Problem Definition and Design Thinking Document

Problem Definition

The project at hand involves designing and setting up a robust data warehouse using IBM Cloud Db2 Warehouse. The primary objective is to create a centralized repository that can bring together data from various sources, perform advanced data integration and transformation, and provide data architects with the necessary tools to explore, analyze, and deliver actionable data for informed decision-making. This project encompasses several key aspects:

Data Warehouse Structure

Objective: Define the schema and structure of the data warehouse to accommodate various data sources.

Understanding: We need to carefully design the data warehouse structure to ensure it can efficiently store and manage data from multiple sources. This involves determining how data will be organized, what tables will be created, and how relationships between them will be established.

Approach: We will start by analyzing the types of data sources and data models required for this project. A thorough understanding of the data's nature will guide us in defining the schema effectively.

Data Integration

Objective: Identify data sources and design a strategy to integrate data seamlessly into the data warehouse.

Understanding: Data may be sourced from diverse systems and formats, including databases, spreadsheets, APIs, and more. Successful integration requires identifying these sources, establishing connections, and devising a strategy to extract data from them efficiently.

Approach: We will conduct a comprehensive assessment of available data sources. We will then design data integration workflows, considering factors like data refresh frequency, data quality, and error handling.

ETL Processes

Objective: Plan and implement ETL (Extract, Transform, Load) processes to extract, transform, and load data into the warehouse.

Understanding: ETL processes are crucial for data preparation. Extracting data from source systems, transforming it to meet the warehouse's schema, and loading it accurately are essential for maintaining data integrity.

Approach: We will create ETL pipelines tailored to the specific requirements of each data source. Transformation logic will be applied to standardize data formats and ensure consistency within the warehouse.

Data Exploration

Objective: Design queries and analysis techniques to empower data architects to explore and analyze data.

Understanding: Data architects and analysts require user-friendly tools and interfaces to explore data. This involves creating custom queries and analysis methods that simplify the process.

Approach: We will explore tools and platforms that allow for interactive data exploration. Creating predefined queries and reports will streamline the data analysis process.

Actionable Insights

Objective: Focus on delivering actionable insights by enabling informed decision-making based on data.

Understanding: The ultimate goal of this project is to provide actionable insights that guide decision-making. This requires visualizations, dashboards, and reporting mechanisms that effectively communicate findings.

Approach: We will collaborate closely with stakeholders to understand their specific requirements for actionable insights. Visualization tools and reporting frameworks will be chosen to create user-friendly dashboards and reports.

Design Thinking Approach
The design thinking approach for this project involves an iterative and user-centric process that considers the needs of data architects, analysts, and other stakeholders. We will follow these key steps:

Empathize: Understand the needs and pain points of data users. Conduct interviews, surveys, and workshops to gather insights.

Define: Clearly define the project scope, objectives, and success criteria. Create user personas and user stories to guide development.

Ideate: Brainstorm and generate ideas for data warehouse structure, integration strategies, ETL processes, data exploration tools, and actionable insights.

Prototype: Create prototypes and mockups of data warehouse structures, ETL pipelines, data exploration interfaces, and visualization dashboards for feedback and validation.

Test: Iterate on prototypes based on user feedback and test them with actual data to ensure functionality and usability.

Implement: Develop and implement the data warehouse, ETL processes, and data exploration interfaces based on the validated designs.

Test and Iterate: Continuously test and improve the system, incorporating feedback from users and stakeholders.

Deliver: Provide training and documentation to users and ensure a smooth transition to the new data warehouse system.

Evaluate and Evolve: Monitor system performance and user satisfaction, making necessary adjustments and improvements over time.

By following this design thinking approach, we aim to create a data warehouse that not only meets technical requirements but also effectively serves the needs of the people who rely on it for making data-driven decisions. This iterative and user-centric process will help ensure the success of the project.