

TECHNOLOGIES BIG DATA

ISGA-Marrakech

Vizle

2019-2020



<https://vizle.offnote.co>

Contact us: vizle@offnote.co

This document was generated automatically by **Vizle**

Your **Personal Video Reader Assistant**

Learn from Videos **Faster** and **Smarter**

VIZLE **PRO / BIZ**

PDF, PPT ~~Watermarks~~

- Convert *entire* videos
- *Customize* to retain all essential content
- Include Spoken *Transcripts*
- Customer support

Visit <https://vizle.offnote.co/pricing> to learn more

VIZLE **FREE PLAN**

PDF only ~~Watermarks~~

- Convert videos *partially*
- Slides may be *skipped**
- Usage restrictions
- No Customer support

Visit <https://vizle.offnote.co> to try free

Login to Vizle to unlock more slides*

BIG DATA ?

Big Data est une trace numérique que nous générons dans cette ère numérique. Cette trace numérique est constituée de toutes les données capturées lorsque nous utilisons la technologie numérique. L'idée de base derrière l'expression Big Data est que tout ce que nous faisons laisse de plus en plus une trace numérique que nous pouvons utiliser et analyser pour devenir plus intelligent.

Bernard Moore

Big Data fait référence aux volumes de données dynamiques, volumineux et disparates créés par des personnes, des outils et des machines. Elle nécessite une **technologie nouvelle**, innovante et évolutive pour **collecter, héberger et analyser** analytiquement la grande quantité de données collectées afin d'obtenir des informations en **temps réel** sur les clients, les risques, les bénéfices, la performance, la gestion de la productivité et la valeur actionnariale.

Ernst and Young

Big Data sont des ressources d'information high-volume, high-velocity, and/or high-variety qui exigent des formes de traitement de l'information rentables et innovantes permettant une meilleure compréhension, prise de décision et automatisation des processus.

Cabinet d'études Gartner

Big Data comme une collection de données provenant de **sources traditionnelles et numériques de l'intérieur et de l'extérieur** d'une entreprise qui représentent **une source continue** de découverte et d'analyse.

Lisa Arthur (collaboratrice de Forbes)

LES CAPACITÉS DE STOCKAGE

Sources: McKinsey report – Digital era, data growth, estimates and, Gartner, IDC estimates



- **Yottabyte** (1 000 000 000 000 000 000 000 000 000 Bytes) = 1024 ZB

Les noms après Yotta :

- **Xenottabyte** (1 000 000 000 000 000 000 000 000 000 Bytes)
- **Shilentnobyte** (1 000 000 000 000 000 000 000 000 000 000 Bytes)
- **Domegemegrottebyte** (1 000 000 000 000 000 000 000 000 000 000 000 Bytes)

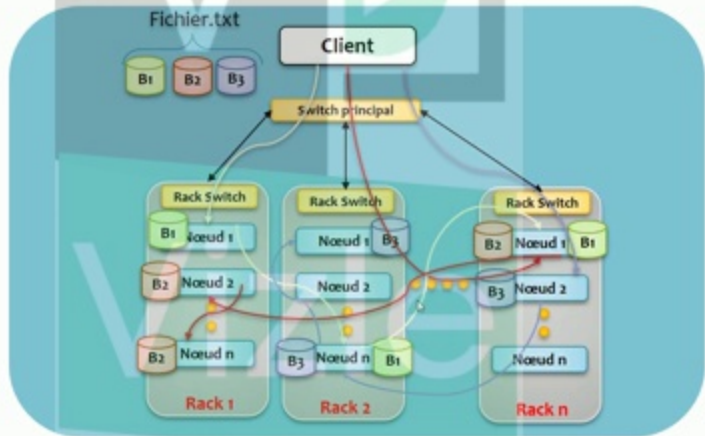
PRINCIPAUX COMPOSANTS DU HADOOP

- **HDFS : Hadoop Distributed File System**
 - Système de stockage distribué des fichiers.
- **Hadoop MapReduce :**
 - MapReduce est un framework :
 - Il permet d'écrire des applications qui traitent une grande quantité de données structurées et non structurées stockées dans le HDFS.
 - Les programmes MapReduce sont par nature **parallèles**.
- **Hadoop YARN (Yet Another Resource Negotiator) :**
 - Fournit la gestion des ressources
 - YARN est appelé en tant que système d'exploitation de Hadoop car il est responsable de la gestion et de la surveillance des jobs .



LES CONCEPTS DU HDFS

- RÉPLICATION
 - Hadoop HDFS crée des copies dupliquées de chaque bloc de données. Ceci s'appelle la réplication. Tous les blocs sont répliqués et stockés sur différents nœuds du cluster.





<https://vizle.offnote.co>

Contact us: vizle@offnote.co

This document was generated automatically by **Vizle**

Your **Personal Video Reader Assistant**

Learn from Videos **Faster** and **Smarter**

VIZLE PRO / BIZ

PDF, PPT Watermarks

- Convert *entire* videos
- *Customize* to retain all essential content
- Include Spoken *Transcripts*
- Customer support

Visit <https://vizle.offnote.co/pricing> to learn more

VIZLE FREE PLAN

PDF only Watermarks

- Convert videos *partially*
- Slides may be *skipped**
- Usage restrictions
- No Customer support

Visit <https://vizle.offnote.co> to try free

Login to Vizle to unlock more slides*