ASSIGNMENT-I

① Write difference between OLAP and OLTP?

OLAP	OLTP
① It involves historical processing of information.	① It involves day to day process.
② OLAP system are used by knowledged worker such as executives, managers and analysist.	② OLTP systems are used by clerks, database architecture, or database professionals
③ It is used to analyse the business.	③ It is used to run the business.
④ It focus on information out.	④ It focus on data in (or) application orient.
⑤ It is based on star schema, fact consolation schema.	⑤ It is based on entity-relation -ship model.
⑥ It contains historical data.	⑥ It contain current data.
⑦ It provides summerized and consolated data.	⑦ It provides primitive and highly detailed data.
⑧ It provides summerized and multidimensional view of data.	⑧ It provides detailed and flat relational view of data.
⑨ The number of users is in hundereds	⑨ The number of users is in thousands
⑩ The number of records assigned is in millions	⑩ The number of records access/assigned is in tens

(11) Data warehouse size is from 100GB to 100TB	(11) Database size is from 100MB to 100GB.
(13) These are highly flexible.	(12) It provide high performance.

(2) Define Data Warehouse. List and explain its characteristics.

Data Warehouse:

* The term "Data Warehouse" was first coined by 'Bill Inmon' in 1990. According to him data ware-house is a subject oriented, integrated, time-variented and non-volatile collection of Data.
* A data warehouse provide us generalised and consolidated data in multidimensional view or ways and also provide us online Analytical Processing tools (OLAP).

Characteristics of DataWarehouse:

There are four features in data warehousing.

(1) Subject Oriented: A data warehouse is a subject oriented because it provides information around a subject rather than the organisation ongoing operations.

A data warehouse doesnot focus on ongoing operations rather it focus on moulding and analysis

of data for decision making.

② Integrated: A data warehouse is a constructed by integrating data from heterogeneous source such as relational database, file etc.

This integration enhance the effective analysis of data.

③ Time-Varient: The data collected in a data ware-house is identified with a particular time period. The data providing information from historical point of view.

④ Non-volatile: Non-volatile means the previous data is not earased when new data is added.

Data warehouse is kept seperate from the operational data base and therefore frequent change in operational database doesnot reflected in data warehouse.

---

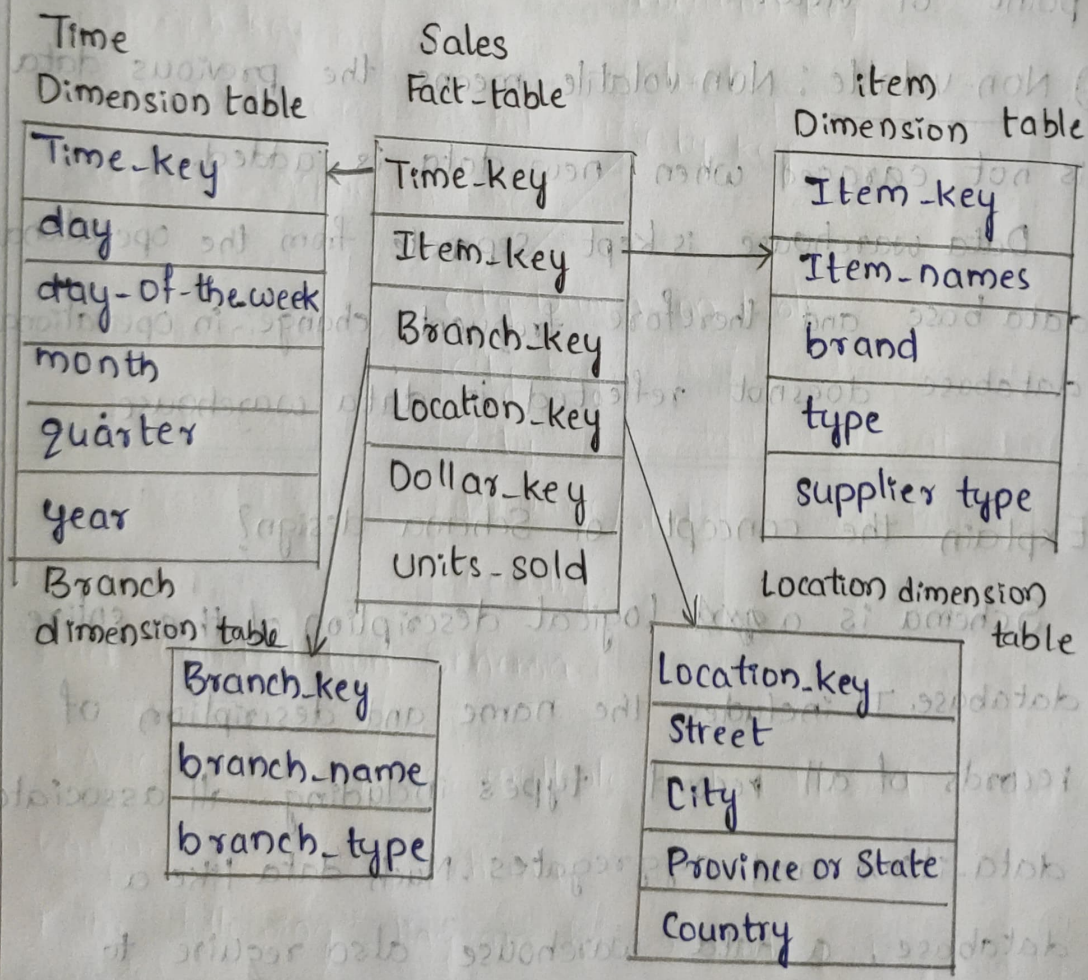
③ Explain the concept of Schema design?

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 data like a database, a data warehouse also require to maintain a Schema. A database uses relational

model, which a data warehouse uses star, snowflake and fact constellation schema. In this, we will discuss the schemas.

Star Schema:

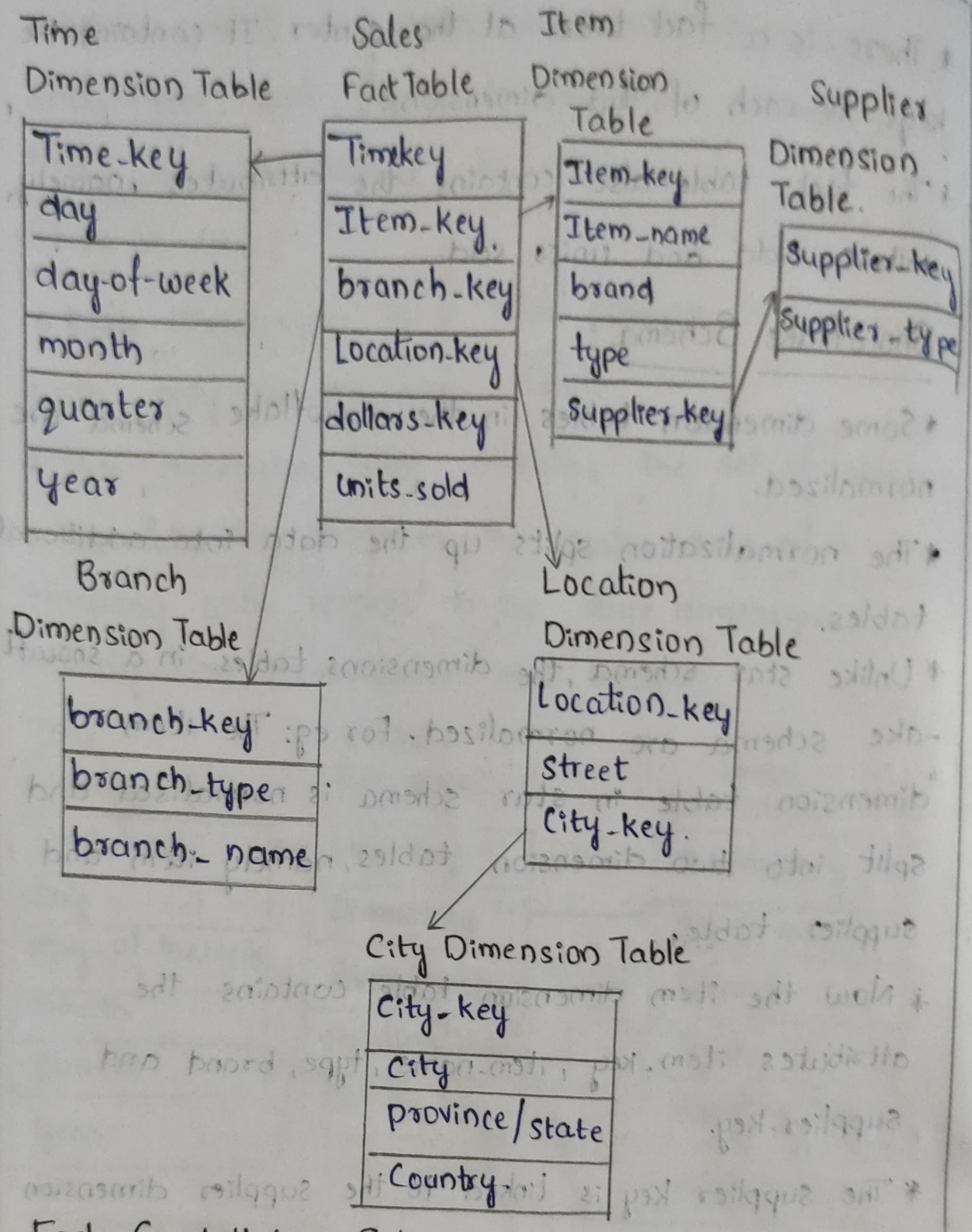
* Each dimension in a star schema is represented with only one-dimension table.
* This dimension table contains the set of attributes
* The following diagram - The sales data of a company with respect to the four dimensions, namely time, item, branch and location.



* 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 tables in a snowfl-ake schema are normalized. For eg: The item dimension table in star schema is normalised 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. In this the attributes are supplier_key and supplier type.
* Due to normalisation in the snowflake schema, the redundancy is reduced and therefore, it becomes easy to maintain and the save storage space.



Fact Constellation Schema:

* A constellation has multiple fact tables. It is also known as galaxy schema.
* The following diagram has 2-fact tables, namely sales and shipping.

Time dimension Table

Time_key
day
day_of_week
month
quater
year

Sales Fact Table

Time_key
Item_key
branch_key
location_key
dollar_sold
units_sold

Item dimension-table

Item_key
Item_name
brand
type
supplier_key

Shipping Fact Table

Item_key
Time_key
shipper_key
from_location
to_location
dollars_cost
units_shipped

Branch Dimension Table

branch_key
branch_name
branch_type

Location Dimension Table

location_key
Street
city
province/state
country.

Shipper Dimension Table

Shipper_key
Shipper_name
Location_key
Shipper_key

* The sales ~~fact~~ fact table is same as that in star schema.
* The shipping fact table also contains two measures namely dollars sold and units shipped.
* It is also possible to share dimension tables between fact tables. For eg: Time, Item, Location, are shared between the sales and shipping fact tables.

④ Explain about different architectures in Data Warehouse.

Data Warehouse Architecture:

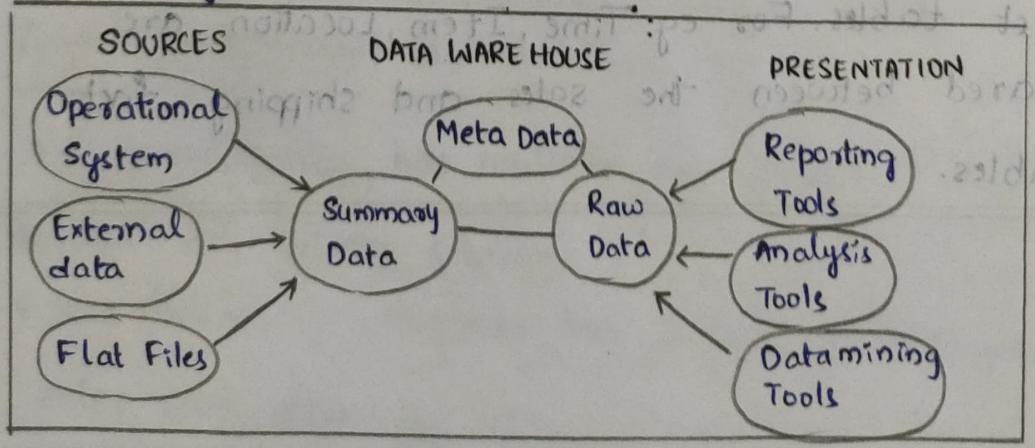
There are 3 ways you can construct a data warehouse system. These approaches are classified by the no.of tiers in the architecture. They are.

① Single Tier Architecture
② Two Tier Architecture
③ Three Tier Architecture

Single Tier Architecture:

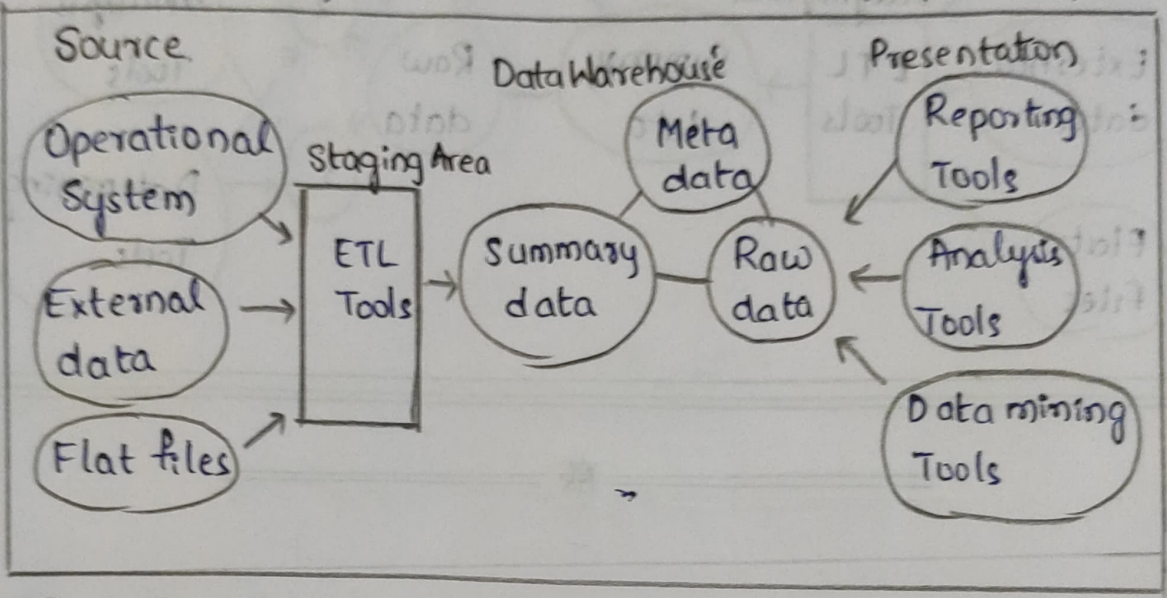
The single-Tier architecture is not a frequently practiced approach. The main goal of having such an architecture is to remove redundancy by minimizing the amount of data stored.

It's primary dis-advantage is that it doesn't have a component that seperates analytical and transactional processing.



Two-Tier Architecture:

It includes a staging area for all data source, before the data warehouse layer. By adding a staging area between the sources and the storage repository, you ensure all data loaded into the warehouse is cleansed and in the appropriate format.



Three Tier Architecture:

It is most widely used architecture for data warehouse systems.

1) The bottom tier: The database of the warehouse, where the cleansed and transformed data is loaded.

2) The middle tier: The application layers giving an abstracted view of the database. It arranges the data to make it more suitable for analysis. This is done with an OLAP server, implemented using the ROLAP or MOLAP model.

3) The top tier: is where the user accesses and interact with the data. It represents the front-end client layer.

You can use reporting tools, Query, Analysis or data mining tools.