

Assignment - I1 Define the following terms

Business Intelligence: is a set of processes, architectures and technologies that convert raw data into meaningful information that drives profitable business.

Data Mart: is focused on a single functional area of an organization and contains a subset of data stored in a data warehouse.

closed frequent Itemset: is ~~is~~ It is a frequent Itemset that is both closed and its support is greater than or equal to minsup.

Outlier Analysis: is a ~~an~~ element of data set that distinctly stands out from the rest of the data.

2 List and describe the five primitives for specifying a data mining task.

1) set of task-relevant data to be mined.

This is the position of database in which the user is interested.

2) kind of knowledge to be mined

It refers to the kind of functions to be performed

1) characterization

2) Discrimination

3) classification

4) prediction

5) clustering.

170280116018

3) Background knowledge: The background knowledge allows data to be mined at multiple levels of abstraction.

4) Interestingness measures and thresholds for pattern evaluation: This is used to evaluate the patterns that are discovered by the process of knowledge discovery.

5) Representation for visualizing the discovered patterns.

1) Rules

2) Tables

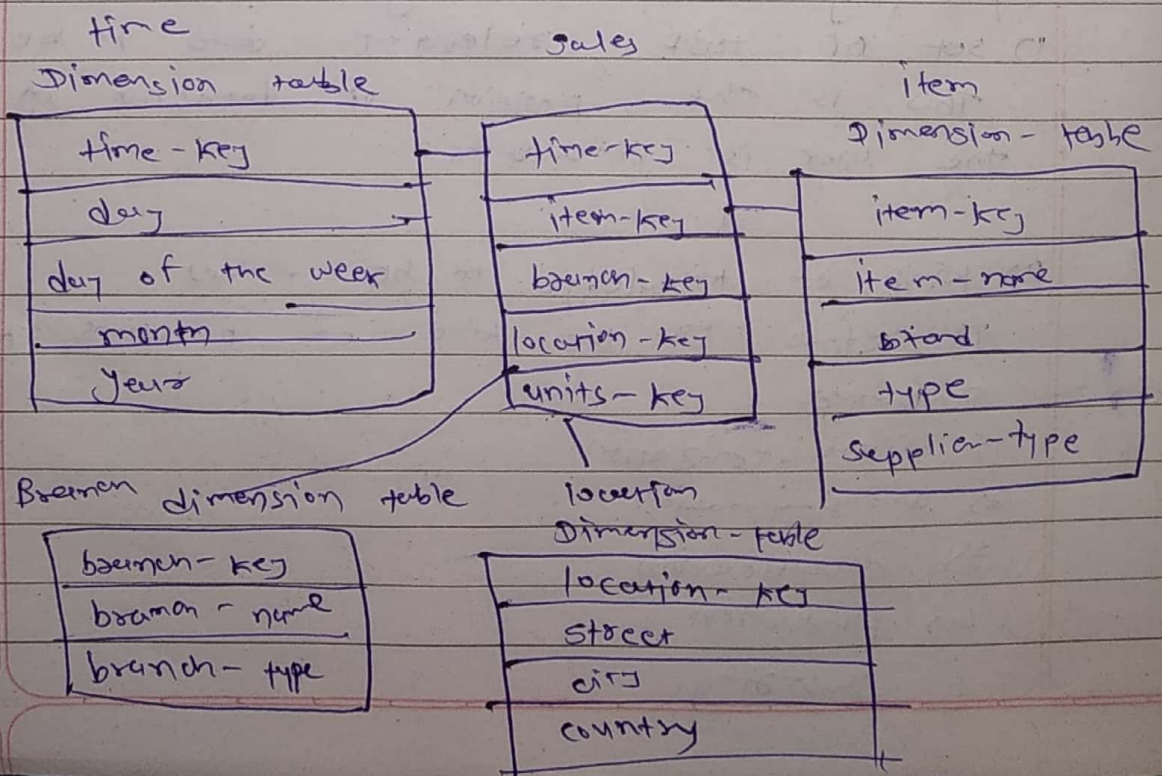
3) Charts

4) Graphs

Representations.

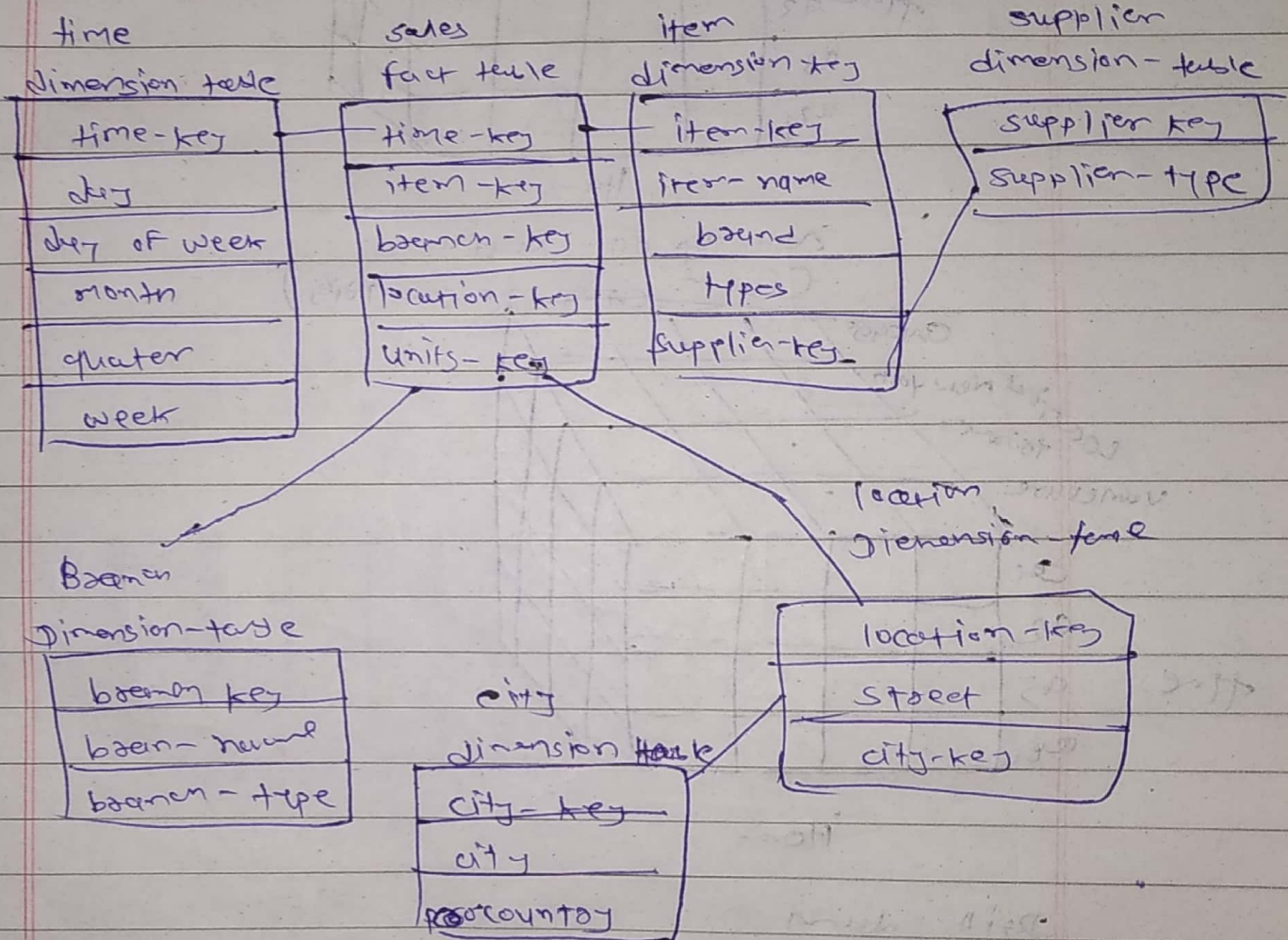
3. Explain star schema and snowflake schema with example.

— Star schema



snowflake schema

some dimension tables in the snowflake schema are normalized.

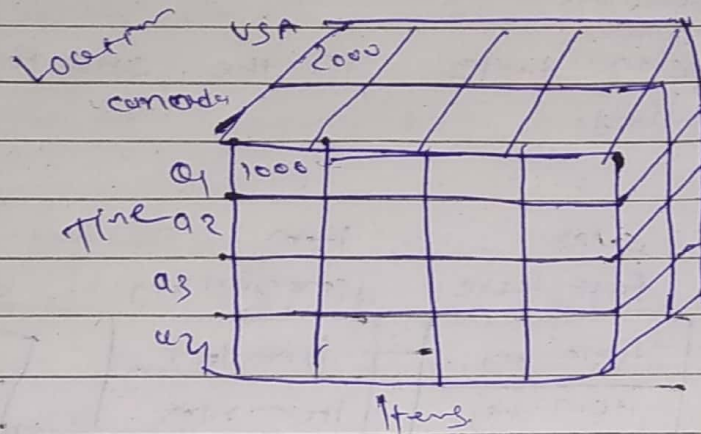


4. Explain various OLAP operations.

Roll-up

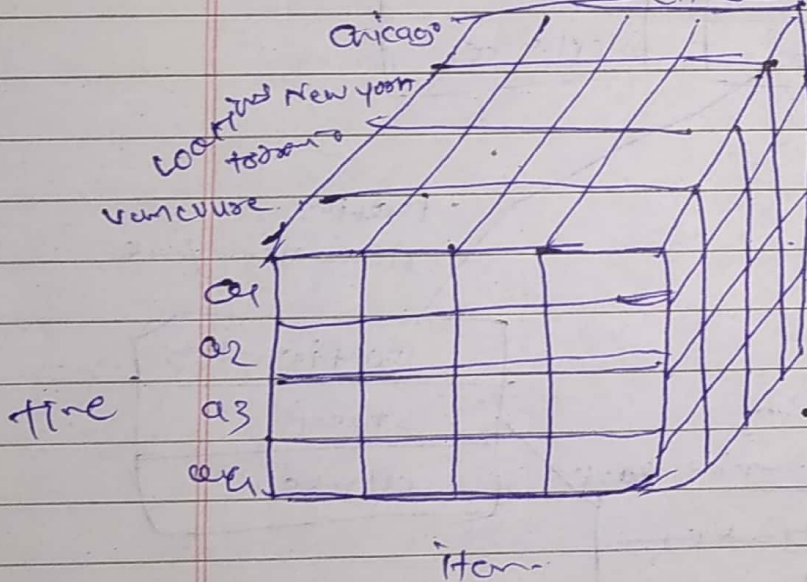
- Roll up performs aggregations on a data cube in any of the following ways.

- 1) By climbing up a concept hierarchy for a dimension.
- 2) By dimension reduction.



Roll up on location

(from cities to countries)

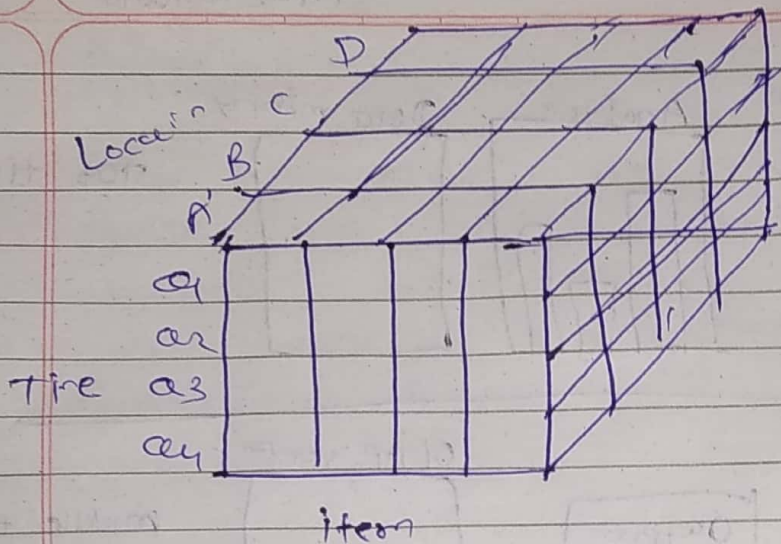


Drill down

- 1) By stepping down a concept hierarchy for a dimension
- 2) By introducing a new dimension.

slice

The slice operation selects one particular dimension from a given cube and provides a new sub-cube.



slice

for time
= R_i

Location

A				
C				
B				
A	505	825	1m	400

item.

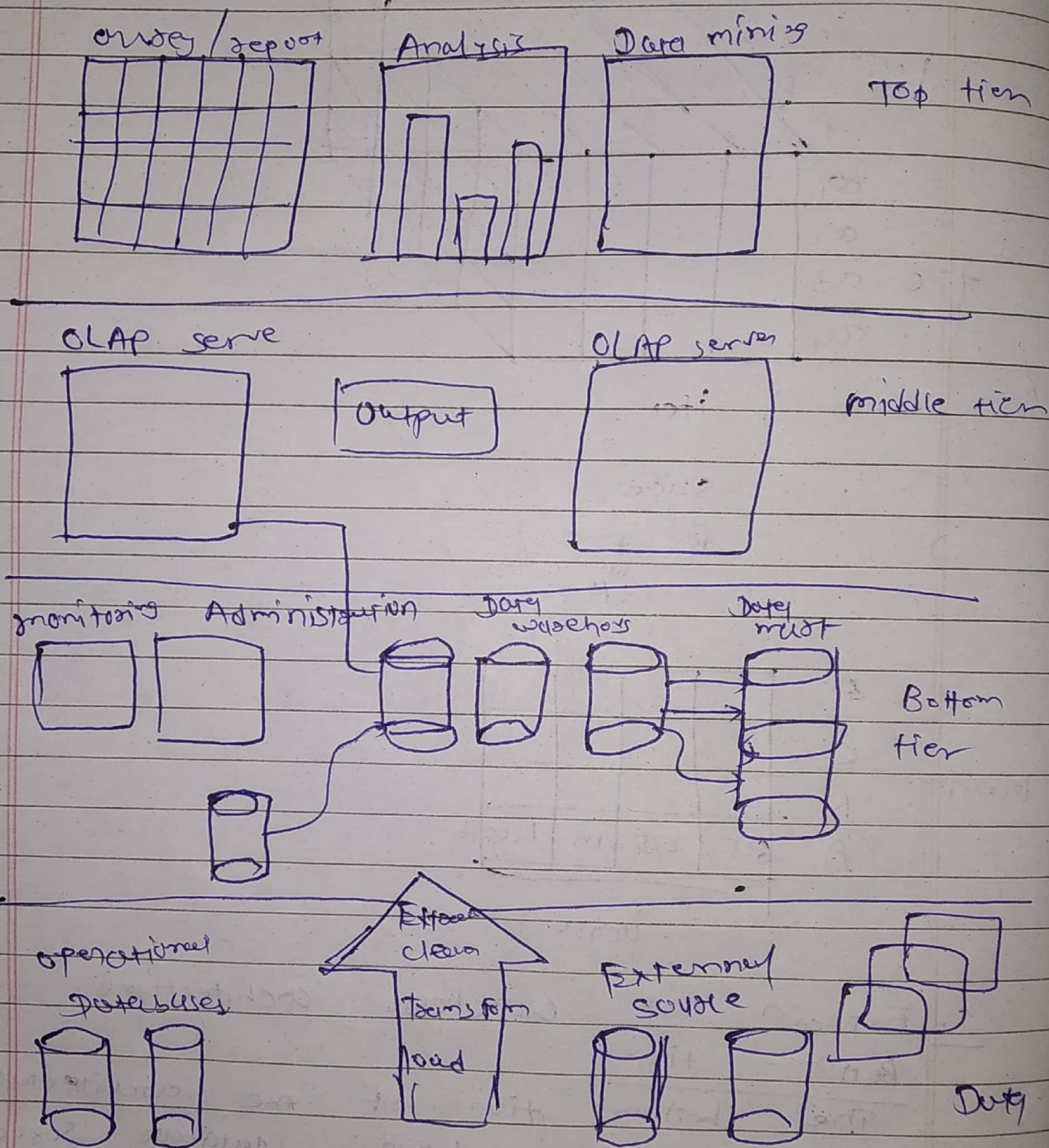
5. Explain Data warehouse architecture

Bottom - tier

The bottom tier of the architecture is the data warehouse database server.

It is the relational database system. we use the backend tools and utilities to feed data into the bottom tier.

170280116018



Middle tier

- In the middle tier, we have the OLAP server that can be implemented in either of the following ways
- 1) By Relational OLAP which is an extended relational database management system.

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.