

# **KGiSL INSTITUTE OF TECHNOLOGY**

**INSTITUTION CODE : 7117**

**NAAN MUDHALVAN**

**PROJECT TITLE:**

*Data Warehousing with IBM Cloud Db2 Warehouse.*

**MENTOR:**

*Mrs. R. Indu Poornima*

**TEAM MEMBERS:**

- 1.** *Keerthana. J*
- 2.** *Saruthi. T*
- 3.** *Kishore. K*
- 4.** *Jeevanandham. S*

# **IMPLEMENTING ADVANCED ANALYTICS AND MACHINE LEARNING MODELS FOR PREDICTIVE ANALYSIS IN THE DATA WAREHOUSE**

## **PROJECT DOCUMENTATION**

### **PROJECT OBJECTIVE:**

*The objective of this data warehousing project is to establish an efficient and comprehensive data management system using IBM Cloud Db2 Warehouse. The primary aim is to consolidate data from multiple sources, integrate it into a unified repository, and empower data architects to extract valuable insights for informed decision-making.*

### **DESIGN THINKING PROCESS AND DEVELOPMENT PHASES:**

**1.Planning Phase:** *Defined the project scope, objectives, and requirements, including the identification of data sources and integration strategies.*

**2.Design Phase:** *Created a structured data warehouse schema, designed for optimal data integration, and formulated a comprehensive ETL strategy.*

**3.Development Phase:** *Implemented the data warehouse using Python and IBM Db2 Warehouse, incorporated ETL processes for data extraction, transformation, and loading, and enabled data exploration and analysis through SQL queries.*

## **DATA WAREHOUSE STRUCTURE:**

*The data warehouse comprises four primary tables: Customer, Product, Orders, and Order\_Items, interconnected through well-defined relationships. This structure facilitates efficient data retrieval and analysis, enabling a holistic view of the organization's operations.*

## **DATA INTEGRATION STRATEGIES:**

*The project incorporates diverse data integration strategies, including direct CSV data import and ETL processes for seamless data extraction from external databases. These strategies ensure a consistent flow of information into the data warehouse, allowing for comprehensive data analysis.*

## **ETL PROCESSES:**

*The ETL processes involve the extraction of data from multiple sources, transformation to adhere to the data warehouse schema, and the subsequent loading of this transformed data into IBM Cloud Db2 Warehouse. This systematic approach ensures data accuracy and consistency.*

## **DATA EXPLORATION TECHNIQUES:**

*Data architects can explore and analyze data within Db2 Warehouse using SQL queries and advanced analysis techniques. These capabilities enable the discovery of meaningful patterns and trends, supporting data-driven decision-making processes within the organization.*

## **SUBMISSION:**

### **GitHub Repository Link:**

*Link to the GitHub Repository*

## **DEPLOYMENT AND MANAGEMENT INSTRUCTIONS:**

- 1.Clone the repository to your local machine.*
- 2.Set up IBM Cloud Db2 Warehouse with the necessary configurations.*
- 3.Configure the required environment variables for establishing the database connections.*
- 4.Execute the provided Python scripts to initiate the data warehouse setup and run the ETL processes seamlessly.*
- 5.Access the data warehouse through Db2 Warehouse to explore and analyze the data using SQL queries.*

## **README FILE:**

*Please refer to the README file in the GitHub repository for comprehensive instructions on navigating the data warehouse, updating content, managing dependencies, and executing the provided scripts for data integration and analysis.*