Name: Wu, Hana Date: 12/19/2022

NYU NetID: yw6775

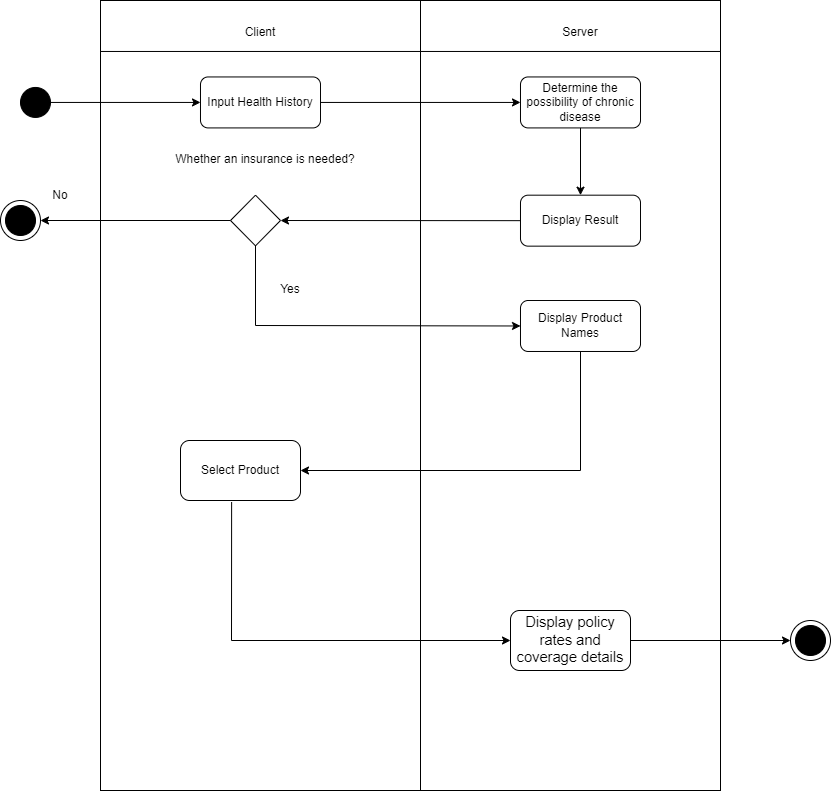
Section: 001

**Project Part 4**

**Professor’s Comments:**

**Affirmation of my Independent Effort:** *Hana Wu*

The W company is an insurance company working on a data-driven workflow-based database application to estimate possibilities of chronic disease and provide insurance quotes. Customers can provide their health history and will be given results of whether they have the potential of having a chronic disease. They can also obtain insurance quotes and products from the application once they think there is a need for insurance. The application could store information about different insurance products, as well as policy rates and coverage details. It could also track the status of quotes and policies, allowing customers to view and manage their coverage online.



When the client inputs their health history, the application should return whether they have a possibility of chronic disease with a trained machine learning model. After that, the user input whether they need insurance. If yes, the system will display a list of insurance products. The customer can choose which product they are interested in and the system will queue the specific product with correlated index and display policy rates and coverage details of the selected insurance product.

The machine learning strategy has been used to predict the likelihood of the user's chronic disease by analyzing the user input features. The pre-trained model has been used to predict the result.

The W database team has collected data from people with chronic diseases and trained a model for predicting the likelihood of the risk of chronic disease. When the user runs the program, the system will ask the user to input their health information. Based on the collected information, the system will predict the whether the user has a risk of chronic disease and show results on the screen. After that, the program will ask the user whether they are interested in any health insurance product. If the answer is yes, the system will list all insurance products and ask which product the user is interested in. After the user selected the product, the system will list all policy rates and coverage details.

The first step in the process is to gather and ingest the data from various sources into the OLTP/ODS database. Those structured and unstructured data can be extracted, transformed, and loaded into the database system. After that, insights should be generated from the data and we can do some preprocessing. Finally, the data should be connected to the client so that users can access the data and retrieve the information they need. Query optimization involves analyzing the database queries and identifying opportunities to optimize them for better performance. This can be done using techniques such as indexing, partitioning, and query rewriting.

The reference architecture should include the following aspects:

* Data Quality Management: Data cleansing, data validation, and data integration to ensure that the data is accurate and consistent.
* Prevention of Data Losses and Leakage: Data backup and recovery, data encryption, and data access controls to prevent unauthorized access to the data.
* Management of the Data Lifecycle: Data ingestion, data transformation, data storage, data access, and data disposal.

Github Site: https://github.com/wuyuhan1218/DBProject.git