Collaborators: Jiayu Gu, Denghao Sun

Date: November 16, 2023

NetID: jg7956, ds6963

Section: 001

**Database Systems Project Part II**

**Logical Schema Optimization and Unstructured Data Collection**

**Affirmation of our independent efforts: \_\_\_\_Denghao Sun, Jiayu Gu\_\_\_\_\_\_\_\_\_\_\_**

**Logical schema optimization and extension steps taken:**

We choose SQL as our database system.

Entities:

Customers

1. Extension

To enhance the detection of chronic diseases, we have added two types of attributes to the entity 'customers': a boolean attribute and a general attribute. The general attribute encompasses the customer's income and the type of chronic disease they may have. The income helps the insurance company gauge the customer's living standard. Differences in the range of income will correspond to different scores. A higher score will get a lower score. If a customer already suffers from a chronic disease, this information is recorded under the 'type of chronic disease' column, and the column will contain a total counter of disease that customers already suffer from. The boolean attribute is derived from a series of questions asked by the company, which customers complete before selecting their insurance plan. This attribute includes factors such as heredity of chronic diseases, burning the midnight oil (indicating lack of sleep), unhealthy eating habits, and unstable emotional status. A 'yes' response to these questions is recorded as '1' in the customer table, while a 'no' response is recorded as '0'.Additionally, other factors like age, country, urbanization, and globalization are taken into account based on customer information such as age and address. The total score of a customer will be an assessment of risk for insurance company. This comprehensive approach allows for a more nuanced assessment of the customer's health risks and insurance needs.

1. Normalization: Third Normal Form(3NF): The SSN of a customer can determine all other attributions in customers. Therefore, there is no optimization step needed for the customer entity.
2. Database system: The customers table shows all information for customers that the insurance company needs and will store in SQL with the form that schema we create.

BillingAccount

1. Normalization: Second Normal Form (2NF): Identified non-key attributes that were not fully dependent on the primary key. For instance, attributes like BillingAddress1, BillingAddress2, BillingCity, BillingState, and BillingZip are all part of the billing address and dependent on BAcctID. These were then extracted into a separate BillingAddress table. Also considered SpecialHandlingCode for separation if it's not directly dependent on BAcctID.

Account

1. Extension: We add an attribution SSN as the foreign attribution in the account entity to connect the entity of account and customers.
2. Normalization:Second Normal Form (2NF): Looked at dependencies and realized that attributes like LocationAddress, LocationCity, LocationState, LocationZip, Location phone and MultiLocationAccountFlag are all about location and depend on AccountID. Moved them to a new Location table.
3. Database System: The account table shows all information about the customer's account and it will be stored in a table which form shows in the schema. We use the SSN of customers to define his account information. Moreover, insurance companies can use BAcctID to find the information of the billing account. We use the AccountID as foreign key in the location table to connect the table location and account.

Invoice

1. Normalization Third Normal Form (3NF): The invoice number can determine all other attributions. Therefore, there is no optimization step needed for the customer entity.
2. Database System: We add two foreign attributions in invoice tables– SSN and BAcctName. The insurance company can check the billing information of a specific customer base on ssn and send an invoice about payment to him.

Claim

1. Normalization Third Normal Form (3NF): The claim number can determine the claim date, settlement date and wellness eligibility date. The claim image is a new table but belongs to the claim.
2. Database System: The claim connects customer, contract and product. When a customer makes a claim, the insurance company will check the claim image and contract information of the customer by using SQL to search the ssn and contract number. We use the claim number as foreign key in the claim image table to connect the claim image and the claim.

Product

1. Normalization Third Normal Form (3NF): The primary key for the product is{LineOfBusiness,Series name,PlanName}. There is no repeat attribution, partial dependency and the primary key can determine all of the other attributes. Therefore, there is no optimization step needed for the product entity.
2. Database System: We create a new table to connect the relationship between account and the product. The insurance company can search the ID of the account to see if the product of the account has already ended. The information in the product table is all of the products that the company provides. All of the information will be stored in the SQL and can be extracted in the future.

Contract

1. Normalization Second Normal Form (2NF): Identified non-key attributes that were not fully dependent on the primary key. The CreditCardNo, ExpirationDate, CardType, BankingTransitNumber, BankingAccountType, BankingAccountNumber and PremiumPaymentLimit don’t depend on the contract number. Therefore, we extracted them into a separate payment table. We set CreditCardNo as the primary key, and CreditCardNo is a foreign key in the table contract. The insurance company can check the information of payment based on the credit card number. There are two tables about the different policy of contract and also reference to the main contract table.
2. Database System: We reference the table product by setting the {LineOfBusiness,Series name,PlanName} as foreign key in table contract. If an insurance company wants to check the product which is included in the contract, they can search the LineOfBusiness, Series name and PlanName to get the corresponding information in the table product. We also set the SSN as foreign key in the table contract. The insurance company can check the information of customers for a specific contract based on the SSN.