GOVERNMENT ENGINEERING COLLEGE, PATAN

Subject: DMBI-2170715 Semester: VII

Assignment-2

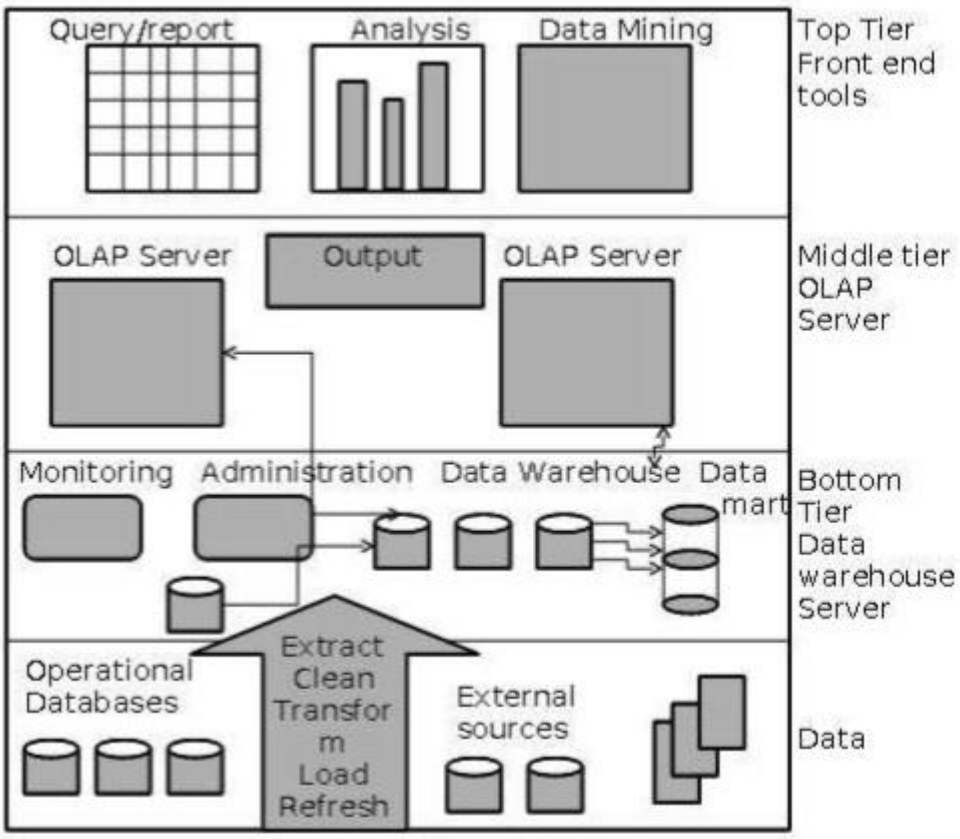
Q1) Explain Data Warehouse architecture in detail.

ANS) Data warehouse architecture refers to the design of an organization's data collection and storage framework. Because data needs to be sorted, cleaned, and properly organized to be useful, data warehouse architecture focuses on finding the most efficient method of taking information from a raw set and placing it into an easily digestible structure that provides valuable BI insights.

Generally a data warehouses adopts a three-tier architecture. Following are the three tiers of the data warehouse architecture.

- **Bottom Tier** − The bottom tier of the architecture is the data warehouse database server. It is the relational database system. We use the back end tools and utilities to feed data into the bottom tier. These back end tools and utilities perform the Extract, Clean, Load, and refresh functions.
- **Middle Tier** − In the middle tier, we have the OLAP Server that can be implemented in either of the following ways.
    - By Relational OLAP (ROLAP), which is an extended relational database management system. The ROLAP maps the operations on multidimensional data to standard relational operations.
    - By Multidimensional OLAP (MOLAP) model, which directly implements the multidimensional data and operations.
- **Top-Tier** − This tier is the front-end client layer. This layer holds the query tools and reporting tools, analysis tools and data mining tools.

The following diagram depicts the three-tier architecture of data warehouse –



Q2) Explain Business Intelligence architecture and its components

ANS) Business intelligence architecture refers to the infrastructure that organizations use to define their data collection streams, information administration, and all the technology that supports their business intelligence.

Creating a sustainable architecture depends on understanding the different components that are involved with developing successful business intelligence tools. The process is broadly divided into three areas: data collection, information management, and business intelligence.

The first area refers to the different channels and methods of collecting data from activities carried out within your organization. This includes understanding which data different users need to meet their requirements, as well as a clear idea of the quality, type, and currency of the data. This step is vital for adding value as the right data produces the best BI insights.

The second major component is data management. This covers various aspects of integrating data, scrubbing datasets, and fashioning the overall structures that will house and administer data.

Finally, business intelligence is the part of an organization's architecture that analyzes properly organized data sets to produce insights. This area involves using real-time analytics, data visualizations, and other BI tools.

Q3) What do you mean by OLTP and OLAP? Differentiate OLTP and OLAP system.

ANS) OLTP and OLAP both are the online processing systems. OLTP is a transactional processing while OLAP is an analytical processing system. OLTP is a system that manages transaction-oriented applications on the internet for example, ATM. OLAP is an online system that reports to multidimensional analytical queries like financial reporting, forecasting, etc.

OLTP	OLAP
It is an online transactional system and manages database modification.	It is an online data retrieving and data analysis system.
Insert, Update, Delete information from the database.	Extract data for analyzing that helps in decision making.
OLTP and its transactions are the original source of data.	Different OLTPs database becomes the source of data for OLAP.
OLTP has short transactions.	OLAP has long transactions.
The processing time of a transaction is comparatively less in OLTP.	The processing time of a transaction is comparatively more in OLAP.
Simpler queries.	Complex queries.
Tables in OLTP database are normalized (3NF).	Tables in OLAP database are not normalized.
OLTP database must maintain data integrity constraint	OLAP database does not get frequently modified. Hence, data integrity is not affected.

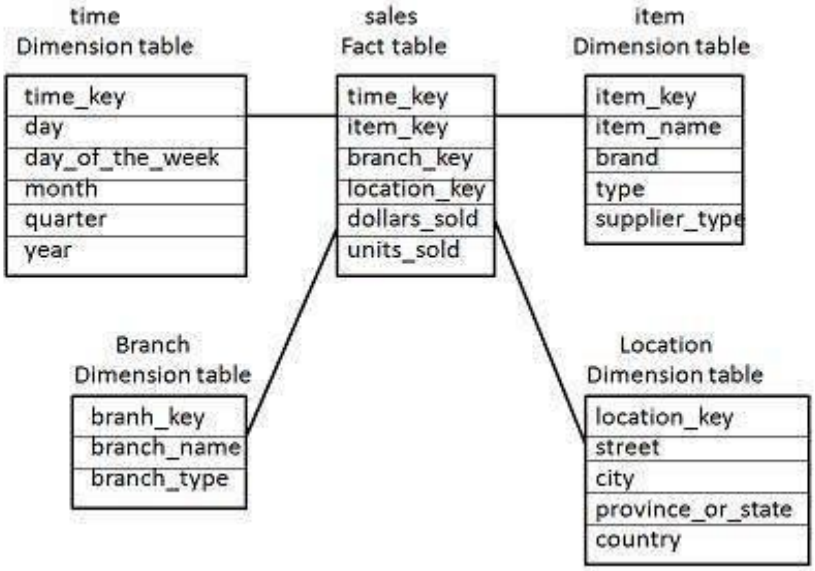
Q4) List out different schemas of data warehouse and explain each in detail. OR Explain Star

ANS) Schema is a logical description of the entire database. It includes the name and description of records of all record types including all associated data-items and aggregates. Much like a database, a data warehouse also requires to maintain a schema. A database uses relational model, while a data warehouse uses Star, Snowflake, and Fact Constellation schema.

Star Schema

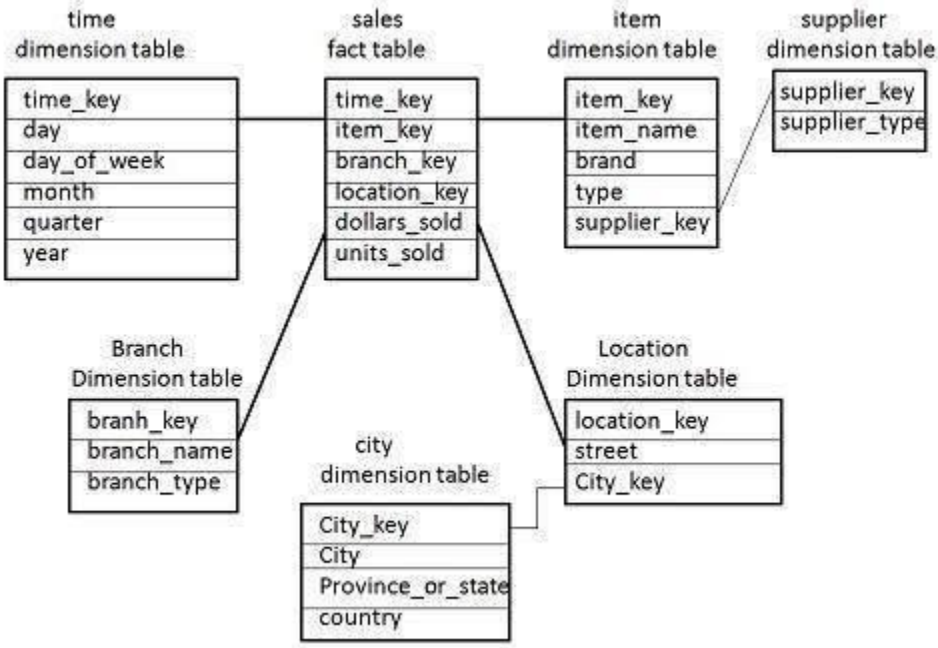
- Each dimension in a star schema is represented with only one-dimension table.
- This dimension table contains the set of attributes.
- The following diagram shows the sales data of a company with respect to the four dimensions, namely time, item, branch, and location.

- There is a fact table at the center. It contains the keys to each of four dimensions.
- The fact table also contains the attributes, namely dollars sold and units sold.



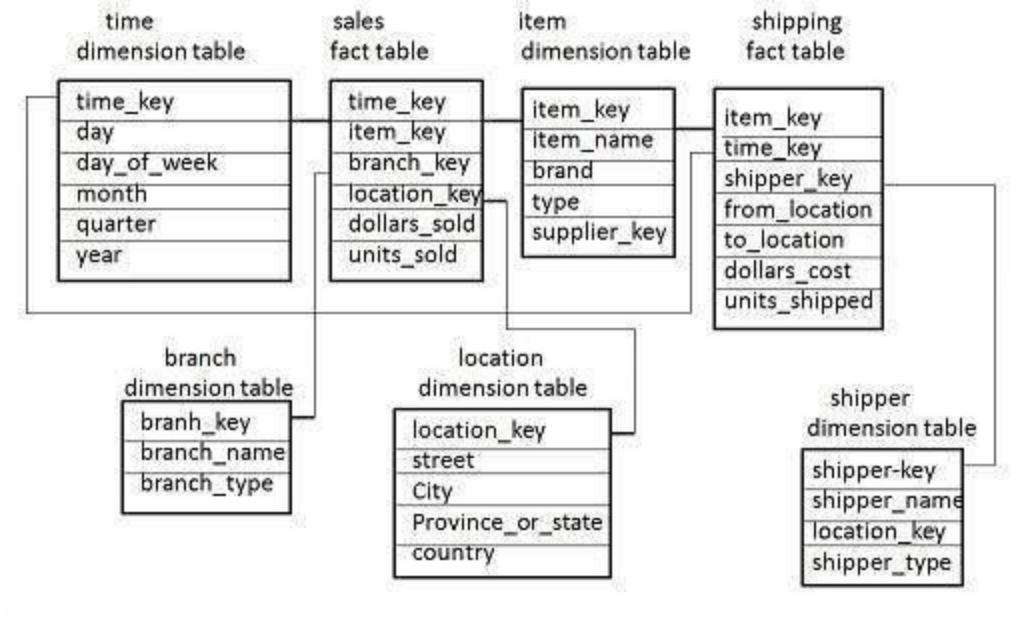
Snowflake Schema

- Some dimension tables in the Snowflake schema are normalized.
- The normalization splits up the data into additional tables.
- Unlike Star schema, the dimensions table in a snowflake schema are normalized. For example, the item dimension table in star schema is normalized and split into two dimension tables, namely item and supplier table.
- Now the item dimension table contains the attributes item_key, item_name, type, brand, and supplier-key.
- The supplier key is linked to the supplier dimension table. The supplier dimension table contains the attributes supplier_key and supplier_type.



Fact Constellation Schema

- A fact constellation has multiple fact tables. It is also known as galaxy schema.
- The following diagram shows two fact tables, namely sales and shipping.
- The sales fact table is same as that in the star schema.
- The shipping fact table has the five dimensions, namely item_key, time_key, shipper_key, from_location, to_location.
- The shipping fact table also contains two measures, namely dollars sold and units sold.
- It is also possible to share dimension tables between fact tables. For example, time, item, and location dimension tables are shared between the sales and shipping fact table.

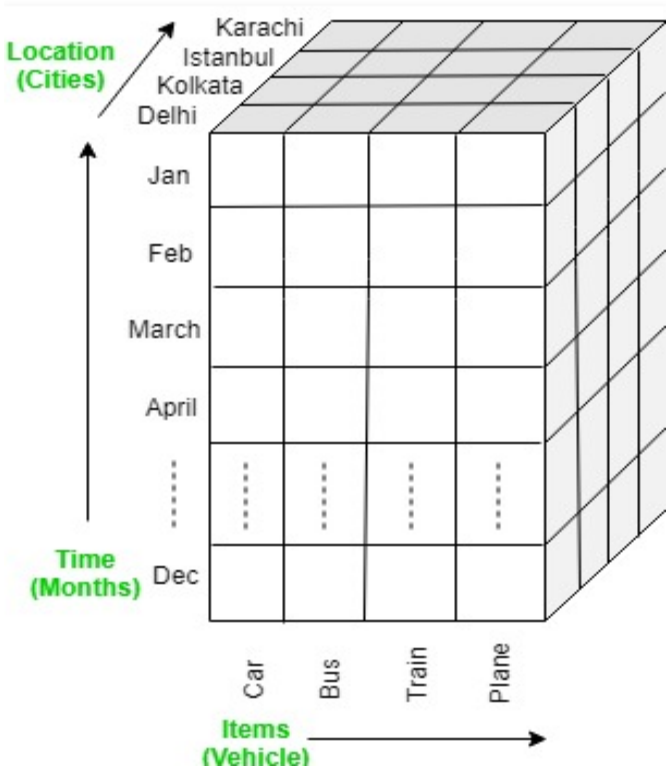


Q5) Explain different OLAP operations in detail

ANS) 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

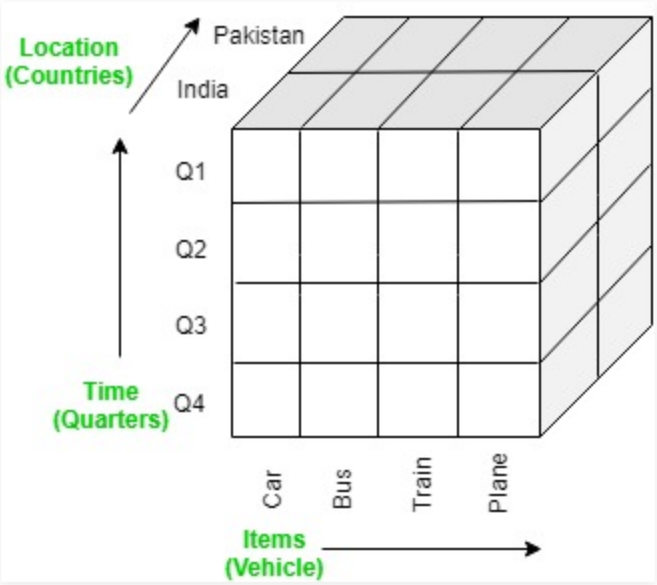
   In the cube given in overview section, the drill down operation is performed by moving down in the concept hierarchy of *Time* dimension (Quarter -> Month).



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

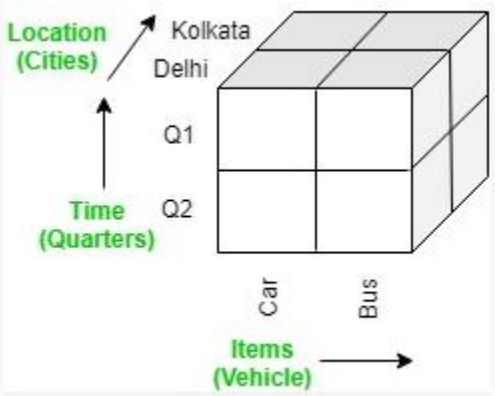
- Climbing up in the concept hierarchy
- Reducing the dimensions

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).

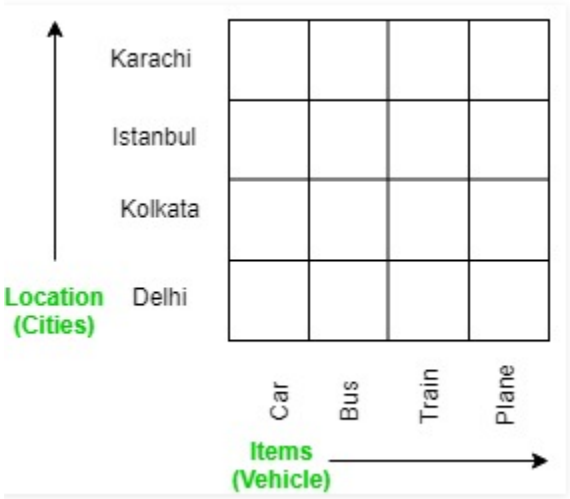


3)Dice: It selects a sub-cube from the OLAP cube by selecting two or more dimensions. In the cube given in the overview section, a sub-cube is selected by selecting following dimensions with criteria:

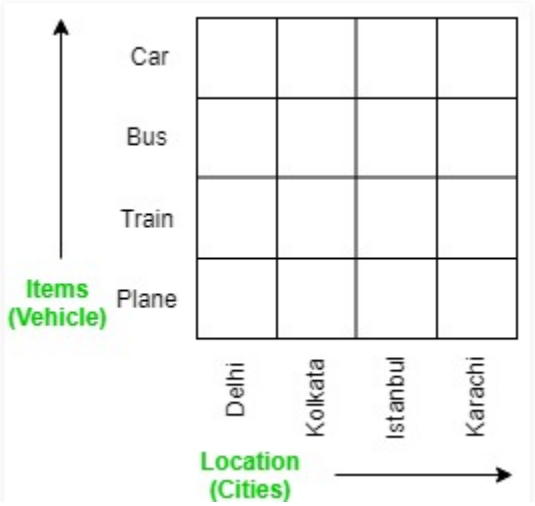
- Location = “Delhi” or “Kolkata”
- Time = “Q1” or “Q2”
- Item = “Car” or “Bus”



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.



Q6) Explain types of OLAP servers. OR explain ROLAP, MOLAP and HOLAP in detail.

ANS) 1)ROLAP:

OLAP stands for Relational Online Analytical Processing. ROLAP stores data in columns and rows (also known as relational tables) and retrieves the information on demand through user submitted queries. A ROLAP database can be accessed through complex SQL queries to calculate information. ROLAP can handle large data volumes, but the larger the data, the slower the processing times.

Because queries are made on-demand, ROLAP does not require the storage and pre-computation of information. However, the disadvantage of ROLAP implementations are the potential performance constraints and scalability limitations that result from large and inefficient join operations between large tables. Examples of popular ROLAP products include Metacube by Stanford Technology Group, Red Brick Warehouse by Red Brick Systems, and AXSYS Suite by Information Advantage.

2)MOLAP:

MOLAP stands for Multidimensional Online Analytical Processing. MOLAP uses a multidimensional cube that accesses stored data through various combinations. Data is pre-computed, pre-summarized, and stored (a difference from ROLAP, where queries are served on-demand).

A multicube approach has proved successful in MOLAP products. In this approach, a series of dense, small, precalculated cubes make up a hypercube. Tools that incorporate MOLAP include Oracle Essbase, IBM Cognos, and Apache Kylin.

Its simple interface makes MOLAP easy to use, even for inexperienced users. Its speedy data retrieval makes it the best for “slicing and dicing” operations. One major disadvantage of MOLAP is that it is less scalable than ROLAP, as it can handle a limited amount of data.

3)HOLAP:

HOLAP stands for Hybrid Online Analytical Processing. As the name suggests, the HOLAP storage mode connects attributes of both MOLAP and ROLAP. Since HOLAP involves storing part of your data in a ROLAP store and another part in a MOLAP store, developers get the benefits of both.

With this use of the two OLAPs, the data is stored in both multidimensional databases and relational databases. The decision to access one of the databases depends on which is most appropriate for the requested processing application or type. This setup allows much more flexibility for handling data. For theoretical processing, the data is stored in a multidimensional database. For heavy processing, the data is stored in a relational database.

Microsoft Analysis Services and SAP AG BI Accelerator are products that run off HOLAP.