

Big Data : caractéristiques

BIG DATA



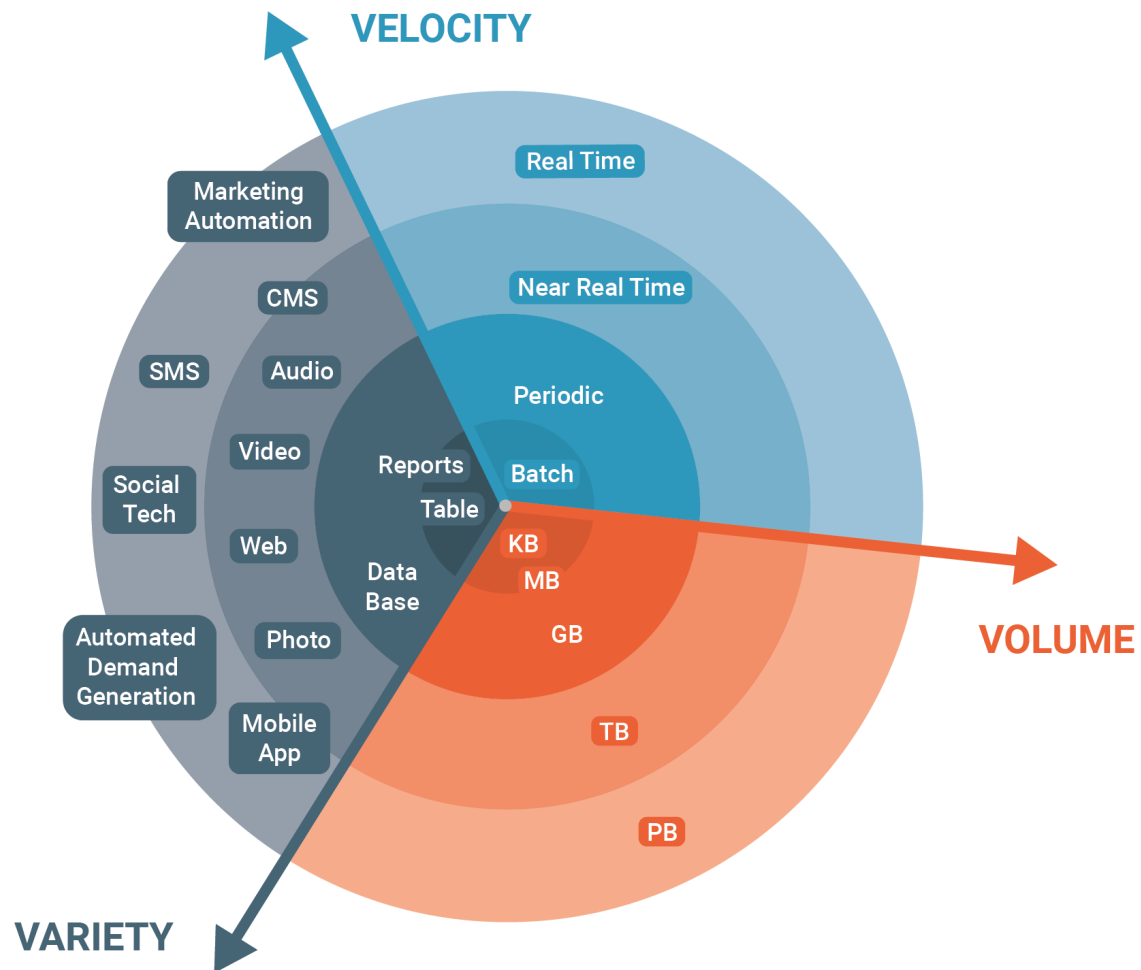


Semifir

Les 3 V

Le Big Data est repose sur le principe des 3 V :

• The 3 V's of Big Data



- **V**olume
- **V**itesse
- **V**ariété

Volume

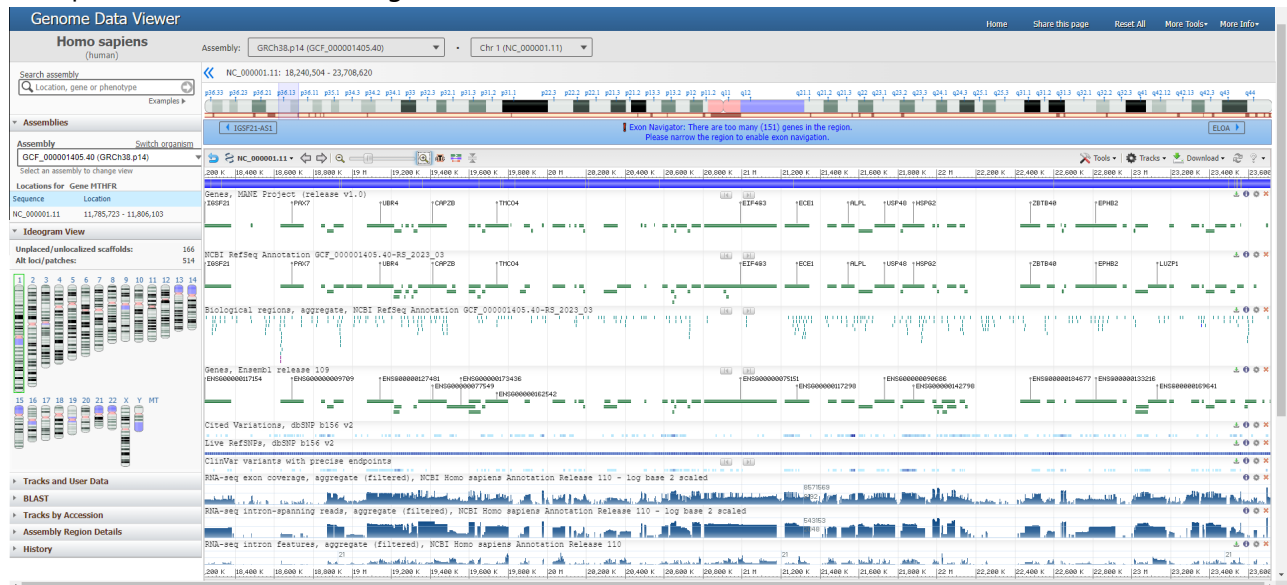
Le plus evident des 3V, le **volume** represente la masse croissante de données.

Neessite des besoins spécifiques pour :

- Stocker les données
- Transporter et echanger les données
- Traiter et analyser les données

Volume : producteurs

- Entreprises commerciales (Google, Amazon, Facebook, Twitter, etc.)



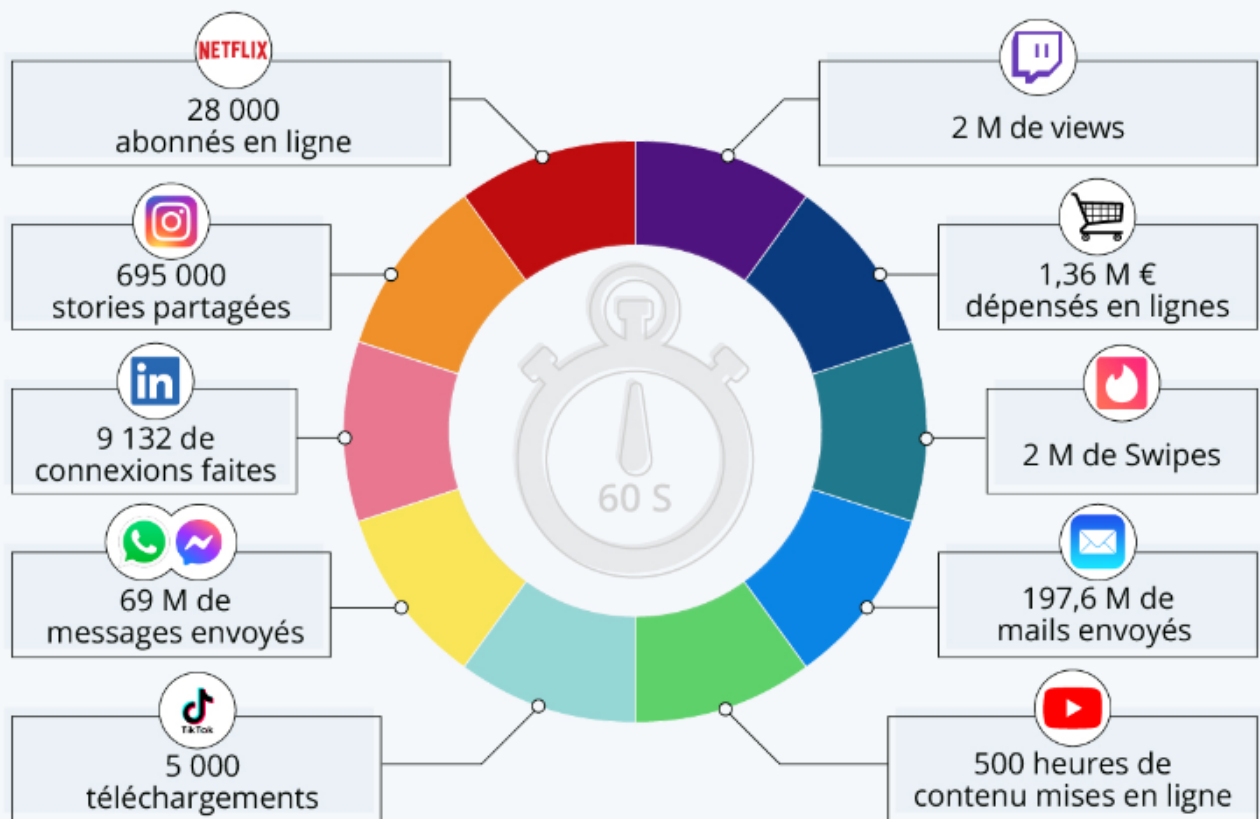
- Analyses de données de santé (les données médicales)
(<https://www.ncbi.nlm.nih.gov/genome/guide/human/>)
- Analyses meteorologiques/spatiales

Vitesse

Le **vitesse** est la rapidité avec laquelle les données sont générées, traitées ou modifiées.

Une minute sur Internet en 2021

Estimation de l'activité et des données générées sur Internet en l'espace d'une minute



Source : Lori Lewis via AllAccess



statista

- Reponds aux besoins des processus chronosensibles (bourses, stream, etc.)
- Risque pour l'Homme de perdre le contrôle sur les données (traitement à la nanoseconde).

Variété

La **variété** est la diversité des types de données.



- Données structurées (tableaux, fichiers, etc.) **20%**
- Données semi ou non-structurées (XML, JSON, textes, images, etc.) **80%**

Variété : données structurées

Données qui à une structure prédéfinie (tableau, fichier, etc.)

- Ce sont toutes les données observables avec un schéma de typage bien définis et fixe.
- Généralement, elles sont observables en tableau ou base de données relationnelles traditionnelles. (exemple : adresse, ventes d'articles, etc.)

Variété : données non-structurées

Données qui sont qualitatives et non quantitatives.

- Ce sont toutes les données qui ne sont pas observables avec un schéma de typage bien définis et fixe.
- Sont une conglomération de nombreuses données de différents types qui sont stockées dans leurs formats en mode natif (video, audio, etc.)

Variété : données semi-structurées

Données intermédiaire entre les données structurées et non-structurées.

- Ce sont des données qui ont une structure mais qui ne sont pas fixe ou qui n'ont pas de structure prédéfinie mais possèdent des métadonnées typées fixes. (exemple : XML, JSON, etc.)

Big Data : Extension des 3V

Aujourd'hui, les principes des 3V se sont largement étendus. Par extension, on y trouve d'autres grands V comme :

- **V**olatilité

- **V**aleur
- **V**ulnérabilité
- **V**éracité
- **V**alidité
- **V**isibilité

Volatilité

Notion de durées de vie d'une données.

- L'estimation de la durée de vie d'une donnée,



- de sa prise en charge et son traitement
- de son obsolescence sont a prendre en compte.

Valeur

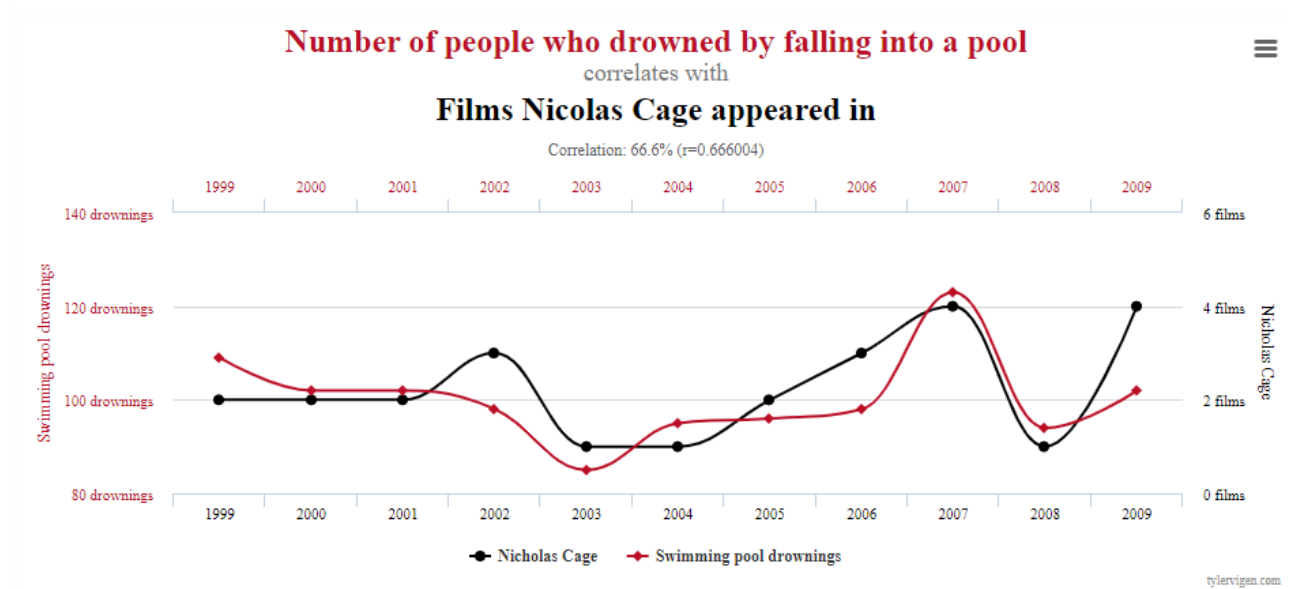
Notion de profit à tirer d'une données.

- BIG DATA**
- 
- to
- 
- BIG VALUE**

- ## Vulnérabilité

- ## Véracité

8 / 10



- Data sources: Centers for Disease Control & Prevention and Internet Movie Database

Necesite des outils de vérification et de validation des données. (recoupement, enrichissement, etc.).

(<https://www.tylervigen.com/spurious-correlations>).

Validité

Notion de conformité et de précision des données.

- Notions en parallèle avec la véracité.
- 60% du temps d'un scientifique est consacré au nettoyage de ses données avant de pouvoir effectuer une analyse.

Visibilité

Notion de visibilité des données.

- [illegible]

- ## Technologies du Big Data