Luis Medina

012020203

Data Management – D597

Task 1: Relational Database Design and Implementation

**Part 1: Design Document**

**Part A:**

1. Describe a business problem that can be solved with a database solution and is in alignment with the chosen scenario.

HealthFit Innovations has been experiencing some challenges with its increasing volume, variety and velocity of health-related data collected. The way the platform collects data is through various sources that include wearables, electronic health records (EHR’s), medical imaging systems, and patient-reported outcomes. With the current way it is, it lacks the ability to scale and flexibility, which can lead to problems. For example, problems with integration and data processing failures.

1. Purpose a data structure to solve the identified business problem.

To solve the identified business problem, a relation database structure is proposed and modeled through an Entity Relationship Diagram (ERD). By creating an ERD, it will be able to support scalable, secure, and flexible integration of the data collected from the health devices and patient-reported inputs. The database that will be created will include these tables to support the data, fitness records, patients, medical conditions, medications, allergies, appointments, and trackers. Each table will have a primary and most of the tables will have a foreign key to connect them to other tables. By doing this, it will reduce data redundancy, enforce data integrity and support fast querying regardless of the size of the data.

1. Justify why a database solution will solve the identified business problem.

Using a database solution will provide HealthFit Innovations with scalable, structured, and integrated solutions for managing lots of data, that will be collected from the wearables. It will also, improve performance and provide support with any analytical work for the future.

1. Explain how the business data will be used within the database solution.

All the data collected from the wearables that the customers are wearing, will have a fast and reliable way to store all the data that is gathered. It will also make easy to access the data using SQL commands to retrieve information that an analyst will need.

**Part B: Data Model**

**A screenshot of a computer

AI-generated content may be incorrect.**

**Part C: Objects and Storage**

**Part D: Scalability**

Like I mentioned before that as HealthFit’s user base continues to grow and the value collected from the devices increases, it’s extremely important that the database supports both vertical and horizontal scalability. As, the data continues to flow in from the devices the ERDs offer a great opportunity through the tables created can make the data more manageable, then with primary keys and foreign keys, all the tables will be connected making easy to gather or combine data. Another example is with the tables Health\_metric and activity\_log can be partitioned independently, as they will grow quickly.

**Part E: Privacy and Security**

Privacy and security are going to be one of the most important parts that need to be addressed, as the data being used is going to be personal and health-related data. Since, the data is sensitive data there needs to be limited access to the database, so that not just anyone can see private information.

**Part 2: Implementation**

**F1: Database Instance**

A screenshot of a computer

AI-generated content may be incorrect.

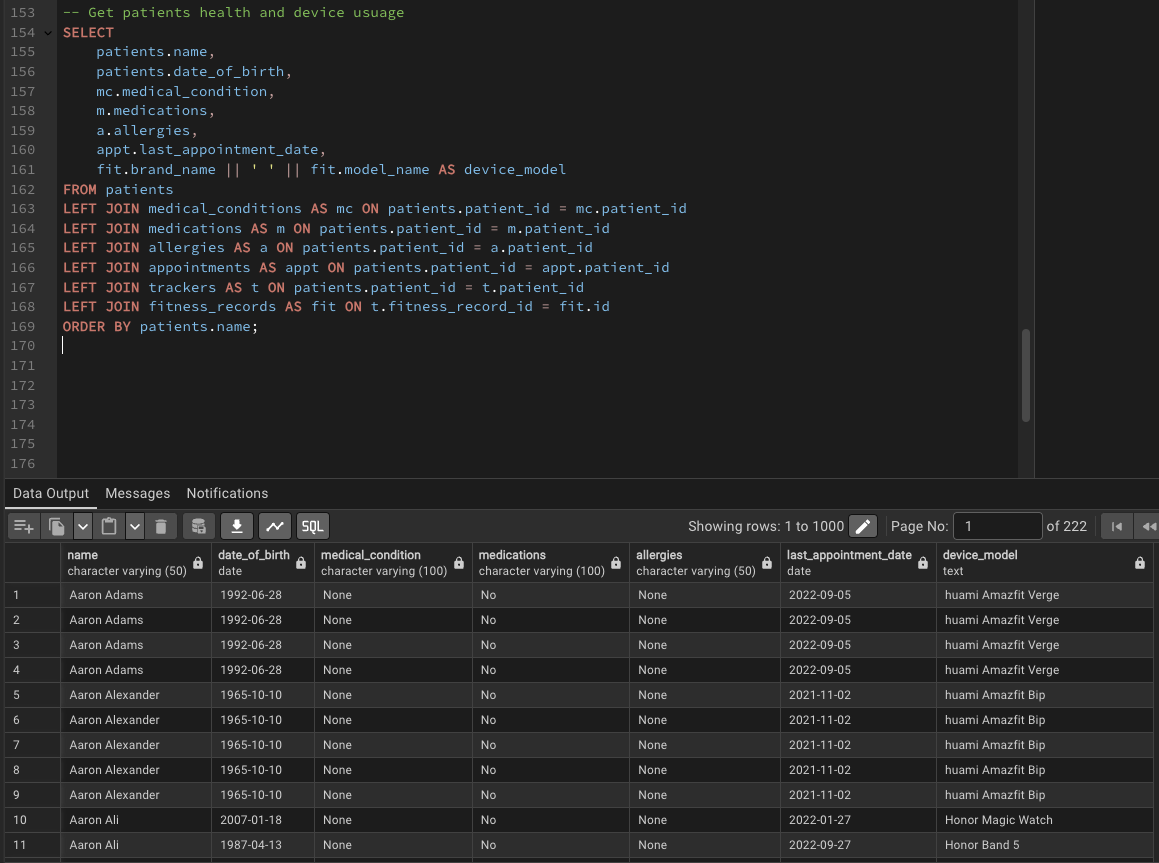
**F2: Insert Records**

Now I will create the SQL script to import the data records from the first scenario CSV files into the database.

A screenshot of a computer

AI-generated content may be incorrect.

**F3: Queries**



A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

**F4: Optimization**

**A screen shot of a computer program

AI-generated content may be incorrect.**

**A screenshot of a computer

AI-generated content may be incorrect.**

**A screenshot of a computer

AI-generated content may be incorrect.**

**A screenshot of a computer

AI-generated content may be incorrect.**

**H: Sources**

For the sources used it was mostly the course material that was provided and the webinars that I attend throughout the course.

Malik, Upom. *SQL for Data Analytics: Perform Fast and Efficient Data Analysis with the Power of SQL.,* Packt Publishing, 2019.