

Chapitre 1 : Introduction au Big Data



- L'informatique c'est l'automatisation du traitement de l'information, création d'informations nouvelles à partir d'une saisie initiale
- Dans la version archaïque de l'informatique, il y a avait très peu de données et elles étaient structurées selon les besoins du processus
- Actuellement, à l'ère des réseaux sociaux, les données sont devenues massives et chaque jour nous générons environ **330 millions de téraoctets** de données qui sont pour la plupart non structurés

Une galaxie d'informations

- Les données sont le nerf de la guerre pour les entreprises, qui leur permettent de prendre des décisions en adéquation avec le contexte
- Données toujours plus massives et diversifiées
 - Les données couramment produites en un an représentent plus en volume que l'ensemble des données produites depuis le début de l'informatique
 - Les données prennent des formes très variées et souvent non structurées ou semi-structurées et donc difficiles à analyser de manière uniforme

- Les sources de données sont multiples :
 - Les sites internet
 - Les bases de données
 - Les téléphone
 - Les serveurs
 - Etc
- Ces sources sont alimentées par des capteurs très variés
 - Des sondes de capture de phénomène physique
 - Des scanners répertoriant divers objets (images, produits, etc)
 - Des saisies utilisateurs telles que les posts des réseaux sociaux
 - Tout autre outil permettant d'enregistrer une donnée

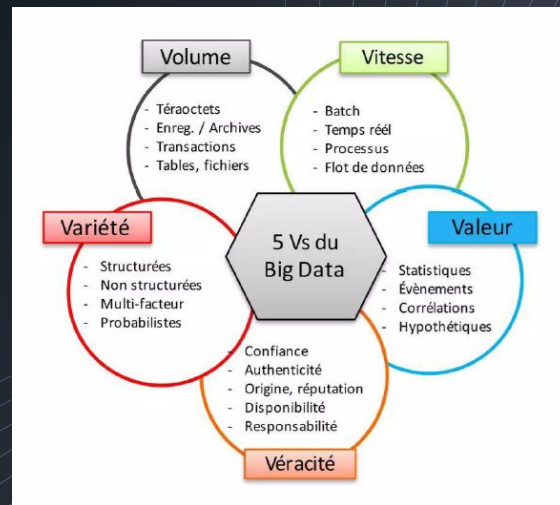
Challenges

- Réunir un grand volume de données variées pour trouver de nouvelles idées
- Capturer des données créées rapidement
- Sauvegarder toutes ces données
- Traiter ces données et les utiliser

Les 5 V du Big Data

- Extraction d'informations et décisions prises sur la base des données récupérées, caractérisées par les 5 V :

- Volume (Volume)
- Variété (Variety)
- Vitesse (Velocity)
- Véracité (Veracity)
- Valeur (Value)



- Le prix de stockage des données a beaucoup baissé ces dernières décennies :
 - ~ 100 000 \$ / Go en 1980
 - ~ 0.05 \$ / Go actuellement
- Facilités pour stocker des quantités importantes de données, pourtant, il est important de choisir de ne stocker que ce qui est utile pour pouvoir facilement parcourir les données

- Les données prennent des formes extrêmement variées
- La plupart des données existantes sont non structurées ou semi-structurées
- Une même donnée peut être stockée sous plusieurs formes
 - Un appel téléphonique peut être transcrit ou enregistré vocalement pour conserver l'intonation

- La génération de nouvelles données est très rapide
 - Nécessité de stocker des quantités massifs d'informations
 - La vitesse de traitement est primordiale pour suivre le rythme
- Risque de perte de données si les données ne sont pas stockées à leur arrivée
- Une information peut avoir une durée de vie très courte et doit être traitée en conséquence
 - Exemple : Relayer un bon plan sur un site marchand

- Fait référence à la fiabilité des données
- Avec l'augmentation de la quantité, la qualité et la précision de l'information se perdent
- Les solution Big Data doivent remédier à cela
 - Grande rigueur dans l'organisation de la collecte
 - Recoupement, croisement et enrichissement des données
 - Volonté de garantir la sécurité et l'intégrité des données

- Le V le plus important
- Il faut transformer toutes les données en valeur exploitables
- L'objectif du Big Data est d'atteindre des objectifs stratégiques de création de valeur pour les clients et pour l'entreprise dans tous les domaines d'activité

Approche traditionnelle vs Big Data

Approche traditionnelle

Les besoins métier guident la conception

- 1) Le responsable métier définit les besoins et les questions à se poser
- 2) Le service IT conçoit une solution avec un ensemble de structures et de fonctionnalités
- 3) Le responsable métier utilise les outils pour répondre au besoin
- 4) Un nouveau besoin émerge et le cycle recommence

Approche Big Data

Les sources d'info guident la découverte créative

- 1) Le responsable métier et IT identifient les sources de données disponibles
- 2) Le service IT fournit une plateforme qui permet une exploration créative des données
- 3) Le responsable métier explore les données et définit les questions à poser
- 4) De nouvelles idées conduisent à intégrer de nouvelles données / technologies

Cohabitation des approches

- La question n'est pas :
 - Que dois-je choisir entre l'approche classique et l'approche Big Data
- Mais plutôt :
 - Comment faire cohabiter les deux approches, les faire fonctionner ensemble

