

# Modélisation de données

## Introduction

AAA\_01a  
v215e

2022-01-10



Christina.Khnaisser@USherbrooke.ca  
Luc.Lavoie@USherbrooke.ca

© 2018-2021, Μητίς (<http://info.usherbrooke.ca/llavoie>)  
CC BY-NC-SA 4.0 (<https://creativecommons.org/licenses/by-nc-sa/4.0/>)

# Plan

- Contexte
- Modèles
- Futurologie
- Vocabulaire
- Références
- Les colles du prof !
- Bibliographie



# Contexte

- Les domaines d'application
- Une caractérisation des problèmes
  - 3 V traditionnels
  - 4 V : la vision d'IBM en 2013
  - 5 V : la vision d'IBM en 2015
  - 7 V contemporains
- Le complot des scientifiques
- Le complot des gestionnaires

## Contexte

### Les domaines d'application

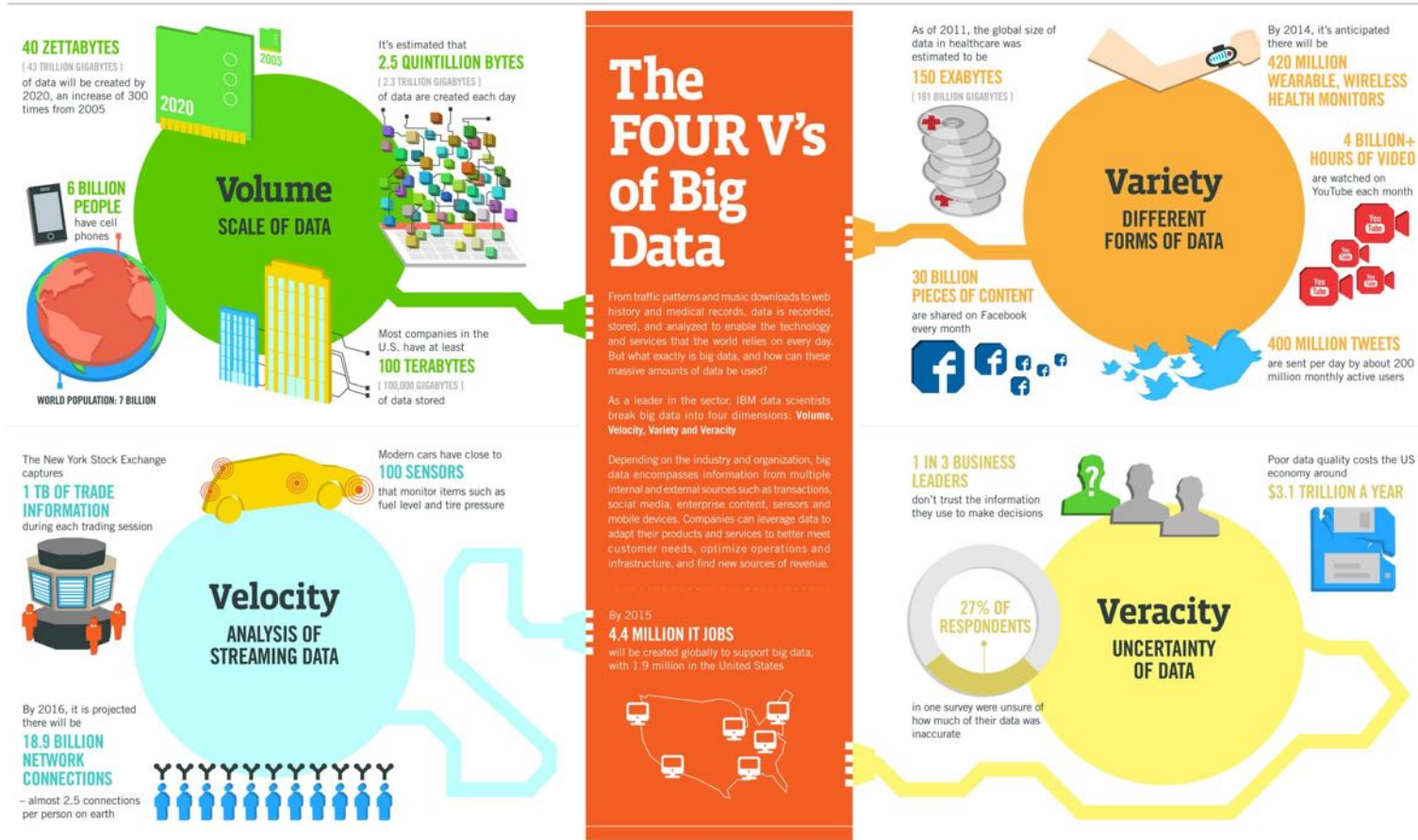
- Besoin de traiter, de conserver et d'analyser de (très) grandes quantités d'informations
  - gouvernements (recensement, impôts, santé...)
  - banques et assurances
  - recherche scientifique (astronomie, chimie, génétique...)
  - télécommunications
  - secteur industriel et manufacturier
  - secteur énergétique
  - grande distribution
  - géomatique
  - agences de renseignement, de propagande et de marketing
  - ...

## Contexte

### Une caractérisation des problèmes (3 V)

- Comment caractériser les problèmes afin de déterminer les solutions les plus adéquates ?
- Grâce au spectre traditionnel des **3 V** !
  - **volume**
  - **variété**
  - **vélocité**

# Une illustration des « V » par IBM en 2013

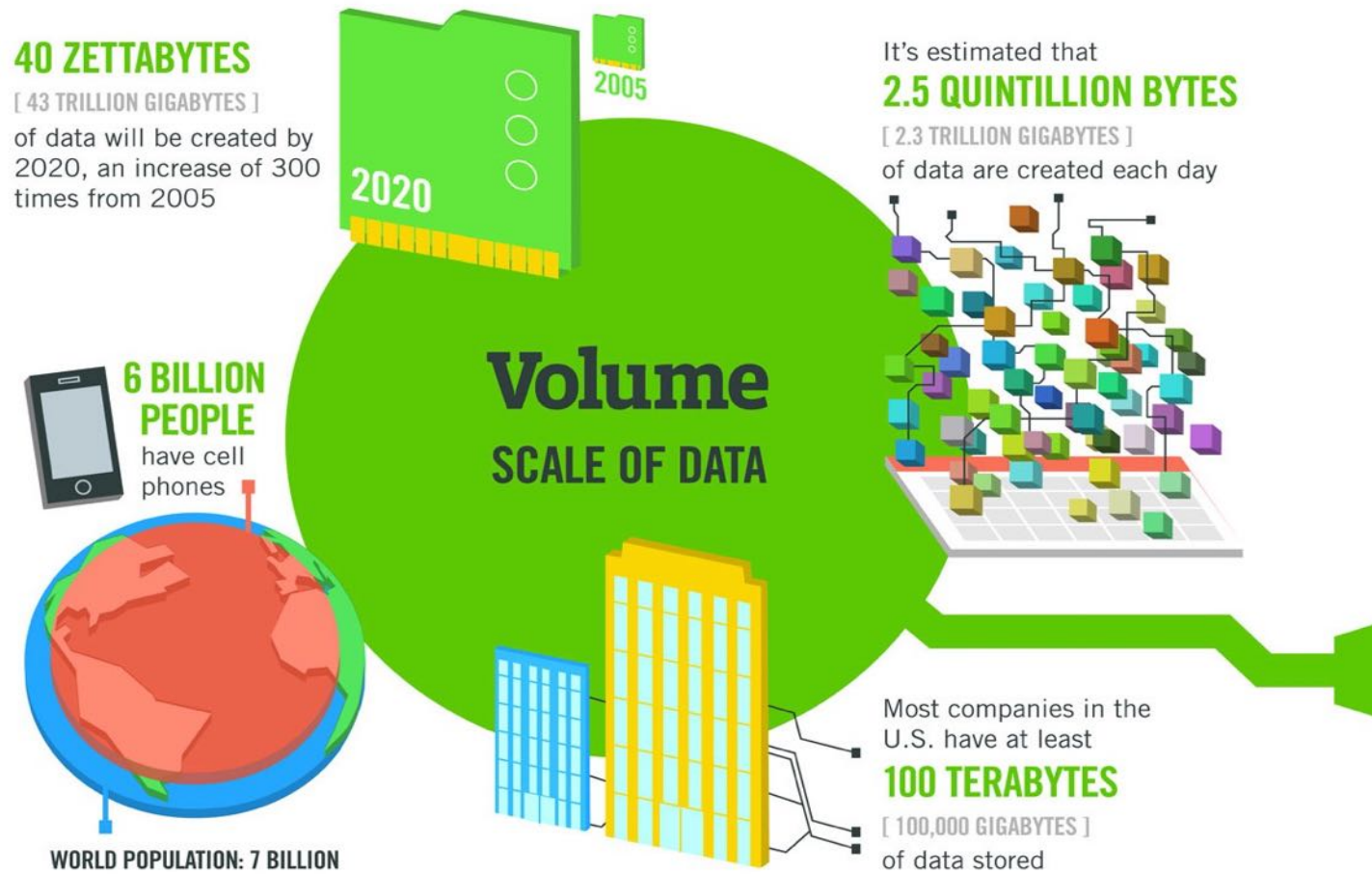


Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTec, QAS



source : <http://www-01.ibm.com/software/data/bigdata/images/4-Vs-of-big-data.jpg>

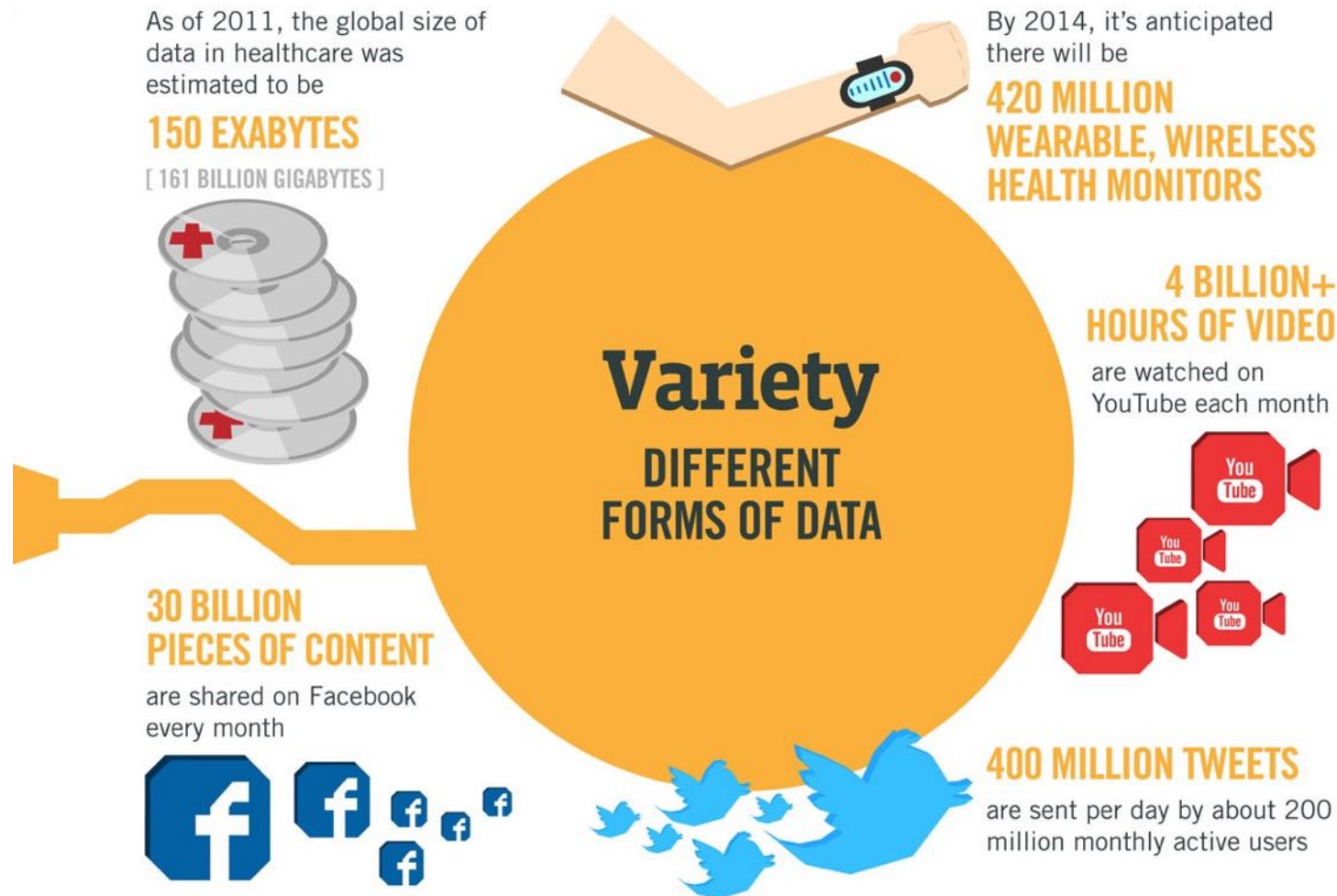
## IBM - Volume : Pour l'année de référence 2013, vraiment ?



source : <http://www-01.ibm.com/software/data/bigdata/images/4-Vs-of-big-data.jpg>



## IBM – Variété : Soyez vigilants et critiques!



source : <http://www-01.ibm.com/software/data/bigdata/images/4-Vs-of-big-data.jpg>



## IBM – Vitesse : Soyez vigilants et critiques!

The New York Stock Exchange captures

**1 TB OF TRADE INFORMATION**

during each trading session



By 2016, it is projected there will be

**18.9 BILLION NETWORK CONNECTIONS**

– almost 2.5 connections per person on earth



**Velocity**  
ANALYSIS OF  
STREAMING DATA



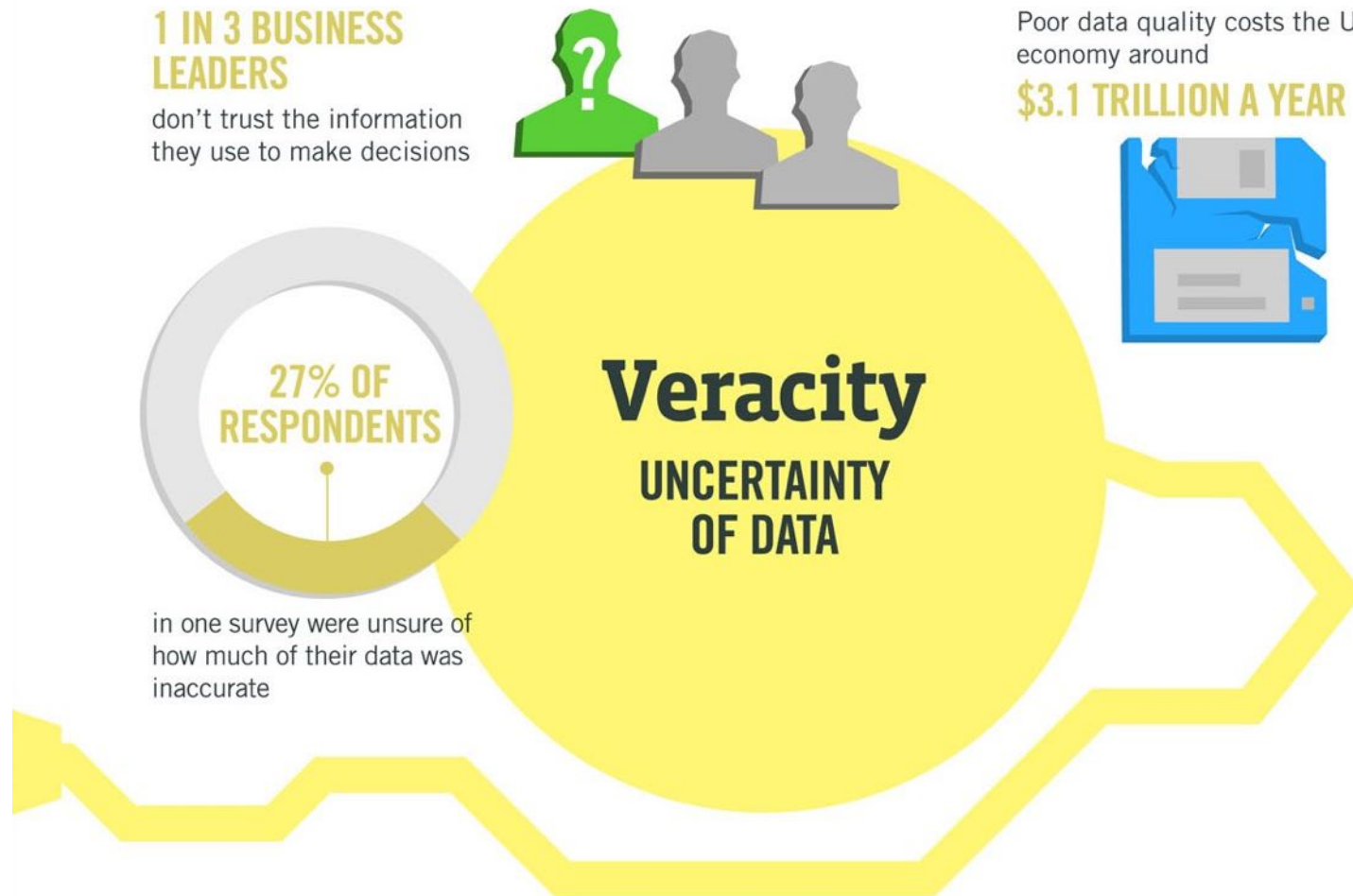
Modern cars have close to

**100 SENSORS**

that monitor items such as fuel level and tire pressure

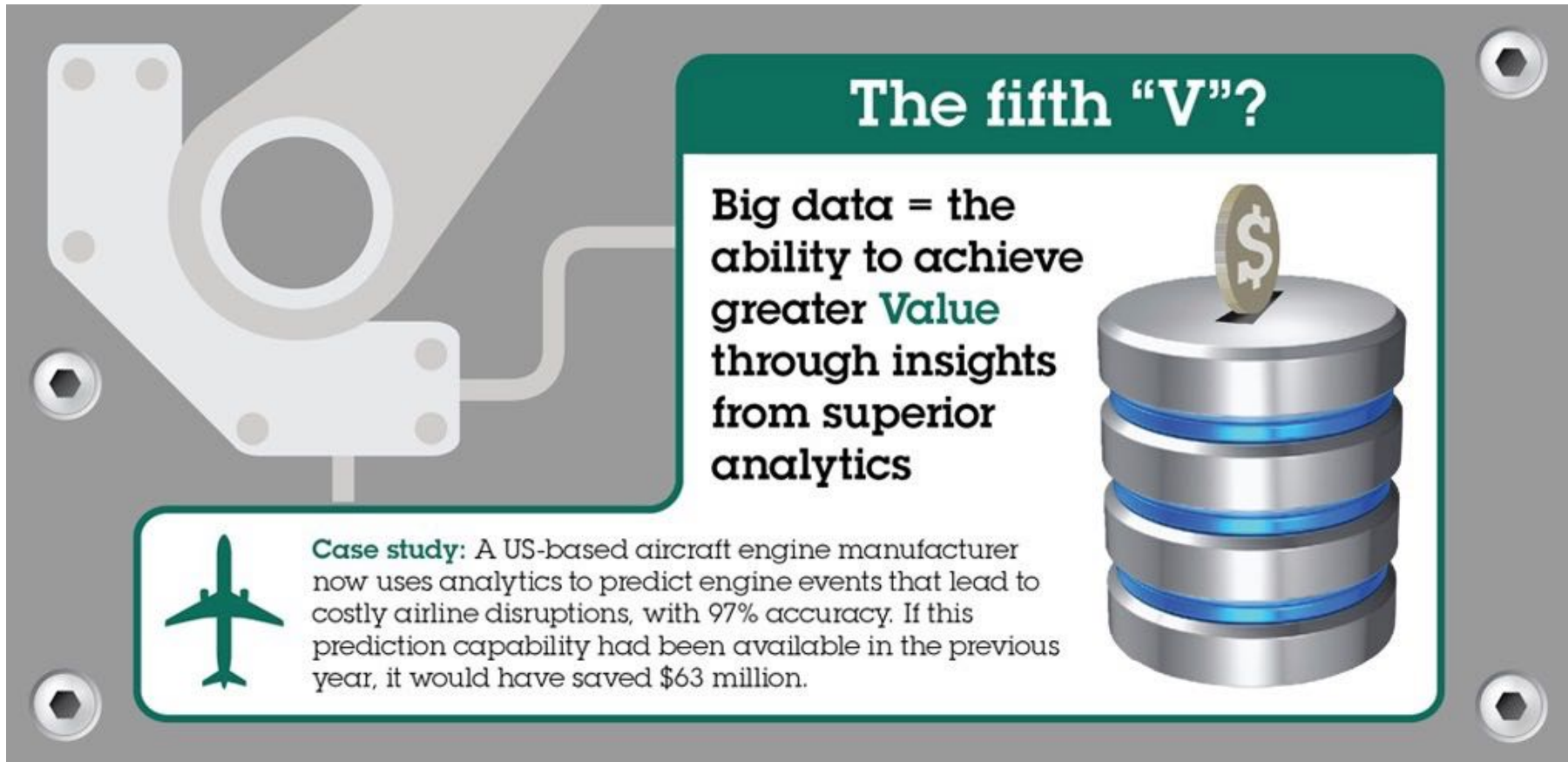
source : <http://www-01.ibm.com/software/data/bigdata/images/4-Vs-of-big-data.jpg>

## IBM – Veracity : Soyez vigilants et critiques !



source : <http://www-01.ibm.com/software/data/bigdata/images/4-Vs-of-big-data.jpg>

## IBM – Valeur : les promesses de 2015...



**The fifth “V”?**

Big data = the ability to achieve greater **Value** through insights from superior analytics

**Case study:** A US-based aircraft engine manufacturer now uses analytics to predict engine events that lead to costly airline disruptions, with 97% accuracy. If this prediction capability had been available in the previous year, it would have saved \$63 million.

The infographic features a green header with the title 'The fifth "V"?' and a white box containing the main text. To the right is a 3D illustration of a stack of five silver data disks with blue horizontal bands, topped with a gold dollar coin. To the left is a green silhouette of an airplane. The background is a grey metallic surface with a large white circular component and several screws.

source :

<http://www.ibmbigdatahub.com/infographic/extracting-business-value-4-vs-big-data>

## Contexte

### Une caractérisation des problèmes (7 V)

- Comment caractériser les problèmes afin de déterminer les solutions les plus adéquates selon le contexte ?
  - > variabilité
- qualité et potentiel
  - > valeur
- emplacement
  - > virtualité
- Comment caractériser les problèmes en regard exigences découlant des lois et règlements quant à la protection et l'utilisation éthique des données ?
  - > vertu

## Contexte

### Une caractérisation des problèmes (7 V)

- **volume**  
(quantité de données devant être stockées)
- **variété**  
(diversité et complexité des types utilisés par le modèle)
- **vélocité**  
(caractérisation du débit de données entrantes et sortantes)
- **variabilité**  
(temporalité, incertitude et perspective des données)
- **valeur**  
(richesse analytique des données)
- **virtualité**  
(distribution des sources de données)
- **vertu**  
(gouvernance des données)

## Contexte

### Tendances

- Intégrer différents types de données
  - Structurées
  - Semi-structurées
  - Non structurées
- Intégrer différentes dimensions
  - Axes temporels
  - Agents
- Intégrer différentes sources de données
  - Objets connectés
  - Médias sociaux

## Contexte

### Le complot des scientifiques

- Une volonté de comprendre « la réalité », puis
  - de documenter cette compréhension,
  - de la transmettre,
  - de l'opérationnaliser,
  - de l'appliquer.
- Une approche qui repose
  - sur le raisonnement axiomatique,
  - afin (notamment) de garantir la réfutabilité,
  - en se fondant sur la logique (du premier ordre)
- Cette approche a donné naissance à la **théorie relationnelle** (de modélisation des données).



## Contexte

### Le complot des gestionnaires

- *Cette approche s'accorde très souvent avec la mission des organisations en ce sens qu'elle constitue un outil puissant pour en gérer optimalement les ressources.*
- *Le méta-complot des gestionnaires :*
  - *Tirer parti des avancées de la théorie relationnelle pour mieux gérer.*

# Modèles

- Principes CMcm
- Trischématisation
- SGBDR
- Modèles et modèles de modèles

## Modèles

### Principe CMcm

- **Cohérence**  
rien ne se contredit.
- **Cohésion**  
tout est lié (tout se tient).
- **Couplage**  
liaison entre deux éléments; en particulier, une liaison peut être une dépendance.
- *Si tous les éléments d'un ensemble sont couplés (transitivement) alors l'ensemble est (totalement) cohésif.*
- Il est (plus) facile de constater l'incohérence dans un ensemble cohésif.
- *Mais la cohésion rend très difficile la modification (l'évolution).*
- On en déduit un principe :  
**Cohérence maximale pour un couplage minimal.**

## Modèles

### Trischématisation, une solution ?

- L'application du principe CMcm au problème de l'intégrité a conduit à l'élaboration du modèle *trischématique*.
- Ce modèle a également permis de répondre à plusieurs préoccupations d'efficacité et d'efficience.
- D'où sa longévité.

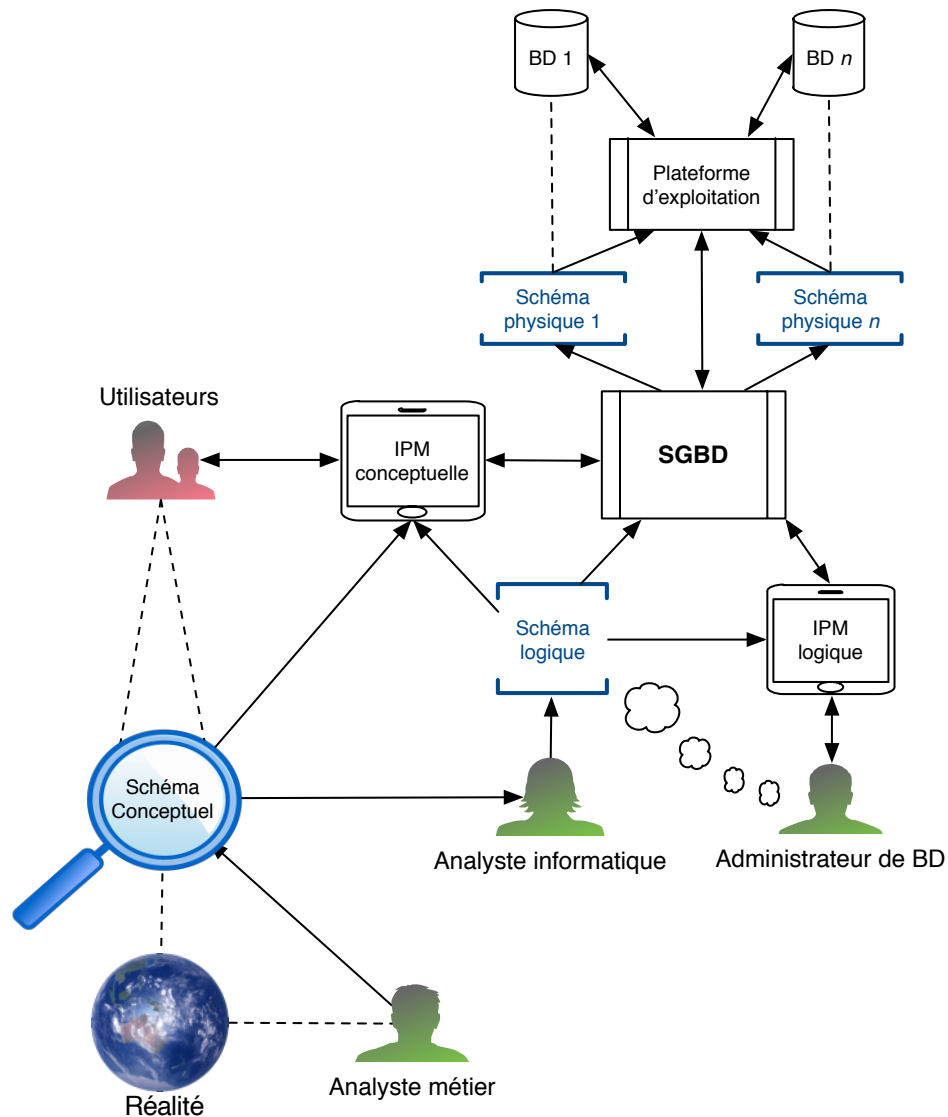


# Trischématisation

## Un exemple, le SGBD

### Schémas :

- **conceptuel**
- **logique**
- **physique**



## Théorie et modèles

### ○ conceptuel

- EA (entité-association)  
ER (*entity-relationship*)
- EAE (EA étendu)  
ERE (*extended ER*)
- Merise
- Ontologique
- ...

### ○ logique

- hiérarchique
- réseau
- relationnel
- objet
- ...

### ○ physique

- adressage dispersé  
(*hashing*),  
arbre de recherche (*B-Tree*)  
...
- horizontal,  
vertical,  
mixte,  
KV,  
RDF  
...

## Trois niveaux sont-ils suffisants ?

- Beaucoup de chercheurs estiment aujourd'hui qu'un quatrième niveau est nécessaire : celui des **modèles de connaissances**.
- D'autres estiment que ce niveau ne fait qu'un avec celui des modèles conceptuels.



## Dans ce cours (1)

- Nous recherchons :
  - une fondation solide à la modélisation de données
  - une grande souplesse de modélisation
  - les moyens de dériver automatiquement une « machine » efficace et efficiente à partir du modèle de données (sans avoir à élaborer le modèle physique)

## Dans ce cours (2)

- Nous nous intéresserons à **une** théorie des schémas logiques :
  - la **théorie relationnelle**
  - et à certains de ses modèles.
  
- Nous nous intéresserons à **une** théorie des schémas conceptuels :
  - la **théorie entité-association**
  - et à certains de ses modèles.

## Modèles et schémas

- Un modèle de données (dérivé d'un modèle d'information s'appuyant lui-même sur un modèle de connaissances, chacun de ces modèles étant décrit à l'aide d'une notation associée à un modèle de modélisation approprié) sera désormais désigné sous le nom de schéma...
- En fait, le mot modèle est
  - surutilisé
  - dans des contextes différents
  - pour décrire des choses différentes
  - ... il faut donc toujours en préciser le sens
- Voilà qui est fait!

# Futurologie

- Un exemple
- De l'esprit critique
- Exercices

## Futurologie

### Un exemple

#### ○ Préambule

- De la difficulté de la prospective
- De la nécessité de l'analyse critique

#### ○ Un exemple

- En 2013, IBM avance un taux de pénétration des appareils mobiles à raison de 6 milliards de personnes sur une population de 7 milliards (voir l'illustration du critère de volume proposée précédemment)
  - a priori, cela est peu vraisemblable (25 % de la population mondiale à moins de 15 ans et au moins 10 % vit dans un territoire non couvert par un réseau téléphonique sans fil)
  - ...

# LA FORTE CROISSANCE DU LARGE BANDE MOBILE SE POURSUIT

Plus de 2 milliards d'abonnements dans le monde d'ici à fin 2013\*

## ■ Amériques

460 millions d'abonnements

Taux de pénétration de 48%

TCAC de 28% (2010-2013)

## ■ Europe

422 millions d'abonnements

Taux de pénétration de 68%

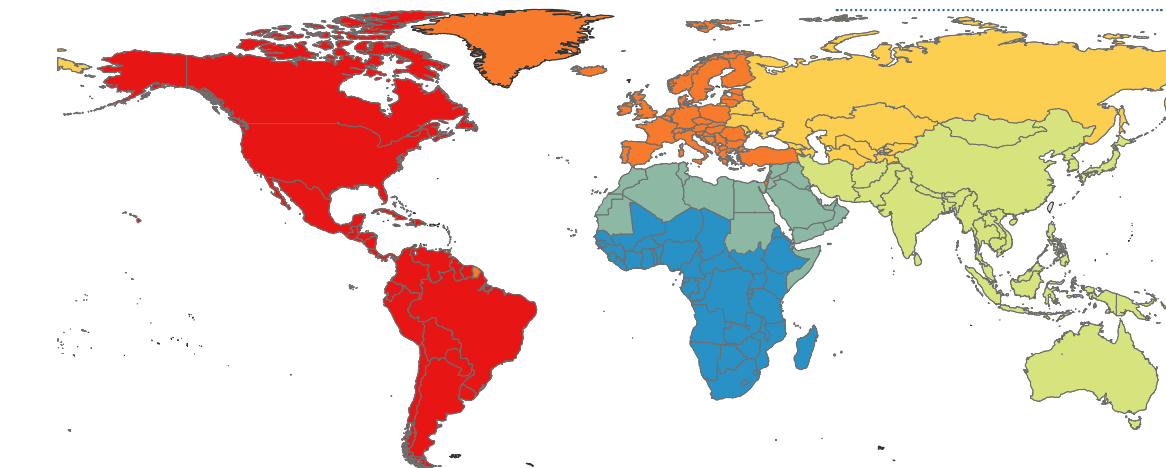
TCAC de 33% (2010-2013)

## ■ CEI

129 millions d'abonnements

Taux de pénétration de 46%

TCAC de 27% (2010-2013)



## ■ Etats arabes

71 millions d'abonnements

Taux de pénétration de 19%

TCAC de 55% (2010-2013)

## ■ Afrique

93 millions d'abonnements

Taux de pénétration de 11%

TCAC de 82% (2010-2013)

## ■ Asie-Pacifique

895 millions d'abonnements

Taux de pénétration de 22%

TCAC de 45% (2010-2013)

Source: Base de données de l'UIT sur les indicateurs des télécommunications/TIC dans le monde

Note: \* Estimation

<https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2013-f.pdf>

## Futurologie esprit critique, téléphone et mobilité

- Une étude plus sérieuse estime à 3,4 milliards de personnes le nombre de possesseurs d'un appareil mobile fonctionnel et connecté en 2013, soit moins de 50 % de la population mondiale. Quant au nombre d'appareils reliés effectivement à Internet, il était inférieur à 2 milliards... et plusieurs de ceux-ci étaient embarqués dans des équipements. De plus, les marges d'erreur sont considérables.
- Voir également
  - <http://www.journaldunet.com/ebusiness/internet-mobile/1009553-monde-le-nombre-d-abonnes-au-telephone-mobile/>



## Futurologie esprit critique et population

- Pour des statistiques relatives à la population mondiale, aux pyramides d'âges, etc., voir :
  - [http://www.prb.org/pdf12/2012-population-data-sheet\\_french.pdf](http://www.prb.org/pdf12/2012-population-data-sheet_french.pdf)
  - <https://www.prb.org/wp-content/uploads/2020/07/PRB2020WPDS-BOOKLET-FR.pdf>
  - [https://population.un.org/wpp/Publications/Files/WPP2019\\_DataBooklet.pdf](https://population.un.org/wpp/Publications/Files/WPP2019_DataBooklet.pdf)
- et comparer les méthodologies et les conclusions.

## Futurologie

### Exercice 1

- Soumettez les illustrations de la caractérisation des « V » proposées par IBM à l'analyse critique.

## Futurologie

### Exercice 2 et 3

- Quelles étaient les prétentions des grands acteurs des télécommunications en 2000 (Cisco, Nokia, Motorola, Ericsson, Nortel, etc.) ?
  - Qu'en est-il advenu ? Pourquoi ?
- Quelles sont les prétentions des grands acteurs du Web en 2022 (GAFAM et les autres) ?
  - Qu'en adviendra-t-il ? Pourquoi ?

# Conclusion

- Références
- Les colles du prof
- Quelques autres références...

## Références

- Bases de données (Databases) :
  - [Date 2004], chapitres 1 et 2
  - [Elmasri 2004], chapitres 1 et 2
  - [Elmasri 2016], chapitres 1 et 2
  - [Ullman 2008], chapitre 1
  - [https://fr.wikipedia.org/wiki/Base\\_de\\_données](https://fr.wikipedia.org/wiki/Base_de_données)
  - <https://en.wikipedia.org/wiki/Database>
- Mégadonnées (Big Data) :
  - [https://fr.wikipedia.org/wiki/Big\\_data](https://fr.wikipedia.org/wiki/Big_data)
  - [https://en.wikipedia.org/wiki/Big\\_data](https://en.wikipedia.org/wiki/Big_data)

## Les colles du prof (2)

- À quels besoins fondamentaux répond une base de données?
- Quelles sont ses « qualités » essentielles?
- Quels sont ses fondements théoriques?
- Quelle est l'architecture d'un SGBD?
- Quelle différence y a-t-il entre un SGBD, une BD et un schéma?

## Les colles du prof (2)



[http://en.wikipedia.org/wiki/Alan\\_Turing#mediaviewer/File:Alan\\_Turing\\_photo.jpg](http://en.wikipedia.org/wiki/Alan_Turing#mediaviewer/File:Alan_Turing_photo.jpg)

## Alan Turing (1912-1954)

Qui était Alan Turing ?

Comment est-il mort ?

Quand a-t-on reconnu publiquement sa contribution au décryptage d'Enigma ?

Qu'est-ce que le prix Turing ?



## Les colles du prof (3)



[https://fr.wikipedia.org/wiki/Marian\\_Rejewski](https://fr.wikipedia.org/wiki/Marian_Rejewski)

### Marian Rejewski (1905-1980)

Qui était Marian Rejewski ?

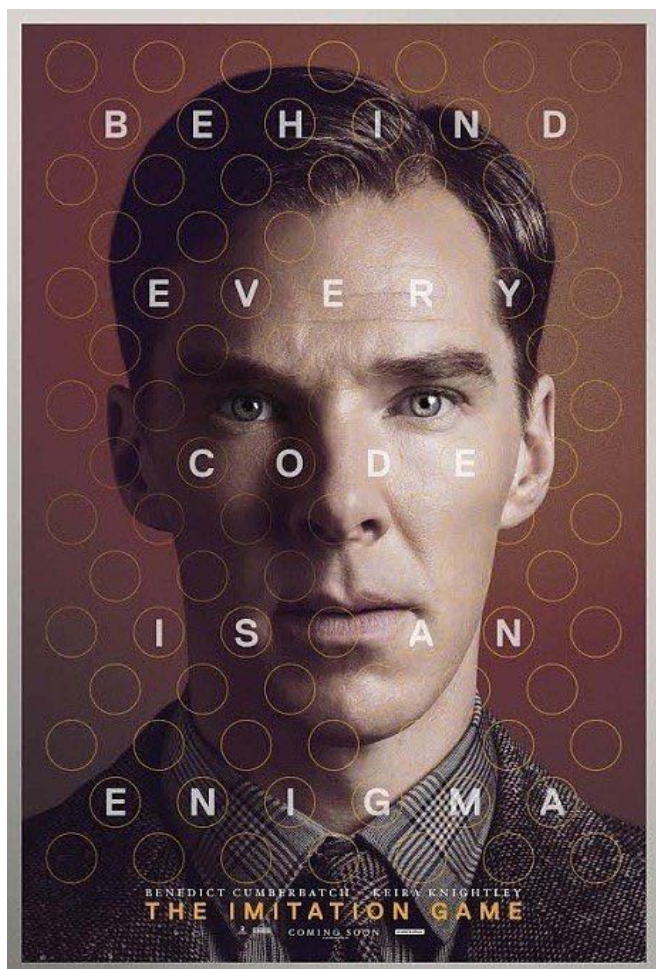
Quel est le lien avec Alan Turing ?

Finalement, qui furent les premiers décodeurs d'Enigma ?

Les deuxièmes ne leur sont-ils pas redevables ?

Il aura fallu attendre jusqu'en 2005 pour que les Anglo-américains reconnaissent la contribution primordiale des cryptologues polonais, dont Rejewski et ses collègues Jerzy Różycki et Henryk Zygalski.

## Imitation Game (2014)



<https://www.imdb.com/title/tt2084970/reference>

Réalisateur :

Morten Tyldum

Scénaristes :

Graham Moore,  
Andrew Hodges

Acteur principal :

Benedict Cumberbatch

Synopsis :

Une version *romancée* d'une partie déterminante de la vie de Turing : le *second* décryptage d'Enigma.

## Sources relatives à Alan Turing

- [https://fr.wikipedia.org/wiki/Alan\\_Turing](https://fr.wikipedia.org/wiki/Alan_Turing) (2022-01-04)
- Jean Lassègue, *Turing*, Paris, Les Belles lettres, 1998. (ISBN 978-2-2517-6014-8)
- Laurent Lemire, *Alan Turing : l'homme qui a croqué la pomme*, Paris, Hachette Littératures, 2004, 191 p. (ISBN 978-2-0123-5618-4)
- Jean Lassègue, *Les Génies de la science*, Pour la Science, n° 29 « Turing... et l'informatique future », nov. 2006 – janv. 2007. (ISBN 978-2-8424-5078-6)
- Andrew Hodges et Douglas Hofstadter, *Alan Turing : The Enigma*, Princeton University Press, 2012, 586 p. (ISBN 978-0-6911-5564-7)

## Sources relatives à Marian Rejewski

- [https://fr.wikipedia.org/wiki/Marian\\_Rejewski](https://fr.wikipedia.org/wiki/Marian_Rejewski)
- [https://pl.wikipedia.org/wiki/Marian\\_Rejewski](https://pl.wikipedia.org/wiki/Marian_Rejewski)
- [https://pl.wikipedia.org/wiki/Bomba\\_kryptologiczna](https://pl.wikipedia.org/wiki/Bomba_kryptologiczna)

## Film

- ***Sekret Enigmy*** (1979) réalisé par Roman Wionczek  
<https://www.imdb.com/title/tt0079878/>

## Série télé

- ***Tajemnica Enigmy*** (1979-1980) réalisé par Roman Wionczek  
<https://www.filmweb.pl/serial/Tajemnica+Enigmy-1979-10013>  
(disponible en CD)

