



Présentation de

BIG DATA

ITAB ACADEMY



Définition du BIG DATA

Pourquoi le BIG DATA?

Les données produites jusqu'en 2003 étaient de 5 milliards de gigaoctets.

Cela peut remplir un terrain de football entier si les données sont insérés dans des disques durs.



- Pourquoi le BIG DATA ?
- The data that was produced till 2003 was **5 billion gigabytes**.
- This may **fill an entire soccer field** if put in hard drives.

Pourquoi le BIG DATA?

Reseaux sociaux:

Facebook compte 2,27 milliards d'utilisateurs actifs par mois.

Twitter compte 336 millions d'utilisateurs actifs par mois.

Données des capteurs:

Données générées par les CAMERAS, IOT Devices, etc.

Fichier LOG:

Fichiers générés par les machines.



- Pourquoi le BIG DATA ?
- **Data volume:**
 - Social media:
 - Facebook has 2,27 billion monthly active users.
 - Twitter has 336 million of monthly active users.
 - Sensor data:
 - Data generated by CAMERA, IOT Devices etc..
 - Log files:
 - Files generated by the machines.

Pourquoi le BIG DATA?

Des médias sociaux:

Beaucoup d'entreprises veulent analyser les sentiments de leurs clients.

Données du capteur:

- ✓ Les gouvernements veulent détecter les crimes en utilisant des caméras de rue et signaler les personnes et les marchandises suspectes aux postes frontières et aux aéroports.
- ✓ Analyse automatique et notification de la police.

Fichiers journaux:

- ✓ Les entreprises de sécurité veulent détecter les attaques avant qu'elles ne se produisent à l'aide de ces fichiers journaux

- Pourquoi le BIG DATA ?
- **Data Analytics:**
 - **Social media:**
 - A lot of companies wants to analyze the sentiments of their clients.
 - **Sensor data:**
 - Governments wants to detect crimes using street cameras and flag suspicious people and cargo at border crossings and airports.
 - Analyze automatique et notification de la police.
 - **Log files:**
 - Security companies wants to detects attacks using these log files.

C'est quoi le BIG DATA?

BIG DATA désigne un ensemble de données volumineuses qui ne peuvent pas être traités à l'aide de techniques informatiques traditionnelles.

BIG DATA ne désigne pas simplement des données, c'est un sujet complet impliquant divers outils, techniques et frameworks.



- C'est quoi le BIG DATA?
- Big data means is a collection of large datasets that **cannot be processed** using traditional computing techniques.
- Big data is not merely a data, rather it has become a **complete subject**, which involves various tools, techniques and frameworks.

BIG DATA DEFINITION

- ▶ The **industry-standard** way to describe big data is with the “Four or Five Vs”:

- ▶ Volume,
- ▶ Velocity,
- ▶ Variety ,
- ▶ Veracity
- ▶ And Valeur.



C'est quoi le BIG DATA?

Volume

- ✓ Le terme Big Data fait référence à de très grandes quantités de données.

Big Data Is All Data

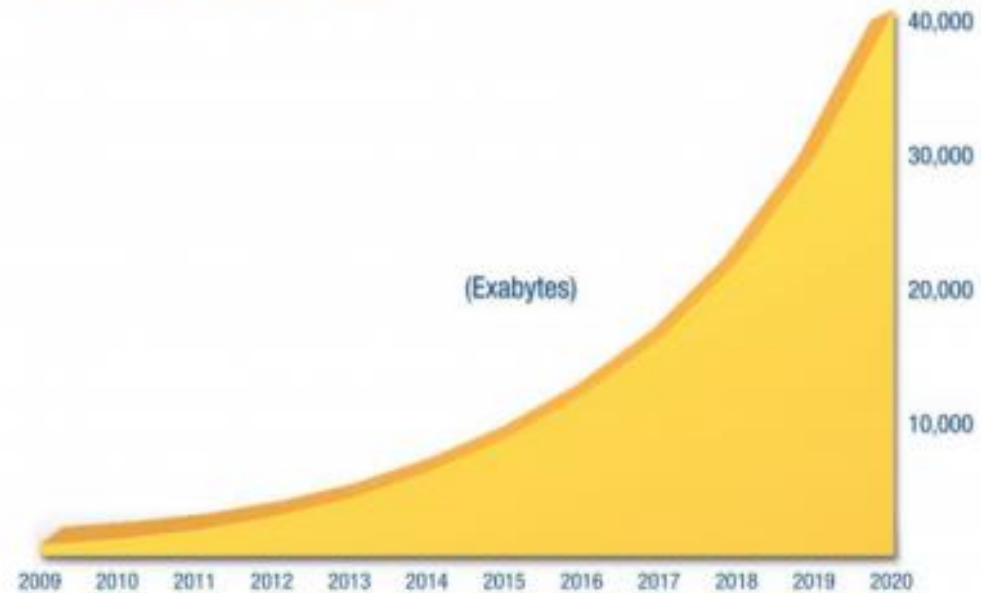


- C'est quoi le BIG DATA?

- **Volume**

- The term big data refers to very large quantities of data.

The Digital Universe: 50-fold Growth from the Beginning of 2010 to the End of 2020



Source: IDC's Digital Universe Study, sponsored by EMC, December 2012

What is Kilo, mega, giga, tera, peta, exa, zetta (2 to the power 10 or 1024)

C'est quoi le BIG DATA?

Rapidité:

- ✓ Les données sont créées à un rythme très rapide et doivent être traitées très rapidement.
- ✓ Par exemple, le flux de données provenant de sources des réseaux sociaux représente des données volumineuses à grande vitesse.

Big Data Is All Data



- C'est quoi le BIG DATA?
- **Velocity:**
- Data is being created at a very rapid pace and needs to be processed very quickly.
- For example, the stream of data coming from social media feeds represents big data with a high velocity.



C'est quoi le BIG DATA?

Le BIG DATA proviennent d'une grande variété de sources et se présentent sous différents formats. Un référentiel Big Data peut inclure:

- ✓ fichiers texte,
- ✓ images,
- ✓ vidéo,
- ✓ fichiers audio,
- ✓ des présentations,
- ✓ feuilles de calcul

Big Data Is All Data



- C'est quoi le BIG DATA?
- **Variety**
- Big data comes from a wide variety of sources and resides in many different formats. A big data repository might include:
 - text files,
 - images,
 - video,
 - audio files,
 - presentations,
 - spreadsheets



C'est quoi le BIG DATA?

✓ Véracité

La qualité des données capturées peut varier considérablement, ce qui affecte l'analyse précise.

Big Data Is All Data



• C'est quoi le BIG DATA?

• Veracity:

- The data quality of captured data can vary greatly, affecting the accurate analysis.

Min	Max	Mean	SD
4.3	?	5.84	0.83
2.0	4.4	3.05	50000000
15000	7.9	1.20	0.43
0.1	2.5	?	0.76

C'est quoi le BIG DATA?

✓ valeur

Le Big Data favorise l'innovation et crée ainsi de la **valeur**.

Big Data Is All Data



- C'est quoi le BIG DATA?

- **valeur:**

- Le Big Data possède un autre V qui mérite d'être pris en compte.
- Le cinquième V du Big Data est la Valeur.
- C'est le résultat de décisions commerciales intelligentes tirées parti des quatre V précédents.
- L'objectif ultime d'une organisation est de :
 - produire de la valeur sous forme de décisions commerciales plus rapides et plus éclairées, optimiser l'utilisation des ressources et découvrir de nouvelles opportunités de marché.



BIG DATA Use Cases

BIG DATA use Cases



- **BIG DATA Use Cases**
- 360° View of the Customer
- Fraud Prevention
- Data Warehouse Offload
- Price Optimization
- Recommendation Engines
- Social Media Analysis and Response (ex. rumors)
- Preventive Maintenance and Support
- Internet of Things
- Medicine



Vu à 360° du Client

Imaginez un tableau de bord à 360 ° du client. Il peut inclure des données provenant de différentes sources:

- ✓ Données démographiques, telles que noms de clients, adresses,
- ✓ informations de vente.
- ✓ Informations de la solution de gestion de la relation client (CRM) sur:
 - ✓ interactions passées avec l'entreprise
 - ✓ et même fournir des liens vers les transcriptions des appels récents,
 - ✓ messages électroniques
 - ✓ ou des sessions de discussion.
 - ✓ Historique du support
- ✓ Pages du site Web récemment visité par un client -> fournissant des indices précieux sur la raison pour laquelle un client pourrait appeler.
- ✓ Derniers articles publiés par le client sur les réseaux sociaux.

Toute cette information aiderait évidemment à préparer le personnel de l'entreprise à mieux interagir avec le client.

- **360° View of the customer**
- Imagine a dashboard with a **360° of the customer**. It can include data from various sources:
 - Demographic data, like customers' names, addresses,
 - sales information.
 - Information from the customer relationship management (CRM) solution about:
 - past interactions with the firm
 - and even provide links to transcripts of recent calls,
 - email messages
 - or chat sessions.
 - Support history
 - Pages of the company website a customer had recently visited -> providing valuable clues about the reason a customer might be calling.
 - Customer's recent social media posts.



Vu à 360° du Client

Supposons qu'un client appelle un opérateur télécom. L'agent dispose d'un tableau de bord avec:

- ✓ Ce que le client a commenté à propos de l'opérateur Hier sur sa page facebook.
- ✓ Les pages vu par le client sur le site Web de l'opérateur.

Toute ces information aiderait évidemment à préparer le personnel de l'entreprise à mieux interagir avec le client.

- **360° View of the customer**
- Suppose a client call a telecom operator. The operator agent has a dashboard with:
 - What the client has commented about the Operator yesterday in his facebook page.
 - What Webpages the client has seen of the operators Website.



Vu à 360° du Client

Si le tableau de bord est utilisé avec des outils d'analyse avancés ou d'apprentissage automatique, il peut alors:

- ✓ suggérer des opportunités de vente croisée (vente d'articles complémentaires) ou de vente incitative (vente d'autres produits coûteux) sur les produits,
- ✓ ou s'il détecte qu'un client pourrait être en danger de passer à un concurrent,
- ✓ cela pourrait suggérer des remises potentielles pouvant abaisser le tarif du client.
- ✓ Certains outils peuvent même analyser le langage des clients pour détecter leurs émotions actuelles et suggérer des réponses appropriées aux agents des ventes ou du service clientèle.

- **360° View of the customer**
- If the dashboard is used with advanced analytics or **machine learning tools**, then it could:
 - suggest opportunities for **cross-selling** (selling complementary items) or **upselling** (selling other expensive products) customers on products,
 - or if it detects that a **customer might be in danger** of defecting to a competitor,
 - it might suggest potential **discounts** that could lower the customer's rate.
 - Some tools can even analyze customers' language to **detect their current emotions** and suggest appropriate responses to sales or customer service agents.



Vu à 360° du Client

Cela peut sembler futuriste?

Mais de **nombreuses entreprises** disposent déjà de systèmes comme celui-ci et les utilisent pour améliorer la satisfaction de leurs clients et augmenter leurs revenus et leurs marges.

- **360° View of the customer**
- This seems futuristic?
- but **many companies** today already have systems like this one in place, and they are using them to improve customer satisfaction and increase revenues and margins.



Prévention de fraude

Sans BIG DATA

Les émetteurs de cartes de crédit utilisaient des systèmes basés sur des règles pour les aider à signaler les transactions potentiellement frauduleuses.

si une carte de crédit était utilisée pour louer une voiture à Asila , mais que le client vivait à Paris, un agent du service clientèle pourrait appeler pour confirmer.

- Fraud prevention
- **Without BIG DATA**
 - Credit card issuers were using **rules-based systems** to help them flag potentially fraudulent transactions.
 - if a credit card were used to rent a car in Hawaii, but the customer lived in Omaha, a customer service agent might call to confirm.



Prévention de fraude

Avec BIG DATA

Un système sophistiqué de prévention de fraude pourrait vérifier que le client a récemment acheté:

billets d'avion,
crème solaire
et un nouveau maillot de bain avant la location de voiture.

Sur la base de modèles historiques, un système d'analyse prédictive ou d'apprentissage automatique pourrait indiquer que la voiture de location est donc **moins susceptible de constituer un achat frauduleux.**

- Fraud prevention
- **With BIG DATA**
 - A sophisticated fraud prevention system might be able to see that the customer had recently purchased:
 - airline tickets,
 - sunscreen
 - and a new swimsuit before the rental car purchase.
 - Based on historical patterns, a predictive analytics or **machine learning** system would be able to tell that the rental car was thus **less likely to be a fraudulent purchase.**



Prévention de fraude

Avec BIG DATA

La fraude a tendance à être concentrée dans certaines régions géographiques, souvent près des aéroports, ce qui permet aux criminels de transporter facilement les biens volés.

Les codes postaux les plus risqués ont tendance à changer avec le temps.

L'analyse de données volumineuses peut examiner les enregistrements antérieurs de transactions frauduleuses et identifier rapidement les tendances de changement.

Identifiez où les criminels résideront dans le futur.

- Fraud prevention

- With BIG DATA



- Fraud tends to be **concentrated** in certain geographic regions—often near airports, which make it easy for criminals to move stolen goods.
- Zip codes are **riskiest** tends to **change over time**.
- Big data analytics can look at past records of fraudulent transaction and
 - quickly identify changing trends.
 - Identify where criminals will be **residing in the future**.



Renseignement de sécurité

Les entreprises utilisent également les analyses de données volumineuses pour les aider à arrêter les pirates informatiques et les cyber-attaquants.

Les serveurs d'une entreprise génèrent des fichiers journaux, d'autres sont disponibles à la vente.

De nombreuses organisations analysent toutes les informations des fichiers logs pour prévenir et détecter les attaques en temps réel.

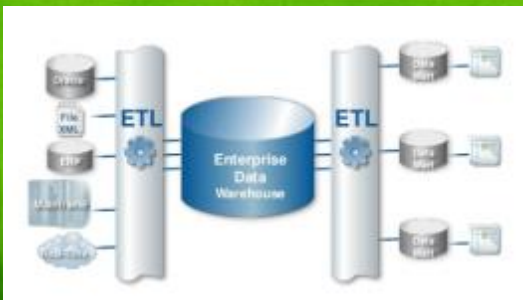
- **Security intelligence**
- Organizations are also using big data analytics to help them stop hackers and cyber attackers.
- Servers in a company are generating log files, other log files are available for sale.
- Many organizations analyze all of the internal and external log information to prevent and detect attacks in real time.

Problème

Il est courant de disposer d'un entrepôt de données facilitant les efforts de Business Intelligence (BI).

La technologie d'entrepôt de données a tendance à être très coûteuse à l'achat et à l'exploitation.

Les solutions d'entrepôt de données n'ont pas toujours été en mesure de fournir les performances souhaitées.



- **Data Warehouse Offload**
- Problem
 - It is common to have a data warehouse that facilitates the business intelligence (BI) efforts.
 - Data warehouse technology tends to be very costly to purchase and run.
 - The data warehouse solutions haven't always been able to provide the desired performance.

Solution

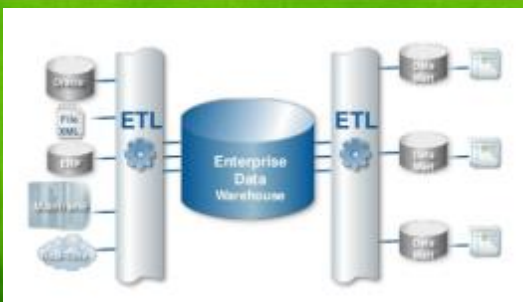
De nombreuses entreprises utilisent une solution Big Data Open Source telle que Hadoop pour remplacer ou compléter leurs entrepôts de données.

Les solutions basées sur Hadoop offrent souvent des performances beaucoup plus rapides tout en réduisant les frais de licence et autres coûts.

- **Data Warehouse Offload**

- **Solution**

- Many enterprises use an open source big data solution like Hadoop to replace or compliment their data warehouses.
- Hadoop-based solutions often provide much faster performance while reducing licensing fees and other costs.



Moteurs de recommandation

Lorsque vous faites des achats sur le site d'Amazon, vous prenez probablement comme acquis que le site Web suggérera des articles similaires qui pourraient vous intéresser.

Bien entendu, la possibilité d'offrir ces recommandations découle de l'analyse BIG DATA pour analyser les données historiques.

- **Recommendation Engines**
- When you are shopping for products from Amazon, you probably now take it for granted that the website will suggest similar items that you might enjoy.
- Of course, the ability to offer those recommendations arises from the use of big data analytics to analyze historical data.



Moteurs de recommandation

Ces moteurs de recommandation sont devenus si courants sur le Web que de nombreux clients les attendent maintenant lorsqu'ils effectuent des achats en ligne.

De plus, les entreprises qui n'ont pas exploité leur Big Data de cette manière risquent de perdre des clients au profit de concurrents ou de perdre des opportunités de vente incitative ou de vente croisée (règles d'association).

- **Recommendation Engines**
- These recommendation engines have become so commonplace on the Web that many customers now expect them when they are shopping online.
- And organizations that haven't taken advantage of their big data in this way may lose customers to competitors or may lose out on upsell or cross-sell opportunities (association rules).



Analyse et réponse des médias sociaux

Les entreprises sont censées surveiller ce que les gens disent à leur sujet (Facebook, Twitter, Instagram) sur les médias sociaux et réagir de manière appropriée - Analyse des sentiments.

Ex. Le ministère du tourisme – sentiment analysis des touristes

S'ils ne le font pas, ils perdent rapidement des clients.

- **Social Media Analysis and Response**
- Companies are expected to monitor what people are saying about them (Facebook, Twitter, Instagram) in social media and respond appropriately — Sentiment Analysis.
- If they do not, they quickly lose customers.





L'Internet des objets

L'Internet des objets fait référence au réseau en pleine expansion d'objets connectés capables de collecter et d'échanger des données à l'aide de capteurs intégrés.

- ✓ **Maison intelligente:** La maison intelligente est probablement l'application IoT la plus populaire à l'heure actuelle. Contrôle des portes et des fenêtres à l'aide d'Internet (domotique).
- ✓ **Appareils portables:** la montre Apple Watch et d'autres montres intelligentes sur le marché ont transformé nos poignets en étui pour smartphone en permettant la messagerie texte, les appels téléphoniques, etc.

- **Internet of Things**
- The Internet of Things refers to the rapidly growing network of connected objects that are able to collect and exchange data using embedded sensors.
 - **Smart Home:** The smart home is likely the most popular IoT application at the moment. Control doors and windows using internet (domestics).
 - **Wearables:** The Apple Watch and other smartwatches on the market have turned our wrists into smartphone holsters by enabling text messaging, phone calls, and more.



L'Internet des objets

- ✓ **Villes intelligentes:** L'Internet des objets peut résoudre les problèmes de congestion routière et réduire le bruit, la criminalité et la pollution. (reconnaissance du visage en Chine, caméra acoustique qui détecte des voitures qui klaxonnent en Chine et envoie l'amende)
- ✓ **Voiture connectée:** Ces véhicules sont équipés d'un accès Internet et peuvent partager cet accès avec d'autres, tout comme la connexion à un réseau sans fil dans une maison ou un bureau. (papier en Turquie)

- **Internet of Things**

- **Smart Cities:** The Internet of Things can solve traffic congestion issues and reduce noise, crime, and pollution. (face recognition in China, acoustic camera that detects cars honking)
- **Connected Car:** These vehicles are equipped with Internet access and can share that access with others, just like connecting to a wireless network in a home or office. (paper in Turkey)



L'Internet des objets

Exemple de réfrigérateur intelligent:

- ✓ Programmer les repas pour toute la semaine.
- ✓ Réfrigérateur intelligent peut demander automatiquement la livraison de nourriture pour les ingrédients manquants.
- ✓ Supposons que nous avons accès aux dossiers médicaux des clients et on sait ce qu'ils mangent.
- ✓ Et si nous connaissions cette information pour des millions d'utilisateurs?
- ✓ Nous pouvons savoir, en appliquant l'apprentissage automatique, que certains aliments peuvent causer certaines maladies.

- **Internet of Things**
- Example of smart fridge:
 - Setting meal program for whole week.
 - Smart fridge can automatically ask for food delivery for missing ingredients.
 - Suppose we have medical records of the family.
 - What if we know this info for millions of users?
 - We can know by applying machine learning that certain food can cause certain diseases.



- **Medicine - Diagnostics**
- A 2018 BBC story revealed that in the US alone, diagnostic errors result in 40,000-80,000 deaths.
- Machine Learning along with historical data can allow them to detect early signs of an issue and enable the doctors to provide preventive care and better treatment to the patients.
- Additionally, this data can also be used by medical researchers to diagnose chronic diseases at early stages and identify treatment options that have proven success records.

Autres cas d'utilisation de BIG DATA

Les agriculteurs utilisent le BIG DATA pour trouver le meilleur moment pour planter ou récolter.

Les équipes sportives professionnelles utilisent des analyses pour déterminer qui devrait figurer sur la liste des joueurs afin d'améliorer leurs performances.



Les traders financiers utilisent le BIG DATA pour déterminer quand acheter ou vendre.

- **Other BIG DATA uses cases**
 - Farmers use big data to find the best time to plant or harvest.
- 
- A red tractor with a green field in the background, illustrating the use of big data in agriculture.
- Professional sports teams use analytics to decide who should be in the list of players to help improve player performance.
- 
- A person sitting at a desk with multiple computer monitors displaying data, illustrating the use of big data in financial trading.
- Financial traders use big data to determine when to buy or sell.