# Phase 5:Project Documentation

This encapsulates the comprehensive journey from problem definition to robust data warehouse implementation. This section summarizes objectives achieved, including data structuring, integration, analytics, schema development, and seamless ETL execution. It highlights a holistic approach ensuring a reliable data foundation. The forthcoming content details achievements, methodologies, and outcomes from each phase, shaping the project's success and significance.

## Problem Definition and Design Thinking

**Objectives**:

- Data Warehouse Design: Structured and implemented a robust data warehouse using IBM Cloud Db2 Warehouse.
- Data Integration: Seamlessly integrated diverse data sources through effective ETL (Extract, Transform, Load) processes.
- Data Exploration: Provided user-friendly querying tools for effective data exploration.

**Achievements**:

- Data Warehouse Structure: Successfully designed and implemented a comprehensive structure involving data sources, staging area, warehouse database, ETL processes, and metadata repository.
- Data Integration: Integrated and harmonized data from various sources into a unified repository through ETL processes.
- Data Exploration: Enabled user-friendly data exploration and effective querying, ensuring actionable insights for stakeholders.

## Innovation

### Objective:

- Advanced Analytics: Incorporated advanced analytics tools or machine learning models for predictive analysis within the data warehouse.

### Accomplishments:

- Advanced Analytics Implementation: Utilized the Scikit-learn library in Python to introduce machine learning models for predictive analysis.
- Data Preparation: Prepared and engineered data for analysis by removing duplicates and performing feature engineering.
- Model Training and Evaluation: Trained a Random Forest Classifier model and evaluated its performance, enabling data-driven decisions based on predictive analysis.

## Development Part 1 - Schema and Structure

### Objective:

- Structured Schema Implementation: Defined and structured schema for efficient storage and analysis of sales data.

### Accomplishments:

- Table Definition: Created tables for customers and orders, ensuring a structured way to store and analyze sales data.
- Data Integration Strategy: Executed ETL processes for data integration, ensuring data consistency and reliability in the integrated data warehouse.

## Development Part 2 - ETL Processes

### Objective:

- ETL Implementation: Implemented Extract, Transform, Load processes for data extraction and integration.

### Accomplishments:

- ETL Execution: Successfully executed ETL methodologies, collecting, transforming, and loading data from diverse sources into the data warehouse.
- Data Quality Assurance: Ensured accurate, consistent, and well-structured data foundation for analysis through seamless ETL processes.

This comprehensive approach covered everything from defining the problem and designing the data warehouse to integrating diverse data sources, implementing advanced analytics, and executing ETL processes. It ensured that the project encompassed all essential stages to meet its objectives effectively.