NAME: YASIR FAHEEM

ID # 6991

SUBJECT: DATA WAREHOUSINNG

Q NO 1: Describe the Architecture of Datawarehouse?

Ans:

Data ware house is complex as it is information system that contains historical and commutative data from multiple sources.

There are three following approaches for constructing data warehouse.

1: Single tier architecture:

It is use for to minimize the amount of data stored. Through this type we can remove the overflow of the data but this architecture is not use.

2: Two tier architecture:

This architecture separates the data warehouse and physical available source. It is also a small architecture and it is not use in practice due to connectivity problem.

3: Three tier architecture:

This architecture is mostly used in practice.

It has three tier top, bottom and middle tier.

Top tier:

Through with the help of this layer we can connect and throughout the data from the data ware house.

It could be query tools, reporting tools, managed query tools, analysis and data mining tools.

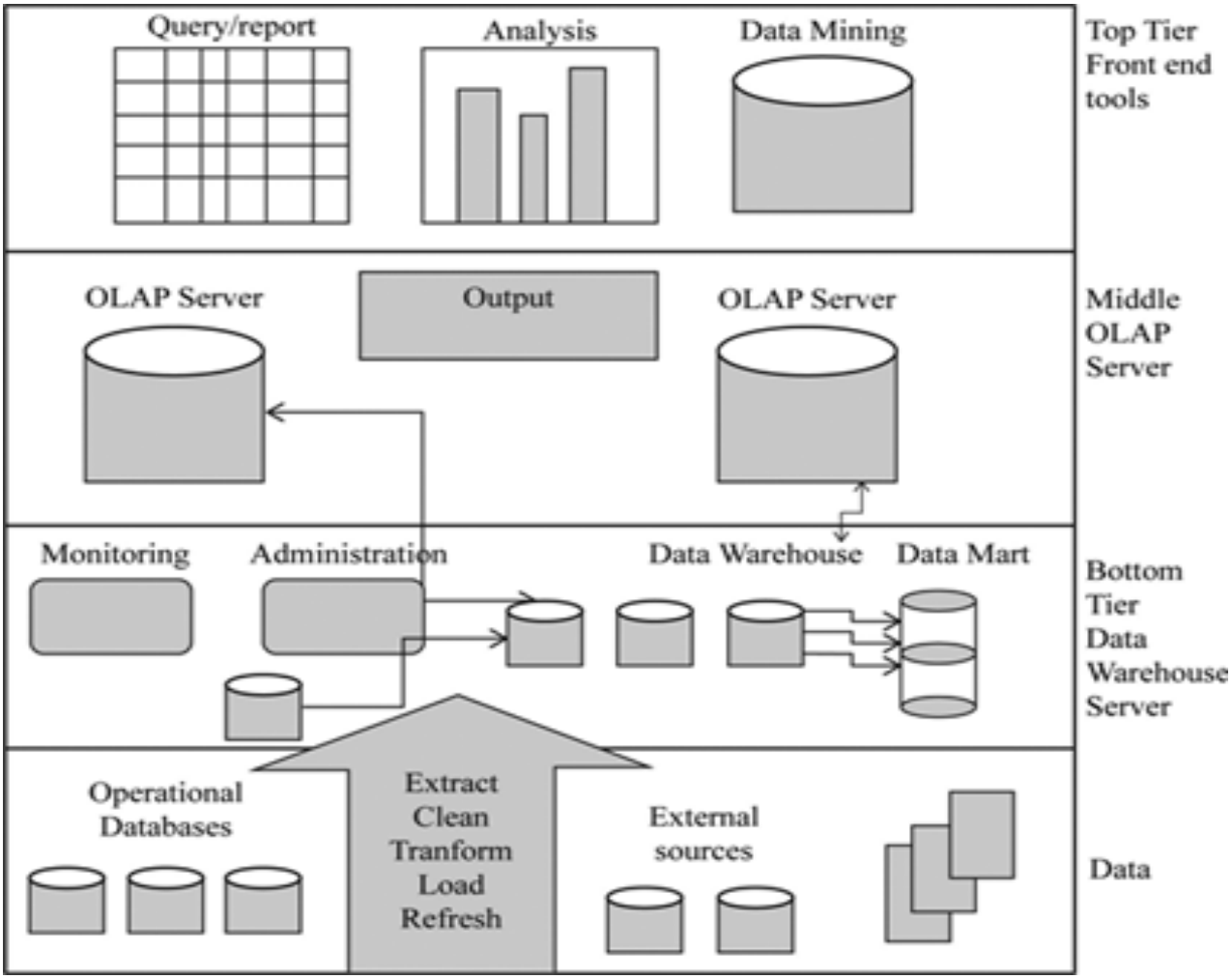
Middle tier:

This is OLAP server which is implemented through either ROLAP or MOLAP model. This is place between the top and bottom so therefore it is called middle tier.

Bottom tier:

It can be create a relation between the data base system. We can transform and loaded data through this layer.

Data warehouse architecture



Q NO 2: Describe star schema with example of any relevant database structure and its representation?

Ans:

Schema:

Schema is the structure which we make our data warehouse.

Star schema:

Star schema is that in which each dimension his own table. Like single, large and central fact. In every dimension his one tuple and its additional many attributes. It consist one or more fact table. It can import mostly cases of snowflake schema.

This type of schema is basic and it mostly used in the data ware house. Through this schema easy to understand, easy to define their hierarchies and reduces no of physical joins.

It is called star schema because it like looks as the star. And it is easy to understand and is simplest form of data ware house.

In star schema the one fact table and they surrounded by dimension table.

Fact table:

Measurement or metric of specific event can be record in fact table. It consist numeric value and foreign keys. It can be defined as one of three types.

- **Transaction fact table**

  It can be record fact about specific event.

- **Snapshot fact tables**

  It can be record fact at given point in a time

- **Accumulating snapshot tables**

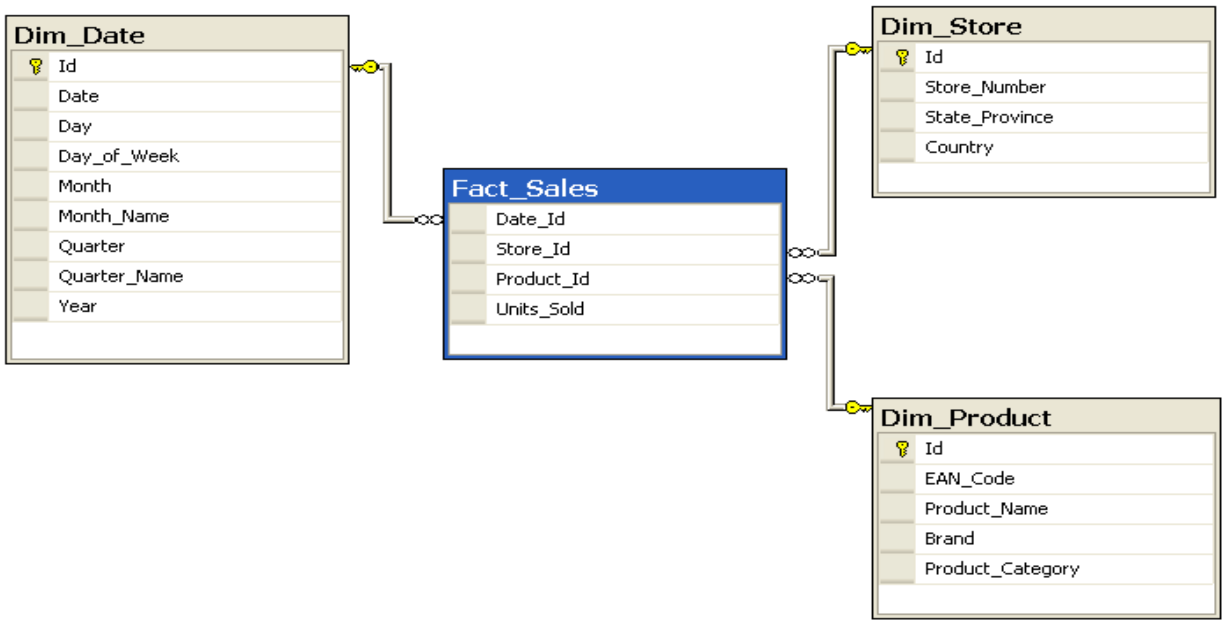
  It can be record aggregate fact at given point in a time.

Dimension table:

It has small number of record compared to fact table. But every record to describe fact data has large number of attribute. It can define many type of characteristic.

- Time dimension tables describe time at the lowest level of time granularity for which events are recorded.
- Geography dimension tables describe location data, such as country, state, or city
- Product dimension tables describe products
- Employee dimension tables describe employees, such as sales people
- Range dimension tables describe ranges of time, dollar values or other measurable quantities to simplify reporting

EXAMPLE:



Fact sales is the fact table and there are three **dimension** tables **Dim date**, **Dim store** and **Dim product**

Each dimension table has a primary key on its id column, relating to one of the columns of the **Fact sales** table's three-column primary key (**Date id, store id and product id).** The non-primary key **units sold** column of the fact table in this example represents a measure or metric that can be used in calculations and

analysis. The non-primary key columns of the dimension tables represent additional attributes of the dimensions

Q NO 3:

Describe snowflake schema with example of any relevant database structure and its representation?

Ans:

Snow flake schema:

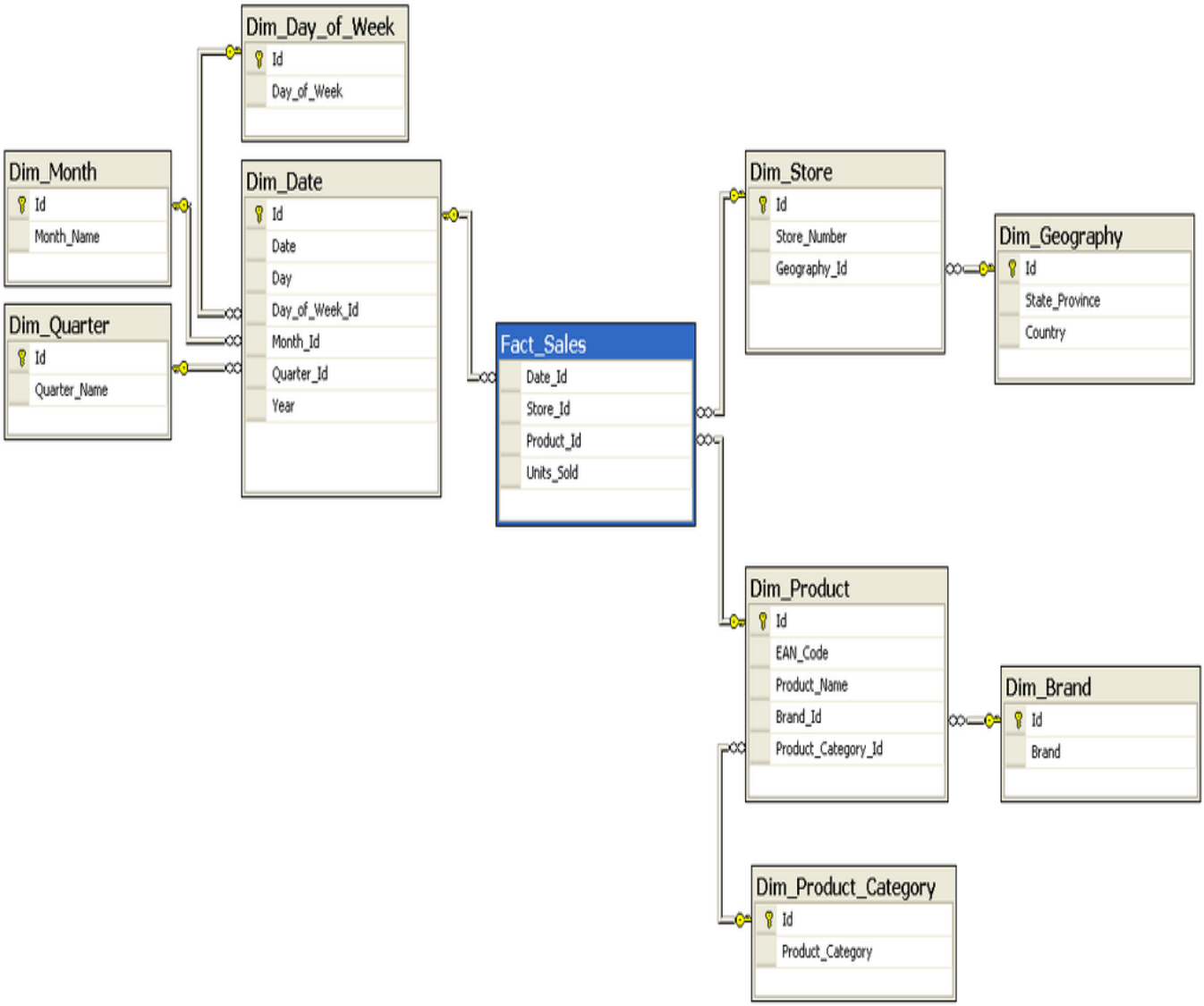
It is the extension of star schema. It has additional dimension table which are surrounded by other dimension table. The data can be split into additional dimension.

There hierarchies are divided into separate tables. It has very complex data base design because it is very large. It has very low level redundancy. In this the cube processing is very slow due to complexity. It is also has one fact table which is surrounded by additional dimension like as schema but different is that in this there dimension are surrounded by the other dimension.

It is the logically arrangement of table in multidimensional data base. Its diagram is look like snowflake shape.

It is the method which normalized the dimension table in star schema. Snowflake schema is same as star schema. The star schema considers is the special case of snow flake schema.

EXAMPLE:



The following example query is the snowflake schema equivalent of the star schema example code which returns the total number of television units sold by brand and by country for 1997. Notice that the snowflake schema query requires many more joins than the star schema version in order to fulfill even a simple query. The benefit of using the snowflake schema in this example is that the storage requirements are lower since the snowflake schema eliminates many duplicate values from the dimensions themselves.

```sql
SELECT
        B.Brand,
        G.Country,
        SUM(F.Units_Sold)
FROM Fact_Sales F
INNER JOIN Dim_Date D              ON F.Date_Id = D.Id
INNER JOIN Dim_Store S             ON F.Store_Id = S.Id
INNER JOIN Dim_Geography G         ON S.Geography_Id = G.Id
INNER JOIN Dim_Product P           ON F.Product_Id = P.Id
INNER JOIN Dim_Brand B             ON P.Brand_Id = B.Id
INNER JOIN Dim_Product_Category C ON P.Product_Category_Id = C.Id
WHERE
        D.Year = 1997 AND
        C.Product_Category = 'tv'
GROUP BY
        B.Brand,
        G.Country
```