1. **What is difference between Data Warehouse and DataMart?**

# Both **Data Warehouse** and **Data Mart** are used for store the data.

**Data Warehouse** is the type of database which is **data-oriented** in nature. while, **Data Mart** is the type of database which is the **project-oriented** in nature.

Data warehouse is large in scope whereas Data mart is limited in scope.

| **Sl.No** |  | **Data Warehouse** | **Data Mart** |
| --- | --- | --- | --- |
| 1. |  | Data warehouse is a Centralised system. | While it is a decentralised system. |
| 2. |  | In data warehouse, lightly denormalization takes place. | While in Data mart, highly denormalization takes place. |
| 3. |  | Data warehouse is top-down model. | While it is a bottom-up model. |
| 4. |  | To built a warehouse is difficult. | While to build a mart is easy. |
| 5. |  | In data warehouse, Fact constellation schema is used. | While in this, Star schema and snowflake schema are used. |
| 6. |  | Data Warehouse is flexible. | While it is not flexible. |
| 7. |  | Data Warehouse is the data-oriented in nature. | While it is the project-oriented in nature. |
| 8. |  | Data Ware house has long life. | While data-mart has short life than warehouse. |
| 9. |  | In Data Warehouse, Data are contained in detail form. | While in this, data are contained in summarized form. |
| 10. |  | Data Warehouse is vast in size. | While data mart is smaller than warehouse. |
| 11. |  | The Data Warehouse might be somewhere between 100 GB and 1 TB+ in size. | The Size of Data Mart is less than 100 GB. |
| 12. |  | The time it takes to implement a data warehouse might range from months to years. | The Data Mart deployment procedure is time-limited to a few months. |
| 13. |  | It uses a lot of data and has comprehensive operational data. | Operational data are not present in Data Mart. |
| 14. |  | It collects data from various data sources. | It generally stores data from a data warehouse. |
| 15. |  | Long time for processing the data because of large data. | Less time for processing the data because of handling only a small amount of data. |
| 16. |  | Complicated design process of creating schemas and views. | Easy design process of creating schemas and views. |

1. **How many Types of Slowly Changing Dimensions, in detailed? SCD**

**Type 0 – Fixed Dimension**

* No changes allowed, dimension never changes

**Type 1 – No History**

* UPDATE and INSERT. If record is available in dimension table, then update else Insert. No history maintained.

**Type 2 – Row Versioning**

* If any changes happened dimension in source, then updated record will be inserted in to dim table and old record will still be there for row versioning. Row Versioning is maintained by ISACTIVE, effective\_dt and expiry\_dt column.

**Type 3 – Previous Value column**

* Track change to a specific attribute, add a column to show the previous value, which is updated as further changes occur

**Type 4 – History Table**

Show current value in dimension table but track all changes in separate table

**Type 6 – Hybrid SCD**

Utilise techniques from SCD Types 1, 2 and 3 to track change

**Q3. What is Fact Table and Dimension Table?**

**What is a Dimension Table?**

* A dimension table contains dimensions of a fact.
* They are joined to fact table via a foreign key.
* Dimension tables are de-normalized tables.
* The Dimension Attributes are the various columns in a dimension table
* Dimensions offers descriptive characteristics of the facts with the help of their attributes

**Difference between Fact Table and Dimension Table:**

| **S.NO** | **Fact Table** | **Dimension Table** |
| --- | --- | --- |
| 1. | Fact table contains the measuring of the attributes of a dimension table. | Dimension table contains the attributes on that truth table calculates the metric. |
| 2. | In fact table, There is less attributes than dimension table. | While in dimension table, There is more attributes than fact table. |
| 3. | In fact table, There is more records than dimension table. | While in dimension table, There is less records than fact table. |
| 4. | Fact table forms a vertical table. | While dimension table forms a horizontal table. |
| 5. | The attribute format of fact table is in numerical format and text format. | While the attribute format of dimension table is in text format. |
| 6. | It comes after dimension table. | While it comes before fact table. |
| 7. | The number of fact table is less than dimension table in a schema. | While the number of dimensions is more than fact table in a schema. |
| 8. | It is used for analysis purpose and decision making. | While the main task of dimension table is to store the information about a business and its process. |

**Q4. How many types of Dimension Table?**

* Conformed Dimension
* Junk Dim
* Degenerate Dimension

**Q5. What is fact table with example?**

In a data warehouse, a fact table is a table that stores the measurements, metrics, or facts related to a business operation.

It is located at the center of a star or snowflake schema and is surrounded by dimension tables.

A fact table has two types of columns: those that contain the facts and those that serve as [foreign keys](https://www.simplilearn.com/tutorials/sql-tutorial/foreign-key) linking to dimension tables.

The [primary key](https://www.simplilearn.com/tutorials/sql-tutorial/primary-key-in-sql) of a fact table is often a composite key made up of all of the foreign keys in the table.

**Q6. What is Star Schema?**

**Star Schema** in data warehouse, is a schema in which the center of the star can have one fact table and a number of associated dimension tables. It is known as star schema as its structure like a star. The Star Schema data model is the simplest type of Data Warehouse schema.

**Q7. What is Snow Flake Schema?**

A **snowflake schema** is a multi-dimensional data model that is an extension of a star schema, where dimension tables are broken down into sub-dimensions. The dimension tables are normalized which splits data into additional tables.

**Q8. What is SCD 0, SCD 1, SCD 2, SCD 3 and SCD 6?**

Type SCD-0 ()

There are situations where you ignore any changes.

For example, when an employee joined an organization,

there are joined related attributes such as joined Designation and Joined Date, etc. that should not change over time.

**Type SCD-1 ()**

The first type of SCD action possible is overwriting.

Here, dimension values are overwritten by new values.

For example, if our customer Mike moves from Delhi to Mumbai,

then his city will be updated with the latest value, i.e Mumbai.

**Type 2 SCD (Row versioning)**

The second type of SCD action is row versioning.

In this type 2 SCD, when values for a current record change,

the current record is marked as closed and a new record gets inserted.

Then, there will be 2 records associated with Mike in the updated table,

but only the latest version will be marked “open”.

Type 3 SCD (Adding previous value column)

The third common SCD type is adding a previous value column.

Here, the previous and current versions are maintained in a single row.

In the customer MIke moving address example we would copy Mike’s ‘Current City’ record into

the ‘Previous City’ and overwrite the Current City record with the new city.

Type 4 SCD

In Type 4, the dimension table has the latest value while its history is maintained in a separate table.

Type 6 SCD

In Type 6, a combination of Type 1, 2 & 3 are used to track changes in dimension.

Generally, Type 6 is adopted in scenarios where multiple parts of a record are slowly changing dimensions,

but using multiple implementations of a single type could lead to issues with rapid inflation of table size.

**Q9. What is OLTP and OLAP?**

**OLTP (Online Transaction Processing):-**

OLTP stands for Online Transaction Processing.

OLTP Database provides day-to-day transaction-oriented applications of an organization.

The main goal of OLTP is data processing not data analysis.

Other examples of OLTP applications are:

* Online banking
* Online airline ticket booking
* Sending a text message
* Order entry
* Add a book to shopping cart

**OLAP (Online Analytical Processing):-**

OLAP stands for Online Analytical Processing.

OLAP systems have the capability to analyse database information of multiple systems at the current time.

The primary goal of OLAP Service is data analysis and not data processing.

|  |  |  |
| --- | --- | --- |
| **Category** | **OLAP (Online Analytical Processing)** | **OLTP (Online Transaction Processing)** |
| Definition | It is well-known as an online database query management system. | It is well-known as an online database modifying system. |
| Data source | Consists of historical data from various Databases. | Consists of only operational current data. |
| Method used | It makes use of a data warehouse. | [It makes use of a standard database management system (DBMS).](https://www.geeksforgeeks.org/introduction-of-dbms-database-management-system-set-1/) |
| Application | [It is subject-oriented. Used for Data Mining, Analytics, Decisions making, etc.](https://www.geeksforgeeks.org/data-mining/) | It is application-oriented. Used for business tasks. |
| Normalized | In an OLAP database, tables are not normalized. | [In an OLTP database, tables are normalized (3NF).](https://www.geeksforgeeks.org/third-normal-form-3nf/) |
| Usage of data | The data is used in planning, problem-solving, and decision-making. | The data is used to perform day-to-day fundamental operations. |
| Task | It provides a multi-dimensional view of different business tasks. | It reveals a snapshot of present business tasks. |
| Purpose | It serves the purpose to extract information for analysis and decision-making. | It serves the purpose to Insert, Update, and Delete information from the database. |
| Volume of data | A large amount of data is stored typically in TB, PB | The size of the data is relatively small as the historical data is archived in MB, and GB. |
| Queries | Relatively slow as the amount of data involved is large. Queries may take hours. | Very Fast as the queries operate on 5% of the data. |
| Update | The OLAP database is not often updated. As a result, data integrity is unaffected. | The data integrity constraint must be maintained in an OLTP database. |
| Backup and Recovery | It only needs backup from time to time as compared to OLTP. | The backup and recovery process is maintained rigorously |
| Processing time | The processing of complex queries can take a lengthy time. | It is comparatively fast in processing because of simple and straightforward queries. |
| Types of users | This data is generally managed by CEO, MD, and GM. | This data is managed by clerksForex and managers. |
| Nature of audience | The process is focused on the customer. | The process is focused on the market. |
| Database Design | Design with a focus on the subject. | Design that is focused on the application. |
| Productivity | Improves the efficiency of business analysts. | Enhances the user’s productivity. |

**Q10. How can we load time Dimension?**

**Q11. What is conformed fact?**

**Q12. Whatever Dimension Table can have numeric value?**

**Q13. What is Surrogate Key?**

**Q14. Steps to build dataware house?**

**Q15. What is ETL and ELT?**

**Extraction, Transform and Load (ETL):**   
ETL is the traditional technique of extracting raw data, transforming it for the users as required and storing it in data warehouses. ELT was later developed, having ETL as its base. The three operations happening in ETL and ELT are the same except that their order of processing is slightly varied. This change in sequence was made to overcome some drawbacks.

1. **Extract:**   
   It is the process of extracting raw data from all available data sources such as databases, files, ERP, CRM or any other.
2. **Transform:**   
   The extracted data is immediately transformed as required by the user.
3. **Load:**   
   The transformed data is then loaded into the data warehouse from where the users can access it.

**Extraction, Load and Transform (ELT):** Extraction, Load and Transform (ELT) is the technique of extracting raw data from the source and storing it in data warehouse of the target server and preparing it for end stream users.

ELT comprises of 3 different operations performed on the data:

1. **Extract:**   
   Extracting data is the technique of identifying data from one or more sources. The sources may be databases, files, ERP, CRM or any other useful source of data.
2. **Load:**   
   Loading is the process of storing the extracted raw data in data warehouse or data lakes.
3. **Transform:**   
   Data transformation is the process in which the raw data source is transformed to the target format required for analysis.

**Q.16 What is Metadata?**

Metadata is data about the data or documentation about the information which is required by the users. In data warehousing, metadata is one of the essential aspects.

Metadata includes the following:

1. The location and descriptions of warehouse systems and components.
2. Names, definitions, structures, and content of data-warehouse and end-users views.
3. Identification of authoritative data sources.
4. Integration and transformation rules used to populate data.
5. Integration and transformation rules used to deliver information to end-user analytical tools.
6. Subscription information for information delivery to analysis subscribers.
7. Metrics used to analyze warehouses usage and performance.
8. Security authorizations, access control list, etc.

Metadata is used for building, maintaining, managing, and using the data warehouses. Metadata allow users access to help understand the content and find data.

### Benefits of OLTP Services

* OLTP services allow users to read, write and delete data operations quickly.
* OLTP services help in increasing users and transactions which helps in real-time access to data.
* OLTP services help to provide better security by applying multiple security features.
* OLTP services help in making better decision making by providing accurate data or current data.
* OLTP Services provide Data Integrity, Consistency, and High Availability to the data.

### Drawbacks of OLTP Services

* OLTP has limited analysis capability as they are not capable of intending complex analysis or reporting.
* OLTP has high maintenance costs because of frequent maintenance, backups, and recovery.
* OLTP Services get hampered in the case whenever there is a hardware failure which leads to the failure of online transactions.
* OLTP Services many times experience issues such as duplicate or inconsistent data.

## what is fact table and dimension table?

**What is a Dimension Table?**

* A dimension table contains dimensions of a fact.
* They are joined to fact table via a foreign key.
* Dimension tables are de-normalized tables.
* The Dimension Attributes are the various columns in a dimension table
* Dimensions offers descriptive characteristics of the facts with the help of their attributes
* No set limit set for given for number of dimensions
* The dimension can also contain one or more hierarchical relationships

## What is the difference between ETL and ELT?

15.

## ****Metadata**** is simply defined as data about data. It means it is a description and context of the data. It helps to organize, find and understand data