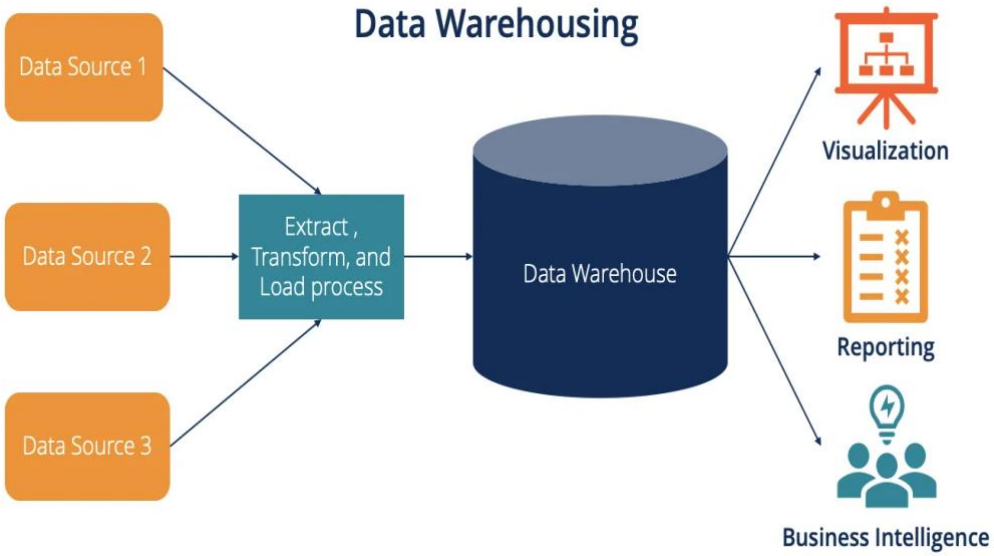
# DATA WAREHOUSE

Date: 17.07.2025

## Introduction to Data Warehousing

A Data Warehouse is a specialized system designed to support data analysis and reporting. It plays a vital role in business intelligence, acting as a centralized platform that collects and integrates data from multiple unrelated sources. The main focus of a data warehouse is to store historical data and enable efficient analytical processing, rather than handling day-to-day transactions.

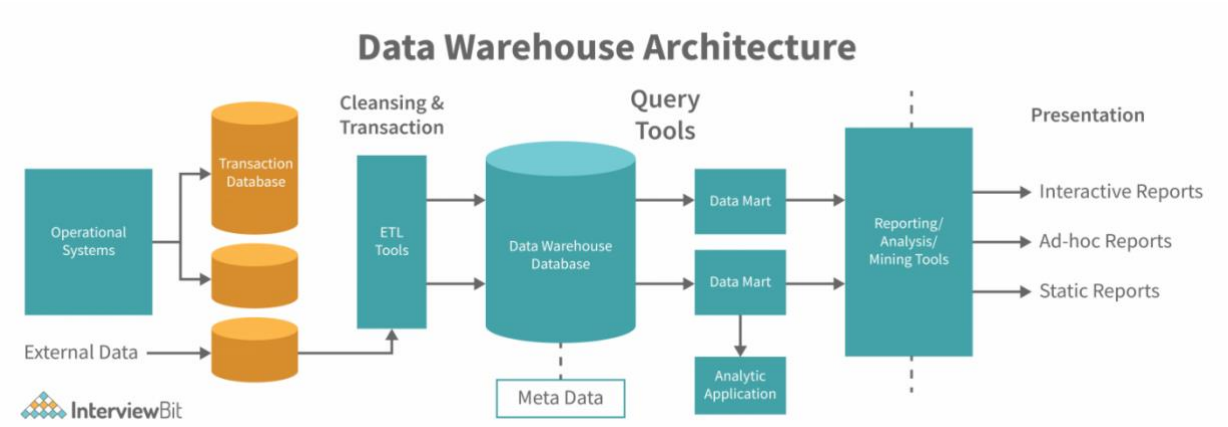



## Why Data Warehousing is Important

1. Unified Data Repository: Brings together information from various databases into a single, consistent source.
2. Historical Tracking: Maintains past records, useful for trends and predictive analysis.
3. Better Decision-Making: Supports strategic planning by providing accurate analytical reports.
4. Optimized for Analysis: Built specifically to manage complex data queries efficiently.
5. Data Accuracy and Consistency: Applies rules to ensure standard formats and reliable data.

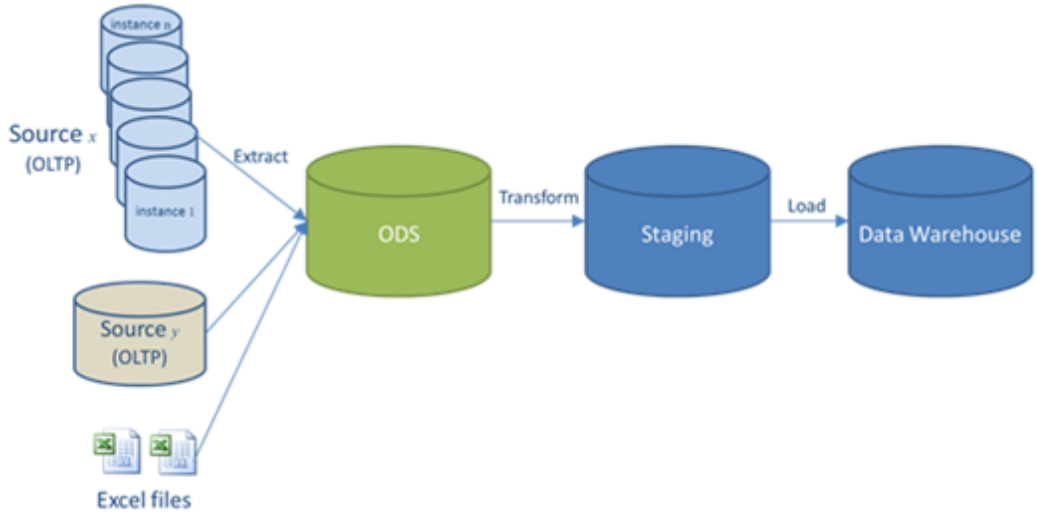
## Architecture of a Data Warehouse

1. Data Input Sources: Includes databases from operations, third-party tools, and external systems.
2. ETL Pipeline (Extract, Transform, Load):
    - Extract: Gathers raw data from different sources.
    - Transform: Converts and cleanses data into a standard structure.
    - Load: Sends the refined data into the warehouse.
3. Data Storage Structure:
    - Staging Area: Temporary space for data before transformation.
    - Warehouse Database: Core storage where final data resides.
    - Data Marts: Smaller, topic-specific segments for departments.
4. Access & Presentation:
    - Tools like dashboards, OLAP systems, and reporting software display results to users.
5. Metadata Layer:
    - Stores details about source systems, ETL rules, data structures, etc.




## What is an ODS (Operational Data Store)?

An Operational Data Store is a storage system used for real-time reporting and operations. Unlike a data warehouse that contains long-term historical data, the ODS manages current, updated information for immediate decision-making.




## Difference Between OLTP and Data Warehousing (OLAP)

| Feature | OLTP (Online Transaction Processing) | Data Warehouse (OLAP) |
|---|---|---|
| Purpose | Routine business processes | Analytical reporting |
| Data Type | Real-time, current | Past, summarized |
| Structure | Highly normalized | De-normalized for fast access |
| Query Style | Fast inserts/updates | Complex analytical queries |
| Users | Operators, Admins | Analysts, Executives |
| System Focus | Speed in processing transactions | Speed in retrieving insights |

## Understanding Data Marts

A Data Mart is a focused section of a data warehouse built for a particular function like sales, marketing, or finance. It helps teams access relevant data quickly. Data marts may operate on their own or as part of the larger warehouse system.

## Data Mart vs Data Warehouse

| Attribute | Data Mart | Data Warehouse |
|---|---|---|
| Focus Area | Specific business unit | Enterprise-wide |
| Size | Smaller | Larger |
| Data Sources | Limited, usually one | Broad and varied |
| Development Time | Faster to implement | Requires more time |
| Cost | Less expensive | More costly due to scale |

## Life Cycle of a Data Warehouse

1. Requirement Gathering: Identify business goals and what data is needed.
2. Modeling the Data: Design data models—conceptual to physical.
3. Building ETL Process: Develop the data flow and transformation setup.
4. Deployment: Install infrastructure and launch the system.
5. Testing: Verify accuracy, speed, and functionality.
6. Ongoing Maintenance: Fix issues, apply updates, and monitor performance.
7. Expansion: Adapt to future needs, add more sources, and scale up features.