Ce contenu est destiné pour les Ingénieurs Affaires

# Data pour les non techniques.

**Objectif :** Que ce soit Business Intelligence ou Big Data, l’objectif final, c’est toujours l’aide à la décision que ce soit amélioratif ou Prédictif ou Curatif

**Business Intelligence**

Business Intelligence, en générale le périmètre de données reste dans le domaine de l’entreprise fournit ou générer par les activités de l’entreprise. (Ventes, Client, …)

Un **job** c’est quoi ?

Un job c’est un traitement ou programme qui ne fait qu’une chose à la fois.

Dans le domaine de data, lorsqu’on est encore en amont de la restitution de données. Ces traitements concernent toujours les flux de données.

**ODS :** Operational Data Storage ( Données Opérationnelles)

Pour effectués ces tâches là on va créer plusieurs jobs soit en utilisant des outils dédiés comme (Talend DI (Data integration), Datastage, Informatica, SSIS…)

Soit via les langages de programmation comme Python , Java ou Scala, R …

Outils Data visualisation ( Qlik View, Qlik Sense, Power BI, …)

ou langages Python

Jobs ETL ou Langages

Jobs ETL ou Langages

**Partie 4**

**Partie 3**

**Partie 2**

**Partie 1**

Data Ware house (DWH) ou entrepôt de données

Data Ware house (DWH) ou entrepôt de données

**ODS**

**DWH**

**Datamarts**

**Data visualisation**

- Courbes

- Graphes

Datamarts : c’est comme un data ware house mais subdivisé selon les domaines métiers (Finances, Marketing …)

Centralisation de données dans une seule Base de données relationnelle en données brutes structurées

**Données structurées**

Fichiers CSV, TXT

SAP

Base De données relationnelles(Oracle, SQL Server …)

Les logiciels Métiers

Sites Web

**Sources**

**Pipeline de données :** La création de ces pipelines de données c’est le métier de **Data engineer**

**Partie 1**

**Partie 2**

**Partie 3**

**Data Pipeline**

**Pipeline de données** c’est l’ensembles des jobs (Partie 1 + Partie 2 + Partie 3)en allant des sources jusqu’à l’arrivé des données aux Datamarts ou DataWarehouse

Vous voyez qu’il y a un ordre à suivre ainsi, pour pouvoir suivre cet ordre, on doit utiliser des outils dédiés ou des ordonnanceurs ou en fabriqué avec les langages d’automatisations des taches comme Shell ou Python. Mais ces Ordonnanceurs-là, se différencient selon les outils ou langage que vous avez utilisé lors du développement de vos jobs.

Un DataEngineer ou Ingénieur de données c’est celui qui met en place le pipeline de données jusqu’à l’arrivé de données une base de données décisionnelle (DataWare House ou Datamart ou Datalake). C’est-à-dire qu’il va créer les jobs de l’acquisition de données jusqu’à l’arrivée dans une base de données décisionnelles et les orchestrer ou ordonnancer pour avoir le pipeline de données

Les compétences qu’il devrait avoir :

Connaissance d’un ou des langages de programmations(Python, Java, Scala) pour créer les jobs en cas ou le projet n’a pas d’outils ETL dédié comme (Talend, Datastage, Informatica etc…)