* Q4.3 – 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
    - 1. Star Schema
    - 2. Snowflake Schema
    - 3. Galaxy Schema (or Fact Constellation)
  + B) Draw a schema diagram for the above data warehouse using one of the schema classes listed in (a).
    - Star Schema Diagram:



* + 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 2010?
    - *Roll-up by dimension reduction* to remove *patient* dimension
    - *Roll-up time dimension* to *year*
    - *Slice* *time* dimension using criterion *time* = “2010”
  + D) To obtain the same list, write an SQL query assuming the data are stored in a relational database with the schema *fee* (*day, month, year, doctor, hospital, patient, count, charge*).

**SELECT** doctor, charge

**FROM** fee

**WHERE** year = 2010;

* Q4.4 – Suppose that a data warehouse for *Big\_University* consists of the four dimensions *student, course, semester,* and *instructor*, and two measures *count* and *avg\_grade*. 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 *operations* (e.g. roll-up from *semester* to *year*) should you perform in order to list the average grade of CS courses for each *Big\_University* student.
    - *Roll-up by dimension reduction* to remove *instructor* dimension
    - *Roll-up by dimension reduction* to remove *semester* dimension
  + C) If each dimension has five levels (including *all*), such as “*student* < *major* < *status* < *university* < *all*”, how many cuboids will this cube contain (including the base and apex cuboid)?

Total number of cuboids = 54 = 625 cuboids

* Q4.5 – 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, which 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 you perform in order to list the total charge paid by student spectators at *GM\_Place* in 2010?
    - *Roll-up by dimension reduction* to remove *game* dimension
    - Drill down *date* dimension to *year*
    - *Dice* operation based on the selection criteria of three dimensions, *spectator* = “students”, *location* = “GM\_Place”, and *year* = 2010
  + C) *Bitmap indexing* is useful in data warehousing. Taking this cube as an example, briefly discuss advantages and problems of using bitmap index structure.

Bitmap indexing is used to speed up processing of multidimensional data cubes. It does this by reducing operations to bit arithmetic. It is typically used in domains with low cardinality. In regard to this data cube, *spectator* would be a good example of low cardinality because it is limited to three distinct values. However, it is not recommended for domains with high cardinality.

