Name: Chenhao Jin

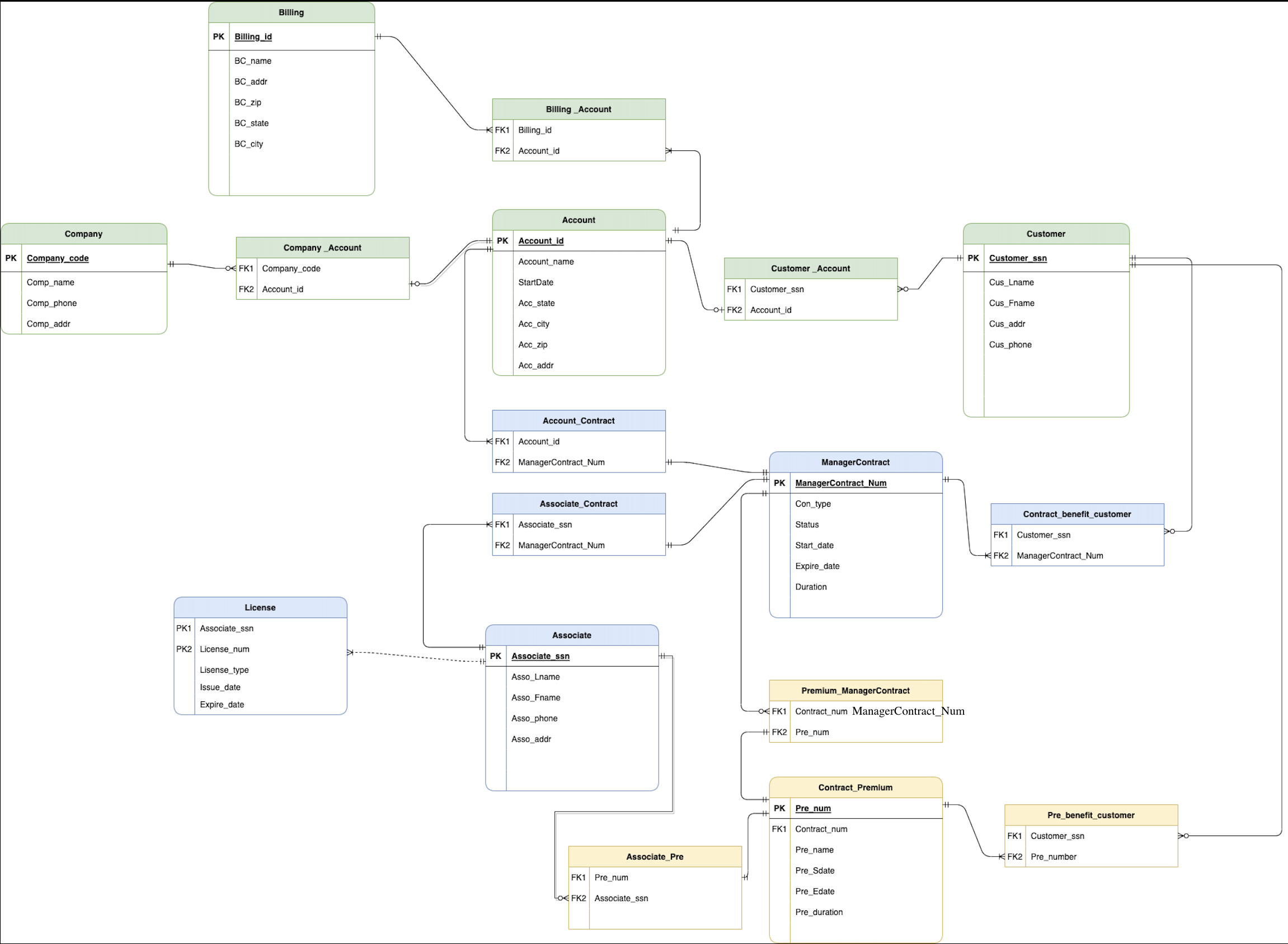
NetID: cj2234

Section: 001

Database System Project Part 2

MongoDB

I used MongoDB Atlas for Data Lake. The clod provider is Azure(Virginia cluster). I set one database in the cluster, the data will be inserted into this database and process. The data will be restructured. I used PyMongo to connect Atlas and manipulate data. The cluster contains static text data. Patients data will be collected and proceed in data lake.The data will be use for training a random forest regressor model to predict the probability of getting chronic disease based on the customer condition who in the MySQL database. Storing the data in the cloud in a good and safe way to train machine learning model. MongoDB supports scalability and high performance on data access

**For Logical Schema**