**United Institute of Technology**

**7145**

**PHASE-3**

**Data Warehousing with IBM Cloud Db2 Warehouse**

**Submitted by:**

R Maha

M Sanjana

V Kavin

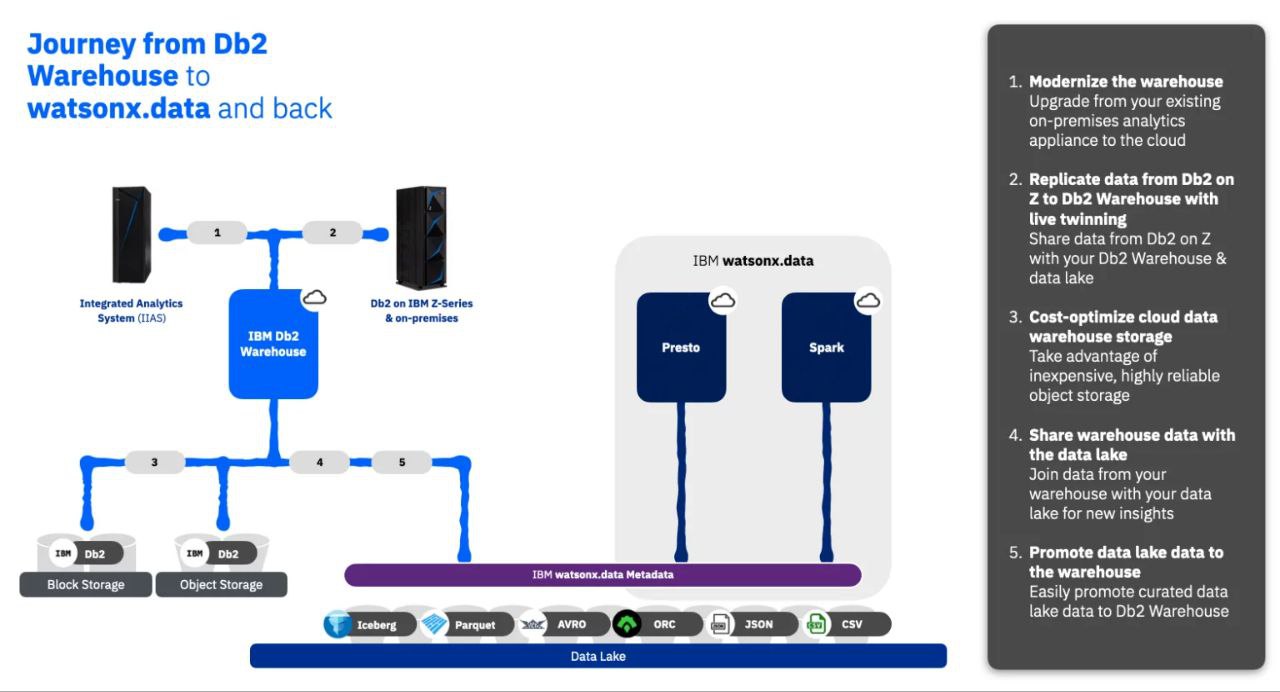
K Navin

S Kiran

S Karthik

**Data Warehousing with IBM Cloud Db2 Warehouse**

**Development part** **1**



IBM Cloud Db2 Warehouse on Watson is a cloud-based data warehousing solution offered by IBM. It is designed to help organizations store, manage, and analyze large volumes of data to make data-driven decisions. Here are some key features and information about it:

1. Data Warehousing: IBM Cloud Db2 Warehouse is specifically designed for data warehousing, which means it's optimized for storing and querying structured data.

2. Cloud-Based: It's available as a cloud service, so you can scale your data warehousing needs based on demand without the need to manage physical hardware.

3. Integration with Watson: It's part of the IBM Watson family, which means you can integrate it with other Watson services for advanced analytics and AI capabilities.

4. SQL Compatibility: Db2 Warehouse supports SQL, making it relatively easy for users with SQL skills to work with the data.

5. Scalability: You can scale your resources up or down as needed, ensuring that you have the appropriate computing power for your data warehouse.

6. In-Memory Processing: It often uses in-memory processing to speed up queries and analysis.

7. Security: IBM places a strong emphasis on security, which is critical when dealing with sensitive data.

8. Analytics: You can run complex analytics and reporting on the data stored in Db2 Warehouse.

9. ETL (Extract, Transform, Load): It can integrate with ETL tools to bring data from various sources into your data warehouse.

10. Data Governance: It provides tools and features to help with data governance and compliance.

To get started with IBM Cloud Db2 Warehouse on Watson, you would typically need an IBM Cloud account and then provision the service.

**IBM Cloud Db2 Warehouse can be integrated with IBM Cloud Object Storage to enhance its data warehousing capabilities. Here's how this integration can be useful:**

1. Data Storage: You can use IBM Cloud Object Storage as a highly scalable and cost-effective storage solution for your data warehouse. This allows you to store vast amounts of data securely and access it as needed.

2. Data Backup: Storing data in IBM Cloud Object Storage provides a reliable backup solution for your data warehouse, ensuring data durability and availability.

3. Data Ingestion: You can easily ingest large datasets into your Db2 Warehouse from IBM Cloud Object Storage. This simplifies the process of loading data into your data warehouse for analysis.

4. Data Archiving: Older or less frequently accessed data can be moved to object storage, reducing storage costs in your data warehouse while still keeping the data accessible.

5. Scalability: The combination of Db2 Warehouse and IBM Cloud Object Storage allows you to scale your data warehousing resources and storage needs independently, making it a flexible and cost-efficient solution.

6. Cost-Effective Storage: IBM Cloud Object Storage is designed for cost-effectiveness, making it a suitable choice for long-term data storage.

7. Data Security: You can take advantage of the security features and access controls provided by IBM Cloud Object Storage to protect your data.

To set up this integration, you would typically configure your Db2 Warehouse to connect to IBM Cloud Object Storage by providing the necessary credentials and information. Then, you can use SQL or other methods to interact with data stored in object storage.

**Program**

To interact with IBM Cloud Db2 Warehouse using Watson services, you'll need to use code. Below is an example of Python code that demonstrates how to connect to an IBM Db2 Warehouse on IBM Cloud and perform a simple SQL query using the ibm\_db library. Additionally, it utilizes IBM Watson services, specifically Watson Studio, for analytics and visualization. Note that you need to have the necessary credentials for both the Db2 Warehouse and Watson Studio services.

python

import ibm\_db

from ibm\_db import connect

import ibm\_db\_dbi

db\_credentials = {

"hostname": "your-db-hostname",

"username": "your-db-username",

"password": "your-db-password",

"port": "50000",

"database": "your-db-name",

"security": "SSL"

}

ws\_credentials = {

"apikey": "your-api-key",

"url": "https://api.us-south.dataplatform.cloud.ibm.com"

}

dsn = (

"DRIVER={{IBM DB2 ODBC DRIVER}};"

"DATABASE={0};"

"HOSTNAME={1};"

"PORT={2};"

"PROTOCOL=TCPIP;"

"UID={3};"

"PWD={4};"

).format(db\_credentials["database"], db\_credentials["hostname"], db\_credentials["port"], db\_credentials["username"], db\_credentials["password"])

conn = ibm\_db.connect(dsn, "", "")

conn\_dbi = ibm\_db\_dbi.Connection(conn)

query = "SELECT \* FROM your\_table\_name"

results = conn\_dbi.execute(query).fetchall()

import pandas as pd

import matplotlib.pyplot as plt

df = pd.DataFrame(results, columns=[i[0] for i in conn\_dbi.execute(query).description])

df['column\_to\_plot'].plot(kind='bar')

plt.title("Data Analysis")

plt.xlabel("X-axis Label")

plt.ylabel("Y-axis Label")

plt.show()

conn\_dbi.close()

This code connects to your Db2 Warehouse on IBM Cloud, performs a SQL query, and retrieves the results. It then uses Watson Studio, pandas, and matplotlib for data analysis and visualization.