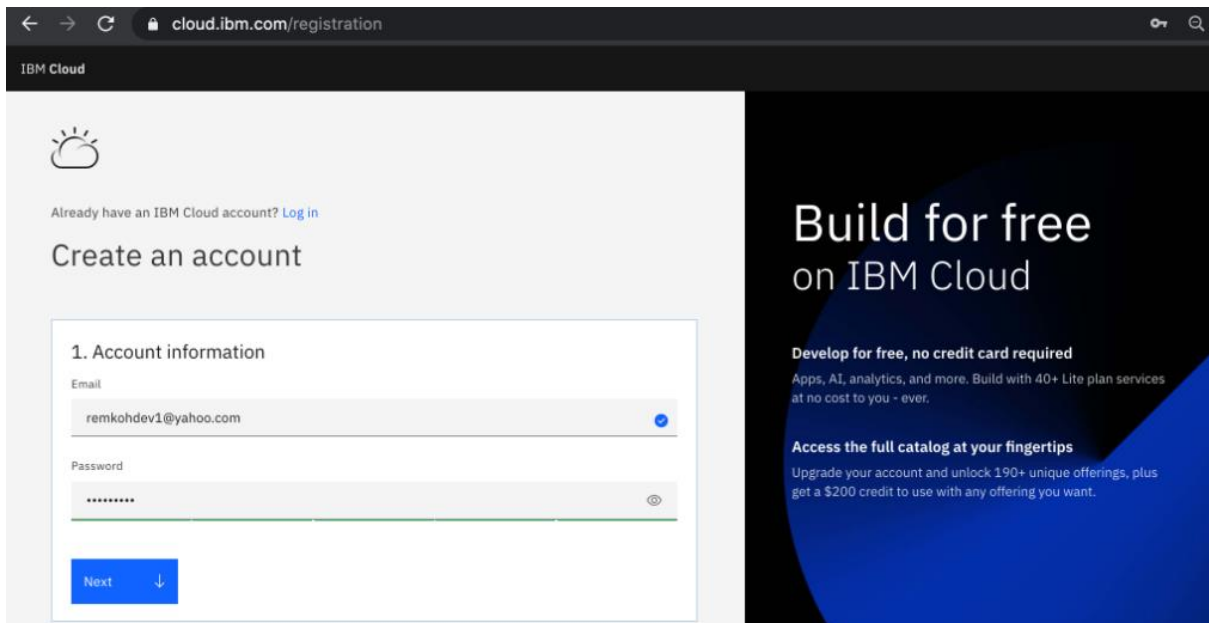


## CAD PHASE-3

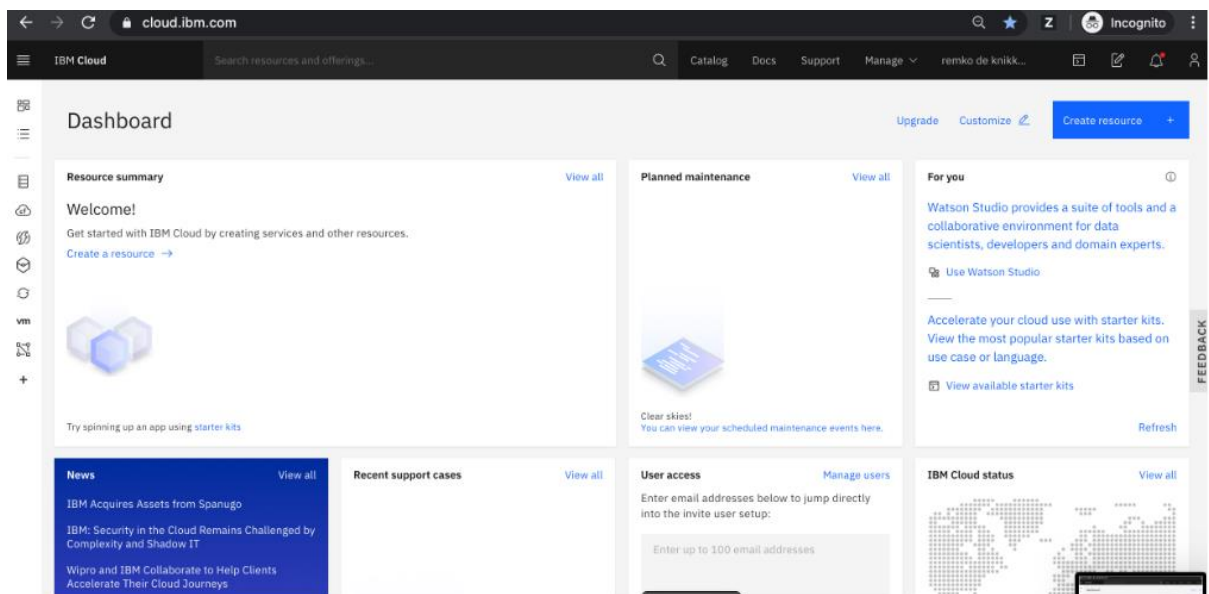
### PROJECT SUBMISSION

#### Big Data Analysis with IBM Cloud Databases

Create an account:

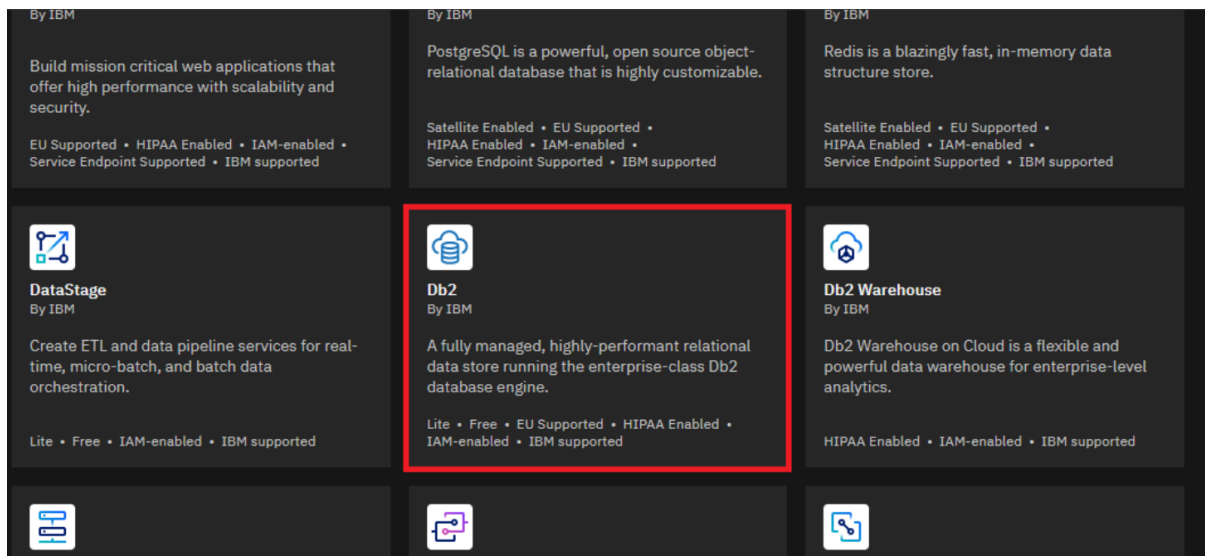


The screenshot shows the IBM Cloud registration page at [cloud.ibm.com/registration](https://cloud.ibm.com/registration). The page has a dark header with the IBM Cloud logo. The main content area is split into two sections. On the left, there's a light gray box with the heading "Create an account" and a "Log in" link for existing users. Below this is a form titled "1. Account information" with fields for "Email" (containing "remkohdev1@yahoo.com") and "Password" (masked with dots). A blue "Next" button is at the bottom of the form. On the right, there's a dark blue section with the text "Build for free on IBM Cloud". It includes the subtext "Develop for free, no credit card required" and "Access the full catalog at your fingertips".



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.

Connect to database

Alias: Db2

Rdbms: Db2

driver: db2jcc.jar  
jdbc:db2://<HOST>:<PORT>/<DB>  
Class com.ibm.db2.jcc.DB2Driver

Standard Advanced

Where is the database located ?

☒ This computer, default port ☐ Let me choose

Host: localhost Port: 50000

Host is up

Authentication

Please enter the database user and password

User: db2admin Password: ..... ☒ Remember

Database

Which database to use ?

Database: test

Help Connect Cancel

IBM Db2 database

Server: TestIBMDb2server.contoso.com:40000

Database: NORTHWD2

Data Connectivity mode

☒ Import ☐ DirectQuery

Advanced options

OK Cancel

## 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.