

**By**

**Sunil Raj Thota  
ALY6030 - Data Warehousing and SQL**

**Class Name: Fall 2021 CPS Quarter**

**Class Number: 202215**

**CRN: 70526**

**Week 3 Assignment 3**

**thota.su@northeastern.edu**

**Introduction:**

In this Assignment, As a consultant engaged by a healthcare network, we are tasked with compiling a list of the top 10 hospitals with Intensive Care Units (ICU) and Surgical Intensive Care Units (SICU) based on the total number of licensed, census, and staffed beds. The information for the study is given in three separate CSV files named business.csv, bed\_type.csv, and bed\_fact.csv accordingly. We must establish a database and analyze the generated results based on the provided requirements by specifying the primary and foreign keys. I have also shared each step with dedicated screenshots and comments where ever required. I aim to perceive these ideas by using various sorts of joins, aggregating methods, and variables. For this report, I have SQL as the language to interprets the results by creating a database schema and defining the table attributes with their data types.

**Analysis:**

**Step 1: Identify the dimensions (first three only) from the two-dimension tables**

A dimension table is a table in a data warehouse's star schema. The properties, or dimensions, that describe the items in a fact table are stored in a dimension table. The dimensions derived from the two data sets are as follows:

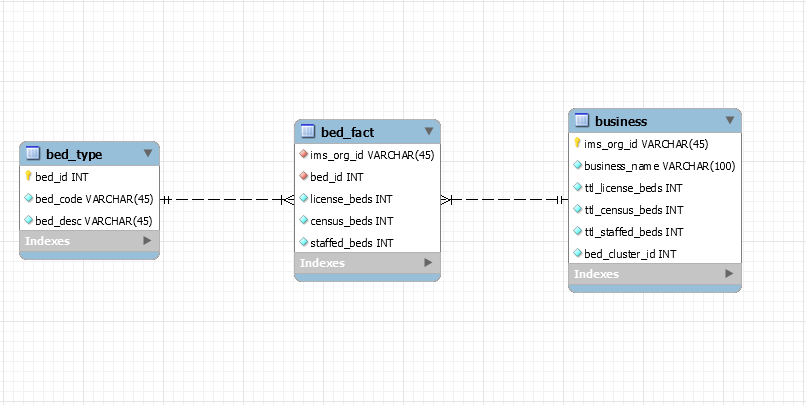
* Ims\_org\_id, business\_name, and bed\_cluster\_id for 'business’ table
* Bed \_id, bed\_code, and bed\_desc are the columns in the 'bed\_type' table

**Step 2: Identify the Facts (first three only) from the single fact table**

A fact table is the core table of a data warehouse's star schema. A fact table is a denormalized table that stores quantitative data for analysis. A fact table is used in conjunction with dimension tables. 'license\_beds,' 'census\_beds,' and 'staffed\_beds' are the facts in the 'bed\_fact' database.

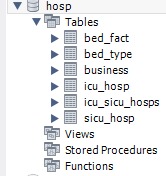
**Step 3: Sketch out a Star Schema**

A star schema is a data warehousing model that has one or more fact tables at the center and indexes any number of dimensional tables.



**Step 4: Implement the schema as a database**

I’ve created a new schema called 'hosp’ as a database. A screenshot of the same can be found below.



**Step 5: Write and run SQL code for the database that will provide answers for the questions posed below:**

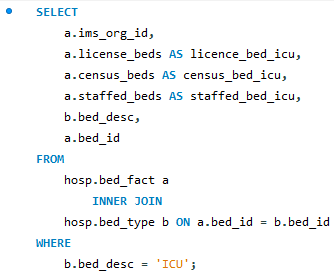
**Construct a list of top 10 Hospitals/Medical Centers that have SICUs (Surgical Intensive Care Units) and ICUs (Intensive Care Units) for the following categories of "beds" (one bed can accommodate one patient):**

1. **Total licensed beds (total beds allowed by state license)**
2. **Total census beds (total beds at the hospital)**
3. **Total staffed beds (total beds for which staffing, e.g., physicians and nurses, exists)**

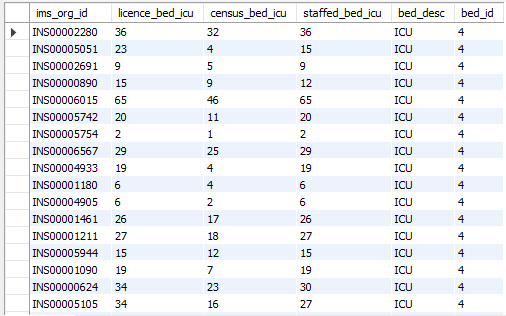
**Solution:**

1. **ICU**

**Code:**

****

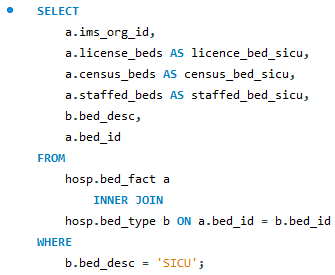
**Results:**

****

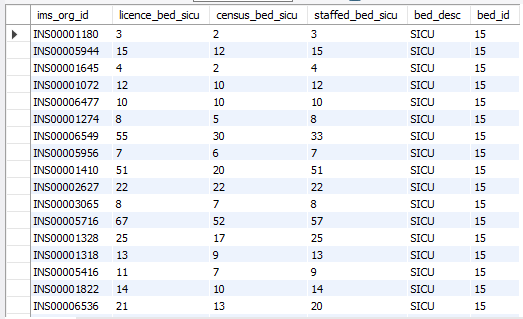
We utilized the INNER JOIN on two tables, ‘bed\_type' and ‘bed\_fact,' to acquire the output of all the hospitals in the US that have an ICU. As a consequence of the query above, we obtain 1000 rows as a result.

1. **SICU**

**Code:**

****

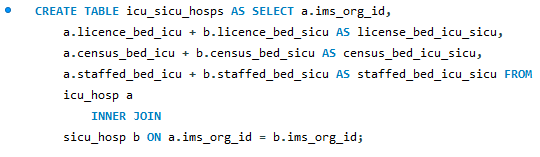
**Result:**

****

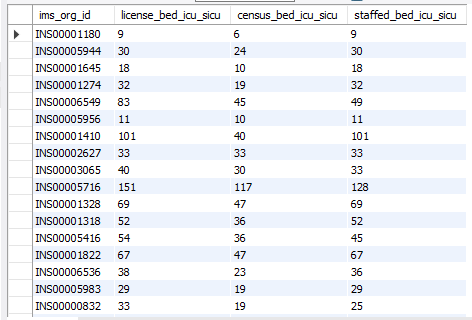
We utilized the INNER JOIN on two tables, ‘bed\_type' and ‘bed\_fact,' to acquire the output of all the hospitals in the US that have an SICU. As a consequence of the query above, we get 225 rows as a result.

1. **ICU and SICU**

**Code:**

****

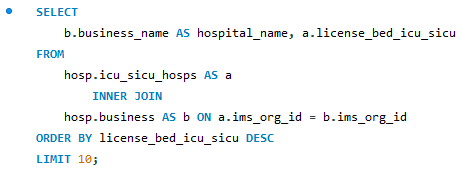
**Result:**



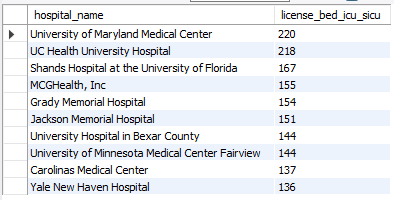
We get 194 items as a result of the above query, which is common to both ICU and SICU. We also added the sum of the two bed\_id's composed to get the total number of license beds, census beds, and staffed beds.

1. **List of top 10 hospitals with licensed beds ordered descending by the sum of ICU license beds + SICU license beds**

**Code:**

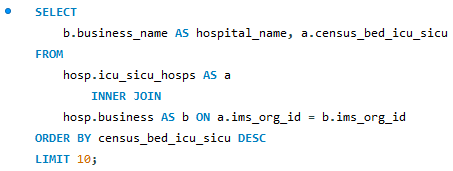


**Result:**

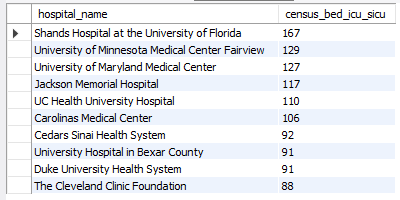


1. **List of top 10 hospitals with census beds ordered descending by the sum of ICU census beds + SICU census beds**

**Code:**

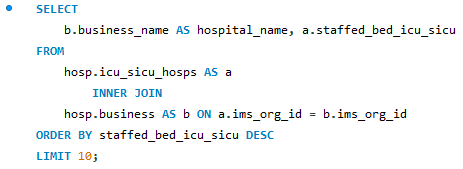
****

**Result:**

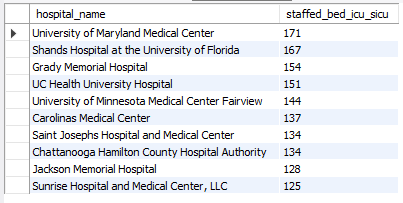
****

1. **List of top 10 hospitals with staffed beds ordered descending by the sum of ICU staffed beds + SICU staffed beds**

**Code:**

****

**Result:**

****

**Step 6: Include an interpretation and explanation of all the results for your discussions with the client who hired you.**

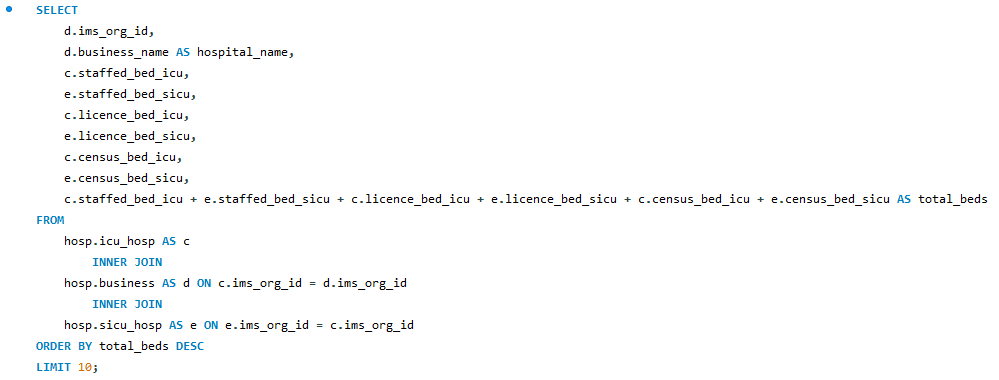
**Solution:**

We may deduce the following from the results received from the three output tables:

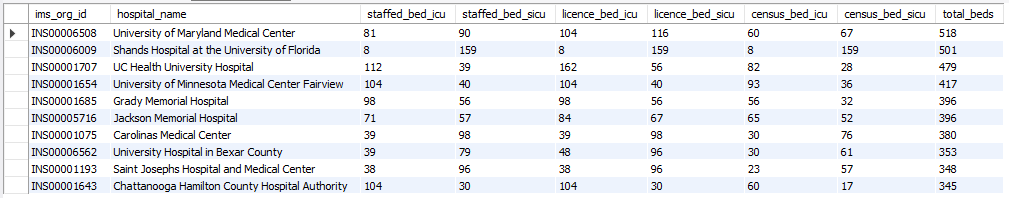
* The University of Maryland Medical Center has the greatest number of ICU and SICU licenses and staffed beds
* The Shands Hospital at the University of Florida has the largest number of total ICU and SICU census beds. It's also in the top three for having a license and having staffed beds
* University of Maryland Medical Center, UC Health University Hospital, Shands Hospital at the University of Florida, Jackson Memorial Hospital, University of Minnesota Medical Center Fairview, and Carolina Medical Center were all mentioned in the top ten rankings. For license beds, census beds, and staffed feds for ICU + SICU beds, these facilities were among the top ten
* University of Maryland Medical Center, UC Health University Hospital, Shands Hospital at University of Florida, Jackson Memorial Hospital, University of Minnesota Medical Center Fairview, Carolina Medical Center, Grady Memorial Hospital, MCG Health, Inc., and University Hospital in Bexar County were all listed in multiple top 10 lists

**Step 7: Counts of each bed type for each for the Top 10 hospitals per Story section (this is just a subset of the entire set of results)**

**Code:**

****

**Result:**

****

**Conclusion:**

I am glad to choose this dataset because it allowed me to experiment with various methods for creating tables and populating records, such as joins and defining the Primary Keys, Foreign Keys, Star Schema ER. As a result of the aforementioned study, the top ten hospitals for each type of bed (staffed bed, licensed bed, and census bed) for the two categories of ICUs (intensive care unit and surgical intensive care unit) were identified. The tables were built using the inner joints to define the top 10 hospitals succinctly by using significant aspects out of each table. The University of Maryland appeared to have the most beds in each category of intensive care unit. A supplementary table was constructed that shows the top 10 hospitals based on the total number of various types of beds throughout all departments. I'd like to try my hand at more complex datasets with advanced star schema designs.