DWDM Assignment-1:-

B-Mohan Sai Dinesh 20pa1a0521

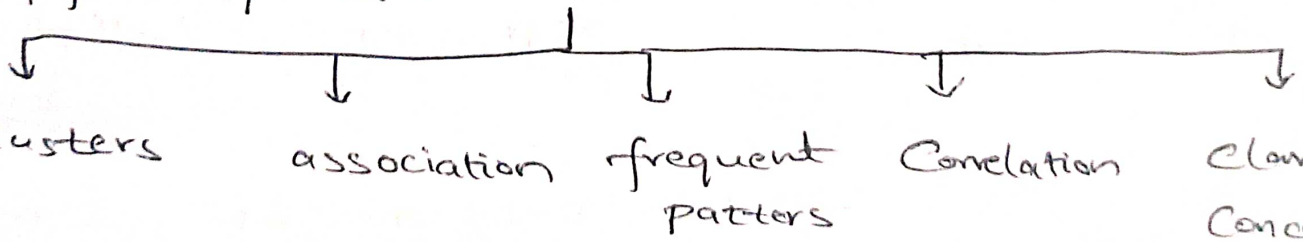
1) Data Mining:-

Data Mining is the process of sorting through large data sets to identify patterns and relationships that can help solve bussiness problems through data analysis.

Different types of data can be mined in data mining. Based on data functionalities, patterns are classified into two categories.

- Descriptive
- Predictive

Descriptive:- It deals with general characteristics and converts them into relevant and helpful information



Clusters:- This is for information of a grp of similar data points. Each point in the collection is somewhat similar but very different from other members of different groups.

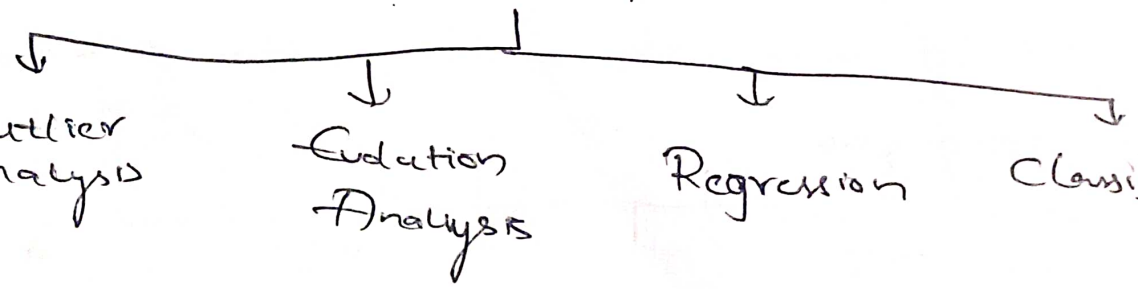
Class :- For instance, in a library, the class or items for borrowed items include books and journals

Frequent patterns :- These are data points that occur more often in dataset

Association :- It shows the relationship b/w data & pre-defined association rules.

Correlation :- It is performed to find statistical correlation b/w two data points.

Predictive Patterns :- It predicts future values by analyzing the data patterns and their outcomes based on the previous data.



Classification :- It predicts the label of unknown data points with help of known data points.

Regression :- It is used to find the missing numeric values

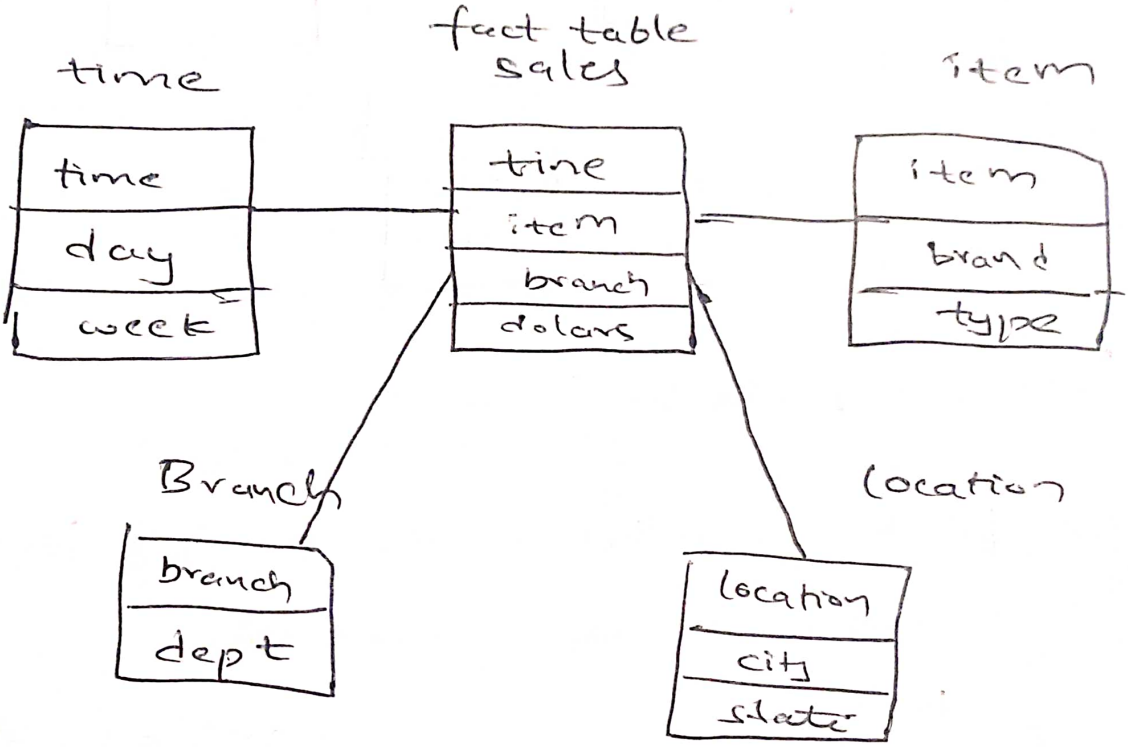
Outlier :- It is different from all the other points in the dataset.

2) Schema is a logical description of the entire database.

Star Schema:-

→ Each dimension in this schema is represented by 1D table.

→ the table contains the set of attributes



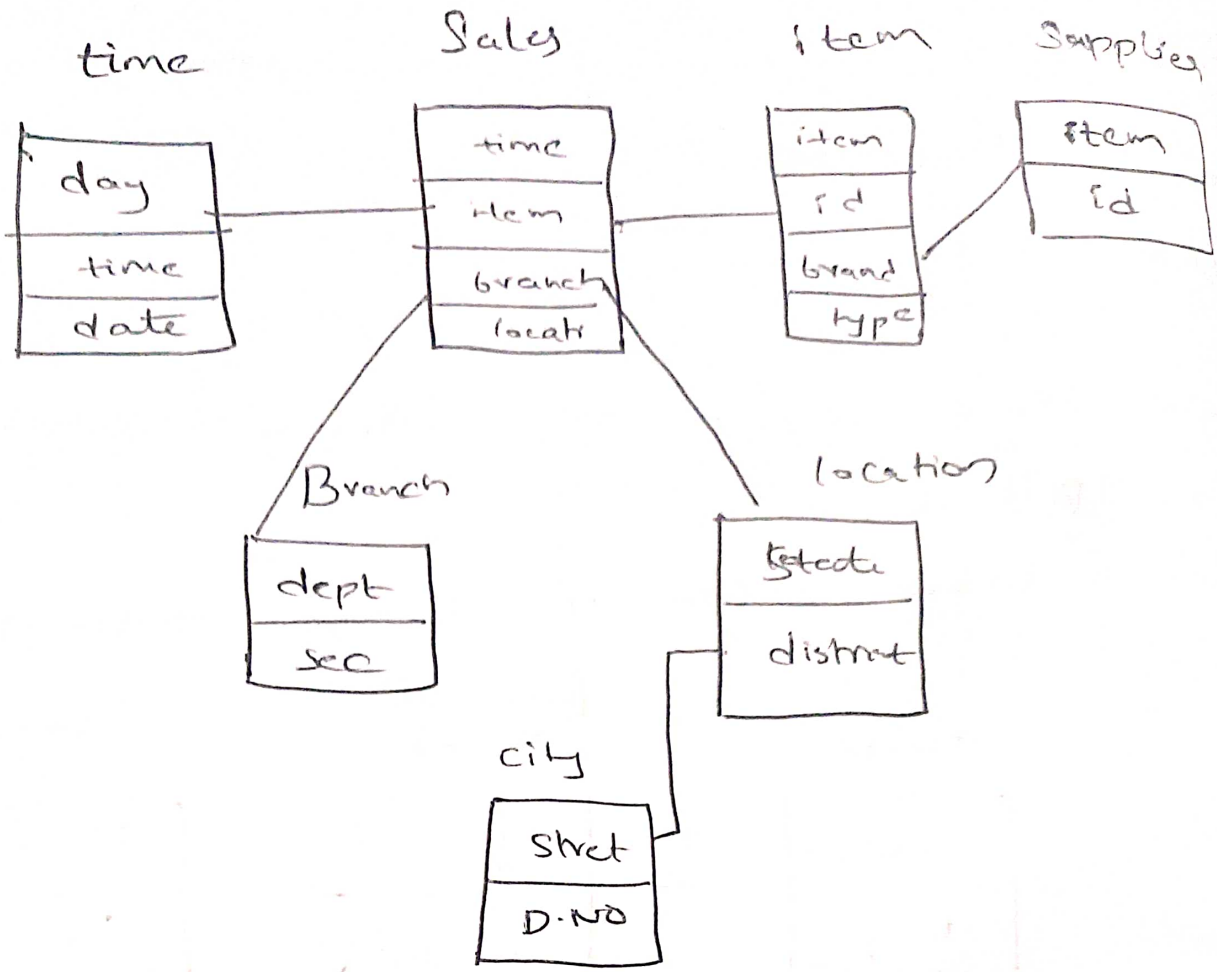
→ there is fact table at the centre. It contains the keys to each of 4 Dimensions

Snowflake Schema:-

→ Some tables are normalized

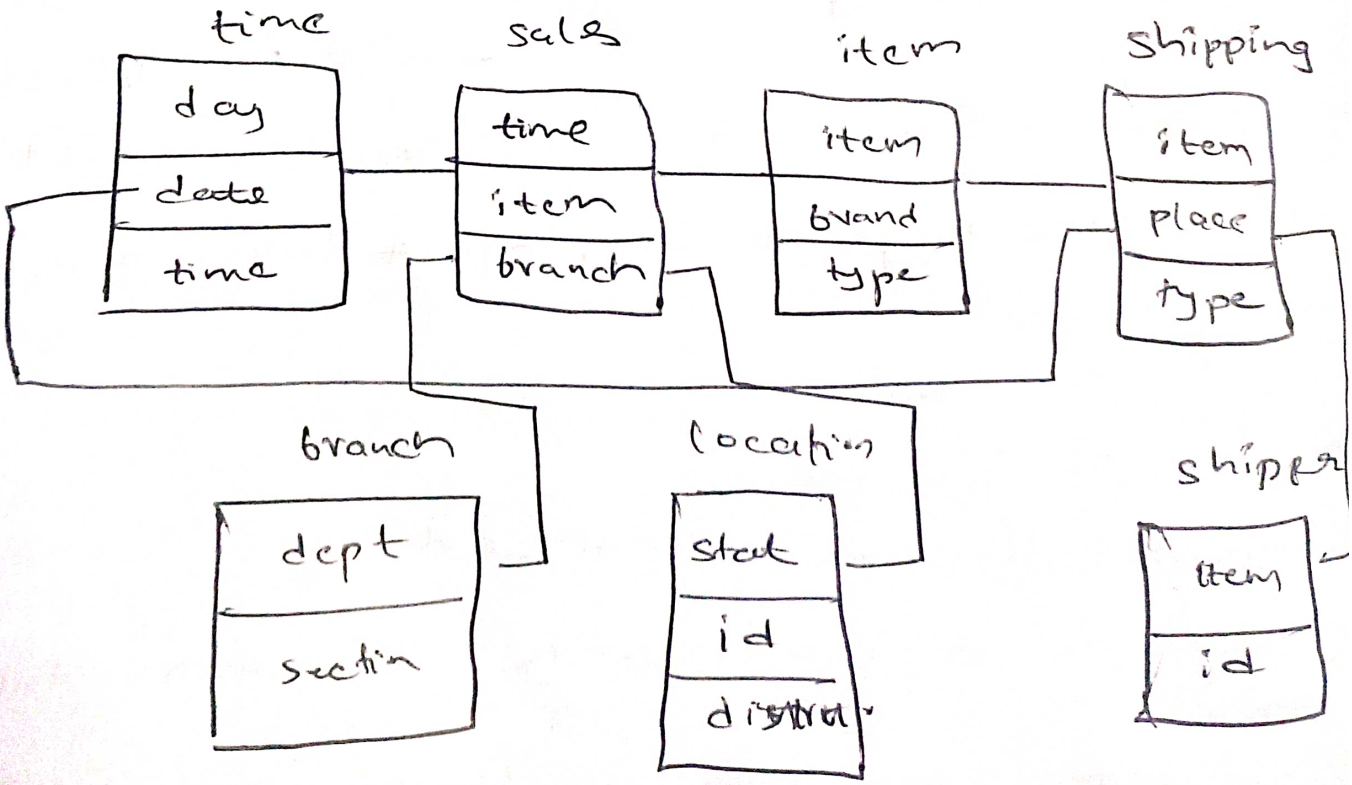
→ the normalization splits into additional tables

→ Due to normalization, the redundancy is reduced, and easy to save storage space.



fact Constellation Schema:

→ a fact constellation has multiple fact tables also known as galaxy schema



3) Apriori Algorithm:- It refers to algorithm which is used to calculate the association rules b/w objects. It operates on a database containing a huge no. of transactions.

Aprori algorithm helps customers to buy their products with ease and increases their sales performance on particular store

Ex:

T_ID	Rice	Pulse	Oil	Apple
t_1	1	1	1	0
t_2	0	1	1	1
t_3	0	0	0	1
t_4	1	1	0	1
t_5	1	1	1	0
t_6	1	1	1	1

Step 1:-

Rice	4
Pule	5
Oil	4
Apple	4

It indicates the products frequently bought

Step 2: create a pair of products

RP → 4	PO → 4
RO → 3	PM → 3
RM → 2	OM → 2

step 3: Implement the support to 50%- and confi having products more than 50%- i.e 3.

we can get RP, PO, RO, PM.

step 4: Now set of three products

→ RP and RO give RPO

→ PO and PM give POM.

step 5: i.e., RPO → 4

POM → 3

∴ If you implement threshold, you can figure out that customers set of product is RPO.

———— * ————