

A **data warehouse** is a centralized repository that stores large amounts of data from multiple sources in a structured and organized way. It is designed to support business intelligence (BI) activities, such as data analysis, reporting, and decision-making.

system that reflects the condition of the **warehouse in real time** is referred to as realtime data warehousing. If you perform a query on the real-time data warehouse to learn more about a specific aspect of the company or entity described by the warehouse,

The goal of **dimensional modelling** is to optimise the database so that data can be retrieved more quickly. In a data warehouse, a dimensional model is used to read, summarise, and analyse numeric data such as values, balances, counts, weights, and so on

**Dimensional modeling** is a data modeling technique used to design data warehouses. It organizes data into a structure that makes it easier to retrieve and analyze for reporting and business intelligence purposes.

In simple terms, a **data cube** in OLAP is a structured way to organize data that makes it easy to analyze from different angles

**Data cube** is essentially a multidimensional array, where each dimension represents a different attribute or feature of the data, and the cells within the cube contain aggregated values (such as sums, averages, or counts).

**Data modeling** is the process of creating a visual representation or blueprint of a complex data system. It defines how data is structured, organized, and related to each other, facilitating communication between data architects, developers, business analysts, and stakeholders.

**Data Modeling:** Design conceptual, logical, and physical data models

The advantage of using **Oracle** is its strong performance, scalability, and advanced features for managing large-scale databases, making it ideal for complex and mission-critical applications.

A **Data Lake** is a large-scale storage repository for structured, semi structured, and unstructured data. It's a location where you can save any type of data in its original format, with no restrictions on account size or file size

Yes, a fact table can contain NULL values, but it is generally not recommended or common practice in data warehousing. NULL values may exist in a fact table under certain circumstances, but they are often avoided to ensure data quality and consistency

A **dimension table** is a table in a data warehouse that contains descriptive, textual, or categorical data related to the facts (measurable data) stored in the fact table

A **Fact Table** is a central table in a star schema or snowflake schema of a data warehouse. It stores quantitative data (measurable facts) for analysis and contains foreign keys to dimension tables.

### **Factless Fact Table:**

A Factless Fact Table is a type of fact table that does not have any measurable facts or numeric data. Instead, it is used to capture the occurrence of events or describe many-to-many relationships between dimension tables.

### **1. ODS (Operational Data Store):**

- It's a type of database that stores current and recent data from different systems.
- Used for quick and simple reports based on up-to-date data.
- Helps in making quick operational decisions.

### **2. OIS (Operational Information Store):**

- Similar to ODS but may have more processed and enriched data.
- Focuses on making data more useful for complex reports and fast access.

So, **ODS** gives you fresh, detailed data for quick access, while **OIS** is often more refined for better reporting and analysis.

---

**Triggers** are special types of stored procedures in a database that automatically run (or "fire") when certain events occur on a table or view. Triggers can respond to data changes like **INSERT, UPDATE, or DELETE** operations, and they are often used to enforce business rules, validate data, or maintain audit trails.

```
CREATE TRIGGER trg_AfterUserUpdate  
ON Users  
AFTER UPDATE  
AS
```

```
BEGIN
    INSERT INTO UserAudit (UserID, ChangeTime, OldName, NewName)
    SELECT
        i.UserID,
        GETDATE(),
        d.Name, -- Old value
        i.Name  -- New value
    FROM
        inserted i
    INNER JOIN
        deleted d ON i.UserID = d.UserID;
END;
```

---

```
WITH RankedAmounts AS (
    SELECT
        CustomerID,
        version,
        amount,
        ROW_NUMBER() OVER (PARTITION BY CustomerID ORDER BY version DESC) AS
        RowRank
    FROM YourTable
)
SELECT
    CustomerID,
    version,
    amount
FROM RankedAmounts
WHERE RowRank = 1;
```

- **Data Definition Language DDL** manages the structure and schema of the database.

**CREATE , ALTER ,DROP ,TRUNCATE ,Rename**

- **Data Manipulation Language DML** manipulates the data within the tables.

**SELECT , INSERT , UPDATE , DELETE**

- **Data Control Language DCL** controls access and permissions for users interacting with the database.

**GRANT, REVOKE**

**TCL (Transaction Control Language)**

**COMMIT, TRANSACTION , ROLLBACK , SAVEPOINT**

- 
- **TRANSACTION:** While SQL does not have a specific "TRANSACTION" command, a transaction refers to the sequence of operations (such as INSERT, UPDATE, DELETE) that are treated as a single unit. It ensures that all the operations in the transaction are executed successfully, or if any operation fails, the transaction can be rolled back.

- **ROLLBACK:** This is a **Transaction Control** command. It is used to undo the changes made by the current transaction, reverting the database to its previous state before the transaction started.
- 

## **star schema**

When to Use:

When the primary goal is to optimize read operations and improve query performance.

Suitable for data marts where speed and simplicity are more critical than storage efficiency

## **Snowflake Schema**

When storage optimization is more important, and there is a need for better data integrity and normalization. o Suitable for large and complex data warehouses where data storage efficiency is crucial

**SSIS (SQL Server Integration Services)** is a data integration and workflow tool used to perform data migration and transformation tasks. It is part of Microsoft's SQL Server and is used to move data between different sources and destinations, transform data into a usable format, and automate data-related workflows.

In simple terms,

**SSIS (SQL Server Integration Services)** has four main parts:

- 1. Control Flow:** This is the brain of SSIS, deciding what tasks to run and in what order (like a to-do list).
- 2. Data Flow:** This is where the real work happens—it moves data from one place to another, cleaning and changing it along the way.
- 3. Event Handlers:** These are like safety nets; they catch problems or special events and let you take action when something goes wrong.
- 4. Package Explorer:** Think of it as a map; it shows everything in your SSIS package and how it's all connected

**In SQL Server Integration Services (SSIS), containers** are logical groupings that help manage and execute tasks in an SSIS package. They provide structure to the SSIS package and allow for better control over the flow of execution, especially in complex ETL (Extract, Transform, Load) processes. Types of Containers in SSIS

- 1. Sequence Container**
- 2. For Loop Container**
- 3. Foreach Loop Container**

#### 4. Task Host Container

- **VARCHAR2** optimizes storage by using only as much space as each text entry needs.
- For example, 'John' takes up only 4 characters in storage within the `first_name` column, even though it has a max length of 50 characters.
- This makes VARCHAR2 an efficient choice for columns where text lengths vary widely, avoiding wasted space.

Steps to Create a Data Warehouse:

- 1. Requirements Gathering:** Define business needs, data sources, and reporting requirements.
- 2. Data Source Analysis:** Identify and analyze data sources for structure and quality.
- 3. Data Modeling:** Design conceptual, logical, and physical data models.
- 4. ETL Design:** Plan the extraction, transformation, and loading of data.
- 5. ETL Implementation:** Develop, test, and deploy ETL processes.
- 6. Deployment:** Deploy the data warehouse schema and ETL processes.
- 7. Reporting:** Create a data access layer and develop reports and dashboards.
- 8. Optimization:** Optimize performance and monitor system health.
- 9. Maintenance:** Perform ongoing maintenance, monitoring, and updates.