

# NoSQL

---



## **Master Professionnel en Systèmes d'Information**

### **Option Ingénierie des Systèmes d'Information**

Octobre 2019

# SOMMAIRE

---

**I. Les origines du Big Data**

II. Le mouvement NoSQL

III. MongoDB

IV. Neo4j

# NoSQL

---

## **I. Les origines du Big Data**

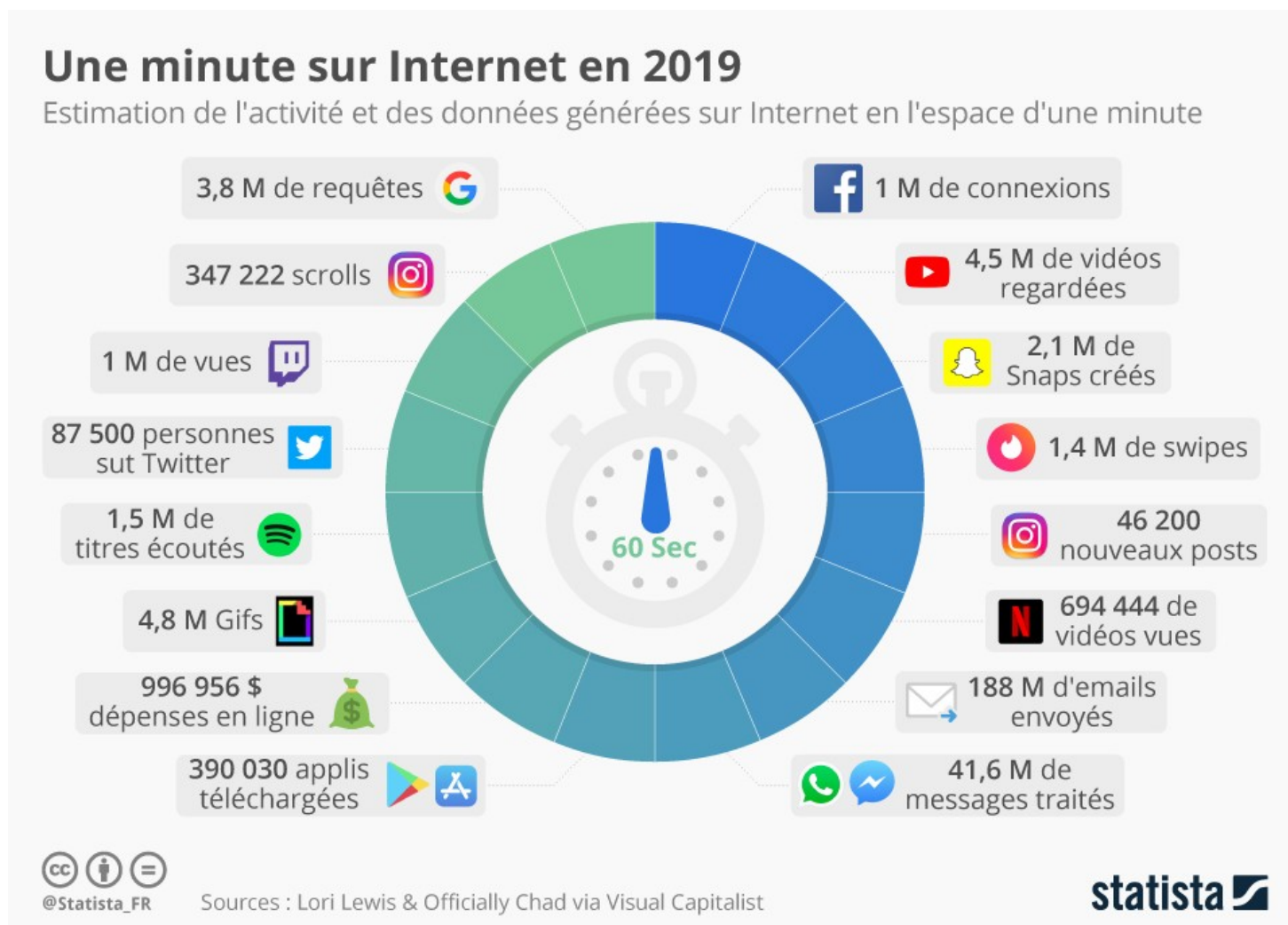
# Les origines du Big data

---

- **Un déluge d'information**
  - Une minute sur Internet
  - Format d'encodage des données
  - Unités de mesure
  - Quelques chiffres
- **Les « 3V » du Big Data**
  - Le Big Data
  - Modèle des « 3V »
  - Les applications du Big Data
- **Les avancées technologiques**
  - De nouveaux besoins
  - Les nouveaux services et architectures
  - Les nouveaux modèles de données

# Un déluge d'information

- Une minute sur Internet



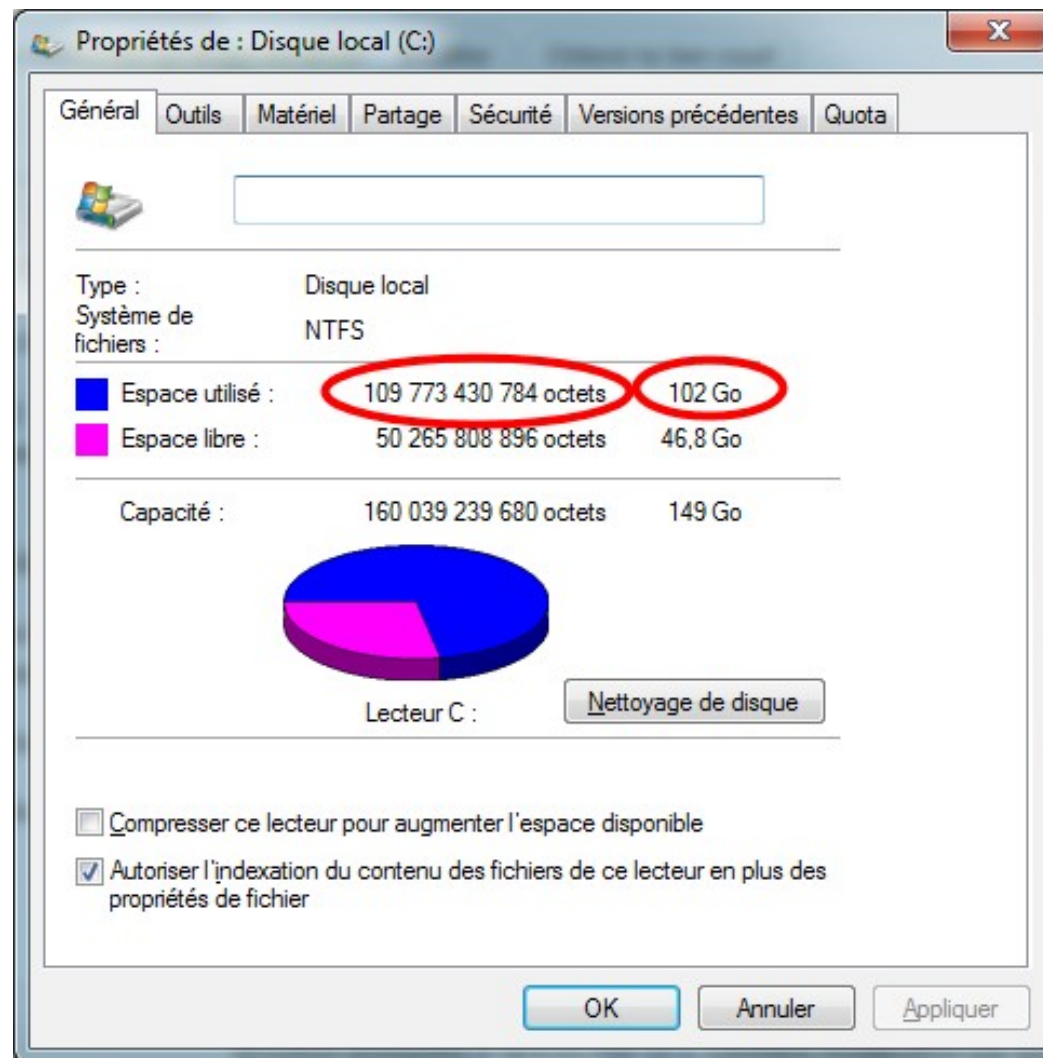
# Un déluge d'information

---

- **Format d'encodage des données**
  - Les données (texte, image, son) sont codées en binaire.
  - Suite de **bits** (0 ou 1).
  - Normes de codage des jeux de caractères:
    - **Norme ASCII** : 7 bits pour un caractère
    - **Norme ISO/CEI 646** : 8 bits = 1 octet pour un caractère
    - **Norme Unicode UTF-8** : variable de 1 à 4 octets pour un caractère.

# Un déluge d'information

- Unités de mesure



# Un déluge d'information

- Unités de mesure

Multiples de l'octet :  
préfixes décimaux du SI et mésusages

Nom	Symbole	Valeur	Mésusage <sup>a</sup>
kilooctet	Ko	$10^3$	$2^{10}$
mégaoctet	Mo	$10^6$	$2^{20}$
gigaoctet	Go	$10^9$	$2^{30}$
téraoctet	To	$10^{12}$	$2^{40}$
pétaoctet	Po	$10^{15}$	$2^{50}$
exaoctet	Eo	$10^{18}$	$2^{60}$
zettaoctet	Zo	$10^{21}$	$2^{70}$
yottaoctet	Yo	$10^{24}$	$2^{80}$

Multiples de l'octet :  
préfixes binaires

Nom	Symbole	Valeur
kibioctet	Kio	$2^{10}$
mébioctet	Mio	$2^{20}$
gibioctet	Gio	$2^{30}$
tébioctet	Tio	$2^{40}$
pébioctet	Pio	$2^{50}$
exbioctet	Eio	$2^{60}$
zébioctet	Zio	$2^{70}$
yobioctet	Yio	$2^{80}$

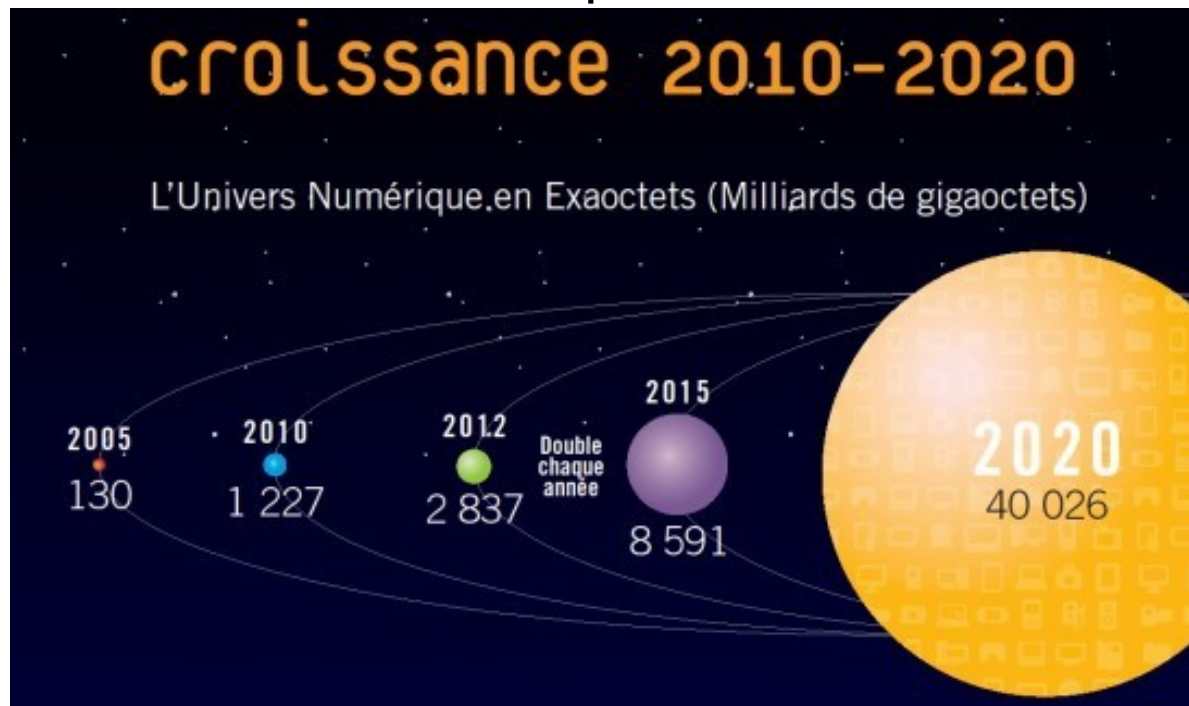


# Un déluge d'information

---

- Quelques chiffres

- En 1992, **100 Go / jour**
- En 2013, **29.000 Go / seconde**
- En 2016, 90% des données existantes avaient été créées au cours des deux années précédentes.



# Les « 3V » du Big Data

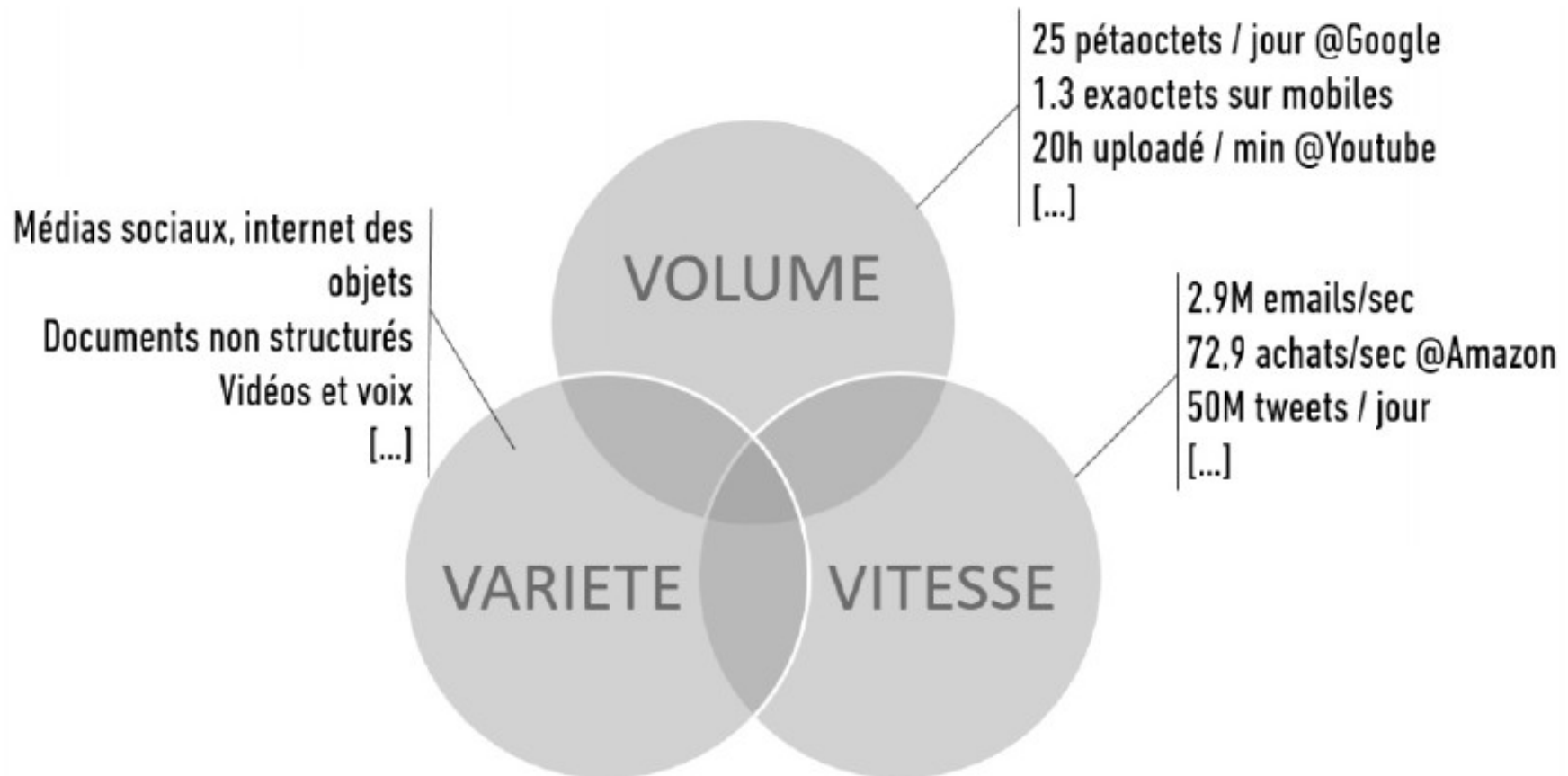
---

- **Le Big Data**

- Le terme « Big Data » que l'on peut traduire par « **mégadonnées** » ou « **données massives** », fait référence à cette explosion des données.
- Il se caractérise par le modèle dit des « 3V »,
  - **Volume**,
  - **Variété**,
  - **Vélocité**

# Les « 3V » du Big Data

- Modèle des « 3V »



# Les « 3V » du Big Data

---

- **Modèle des « 3V »**
  - **Le Volume** fait référence à la quantité d'informations, trop volumineuse pour être acquise, stockée, traitée, analysée et diffusée par des outils standards.
  - Ce caractère peut s'interpréter comme le traitement d'objets informationnels de grande taille ou de grandes collections d'objets

# Les « 3V » du Big Data

---

- **Modèle des « 3V »**

- **La Variété** fait référence à l'hétérogénéité des formats, de types, et de qualité des informations.
- Ces informations peuvent présenter des formes complexes du fait qu'elles trouvent leurs origines:
  - dans des capteurs (température, vitesse du vent, hygrométrie, tours/mn, luminosité...),
  - dans des messages échangés (e-mails, médias sociaux, échanges d'images, de vidéos, musique),
  - Sur les sites Web

# Les « 3V » du Big Data

---

- **Modèle des « 3V »**
  - **La Variété**
  - Différents niveaux de structuration des données

Niveau de structuration	Modèle de données	Exemples
Structuré	Système de données relationnel objet/colonne	Base de données d'entreprise...
Semi-structuré	XML, JSON, CSV, logs	API Google, API Twitter, web, logs...
Non structuré	Texte, image, vidéo	web, e-mails, documents...

# Les « 3V » du Big Data

---

- **Modèle des « 3V »**

- **La Vélocité** fait référence à l'aspect dynamique et/ou temporel des données, à leur délai d'actualisation et d'analyse.
- Les données ne sont plus traitées, analysées, en différé, mais en temps réel ou quasi réel. Elles sont produites en flots continus, sur lesquels des décisions en temps réel peuvent être prises.
- Ce sont les données notamment issues de capteurs, nécessitant un traitement rapide pour une réaction en temps réel.

# Les « 3V » du Big Data

---

- **Les applications du Big Data**

- **Analyser les données en mouvement**

- Le suivi en temps réel et à forte réactivité de clients (commerce de détail, téléassistance)

- **Traiter un volume conséquent de données**

- Détection de fraudes, repérage de clients indéliques ou manipulateurs

- **Découvrir et expérimenter**

- Profilage de nouveaux comportements
    - Expérimenter l'impact d'un nouveau produit, son ressenti.

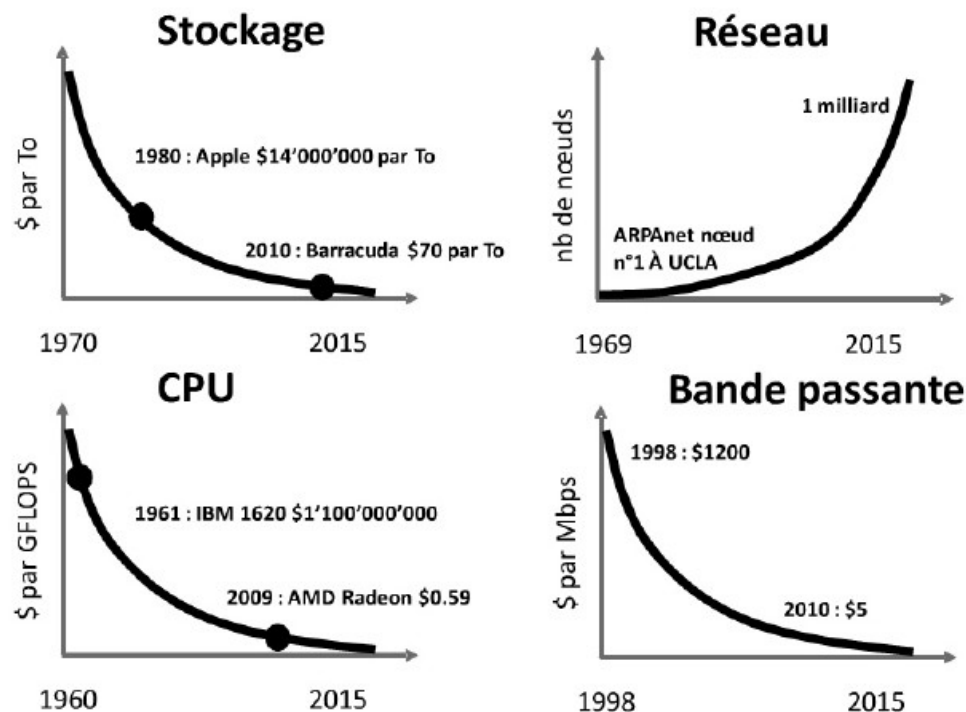
- **L'analyse prédictive et l'intelligence Artificielle**



# Les avancées technologiques

- **De nouveaux besoins**

- Essentiellement en termes de capacité de **stockage** et de **calcul**.
- Accompagnés par une baisse des coûts



# Les avancées technologiques

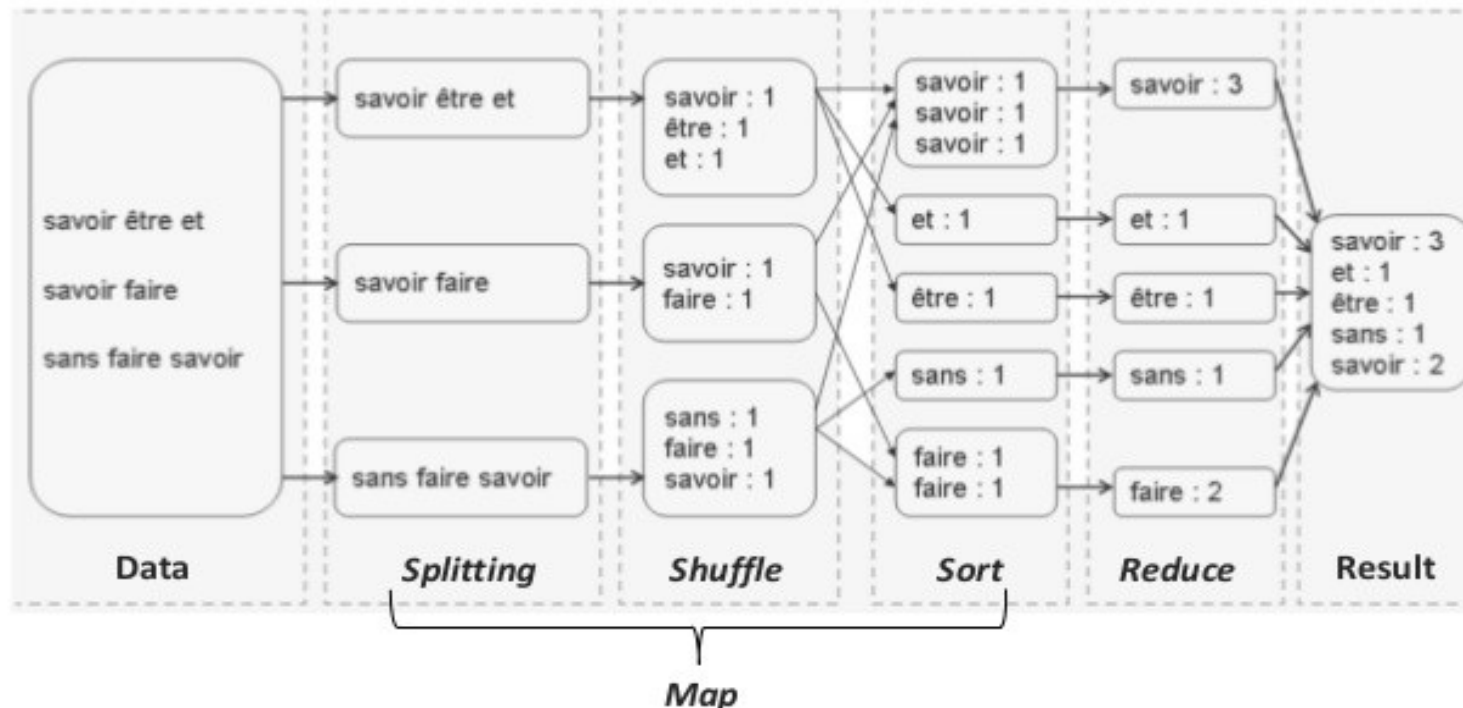
---

- **Les nouveaux services et architectures**
  - Le **Cloud Computing** a grandement facilité l'accès aux infrastructures
    - Ressources ajustables
    - Facturation à la consommation
    - Capacité de stockage illimitée
  - Les **architectures distribuées** pour le traitement et le stockage.
  - Exemple : La plateforme **Hadoop** avec
    - Son système de fichiers distribués **HDFS**,
    - Son Modèle de calcul distribué **MapReduce**

# Les avancées technologiques

- Les nouveaux services et architectures

## MapReduce Principe



# Les avancées technologiques

---

- **Les nouveaux modèles de données**

- Les SGDB (MySQL, SQL Server ou Oracle) basés sur le modèle relationnel, ne peuvent pas répondre aux exigences (les 3V) du Big Data.
- Les données doivent être stockées autrement que sous forme de tables, afin que leur traitement puisse répondre aux exigences du Big Data
- Les Bases de données **NoSQL** répondent mieux à ces exigences. Elles ont été introduites par les géants du Web, Google, Amazon, Facebook.

# Les avancées technologiques

---

- **Les nouveaux modèles de données**
  - Les bases de données NoSQL (**Not Only SQL**) se caractérisent par les propriétés suivantes :
    - **Forte distribution** des données
    - « **Schema-less** » ou schéma dynamique
  - Elles se divisent en quatre catégories
    - Modèle orienté **clé/valeur**
    - Modèle orienté **colonne**
    - Modèle orienté **document**
    - Modèle orienté **graphe**

# Les avancées technologiques

- Les nouveaux modèles de données

