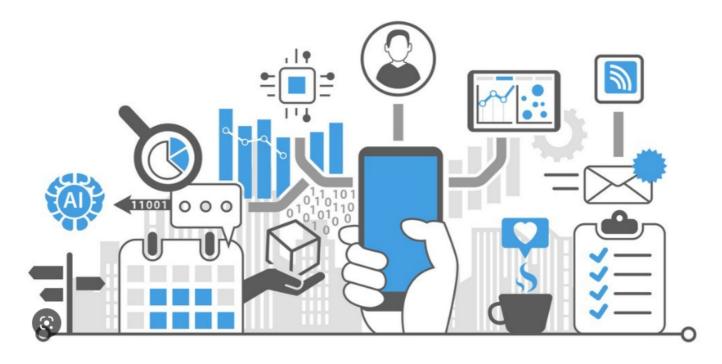
# Big Data

# **BIG DATA**





C'est quoi le Big Data?



### Big Data: Constat

- Explosion des capacités de stockage et de calculs des ordinateurs
- Augmentation exponentielle du nombre de données générées et stockées
- Augmentation exponentielle du nombre de données connectées
- Developpment des technologies de virtualisation et de cloud computing

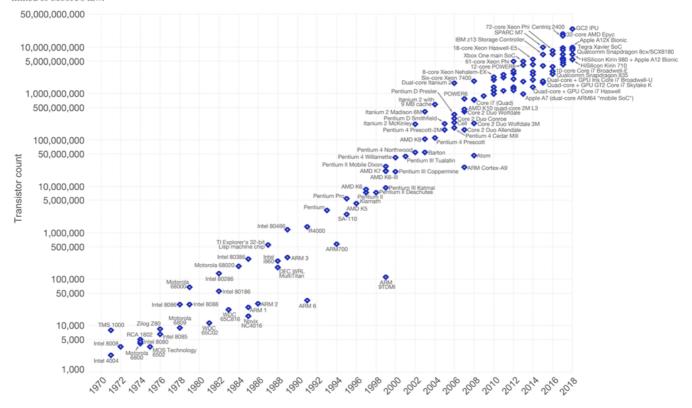
## Big Data : loi de Moore

Evolution de la puissance de calculs des processeurs.

#### Moore's Law – The number of transistors on integrated circuit chips (1971-2018)



Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important as other aspects of technological progress – such as processing speed or the price of electronic products – are linked to Moore's law.

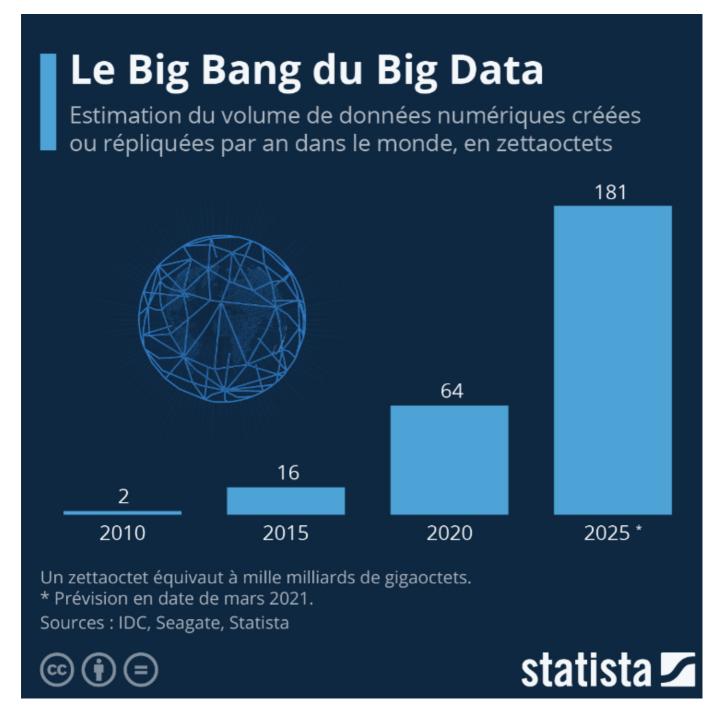


Data source: Wikipedia (https://en.wikipedia.org/wiki/Transistor\_count)
The data visualization is available at OurWorldinData.org. There you find more visualizations and research on this topic.

Licensed under CC-BY-SA by the author Max Roser.

### Big Data : Volume de données

Generation sans cesse croissante de données.



#### rappel:

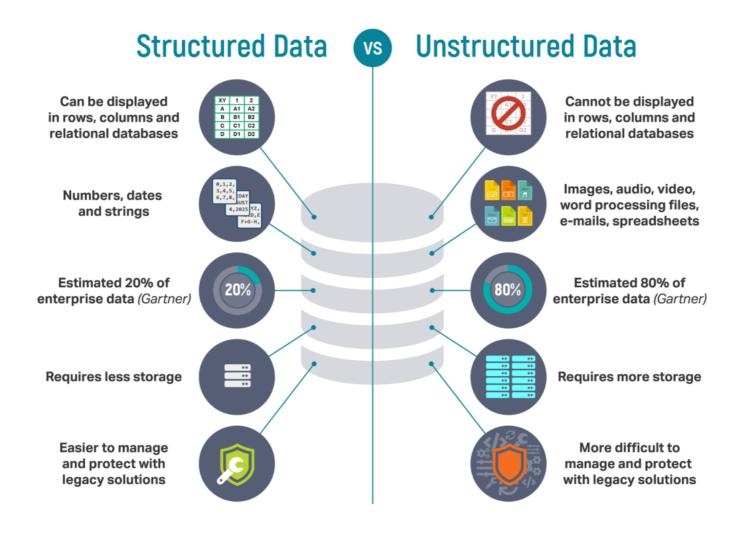
• 1 Zo = 1000 000 000 000 Go

• 1 Zo = 10^21 octets

Big Data : données connectées

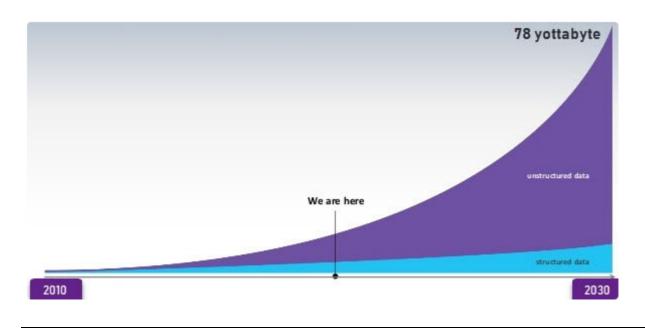
## Une minute sur Internet en 2021 Estimation de l'activité et des données générées sur Internet en l'espace d'une minute 28 000 2 M de views abonnés en ligne 695 000 1,36 M € dépensés en lignes stories partagées 9 132 de 2 M de Swipes connexions faites 197.6 M de 69 M de mails envoyés messages envoyés 5 000 500 heures de contenu mises en ligne téléchargements Source: Lori Lewis via AllAccess statista 🔽

Structurées vs non structurées



#### Big Data: Données non structurées

Untilisation de données semi ou non-structurées. de plus en plus importantes



Limitation des systemes classiques

- Données trop voluineuses pour être stockées sur un seul serveur.
- Données trop variées pour être stockées facilement.
- Données changeantes trop rapidements pour être stockées/traitées facilement.
- Données trop complexes pour être traitées facilement.

#### Le Big Data

Ensemble de technologies et de méthodes permettant de stocker, traiter et analyser des données massives, variées et changeantes.

But : permettre un traitement efficace de données massives, variées et changeantes à un coup financier, humain et temporel raisonnable.

#### Big Data: historique

- Concept ancien issu des années 70 avec les premiers datacenter
- le terme serait apparu en 1997.

#### Big data: tous est informations

Principe que l'homme et la totalité du monde qui l'entour peuvent être représentés comme :



- « des ensembles informationnels, dont la seule différence avec la machine est leur niveau de complexité.»
- « La vie deviendrait alors une suite de 0 et de 1, programmable et prédictible ».

#### Big Data: MapReduce et NoSQL

• Concretiation dans les années 2005 par **Google** qui deploye un algorithme sur des operations massives.

le MapReduce qui deviendra le projet hadoop.

• Dans les années 2009, le deployement de stockage open-source, distribués et non-relationnels

Le NoSQL.

#### Big Data: Concepts

Repose sur le concept de parralelisme et de distribution des algorithmes de traitement et du stockage des données.

Diviser pour mieux régner. C'est à dire repartir le stockage ou le traitements des données massives sur plusieurs machines.

#### Big Data: Domaines d'application

L'utilisation du Big Data est très large et touche de nombreux domaines :

- Commerce : analyse des données clients, prévision, recommandation, etc.
- Finance : analyse des données financières, prévision, etc.
- Santé: analyse des données médicales, prévision, etc.

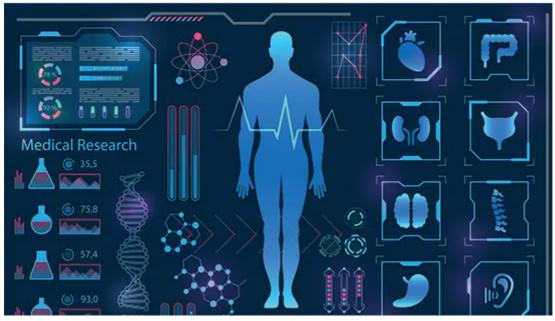
### Big Data: Domaines d'application

L'utilisation du Big Data est très large et touche de nombreux domaines :

- Energie : analyse des données énergétiques, prévision, etc.
- Transport : analyse des données de transport, prévision, etc.
- Agriculture : analyse des données agricoles, prévision, etc.

#### Exemple: en santé

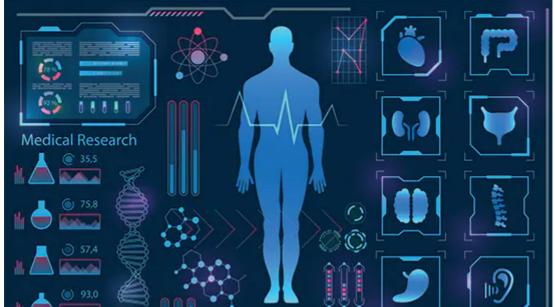
• Recherche sur l'effet d'un médicament dans une maladie.



- Collecter des données à partir de milliers/millions de patients.
- Sources multiples et variées (dossiers médicaux électroniques, des bases de données d'assurance maladie, des registres et des enquêtes de santé publique).

#### Exemple: en santé

- Analyses des données.
- Identifier tendances, corrélations avec les modèles, efficacité du médicament, facteurs de risque,



traitements

associés, et les impacts sur la qualité de vie.

 Aide à decision, adaptabilité des traitements, prévision des effets secondaires ou de l'evolution de la maladie.

### Big Data: Domaines technique

Croisement de nombreuses spécialités techniques :

- Informatique transactionnelle (principe ACID, etc.)
- Informatique décisonnelle (BI, prise de decision, etc.)
- Informatique en temps reel (temps de reponse critique)

#### Big Data: Domaines technique

Croisement de nombreuses spécialités techniques :

- Stockage et tri des données (besoin volume, rapidité, etc.)
- Traitement et analyse des données (catégorisations, synthèse, prédictions, représentations, etc.)

Carracteriques du Big Data.