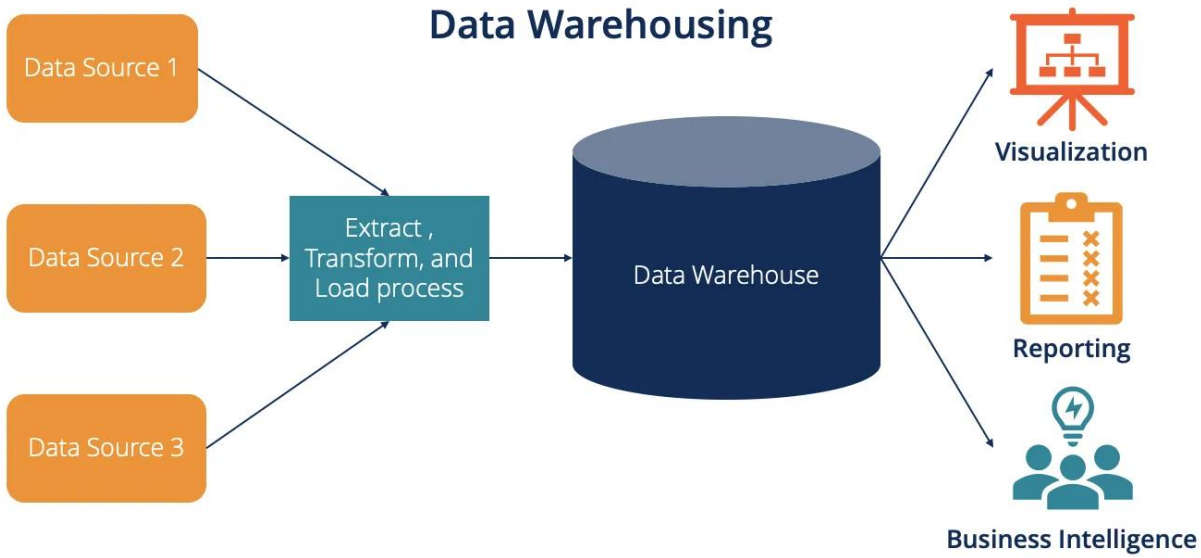
1. An Introduction to Data Warehousing

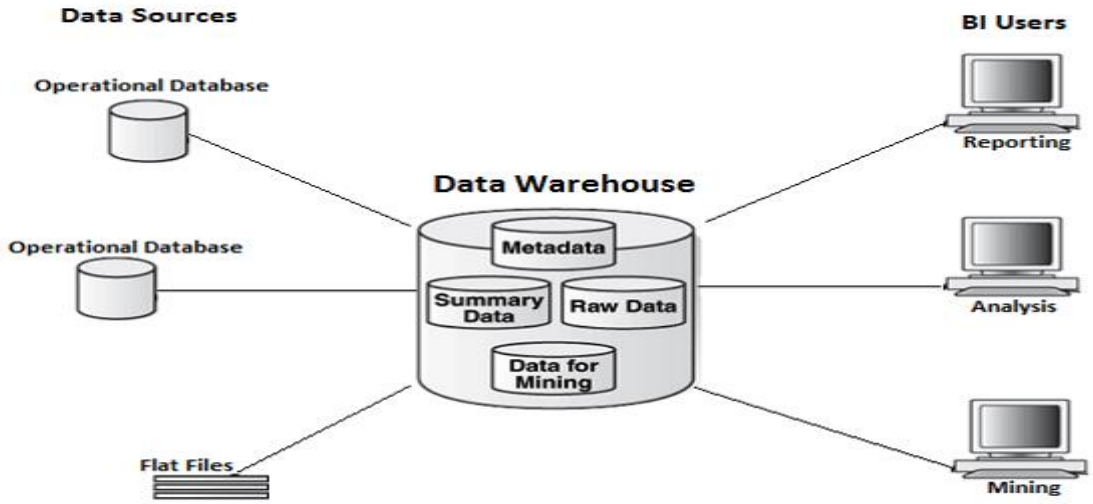
Data Warehousing is a technology that collects, stores, and manages large volumes of historical data from various sources to support business analysis and decision-making. Unlike operational databases, data warehouses are optimized for reading and analyzing data rather than transactional processing.



2. Purpose of Data Warehouse

Here are the 3 main purposes of a Data Warehouse:

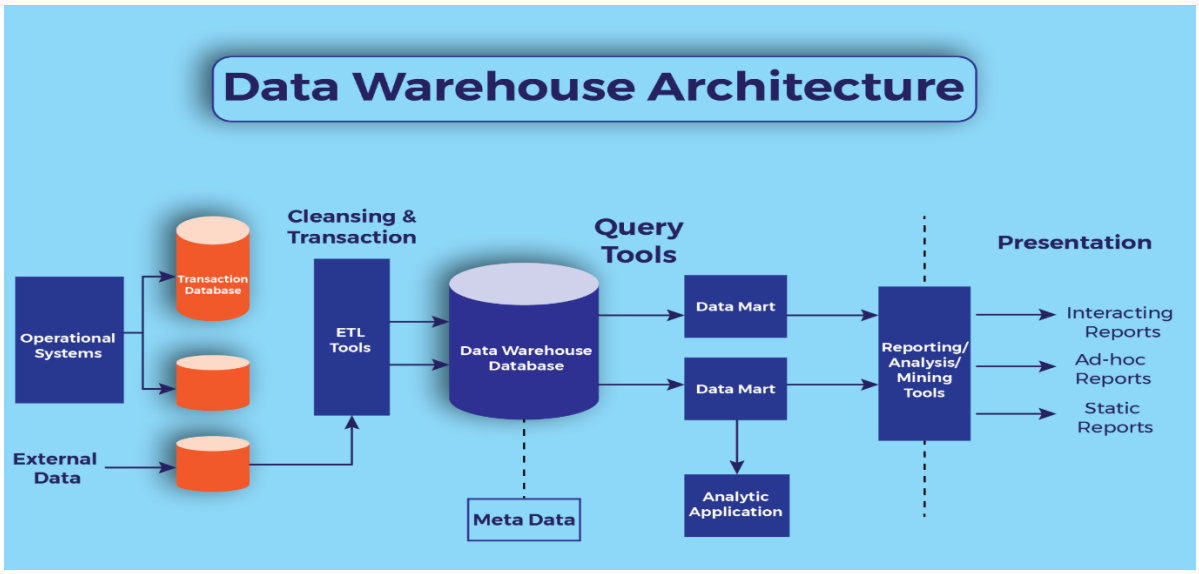
1. Centralized Data Storage:
   Integrates data from multiple sources into a single, unified repository.
2. Support for Decision Making:
   Provides accurate, historical, and analytical data to help businesses make informed decisions.
3. Efficient Data Analysis & Reporting:
   Enables fast, complex queries, trend analysis, and business intelligence reporting.



3. Data Warehouse Architecture

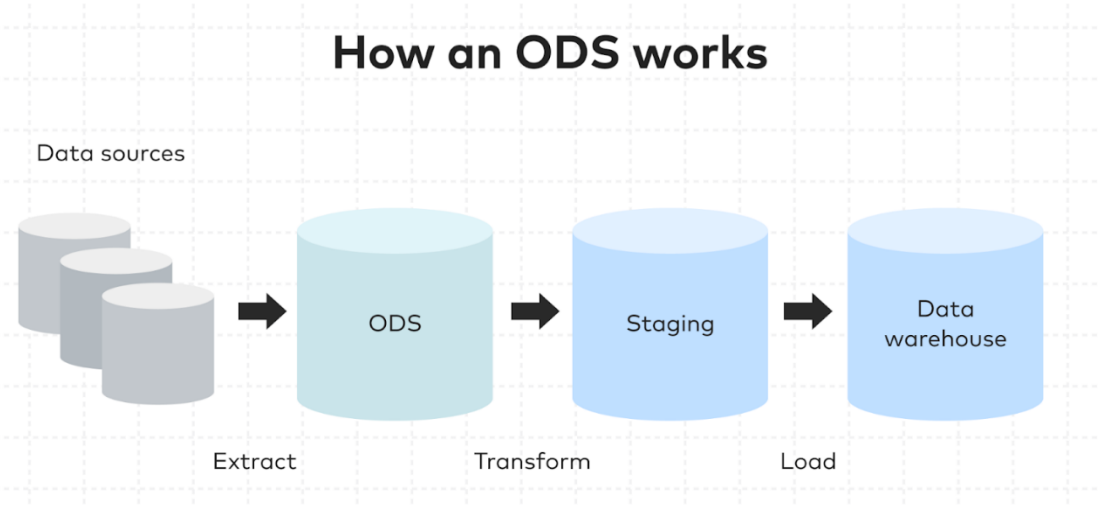
Typically includes three layers:

- Data Source Layer (operational DBs, flat files)
- Data Staging Layer (ETL process)
- Presentation Layer (data marts, OLAP cubes, reporting tools)



4. Operational Data Store (ODS)

An ODS is an intermediate layer between operational systems and a data warehouse. It stores current operational data in a central database for short-term decisions and is often updated in real-time.



5. OLTP vs Warehouse Applications

OLTP (Online Transaction Processing): Optimized for quick insert/update/delete operations. Used for day-to-day operations.

Data Warehouse: Optimized for read-heavy operations, complex queries, and data analysis.

OLAP	OLTP
Involves historical processing of information.	Involves day-to-day processing.
OLAP systems are used by knowledge workers such as executives, managers and analysts.	OLTP systems are used by clerks, DBAs, or database professionals.
Useful in analyzing the business.	Useful in running the business.
Contains historical data.	Contains current data.
Provides summarized and multidimensional consolidated data.	Provides primitive and highly detailed data.
Number or users is in hundreds.	Number of users is in thousands.
Number of records accessed is in millions.	Number of records accessed is in tens.
Database size is from 100 GB to 1 TB	Database size is from 100 MB to 1 GB.
Highly flexible.	Provides high performance.
Based on Star Schema, Snowflake, Schema and Fact Constellation Schema.	Based on Entity Relationship Model.

6. Data Marts

A data mart is a subset of a data warehouse focused on a specific business line or team (e.g., sales, HR). It provides quicker access and is easier to maintain for specific needs.

7. Data Marts vs Data Warehouses

Data Warehouse: Enterprise-wide, contains integrated data from all sources.

Data Mart: Department-specific, contains focused datasets.

8. **Data Warehouse Life Cycle**