

maj 2016

# Big Data and Big Data Analytics

*Compréhension du principe*

Par Dr. Samia GAMOURA-CHEHBI

# GREEN SUPPLY CHAIN MANAGEMENT, quoi ?

*On connait le SCM, mais c'est quoi le GSCM ?*

Ensemble de données avec les caractéristiques suivantes :

- Data Sources (Variety – V1)
- big data acquisition (Velocity – V2),
- big data storage (Volume – V3),
- big data analysis (Veracity – V4).



**Variety ?**  
Données hétérogènes

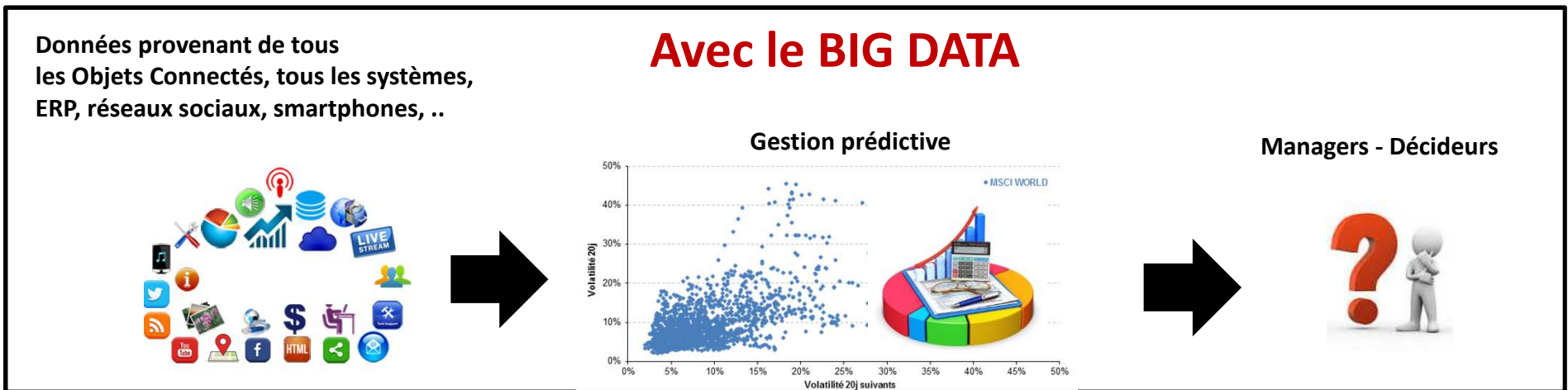
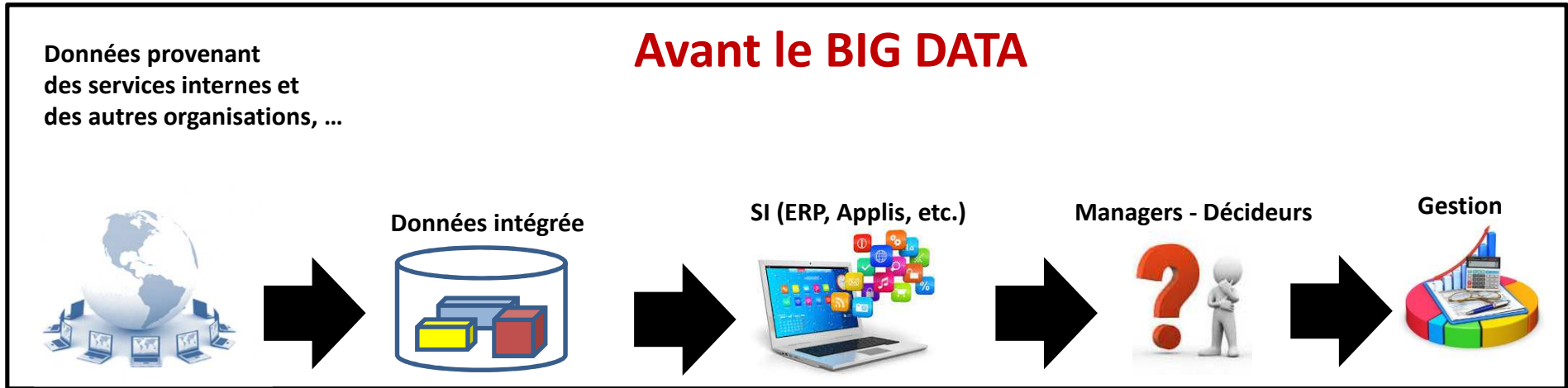
**Velocity ?**  
Vitesse d'acquisition

**Volume ?**  
Stockage démesuré

**Veracity ?**  
Analytique incertaine

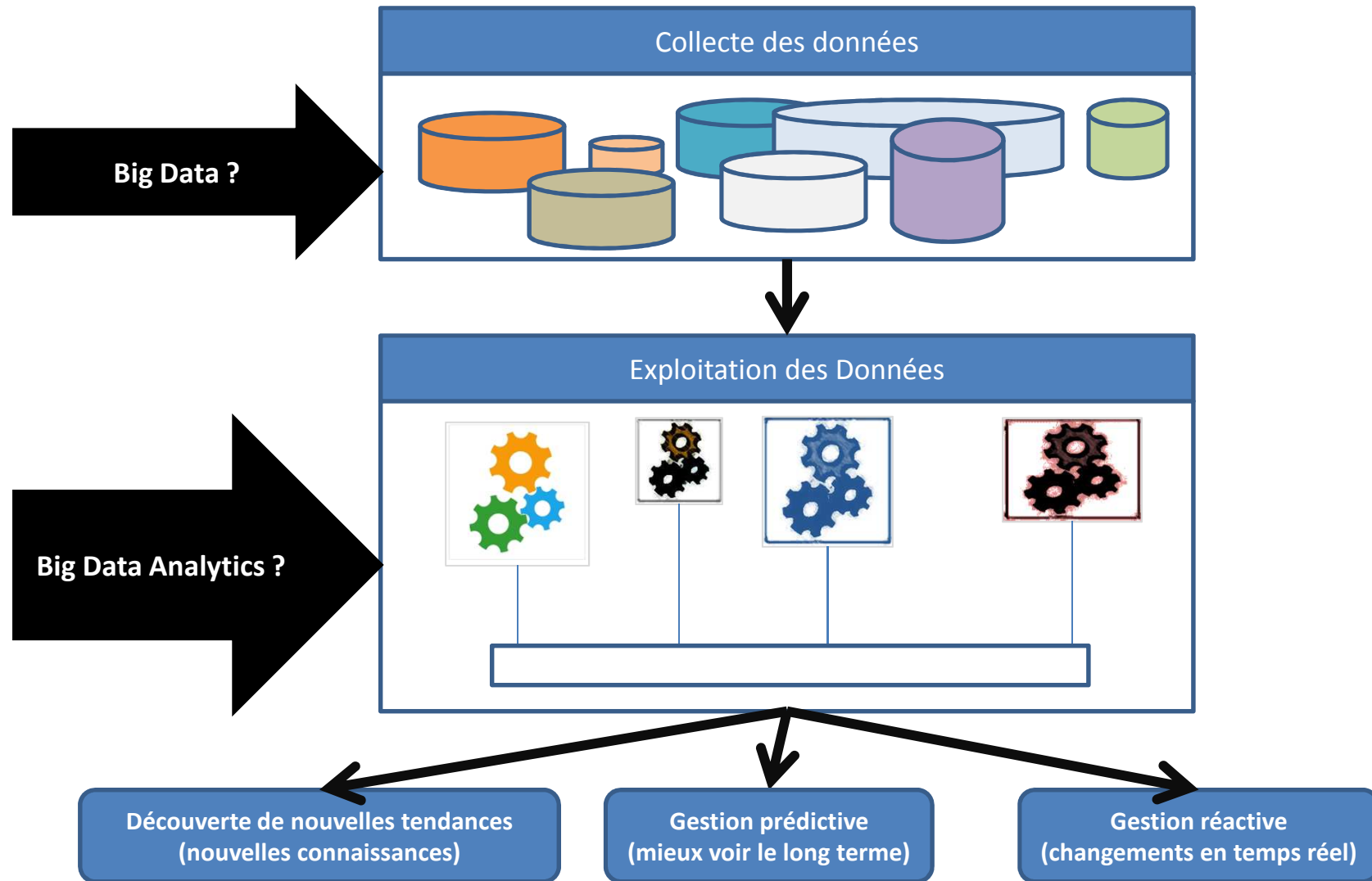
# BIG DATA, quoi ?

Qu'est ce qui a changé pour les entreprises, avec l'arrivée du BD ?



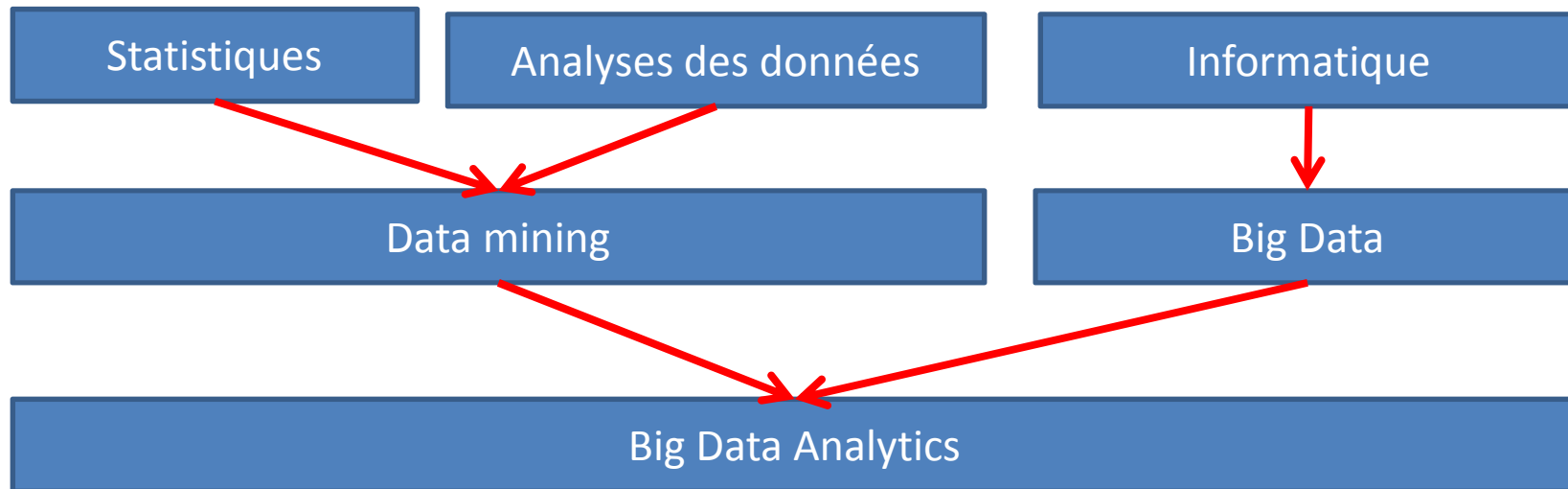
# Et le BIG DATA ANALYTICS !?

*On connaît le BD maintenant, alors c'est quoi le BDA?*



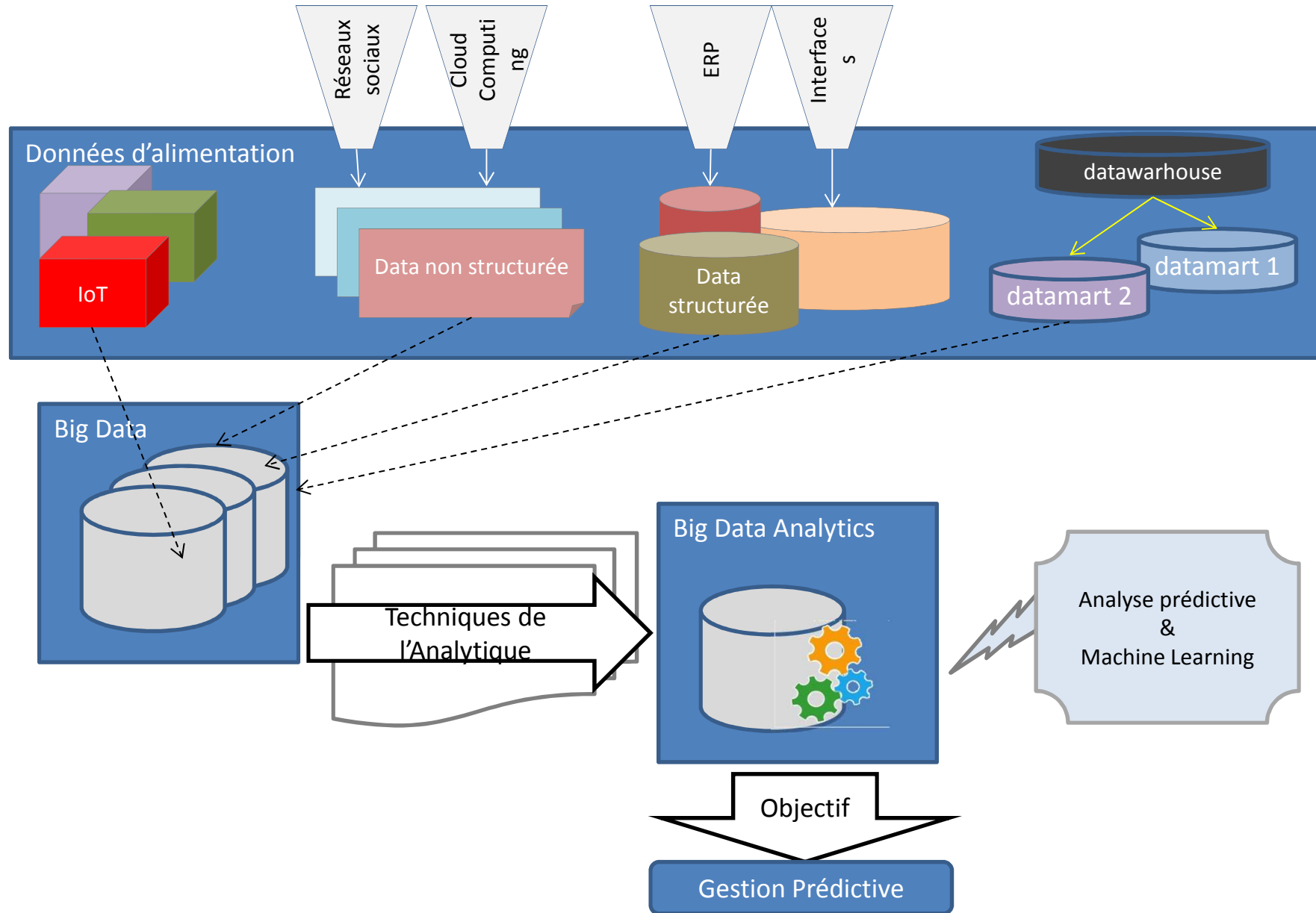
## Lien avec l'existant ?

*Comment les SI ont évolué pour muter au BDA?*



# Et la solution système, comment ?

Comment peut-on concevoir la solution informatique basée sur la BDA?



# Sources des données ?

Où trouver les masses de données, en repositories ou en flux temps-réel ?



POS : Points of Sales



PCR : Payment Cards Readers



GPS : Global Positioning System



SM : Social Media



IS : Information Systems



IoT : Internet of Things



RFID : Radio Frequency ID



Prévisions météo



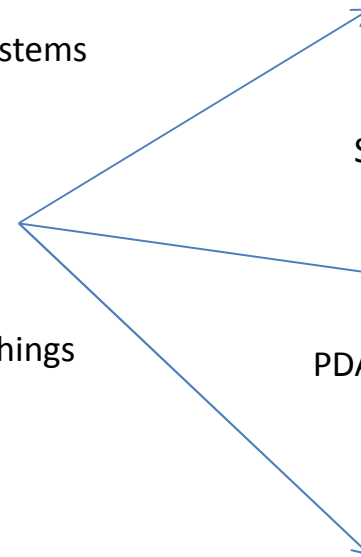
SP: smartphones



PDA: plannings mobiles



VC: véhicules connectés



# Problématiques liés à la recherche ?

En recherche académique ou industrialo-académique, quels sont les problèmes liés à la GSCM-BA

## Problèmes

Jeux de données ?

- Volumétrie : Où trouver des masses de données ?
- Disponibilité : Quelles entreprises, mettraient leurs données à disposition ?



## Solutions ?!

L'utilisation de référentiels de données:

- A partir d'entreprises impliquées dans des projets de recherche (Facebook, Amazon, etc.),
- A partir de projets de recherches académiques en open sources,
- De plus en plus de repositories de BDA (référentiels) sont disponibles :

Ex: <https://archive.ics.uci.edu/ml/index.html>

Ex: <https://datahub.io/dataset/master-global-garment-supply-chain>

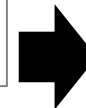
Ex: <https://supplychainco2.dge.carnegiescience.edu/>

- Mise en œuvre complexe
- infrastructure complexe
- choix technologique



- Réutilisation des infrastructures existantes, puis les étendre ?
- bases de données relationnelles montées en parallèles ?
- Repositories ?
- Cloud Computing ?

- Confidentialité des données
- Données ouvertes pour bénéficier du BD ?



- Système de GET without POST , est-il possible ? Si oui, alors données restreintes !
- Quelle donnée déposer à tous et quelle donnée garder confidentielle !



**Merci**

---