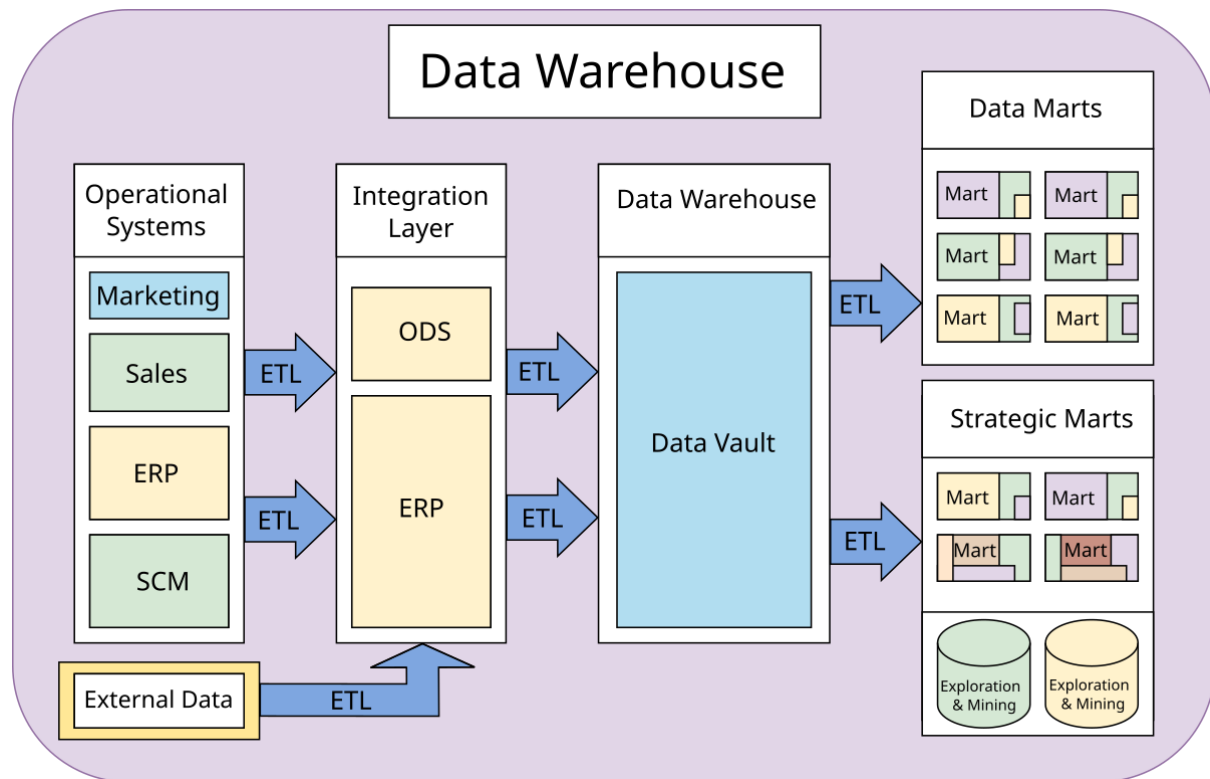


## An Introduction to Data Warehousing

From what we understood in class, a data warehouse is like a big storage system where data from different sources is brought together. It is mostly used for analysis and reporting, not for daily operations. It helps us make sense of large amounts of data from the past.



## Purpose of Data Warehouse

As discussed in class, the main reasons for using a data warehouse are:

- To combine data from multiple sources into one place.
- To make sure the data is consistent and accurate.
- To store historical data and analyze trends.
- To help managers and analysts make better decisions.
- To reduce the load on operational systems when generating reports.

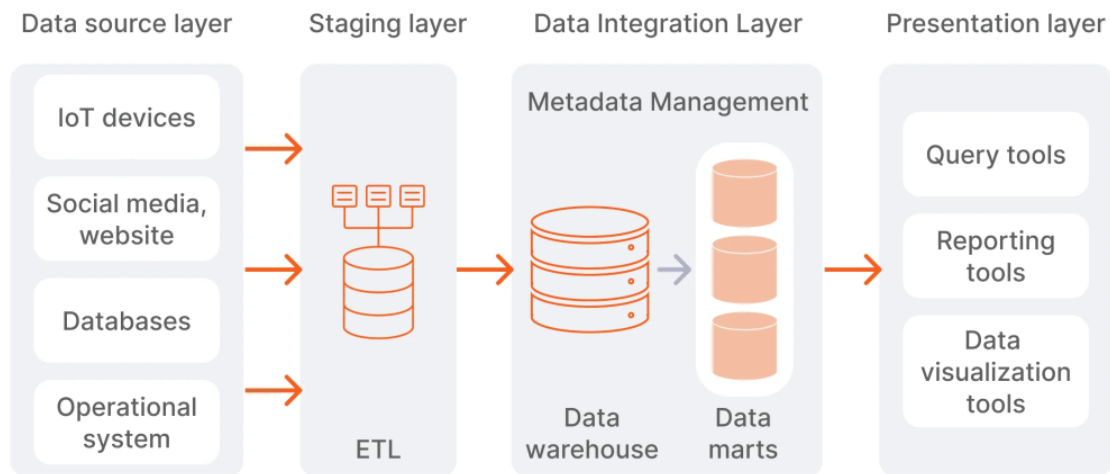
## Data Warehouse Architecture

A typical data warehouse architecture includes the following layers:

1. **Data Source Layer:** Includes operational databases, flat files, and external sources.
2. **Data Staging Layer:** Data is extracted, transformed, and loaded (ETL) here.
3. **Data Storage Layer:** Data is organized in a schema (star or snowflake).

4. **Data Presentation Layer:** Provides access to data through business intelligence tools.
5. **Metadata and Management Layer:** Stores information about the data and manages the warehouse.

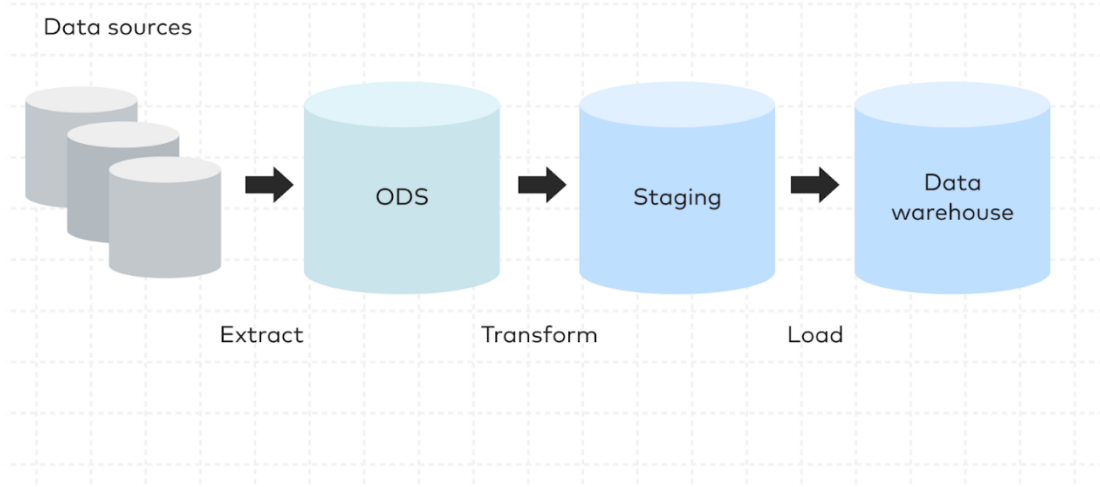
## Data Warehouse Architecture Layers



## Operational Data Store (ODS)


ODS is a kind of mini-warehouse. It stores data from different sources in real-time and is used for day-to-day reporting. It's not as complex or historical as a data warehouse.

## How an ODS works



## OLTP vs Warehouse Applications(OLAP)

OLTP	OLAP
Records high volumes of simple online transactions rapidly	Collects large amounts of data for analytics
Stores data in a relational database	Stores data in multiple dimensions
Used to perform basic database transactions	Used for multi-dimensional data analytics, data mining, and business intelligence
Built to record real-time business operations	Built to analyze business measures, organized in multiple dimensions
Performance is measured by the number of transactions completed per second	Performance is measured by query response times
Consists primarily of current operational data	Consists of historical data from multiple databases
Data tables are normalized	Data tables are not normalized
Requires regular backups and a robust recovery process	Does not require regular backups



### Data Marts

A data mart is like a mini data warehouse focused on one department, like Sales or HR. We discussed three types:

- **Dependent:** Gets data from the main warehouse.
- **Independent:** Gets data directly from sources.
- **Hybrid:** Mix of both.

Benefits:

- Easier to manage.
- Specific to user needs.
- Faster performance.

### Data Marts vs Data Warehouses

Feature	Data Mart	Data Warehouse
Scope	Department-level	Whole organization
Size	Smaller	Larger
Data Source	One/few sources	Multiple systems
Setup Time	Faster	Longer

## Data Warehouse Life Cycle

Here's the basic flow we noted down:

1. **Requirements:** Talk to users and understand what data is needed.
2. **Design:** Plan how data will be stored (schemas).
3. **ETL:** Extract, clean, and load data into the warehouse.
4. **Build:** Set up the warehouse and tools.
5. **Testing:** Make sure the system works properly.
6. **Deployment:** Launch the system for use.
7. **Maintenance:** Keep it updated and fix any issues.

This life cycle helps us understand how a data warehouse is created and maintained over time.

