Chapter 2: Data Warehousing

Implementing a Data Warehouse

1. Gather requirements
2. Create warehouse environments
3. Choose a data model
4. Connect to sources
5. Transform incoming data
6. Create data marts
7. Configure BI and analytics
8. Audit and review

Gather requirements

Implementing a data warehousing is a team project. Ensure to involve all stakeholders

- Business personnel and decision makers
- IT personnel
- Analytics team
- Security and Compliance team

Create warehouse environments

- On-premise: Host on local hardware
- Public cloud: Use a hosted cloud solution, like AWS or Azure
- Private cloud: Host a cloud on your own hardware, or hire a trusted third party
- Hybrid cloud: Either mix on-premise and cloud storage or store data on-premise and use cloud capability for processing and analytics

Create warehouse environments

- On-premise: Host on local hardware
- Public cloud: Use a hosted cloud solution, like AWS or Azure
- Private cloud: Host a cloud on your own hardware, or hire a trusted third party
- Hybrid cloud: Either mix on-premise and cloud storage or store data on-premise and use cloud capability for processing and analytics
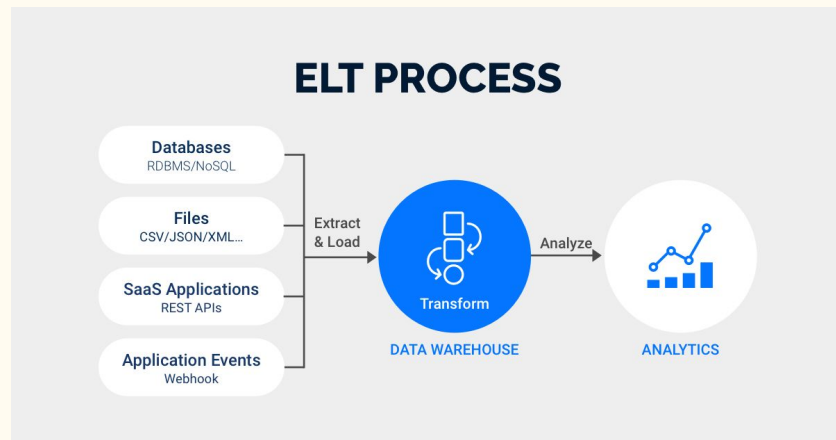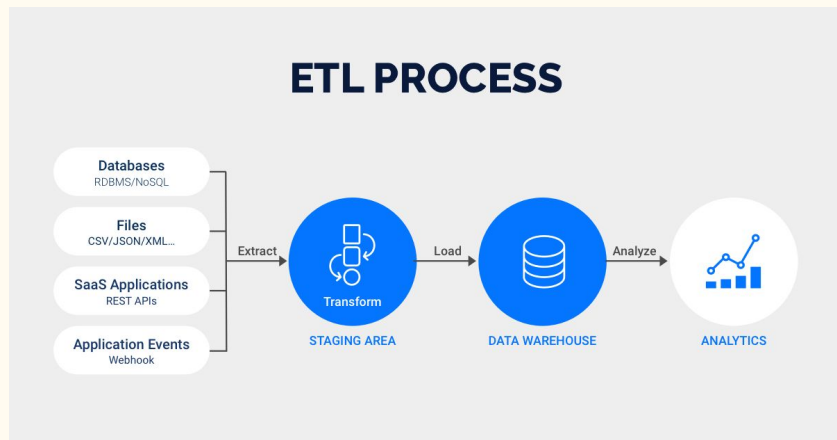
Choose a data model

Choose a model that is suitable for the existing data and is scalable.

Commonly used schemas:

- Star schema
- Snowflake schema
- Galaxy schema
- Constellation schema

Connect to sources, transform incoming data

Extract data from the target source, and then **L**oad it to the data warehouse after/before **T**ransforming the data.
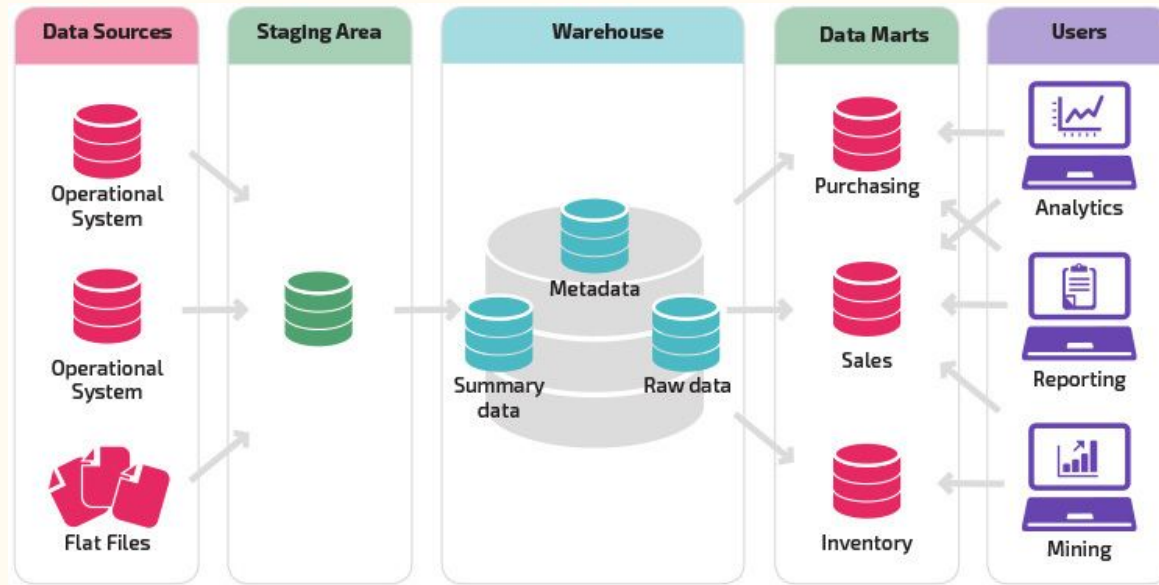
Connect to sources, transform incoming data

ETL tools:

Fivetran, Integrate.io, Matillion, IBM Cognos Data Manager, Apache Airflow, Apache Kafka etc.

Create data marts

Marts are a logical division within the warehouse – a limited view that only shows relevant results.

Configure BI and Analytics

BI and analytics tools help understand trends, identify patterns, and give insights which are crucial for decision making.

Examples of BI/analytics tools:

SAS BI, Microsoft Power BI, Tableau, Sisense, QlikSense, Oracle BI, Zoho Analytics etc.

Audit and review

Measure the quality of your warehouse contents

Identify the discrepancy between the raw data and the transformed data stored in the data warehouse