Romain GÉRARD

Journée des Anciens
10 décembre 2016

Cursus

- DUT Informatique de Bayonne :) (2009-2011)
- INSA de Lyon 2011-2014)
- KAIST – Corée du Sud (2013)

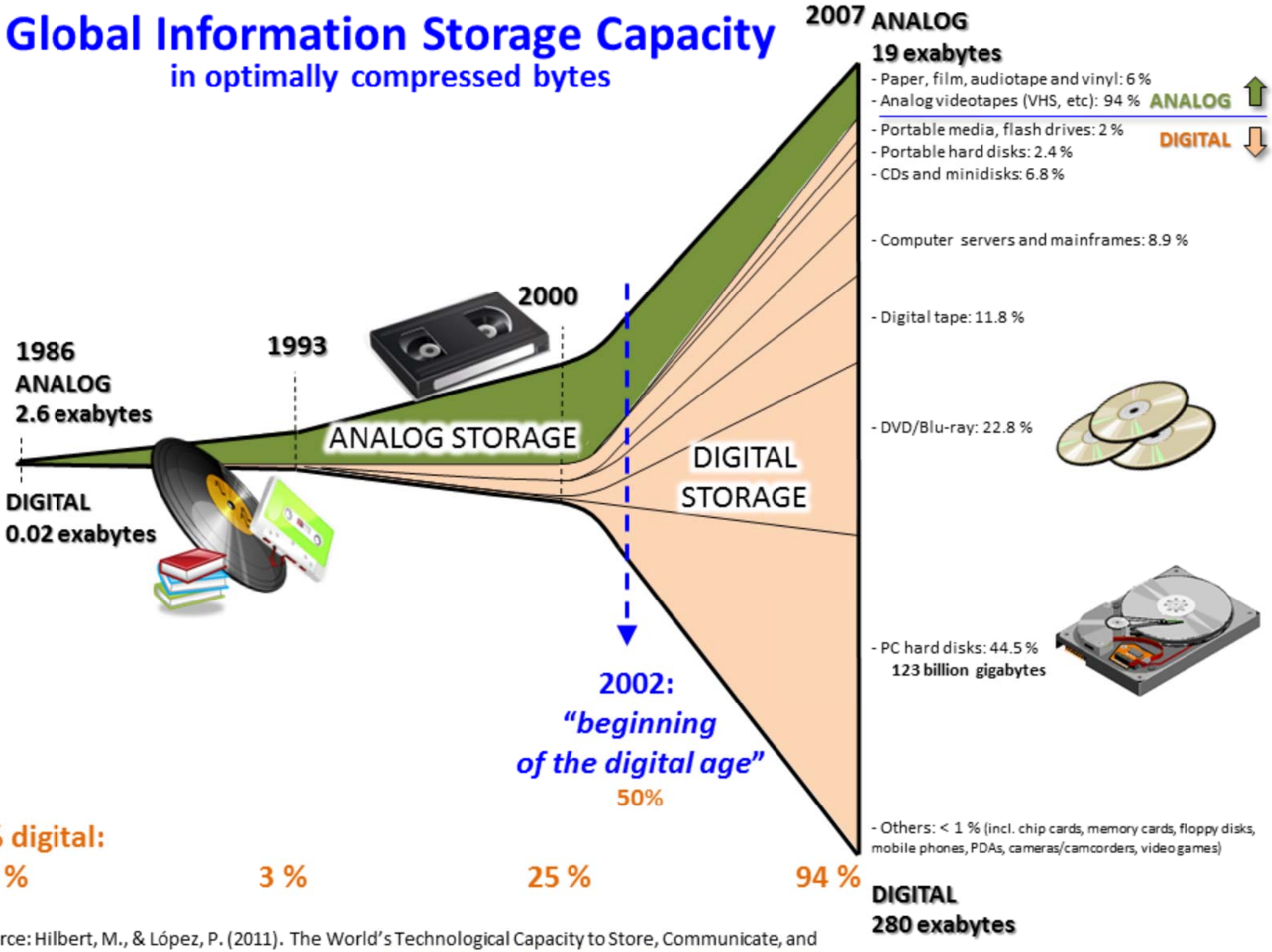
- Amesys puis racheté par ATOS (2014-2016)
  - Wyplay → Ingenieur Middleware (C++/Python)
  - Orange → Ingenieur Big-Data (Java8, Hadoop)
- Criteo (Novembre 2016 - …)
  - Ingénieur Développement et Opération (DevOps)

Sommaire

- Enjeux
- Compétences
- Outils
- Questions

Global Information Storage Capacity

in optimally compressed bytes



Source: Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 60 –65. http://www.martinhilbert.net/WorldInfoCapacity.html

Enjeux (2/3)

TOUS LES 2 ANS

On crée autant d'informations que depuis la nuit des temps à 2003

CHAQUE SECONDE

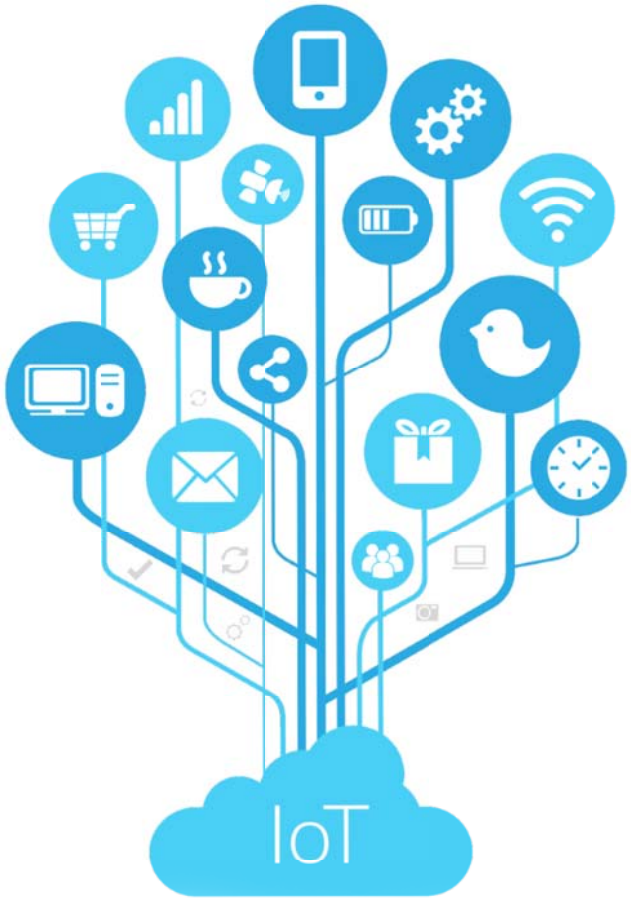
En 2020, chaque humain va générer 1.7 megaoctets d'informations

Maitriser le changement

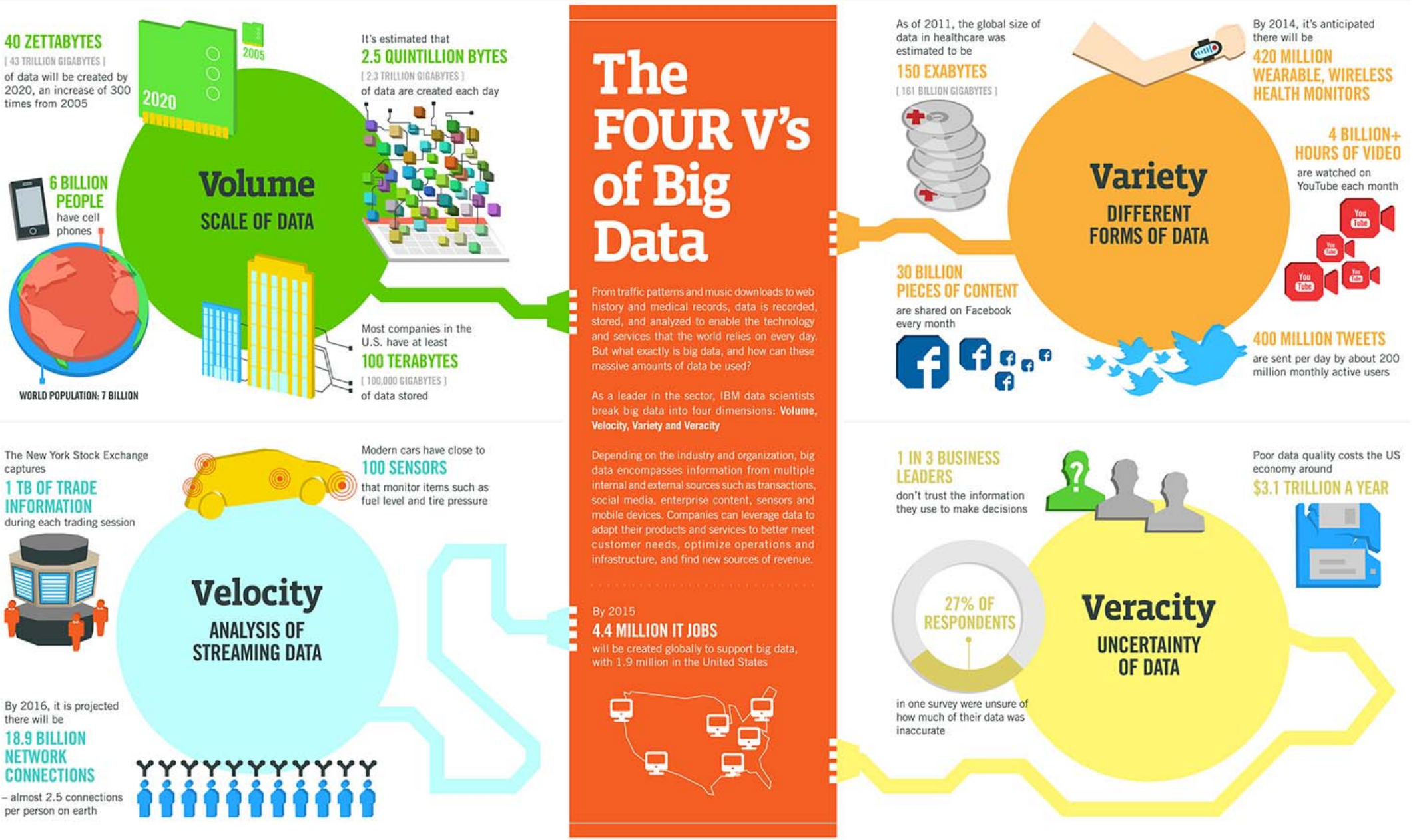
Comment se donner les moyens de traiter et de réagir à cette masse d'informations ?

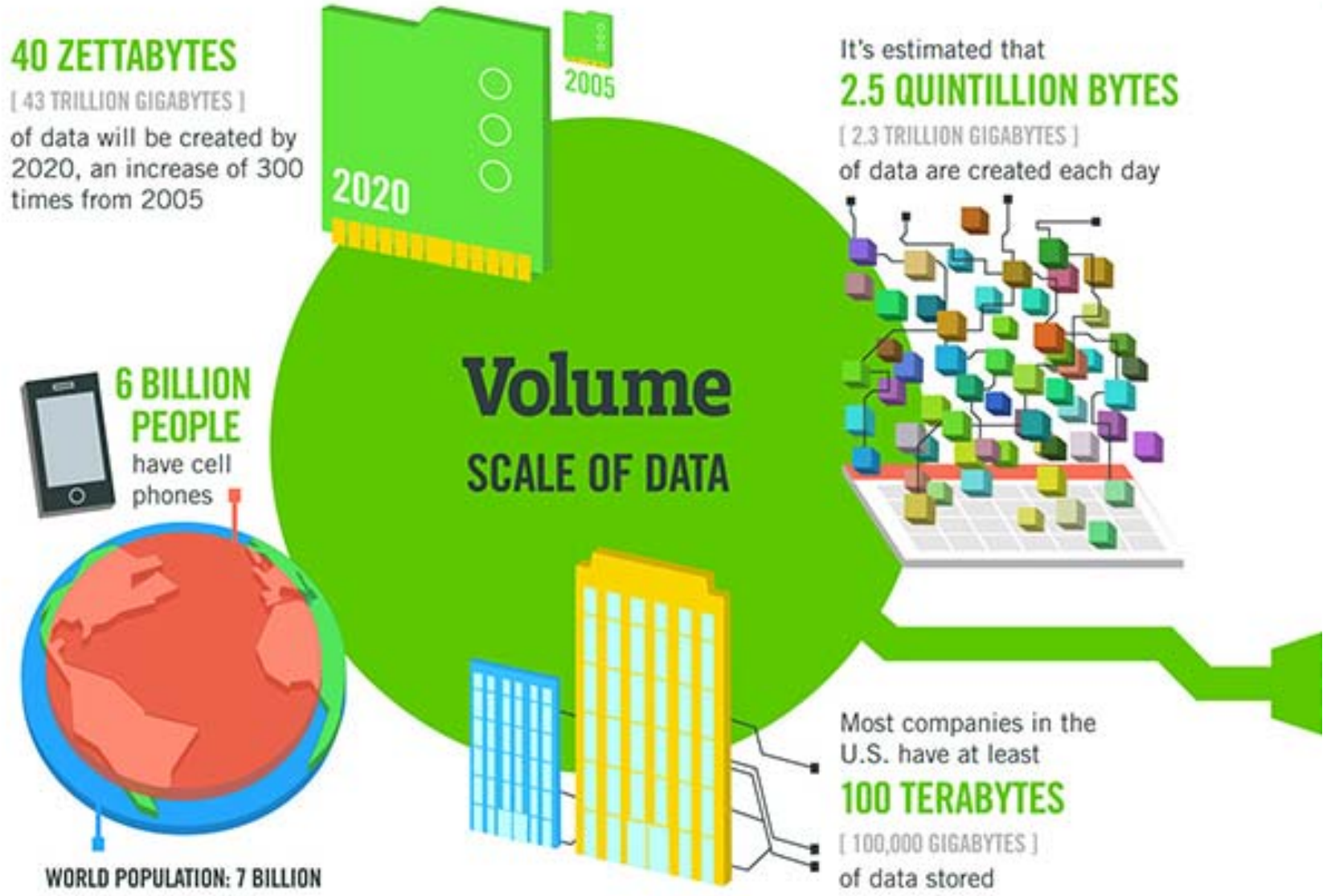
Enjeux (3/3)

- Encore plus de croissance à prévoir grâce aux
  - Objets connectés
    - Puces RFID pour suivre les objets
    - Capteurs en tous genres
  - Médecine
    - Bilan de santé
    - Analyses biologiques/ADN
  - Auto alimentation
    - + on automatise == + de données encore



Compétences




40 ZETTABYTES
[43 TRILLION GIGABYTES]
of data will be created by 2020, an increase of 300 times from 2005
2020
2005
It's estimated that
2.5 QUINTILLION BYTES
[2.3 TRILLION GIGABYTES]
of data are created each day
Volume
SCALE OF DATA
6 BILLION PEOPLE
have cell phones
WORLD POPULATION: 7 BILLION
Most companies in the U.S. have at least
100 TERABYTES
[100,000 GIGABYTES]
of data stored

The New York Stock Exchange captures

1 TB OF TRADE INFORMATION

during each trading session

Modern cars have close to

100 SENSORS

that monitor items such as fuel level and tire pressure

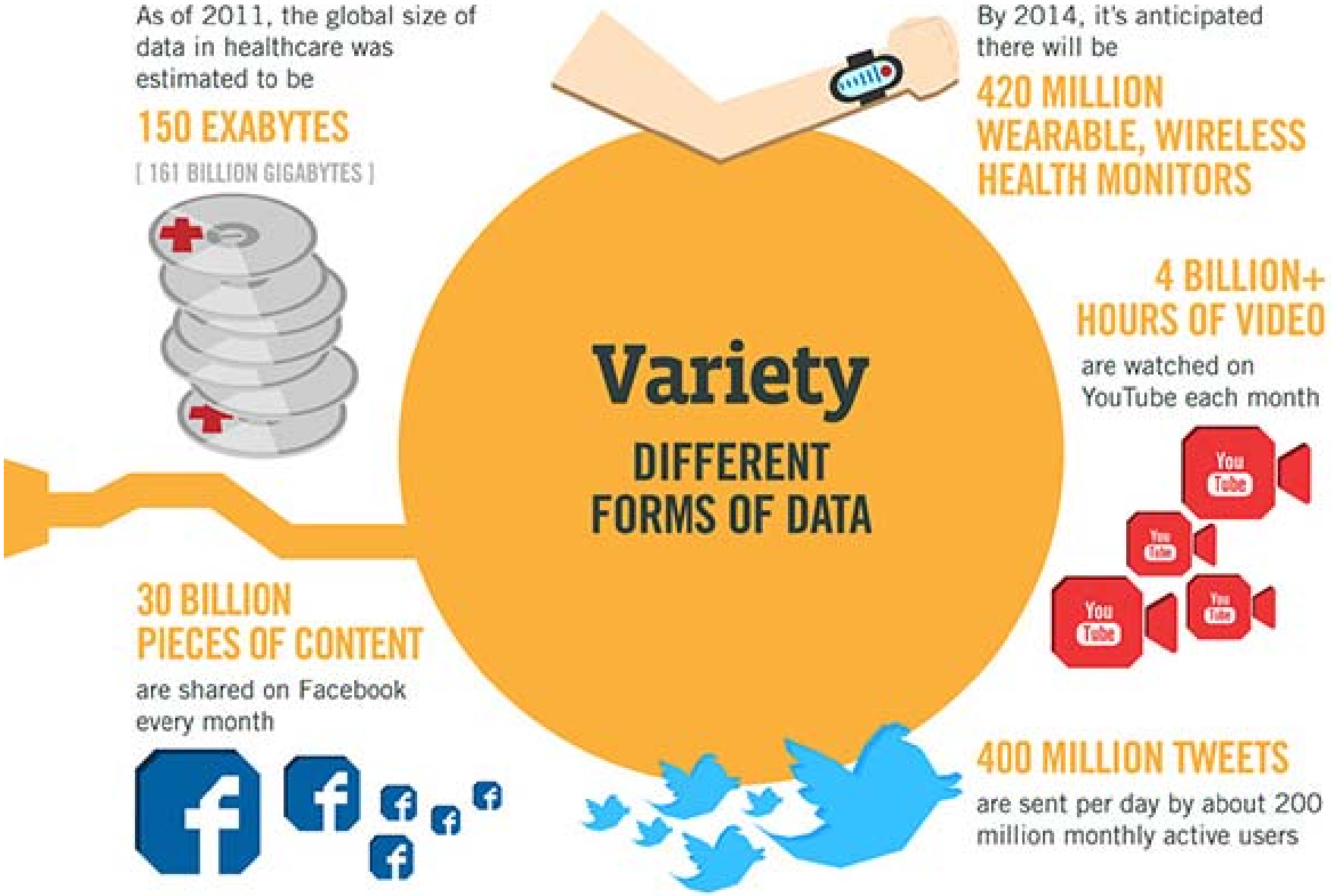
Velocity

ANALYSIS OF STREAMING DATA

By 2016, it is projected there will be

18.9 BILLION NETWORK CONNECTIONS

– almost 2.5 connections per person on earth


As of 2011, the global size of data in healthcare was estimated to be
150 EXABYTES
[161 BILLION GIGABYTES]
By 2014, it's anticipated there will be
420 MILLION WEARABLE, WIRELESS HEALTH MONITORS
Variety
DIFFERENT FORMS OF DATA
4 BILLION+ HOURS OF VIDEO
are watched on YouTube each month
30 BILLION PIECES OF CONTENT
are shared on Facebook every month
400 MILLION TWEETS
are sent per day by about 200 million monthly active users

Veracity

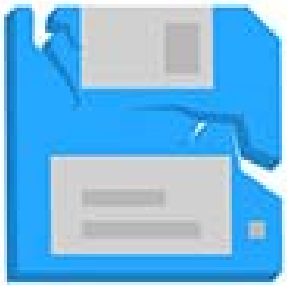
UNCERTAINTY OF DATA

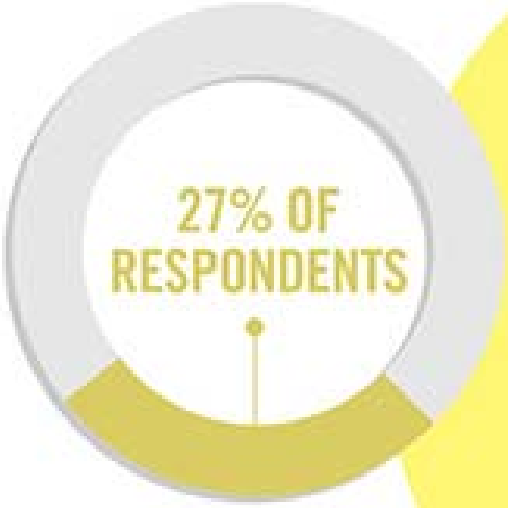
1 IN 3 BUSINESS LEADERS

don't trust the information they use to make decisions

Poor data quality costs the US economy around

$3.1 TRILLION A YEAR





in one survey were unsure of how much of their data was inaccurate

Outils (1/2)

- Système de fichiers distribués
  - 

- Système de calculs distribués
  - 

- Système de traitements évènementiels
  - 

Outils (2/2)

- Système de croisement des données
  - Spark SQL 

- Visualisation
  - D3.js, Tableau, Apache Zeppelin


Questions ?
BIG DATA
VOLUME
DATA SIZE
VELOCITY
SPEED OF CHANGE
VARIETY
DIFFERENT FORMS OF DATA SOURCES
VERACITY
UNCERTAINTY OF DATA