**Practice**

Q1. Suppose that a data warehouse for Big-University consists of the following four dimensions: student, course, semester, and instructor, and two measures count and avg grade. When at the lowest conceptual level (e.g., for a given student, course, semester, and instructor combination), the avg grade measure stores the actual course grade of the student. At higher conceptual levels, avg grade stores the average grade for the given combination.

(a) Draw a snowflake schema diagram for the data warehouse.

(b) Starting with the base cuboid [student, course, semester, instructor], what specific OLAP oper ations (e.g., roll-up from semester to year) should one perform in order to list the average grade of CS courses for each Big-University student.

(c) If each dimension has five levels (including all), such as "student < major < status < university < all", how many cuboids will this cube contain.

Answer:

a)
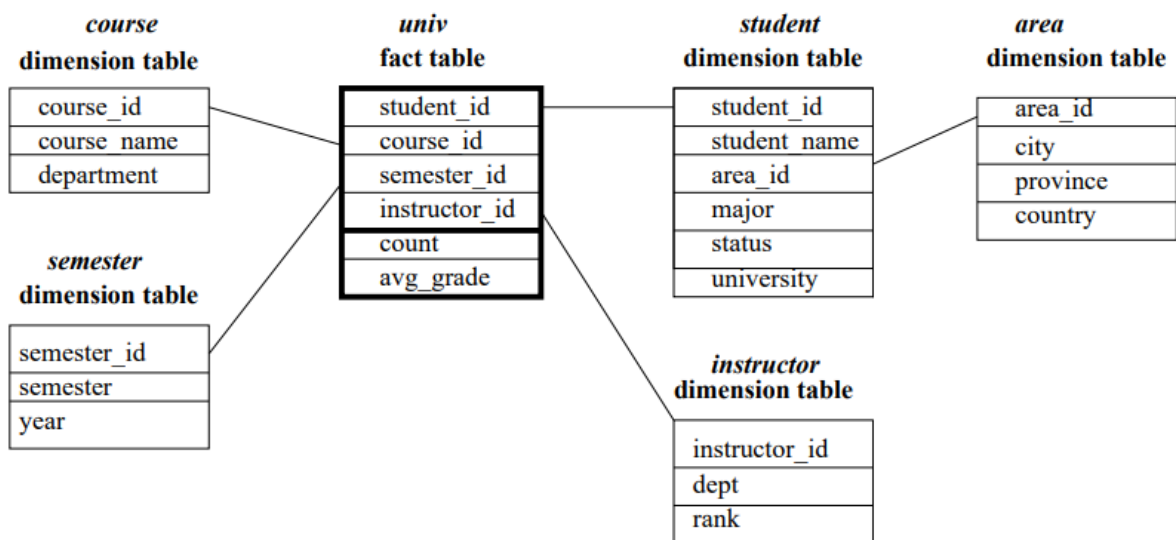
b) The specific OLAP operations to be performed are:
   • Roll-up on course from course id to department.
   • Roll-up on semester from semester id to all.
   • Slice for course="CS"
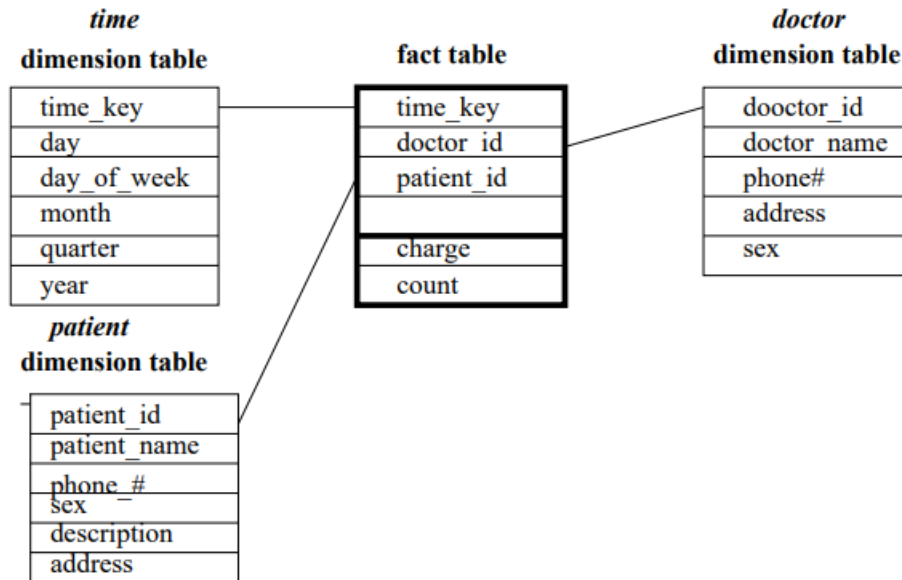c) This cube will contain 5^4 = 625 cuboids.

Q2. Suppose that a data warehouse consists of the 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.

(a) Enumerate three classes of schemas that are popularly used for modeling data warehouses.

(b) Draw a schema diagram for the above data warehouse using one of the schema classes listed in (a).

(c) 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?

Answer:

a) Three classes of schemas popularly used for modeling data warehouses are the star schema, the snowflake schema, and the fact constellation schema.

b)

time dimension table		*doctor* dimension table

fact table

| time_key |
| day |
| day_of_week |
| month |
| quarter |
| year |

| time_key |
| doctor_id |
| patient_id |
| |
| charge |
| count |

| dooctor_id |
| doctor_name |
| phone# |
| address |
| sex |

patient dimension table

| patient_id |
| patient_name |
| phone_# |
| sex |
| description |
| address |

c) The operations to be performed are:
   • Roll-up on time from day to year.
   • Slice for time = 2010.
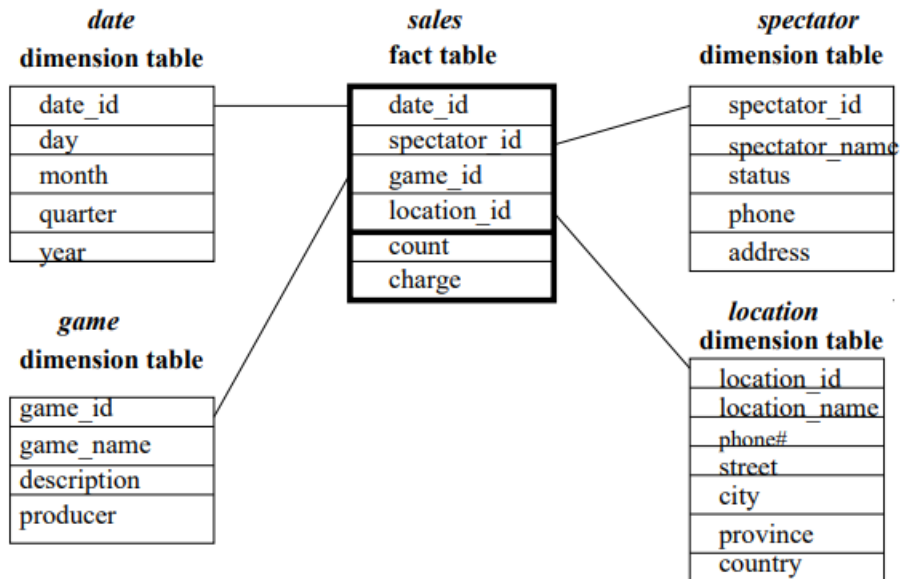   • Roll-up on patient from individual patient to all.

Q3. Suppose that a data warehouse consists of the four dimensions, date, spectator, location, and 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. Spectators may be students, adults, or seniors, with each category having its own charge rate.

(a) Draw a star schema diagram for the data warehouse.

(b) Starting with the base cuboid [date, spectator, location, game], what specific OLAP operations should one perform in order to list the total charge paid by student spectators at GM Place in 2010?
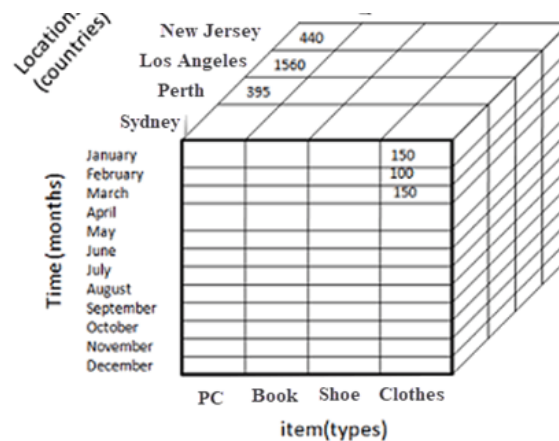
Answer:

a)

date
dimension table

| date_id |
| day |
| month |
| quarter |
| year |

sales
fact table

| date_id |
| spectator_id |
| game_id |
| location_id |
| count |
| charge |

spectator
dimension table

| spectator_id |
| spectator_name |
| status |
| phone |
| address |

game
dimension table

| game_id |
| game_name |
| description |
| producer |

location
dimension table

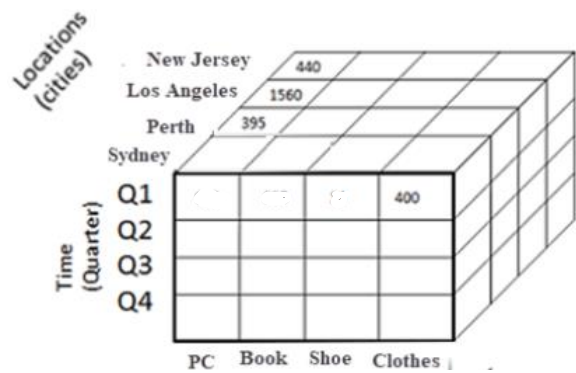| location_id |
| location_name |
| phone# |
| street |
| city |
| province |
| country |

b) The specific OLAP operations to be performed are:

- Roll-up on date from date id to year.
- Roll-up on game from game id to all.
- Roll-up on location from location id to location name.
- Roll-up on spectator from spectator id to status.
- Dice with status="students", location name="GM Place", and year = 2010

Q4. Perform roll up OLAP operation on given cube for the time dimension in quartly bases and pivot operation on your choice of dimension

Answer:

Roll up

Pivot