

Big Data

La numérisation du monde

SI7

Partie 1

Introduction

Introduction

- * Ère industrielle (automatisation)
 - * Ère numérique (digitalisation)
 - * Ère suivante ... ?
-
- * Origine : internet, web 2.0 (généricité)
 - * Big Data (spécialisation)
 - * Création de grandes masse de données
 - * Big Data : littéralement « grosses données »

Introduction

- * Sources : GAFA, NATU, BATX, économie, science, industrie
- * Corrélation possible avec la globalisation croissante
- * Émergence de nouveaux marchés
- * Le contenu est essentiellement produit par les individus (smartphone, tablettes, PC, ...), les machines
- * Infrastructures nécessaires : fibre optique, WIFI, 4G, satellites, ...

Introduction

- * Le Big Data c'est un peu la matière noire de l'univers numérique. C'est une quantité d'information cachées, qu'il faut rendre visible (création de nouveaux marchés, renouvellement des offres, adaptation à l'évolution,...)
- * zettaoctet (10 suivi de 20 zéros).

Introduction

TAILLE EN OCTETS CHAQUE ÉLÉMENT EST MILLE FOIS PLUS VASTE QUE LE PRÉCÉDENT.

UN OCTET = un caractère ou un grain de sable

UN KILOOCTET = une phrase ou quelques pincées de sable

UN MÉGAOCTET = une présentation PowerPoint regroupant 20 diapositives, un petit livre ou une cuillerée à soupe de sable

UN GIGAOCTET = neuf mètres de livres sur une étagère ou une boîte à chaussures remplie de sable

UN TÉRAOCTET = 300 heures de vidéo de bonne qualité, un dixième de la bibliothèque du Congrès ou un grand bac à sable

UN PÉTAOCTET = 350 000 photos numériques ou plus de 1,5 km d'étendue de plage

UN EXAOCTET = la moitié des informations générées dans le monde en 1999 ou une plage qui irait de Nice à Gibraltar

UN ZETTAOCTET = l'inimaginable ou une plage aussi vaste que tous les littoraux du monde entier

Introduction

- * Collecte = nécessité d'avoir des émetteurs connectés
- * Potentiel d'adressage élevé (ipV6)
- * Web des objets (M2M) : autant d'adresses que de grains de sable ou d'étoiles dans le ciel

Introduction

- * Big Data engendre :
 - * Infrastructure avec des émetteurs (machines, êtres vivants)
 - * Stockage de l'information
 - * Analyse de l'information : Big Analytics
- * Les temps de latence doivent être courts.

Introduction

- * Gartner Group (2001) analyse le Big Data sous 3 axes :
 - * Volume
 - * Variété
 - * Vélocité
- * A la suite des 3V, on ajoute aujourd'hui 2V
- * Véracité, Valeur

1. Volume

- * Volume de données stockées dans le monde :

- * 1,2 zettaoctets en 2010

- * 1,8 zettaoctets en 2011

- * 2,8 zettaoctets en 2012

- * 40 zettaoctets en 2020

1. Volume

2015

6 A NEW CONNECTED HOME

TRAFFIC GENERATED
by
20 homes in 2015

=

TOTAL TRAFFIC OF
Internet in 1995



2. La variété



2. La variété

- * Est-il bon de mélanger des données de types différents pour en tirer des conclusions ou des tendances ?
- * Travailler sur des données non structurées ?

3. Vélocité

- * Vélocité = vitesse
- * Vitesse de collecte
- * Vitesse d'analyse
- * Réduction des temps de latence
- * Travail en temps réel

4. Véracité

- * Véracité : variabilité = qualité de la donnée
- * Big Data : exigence d'un niveau de qualité élevé

5. Valeur

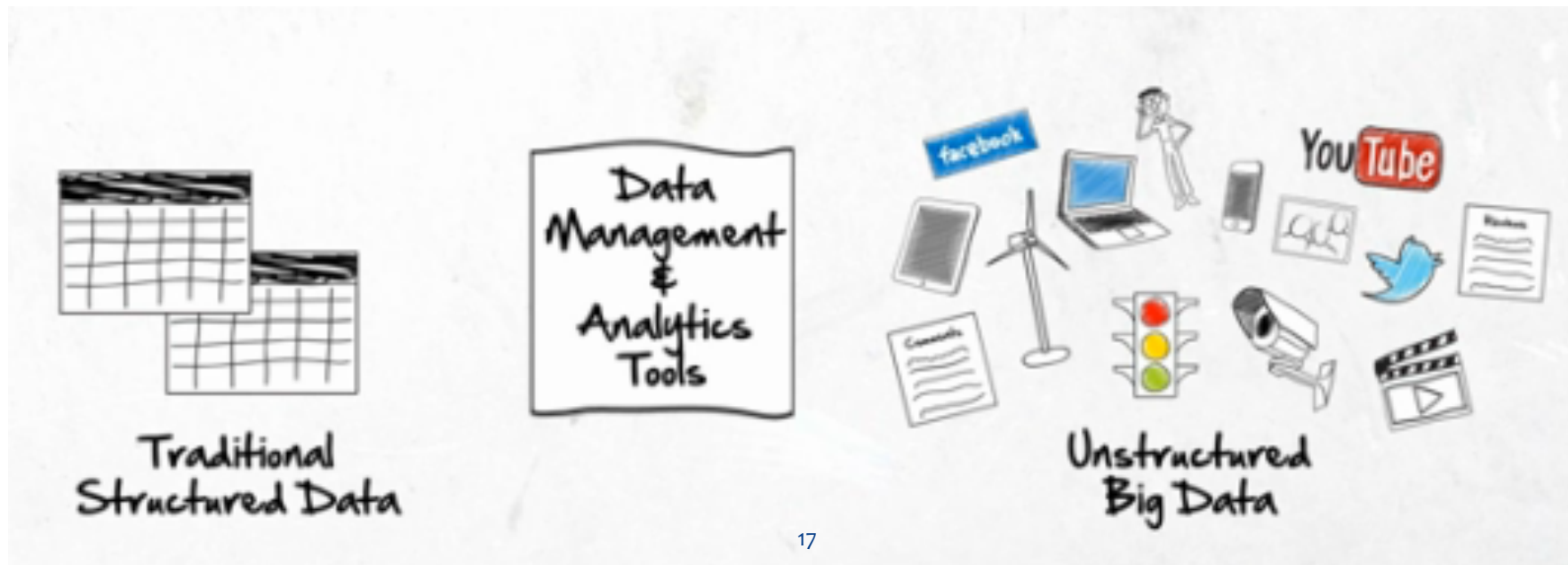
- * Valeur : Quelle question doit-on poser aux données ?
- * La question traitée met en valeur les données.

6. Business Intelligence

- * Big Data :
 - * données de faible densité : volume
 - * Nouveau concept (2010)
- * Business Intelligence :
 - * données de forte densité : peu de valeurs, moins de traitements
 - * Concept ancien (<2000)
- * Outils de calculs, traitements statistiques, agrégation : différents

7. Le modèle de stockage

- * Big Data : modèle relationnel en péril
- * noSQL : structuration, volume



8. Le stockage

- * Cloud Computing : nuagisation
- * Les entreprises doivent faire le choix du nuage
- * Capteurs : vecteurs de stockage (miniaturisation)
- * Acheminement de l'information (infrastructures)

9. Les applications

- * Criminalité : arrêter plus facilement des suspects
- * E-commerce, business : Kraggle, Captain Dash, Beampulse
- * Le divertissement (Entertainment) : 3D
- * Analyse de consommation énergétique : smart grid, linky, ...
- * La santé : quantified self, Google Flu
- * Marchés boursiers : Hyper-Trading
- * Emplois : Viadéo, LinkedIn, MOOCs
- * Education : MOOCs
- * Le flux « stream » de la vie : de la naissance à la mort
- * Le tourisme ...

9. Les applications

- * Gestion bancaire : requêtage illimité (**AMF, ACP, CNIL**)
- * Analyse du sentiment : retour d'expérience
- * Analyse du comportement : comment vendre plus
- * Politique : comment gagner une élection
- * Transhumanisme :
 - * plus d'individus virtuels que réels dans les consoles de jeu.
 - * Humain augmenté
- * Tout cela est déjà bien concret.

9. Les applications

- * Big Data : accélération numérique
- * Il faut également tenir compte des technologies disruptives (Google Glass).
- * Nécessité d'agréger des quantités énormes d'information (adaptation)
- * Nécessité de porter des artifices des extensions afin d'agréger plus rapidement les données (prolongements numériques)

9. Les applications

- * Téléchargement de l'esprit : projet Blue Brain d'IBM
- * Recherche du Graal : vie éternelle = vie numérique
- * « Des individus sur disques durs »
- * Reproduction du fonctionnement synaptique, interconnexion par milliard, simulation en mémoire d'une mémoire ... (la matrice n'est pas loin ... 2040)

10. Les métiers

- * Nouveaux métiers :
 - * Data Journalism
 - * Data Scientist (CDO)
 - * Community Manager
 - * FabLabs : impression 3D

11. Nouvelles tendances

- * Nouvelles tendances de consommation : événements majeurs de la vie
- * Nouveaux risques
- * E-réputation
- * Identités numériques multiples