**Data Mart Project: Neuroscience Research Laboratory**

Huy Nguyen

Graduate SBGE, Seattle Pacific University

ISM 62255: KMBI

Dr. Ryan LaBrie

May 9, 2023

**Business Statement**

The two provided data marts contain current research results and rat weight data respectively. This is a proof of concept for a dedicated database to increase data accessibility to international members of the lab. The rest remains the same, I’ll include the original business problem that this database project seeks to solve:

*Our lab has always and continues to use physical records and Slack – a group chat application for businesses – as our main way of communicating and storing research data. Only when the research calls to use specialty equipment with a built-in database that we start to utilize them; even then, the usage ends when said project ends.*

*This approach has several major problems: (1) Messy archives due to the dispersal of data, (2) over-reliant on Slack, which is a suboptimal tool for keeping track of scheduling and progress.*

*The lack of a centralized database means that record keeping is all over the place. Data can either be stored in various physical binders, text messages, and/or built-in databases of proprietary equipment. While this might not be a problem in the short-term, future data retrieval will be extremely difficult. Furthermore, this unorganization of data might potentially be confusing to new team members.*

*Slack, despite being custom-built for teamwork, at its core is still a text message app. Therefore, any form of data storing will be lost with time as new messages pile up the old ones. Additionally, the free version only allows access to the latest 10, 000 messages which again, might pose some problems down the line.*

*The lab desperately needs to centralize its data to improve data preservation, scheduling efficiency, and clarity for its members.*

**Data Mart Structures**

Due to the complexity of my original ERD, I had to perform some downscaling to avoid any potential problem in converting the ERD to a star schema. I basically created the star schema in Microsoft Access before porting it into SQL Server 2017.

**Facts/Measures**

The only two measures that the original data have were (1) weight data and (2) research results, so I just ported both of them onto SQL Server 2017. It is note-worthing that research results contain results from four separate methods; however, it was easier for me to keep them as one column due to (1) the small scope of both projects and (2) reduce dimension redundancy since they would all share the same dimensions due to the supertype/subtype nature of the original database.

Both measures share the same dimensions due to the aforementioned supertype/subtype structure. I found that one can implement a star schema with multiple fact tables in it but decided against it to maintain the small manageable scope of this project. Consequently, I ended up with two similar cubes.

**Dimensions**

Dimensions include (1) rat information, (2) research methods, (3) time, and (4) researcher/research assistant performing the experiment(s). These dimensions are chosen because they are relevant to the research of our neuroscience lab. I simplify the schedule aspect of the ERD to an ideal condition that is convenient for the requirement of this assignment as opposed to the emulating approach I took last quarter.

**3Ws**

**What Went Right?**

The cube-building process itself was a bit too straightforward. It was an enjoyable process after the pain and frustration of installing the SQL Server and building the tutorial cube. I really do appreciate that you, Professor LaBrie, taking it easy on us.

**What Did Not Go Right?**

Not knowing the details of this assignment, the original ERD was not as straightforward to convert to a star schema as I would like. Additionally, I also manually input data to populate the time dimension. Thankfully, I was able to speed up the process using Excel. That said, it is nothing compared to the data cleaning/transforming process for data mining.

**What Would I Do Differently?**

In an ideal condition where building data mart(s) was the sole focus of this course, I would actually attempt to convert my original ERD to a star schema as well as attempt to build a multiple-fact-tables cube. Other than that, I am quite please with the results.