**Individual contribution report – Data storage solutions**

**ROBIN ROY**

* Introduction:

The dataset selected for the analytics is the dataset which contains the historical sales of supermarket company which operates its sales in different branches in different cities named as A,B and C and the data shows the customer purchase details collected over a period of three months.

The main purpose of the project is to come to meaningful conclusions which will benefit the business as there are already competitors in popular cities providing the same service.

We designed and created a data warehouse through ETL(extract, transform, load), we created several reports generated through the analysis of data using SSRS and visualized the reports in dashboard using a the tool called “tableau” which is an important reporting software used for effective data visualization and analytics. We used Neoj4 which is a graph database system to demonstrate graph databases for our dataset.

The tasks for the successful contribution of the project were equally divided among the three team members. This report rapidly explains about the stages that I have experienced while completing the assignment and my dedication for the same. The tasks I completed are following:

* Studying the dataset selected for the analysis
* Using MS SQL Server to draw schema diagram.
* Creating Fact table and dimension tables
* Creating ETL
* Creating Report
* Creating PowerPoint

There is a requirement for developing a suitable schema for analysing the data warehouse. A schema in a SQL database is an assortment of logical structures of information. I First connected to the database engine and navigate to the database we are working on and created new database diagram. I created the fact table which contains measurements, metrics or facts about the dataset selected and it is surrounded by the dimensions of the fact, we connect the dimension table to the fact table and are located at the edges of the star schema, it helped me to create a logical grouping of database objects. It is also very important to create the ETL where I used the data pipeline to collect data from the various sources, transform the data according to business rules, and load it into a destination data store. I understood the control flow and the data flow which ensures the orderly processing of the set of tasks

Challenges faced:

* Ensuring data integrity while formatting the data values
* Deleting duplicating data and cleaning the database
* Loss of data
* Studying the dataset understanding-R sales and purchases

Conclusion:

I understood how important is to analyse data for business operations. It is very import to study the data collected from various operation of businesses and create notable information. It is very important to maintain data integrity.

The reports which are the documents created as result of analysing the data helped us to understand the dataset better. Creating dashboard using Tableau simplifies the process of data visualization. Graph database using Neoj4 helped in creating graph database from reports. We came to several conclusions that will help the business in future.