**Data Modeling and Schema Design**

Design of the data model and schema was the first key component of the project. This was a key step toward the creation of the trusted foundation for the fraud detection data warehouse. We started by choosing the right schema for the project and chose to create a star schema. Since we would be dealing with large volumes of transactional data, simplicity and efficiency was the choice of the star schema. There are several dimension tables, linked to a central fact table. For example, in the case of storing the quantity that was sold, the fact table was designed to store quantitative data regarding fraud detection however the dimension tables were developed to store descriptive data related to products, customers, and dates.

In this next phase they defined the dimension tables that would be needed for the fraud detection project. The Product dimension was also important in this example: it included product names, categories, and item codes for categorizing transactions. Likewise, the Customer dimension contained customer ID and buyer information which allowed us to keep track of which customers were committing fraud. The attributes important for time-based analysis of the data were stored inside the Date dimension, such as transaction date, month, year, and quarter.

The size of the fact table during the schema design phase posed one of the major challenges because one cannot reveal such huge amounts of data inside a transaction table without fighting query performance. Architecting fact and dimension tables using property relationships helps solve this problem because you only have to store the relevant data. Furthermore, normalization principles constrained data integrity and consistency and were deformalized to generate a schema that will enhance query performance. The aim was to strike fictional balance between sheer efficiency and total clarity, such that a schema could clean large datasets to a point we could use them for analytical purposes.

An important role for strong ETL (Extract, Transform, Load) process was found. This raw data generated from the operational system needed to be accurately transformed and loaded into the data warehouse. There were inconsistencies in the data, blank values, and improper formats that the ETL process had to handle in order to avoid misinformation and bad analysis that will happen in the reporting. To ensure the integrity of the data and thus the schema to be more stable for future analysis and reporting, rules and checks were implemented for clear data transformation rules.

In this phase, team members had been working to exchange ideas with each other to learn whether the schema would be suitable for the project’s mandate. Everyone sat down to figure out each other’s' relationships between tables and to finalize the transformations needed to get data ready to be loaded into the warehouse. Through a team collaboration, the schema design was both efficient and flexible to allow changes in the business requirements and to adapt to a changing data structure.