ALY 6030 - Data Warehousing & SQL - 70526  
Module 3: Dimensional Modeling

Professor: Valeriy Shevchenko

CPS in Informatics

Chongfan Bao

bao.c@northeastern.edu

Northeastern University

10/11/2021

**Abstract**

In project 3, I am about to endeavor to design a star schema based on two dimension tables and a fact table that describe the number of various beds situation in various hospitals. Then, we will discuss the top 10 hospitals for the different bed types of situation by leveraging the MySQL workbench and SQL language. Finally, I need to explain my design ideas and answer clients’ questions.

*Keywords:* star schema, SQL

**Introduction**

In this module 3, I will use the hospital dataset that depicts the number of various categories of beds in different hospital center to conduct data analysis. There are three categories’ beds: licensed beds, census beds and staffed beds. In addition, the bed has 20 sub-types. We can use bed id to distinguish. Next, I am going to draw a star schema in ERD Plus tool. According to this schema, I will build an ERD in MySQL workbench and leverage SQL syntax to query the top 10 hospital beds situations.

**Modeling**

According to two dimension tables and one fact tables, I notice that these tables include the number of beds, hospital name, type of bed, and so on. Also, I find that each table has at least one id column, such as bed\_id, ims\_org\_id, and bed\_cluster\_id, which can help me set as a primary key or foreign key and build dimension tables to connect the fact table. Each id can be as a dimension table. However, in a star scheme that we will draw, it consists of one fact table and two dimension tables. This is because when I import data to MySQL workbench, I cannot find a good corresponding data from the parent table and child table.

From figure 1, I created an ERD diagram, which includes a fact table named hospital bed that has two foreign keys and bed metrics. The ims\_org\_id is both the primary key and the foreign key. These keys are used to connect to the dimension tables. Bed classification dimension table is used to hold the data that describes the type of bed in hospital. Hospital information dimension table is used to hold hospital name and total number of bed data.

Based on this star schema, I will create a database in MySQL workbench and import original dataset from table data import wizard. The database is made of three tables corresponded to one fact table and two dimension tables. Each column represents a attribute in the star schema diagram.

Diagram

Description automatically generated

Figure 1

From figure 2, people can know that I created a hospital\_database\_week3 database that includes three tables: hospital\_bed\_fact, bed\_classfication and hospital information to hold original dataset. Meanwhile, I set up a primary key for each table and the column attribution. However, the table doesn’t have any value and relationship. So, I will set up foreign keys to build relationship in the next step. The primary key cannot be null and other column can be null value

Graphical user interface, text

Description automatically generated

Figure 2

Based on the parent table that I created, we set up two foreign keys: bed\_id\_fk and ims\_org\_id\_fk, which references the bed\_classification table and hospital\_information table seperatively. Also, I set up constrains when we need to manipulate update and delete syntax.

A screenshot of a computer

Description automatically generated with medium confidenceA screenshot of a computer

Description automatically generated with medium confidence

Figure 3

From figure 4, I attempt to import dataset into MySQL workbench. Firstly, we need to choose the file path to import data. I choose the bed\_type file as the original dataset.

Graphical user interface, text

Description automatically generated

Figure 4

From figure 5, I choose the exit table bed\_classfication’s to import date from bed\_type file as original dataset. Meanwhile, I choose truncate table before import.

Graphical user interface, text

Description automatically generated

Figure 5

From figure 6, I can select columns that I would like import data. The original data table column will be align with destination table column that I created in MySQL Workbench.

Graphical user interface, application

Description automatically generated

Figure 6

The figure 7 shows data import results. People can notice how many records are imported. Also, based on the number of records, we can know if the data is imported successfully.

Graphical user interface, text

Description automatically generated

Figure 7

As other two tables, we can follow the same steps to import data. We just need to change the file path and choose the existing table to align with the path file. Then, we select the column that we want to import.

In figure 8, we can find that we have successfully imported the original dataset into MySQL workbench. In the hospital\_bed\_fact table, we import 23135 records. In the hospital\_information table, we import 23147 records. After this, I am able to conduct SQL syntax to export the top 10 hospital for different bed.

Text

Description automatically generated

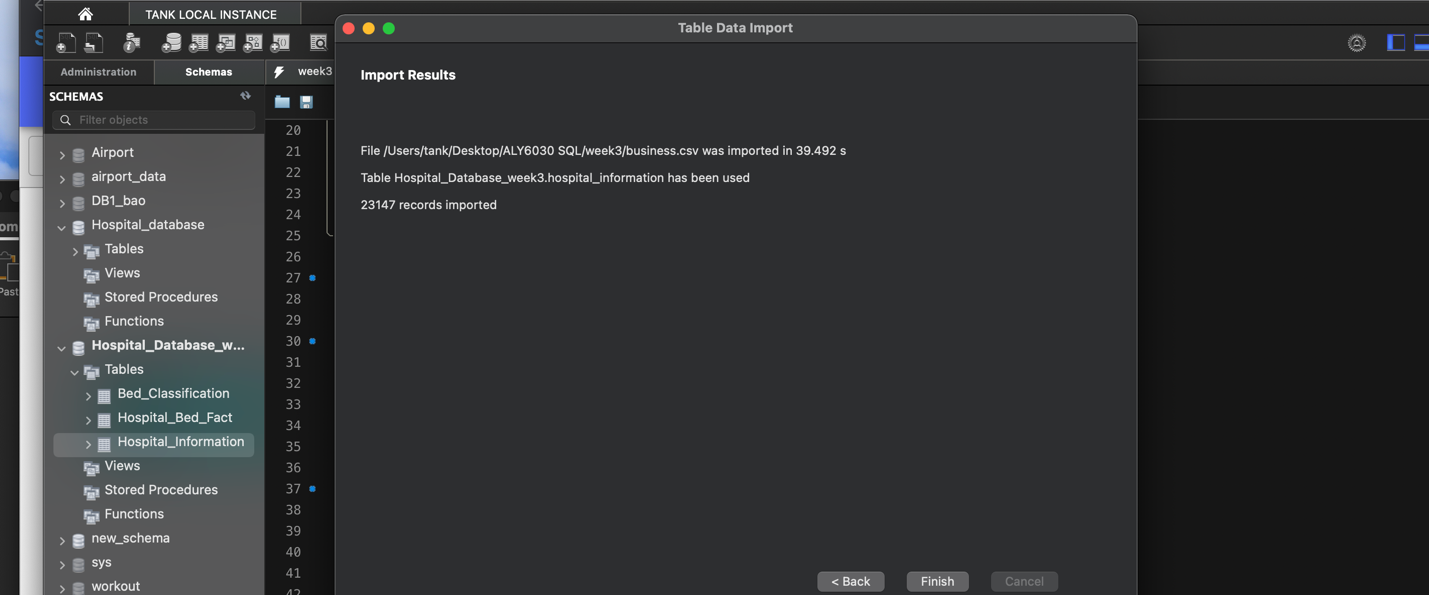


Figure 8

According to figure 9, we can notice that the Saint Marys Hospital has the most ICU type beds. The license\_beds is 264; The census\_beds is 147; The staffed\_bed is 219. At same time, the Brigham and Womens Hospital has the most ICU. The license\_beds is 138; The census\_beds is 124; The staffed\_bed is 138. If we want to check different bed type, we just change the bed\_id number or we can list all type beds. But it will cause too large table and not easy to check.

Graphical user interface, application, table

Description automatically generated

Figure 9

**Conclusion**

In the whole project, I drew a star schema. Based on the schema, I built database to store hospital beds dataset. Then, I used the SQL syntax to query the top 10 hospital for different beds and display one single table. So people can intuitively find which hospital has more bed available.

By this project, the biggest challenge is that I don’t know how to use the case syntax correctly. I made a lot of mistakes. For instance, when I try to use select query with case query, we need use a comma after the select query. Otherwise, the system will reminder the mistakes. There is a red underline under the “case” syntax so that I cannot run the code. However, I only leverage the select query. I don’t need the comma and it works. In addition, if I separate two dimension tables to more than two dimension tables when I draw a star schema, it will be difficult for me to import the same ID of key in parent table and child table to ensure that the data is connected together. However, I am did more good at use the MySQL workbench and SQL syntax because I will repeat steps that I have done in before assignments. In the fugure, I believe that I will be able to use MySQL professionally to solve any business questions.

**References**

WHERE is not valid at this position, expecting EOF.Retrieved from https://cursos.alura.com.br/forum/topico-where-is-not-valid-at-this-position-expecting-eof-132803