## Data Warehousing with IBM Cloud Db2 Warehouse

## Project Title: Data Warehousing

## PHASE 1 SUBMISSION

## Problem Statement: 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 Steps**

**Phase 1: Problem Definition and Design Thinking**

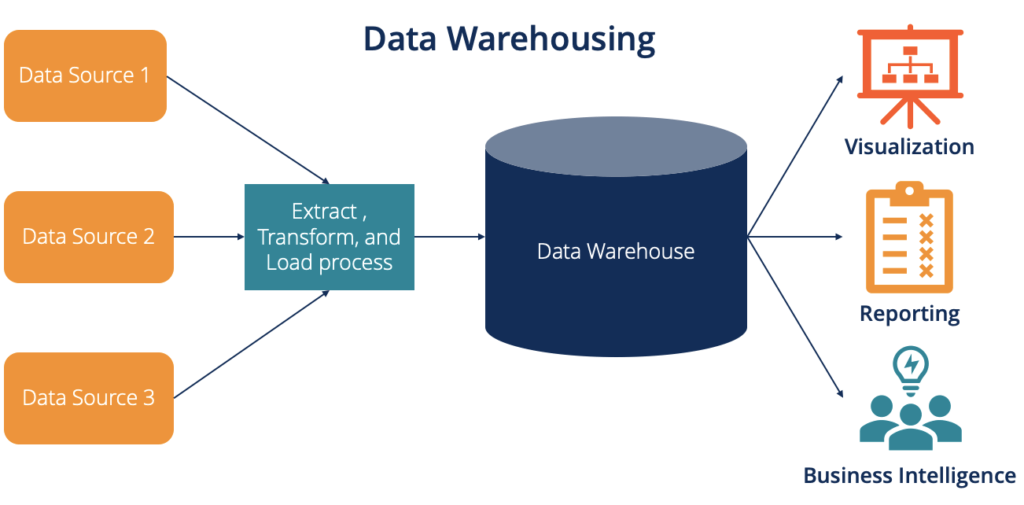
**Problem Definition:** The project involves designing and setting up a robust data warehouse using IBM Cloud Db2 Warehouse. The objective is to bring together data from various sources, perform advanced data integration and transformation, and provide data architects with the tools to explore, analyze, 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.

**DATA WAREHOUSING:**

A data warehouse, also called an enterprise data warehouse (EDW), is an enterprise data platform used for the analysis and reporting of structured and semi-structured data from multiple data sources, such as point-of-sale transactions, marketing automation, customer relationship management, and more. Data warehouses include an analytical database and critical analytical components and procedures.

**ADVANTAGES:**

* Data Centralisation
* Historical Data
* Improved data Quality
* Scalability
* Data Security
* Data Redundancy
* Decision Support
* Faster Query Performance



**STEPS INVOLVED IN DATA WAREHOUSE**:

**EXTRACTION OF DATA** **=** A large amount of data is gathered from various sources.

**CLEANSING OF DATA =** Once the data is compiled, it goes through a cleaning process. The data is scanned for errors, and any error found is either corrected or excluded.

**CONVERSION OF DATA** **=** After being cleaned, the format is changed from the database to a warehouse format.

**STORING IN A WAREHOUSE =** Once converted to the warehouse format, the data stored in a warehouse goes through processes such as consolidation and summarization to make it easier and more coordinated to use. As sources get updated over time, more data is added to the warehouse.

**DESIGN THINKING**

1. **Data Warehouse Structure: Define the schema and structure of the data warehouse to accommodate various data sources**.

**The Schema used in our database is Star Schema.**

* In multiple dimension there is no sub-dimension tables.
* Data Redundancy is high.
* Increased Performance.
* It is easy to understand.

**Structure of Data Warehouse:**

This is mostly used Architecture of Database Warehouse, because it consists of three tiers.

* Top Tier.
* Middle Tier.
* Bottom Tier.

STEPS TO INVOLVED DATA INTERGRATION FOR DATA WAREHOUSING.

1. **\*Define Business Goals\*:**

- Understand your organization's objectives and define specific business goals that the data warehouse will help achieve.

**2. \*Select Cloud Data Warehouse Service\*:**

- Choose a cloud data warehouse service that suits your needs (e.g., AWS Redshift, Google Big Query, Snowflake, Azure Synapse Analytics).

**3. \*Data Source Identification\*:**

- Identify all data sources, both internal and external, that you want to integrate into the data warehouse. These could include databases, SaaS applications, APIs, and more.

**4. \*Data Extraction\*:**

- Set up data extraction processes (ETL or ELT) to extract data from various sources. Consider tools like Apache NiFi, Apache Airflow, or cloud-native data integration services.

**5. \*Data Transformation\*:**

- Transform the extracted data to conform to a unified schema and quality standards. This step often includes data cleansing, data enrichment, and data aggregation.

**6. \*Data Loading\*:**

- Load transformed data into the cloud data warehouse. Optimize loading strategies (e.g., batch, streaming) based on your data volume and latency requirements.

**7. \*Data Modeling\*:**

- Design an appropriate data model (e.g., star schema, snowflake schema) to support analytical queries efficiently.

**8. \*Data Security\*:**

- Implement robust data security measures, including encryption, access controls, and compliance with data privacy regulations.

**9. \*Data Governance\*:**

- Establish data governance practices to maintain data quality, metadata management, and version control.

**10. \*Performance Optimization\*:**

- Continuously monitor and optimize query performance by using indexing, partitioning, and materialized views.

**11. \*Backup and Recovery\*:**

- Implement regular data backups and a disaster recovery plan to protect against data loss.

**12. \*Scaling\*:**

- Ensure your data warehouse can scale both vertically and horizontally to accommodate growing data volumes and user demands.

**13. \*Monitoring and Logging\*:**

- Set up monitoring and logging solutions to track system health, query performance, and user activity.

**14. \*Business Intelligence Tools\*:**

- Integrate business intelligence (BI) tools like Tableau, Power BI, or Looker to create reports and dashboards for business users.

**15. \*User Training and Support\*:**

- Train your team and end-users on how to access and leverage the data warehouse effectively.

**16. \*Documentation\*:**

- Maintain comprehensive documentation of your data warehouse architecture, data models, and processes.

**17. \*Iterate and Refine\*:**

- Continuously assess your data warehouse's performance and relevance to evolving business needs. Make improvements as necessary.

**18. \*Cost Management\*:**

- Keep an eye on costs associated with cloud resources and optimize resource allocation to control expenses.

**19. \*Data Integration with Business Workflows\*:**

- Integrate data access and analytics capabilities into business workflows to enable real-time decision-making.

**20. \*Data Governance and Compliance\*:**

- Ensure ongoing compliance with data governance policies and regulations.

**TEAM MEMBERS:**

**JEYALAKSHMI SUBIKSHA S**

**HUDHA N**

**GAYATHRI S**