

Introduction au Big Data

Pr. HIBBI Fatima-Zohra

f.hibbi@emsi.ma

Plan

- Définitions
- Exemples d'utilisations
- Données au repos – Données en mouvement.
- Caractéristiques du Big Data
- Technologie Big Data

BIG DATA

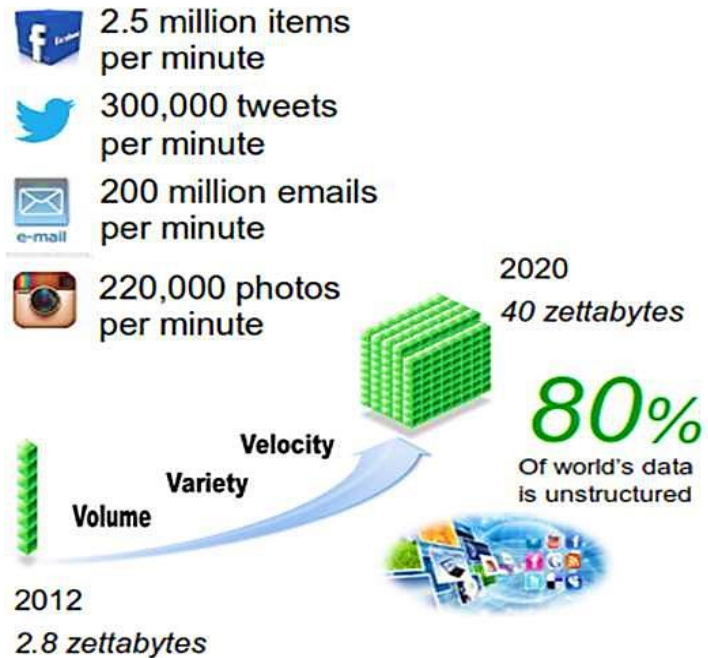


Contexte

- § L'univers digital devient de plus en plus **large** et interconnecté
- § Les limites de l'utilisation d'un SGBDR:
 - Bases de données traditionnelles SQL sont monolithiques (traitement de multiples tâches connexes).
 - Problème de nombre d'utilisateur
 - Problème du nombre élevé des relations

Contexte

§ La croissance des données:



Organizations need deeper insights

1 in 3

Business leaders frequently make decisions based on information they don't trust, or don't have

1 in 2

Business leaders say they don't have access to the information they need to do their jobs

83%

of CIOs cited "Business intelligence and analytics" as part of their visionary plans to enhance competitiveness

60%

of CEOs need to do a better job capturing and understanding information rapidly in order to make swift business decisions

Information is at the center of a new wave of opportunity

Constat

- L'utilisation des données numériques devient plus en plus compliqué avec la variété et le volume de données circulées

Objectif :

L'amélioration des processus, aider à la prise de décision, la prévision, etc.

Solution:



Définition

Le cabinet d'études Gartner définit la Big Data comme suit:

« Big Data is high-volume, high-velocity, and/or high-variety information assets that demand cost effective, innovative forms of information processing that enable enhanced insight, decision making and process automation »

Définition

- Quelques exemples d'utilisation de la Big Data

- Science
- Astronomy
- Atmospheric science
- Genomics
- Biogeochemical
- Biological
- Social
- Social networks
- Social data
 - * **Twitter**
 - * **Facebook**
 - * **LinkedIn**
- Commercial
- Web / event / database logs
- Sensor networks
- Internet text and documents
- **Medical** records
- Photographic archives
- Video / audio archives
- Government
- Military and homeland security surveillance

Exemple de volumétrie

- 13 petabytes

Quantité pouvant être téléchargée d'Internet en deux minutes, si environ 300 millions de personnes se trouvaient connectées simultanément.

- 124 exabytes

Quantité des données des data center dans le monde en 2018

- 40 zettabyte

Quantité de données en 2020

Value	<u>Decimal</u>	
		<u>Metric</u>
1000	kB	<u>kilobyte</u>
1000 ²	MB	<u>megabyte</u>
1000 ³	GB	<u>gigabyte</u>
1000 ⁴	TB	terabyte
1000 ⁵	PB	petabyte
1000 ⁶	EB	exabyte
1000 ⁷	ZB	zettabyte
1000 ⁸	YB	yottabyte
1000 ⁹	RB	ronnabyte
1000 ¹⁰	QB	quettabyte

Système d'unité

L'état des données

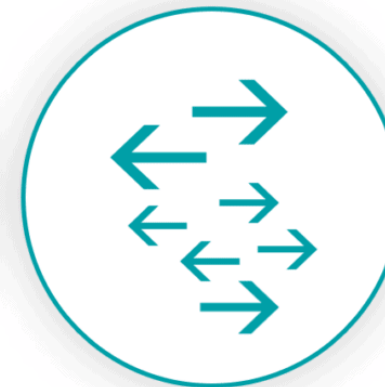
- § Donnée au repos: (Données stocké dans des supports physique). Accessible/ Non accessible.
- § Donnée en transit: Des apps de travail collaboratif(slack, MsTeams,etc)
- § Donnée en cours d'utilisation: l'utilisation des apps ouverts: données qui utilisent les mémoires (chargement des fichiers).

LES TROIS ÉTATS DES DONNÉES

AU REPOS



EN TRANSIT



EN COURS D'UTILISATION



Les caractéristiques du Big Data

Il n'y a pas de définition unique du Big Data, mais certains éléments sont communs à toutes les définitions: la vitesse, le volume, la variété.

Extraire, de manière rapide et rentable, des connaissances à partir de données volumineuses, variées et d'une vitesse élevée.

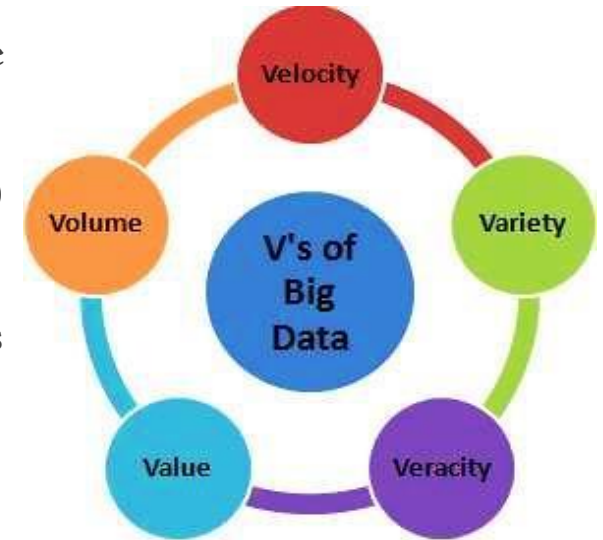
- Volume**: varie de terabytes à zettabytes. On doit traiter efficacement ce volume croissant.

- Variété**: Gérer divers types et structures de données (Base des données structurée) et non structurée (Image, vidéo, signaux)

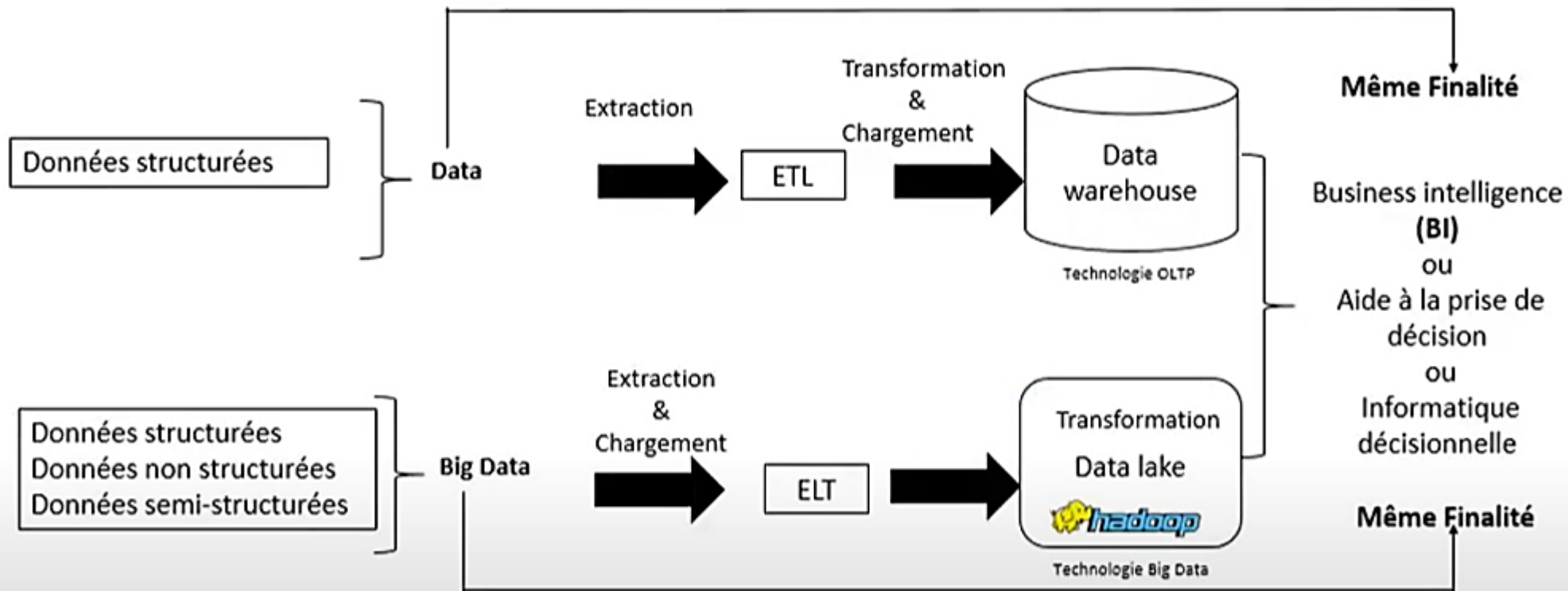
- Vitesse (vitesse)**: Analyser les flux de données et les grands volumes de données persistantes.

- Véracité**: authenticité et fiabilité des données (Sources Fiables)

- Valeur**: le point le plus important, une solution Big Data doit apporter une valeur ajoutée pour l'entreprise en répondant à des objectifs commerciaux ou Marketing qui orientent l'utilisation des Big Data



Big data Vs système décisionnel (BI)

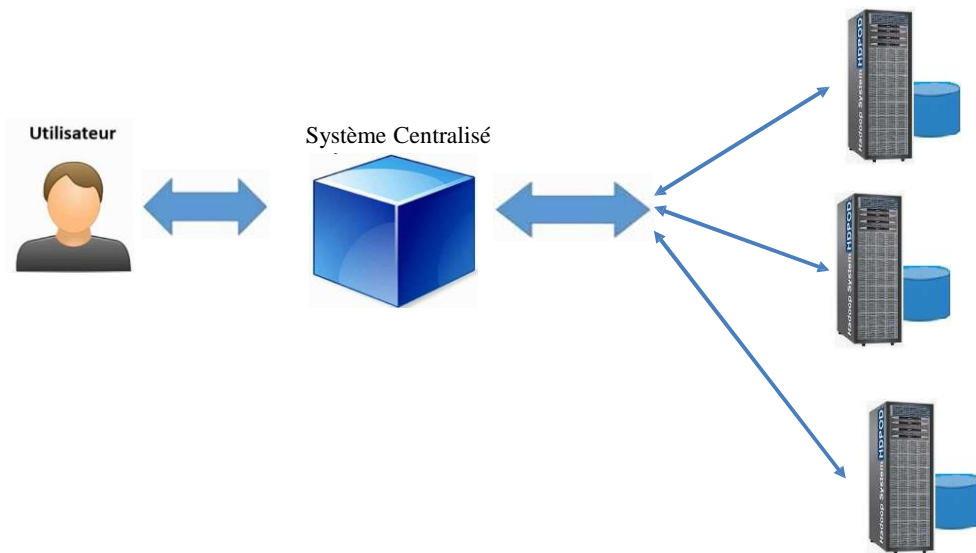


Les avantages et Projet Big data

- § Aide intelligente pour la prise de décision,
 - § Réductions des coûts
 - § Réductions de temps
 - § Optimisation des processus de l'entreprise et développement de nouveaux produits ou services
- Déforestation : Projet Planetary SKIN(Par la NASA, CISCO, ...)
 - Suivi astronomique en direct : Projet LSST (Large Synoptic Survey Telescope) (20 Tb chaque nuit)
 - Traitement du Cancer: projet ICGC (International Cancer Genome Consortium), analyse de plusieurs BD sur des tumeurs de 50 types de cancers

Approche Big Data

§ Répartition des données et des traitements



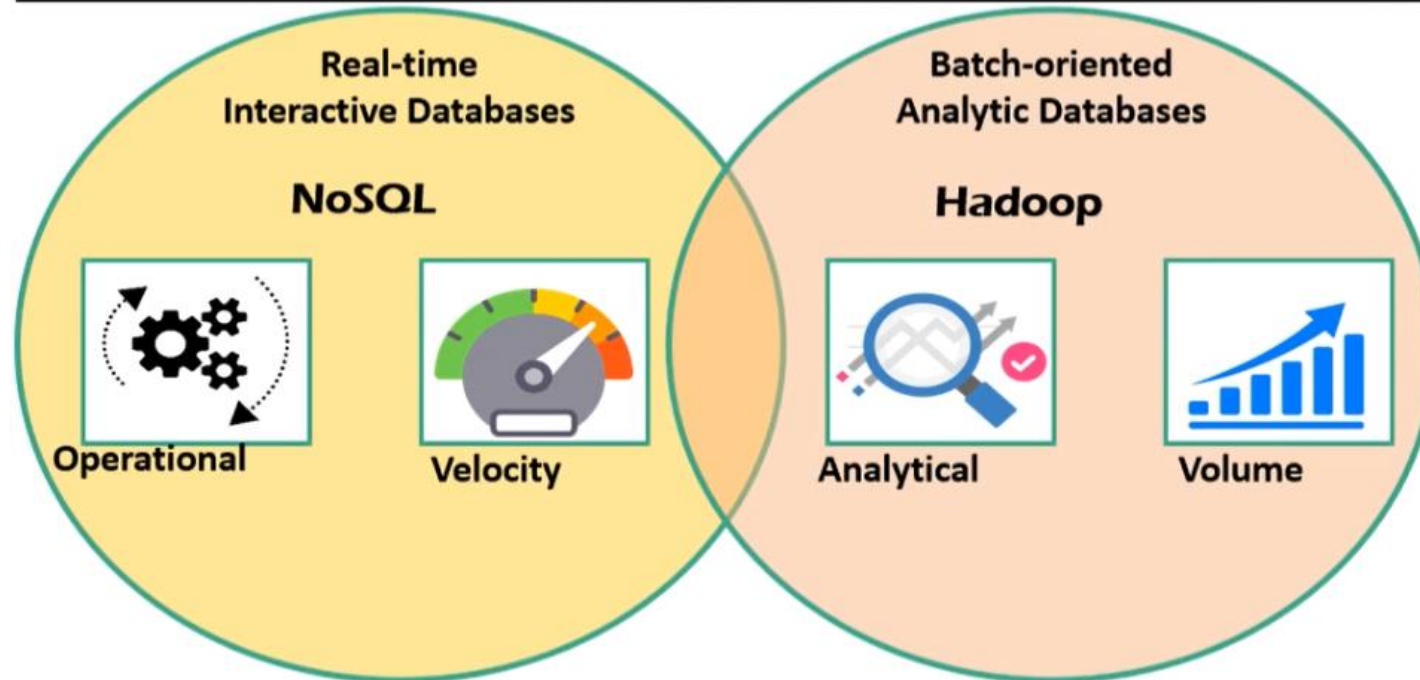
Fonctionnalité du Big Data

L'explosion de la quantité des données générées par les applications nous a permis de chercher les moyens suivants:

- § Transmettre et diffuser les données entre les applications distribuées (Data Broker: *transmission des données à partir du comportement des utilisateurs*) {outil: [Kafka Open source](#)}
- § Stocker et sécuriser les données {outil: [Hadoop hdfs](#), [NoSql](#)}
- § Traiter et analyser les données: Batch processing et stream processing {outil: [Map reduce/Spark](#), [Spark/Kafka](#)}
- § Etc.

Technologies de Big Data

Big Data = Operational + Analytic (NoSQL + Hadoop)



- Online
- Web/Mobile/IOT Apps
- Millions of Customers/Consumers

- Offline
- Analytics Apps
- Hundreds of Business Analysts