

Data Warehousing with IBM Cloud Db2 Warehouse

921821104051: YASVINDHINI U

PHASE -1

Overview:

Phase 1 document for the project “**Data Warehousing with IBM Cloud Db2 Warehouse**” from IBM on the Naan Mudhalvan Scheme.

Project Title:

Data Warehousing with IBM Cloud Db2 Warehouse.

Project Description:

Design and set up a robust data warehouse using IBM Cloud Db2 Warehouse. Bring together data from various sources to unlock valuable business insights. Perform advanced data integration and transformation effortlessly. Empower data architects to explore, analyze, and deliver actionable data for informed decision-making!

Project Definition:

The project involves designing and setting up a robust data warehouse using IBM Cloud Db2 Warehouse. The objective is to bring together data from various sources, perform advanced data integration and transformation, and provide data architects with the tools to explore, analyze, and deliver actionable data for informed decision-making. This project encompasses defining the data warehouse structure, integrating data sources, performing ETL (Extract, Transform, Load) processes, and enabling data analysis

Design Thinking

1.Data warehouse structure :

Data warehouse schema is a description, represented by objects such as tables and indexes, of how data relates logically within a data warehouse.

A data warehouse schema typically involves a star or snowflake structure. In a star Schema, a central fact table is surrounded by dimension tables. In a snowflake schema, dimensions are normalized into multiple related tables. The Choice depends on your specific needs and the nature of your data. Make sure to identify key attributes and relationships between data sources to design an Efficient structure.

2.Data Integration :

Data warehouse integration combines data from several sources into a single, unified warehouse. The data warehouse can be accessed by any department within an organization, and the data can be easily structured into spreadsheets or tables for research and analysis purposes

3.ETL processes

Plan and implement ETL processes to extract, transform, and load data into the warehouse.

To design effective ETL processes, start by understanding your data sources, defining transformation logic, and ensuring compatibility with the warehouse schema.

Extract the relevant data from the source database. Transform the data so that it is better suited for analytics. Load the data into the target database.

4. Data Exploration:

Data collection: This refers to the process of gathering data from various sources, such as sensors, devices, social media, websites, and more.

Transformation: When the data is already collected then it should be cleaned and transformed before storing.

Data storage: This refers to the systems and technologies used to store and manage data, such as databases, data lakes, and data warehouses.

Analytics: This refers to the tools and techniques used to analyze and interpret data, such as statistical analysis, machine learning, and visualization.

Query for Schema Evolution:

❓ `SHOW CREATE TABLE table_name;`

❓ Helps track changes in data structure over time.

5. Actionable Insights

Focus on delivering actionable insights by enabling informed Decision-making based on data.

Actionable insights empower better decision-making, transforming data into strategic advantages.