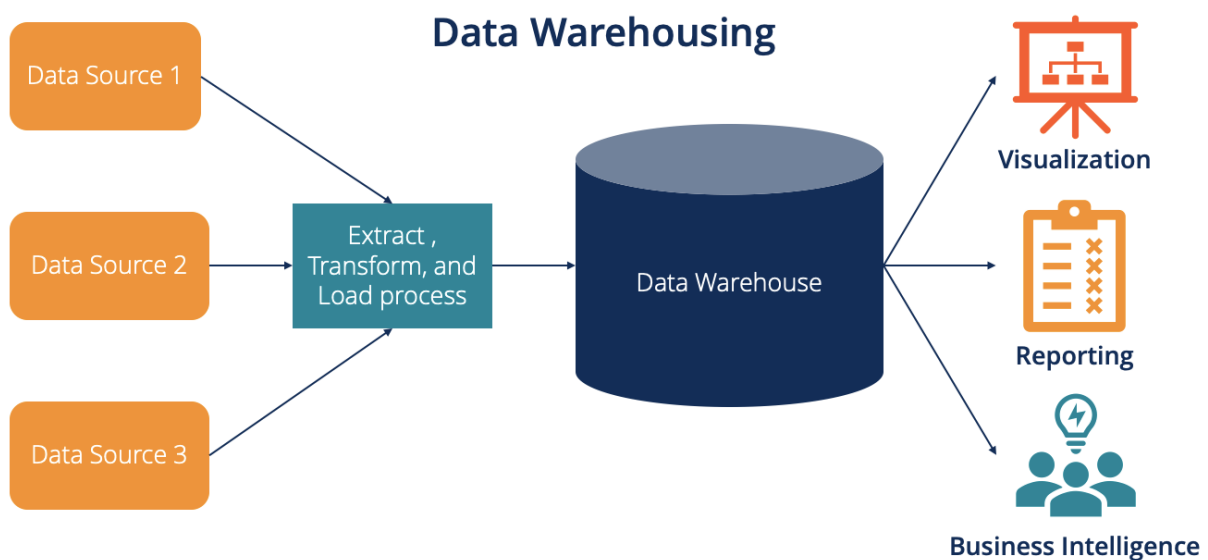


## Assignments on Data warehouse

### An Introduction to Data Warehousing

Data Warehousing is a system used for reporting and data analysis, and is considered a core component of business intelligence. It is a central repository of integrated data from one or more disparate sources. Data warehouses store historical data and are optimized for read access and analytical queries.



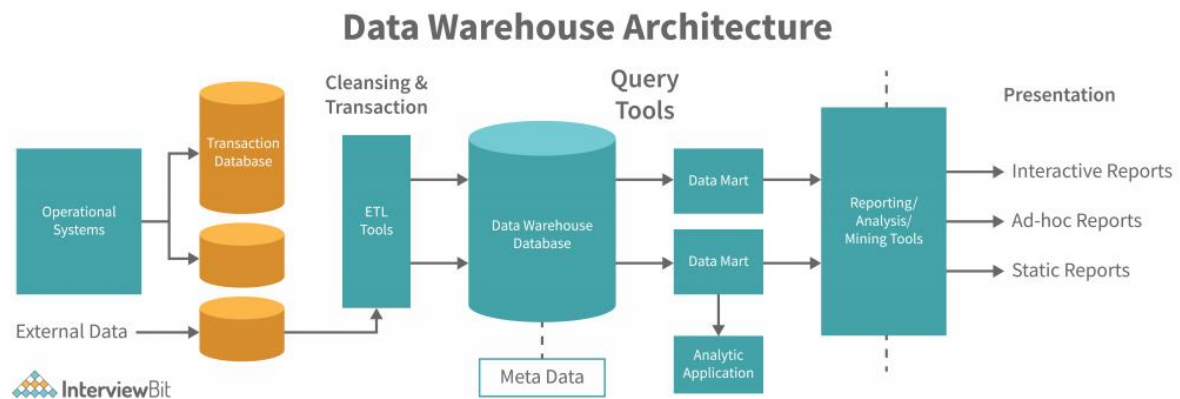
### Purpose of Data Warehouse



1. **Centralized Data Storage:** Combines data from various sources into one central location.
2. **Historical Analysis:** Stores historical data for trend analysis and forecasting.
3. **Improved Business Intelligence:** Provides insights for decision-making and strategic planning.
4. **Performance:** Optimized for complex queries and data analysis, not everyday transactions.
5. **Data Quality and Consistency:** Enforces standards, consistency, and data integrity.

### **Data Warehouse Architecture**

1. **Data Sources:** Operational databases, external sources, etc.
2. **ETL Process (Extract, Transform, Load):**
  - **Extract:** Pull data from various sources.
  - **Transform:** Convert into the right format.
  - **Load:** Push data into the warehouse.
3. **Data Storage Area:**
  - Staging Area
  - Data Warehouse Database
  - Data Marts
4. **Presentation Layer:**
  - Reporting tools, dashboards, OLAP tools.
5. **Metadata:** Information about data sources, transformation rules, etc.



## Operational Data Store (ODS)

An ODS is an intermediate storage area used for operational reporting. Unlike a data warehouse, which stores historical data, ODS contains current or near-real-time data and is used for routine operational tasks.

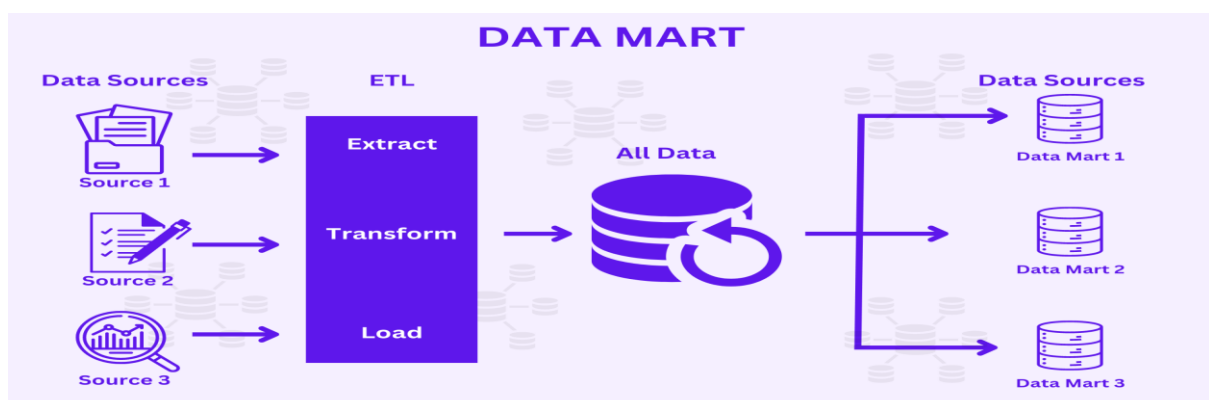


## OLTP vs Warehouse Applications

Feature	OLTP (Online Transaction Processing)	Data Warehouse (OLAP)
Purpose	Daily operations	Analysis and reporting
Data Type	Current, up-to-date	Historical, aggregated
Normalization	Highly normalized	De-normalized
Query Type	Short, fast inserts and updates	Complex queries and analytics
Users	Clerks, DBAs	Analysts, Managers
Performance Focus	High transaction throughput	High query performance

## Data Marts

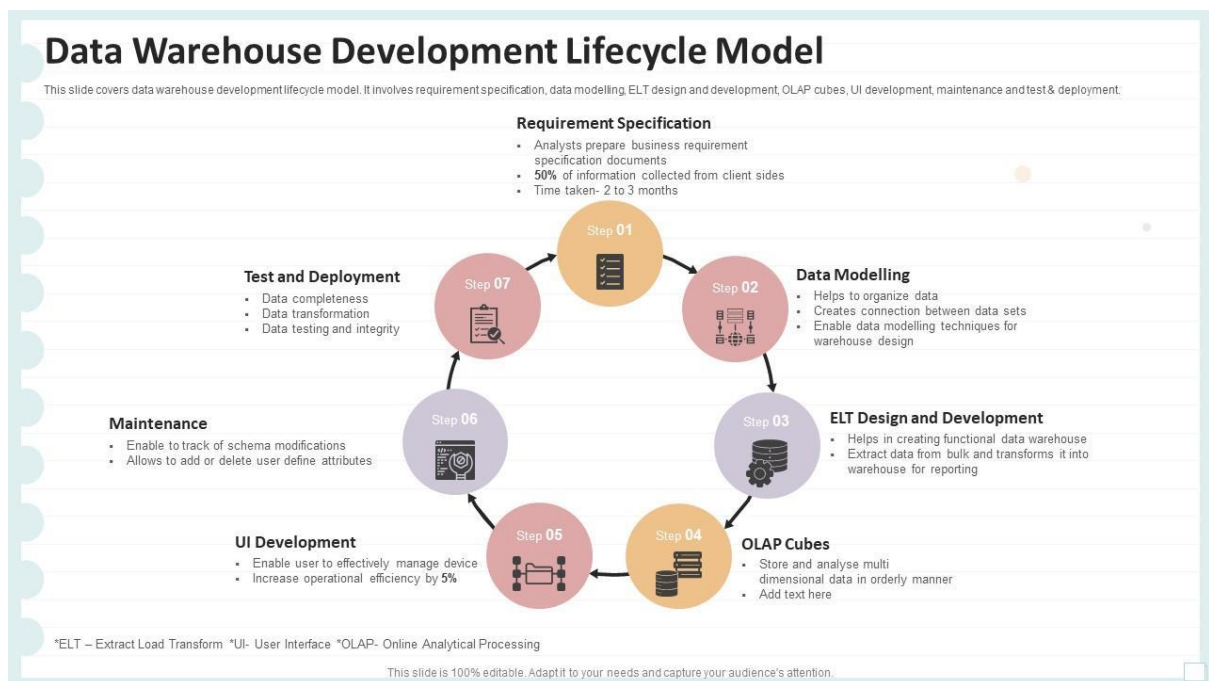
A data mart is a subset of a data warehouse focused on a specific business area such as sales, finance, or marketing. It is designed for a specific group of users and can be independent or dependent (sourced from the main data warehouse).



## Data Marts vs Data Warehouses

Feature	Data Mart	Data Warehouse
Scope	Specific department	Entire organization
Size	Smaller	Larger
Data Source	One or few sources	Multiple, enterprise-wide
Development Time	Shorter	Longer
Cost	Lower	Higher

## Data Warehouse Life Cycle



- Requirement Analysis:** Understand business goals and data needs.
- Data Modeling:** Design conceptual, logical, and physical models.
- ETL Design and Development:** Set up ETL tools for data flow.

4. **Data Warehouse Deployment:** Set up storage, infrastructure, and deploy.
5. **Testing and Validation:** Ensure data integrity, performance, and usability.
6. **Maintenance and Support:** Regular updates, error fixing, performance tuning.
7. **Evolution and Scaling:** Improve architecture, add new sources, and enhance features.