Big Data Analysis with IBM Cloud Databases

PHASE 3: Development Part 1

GIVEN STATEMENT:

Start building the big data analysis solution using IBM Cloud Databases.Create an IBM Cloud account, choose the appropriate database service (e.g., Db2, MongoDB), and set up a database instance.

Develop queries or scripts to explore and analyze the selected dataset. Perform basic data cleaning and transformation as needed.

I understand the importance of your project, and I'm here to help. To get started with your big data analysis project using IBM Cloud Databases, follow these steps:

1. Create an IBM Cloud Account:

If you don't have an IBM Cloud account, sign up for one. You can do this by visiting the [IBM Cloud website] (**https://cloud.ibm.com/registration**) and following the registration process.

2. Choose the Appropriate Database Service:

Select the IBM Cloud Database service that best suits your project's needs. As mentioned earlier, you can choose between Db2 or MongoDB, depending on your dataset and requirements.
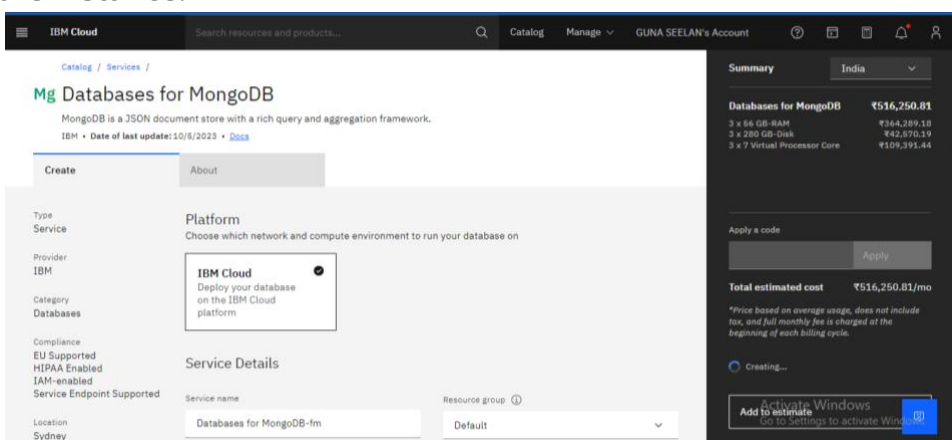
3. Set Up a Database Instance:

   For Db2:
- Log in to your IBM Cloud account.
- From the IBM Cloud dashboard, click on the "Create Resource" button.
- In the catalog, select "Databases" and then "Db2."

- Follow the on-screen instructions to configure your Db2 database instance, including specifying the instance name, region, and other settings.
- Create the instance.
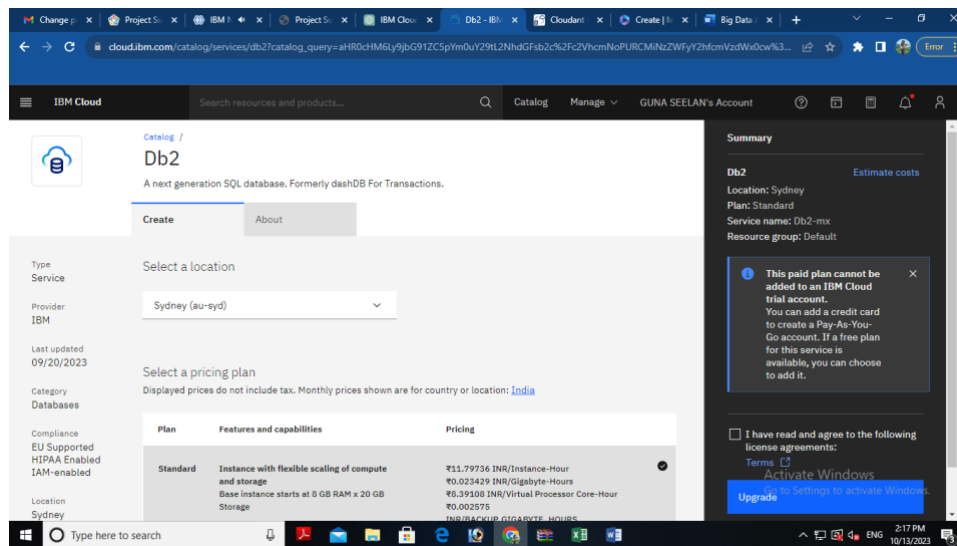
For MongoDB:
- Log in to your IBM Cloud account.
- From the IBM Cloud dashboard, click on the "Create Resource" button.
- In the catalog, select "Databases" and then "MongoDB."
- Follow the on-screen instructions to configure your MongoDB database instance, including specifying the instance name, region, and other settings.
- Create the instance.

## 4. Develop Queries or Scripts:

After setting up your database instance, you can start developing queries or scripts to explore and analyze your dataset. The type of queries and scripts you write will depend on the nature of your dataset and your analysis goals. You can use SQL for Db2 or MongoDB's query language for MongoDB.
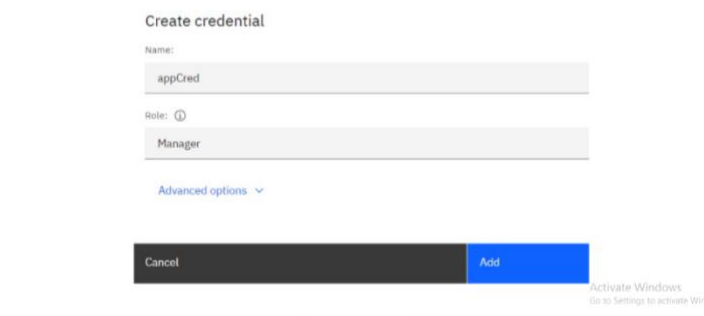
Creating Service Credentials the IBM DB2 database

- ♦ In the resource list screen of IBM Cloud, click on the DB2 service (displayed under Services and software category) that you created
- ♦ From the service page, select the menu option "**Service Credentials**" to create / access the credentials of the db2 database
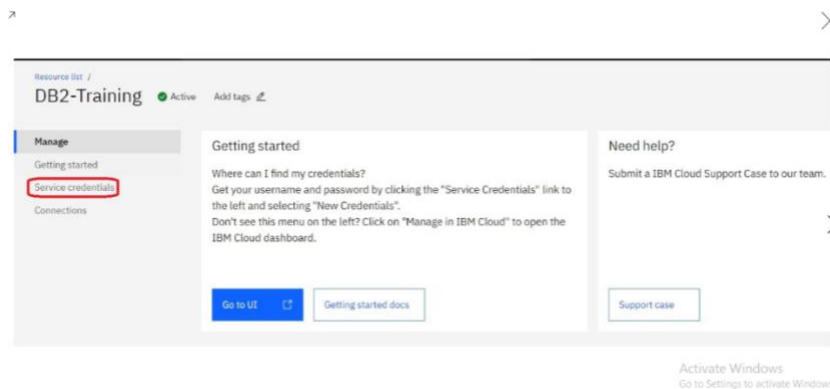
- ♦ Click on **New Credential** button in the Service Credential page to create a new credential
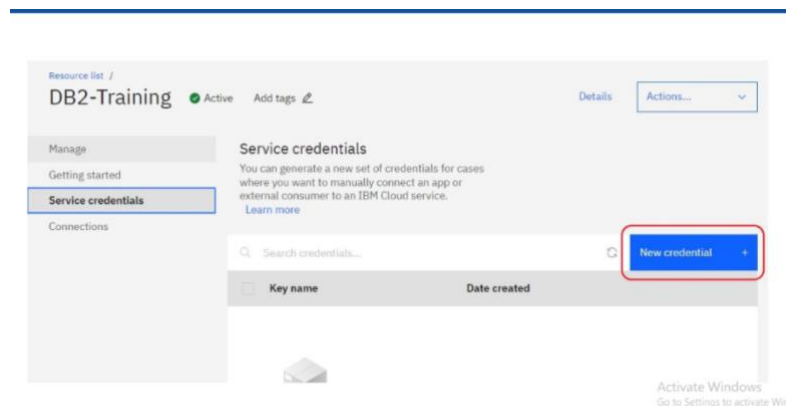
- ♦ Provide the any name for service credential (e.g. **appCred**) and click on **Add**

♦ New credential gets created and is displayed. Expand the newly to created credential to get the all the details that is required for client application to connect to the database. Note down the value for the following properties separately, which we will use it later to configure our application to connect to this database.

Property Name	Value
Database name	*<database> [e.g. bludb]*
Host name	*<hostname>*
Port	*<port>*
User Name	*<username>*
Password	*<password>*

```
"db2": {
  "authentication": {
    "method": "direct",
    "password": "                    ",
    "username": "              "
  },
},
```

Activate Windows

>

```
"hosts": [
  {
    "hostname": "fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud",
    "port": 32731
  }
]
```
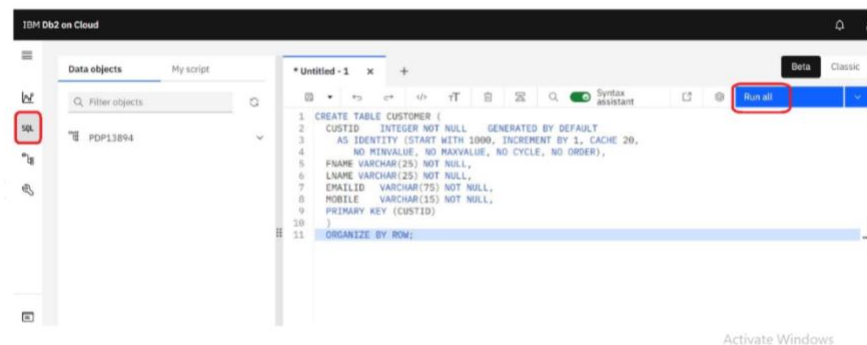
3. Setting up IBM DB2 database
♦ In the resource list screen of IBM Cloud, click on the DB2 service (displayed under Services and software category) that you created, if the page is not already opened.
♦ From the service page, select the menu option "**Manage**" and click on Go to UI to launch the DB2 console
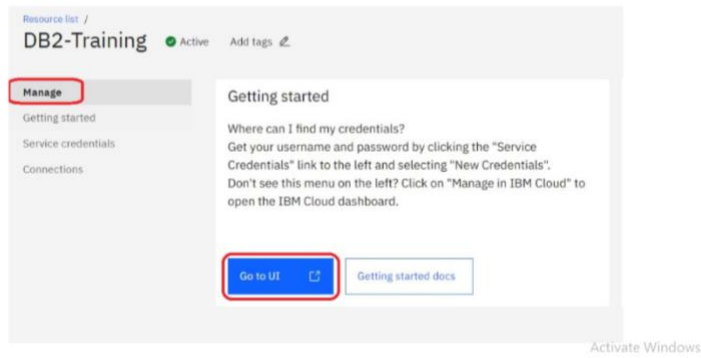
♦ IBM DB2 on cloud console is opened. To create database objects, click on SQL menu option from the left-side menu.

♦ SQL editor is opened up for you. Type the query that you want to execute in the SQL editor and click **Run all**

♦ The status of the query execution is displayed at the bottom of the SQL editor as shown below

The above steps can be followed to create any more database objects in future.

4. Downloading DB2 SSL Certificate and converting to PEM format

♦ In the console for IBM DB2, click on the spanner like icon which denotes Administration. On the resulting page, click on Download SSL Certificate button to download the DB2 certificate as shown below
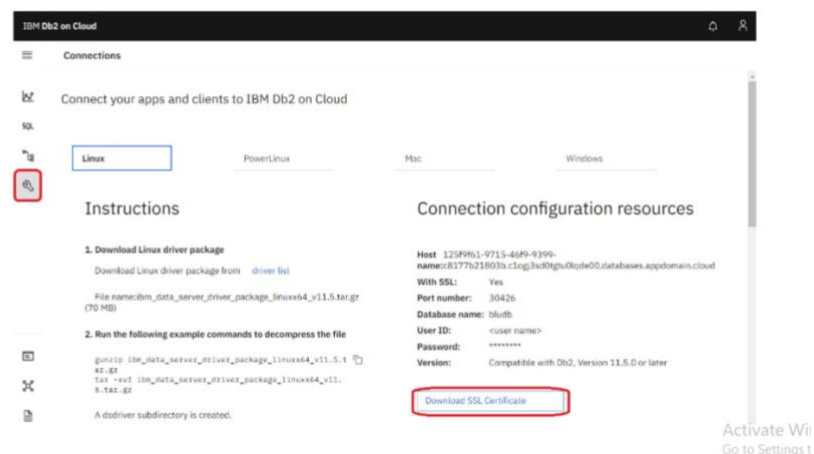
The SSL Certificate gets downloaded into the local machine, which is in DER format (cert file). To convert the cert file to PEM format, we can use the link SSL Converter - Convert SSL Certificates to different formats.

♦ In the SSL Converter website specify the following
♦ **Certificate File to Convert:** Upload the downloaded certificate file
♦ **Type of Current Certificate:** DER/Binary
♦ **Type To Convert To:** Standard PEM
♦ Click on **Convert Certificate** button to download the certificate in PEM format.
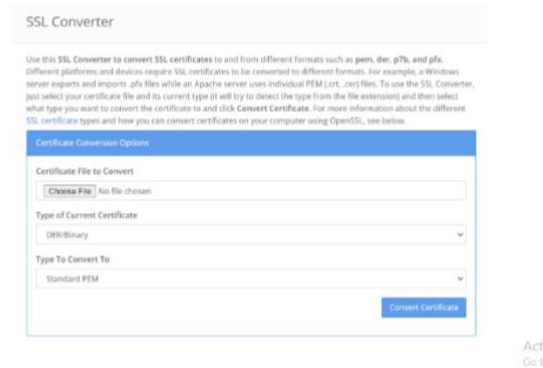
In this blog, we have seen how to subscribe to DB2 service on IBM Cloud, setup the database and create service credentials & certificate for application connectivity. In another blog, we will focus on using these details to configure ACE Cloud connector for DB2 to connect and use this database as part of solution development.

5. Perform Data Cleaning and Transformation:

As part of your data analysis, you may need to perform data cleaning and transformation. This can involve removing duplicates, handling missing data, and converting data types. The specific data cleaning and transformation tasks will depend on your dataset and analysis requirements.

Remember that I can provide guidance, answer questions, and help with SQL queries or MongoDB queries if you encounter specific issues during your project. Feel free to ask for assistance with any part of your project, and I'll do my best to help you successfully complete it.

Sample SQL Queries for Data Exploration and Analysis:

Retrieve Data from the Employee Table:

SELECT *

FROM employee_table;

Calculate the Average Salary:

SELECT AVG(salary) AS average_salary

FROM employee_table;

Find the Highest-Paid Employee:

SELECT first_name, last_name, salary

FROM employee_table

ORDER BY salary DESC

LIMIT 1;

Sample SQL Query for Data Cleaning (e.g., Remove Duplicates):

To remove duplicates based on a specific column (e.g., employee_id):

DELETE e1

FROM employee_table e1

INNER JOIN employee_table e2

ON e1.employee_id = e2.employee_id

WHERE e1.rowid > e2.rowid;

Sample SQL Query for Data Transformation (e.g., Update Date Format):
To update date format (assuming date_column is in the format 'MM/DD/YYYY'):

UPDATE employee_table

SET date_column = TO_DATE(date_column, 'MM/DD/YYYY');