SQL Analytics for Healthcare Data

Project Description:

Health care data to generate instant reporting for their claim status, number of providers and patients diagnosed details for diseases based on ICD10, ICD9 codes drug NDC, Provider NPI, etc. The data is available in cloud data warehouse. We are making complex SQL script to export report in excel and format also validate the results with the open resources which is available in govt sites before sending it to client. The main objective of this project is the client wants to know about their product market and the claim status.

Technologies Used:

Snowflake, Databricks, Excel, SQL

Requirements:

Instant reporting for client requirements in different categories in different time period.

Time Period	Categories
Every Day	NPI
Weekly	NDC
Last 15days	Claims
Monthly	ICD9
Quarterly	ICD10
Half Yearly (Last 6 Months)	Drug Name
Yearly (Last 12 Months)	ProductName

Challenges Faced:

The main challenge is we need to understand the data. Because the data is coming from different sources like all scripts, ERD, etc. The data is available in cloud data warehouse in different level. We need relate the data and combine to give the expected solution. Based on client requirement we need to connect multiple tables with complex SQL queries. If we join multiple tables directly in single query it will take time also occupy more memory space in cloud. It will affect the other persons script in cloud sometimes it breaks.

Solutions:

We provide more scalable and time-consuming solutions to overcome the challenges what we have faced. Create temporary table to export the necessary data which is needed for the client requirements. That would avoid making more joins in actual data table. Split the query to export the data from different data sources. It is the easy way also to reduce query running time. Finally merge the data and get the report.

After export report we will validate with real-time open source data which is available in online. And format it as easy way of understanding the data.