ALY6030 - Data Warehousing and SQL

CRN: 71719

Module 3 Assignment

By: Karan Daiya

Introduction:

This assignment contains tasks to design star schema where there is one fact table and few other dimensional tables. Star schema includes all measurement tables to be assorted to the fact table. Using this schema, we will construct ERD in an online tool called ERD Plus and aftyer that we will remake the whole schema in Workbench. This will make a database where we are able to import our information and run SQL queries. We are going to utilize given csv files that are business.csv, bed\_type.csv and bed.fact.csv.

Tasks:

Identify Dimensions:

Dimension tables contain columns that depict the fact records within the fact table. A few of these columns give descriptive data. Other columns specify how the information within the fact table is summarized to supply valuable data. Dimension tables contain hierarchies that offer assistance to summarize data. Dimension tables are smaller, denormalized lookup tables that contain descriptive columns simply reference after you define queries. Table “business” and “bed\_type” are the dimension tables from above given three tables. Business table has attributes such as,

ims\_org\_id

business\_name

ttl\_license\_beds

ttl\_census\_beds

ttl\_staffed\_beds

bed\_cluster\_id

From these, business\_name is one of the dimensions. The table “bed\_type” has following attributes,

bed\_id

bed\_code

bed\_desc

from which bed\_code and bed\_desc are the dimensions.

Identify Facts:

Fact table can store different types of measures to be added in the dimension such as additive, non-additive, semi-additive. From the above three, “bed\_fact” is the fact table and following are the attributes,

ims\_org\_id

bed\_id

license\_beds

census\_beds

staffed\_beds

From this, license\_beds, census\_beds and staffed beds are the facts.

Sketching a Star Schema:

Now, to design a Star Schema we will use ERDPlus tool. ERDPlus is a database modeling tool for creating Entity Relationship Diagrams, Relational Schemas, Star Schemas, and SQL DDL statements. Here, we will create Star schema in MySQL Workbench and will create the whole database through ER Diagram only by using Reverse Engineering Tool.

Graphical user interface, table

Description automatically generated

Here, I have created the above given three tables by just drag and drop tool of creating a new table. In bed\_type table I have assigned primary key to bed\_id and in business table I have assigned primary key to ims\_org\_id. Now we will assign the foreign keys and their relationship to the tables as shown below.

Graphical user interface, text, application

Description automatically generated

Diagram

Description automatically generated

Now finally we will use the Table Import Wizard to add the table file.

Graphical user interface, text, application

Description automatically generated

Questions:

**1) List the top 10 hospitals with most licensed beds.**

SELECT business.business\_name AS healthcare\_centre,

SUM(bed\_fact.license\_beds) AS licensebed\_sum

FROM business INNER JOIN bed\_fact

ON business.ims\_org\_id = bed\_fact.ims\_org\_id

WHERE bed\_fact.bed\_id IN(4,15)

ORDER BY SUM(bed\_fact.license\_beds) DESC LIMIT 10;

**2) List the top 10 hospitals with highest number of beds for staff members.**

SELECT business.business\_name AS healthcare\_centre,

SUM(bed\_fact.staffed\_beds) AS staffbed\_sum

FROM business INNER JOIN bed\_fact

ON business.ims\_org\_id = bed\_fact.ims\_org\_id

WHERE bed\_fact.bed\_id IN(4,15)

GROUP BY business.business\_name

ORDER BY SUM(bed\_fact.staffed\_beds) DESC LIMIT 10;

**3) List the top 10 hospitals with most census beds.**

SELECT business.business\_name AS healthcare\_centre,

SUM(bed\_fact.census\_beds) AS censusbed\_sum

FROM business INNER JOIN bed\_fact

ON business.ims\_org\_id = bed\_fact.ims\_org\_id

WHERE bed\_fact.bed\_id IN(4,15)

GROUP BY business.business\_name

ORDER BY SUM(bed\_fact.census\_beds) DESC LIMIT 10;

Conclusion:

A star schema comprises of fact tables and dimension tables. Fact tables contains the quantitative or genuine information around a business–the data being questioned. This data is often numerical, added substance measurements and can comprise of numerous columns and millions or billions of lines. Dimension tables are usually smaller and hold expressive data that reflects the dimensions, or attributes, of a trade. SQL questions at that point utilize joins between fact and dimension tables and constraints on the data to return chosen information.