerciales et informatiques. Il peut, par exemple, vous aider à appréhender les comportements en matière de commerce électronique, capter les opinions sur les réseaux sociaux et via les interactions avec le support client, ou à comprendre les méthodes de corrélation statistique et leur pertinence pour les données client, produit, fabrication et ingénierie.

2. Réduisez le manque de compétences grâce aux normes et à la gouvernance

L'un des principaux obstacles à tirer profit de votre investissement dans le big data est le manque de compétences. Vous pouvez réduire ce risque en veillant à ce que les technologies, considérations et décisions liées au Big Data soient ajoutées à votre programme de gouvernance informatique. La normalisation de votre approche vous permettra de maîtriser les coûts et de tirer parti des ressources. Les entreprises mettant en œuvre des solutions et stratégies de Big Data doivent évaluer leurs besoins en matière de compétences en guide de préambule et identifier en amont toute pénurie potentielle de compétences. Vous pouvez y remédier en formant les ressources existantes, en embauchant ou en faisant appel à un cabinet de consultants.

3. Optimiser le transfert de connaissances avec un centre d'excellence

Utilisez une approche de centre d'excellence pour partager les connaissances, contrôler la supervision et gérer les communications du projet. Que le Big Data soit un investissement nouveau ou en expansion, les coûts fixes et indirects peuvent être partagés dans l'entreprise. Cette approche permet d'augmenter les capacités en termes de Big Data et la maturité de l'architecture globale des informations de manière plus structurée et systématique.

4. Les meilleurs résultats sont l'alignement des données non structurées avec les données structurées

Il est certainement utile d'analyser le big data seul. Vous obtiendrez néanmoins des perspectives plus larges en connectant et en intégrant le Big Data de faible densité avec les données structurées que vous utilisez déjà actuellement.

Que vous recueilliez du Big Data concernant des clients, des produits, des équipements ou l'environnement, l'objectif est d'ajouter des points de données plus pertinents à vos synthèses principales et analytiques, afin d'améliorer les conclusions. Par exemple, il y a une différence entre s'intéresser aux opinions de tous les clients et seulement à celles de nos meilleurs clients. C'est la raison pour laquelle de nombreuses personnes considèrent le Big Data comme l'extension de leurs fonctions de business intelligence existantes, de leur plateforme de data warehouse et de l'architecture de leurs informations.

N'oubliez pas que les processus et les modèles d'analyse du Big Data peuvent être à la fois humains et basés sur des machines. Les fonctions d'analyse du Big Data englobent les statistiques, l'analyse spatiale, la sémantique, la découverte interactive et la visualisation. Les modèles d'analyse permet de corrélér différents types et sources de données afin de créer des associations et de faire des découvertes pertinentes.

5. Planifiez votre laboratoire de recherche pour la performance

Il n'est pas toujours simple de trouver du sens à vos données. Parfois, nous ne savons même pas ce que nous recherchons. C'est prévu. La direction et l'informatique doivent prendre en charge ce « manque de direction » ou « manque d'exigences claires ».

Dans le même temps, il est important que les analystes et les spécialistes des données travaillent étroitement avec l'entreprise pour comprendre les principales lacunes et exigences en matière de connaissances. Afin de favoriser l'exploration interactive des données et l'expérimentation avec des algorithmes statistiques, vous avez besoin d'espaces de travail ultra-performants. Assurez-vous que les environnements sandbox disposent du support dont ils ont besoin et sont correctement gérés.

6. Adoptez le modèle d'exploitation cloud

Les processus et les utilisateurs de big data ont besoin d'accéder à un large éventail de ressources pour l'expérimentation itérative et l'exécution de tâches de production. Une solution de Big Data comprend tous les aspects liés aux données, notamment les transactions, les données principales, les données de référence et les données de synthèse. Les bacs à sable analytiques doivent être créés à la demande. La gestion des ressources est cruciale pour garantir la maîtrise du flux de données dans son intégralité, notamment le traitement préalable et postérieur, l'intégration, la synthèse au sein de la base de données et la modélisation analytique. Une mise en service d'un cloud privé et public et une stratégie de sécurité bien planifiées jouent un rôle primordial dans la prise en charge de ces besoins en évolution.

En savoir plus sur le big data chez Oracle

- Essayez un atelier gratuit sur le Big Data



Contactez
l'équipe
commerciale

- Ressources pour 
- Pourquoi choisir Oracle ? 
- Formation 
- Nouveautés 
- Nous contacter 



Contacter
l'équipe
commerciale