

B1.3 Big Data

Le terme "données" est un sujet fréquent aujourd'hui, nous le rencontrons ainsi que le terme "big data" tous les jours. Les deux concepts se retrouvent également dans ce texte d'étude, nous allons donc les connaître de plus près.

Qu'est-ce que les données?

Les données peuvent être décrites comme les quantités, caractères ou symboles sur lesquels des opérations sont effectuées par un homme ou un ordinateur, et qui sont stockées et/ou transmises sous forme écrite ou sous forme de signaux électriques et enregistrées sur des supports magnétiques, optiques ou mécaniques. support d'enregistrement.

Exemples : chiffres 2, -4, 0,257, 3,14, $\sqrt{274}$, ... , lettres b, xyz, a, A+B, ..., symbols ☉, ♀, ➔, mots "Être ou ne pas être?", "Je n'aime plus ça", ... , .

Qu'est-ce que le Big Data?

La collecte de données qui est énorme en volume et en taille et qui croît de façon exponentielle avec le temps. Les données sont si volumineuses et complexes qu'aucun des outils de gestion de données traditionnels ne peut les stocker ou les traiter efficacement.

Exemples:

Les « Bourses » sont des exemples de Big Data qui génèrent environ un téraoctet de nouvelles données commerciales par jour.

Plus de 500 téraoctets de nouvelles données transitent chaque jour par les bases de données des sites de médias sociaux, générées en termes de téléchargements de photos et de vidéos, de messages, etc.

Les Big Data sont des ensembles de données dont la taille ou le type dépasse la capacité des bases de données relationnelles traditionnelles à capturer, gérer et traiter les données avec une faible latence. Les caractéristiques des mégadonnées comprennent un volume élevé, une vitesse élevée et une grande variété. Les sources de données deviennent plus complexes que celles des données traditionnelles car elles sont pilotées par l'intelligence artificielle (IA), les appareils mobiles, les médias sociaux et l'Internet des objets (IoT). Par exemple, les différents types de données proviennent de capteurs, d'appareils, de vidéos/audio, de réseaux, de fichiers journaux, d'applications transactionnelles, du Web et des médias sociaux, dont la plupart sont générés en temps réel et à très grande échelle.

Les Big Data sont des ensembles de données dont la taille ou le type dépasse la capacité des bases de données relationnelles traditionnelles à capturer, gérer et traiter les données avec une faible latence. Les caractéristiques des mégadonnées comprennent un volume élevé, une vitesse élevée et une grande variété.



„Freepig.com“

Les sources de données deviennent plus complexes que celles des données traditionnelles car elles sont pilotées par l'intelligence artificielle (IA), les appareils mobiles, les médias sociaux et l'Internet des objets (IoT).

Par exemple, différents types de données proviennent de capteurs, d'appareils, de vidéos/audio, de réseaux, de fichiers journaux, d'applications transactionnelles, du Web et de médias sociaux, dont la plupart sont générés en temps réel et à très grande échelle.

Grâce à l'analyse des mégadonnées, vous pouvez finalement alimenter une prise de décision, une modélisation et une prévision meilleures et plus rapides des résultats futurs et une intelligence économique améliorée. Les entreprises peuvent accéder à un grand volume de données et analyser une grande variété de données pour obtenir de nouvelles informations et prendre des mesures.

L'analyse des données provenant de capteurs, d'appareils, de vidéos, de journaux, d'applications transactionnelles, du Web et des médias sociaux permet à une organisation d'être axée sur les données.

Types de Big Data

Il existe trois types de mégadonnées : structurées, non structurées et semi-structurées.

Données structurées : toutes les données qui peuvent être stockées, consultées et traitées sous la forme d'un format fixe. De nos jours, la taille de ces données augmente considérablement, les tailles typiques étant de l'ordre de plusieurs zettaoctets (un milliard de téraoctets forme un zettaoctet).

Exemples : Les données stockées dans un système de gestion de base de données relationnelle sont un exemple de données « structurées ». Données dans des tableaux, par exemple, horaire, tableaux de rémunération, liste de mots-clés à la fin de la publication, etc.

Une table « employé » dans une base de données est un exemple de données structurées.

Données non structurées : toutes les données dont la forme ou la structure est inconnue sont classées comme données non structurées. En plus de leur taille énorme, les données non structurées posent de multiples défis en termes de traitement pour en tirer de la valeur.

Exemple : Une source de données hétérogène contenant une combinaison de fichiers texte, d'images, de vidéos, etc.

De nos jours, les organisations disposent d'une multitude de données non structurées avec des informations intéressantes disponibles, mais elles ne savent pas comment en tirer des valeurs cachées puisque ces données sont sous leur forme brute ou dans un format non structuré.

Données semi-structurées : peuvent contenir les deux formes de données.

Exemples : Pages web personnelles : contiennent des textes, des images, des vidéos, des liens vers des réseaux sociaux, des liens pour communiquer avec des amis, etc. Un autre exemple est les pages web de boutiques en ligne avec publicité, terminaux de paiement.

Les caractéristiques du Big Data sont communément appelées les quatre volumes du Big Data.

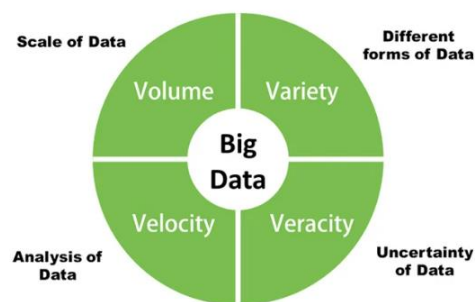
Le volume de données fait référence à la taille des ensembles de données qui doivent être analysés et traités, qui dépassent désormais souvent les téraoctets et les pétaoctets . Le volume considérable de données nécessite des technologies de traitement distinctes et différentes des capacités de stockage et de traitement traditionnelles. Cela signifie que les ensembles de données dans le Big Data sont trop volumineux pour être traités avec un processeur d'ordinateur portable ou de bureau ordinaire.

La vitesse du Big Data fait référence à la vitesse à laquelle les données sont générées. Les données à grande vitesse sont générées à un rythme tel qu'elles nécessitent des techniques de traitement distinctes (distribuées). Un exemple de données générées à grande vitesse serait les messages Twitter ou la communication Facebook.

La variété du Big Data rend le Big Data vraiment gros. Le Big Data provient d'une grande variété de sources et est généralement l'un des trois types suivants : données structurées, semi-structurées et non structurées. La variété des types de données nécessite souvent des capacités de traitement distinctes et des algorithmes spécialisés. Un exemple d'ensembles de données très variés serait les fichiers audio et vidéo générés à divers endroits dans une ville.

La vérocité des Big Data fait référence à la qualité des données qui sont analysées. Les données de haute vérocité contiennent de nombreux enregistrements qui sont précieux à analyser et qui contribuent de manière significative aux résultats globaux. Les données de faible vérocité, en revanche, contiennent un pourcentage élevé de données sans signification. Le non précieux dans ces ensembles de données est appelé bruit. Un exemple d'ensemble de données à haute vérocité serait les données d'une expérience ou d'un essai médical.

Les données à volume élevé, à grande vitesse et à grande variété doivent être traitées avec des outils avancés (analyses et algorithmes) pour révéler des informations significatives. En raison de ces caractéristiques des données, le domaine de connaissances qui traite du stockage, du traitement et de l'analyse de ces ensembles de données a été étiqueté Big Data.



Les données à volume élevé, à grande vitesse et à grande variété doivent être traitées avec des outils avancés (analyses et algorithmes) pour révéler des informations significatives. En raison de ces caractéristiques des données, le domaine de connaissances qui traite du stockage, du traitement et de l'analyse de ces ensembles de données a été étiqueté Big Data.

Avantages du traitement du Big Data

- Les entreprises peuvent utiliser des renseignements extérieurs lors de la prise de décisions.
- L'accès aux données sociales des moteurs de recherche et des sites comme Facebook ou Twitter permet aux organisations d'affiner leurs stratégies commerciales.
- Amélioration du service client.
- Identification précoce des risques pour les produits/services, le cas échéant.
- Meilleure efficacité opérationnelle.
- Les technologies Big Data peuvent être utilisées pour créer des entrepôts de données.

Résumé

Big Data est un terme utilisé pour décrire une collection de données de taille énorme et qui croît pourtant de façon exponentielle avec le temps. Le Big Data peut être 1) structuré, 2) non structuré, 3) semi-structuré. La vitesse du Big Data fait référence à la vitesse à laquelle les données sont générées. Les données à grande vitesse sont générées à un rythme tel qu'elles nécessitent des techniques de traitement distinctes. La variété de Big Data provient d'une grande variété de sources et est généralement de l'un des trois types suivants : données structurées, semi-structurées et non structurées. La variété des types de données nécessite souvent des capacités de traitement distinctes et des algorithmes spécialisés. La vérocité du Big Data fait référence à la qualité des données qui sont analysées : on distingue les données à haute vérocité et les données à faible vérocité. Volume, variété, vitesse et variabilité sont les quatre caractéristiques du Big Data. Un service client amélioré, une meilleure efficacité opérationnelle, une meilleure prise de décision sont quelques avantages du Big Data.

Liens vers des sujets pertinents

https://en.wikipedia.org/wiki/Big_data

<https://www.guru99.com/what-is-big-data.html>

<https://www.gartner.com/en/information-technology/glossary/big-data>

Mots clés

quantités

personnages

symboles

Big Data

intelligence artificielle

appareils mobiles

réseaux sociaux

Internet des objets

piloté par les données

données structurées

données non structurées

données semi-structurées

quatre volumes de Big Data

vitesse du Big Data

variété de mégadonnées

véracité du Big Data