ASSIGNMENT 2

Payal Sharma

Hospital ICU/SICU Bed Analysis

Northeastern University: College of Professional Studies

ALY6030: Data Warehousing and SQL

Nov 23rd, 2025

1. Dimensions in the Dimension Tables

1.1 bed_type table

Columns in bed_type:

- bed_id
- bed_code
- bed_desc

In dimensional modeling terms:

- A **fact** is usually a numeric measure that we can aggregate (sum, average, etc.).
- A **dimension** is descriptive information that gives context to the facts, such as IDs, codes, or text descriptions.

For bed_type:

- **bed_id** – This is the primary key that uniquely identifies each bed category (for example, ICU or SICU). It is a dimension key.
- **bed_code** – This is a short code for the bed type (for example, "IC" or "SI"). It is descriptive, so it behaves as a dimension attribute.
- **bed_desc** – This is a longer text description of the bed type (for example, "Intensive Care Unit"). This is also a dimension attribute.

Conclusion for bed_type:

- All three fields in this table are **dimensions**: bed_id, bed_code, and bed_desc.
- There are **no fact variables** in bed_type, because none of the columns is a numeric measure that we would aggregate.

1.2 business table

Columns in business:

- ims_org_id
- business_name
- ttl_license_beds
- ttl_census_beds
- ttl_staffed_beds
- bed_cluster_id

How each one behaves:

- **ims_org_id** – Unique identifier for each hospital. This is the primary key and acts as a **dimension key**.

- **business_name** – The hospital name. This is descriptive, so it is a **dimension attribute**.
- **bed_cluster_id** – A categorical grouping based on beds. This is also descriptive and works as a **dimension attribute**.

The remaining three are clearly measures:

- **ttl_license_beds** – Total licensed beds in the hospital (numeric, can be summed). This behaves like a **fact**.
- **ttl_census_beds** – Total census beds (numeric). Also a **fact**.
- **ttl_staffed_beds** – Total staffed beds (numeric). Again a **fact**.

Conclusion for business:

- The three **dimension variables** I am selecting (as the assignment asks) are:
    1. ims_org_id – primary key / dimension key
    2. business_name – descriptive dimension
    3. bed_cluster_id – categorical dimension
- The **fact-type variables** in this table are:
    - ttl_license_beds
    - ttl_census_beds
    - ttl_staffed_beds

---

2. Fact Variables in the Fact Table (bed_fact)

Columns in bed_fact:

- ims_org_id
- bed_id
- license_beds
- census_beds
- staffed_beds

Interpretation:

- **ims_org_id** – Foreign key linking to business. This tells us *which* hospital the row belongs to. It is a **dimension key**, not a fact.
- **bed_id** – Foreign key linking to bed_type. This tells us *which bed type* (ICU, SICU, etc.) the row is about. Again, this is a **dimension key**.
- **license_beds** – Number of licensed beds of that type at that hospital. Numeric and can be aggregated → **fact**.

- **census_beds** – Number of census beds of that type at that hospital. Numeric and aggregatable → **fact**.
- **staffed_beds** – Number of staffed beds of that type at that hospital. Numeric and aggregatable → **fact**.

Conclusion for bed_fact:

- **Dimension fields inside the fact table:** ims_org_id, bed_id
- **Fact fields (three requested):**
    1. license_beds
    2. census_beds
    3. staffed_beds

These are the measures we sum when building the Top 10 hospital lists.

---

3. Star Schema Structure (Conceptual Explanation)

For this assignment, the data model naturally forms a **star schema** with:

- **One central fact table:** bed_fact
- **Two dimension tables:** business and bed_type

Fact table: bed_fact

- Keys: ims_org_id, bed_id
- Facts: license_beds, census_beds, staffed_beds

Each record in bed_fact answers questions like:

"How many ICU/SICU beds of a given type does this hospital have (licensed, census, staffed)?"

Dimension table: business

- Key: ims_org_id (primary key)
- Main attributes used as dimensions: business_name, bed_cluster_id
- It gives hospital-level context such as the hospital's name and cluster.

Dimension table: bed_type

- Key: bed_id (primary key)
- Attributes: bed_code, bed_desc
- It tells us what type of bed we are looking at (ICU, SICU, etc.).

Relationships:

- bed_fact.ims_org_id → business.ims_org_id
- bed_fact.bed_id → bed_type.bed_id

If I draw this in MySQL Workbench using reverse engineering:

- bed_fact sits in the center as the fact table.
- business and bed_type are connected around it as dimension tables.

That is exactly the classic star pattern: one central numeric table, surrounded by descriptive tables.

---

4. High-Level Summary of the Top-10 Outputs

(The actual numbers and rows come from the SQL queries. Here I am just explaining what those results mean, which is what leadership cares about.)

4.1 Hospitals with ICU or SICU beds (either type)

In step 6a, I looked at all hospitals that have **ICU beds** (bed_id = 4), **SICU beds** (bed_id = 15), or both. For those hospitals, I created three separate Top 10 lists:

- Top 10 by **total ICU/SICU licensed beds**
- Top 10 by **total ICU/SICU census beds**
- Top 10 by **total ICU/SICU staffed beds**

Across these lists, a relatively small group of large hospitals consistently appears at the top. Examples include:

- University of Maryland Medical Center
- Vidant Medical Center
- Phoenix Children's Hospital
- UC Health University Hospital
- The Methodist Hospital
- Shands Hospital at the University of Florida
- Dallas County Hospital Association

These hospitals show up on multiple lists, which tells me they handle a large share of ICU/SICU bed volume when we look at the system as a whole.

4.2 Hospitals that have both ICU and SICU bed types

In step 7a, the focus shifts to hospitals that have **both** bed types:

- At least one ICU bed (bed_id = 4)
- At least one SICU bed (bed_id = 15)

For this narrower group of hospitals, I again created three Top 10 lists:

- Top 10 by total ICU/SICU licensed beds

- Top 10 by total ICU/SICU census beds
- Top 10 by total ICU/SICU staffed beds

From this drill-down, a core set of hospitals shows up again and again, such as:

- University of Maryland Medical Center
- Shands Hospital at the University of Florida
- UC Health University Hospital
- University of Minnesota Medical Center Fairview
- Jackson Memorial Hospital
- Carolinas Medical Center
- Grady Memorial Hospital

These hospitals are not only high-volume overall, but they also run both ICU and SICU units, which makes them especially relevant for a staffing intervention focused on critical care.

---

5. Interpretation for Leadership (Step 6b)

Here is how I would explain the findings to leadership in plain language.

1. **Critical care capacity is concentrated in a limited number of hospitals.**
   The Top 10 lists for licensed, census, and staffed ICU/SICU beds are dominated by a small set of large tertiary and academic hospitals. These facilities carry a big portion of the intensive care workload in the system.
2. **Some hospitals are consistent high performers across multiple measures.**
   Certain hospitals, like **University of Maryland Medical Center**, **Vidant Medical Center**, and **Shands Hospital at the University of Florida**, appear in the Top 10 for several of the metrics. That consistency suggests that they have:
   - Large numbers of ICU/SICU beds
   - High occupancy or census levels
   - Significant staffing dedicated to these units
3. **Why this matters for the nurse staffing intervention.**
   If leadership wants to test whether adding nurses improves outcomes and operations, it makes the most sense to run the pilot in hospitals that:
   - Already handle a large volume of ICU/SICU patients
   - Have both ICU and SICU units in place
   - Can generate enough data in a reasonable time frame

The high-volume hospitals identified in the Top 10 lists meet these criteria. A change in staffing at these sites would affect a meaningful number of patients and give leadership strong evidence to evaluate whether the intervention is worth scaling.

---

6. Final Recommendation to Leadership (Step 7b)

Leadership asked for **one or two** pilot hospitals. Based on all the analyses (both the "ICU or SICU" lists and the "both ICU and SICU" lists), my recommendation is:

Recommended pilot site 1: University of Maryland Medical Center

Reasons:

- It appears in the Top 10 across almost every list:
    - ICU/SICU licensed beds
    - ICU/SICU census beds
    - ICU/SICU staffed beds
    - And within the smaller group of hospitals that have both ICU and SICU bed types.
- Its ICU/SICU bed counts are consistently high, which indicates a large and steady critical care workload.
- Implementing the nurse staffing intervention here is likely to:
    - Affect a significant number of patients
    - Generate enough before-and-after data to judge the impact clearly

Overall, this hospital is an excellent "flagship" site for the pilot.

Recommended pilot site 2: Shands Hospital at the University of Florida

Reasons:

- Shands also ranks high in several Top 10 lists, especially for ICU/SICU census and staffed beds, and it has both ICU and SICU units.
- Its licensed, census, and staffed ICU/SICU bed numbers are all relatively high and balanced, suggesting a mature and busy critical care operation.
- Running the pilot here provides:
    - A strong comparison to University of Maryland Medical Center
    - Insights into how intervention works in a different region and patient population

---

Overall summary of the recommendation

I recommend that leadership select:

- **University of Maryland Medical Center**, and
- **Shands Hospital at the University of Florida**

as the two pilot hospitals for the nurse staffing intervention.

These hospitals:

- Have large ICU and SICU capacities
- Show consistently high ICU/SICU activity across licensed, census, and staffed beds
- Are likely to generate clear, interpretable results that leadership can use to decide whether and how to roll out the intervention more broadly.