



# Gestion de Projet Big Data & Développement d'applications Big Data

EDAH Kodjo  
Consultant Système d'Information, Big Data



# Objectifs

- Comprendre la notion et les spécificités du Big Data
- Connaître les outils de collecte, de traitement et d'exploitation des données
- Savoir utiliser les outils de visualisation des données (Dataviz)
- Piloter et maîtriser les risques des projets

# Objectifs

- Connaître les technologies de l'écosystème Hadoop
- Connaître le langage python et utiliser les librairies de machine learning
- Savoir utiliser les outils de visualisation des données (Dataviz)



# Partie 1 : Définition et les enjeux du big data



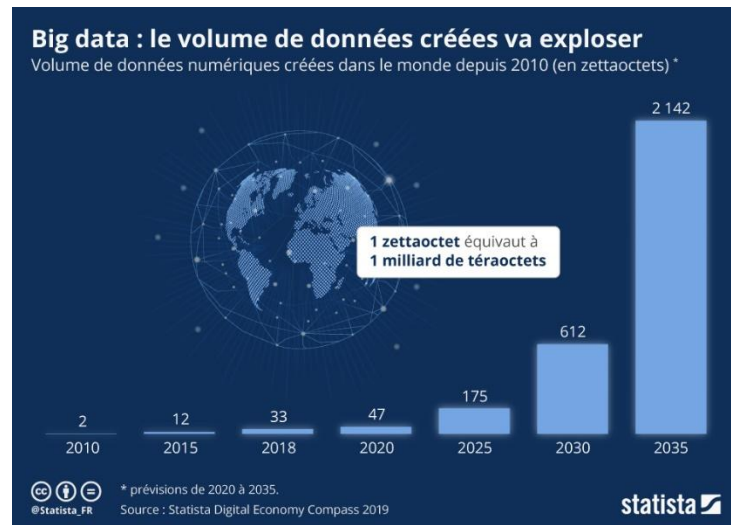
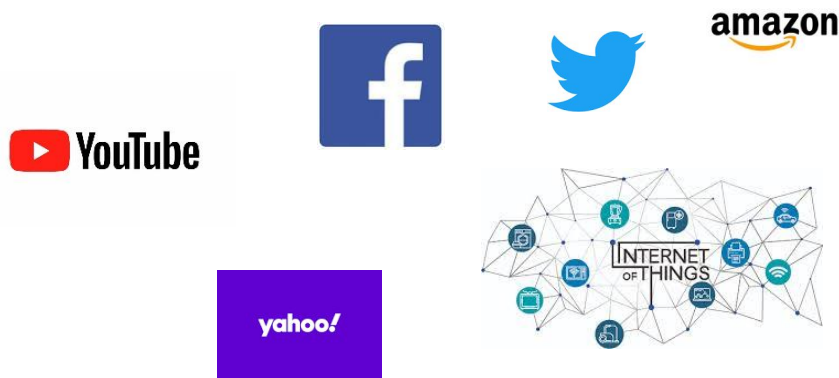
# Partie 1 : Définition et les enjeux du big data

- Définition et contexte du Big Data
- Les spécificités d'un projet Big Data
- Propriété de la donnée,  
environnement juridique du  
traitement, sécurité



# Définition et contexte du Big Data

- Big data : signifie littéralement « mégadonnées » ou « Grosses données »
- Octobre 1997 dans la bibliothèque numérique de l'ACM
- Augmentation massive des données



# Qu'est-ce que le Big Data

## Caractéristiques du big data : les 3V

- La définition de Gartner, circa 2001 : le Big Data regroupe des données présentant une grande **variété**, arrivant en **volumes** croissants, à grande **vitesse**.



# Définition et contexte du Big Data

## Caractéristiques du big data : les **5V**

- Volume : énormément de données (Péta => Exa)
- Vitesse (Vélocité) : temps réel, flot important de données
- Variété : sources, natures et formats hétérogènes (structurées et non structurées)
- Valeur : recherche de profit, corrélations
- Véracité : fiabilité, confiance, authenticité, validité, précision
- ....





# Quelques faits sur les données

- 90% des données mondiales ont été créées au cours des deux dernières années
- Chaque minute, 400h de vidéos sont uploadées sur Youtube
- Août 2015 : 1 milliards d'utilisateur par jour sur Facebook
- 75% des entreprises comptent implémenter l'IA et le Big Data

<https://www.lebigdata.fr/chiffres-big-data/>

<https://www.webrankinfo.com/dossiers/youtube/chiffres-statistiques>

# Quelques faits sur les données

Voici les **10 métiers les plus recherchés en France en 2019** selon Glassdoor :

1. **Data Scientist** – Voir les [offres d'emploi](#)
2. **Commercial** – Voir les [offres d'emploi](#)
3. **Data Analyst** – Voir les [offres d'emploi](#)
4. **Graphiste** – Voir les [offres d'emploi](#)
5. **Product Owner** – Voir les [offres d'emploi](#)



# Cas d'utilisation du big data

- La recommandation et le développement de produits :
  - ☐ Netflix : anticiper la demande des clients grâce aux modèles prédictifs
  - ☐ Amazon : livraison anticipée
  - ☐ Uber (Uber pool)



Uber

<https://www.oracle.com/fr/big-data/what-is-big-data.html#link3>

<https://www.sncf.com/fr/resseau-expertises/direction-materiel/pionniers-maintenance-predictive>

\*<https://www.altenrecrute.fr/blog-alten/decryptages/airbus-skywise-la-maintenance-predictive-de-derniere-generation>

# Cas d'utilisation du big data

- La maintenance prédictive : Analyse des données structurée ( année de fabrication, marque, modèle ...) et non structurée ( millions d'entrée de journaux, la température, les messages d'erreur ....) peuvent permettre de déployer la maintenance de manière plus rentable et d'optimiser le temps de fonctionnement des pièces et équipements

❑ SNCF



❑ Airbus (Skywise Predictive Maintenance)\*

**AIRBUS**

<https://www.oracle.com/fr/big-data/what-is-big-data.html#link3>

<https://www.sncf.com/fr/resseau-expertises/direction-materiel/pionniers-maintenance-predictive>

\*<https://www.altenrecrute.fr/blog-alten/decryptages/airbus-skywise-la-maintenance-predictive-de-derniere-generation>



# Cas d'utilisation du big data

- **Expérience client**

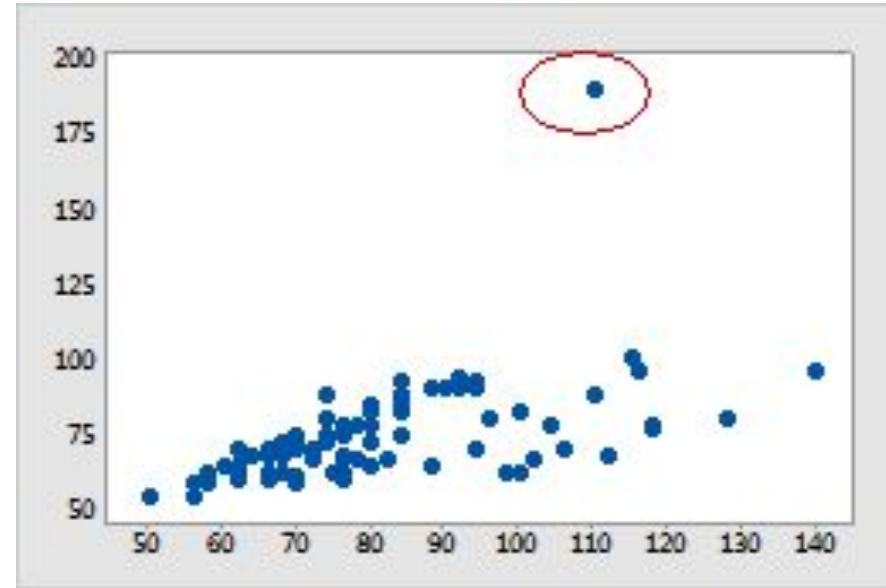
- ☐ **Collecter les données (réseaux sociaux, visites web, appels, etc.)**
- ☐ **Qualifier l'expérience vécue par le client**



# Cas d'utilisation du big data

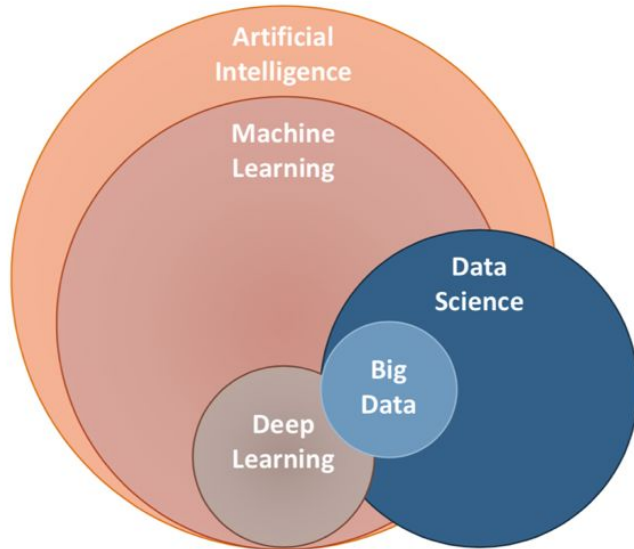
- **Fraude et conformité:**

Identifier des modèles dans les données qui indiquent une fraude et à agréger de grands volumes d'informations permettant d'accélérer le reporting réglementaire

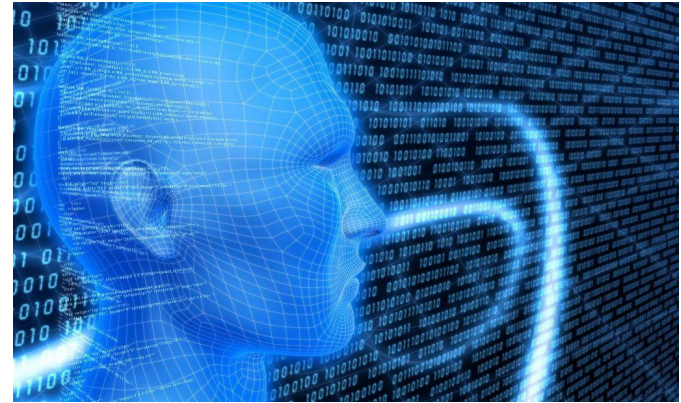


# Cas d'utilisation du big data

- Machine Learning



Source : inconnu



<https://mytf1-embedded.tf1.fr/embedplayer/13737226?context=MYTF1#amp=1>

<https://www.oracle.com/fr/big-data/what-is-big-data.html#link3>

# Les domaines d'application du big Data

## ☐ La santé :

- médecine préventive et personnalisée
- Détection d'épidémie de grippe à partir des recherches des internautes

## ☐ L'IA Bluedot avait prédit l'épidémie Coronavirus



## ☐ Le transport

- Analyse des pass permettant la géolocalisation des personnes
- Modelisation des déplacements des populations pour ajuster les horaires et infrastructures



Source : <https://i2.wp.com/europe-limousin.eu/wp-content/uploads/2019/12/transport.png?ssl=1>



# Les domaines d'application du big Data

## ❑ Energie

- ❑ Gestion de réseaux énergétiques complexes via les réseaux électriques intelligents (smartgrids)



Source : <http://melec64.free.fr/wp-content/uploads/2017/02/smart-grid-800x445.jpg>

## ❑ Aéronautique

- ❑ Données météoriques et capteurs permettant de déterminer les couloirs adéquats en temps réel



Source : [https://influenceursduweb.org/wp-content/uploads/2019/04/photos-de-couv\\_581x200.jpg](https://influenceursduweb.org/wp-content/uploads/2019/04/photos-de-couv_581x200.jpg)



# Les domaines d'application du big Data

## ☐ Sécurité

- ☐ Repérer les éventuels prémices d'une attaque ou d'une faille informatique.
- ☐ Proactive vs Réactive



Source : <https://www.ivation.fr/wp-content/uploads/2018/07/demo2.lacentraleduweb.net-pres-de-80-des-entreprises-touchees-par-des-attaques-informatiques-en-2017-7.jpg>

## ☐ Finance et banques

- ☐ Détection fraude à la carte
- ☐ Personnalisation des services
- ☐ Marketing prédictif



Source : [https://www.professioncgp.com/uploads/media/cache/articles\\_large/articles/0001/06/5281\\_articles\\_large.jpg](https://www.professioncgp.com/uploads/media/cache/articles_large/articles/0001/06/5281_articles_large.jpg)



# Les domaines d'application du big Data





# Spécificité d'un projet big data





# Spécificité d'un projet big data

DATA

INFORMATION

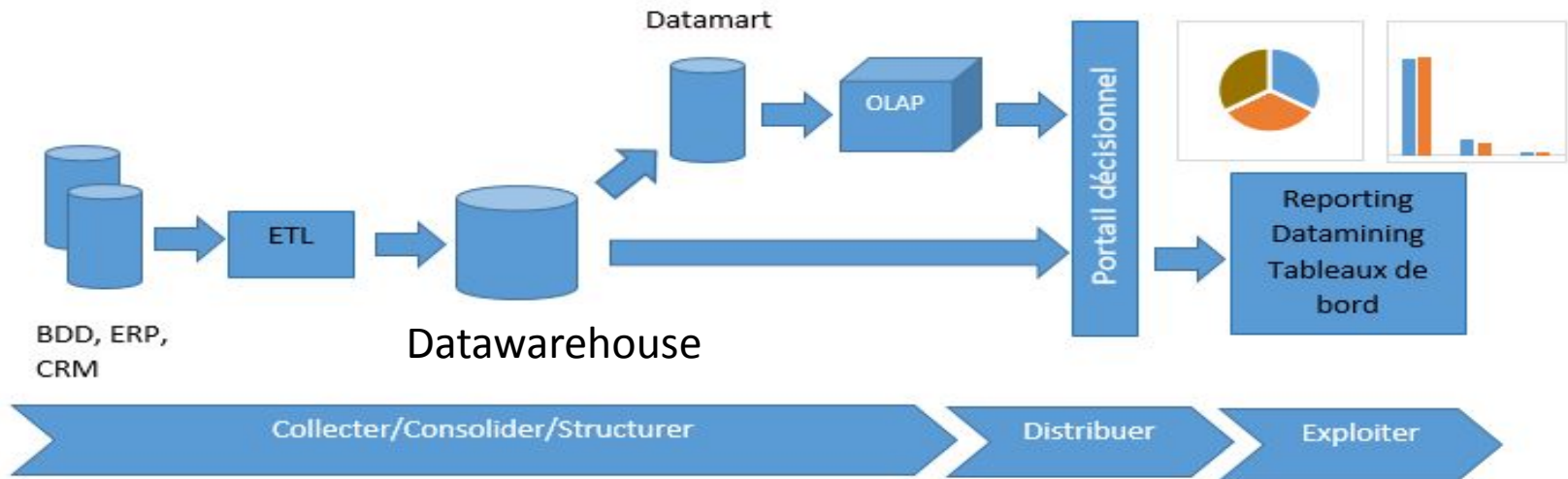
KNOWLEDGE

DECISION



# Avant le big data ... Business Intelligence

- Ensemble des technologies (outils et méthodes) permettant aux entreprises d'analyser les données au profit de leurs prises de décisions.



# Business Intelligence (BI) et Big Data

## Point Commun

- ❑ Récupérer et traiter des données pour atteindre de meilleures performances
- ❑ Utiliser les données pour aider les entreprises dans leur prise de décision



# Différence entre Business Intelligence (BI) et Big Data

BI		Big data
Descriptive	Approche	Exploratrice
Datawarehouse	Stockage	Système distribué
Généralement Structuré	Type de donnée	Structuré et non structuré
Métiers	Compétence	Data scientist, Métiers
Historique	Temps	Historique et en temps réel





# Technologies, compétences et métiers



# Technologies, compétences et métiers

	Python	R	Scala	Java
Popularité	+++	++	+	+
Librairies analytiques	+++	+++	++	+
Scalabilité	++	+	+++	++
Machine Learning	+++	+++	++	++
Compétences	++	++	+	+++

Source : M Roodnejad, S. Sakly - 2018



# A partir de quand peut-on parler de Big data ?

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



# A partir de quand peut-on parler de Big data ?

- ❑ 100 To ? Par jour/mois/année ?
- ❑ Les données considérées comme « Big » au début 2000 sont aujourd'hui requalifiées de « Small »
- ❑ La notion de « Big » est donc relative !!!
- ❑ *Rappel : Le big data consiste à traiter un ensemble de données plus ou moins structurées, volumineuses qu'elles sont difficiles à travailler avec des outils classiques de gestion de base de données*





# A vous de jouer (30min)

---

- ☐ Chercher un domaine d'application du Big data
- ☐ Chercher une technologie utilisée dans le domaine du Big Data
- ☐ Une actualité sur le Big Data



Questions ?

---

Merci

