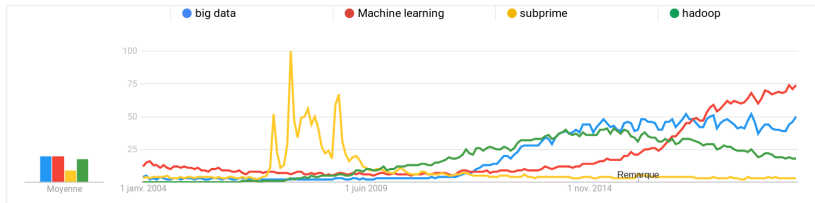


Machine Learning , méthodes et solutions

Sources de données

Sources de données



Sources de données

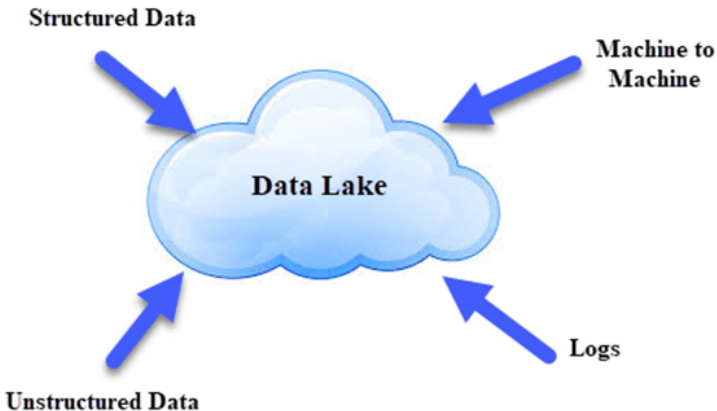
Les différents formats de sources de données :

- Fichiers (TXT, CSV, Xml, Json, ...)
- BDD Relationnelles : SQL
- BDD “pas que” Relationnelles : NoSQL (Représentation non-tabulaire des données)



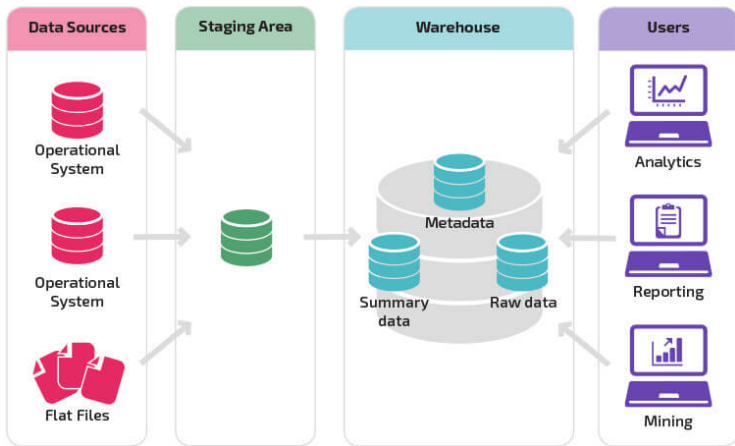
Sources de données

Dans un premier temps, les données internes de l'entreprise

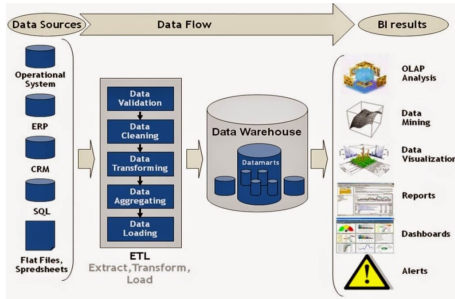


Sources de données

Entrepot de données sur les clients (Datawarehouse)



Extract Transform Load (ETL)



Orientés **synchronisation**

- Blendo
- Fivetran
- Segment
- Stitch
- ...

Orientés **Transformations**

- DataVirtuality
- Alooma
- Etleap
- Xplenty
- ...

Sources de données

Processus de gestion (ERP)



Forme des données :

- logs d'activité (tabulaire)
- IOT (capteurs)
- Réseaux sociaux
- Open-data (data-gouv.fr, geonames.org, ...)
- Web crawling

Bases de données sur les clients disponibles :

- Institut de sondage
- Réseaux sociaux
- Profils clients
- ...

Des places de marché en ligne de données :

- Microsoft Azure Marketplace
- Datamarket
- Data Publica
- ...