

# Base De Données Avancée

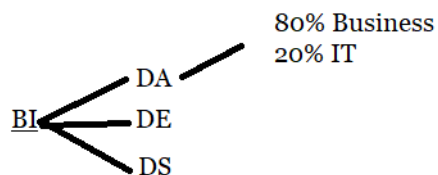
|         |  |
|---------|--|
| ➤ Event |  |
| ➤ Tasks |  |

## ▼ DA,DS,DE

DA :Data analyst

DS : Data Scientist

DE : Data Engreener



- différence entre DA et DS

Un **Data Analyst** , comme son titre l'indique, a pour rôle d'analyser les données. Le **Data Scientist** va plus loin, et possède une expertise métier et des compétences en « **Data** Visualization » (visualisation **de** données).

⇒le ds analyse le données plus intelligente par rapport au da

- Quelle différence entre data Engineer et data scientist ?

Alors que les **data scientists** utilisent leurs compétences pour créer des modèles et résoudre des problèmes, les **data engineers** construisent et gèrent l'infrastructure qui se situe **entre** les sources de données et l'analyse des données

## ▼ Types D'analyses:

### 1. Analyse Descriptive:

exp: résultat bac par région par spécialité dans l'année 2020 ⇒pas intelligente mais permet de mieux comprendre. analyse quantitative

### 2. Analyse Diagnostique:

exp pourquoi les résultats du bac du Zaghouen sont basses? ⇒ analyse qualitative

### 3. Analyse Prédicative:

exp: on a remarqué que Zaghouen va rester dans le même niveau mais Kirouen a augmenté . ⇒ twakaat haja ou fasartha

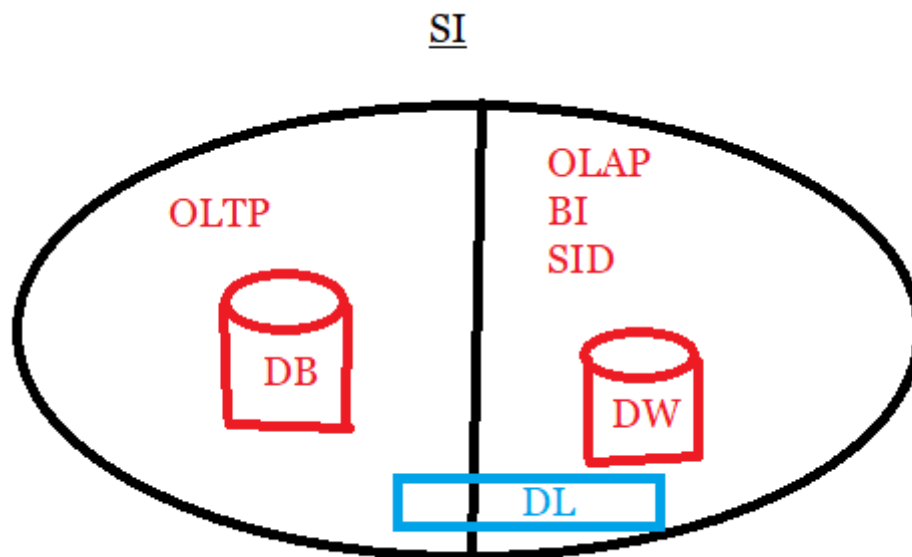
### 4. Analyse Perspective:

quelles sont les actions à prévoir

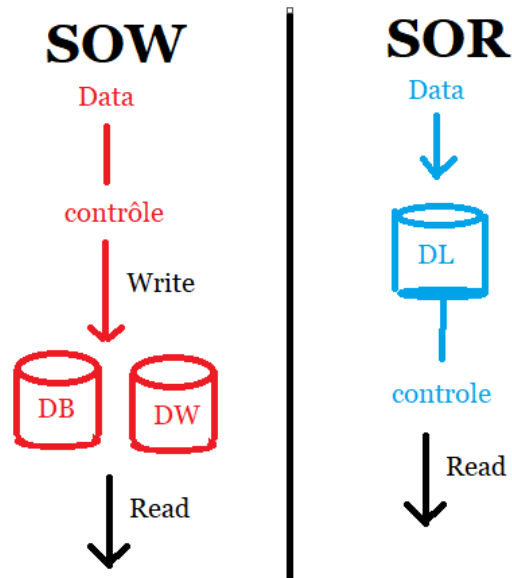
### 5. Analyse cognitive:

exp: la conduite automatique de la voiture

## ▼ Mots clés



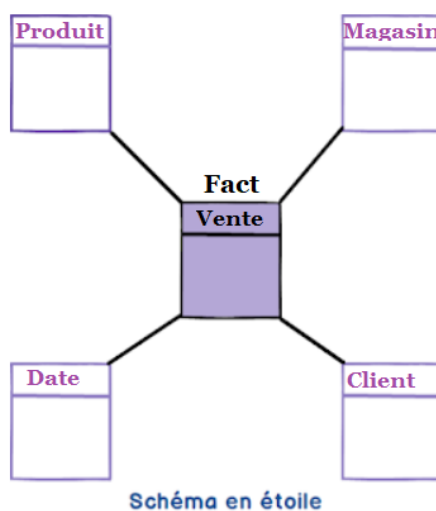
- **OLTP** :c'est le sous système qui s'occupe de l'activité quotidienne de l'entreprise
- **DATA WAREHOUSE**: conteneur de données structurée relationnel orienté vers l'analytique
- **Data Lake**: conteneur de données non structurée et structurée orienté 80% analytique 20% prod exp Mongo DB
- **Schema on write**: structure de données qui peut correspond au DB ET DL
- **Schema on read**: quand les données arrivent il n'y a pas de contrôle correspond au DL.



▼ **Modèle en flocon de neige:**

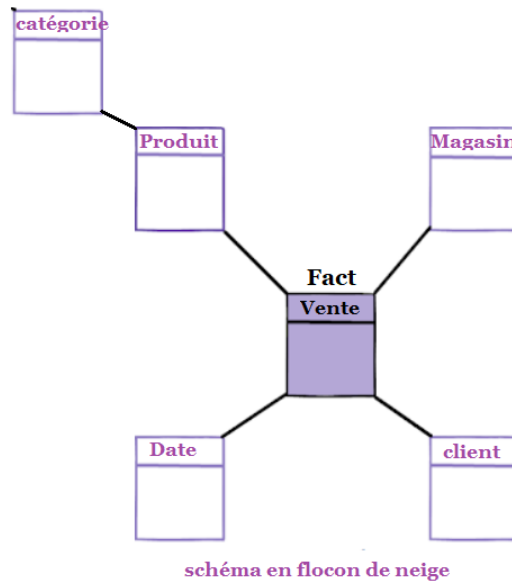
modèle de données pour le DW

- différence entre modèle en étoile et en flocon de neige:



**Le schéma en étoile**

est le paradigme de modélisation simple et commun dans lequel l'entrepôt de données comprend une table de faits avec une seule table pour chaque dimension.



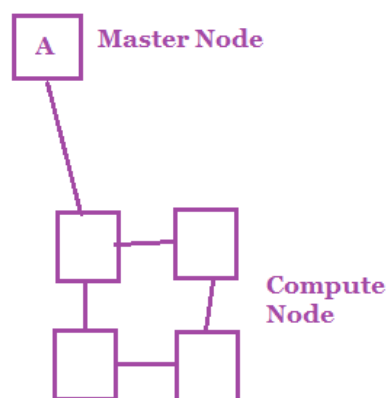
### Le schéma en flocon de neige

est un type de **schéma en étoile**

qui inclut la forme hiérarchique des tables dimensionnelles.

- **vélocité**: vitesse d'acquisition de données
- **Cluster**: un élément d'un système qui n'a pas d'autonomie individuelle  
 ⇒ un cluster est un ensemble de machines interconnectés pour donner une solution compacte (pour collaborer) quand au réseau est un ensemble de données connectées pour échanger des informations.

exemple :



- **Etl**: mécanique / process qui transfert de données de l'OLTP au OLAP  
 exemples de logiciels ETL:
  - ibm datastage

- siss
- odi
- **BIG DATA:** domaine de l'informatique qui traite les données massives variées avec vélocité. (grand volume, variété de structure, grande vélocité qui caractérisent les projets contemporains )