

ETL (Extract – Transform -Load)

Presented by:

AYMANE MAGHOUTI (Data Engineering Student (ID2))

Outline:

- Context (Problem)
 - Objectif
 - ETL Pipeline
 - ETL steps
- Implement your ETL process
 - Native Method
 - Using ETL Tools
 - Process Type
 - ETL vs ELT
 - Demo (Mini project)



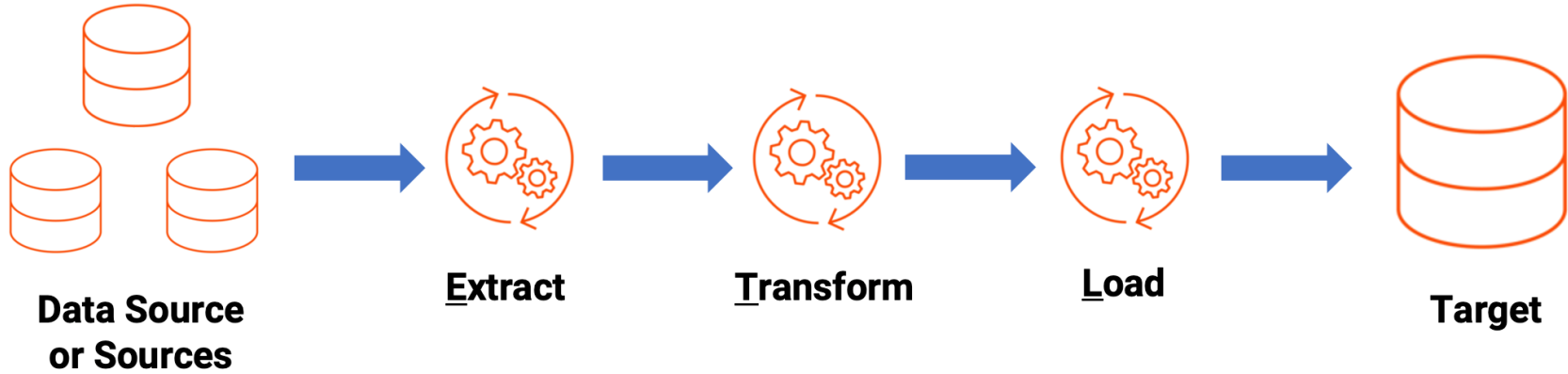
Context (Problem)

Objectif

- Powering data warehouses
- Data integration



ETL Pipeline





ETL steps

Step 1 : Extraction

◎ Data Source :

- Relational database (MySQL, ORACLE, PostgreSQL, etc.)
- Non-relational database (MongoDB...)
- WEB API (Youtube API)
- Web application (using scrapping instead of api)
- JSON, CSV, XML....
- Social media (FB, Instagram, ...)



Extraction Type:

- **Full Extraction:** extracting all the data from the source system each time the ETL process runs.
- **Incremental Extraction:** retrieving only the data that has been added or modified since the last extraction.
- **Partition Extraction:** dividing the dataset into smaller partitions or subsets based on certain criteria.



Step 2 : Transform

- **Data cleaning** : removing duplicates, null_values...
- **Data enrichment**: adding additional information to the data
- **Data Transformation**: converting the data into a format that is suitable for loading into the target system.
- **Data validation**: ensure that the data being used is reliable and trustworthy.

Step 3 : Loading

- **Full Load** : Loading all the data from the source into the target system each time, replacing existing data entirely.
- **Incremental Load** : Loading only the new or changed data since the last load, minimizing processing time and resource usage.
- **Delta Load** : Loading only the changes (or "delta") between the source and target systems since the last load, typically used for real-time or near-real-time data synchronization.

- Extraction



ORACLE



- Transformation



- Loading



Implement your ETL process

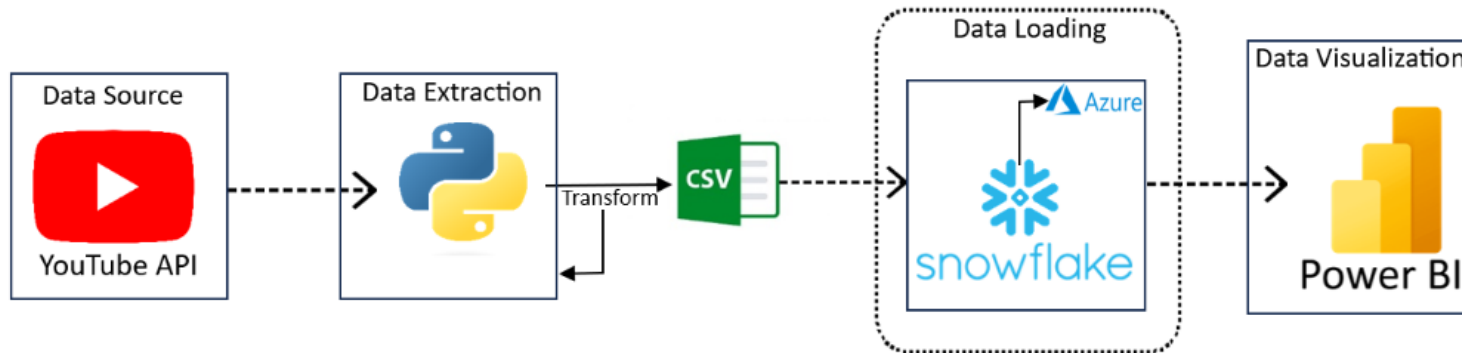
- Define the source
- Choose the data to extract
- Plan the transformation
- Choose the target

Native Method

- Using programming languages (ex. Python)

Example:

YouTube Data Pipeline

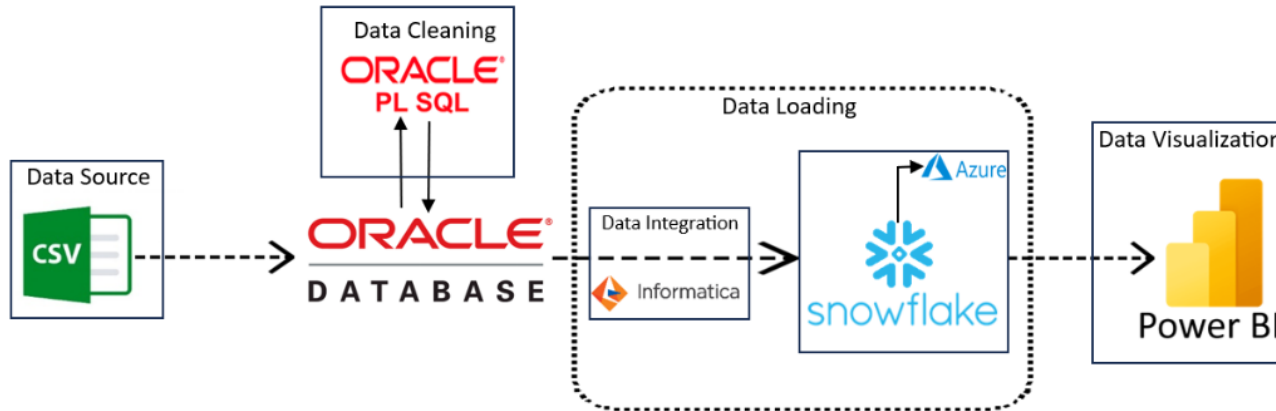


By Aymane -MG

Using ETL Tools

- Informatica
- Talend
- SSIS

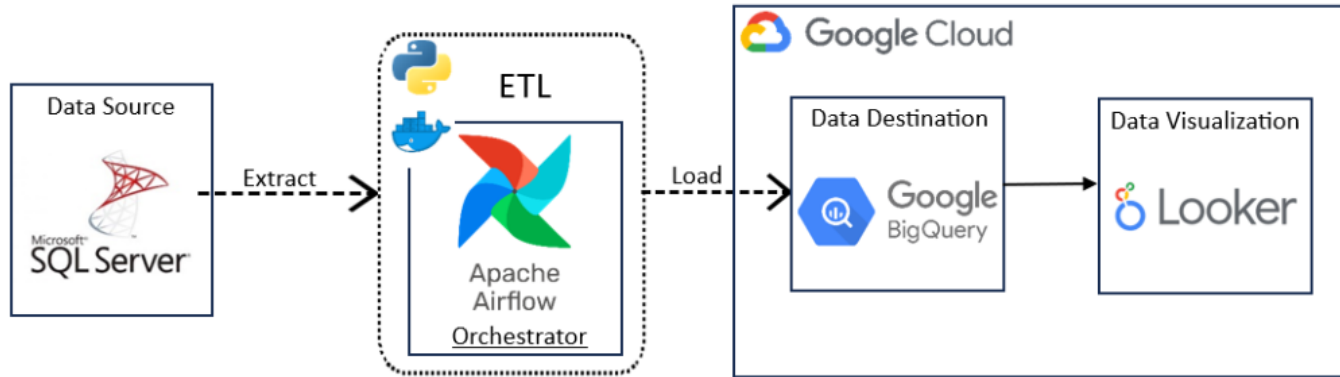
Example



By Aymane -MG

Orchestration

Example:



By Aymane -MG

Process Type

batch processing



stream processing

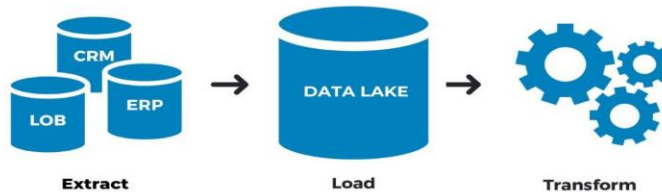


ETL vs ELT

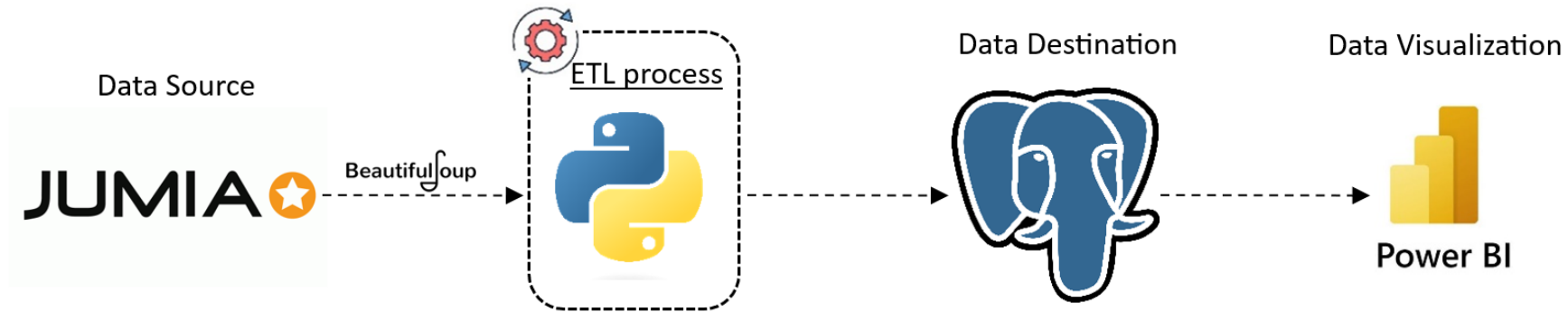
ETL

VS.

ELT



Demo (Mini project)



By Aymane-MG



Thank you for your attention !

