1. What is OLTP and OLAP?

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

* Online Transaction Processing refers to a type of system designed for managing and processing large volumes of online transactions in real-time.
* It focuses on transactional data processing.
* These systems emphasize rapid data entry, retrieval, and modification, catering to concurrent users and ensuring data integrity.
* Examples Of OLTP are retail sales systems, banking transactions, and online booking platforms.

**OLAP (Online Analytical Processing):**

* Online Analytical Processing, is a category of software tools and technologies used for analysing multidimensional data from various perspectives.
* OLAP deals with querying, reporting, and analysing large datasets to gain insights into business trends, patterns, and relationships.
* In OLAP systems, data is typically stored in a multidimensional database, often called a data cube or OLAP cube.
* OLAP systems are commonly used in business intelligence (BI) applications for tasks such as financial reporting, sales analysis, and trend forecasting.

1. Differences between OLTP and OLAP.

Ans:

|  |  |  |
| --- | --- | --- |
| Aspect | OLTP | OLAP |
| Purpose | Manage transactional data for day-to-day operations | Analyze historical data for decision-making |
| Type of Processing | Transaction-oriented, real-time processing | Analytical processing, typically batch or periodic |
| Database Structure | Typically relational database structure | Multidimensional or dimensional model |
| Usage | Supports frequent, short transactions | Supports complex, ad-hoc queries and analysis |
| Data Model Focus | Normalized data model | Denormalized or star/snowflake schema data model |
| Query Complexity | Simple queries for data retrieval and manipulation | Complex queries involving aggregation and analysis |
| Data Volume | Handles relatively low to moderate data volumes | Handles large volumes of historical data |
| Concurrency | Supports high concurrency with short transactions | Handles fewer concurrent users but complex queries |
| Performance | Optimized for transaction speed and throughput | Optimized for query performance and analysis |
| Examples | Point-of-sale systems, banking transactions | Business intelligence, data warehousing systems |

1. **Database Normal Forms (5 Normal Forms)**

**Ans:**

The database normalization process involves organizing the attributes and tables of a relational database to minimize redundancy and dependency. There are five normal forms (1NF, 2NF, 3NF, BCNF, and 4NF) that represent increasing levels of normalization. Here's a brief overview of each:

1. First Normal Form (1NF):

* Eliminates repeating groups within a table.
* Ensures that each column contains atomic (indivisible) values.
* Ensures each row is unique.

2. Second Normal Form (2NF):

* Must be in 1NF.
* Eliminates partial dependencies.
* Ensures that no non-prime attribute is dependent on only a portion of the primary key.

3. Third Normal Form (3NF):

* Must be in 2NF.
* Eliminates transitive dependencies.
* Ensures that no non-prime attribute is dependent on another non-prime attribute.

4. Boyce-Codd Normal Form (BCNF):

* A stronger version of 3NF.
* Ensures that every determinant is a candidate key.
* Every non-trivial functional dependency is a dependency on a super key.

5. Fourth Normal Form (4NF):

* + Must be in BCNF.
  + Addresses multi-valued dependencies.
  + Ensures that every multi-valued dependency is a trivial dependency.

These normal forms help database designers to create databases that are efficient, scalable, and maintainable by minimizing redundancy and ensuring data integrity.

1. Dimension vs Fact Table and Types of Dimensions.

Ans:

**Dimension Table:**

**Definition**: Dimension tables contain descriptive attributes that provide context and details about the data in a fact table.

**Characteristics:**

* Typically smaller in size compared to fact tables.
* Usually consists of textual data, such as names, descriptions, categories, etc.
* Used to filter, group, and aggregate data in fact tables.

**Example Attributes:** Product name, customer details, location, time, etc.

**Fact Table:**

**Definition:** Fact tables contain numerical measures or metrics, along with keys to dimension tables, providing the quantitative information for analysis.

**Characteristics:**

* Typically larger in size compared to dimension tables.
* Contains numerical data representing business facts or events, such as sales amounts, quantities, etc.
* Each row represents a specific event or transaction.

**Example Measures:** Sales amount, quantity sold, profit, revenue, etc.

**Types Of Dimensions:**

1. Slowly Changing Dimension
2. Conformed Dimension
3. Degenerate Dimension
4. Junk Dimension
5. Role- Playing Dimension
6. Static Dimension
7. Shrunken Dimension.
8. **Star Schema VS Snowflake Schema.**

**Ans:**

|  |  |  |
| --- | --- | --- |
| **Slno** | **Star Schema** | **Snowflake Schema** |
| 1 | In Star Schema, The fact tables and the dimension tables are contained. | In Snowflake Schema, The fact tables, dimension tables as well as sub dimension tables are contained. |
| 2 | It is a top-down model. | It is a bottom-up model |
| 3 | Star schema uses more space. | Snowflake schema uses less space |
| 4 | It takes less time for the execution of queries. | It takes more time than star schema for the execution of queries. |
| 5 | In star schema, Normalization is not used. | Both normalization and denormalization are used. |
| 6 | It’s design is very simple. | It’s design is complex. |
| 7 | The query complexity of star schema is low. | The query complexity of snowflake schema is higher than star schema. |
| 8 | It’s understanding is very simple. | It’s understanding is difficult |
| 9 | It has less number of foreign keys. | It has more number of foreign keys. |
| 10 | It has high data redundancy. | It has low data redundancy. |