

# Big Data Open Data

MMI 2 – TP#2 S4

The background features a dark gray field with two large, abstract geometric shapes: a bright yellow triangle on the left and a light red triangle on the right, both pointing towards the center.

Danielo **JEAN-LOUIS**  
Développeur front-end

# Big Data – En méga chiffres

- Internet génère 2,5 QB de données chaque jour
- facebook génère 100 TB/jour
- Chaque personne génère 1,7 MB de données par seconde
- Chaque jour 65 milliards de messages sont échangés sur WhatsApp

## Sources :

- <https://techjury.net/blog/big-data-statistics/> - anglais
- [https://web-assets.domo.com/blog/wp-content/uploads/2017/07/17\\_domo\\_data-never-sleeps-5-01.png](https://web-assets.domo.com/blog/wp-content/uploads/2017/07/17_domo_data-never-sleeps-5-01.png) - anglais

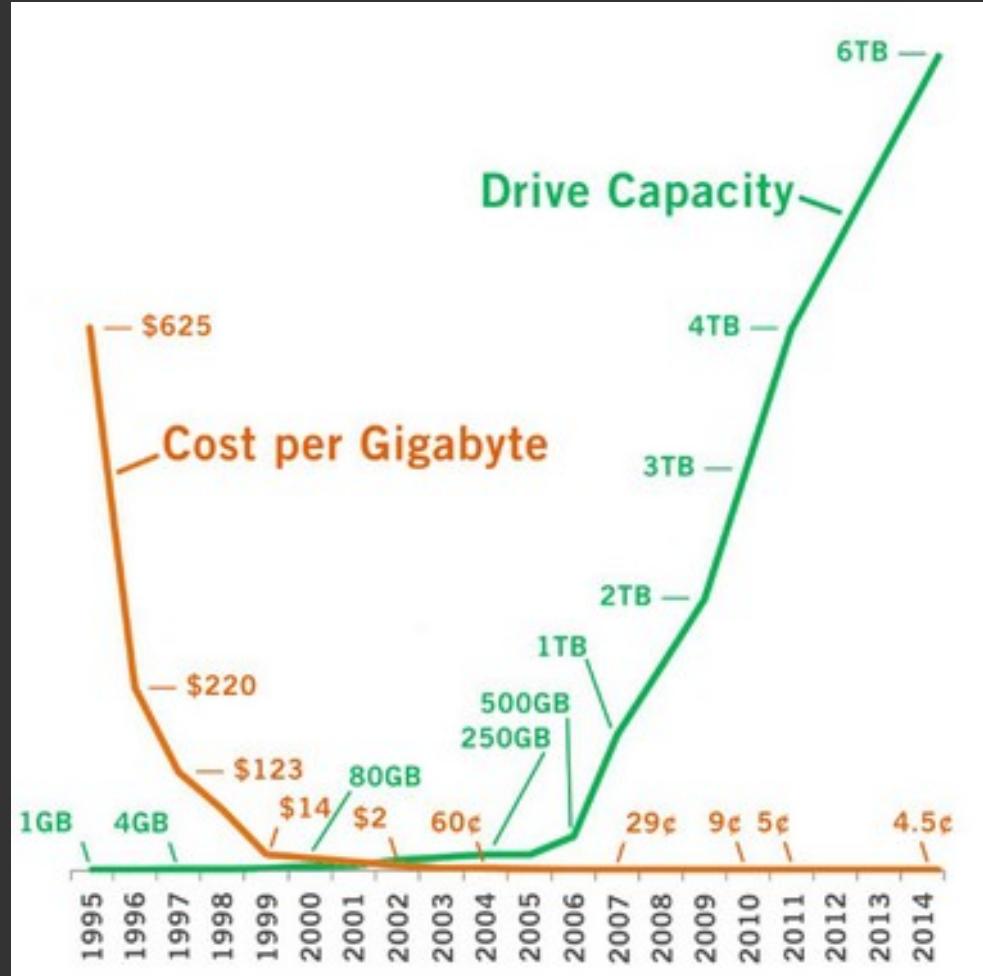
**On stocke ceci où ?**

# Big Data – Stockage

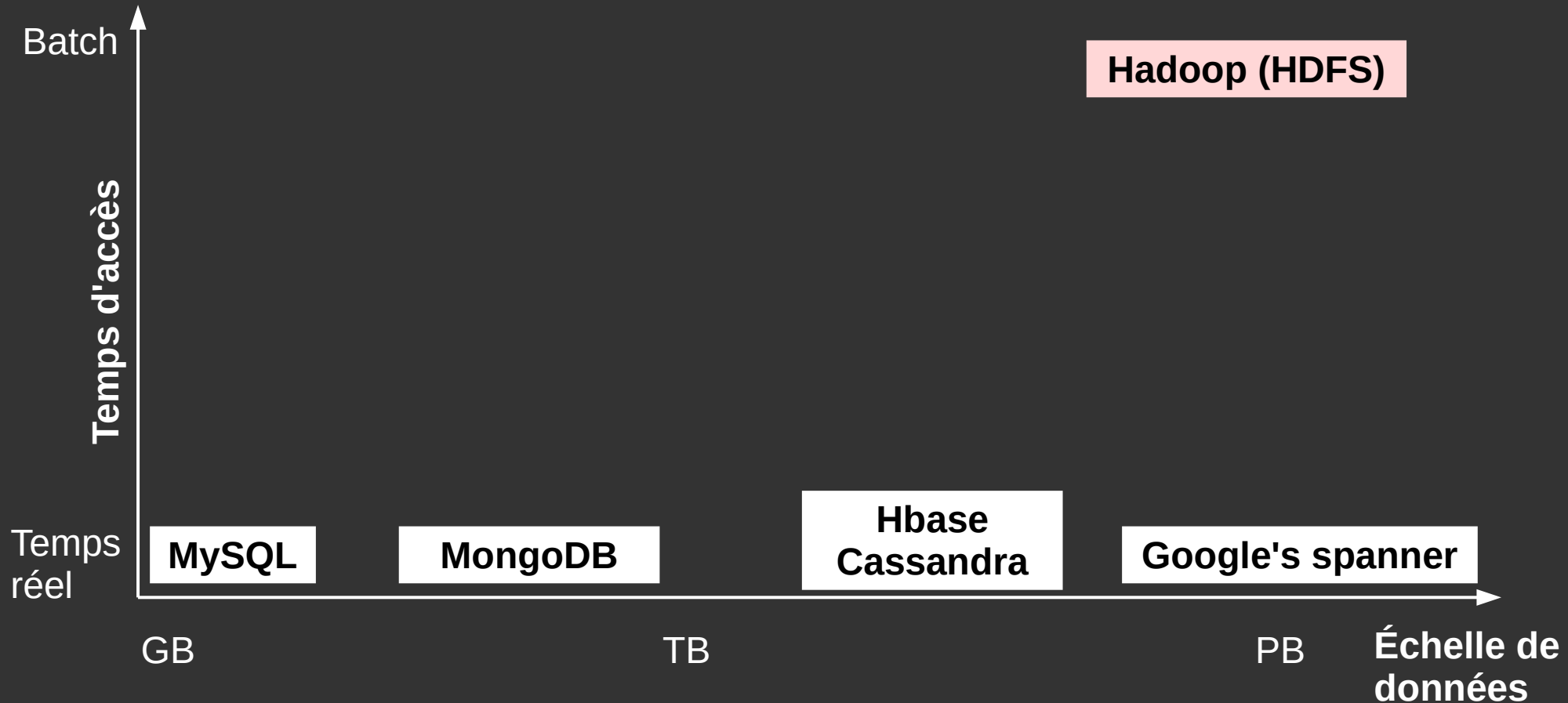
- Ordinateur/smartphone/tablette/DD
- Serveur
- Rack (collection de serveurs)
- Datacenter (collection de racks)
- Cluster (collection de datacenters)

# Big Data – Stockage – Des prix en baisse

- Baisse des prix
- Augmentation des capacités de stockage
- Pas d'espace de stockage, pas de big data



# Big Data – Stockage – Panorama technologique

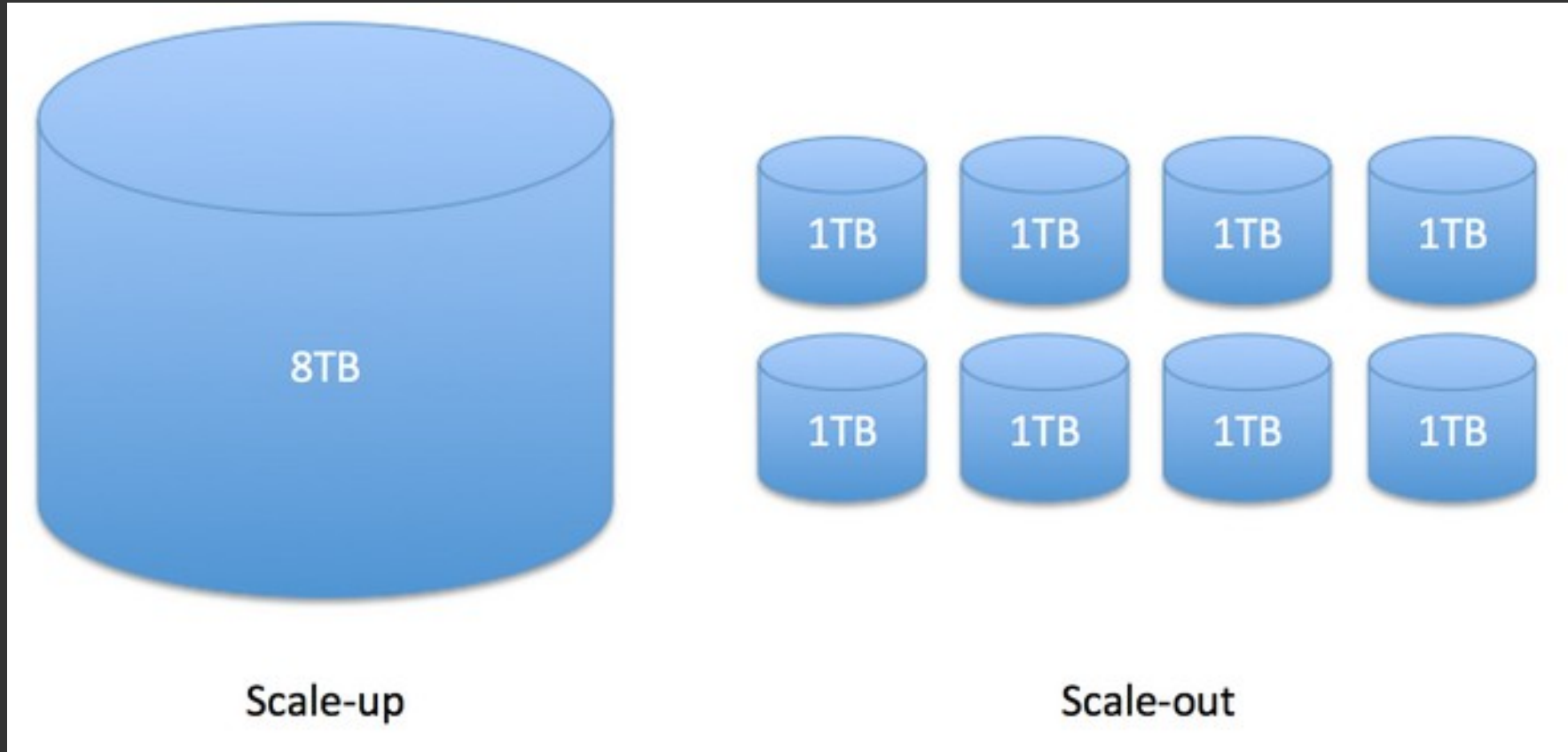


# Stockage – SGBDR vs HDFS

Caractéristiques	SGBDR (Ex: MySQL)	Hadoop (HDFS)
Stockage	Base de données	Fichier
Transactions	Oui	Non (La donnée mise à jour ne sera pas forcément accessible)
Structuration des données	Oui	Non
Données en temps réel	Oui (Flux)	Non (Batch)
Sensible aux pannes	Oui	Non
Mise à l'échelle	Non linéaire (scale-up)	Linéaire (scale-out)

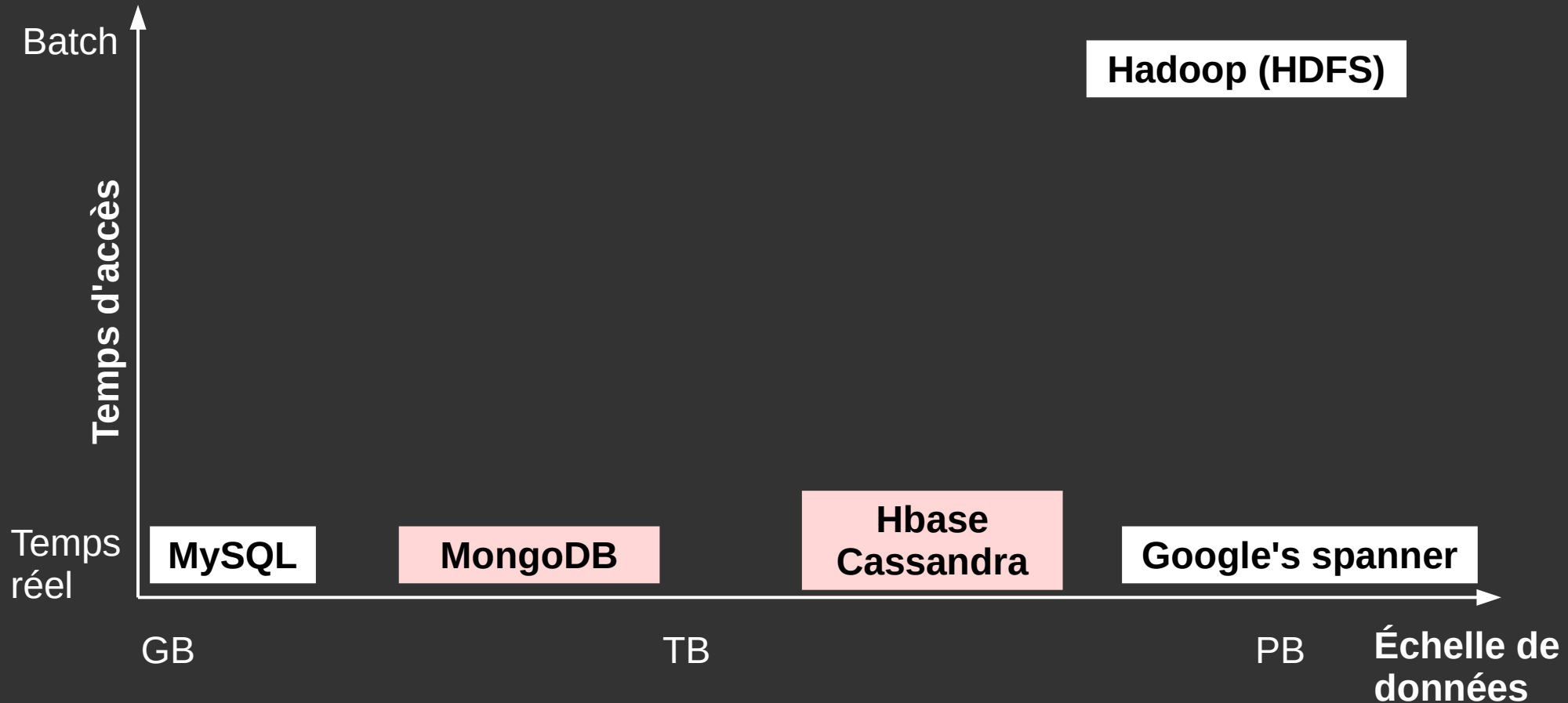


# Big Data – Scale-out vs Scale-up



L'architecture scale-up est beaucoup plus onéreuse

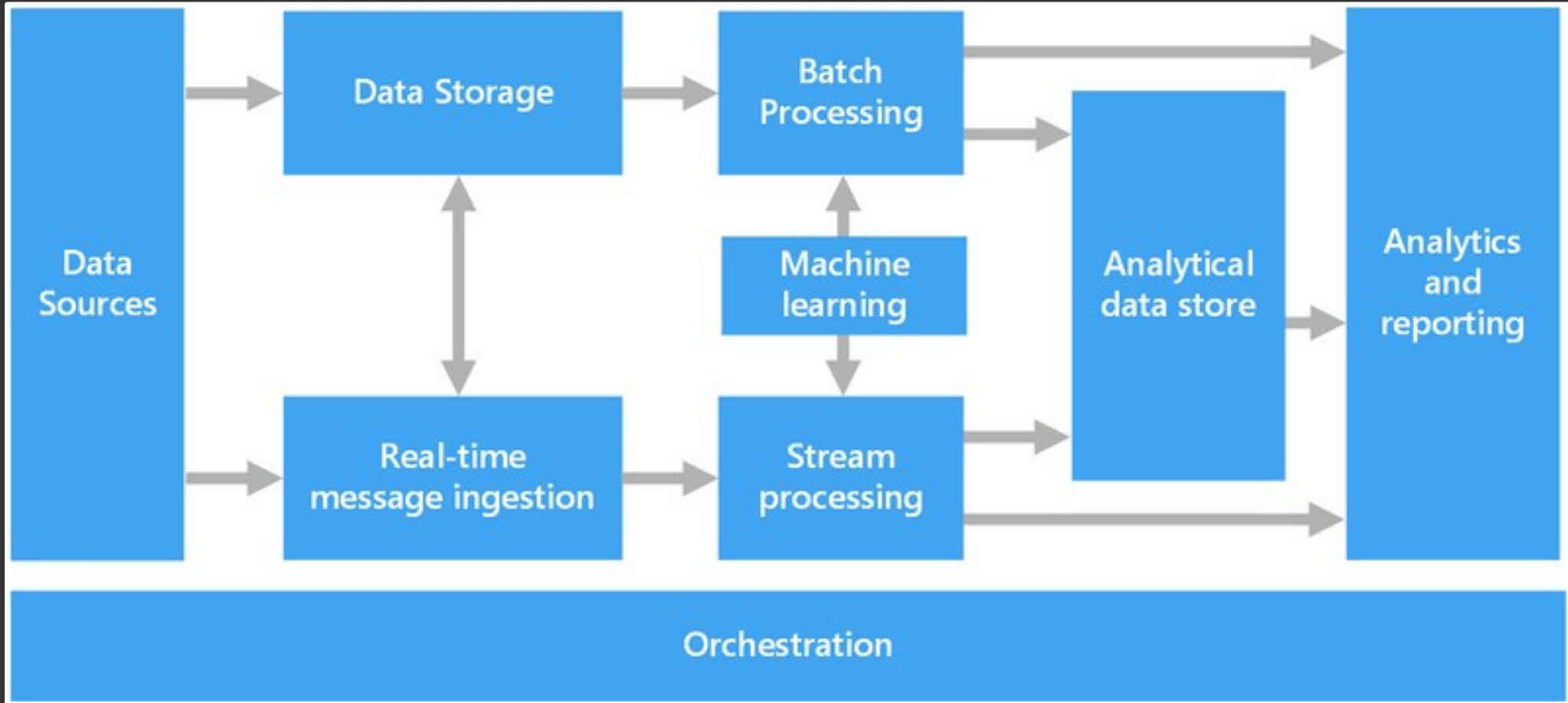
# Big Data – Stockage – Panorama technologique



# Stockage – SGBDR vs NoSQL

Caractéristiques	SGBDR (Ex: MySQL)	NoSQL
Stockage	Base de données	Fichier
Structuration des données	Oui	Facultatif
Jointures	Oui	Non
Sensible aux pannes	Oui	Non
Mise à l'échelle	Non linéaire (scale-up)	Linéaire (scale-out)

# Cheminement de la donnée



# Cheminement de la donnée - Architecture

- La plus répandue : Datalake
  - Plus simple à mettre en place
  - La moins chère
- Kappa / Lambda / SMACK
  - Chacune à des avantages et des inconvénients

## Sources :

- <https://docs.microsoft.com/fr-fr/azure/architecture/data-guide/big-data/>

**Que pouvons-nous faire avec ces  
données stockées ?...**

**De la science, mieux de la  
data-science !**

# Data-science

- Secteur en aval du big data (stockage) et en amont du produit (marketing)
  - Ex : Le pôle d-s va trouver une tendance pour sortir un nouveau produit
- Deux grandes branches :
  - data-architects : définit la plate-forme technique
  - data-analystes : applique des algorithmes sur la donnée

## Sources :

- <https://www.lebigdata.fr/data-science-definition>



**Questions ?**

