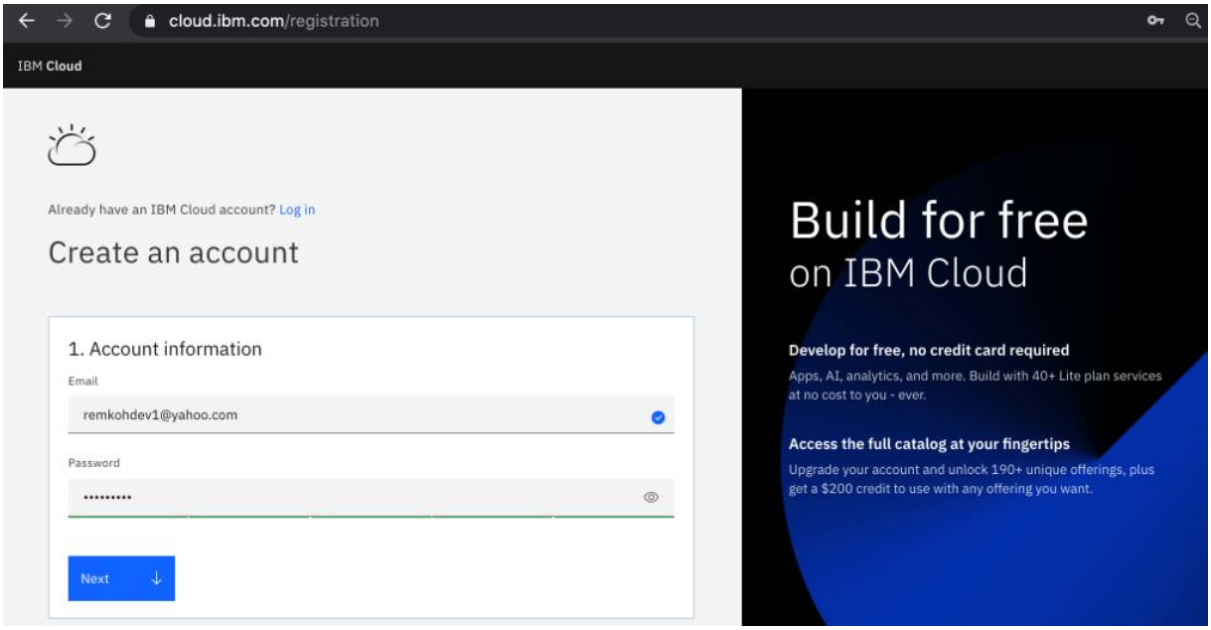
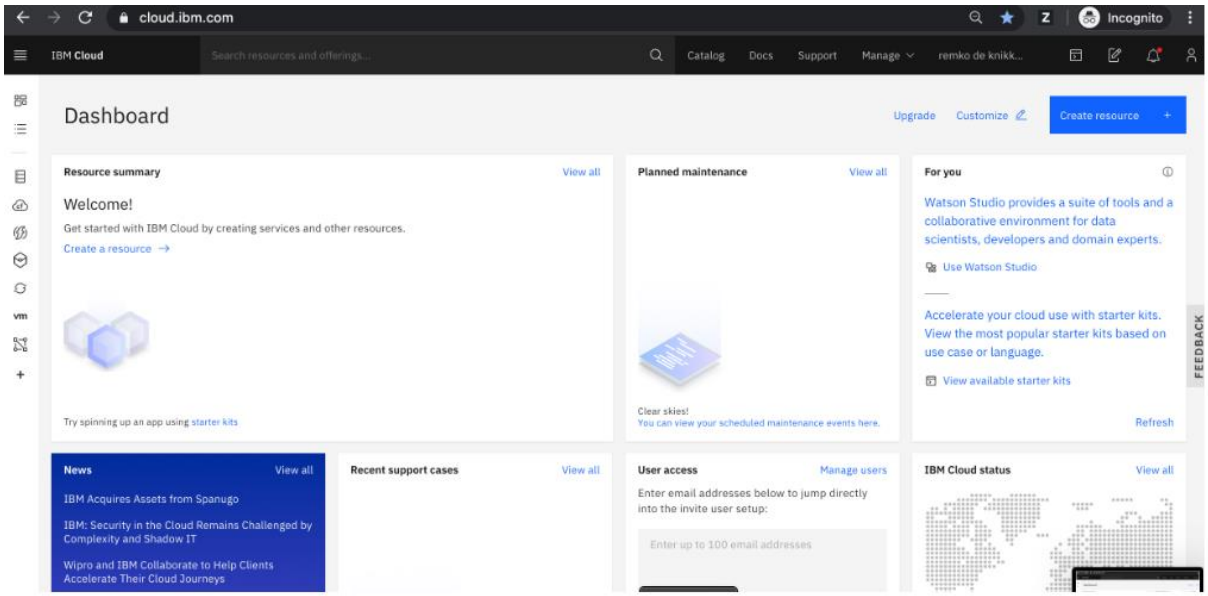
CAD PHASE-3

PROJECT SUBMISSION

Big Data Analysis with IBM Cloud Databases

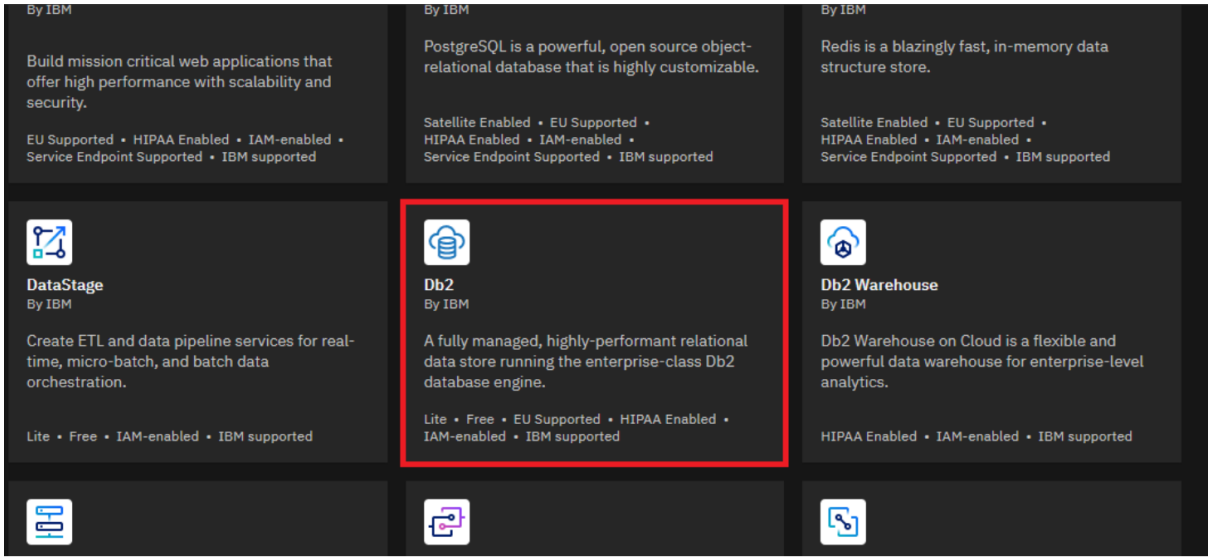
Create an account:





Select the Appropriate Database Service:

IBM Cloud offers various database services, including Db2, MongoDB, Db2 on Cloud, and more. The choice of database service depends on specific project requirements.



Db2: This is a relational database management system (RDBMS) that's suitable for structured data and SQL-based querying.

MongoDB: This is a NoSQL database that's great for unstructured or semi-structured data and allows for flexible schema design..

Set Up a Database Instance:

Follow these steps to set up a database instance:

Log in to IBM Cloud account.

Go to the IBM Cloud Dashboard.

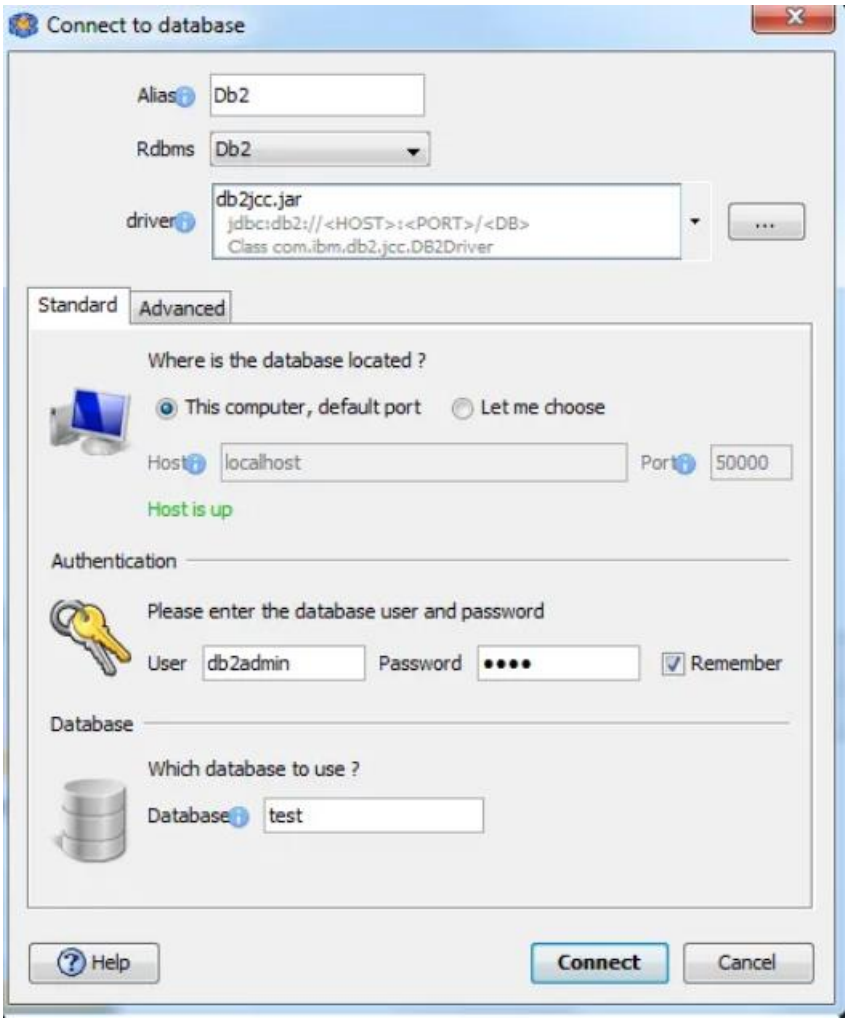
Click on "Create Resource."

Search for and select the database service that is chosen (e.g., Db2 or MongoDB).

Configure the service, specifying details such as instance name, region, resource group, and capacity.

Load Data into the Database:

Depending on the dataset, we may need to import our data into the database. Most database services provide tools and utilities for data import. Ensure that our data is in a format that's compatible with the chosen database service.



Develop Queries or Scripts:

Next, we can start building your queries or scripts for data analysis. We'll typically use SQL for databases like Db2 and a database-specific query language for databases like MongoDB (e.g., MongoDB Query Language). Write the queries that suit your analysis goals, which may include selecting, filtering, aggregating, and joining data.

Data Cleaning and Transformation:

Before analysis, it's essential to clean and transform data as needed. This process may include:

Removing duplicates.

Handling missing data (e.g., filling in missing values or removing incomplete records).

Standardizing data formats.

Aggregating or summarizing data.

Converting data types as necessary.

Perform Analysis:

Execute your queries or scripts to perform the desired analysis on your dataset. This can include generating reports, visualizations, or any other analytical output you need for your project.

Optimize and Scale:

Depending on the scale and complexity of your analysis, we might need to optimize the database configuration or consider scaling up our resources in IBM Cloud to ensure smooth performance.

Monitor and Maintain:

Regularly monitor our database's performance and maintain it as needed. Ensure that our data remains accurate and up to date.