UNIT-2

1.) Explain various concepts of Data Warehouse.

Ans:

Data Warehouse:

A decision support database that is maintained separately from the organization's operational database

Support information processing by providing a solid platform of consolidated, historical data for analysis.

Data warehousing provides architectures and tools for business executives to systematically organize, understand, and use their data to make strategic decisions.

Data warehouse systems are valuable tools in today's competitive, fast-evolving world. In the last several years, many firms have spent millions of dollars in building enterprise-wide data warehouses.

Data Warehouse subject-oriented:

Focusing on the modeling and analysis of data for decision makers, not on daily operations or transaction processing.

Provide a simple and concise view around particular subject issues by excluding data that are not useful in the decision support process.

Data Warehouse integrated:

Constructed by integrating multiple, heterogeneous data sources relational databases, flat files, on-line transaction records Data cleaning and data integration techniques are applied.

- Ensure consistency in naming conventions, encoding structures, attribute measures, etc. among different data sources
- When data is moved to the warehouse, it is converted.

Data Warehouse time-variant:

The time horizon for the data warehouse is significantly longer than that of operational systems

- Operational database: current value data
- Data warehouse data: provide information from a historical perspective (e.g., past 5-10 years) Every key structure in the data warehouse
- Contains an element of time, explicitly or implicitly

- But the key of operational data may or may not contain “time element”.

Data Warehouse non-volatile:

- A physically separate store of data transformed from the operational environment
- Operational update of data does not occur in the data warehouse environment
- Does not require transaction processing, recovery, and concurrency control mechanisms
- Requires only two operations in data accessing:
- initial loading of data and access of data.

2.) What are the needs and goals of data warehouse.

Ans :

Goals of Data Warehousing :

- **To help reporting as well as analysis**
- **Maintain the organization's historical information**
- **Be the foundation for decision making.**

Need for Data Warehouse

Data Warehouse is needed for the following reasons:

1. **Business User: Business users require a data warehouse to view summarized data from the past. Since these people are non-technical, the data may be presented to them in an elementary form.**
2. **Store historical data: Data Warehouse is required to store the time variable data from the past. This input is made to be used for various purposes.**
3. **Make strategic decisions: Some strategies may be depending upon the data in the data warehouse. So, data warehouse contributes to making strategic decisions.**
4. **For data consistency and quality: Bringing the data from different sources at a commonplace, the user can effectively undertake to bring the uniformity and consistency in data.**
5. **High response time: Data warehouse has to be ready for somewhat unexpected loads and types of queries, which demands a significant degree of flexibility and quick response time.**

Benefits of Data Warehouse

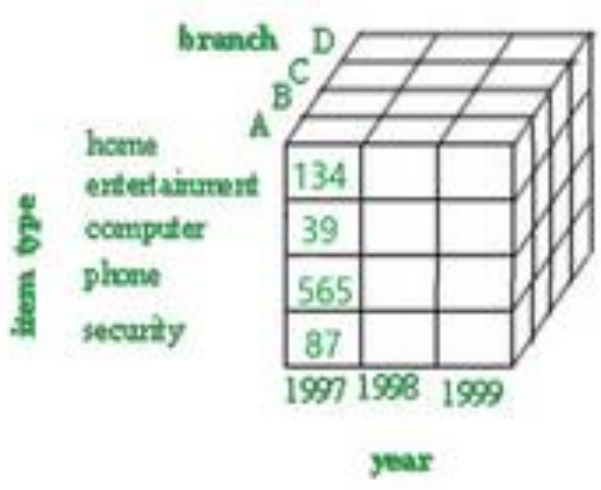
1. **Understand business trends and make better forecasting decisions.**
2. **Data Warehouses are designed to perform well enormous amounts of data.**
3. **The structure of data warehouses is more accessible for end-users to navigate, understand, and query.**
4. **Queries that would be complex in many normalized databases could be easier to build and maintain in data warehouses.**
5. **Data warehousing is an efficient method to manage demand for lots of information from lots of users.**
6. **Data warehousing provide the capabilities to analyze a large amount of historical data.**

3.) What is Data Cube and how to give the representation.

Ans :

Grouping of data in a multidimensional matrix is called data cubes. In Dataware housing, we generally deal with various multidimensional data models as the data will be represented by multiple dimensions and multiple attributes. This multidimensional data is represented in the data cube as the cube represents a high-dimensional space. The Data cube pictorially shows how different

attributes of data are arranged in the data model. Below is the diagram of a general data cube.



The example above is a 3D cube having attributes like branch(A,B,C,D),item type(home,entertainment,computer,phone,security), year(1997,1998,1999) .

Data cube classification:
The data cube can be classified into two categories:

- **Multidimensional data cube: It basically helps in storing large amounts of data by making use of a multi-dimensional array. It increases its efficiency by keeping an index of each dimension. Thus, dimensional is able to retrieve data fast.**
- **Relational data cube: It basically helps in storing large amounts of data by making use of relational tables. Each relational table displays the dimensions of the data cube. It is slower compared to a Multidimensional Data Cube.**

Advantages of data cubes:

- **Multi-dimensional analysis: Data cubes enable multi-dimensional analysis of business data, allowing users to view data from different perspectives and levels of detail.**
- **Interactivity: Data cubes provide interactive access to large amounts of data, allowing users to easily navigate and manipulate the data to support their analysis.**
- **Speed and efficiency: Data cubes are optimized for OLAP analysis, enabling fast and efficient querying and aggregation of data.**
- **Data aggregation: Data cubes support complex calculations and data aggregation, enabling users to quickly and easily summarize large amounts of data.**
- **Improved decision-making: Data cubes provide a clear and comprehensive view of business data, enabling improved decision-making and business intelligence.**

- **Accessibility: Data cubes can be accessed from a variety of devices and platforms, making it easy for users to access and analyze business data from anywhere.**
- **Helps in giving a summarised view of data.**
- **Data cubes store large data in a simple way.**
- **Data cube operation provides quick and better analysis,**
- **Improve performance of data.**

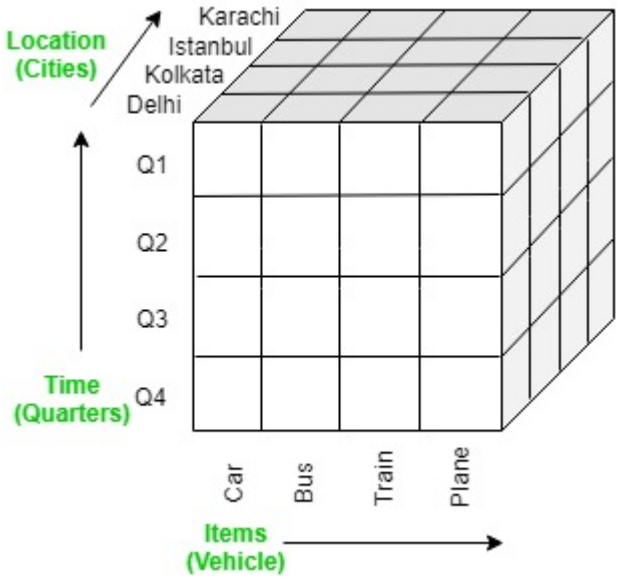
Disadvantages of data cube:

- **Complexity: OLAP systems can be complex to set up and maintain, requiring specialized technical expertise.**
- **Data size limitations: OLAP systems can struggle with very large data sets and may require extensive data aggregation or summarization.**
- **Performance issues: OLAP systems can be slow when dealing with large amounts of data, especially when running complex queries or calculations.**
- **Data integrity: Inconsistent data definitions and data quality issues can affect the accuracy of OLAP analysis.**
- **Cost: OLAP technology can be expensive, especially for enterprise-level solutions, due to the need for specialized hardware and software.**
- **Inflexibility: OLAP systems may not easily accommodate changing business needs and may require significant effort to modify or extend.**

4.) Define OLAP and explain its operations.

Ans :

OLAP stands for *Online Analytical Processing* Server. It is a software technology that allows users to analyze information from multiple database systems at the same time. It is based on multidimensional data model and allows the user to query on multi-dimensional data (eg. Delhi -> 2018 -> Sales data). OLAP databases are divided into one or more cubes and these cubes are known as *Hyper-cubes*.



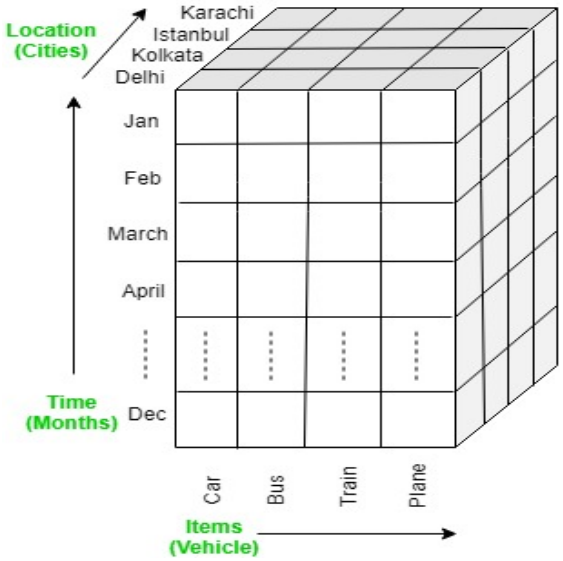
OLAP operations:

There are five basic analytical operations that can be performed on an OLAP cube:

1.)Drill down: In drill-down operation, the less detailed data is converted into highly detailed data. It can be done by:

- **Moving down in the concept hierarchy**
- **Adding a new dimension**

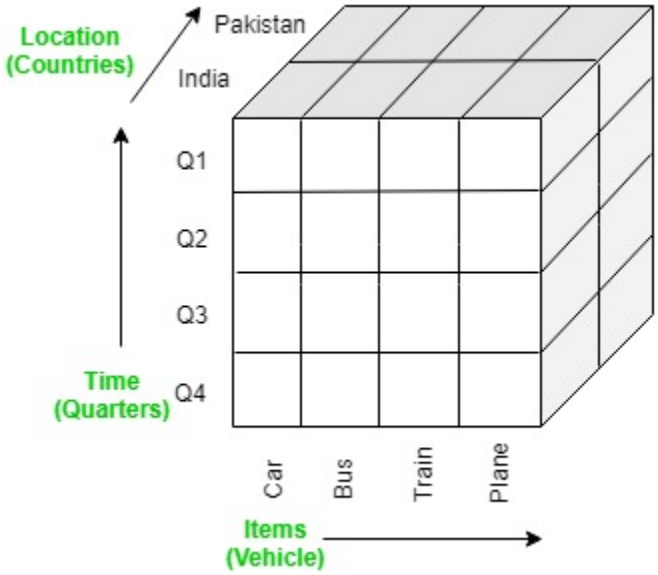
In the cube given in overview section, the drill down operation is performed by moving down in the concept hierarchy of *Time* dimension (Quarter -> Month)



2.) Roll up: It is just opposite of the drill-down operation. It performs aggregation on the OLAP cube. It can be done by:

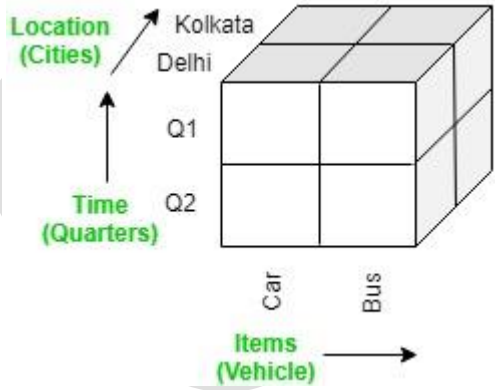
- **Climbing up in the concept hierarchy**
- **Reducing the dimensions**

In the cube given in the overview section, the roll-up operation is performed by climbing up in the concept hierarchy of *Location* dimension (City -> Country).



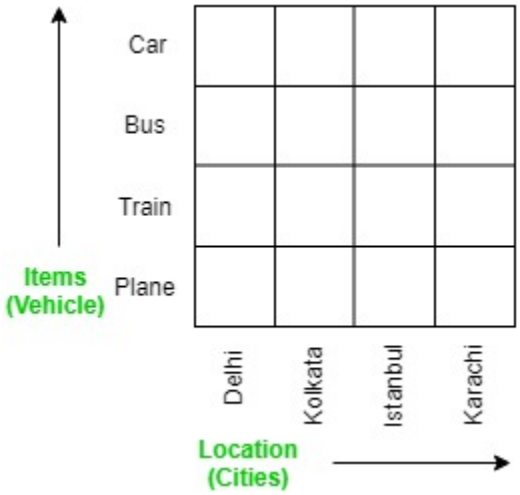
3.) Dice: It selects a sub-cube from the OLAP cube by selecting two or more dimensions. In the cube given in the overview section, a sub-cube is selected by selecting following dimensions with criteria:

- **Location = "Delhi" or "Kolkata"**
- **Time = "Q1" or "Q2"**
- **Item = "Car" or "Bus"**



4.) Slice: It selects a single dimension from the OLAP cube which results in a new sub-cube creation. In the cube given in the overview section, Slice is performed on the dimension Time = "Q1".

5.) Pivot: It is also known as *rotation* operation as it rotates the current view to get a new view of the representation. In the sub-cube obtained after the slice operation, performing pivot operation gives a new view of it.



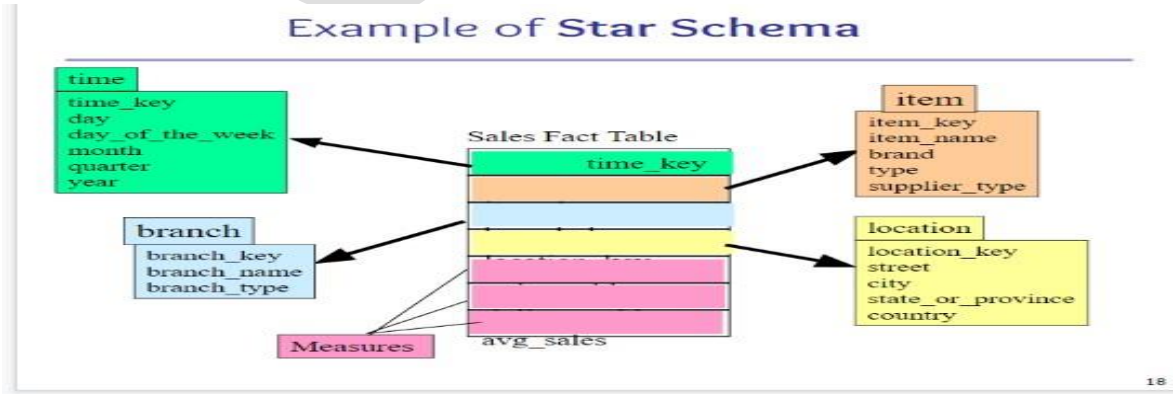
5.) Explain the following.

(i)Star Schema
(ii)Snowflake Schema
(iii)fact constellation Schema

Ans :

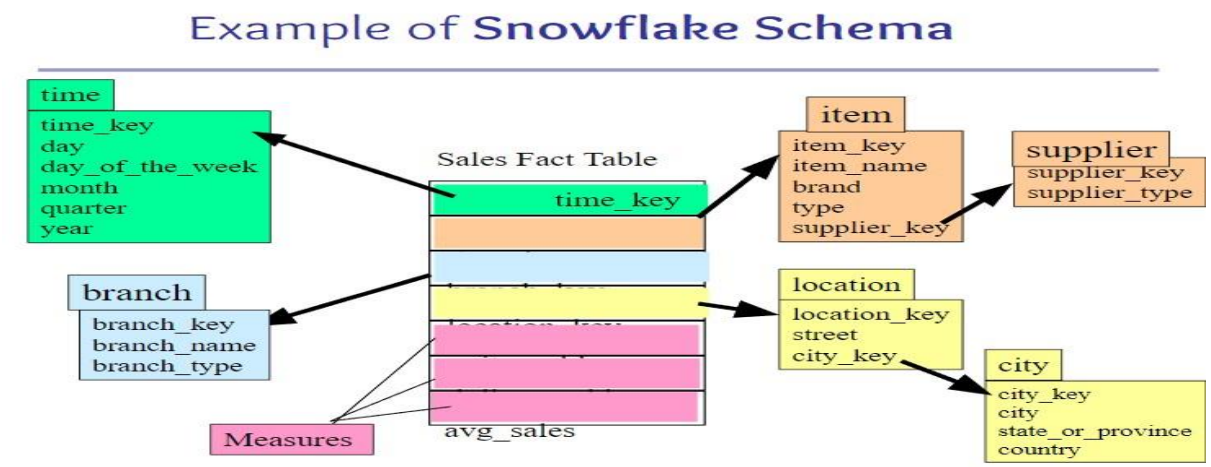
Star Schema :

A fact table in the middle connected to a set of dimension tables



Snowflake schema:

A refinement of star schema where some dimensional hierarchy is normalized into a set of smaller dimension tables, forming a shape similar to snowflake



Fact constellations:

Multiple fact tables share dimension tables, viewed as a collection of stars, therefore called galaxy schema or fact constellation