**徽标

中度可信度描述已自动生成**

**Final Project Assignment — Pharmacy Claims**

QINGRU YANG

ALY6030: Data Warehousing & SQL

Instructor: Zhi He

3/30/2022

**Part 1 Normalization**

Firstly, there are some attributes repeat in the dataset, therefore the dataset does not meet 1nf. (fill\_date1, fill\_date2, fill\_date3, copay1, copay2, copay3, insurancepaid1, insurancepaid2, insurancepaid3). We need to rearrange the data to eliminate duplicate attributes to meet 1nf.

Secondly, there is Partial Dependency, therefore the dataset does not meet 2nf. To meet 2nf, I divided the table into multiple tables, which are 2-dimension tables (dim\_member and dim\_drug\_ndc), and 1 fact table (fact\_drug).

Finally, to meet 3nf, divide dim\_drug\_ndc table into 2 tables (dim\_drug\_form and dim\_drug\_brand\_generic) to eliminate the Transitive Dependency.

Therefore, I divided into 1 fact table(fact\_drug) and 4-dimension tables (dim\_memebr,

dim\_drug, dim\_drug\_form, dim\_drug\_brand\_generic).

图形用户界面, 表格

描述已自动生成

**Questions:**

**Q1: For each fact variable in your fact table, what type of fact is it?**

In the fact table, the fact variables are copay and insurancepaid. They are **additive** because they can be used with any aggregation function.

**Q2: In your fact table, describe the grain in one sentence.**

Each fact row is the patient's drug fill record, as well as their payment history.

**Part 2 Primary and Foreign Key Setup in MySQL**

**Questions:**

**Q1: What are the primary keys you designated for each of your tables? For each PK, is it a natural key or a surrogate key?**

|  |  |  |
| --- | --- | --- |
| **Table** | **Primary Key** | **Type** |
| Fact\_drug | id | Surrogate key |
| Dim\_member | Member\_id | Natural key |
| Dim\_drug | Drug\_ndc | Natural key |
| Dim\_drug\_form | Drug\_form\_code | Natural key |
| Dim\_drug\_brand\_generic | Drug\_brand\_generic\_code | Natural key |

**Q2: What are the foreign keys you designated for each of your tables? For each FK, which table did you reference where that FK is listed as the PK?**

The member\_id is PK of the dim\_member table, and drug\_ndc is PK of dim\_drug\_ndc table. The drug\_brand\_generic\_code is PK of dim\_drug\_brand\_generic table, and drug\_form\_code is PK of dim\_drug\_form\_code. In fact\_drug table, the FK are member\_id, drug\_ndc, drug\_brand\_generic\_code and drug\_form\_code.

**Q****3: For each FK, what did you tell MySQL to in case of deletion or update (CASCADE, SET NULL, or RESTRICT)? Why did you select the option that you did for each FK?**

I choose the SET NULL option for each FK because for this option when we delete the parent table row, it will set the column value to NULL in the child table. In this case it is easier to make the change. However, when the parent changes, CASCADE will propagate the change. If we delete a row, the rows in the constrained table that reference that row will also be deleted. If there are child rows that reference the value of the parent row, RESTRICT will cause you to be unable to delete the given parent row.

**Part 3 Entity Relationship Diagram**

The ERD shows below:

**图示

描述已自动生成**

**Part 4 Analytics and Reporting**

**Q1: How many** **prescriptions were filled for the drug Ambien?**

From below table, we can see that there are 5 prescriptionsfilled for Ambien.

**表格

描述已自动生成**

**Q2: How many unique members are over 50 years of age? And how many prescriptions did they fill?**

表格

描述已自动生成

There are 2 unique members are over 50 years of old. And 7 prescriptions filled.

**Q3:** **For member ID 10004, what was the drug name listed on their most recent fill date? How much did their insurance pay for that medication?**

**表格

描述已自动生成**

**表格

描述已自动生成**

From above table, we can see that: For member\_id 10004, the drug named **Diprosone** listed on their most recent fill date. member\_id 10001, the insurance paid is 55. member\_id 10002, the insurance paid is 130. member\_id 10003, the insurance paid is 322.member\_id 10004, the insurance paid is 712.