

Assignment -2

Ques -1 Define the terms: Data Mart, Enterprise warehouse and Virtual Warehouse.

1) Data Mart :

- Focus only on one single department.
- It contains only one single subject.
- There are two types of data marts.
 - 1) Dependent data mart.
 - 2) Independent data mart.

2) Enterprise data warehouse :

- Collects all of the information about subjects spanning the entire organization.
- Scope is large.
- Contains more than one object.
- Fact consolidation schema.

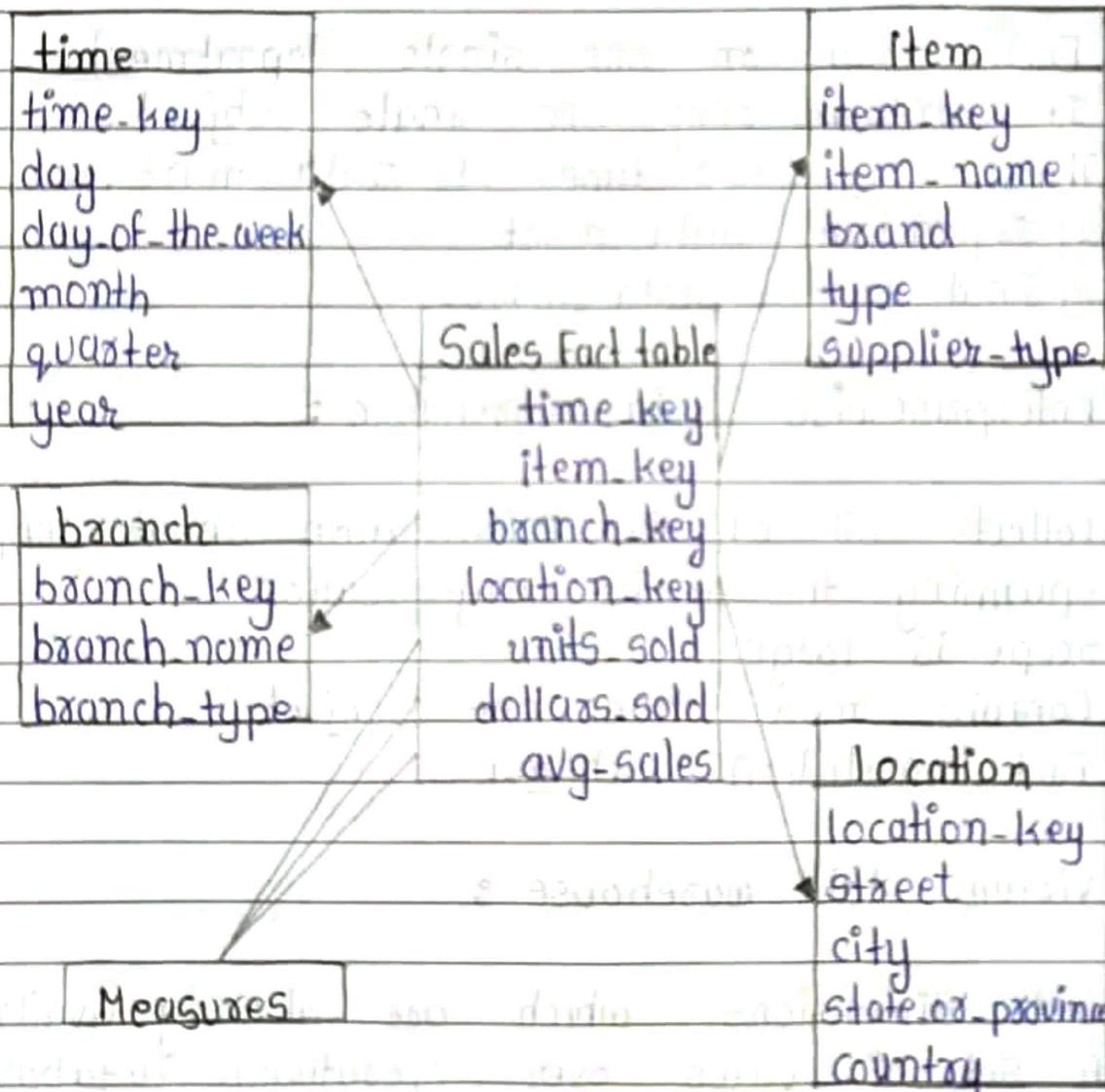
3) Virtual data warehouse :

- Get dimensions which are already available.
- A set of views over operational databases.
- Only some of the possible summary views may be materialized.

Ques-2 Discuss Stars, snowflakes and fact constellation using figures.

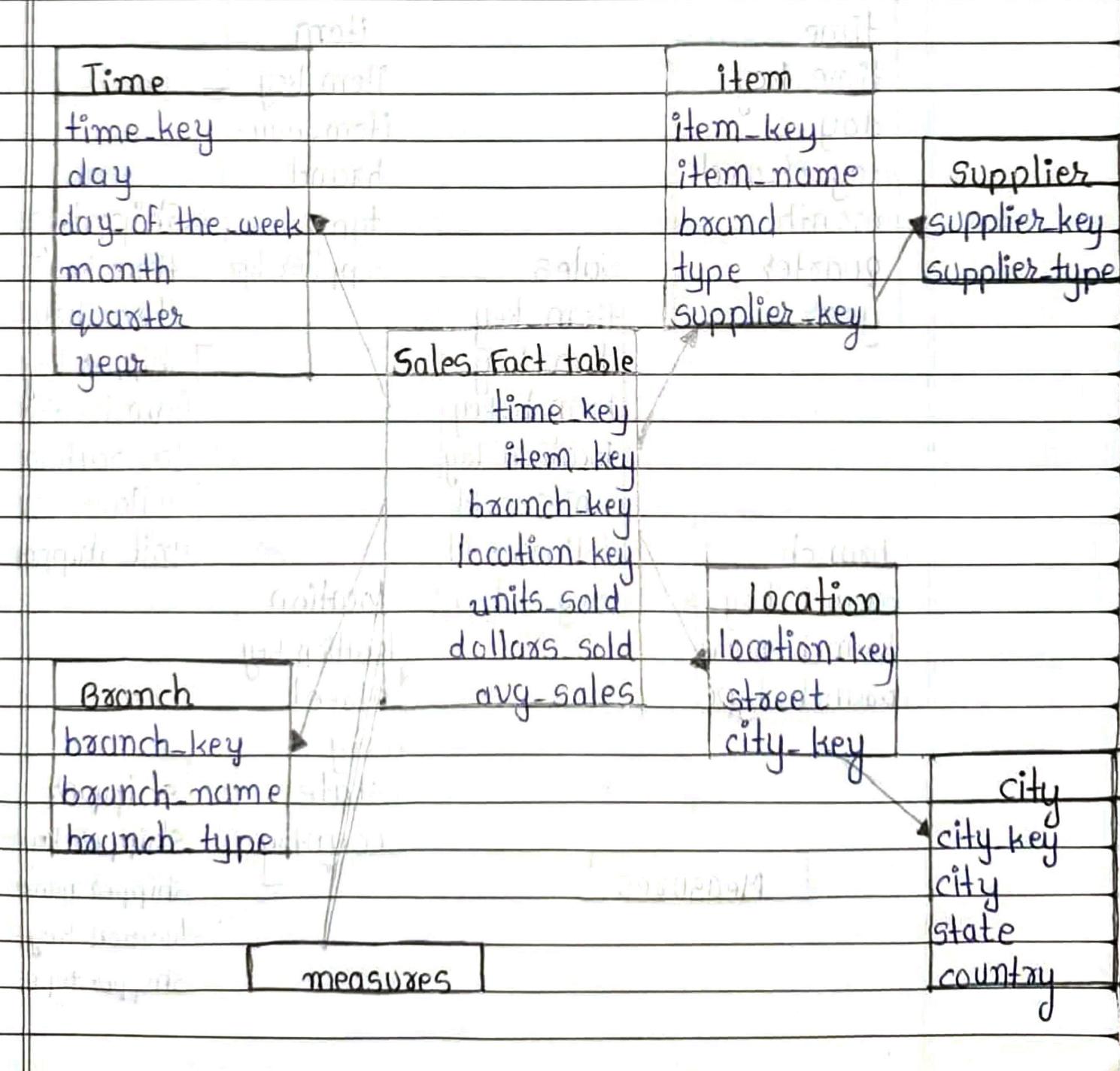
1) Star schema:

→ A fact table in the middle connected to a set of dimension tables.



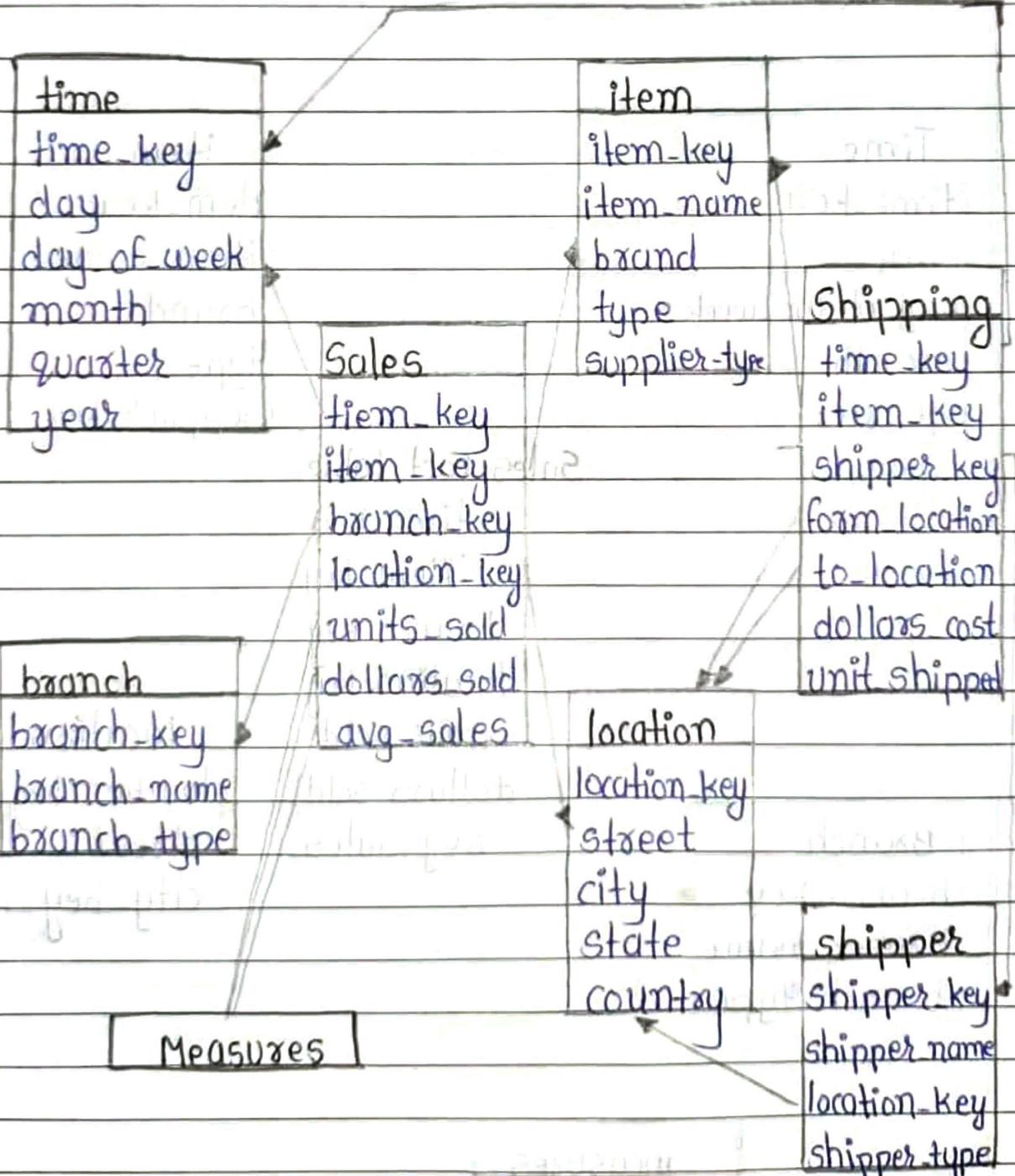
2) Snowflake schema :

→ A refinement of star schema where some dimensional hierarchy is normalized into a set of smaller dimension tables, forming a shape similar to snowflake.



3) Fact constellations :

→ Multiple fact table shares dimension tables, viewed as a collection of stars; therefore called galaxy schema or fact constellation.



Ques-31) List and Explain various OLAP operations.

→ These are 4 main OLAP operations.

1) Roll-up (drill up)

2) Drill down (roll down)

3) Slice and dice

4) Pivot (rotate)

1) Roll-up :

→ summarize data by first method with deepest

→ by climbing up hierarchy by dimension reduction

2) Drill down :

→ reverse of roll-up

→ from higher level summary to lower level summary
detailed data, or introducing new dimensions

3) Slice and dice :

→ project and select

4) Pivot :

→ reorient the cube, visualization, 3D to series of
2D planes

→ There are two more OLAP operations.

1) Drill across

2) Drill through

1) Drill across :

→ involving (across) more than one fact table

2) Drill through :

→ Through the bottom level of the cube to its back-end relational tables (using SQL)

Ques-4 State true/false and justify the answer: Decision Support system is used data warehouse to store data.

→ True.

→ DSS are interactive software based systems intended to help managers in decision making by accessing large volume of information generated from various related information systems involved in organizational business process, such as office automation systems, transaction processing systems, etc.

→ DSS uses the summary information, exceptions, patterns and trends using the analytical models.

→ A decision support system helps in decision making but does not give a decision itself.

- The decision makers compile useful information from raw data, documents, personal knowledge, and/or business models to identify and solve problems and make decisions.
- Hence, we can say that DSS is used data warehouse to store data.

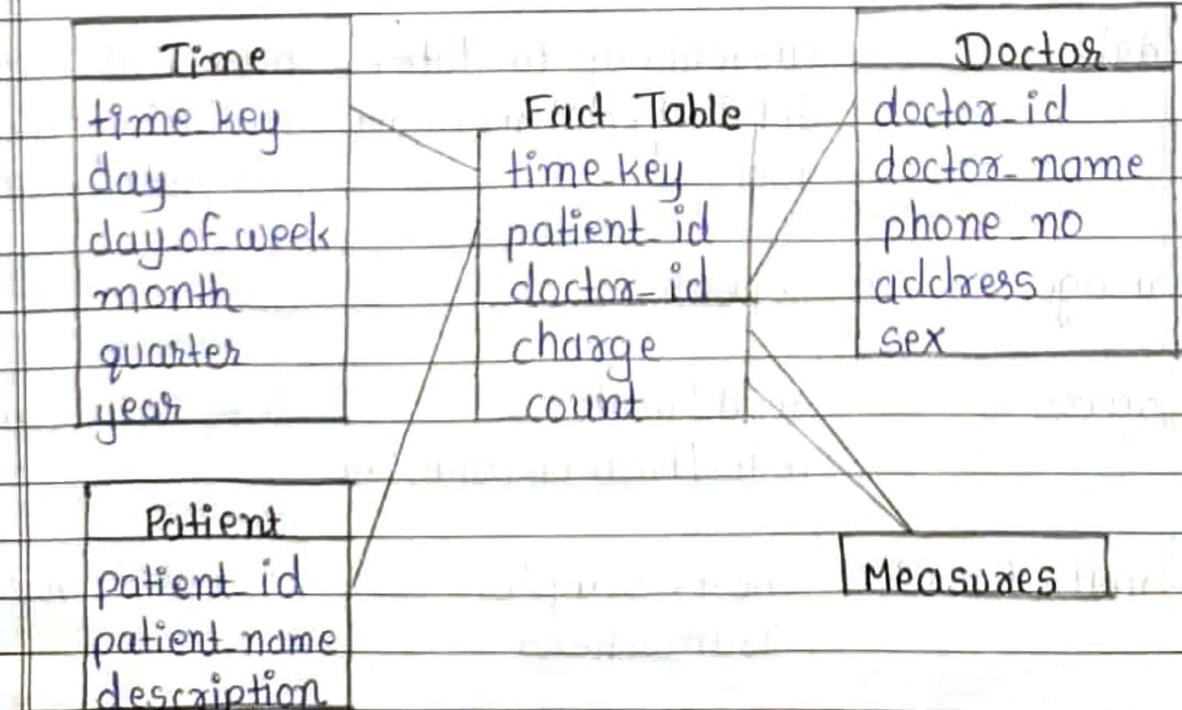
Ques-5 Compare an Online Transaction processing to an Online Analysis Processing.

	OLTP	OLAP
users	clerk, IT professional, knowledge worker	
function	day to day operations	decision support
DB design	application oriented	subject oriented
data	current, up-to-date, detailed, flat relational, isolated	historical, summarized, multidimensional, integrated, consolidated
usage	repetitive	ad-hoc
access	read / write index/hash on prim. key	lots of scans
unit of work	short, simple transactions	complex query

#record access	tens	millions
#users	thousands	hundreds
DB size	100 MB - GB	100 GB - TB
metric	Transaction throughput query throughput, response time	

Ques-6 Suppose that a data warehouse consist of three dimensions time, doctor and patient and the two measures count and charge, where charge is the fee that a doctor charges a patient for a visit.

1) Select appropriate schema diagram for the above data warehouse.



2) Starting with the base cuboid [day, doctor, patient] what specific OLAP operations should be performed in order to list the total fee collected by each doctor in 2004?

→ Roll-up from day → month → year

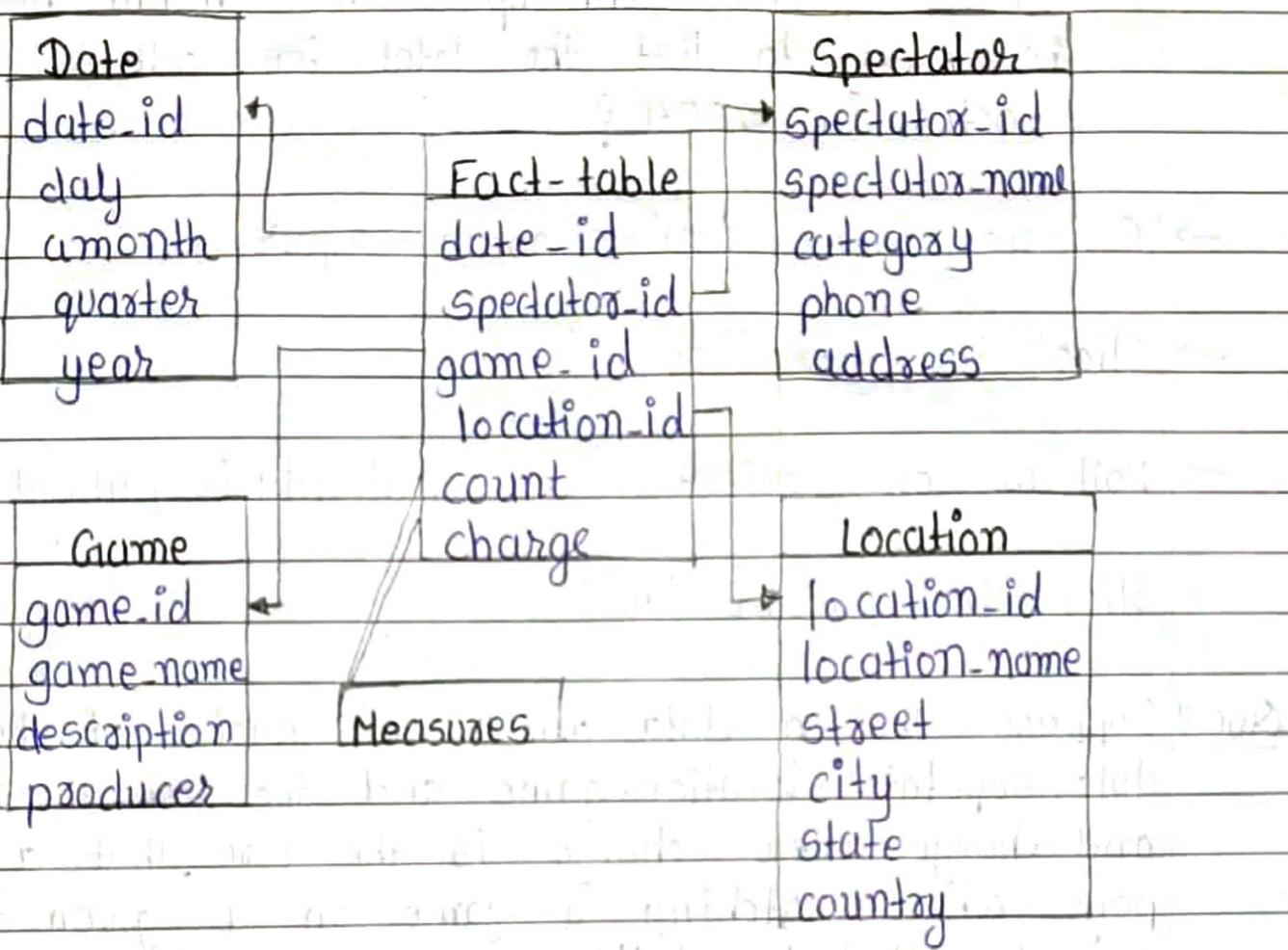
→ Slice for year "2004"

→ Roll-up on patient from individual patient to all

→ Slice for patient "all"

Ques-7 Suppose that a data warehouse consists of the four dimensions date, spectator, location, game and the two measures count and charge where charge is the fare that a spectator pays when watching a game on a given date. Spectator may be students, adults or seniors with each category having its own charge rate.

1) Design a star schema for the data warehouse.



2) Starting with base cuboid what specific OLAP operations should one perform in order to list the total charge paid by student at GM place 2004.

→ Roll up on spectator → category

→ Roll up on date → year

→ Roll up on game → all

→ Dice parameter category - 'student', location-name - 'GM', place', year - '2004'.