

Plan

1. Introduction
2. Caractéristiques du Big Data
3. Chaîne de valeur du Big Data
4. Problèmes d'entreprise résolus par le Big Data

Introduction

EMPREINTES NUMÉRIQUES DEPUIS APPAREILS

- **Ordinateurs** : Toute action effectuée sur un ordinateur, de naviguer sur le Web pour utiliser des applications, des feuilles Derrière les traces numériques. Cela inclut des données telles que l'historique du navigateur , les cookies, les informations de connexion et les journaux d'activité .
- **Appareils mobiles** : Smartphones et tablettes en permanence générer des données, notamment le suivi de localisation , l'utilisation des applications, les messages et les journaux d'appels. Les appareils mobiles sont souvent synchronisé avec des services cloud, créant un enregistrement numérique complet.
- **Appareils portables (par exemple, montres)** : les montres intelligentes et les trackers d'activité physique surveillent l'activité physique activité , fréquence cardiaque , habitudes de sommeil et même niveaux de stress . appareils en continu collecter données relatives à la santé , contribuant à l'empreinte numérique d' un individu .

INTERACTIONS SUR LES RÉSEAUX SOCIAUX

- **Publications et commentaires** : Chaque publication, commentaire, mention « j'aime » et partage sur les plateformes de médias sociaux telles que Facebook, Twitter, Tik Tok et Instagram génère des données. Ces données reflètent les intérêts , les opinions et les relations sociales des utilisateurs .
- **Données comportementales** : les plateformes de médias sociaux suivent le comportement des utilisateurs , par exemple la durée pendant laquelle ils dépenser pour une publication, le contenu avec lequel ils interagissent et ce qu'ils annoncent ils cliquent dessus. Ces données sont utilisées pour personnaliser flux et publicités .

REQUÊTES

- **Recherche Google** : Chaque recherche requête entré dans Google ou autre les moteurs de recherche sont enregistrées et analysées . Ces données révèlent quelles informations les gens recherchent , leurs intérêts et leurs préoccupations .
- **Tendances de recherche** : les moteurs de recherche regroupent les données de millions d'utilisateurs pour identifier les tendances, comme les sujets populaires à un moment donné . Ces informations peuvent être utilisé pour divers fins , y compris marché recherche et prévision analytique .

INTÉGRATION DE CAPTEURS DANS LES OBJETS DU QUOTIDIEN

- **Internet des objets (IoT)** : les capteurs sont désormais intégré dans une large gamme de besoins quotidiens objets , des thermostats intelligents aux réfrigérateurs , en passant par les véhicules et même vêtements . Ceux-ci capteurs recueillir des données sur la manière dont ces les objets utilisés , leurs performances et les conditions environnementales .
- **Smart Cities** : Dans les zones urbaines , des capteurs sont intégrés dans les infrastructures pour surveiller le trafic , la qualité de l'air , la consommation d'énergie , etc. Ces données aident les urbanistes à optimiser ressources et améliorer qualité de vie des résidents .
- **Voitures intelligentes** : générer , collecter et transmettre de vastes quantités de données quantités de données en temps réel.

DONNÉES SUR LES ACTIVITÉS

- **Monde numérique** : les activités en ligne , notamment les transactions de commerce électronique, les services de streaming et les interactions virtuelles , génèrent des quantités massives de données. Ces données comprennent achat historique , habitudes de visionnage et enregistrements de communication.
- **Monde physique** : Dans le monde physique , les données sont générées à travers activités comme les achats en magasin, le suivi GPS et les images des caméras de sécurité . La frontière entre les données numériques et physiques est de plus en plus floue comme plus physique les activités sont suivies numériquement .

TOUT ÇA NOUS AMÈNE AU MONDE DU BIG DATA : DE GRANDS ENSEMBLES DE DONNÉES

- **Empreintes numériques depuis Appareils**
- **Interactions sur les réseaux sociaux**
- **Requêtes**
- **Intégration de capteurs dans les objets du quotidien**
- **Données sur les activités**
- **L' émergence du Big Data**

Qu'est-ce que le Big Data ?

- Le Big Data fait référence à des données extrêmement volumineuses et complexes. ensembles de données que ne peut pas être facilement traité, géré ou analysé avec traitement traditionnel des données outils.
- Un big data se caractérise par un volume, une vitesse et une variété énormes .
- Les défis incluent l'analyse, la capture, la conservation des données, la recherche, le partage, le stockage, le transfert, la visualisation, l'interrogation, la mise à jour et la confidentialité des informations.
- L'analyse des ensembles de données peut trouver de nouvelles corrélations pour « repérer les tendances commerciales, prévenir les maladies, lutter contre la criminalité, etc. »

Qui génère le Big Data ?

Sociale



Suivi et engagement des utilisateurs



Sécurité intérieure



Commerce électronique



Services financiers



Recherche en temps réel

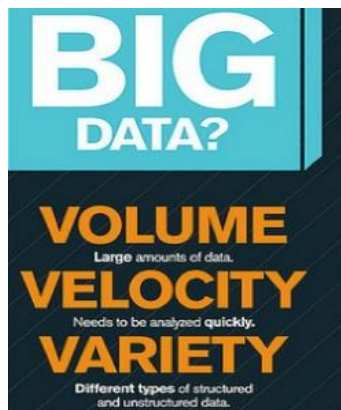
Google



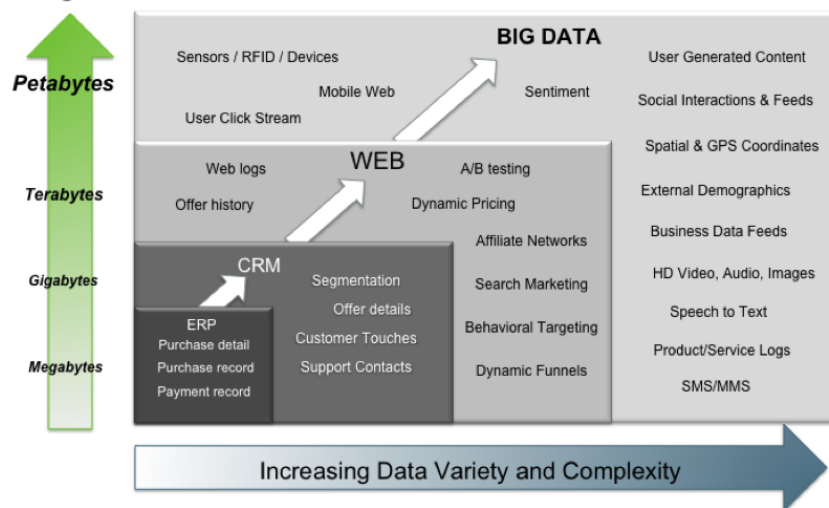
Plan

1. Introduction
2. Caractéristiques du Big Data
3. Chaîne de valeur du Big Data
4. Problèmes d'entreprise résolus par le Big Data

Caractéristiques du Big Data : 3V ...



Big Data = Transactions + Interactions + Observations



Source: Contents of above graphic created in partnership with Teradata, Inc.

Volume

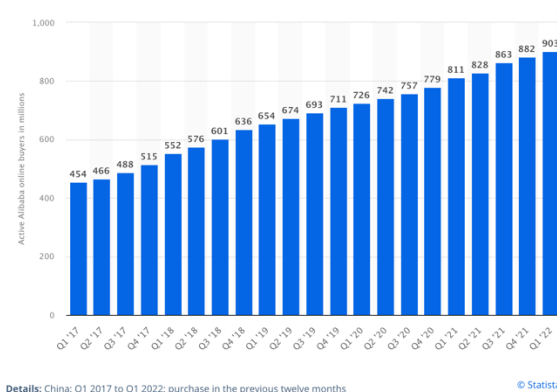
- « volume » fait référence à la pure quantité ou quantité de données qui est être généré , collecté et traité . Volume reflète l'immense échelle de données qui organisations Il est nécessaire de gérer et d'analyser les données , et cela met en évidence les défis posés par la taille même des ensembles de données .
- Le volume de données a augmenté de façon exponentielle ces derniers temps années en raison de diverses facteurs , notamment :
- **Utilisation d'Internet** : l' augmentation nombre d' utilisateurs d'Internet , croissance des services en ligne et expansion des activités numériques contribuer au volume massif de données générées tous les jours .
- **Connecté Appareils** : La prolifération des appareils de l'Internet des objets (IoT) , tels que les capteurs , les appareils intelligents et les objets portables , génère vaste quantités de données comme celles-ci appareils en continu collecter et transmettre des informations.

Exemple : le festival de shopping de la Journée des célibataires

- **Volume de données** : Lors de l' événement Singles' Day 2020 , Alibaba a traité **583 000 commandes par seconde** à son pic . Au cours de l' événement de 11 jours , Alibaba a généré et traité plus de **2,32 milliards de livraisons commandes** . Ce volume massif de transactions a entraîné la génération de vastes quantités de données liées au client comportement , produit préférences , détails de la transaction , paiement méthodes et logistique .
- **Stockage et traitement des données** : gérer ce énorme volume de données, Alibaba utilise ses plateforme de cloud computing avancée , Alibaba Cloud. Les données collectées comprend données structurées (par exemple, montants des transactions , heures et produits) SKU) et des données non structurées (par exemple, client avis , images et vidéos) . Le volume considérable de ces données nécessite solutions de stockage hautement évolutives et parallèles traitement des frameworks comme Hadoop et Spark.
- **Implications** : La capacité de traiter et d'analyser ce volume de données permet à Alibaba d' optimiser logistique , personnaliser client expériences et améliorer la prise de décision en temps réel . Par exemple, l'analyse Les habitudes d'achat pendant la Journée des célibataires peuvent aider à prédire demande de spécificité produits , gérer les stocks plus efficacement et adapter les efforts de marketing aux différents segments de clientèle .

Quantité de données générées par jour

- C'est une dynamique et rapide chiffre changeant en raison de la continuité croissance des technologies numériques, d'Internet et des réseaux connectés appareils .
- taux de création de données mondial était de 10^{18} octets.
 - il est essentiel de noter que ce l'estimation peut varier selon la source et la réalité la quantité de données générées continue d' augmenter au fil du temps.



Nombre de consommateurs actifs annuels à travers Les propriétés d'achat en ligne d'Alibaba
du 1er trimestre 2017 au 1er trimestre 2022 en Chine

Variété (Complexité)

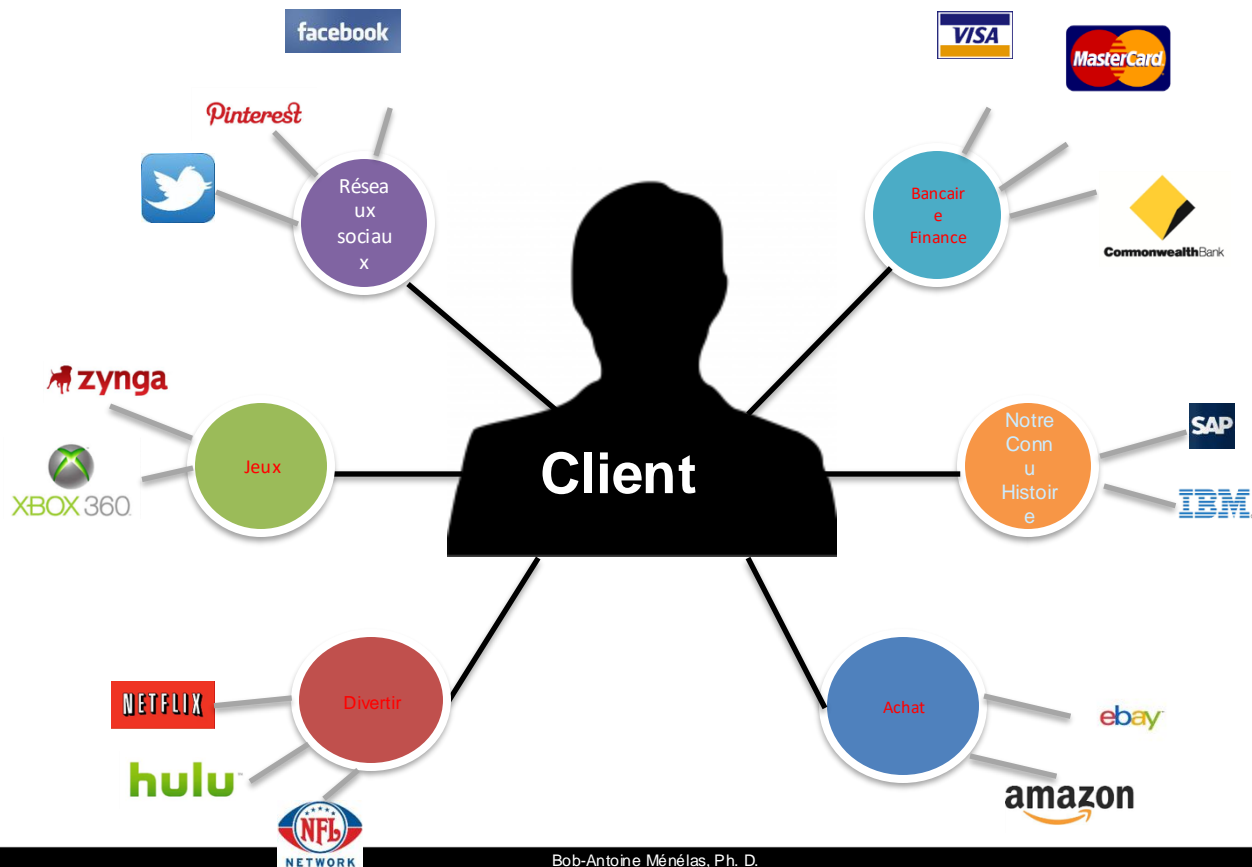
- « **variété** » fait référence aux divers types de données rencontrées . Systèmes de gestion de données traditionnels étaient principalement conçu pour gérer données structurées , qui convient soigneusement dans des tables et des lignes (par exemple, relationnelles bases de données). Dans le big data, la variété des types de données s'est élargie au-delà données structurées à inclure données non structurées et semi- structurées .

1. Données structurées

2. **Données non structurées** : Les données non structurées n'ont pas de modèle de données prédéfini ou ne s'intègrent pas dans un modèle relationnel. base de données . Exemples inclure documents texte , e-mails, vidéos , images, publications sur les réseaux sociaux et autres formes de données non tabulaires . Analyse les données non structurées peuvent fournir des informations précieuses mais présente des défis en raison de sa absence de structure fixe .

3. **semi- structurées** : les données semi- structurées tombent entre données structurées et non structurées . Tandis que il peut ne pas se conformer à une structure rigide comme la structure traditionnelle bases de données , il contient quelques niveau de structure, souvent sous forme de balises, de marqueurs ou de clés. Exemples inclure des fichiers JSON (JavaScript Object Notation) ou XML (eXtensible Markup Language).

Une vue unique sur le client



Surveiller les activités numériques d'un utilisateur

- La surveillance de plusieurs activités numériques d'un utilisateur, telles que les interactions sur les réseaux sociaux, les préférences sportives télévisées et d'autres comportements en ligne, est cruciale pour les organisations pour plusieurs raisons :
 - **Marketing personnalisé et expérience client**
 - Comprendre les préférences
 - En suivant les activités numériques des clients, les entreprises peuvent mieux comprendre leurs préférences, leurs intérêts et leurs comportements. Cela permet aux entreprises d'adapter leurs messages marketing, leurs produits et leurs services pour répondre plus efficacement aux besoins individuels des clients.
 - Campagnes ciblées
 - Connaître les sports que regarde un client ou ce qu'il partage sur les réseaux sociaux permet aux entreprises de créer des campagnes marketing très ciblées. Par exemple, un passionné de sport peut recevoir des promotions liées à des événements sportifs à venir ou à des produits associés, ce qui rend les efforts marketing plus pertinents et augmente les chances d'engagement.

Surveiller les activités numériques d'un utilisateur

— Amélioration de l'engagement et de la fidélité des clients

- Établir des relations
 - Le suivi des activités numériques aide les entreprises à interagir avec leurs clients de manière plus significative en répondant à leurs intérêts spécifiques. Cela peut renforcer les relations, favoriser la fidélité à la marque et encourager les clients à revenir.
- Interactions opportunes
 - En comprenant quand et comment les clients s'engagent dans certaines activités numériques, les organisations peuvent interagir avec eux aux moments les plus opportuns, par exemple lors d'événements en direct ou après une publication importante sur les réseaux sociaux.

Surveiller les activités numériques d'un utilisateur

- Développement de produits et de services améliorés
 - Dépenses publicitaires optimisées
 - Ciblage rentable
 - Éviter le gaspillage
 - Gestion des risques et prévention des crises
 - Suivi du sentiment de marque
 - Prise de décision éclairée
 - Analyse de la concurrence
- En résumé, le suivi des différentes activités numériques d'un client permet aux entreprises de créer une approche plus personnalisée, plus efficace et plus efficiente du marketing, du développement de produits, de l'engagement client et de la gestion des risques. Cette compréhension globale du comportement des clients est essentielle pour conserver un avantage concurrentiel sur le marché actuel axé sur les données.

Vélocité (vitesse)

- « **vélocité** » fait référence à la vitesse à laquelle les données sont générées , traitées et analysées . met en évidence la nature dynamique des données dans le monde numérique d'aujourd'hui , où l'information est généré à des vitesses sans précédent . Plusieurs facteurs contribuer à la vélocité des données :
 1. **Flux de données en temps réel** : De nombreuses applications et systèmes produire des données en temps réel. Cela comprend les publications sur les réseaux sociaux , les transactions financières , les capteurs lectures à partir d'appareils IoT et plus encore. Analyse et traitement ces données en temps réel sont essentiel pour une mise en œuvre opportune décisions fabrication
 2. **Internet des objets (IoT)** : appareils IoT générer des données en continu et, dans de nombreux cas, ces données doivent être traité et agi en temps réel . Exemples inclure des capteurs de ville intelligente , connectés véhicules et applications IoT industrielles .

Big Data en 6 V

- Volume : Dans un environnement Big Data, les quantités de données collectées et traitées sont bien plus importantes que celles stockées dans les bases de données relationnelles classiques.
- Variété : Le Big Data est constitué d'une grande variété de types de données.
- Vitesse : les Big Data parviennent à l'organisation à grande vitesse et à partir de plusieurs sources simultanément.
- **Véracité : Les problèmes de qualité des données sont particulièrement difficiles dans un contexte de big data.**
- **Visibilité/Visualisation : Une fois les big data traitées, nous avons besoin d'un moyen de présenter les données de manière lisible et accessible.**
- **Valeur : En fin de compte, les big data n'ont aucun sens s'ils n'apportent pas de valeur à un objectif significatif.**

Véracité (qualité et confiance)

- *Données = quantité + qualité*
- Lorsque nous parlons de big data, nous faisons généralement référence à sa quantité :
 - Quelle est la capacité d'un système à faire face à la taille considérable des données ?
 - Une requête sur le Big Data est-elle réalisable dans le cadre de nos ressources disponibles ?
 - Comment pouvons-nous rendre nos requêtes exploitables sur le Big Data ?
 - . . .
- **Pouvons-nous faire confiance aux réponses à nos questions ?**
 - Les données sales conduisent régulièrement à des rapports financiers trompeurs et à des décisions de planification stratégique des activités \Rightarrow **perte de revenus, de crédibilité et de clients, conséquences désastreuses**
- *L'étude de la qualité des données est aussi importante que la quantité des données*

Véracité (qualité et confiance)

- « **Véracité** » fait référence à la fiabilité des données. Véracité aborde les problèmes d'incertitude et de qualité qui peuvent survenir fonctionnement avec le big data.
- **Exactitude des données** : Assurer que les données sont précis et exempt de erreurs est une préoccupation fondamentale. Des données inexacts peuvent conduire à des analyses incorrectes donc à de mauvaises prises de décision.
- **Cohérence des données** : Les incohérences dans les données, telles que les variations dans les formats de données ou les informations contradictoires , peuvent entraver l' intégration et l'analyse des ensembles de données .
- **Pertinence des données** : Véracité aussi considère si les données sont pertinentes pour l'analyse en cours. Des données non pertinentes ou obsolètes peuvent introduire du bruit et diminuer la performance globale des analyses.
- **Exhaustivité des données** : Les données manquantes ou incomplètes peuvent affecter la fiabilité des analyses. Il est essentiel de combler les lacunes dans les données pour garantir une analyse complète.
- **Fiabilité des données** : Véracité est aussi liée à la fiabilité des sources de données. Données provenant des sources peu fiables ou non fiables peuvent introduire biais ou inexactitudes .
- Les défis de véracité impliquent des Stratégies d' assurance qualité des données , de nettoyage des données et de gouvernance des données.
 - L'établissement de pratiques de gouvernance des données garantit que il existe des normes et des procédures en place pour maintenir la qualité des données au fil du temps.

Visibilité/Visualisation

- Visibilité fait référence à la capacité d'avoir un aperçu, de surveiller et de comprendre divers aspects des données, des processus et des systèmes . Cette visibilité est crucial pour les organisations afin de prendre des décisions éclairées.
- **Visibilité des données** : les organisations doivent avoir une vision claire compréhension de leurs données - où il est stocké , comment c'est consulté , qui y a accès etc.
- Visibilité des données aide à maintenir la gouvernance des données , à assurer la sécurité des données et à gérer la conformité avec règlements .
- **Visibilité du processus métiers**
- **Visibilité opérationnelle**
- **Visibilité de la chaîne d'approvisionnement**
- **Visibilité de la conformité**

Valeur

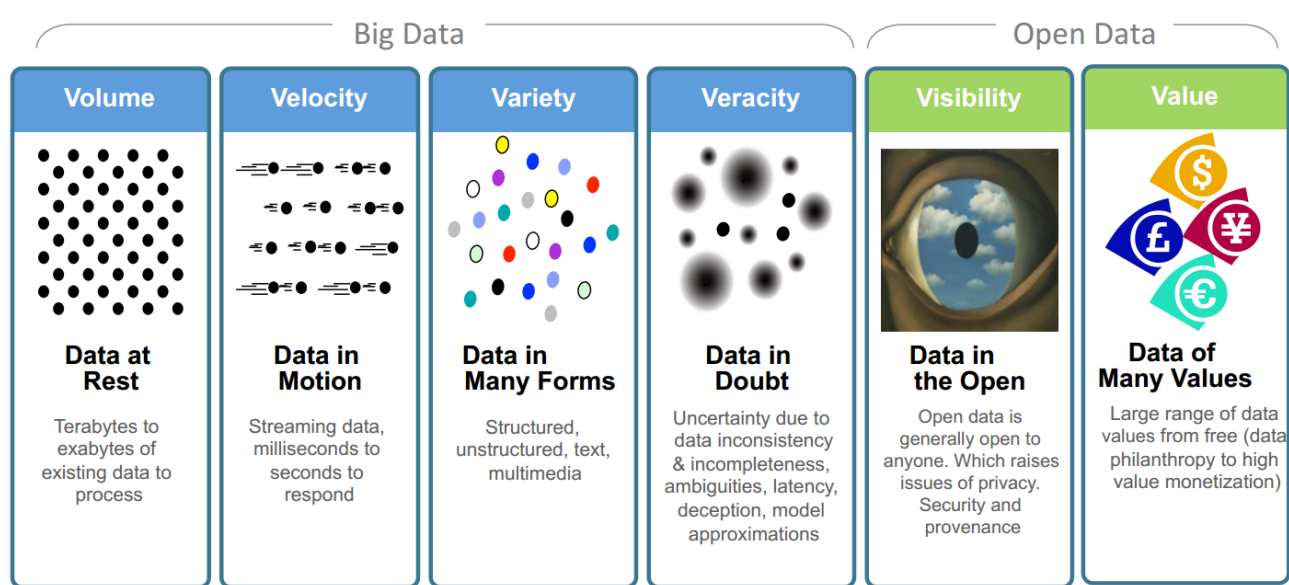
- Les Big Data n'ont aucun sens si elles n'apportent pas de valeur à un objectif significatif.



Valeur

- « valeur » fait référence au but ultime de l'extraction des idées significatives. Il souligne l'importance de transformer les données en résultats exploitables.
- **Dériver des idées** : le principal Le but de l'analyse des big data est d'obtenir des informations précieuses qui peuvent éclairer la prise de décision . En identifiant les modèles, les tendances et les corrélations au sein de grands ensembles de données , les organisations peuvent découvrir des informations qui seraient invisibles avec traditionnel méthodes .
- **Prise de décision éclairée**: Analyse des Big Data fournit une base pour prendre des décisions plus éclairées et stratégiques décisions . Que ce soit dans le domaine applicatif, les données sont la pour permettre une amélioration de l'efficacité , réduite coûts , et améliorer l'ensemble résultats .

Big Data : 6V en résumé



Transformer l'énergie et les services publics grâce au Big Data et à l'analytique. Par Anders Quitzau@IBM

Plan

1. Introduction
2. Caractéristiques du Big Data
3. Chaîne de valeur du Big Data
4. Problèmes d'entreprise résolus par le Big Data

Chaîne de valeur des données

- décrit le processus de bout en bout de la manière dont les données sont créées, collectées, traitées et finalement utilisées pour créer de la valeur au sein d'une organisation.
- **Génération de données** : le processus commence avec la création de données. Cela pourrait être généré à travers divers moyens tels que les interactions des utilisateurs, les appareils IoT, les capteurs, les transactions, etc.
- **Collecte de données** : une fois les données collectées, il doit être collectées et stockées. Cela implique l'agrégation de données provenant de diverses sources dans des référentiels ou bases de données centralisés.
- **Traitement des données** : Les données brutes doivent être transformées et traitées pour être adaptées à l'analyse. Cette étape implique le nettoyage, l'agrégation et la structuration des données.
- **Stockage des données** : Les données traitées sont alors stockées dans des bases de données ou des entrepôts de données. Les solutions de stockage doivent être évolutives et sécurisées pour s'adapter au volume et à la variété des données.
- **Analyse des données** : Des outils et des algorithmes d'analyse sont appliqués aux données pour découvrir des modèles, des tendances et des informations. Cette étape est cruciale pour l'extraction d'informations exploitables à partir des données.
- **Génération d'informations** : Les résultats de l'analyse des données conduisent à la génération d'informations. Ces informations peuvent éclairer la prise de décision, identifier des opportunités et optimiser divers aspects de l'entreprise.
- **Prise de décision** : Sur la base des connaissances acquises, des décisions sont prises pour favoriser les améliorations, les innovations et les actions stratégiques au sein de l'organisation.
- Tout au long de cela, il est important de prendre en compte la gouvernance des données, l'assurance qualité et la sécurité pour garantir que les données utilisées sont fiables, précises et alignées avec les normes et politiques organisationnelles.

Plan

1. Introduction
2. Caractéristiques du Big Data
3. Chaîne de valeur du Big Data
4. Problèmes d'entreprise résolus par le Big Data

Problèmes commerciaux résolus par l'analyse des Big Data

- **Efficacité des processus et réduction des coûts**
- **Gestion de marque**
- **Maximisation des revenus, ventes croisées/ventes incitatives**
- **Expérience client améliorée**
- **Service client amélioré**
- **Identification du taux de désabonnement, recrutement de clients**
- **Identifier de nouveaux produits et opportunités de marché**
- **Gestion des risques**
- **Conformité réglementaire ...**