**4.1 Ans:**

The major difference between update-driven approach and query driven approach is in update-driven approach first the data collected from different heterogeneous systems and combined or summarized by solving all the semantic and syntactic related problems exists among the various databases, and then actual query processing will take place. Where as in query-driven approach the query processing willbe done directly on different databases simultaneously without actually summarizing the data. Here in this query driven approach, the multiple queries triggered on various databases and later combined. So basically during query execution on multiple databases all the execution engines are compete for the resources, degrade the performance and also the combined resultant output further need to be filtered and integrated.

Hence the update driven approach in data warehousing is more preferable than query driven approach.

Query driven approach is more preferable in the cases where the queries are completely rely on current data and don’t care about the data aggregations. And also applying the aggregations and spending significant amount of money in building the data warehouse before query execution where the databases are relatively small in size is not a very good idea as the queries are fully relied on current data and data warehouses don’t build to store current information.

**4.2 Ans:**

**a.** The snowflake schema and fact constellation schema both are under the hood of schema model. The common feature between two is both involves a fact table and a set of dimension tables. And the difference is in snow flake schema, all the dimension tables are normalized, whereas in fact constellation the common dimension tables are shared by a set of fact tables.

The starnet model showed in below figure is not the same as schema model it follows completely different architecture. It basically starts with a centre point from which all the radial lines are originated. The radial lines are demonstrated as dimensions and the points along the way of each radial line called as level of dimension. The overall model represents the concept of hierarchy as it moving towards the dimensions from centre point i.e. stepping towards the level of dimensions. It is mainly useful in OLAP operations.

**b.** The process of finding out the noisy and erroneous data, rectifying them is called Data cleaning. Data transformation, from the name itself suggests that transforming the data from one form to other. Here in Data warehousing study, data transformation is converting all the combined data coming from the different sources or heterogeneous sources to a unified format. One more action plays a very important role in data warehousing data transformation is refresh, which reflects all the data in different data sources to the data warehouse.

**c.**

The discovery driven data cubes are the data cubes with Online Analytical Processing techniques enabled. The main purpose of these cubes is , it has ability to compute the complex aggregations with ease and it even efficiently explores the space of the data cubes which is usually aggregated.

The main purpose of multi feature cubes is, when there is a need to perform complex queries on data cubes. It is useful in complex computation which involves multiple granularities.

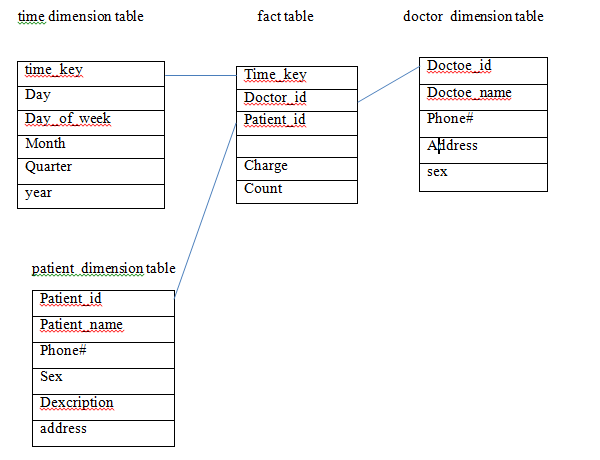
The virtual ware house mainly build upon set of view extracted from operational databases. Empirically the set of views extracted from the databases only few will be useful in real time scenario. It is easy to build but it puts an overhead on operational databases.

**4.3 Ans:**

**a.** There are mainly three classes of schemas are popularly used for modelling data warehouses.

* Star Schema
* Snowflake schema and
* Fact-Constellation schema

**b.**



**c.** The operations to be performed are:

• Generalize the data based on time from day to year

• select the 2004 time slice data.

• Generalize and gather the patient’s data from individual to all.

**d.**

select doctor, SUM(charge) as SUM\_CHARGE

from fee

where year=2004

group by doctor;