# horizontal lineDatabase R&D Exercise

Part 3

I confirm that this is my own work and that use of material from other sources, including the Internet, has been properly and fully acknowledged and referenced.

|  |  |
| --- | --- |
| Name: | Pang, Jinhao |
| Date: | 2022.12.8 |
| NYU ID: | N19475049 |
| Course Section Number: | csci-ga.2433-001 |



**Total in points** (100 points total): \_\_\_\_\_

**Professor’s Comments:**

|  |
| --- |
|  |

**1. Physical database design**

As so many tables need to deploy, we only select several typical designs for part 3.

Diagram

Description automatically generated

Figure 1 Logical design for inter-related data

Here, we deploy three databases. Chronic is the indicator that stores all the information about chronic diseases. For specific chronic diseases, we deploy a table Kidney to store the information on chronic kidney disease. Customer is to have insurance and CustomerImage is used to store the unstructured data like images for customers.

1. Customers

Graphical user interface, text, application, email

Description automatically generated

Figure 2 Physical design of customer

(CustLastName,CustFirstName,CustMiddleName,CustSuffix,CustDOB) can be used as a unique key to finding a particular record, and set index to (CustFirstName,CustMiddleName,CustLastName) for the faster query because this may have duplicates and usually be used to find a customer in a practical situation.

1. CustomerImage

Graphical user interface, text, application, email

Description automatically generated

Figure 3 Physical design of CustomerImage

We have foreign key to link the corresponding customer and reference constraints.

1. Others

Text

Description automatically generated

Figure 4 Partition

We also want to use partition to increase the query speed, where we use different years to find the specific patient easier and provide support and policy.

**2. Use cases and reference architecture**

The use cases roughly demonstrate the workflow-based application for a customer to obtain an insurance quote and a policy shown in Figure 5. Here we use user information, for example, the architecture uses various medical information to acquire the potential possible risk of having a chronic disease and provide appropriate quotes and policies for the customer. The analysis is based on the machine learning model trained by the massive amount of data. Each time customer has another health status, the system will update its information for the following policy and quote to increase the profits and reduce costs. On the other hand, according to the big data analysis, prediction of potential disease can be found and reduce the total cost of the customer.

**Diagram

Description automatically generated**

Figure 5 Use cases for simple application view

**3. Machine Learning Model Creation**

Four models are used to predict the possibility of the chronic kidney disease. The model performance is shown as follows:

Graphical user interface, text, application

Description automatically generated

Figure 6 Model performance

The dataset is chronic kidney disease, a microcosm of all the chronic diseases. We also have chronic disease indicator table as an overview. We can infer the correlation between ages with chronic diseases and other factors. As the number of data increasing, more insights can be inferred with few-shot learning.

MySQL is deployed on the Microsoft Azure Cloud as mentioned in Project Part 2 and machine learning model is shown in the Google Colab. The link is <https://colab.research.google.com/drive/1V2qRofrjW-T4PQMghpGN5y9E6y757C3L?usp=sharing>.