

Db2 on Cloud and Watson Studio Workshop Guide

Version 4

Analyzing airline flight delays

IDUG NA 2024

June 23, 2024

Roger E. Sanders

Principal, Learning Content Development – Data, AI, & Storage

roger.sanders@us.ibm.com

NOTICES AND DISCLAIMERS

IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion. Information regarding potential future products is intended to outline potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision.

The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract.

The development, release, and timing of any future features or functionality described for our products remains at our sole discretion I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results like those stated here.

Information in these presentations (including information relating to products that have not yet been announced by IBM) has been reviewed for accuracy as of the date of initial publication and could include unintentional technical or typographical errors. IBM shall have no responsibility to update this information. **This document is distributed "as is" without any warranty, either express or implied. In no event, shall IBM be liable for any damage arising from the use of this information, including but not limited to, loss of data, business interruption, loss of profit or loss of opportunity.** IBM products and services are warranted per the terms and conditions of the agreements under which they are provided.

IBM products are manufactured from new parts or new and used parts. In some cases, a product may not be new and may have been previously installed. Regardless, our warranty terms apply."

Any statements regarding IBM's future direction, intent or product plans are subject to change or withdrawal without notice.

Performance data contained herein was generally obtained in controlled, isolated environments. Customer examples are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual performance, cost, savings or other results in other operating environments may vary. References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business.

Workshops, sessions, and associated materials may have been prepared by independent session speakers, and do not necessarily reflect the views of IBM. All materials and discussions are provided for informational purposes only, and are neither intended to, nor shall constitute legal or other guidance or advice to any individual participant or their specific situation.

It is the customer's responsibility to ensure its own compliance with legal requirements and to obtain advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulatory requirements that may affect the customer's business and any actions the customer may need to take to comply with such laws. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the customer follows any law.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products about this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products. IBM does not warrant the quality of any third-party products, or the ability of any such third-party products to interoperate with IBM's products. **IBM expressly disclaims all warranties, expressed or implied, including but not limited to, the implied warranties of merchantability and fitness for a purpose.**

The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents, copyrights, trademarks or other intellectual property right.

IBM, the IBM logo, and ibm.com are trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at: www.ibm.com/legal/copytrade.shtml.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates. Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

OpenShift is a trademark of Red Hat, Inc.

UNIX is a registered trademark of The Open Group in the United States and other countries.

© 2024 International Business Machines Corporation. No part of this document may be reproduced or transmitted in any form without written permission from IBM.

U.S. Government Users Restricted Rights — use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM.

Table of Contents

I. Introduction.....	5
The Business Use Case for this workshop	7
II. Provisioning a Db2 on Cloud Lite plan database on the IBM Cloud.....	9
1. Create an IBM account.	9
2. Create an IBM Cloud account.	17
3. Provision a Db2 on Cloud Lite plan database.....	32
4. Create a service credential for the Db2 on Cloud database.	38
5. Query the system catalog using the Db2 on Cloud User Interface (UI).	42
III. Building and populating a Db2 on Cloud database table	51
1. Create a table named AIRLINE_DELAY_CAUSE.	51
2. Populate the AIRLINE_DELAY_CAUSE table.	59
IV. Using IBM Watson Studio to run a Jupyter Notebook.....	68
1. Create an IBM Cloud Pak for Data as a Service account.....	68
2. Create a new project.	72
3. Add a Jupyter Notebook to the project.	74
4. Add a data file to the project.	81
5. Prepare the Jupyter Notebook for execution.....	83
6. Run the Jupyter Notebook.....	92
V. Analyzing the data	95
1. Log in to your IBM Cloud Pak for Data as a Service account.	95
2. Add a new Jupyter Notebook to the project created earlier.	97
Prepare the Jupyter Notebook for execution.....	103
3. Run the Airline_Delay_Analysis Jupyter Notebook	106
VI. Thank you!.....	111

Acknowledgements

Special thanks to **George Baklarz** for his assistance with the **Db2 Magic Commands** and his **Jupyter Notebook** debugging expertise. Without George's help, the workshop exercise would have taken much longer to complete.

And thanks to Danny Arnold for putting together the first two versions of this workshop guide.

I. Introduction

As the COVID-19 pandemic wound down and life returned to normal, the return to business and leisure travel resulted in many challenges for travelers – most notably, an apparent increase in flight delays and cancellations. However, because many people spent several years *not* traveling, have flight delays and cancellations really increased since 2021? Or is this just a historical pattern that people have simply forgotten about? This workshop attempts to answer that.

To determine whether flight delays and cancellations were up in 2022 and 2023, data containing airline on-time statistics and delay causes for the period of January 2013 through December 2023 was downloaded from the **United States Department of Transportation, Bureau of Transportation Statistics** website (https://www.transtats.bts.gov/ot_delay/). This data contains information about every flight, flight delay, flight cancellation, and reason for the delay or cancellation (there can be multiple reasons that contribute to a flight delay or cancellation) for every commercial airport and airline operating in the United States during this time.

This workshop is designed to provide a person with hands-on experience in provisioning a Db2 on Cloud Lite (free) plan database on the IBM Cloud, creating an AIRLINE_DELAY_CAUSE table in the Db2 on Cloud database, loading that table with data obtained from the United States Department of Transportation, Bureau of Transportation Statistics website, and then analyzing the data using IBM Watson Studio in IBM Cloud Pak for Data as a Service and Jupyter Notebook.

This workshop is divided into four main sections, as follows:

1. Provisioning a Db2 on Cloud Lite plan database on the IBM Cloud.

- Create an IBM account and IBMid (*optional – not necessary if you already have a valid IBMid*).
- Create an IBM Cloud account (*optional – not necessary if you already have an IBM Cloud account*).
- Provision a Db2 on Cloud Lite (free) plan database in the IBM Cloud data center of your choice.
IMPORTANT: The Db2 on Cloud Lite plan is only available in the Dallas (us-south) and London (eu-gb) data centers.
- Create a service credential for the newly created Db2 on Cloud database that will be used to perform some of the exercises in this workshop.
- Query the system catalog of the newly created Db2 on Cloud database using the Db2 on Cloud User Interface.

2. Building and populating a Db2 on Cloud database table.

- Create a table named **AIRLINE_DELAY_CAUSE** using the Data Definition Language (DDL) statement stored in a file named **AIRLINE_DELAY_CAUSE_DDL.sql** and the Db2 on Cloud User Interface.
- Populate the **AIRLINE_DELAY_CAUSE** table with data stored in a file named **Airline_Delay_Cause.csv** using the Db2 on Cloud User Interface. (**NOTE:** This file contains airline delay data just for the period of January 2019 through December 2022.)

3. Using IBM Watson Studio to run a Jupyter Notebook

- Create an IBM Cloud Pak for Data as a Service account (*optional – not necessary if you already have an IBM Cloud Pak for Data as a Service account.*)
- Create a new project in Watson Studio/ IBM Cloud Pak for Data.
- Add a Jupyter Notebook named **Build_And_Populate_Table** to the project by importing a pre-built Jupyter Notebook file.
- Add a data file to the project.
- Prepare the Jupyter Notebook for execution by:
 - Adding information found in the Db2 on Cloud service credential created earlier to application variables that are used to establish a database connection.
 - Generating a code snippet that will be used to access the data file that was added to the project.
- Run the **Build_And_Populate_Table** Jupyter Notebook.

4. Analyzing the data

- Log in to your IBM Cloud Pak for Data as a Service account.
- Add another Jupyter Notebook named **Analyze_Delay_Cause** to the project by importing another pre-built Jupyter Notebook file.
- Prepare the new Jupyter Notebook for execution by adding information found in the Db2 on Cloud service credential created earlier to application variables that are used to establish a database connection.
- Run the **Analyze_Delay_Cause** Jupyter Notebook.

The data set used for this workshop

Here is some background information about the data set that is used in this workshop:

- The data used for this workshop is data that has been collected by the *United States Department of Transportation* and consists of information about domestic (US) airline flight delays for the years 2013 through 2023. Information that can be gleaned from this data include the number of on-time flight departures and the number of flights that have been canceled or delayed. According to the definition provided by the Bureau of Transportation Statistics for the data set, a flight is considered delayed if it arrives at its destination airport 15 minutes or more past its scheduled arrival time. For delayed flights, the reason for the delay can be one of the following:
 - Air carrier delay – This category covers flight delays that occur because of mechanical issues, personnel issues (such as a flight crew exceeding their hours), and other related reasons.
 - Aircraft arriving late – This category covers flight delays caused by the incoming flight (that is providing the aircraft for the flight) arriving late from its origin airport.
 - Security delay – This category covers flight delays that are caused by airport security reasons. For example, if a traveler gets past a Transportation Security Administration (TSA) checkpoint after a prohibited item is found, thereby breaching airport security, and causing all passengers to be re-screened.
 - National Aviation System (NAS) delay – This category covers flight delays that are caused by NAS issues, such as air traffic control and related issues.
 - Weather delay – This category covers flight delays that occur because of extreme weather such thunderstorms, wintery weather and icing conditions, tornado warnings, and other weather-related delays.

The Business Use Case for this workshop

A fictional travel company, Global Travel Associates (GTA), has been receiving negative scores on Customer Satisfaction Surveys that are completed by their customers traveling in the United States of America (USA). Their number one complaint is about flight delays that result in missed connections or late arrivals to their destinations.

To try to improve their customer satisfaction scores, GTA has obtained recent flight delay data from the U.S. Department of Transportation's Bureau of Transportation Statistics for the years 2019 through 2022 and 2013 through 2023. Their plan is to analyze this data to determine if flight delays have increased over the past two years (2022 and 2023) and if so, what might be causing the delays. To save on costs, GTA has selected the **Db2 on Cloud ‘Lite’ plan** as their cloud database platform. They also have decided to use the free tier of **IBM Watson Studio in IBM Cloud Pak for Data as a Service**.

The exercises in this workshop are designed to take you through GTA’s process of storing data obtained from U.S. Department of Transportation’s Bureau of Transportation Statistics for the years 2019 through 2022 in a Db2 on Cloud ‘Lite’ plan database and analyzing the data to gain insight on what has been causing flight delays and how flight delays have been trending. You will also see how GTA used a layered approach in examining the data to avoid making an incorrect assumption based on initial analysis results. Finally, you will see how GTA used historical flight delay data (obtained from the same source for the years 2013 through 2023) to see how flight delays observed in 2022 and 2023 compare to flight delays seen in the past.

At the conclusion of the workshop exercise, the participant should know how to provision a Db2 on Cloud database, create and populate a database table using both the Db2 on Cloud User Interface and source code stored in a Jupyter Notebook (executing in IBM Watson Studio on IBM Cloud Pak for Data as Service), and how to use a Jupyter Notebook and Db2 Magic Commands to analyze data quickly.

II. Provisioning a Db2 on Cloud Lite plan database on the IBM Cloud

The premise behind this workshop is that a client wants to store and analyze flight delay information at the lowest possible cost. Consequently, a Db2 on Cloud Lite plan database is used since this offering is available at no charge.

To provision a Db2 on Cloud “Lite” plan database, you must have an IBMid and an IBM Cloud account, both of which can be created free of charge. **If you already have an IBMid and an IBM Cloud account, proceed directly to Step 3.**

1. Create an IBM account.

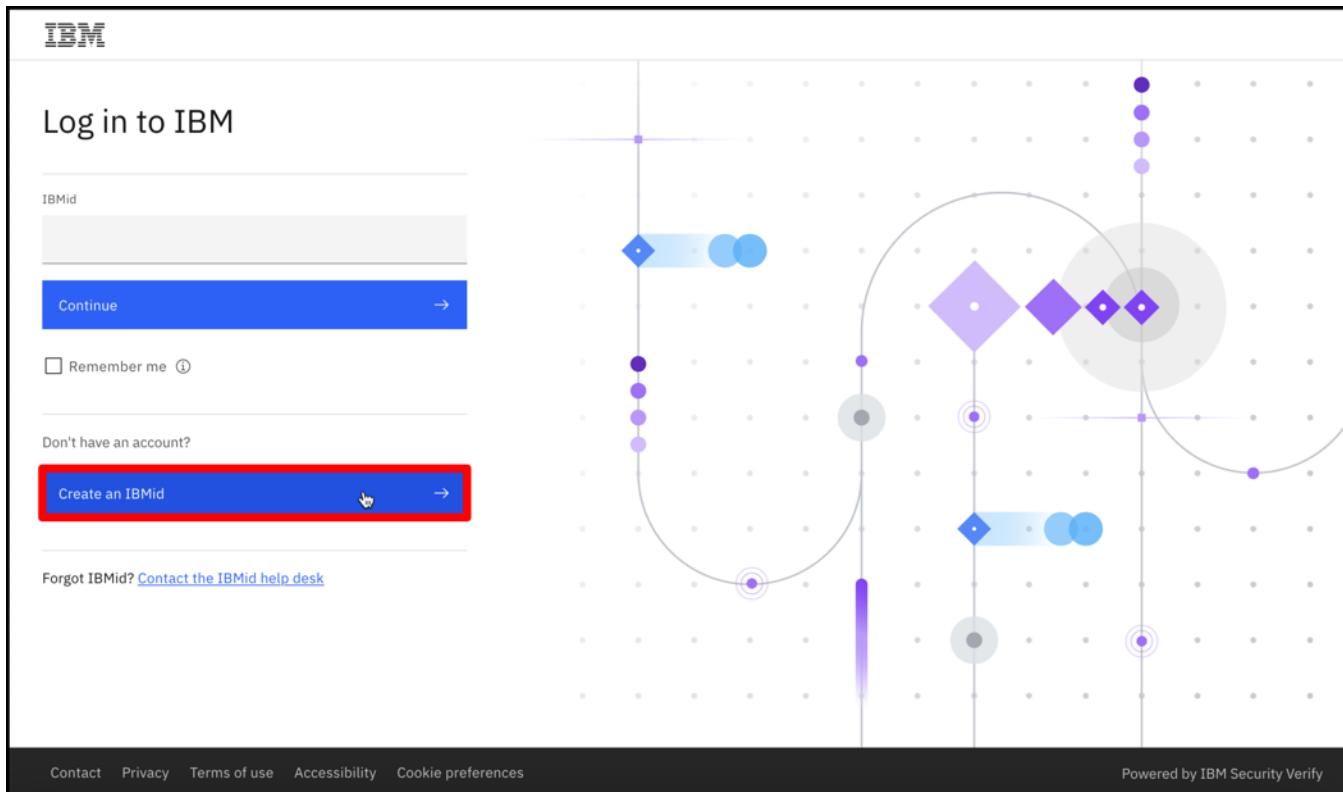
1. Open a web browser and go to the *IBM Account* web site (<https://www.ibm.com/account/us-en/>). Once there, select the **Log in to My IBM** button.

See what's possible with an IBM account

Log in to My IBM →

Manage and Support	Manage profile	Manage products	Get support
	Maintain profile information such as phone number, address, contact email,	Manage your trials, subscriptions, software and hardware purchases from	Find technical documentation, downloads or open a problem ticket

_____ 2. When the *Log in to IBM* window is displayed, locate the **Create an IBMid** button and click on it.



_____ 3. When the *Create an IBM account* screen is displayed, you may be asked to indicate how you want to let IBM use cookies. If that is the case, click on the **Accept all** or the **Required only** button shown in the popup window displayed at the bottom of the screen. (In this example, the **Accept all** button was chosen.)

IBM

Welcome to IBM

Create an IBMID

Already have an IBM account? [Log in](#)

Account information

Sign in with LinkedIn
AutoFill your information

E-mail

Your email address will become your IBMID, which you'll use to log into IBM.com.

First name Last name

Password

About cookies on this site

Our websites require some cookies to function properly (required). In addition, other cookies may be used with your consent to analyze site usage, improve the user experience and for advertising.

For more information, please review your [cookie preferences](#) options. By visiting our website, you agree to our processing of information as described in IBM's [privacy statement](#).

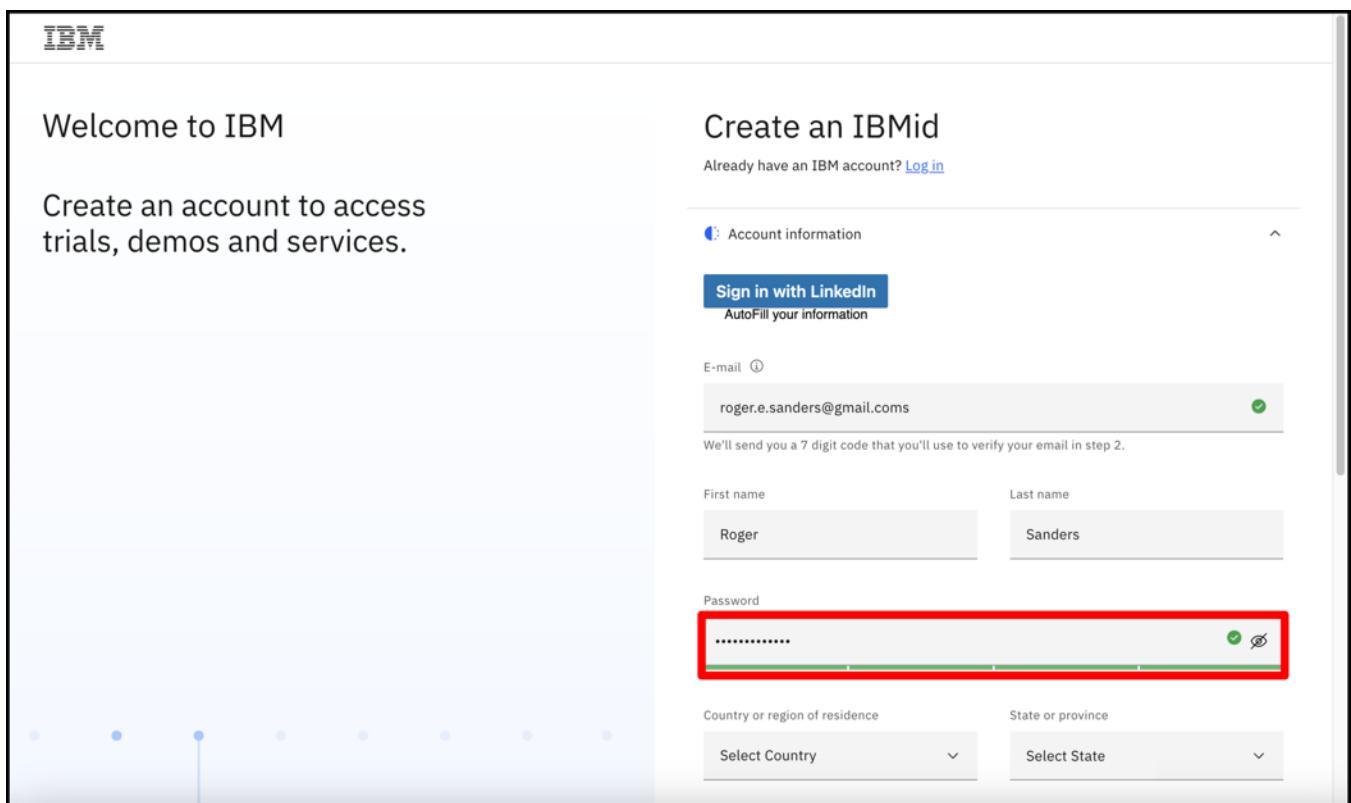
To provide a smooth navigation, your cookie preferences will be shared across the IBM web domains listed [here](#).

Select Country

Accept all

Required only

4. Once you have indicated how you want cookies to be used, provide your **E-mail** address, **First name**, **Last name**, and a **Password** in the appropriate entry fields. Note that the password supplied must be 12-36 characters in length, and must contain at least one uppercase character, one lowercase character, and one number. Special characters are optional, but double-byte characters are NOT allowed. The bar displayed just below the **Password** field will turn green once a sufficiently strong password that meets these criteria has been entered.



The screenshot shows the 'Create an IBMID' page on the IBM website. At the top left is the IBM logo. Below it, the heading 'Welcome to IBM' and the sub-instruction 'Create an account to access trials, demos and services.' To the right, the main title 'Create an IBMID' is displayed above a 'Account information' section. A 'Sign in with LinkedIn' button is present. The form includes fields for 'E-mail' (roger.e.sanders@gmail.com), 'First name' (Roger), 'Last name' (Sanders), and 'Password'. The 'Password' field is highlighted with a red rectangle. Below the password field are dropdowns for 'Country or region of residence' and 'State or province'. At the bottom left, there is a progress bar consisting of several blue dots connected by a vertical line. On the right side of the page, there is a sidebar with a 'Log in' link and a 'AutoFill your information' link.

____ 5. (A) Scroll down and provide the appropriate information in the **Country or region of residence**, **State or province**, and **Company** fields. Also indicate whether you are a student by selecting the appropriate radio button (**Yes** or **No**) under the **Are you a student?** heading. When you are finished, (B) click the Next button.

The screenshot shows the IBM account creation process. On the left, there is a decorative graphic of blue dots connected by lines forming a grid-like pattern. On the right, the form fields are displayed:

- Password:** A password field containing several dots, with a green eye icon and a copy/paste icon to its right.
- Country or region of residence:** A dropdown menu set to "United States of America".
- State or province:** A dropdown menu set to "North Carolina".
- Are you a student?**: A radio button group where "No" is selected (indicated by a red circle).
- Company:** A text input field containing "IBM".
- Next:** A blue button with a white arrow pointing right, highlighted with a red rectangle and a red circle containing the letter "B".
- 2. Verify email:** A collapsed section indicated by a minus sign.
- Create account:** A grey button at the bottom.

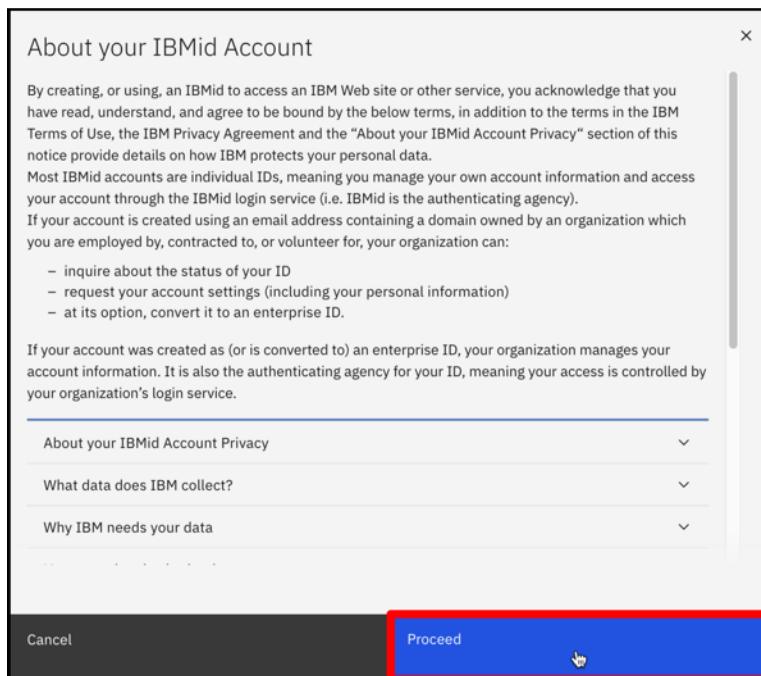
At the very bottom of the page, there is a black footer bar with links: "Contact IBM", "Privacy", "Terms of use", "Accessibility", and "Cookie Preferences".

This will cause the **Account information** section of the screen to be collapsed, and the **Verify email** section to be expanded. A 7-digit confirmation code will then be sent to the email address you provided. (The email will come from *IBM Security* and the subject will be *New User Registration*.) The confirmation code provided is only valid for 30 minutes. If, after 5 minutes, you don't see an email containing the code in your inbox, check your spam/junk email folder.

6. Once you receive an email from *IBM Security*, (A) enter the 7-digit confirmation code provided with the email in the **Verification token** field. Then, (B) locate and click the **Create account** button.

The screenshot shows the 'Create an account to access trials, demos and services.' page. At the top right, there are two sections: 'Account information' and '2. Verify email'. Under '2. Verify email', it says 'We emailed a 7 digit code to roger.sanders@us.ibm.com. This code will expire in 30 minutes.' Below this is a 'Verification token' input field containing '9468149', which is highlighted with a red rectangle and marked with a red circle labeled 'A'. Below the token field is a message: 'Didn't receive the email? Check your spam filter for an email from ibmacct@iam.ibm.com.' It also says 'You can resend the verification token in 41 seconds'. Further down, there's a checkbox for 'Email' and a link 'I accept the product [Terms and Conditions](#)'. At the bottom right is a blue 'Create account' button, which is highlighted with a red rectangle and marked with a red circle labeled 'B'.

7. When the *About your IBMid Account* popup appears, read through the terms and conditions of your IBM account. Then, when you are ready to continue, click the **Proceed** button.



Once this is done, you may see the following message on the screen: **"Please wait while we create your IBMid and set up the service subscription."**

____ 8. When your IBMid and service subscription have been created, you will receive a second email from *IBM Security* with the subject *Registration successful* telling you your IBMid has been activated. For added security, you will also receive a separate email – again from *IBM Security*, but with the subject *Verify your identity* – that contains a 6-digit email verification code. (This code can only be used once and will expire after 20 minutes.)

____ 9. When prompted, **(A)** enter the 6-digit email verification code you received in the **Enter email code** entry field. Then, **(B)** click the **Verify** button.

IBM

Enter code sent to your email

For added security, we sent a 6-digit code to **rog*****@us.ibm.com**.

Please enter the code below within 20 minutes

Enter email code

3445-.....

A

Verify

B

Didn't receive the email?

Check your spam filter for an email from

ibmacct@iam.ibm.com

[Resend code](#)

[Contact](#) [Privacy](#) [Terms of use](#) [Accessibility](#)

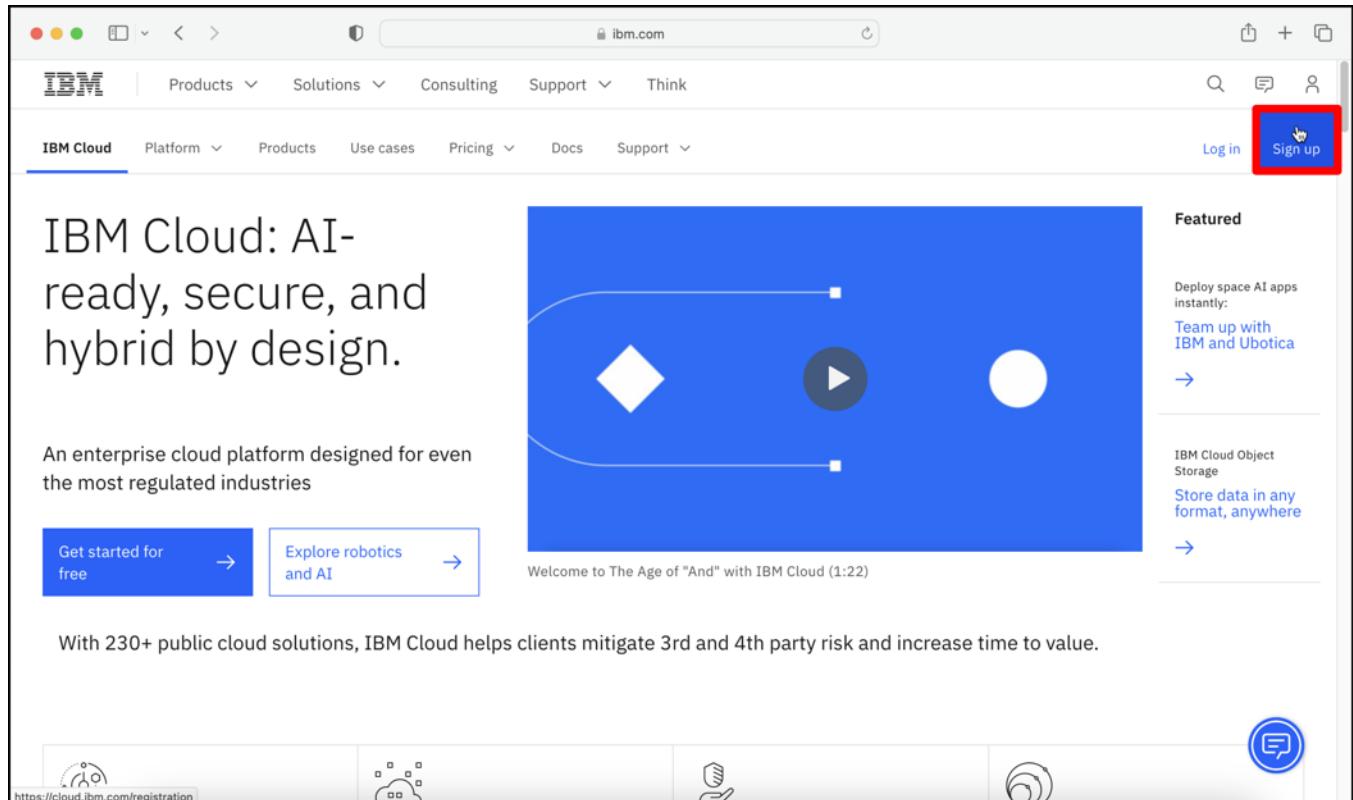
Once the email code you provided has been verified, you should see a screen that looks something like this:

The screenshot shows a web interface for managing applications. At the top, there is a navigation bar with links for 'IBMid', 'App center', and 'My accesses'. On the right side of the header are icons for notifications and user profile. Below the header, the page title is 'My apps'. A search bar contains the placeholder text 'What app are you looking for?'. To the right of the search bar are buttons for 'Sort by A-Z' and a filter icon. The main content area displays a message: 'No applications are available.' accompanied by an illustration of a smartphone, a laptop, and a plant. Below the illustration, a sub-message states 'You do not have any apps configured at this time.' At the bottom of the page, there is a copyright notice: '© Copyright IBM Corp. 2017, 2020. All rights reserved.'

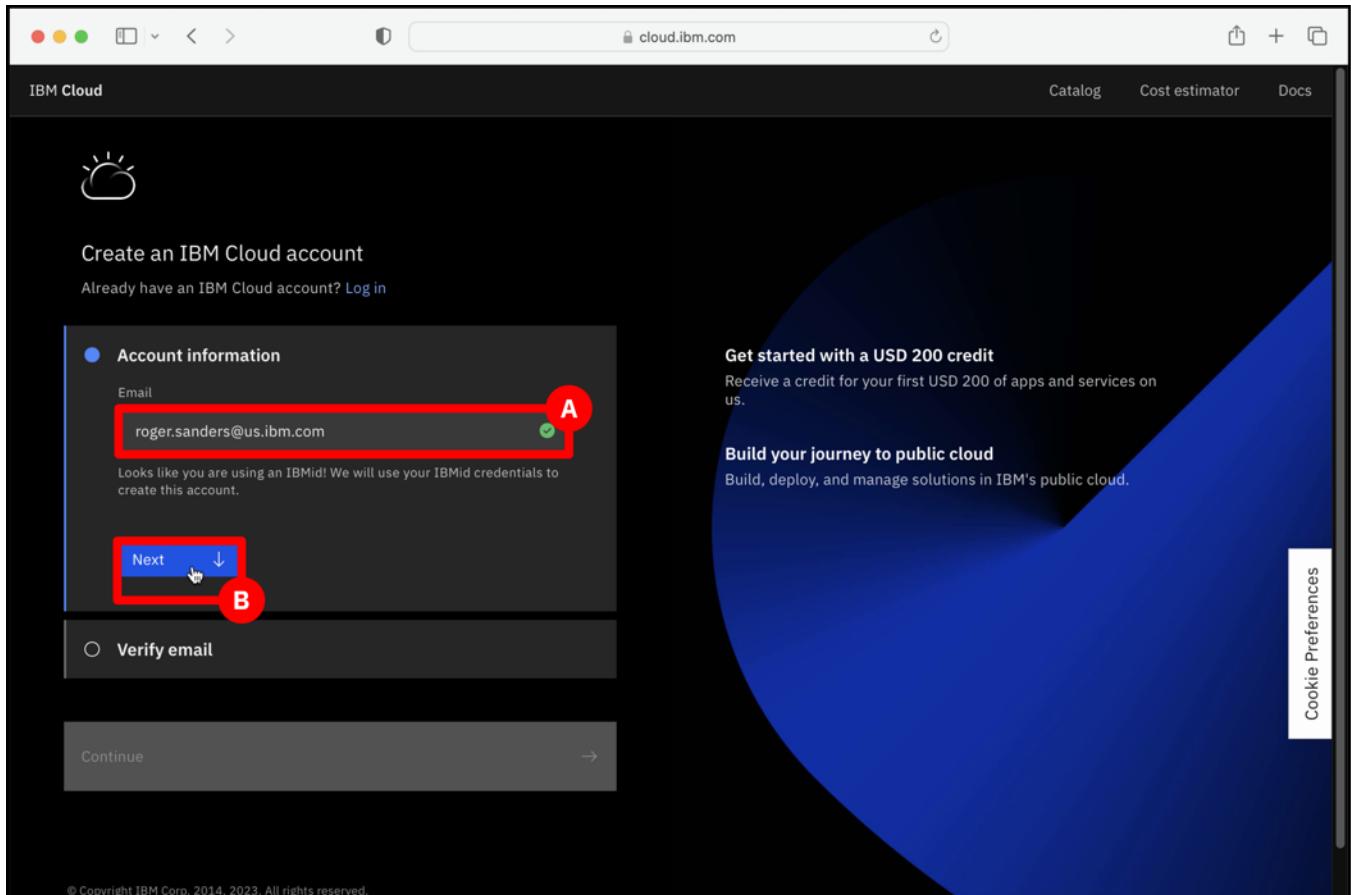
Congratulations! You now have an **IBMid**, which you will now use to create an IBM Cloud account (and later, to create an IBM Cloud Pak for Data as a Service account).

2. Create an IBM Cloud account.

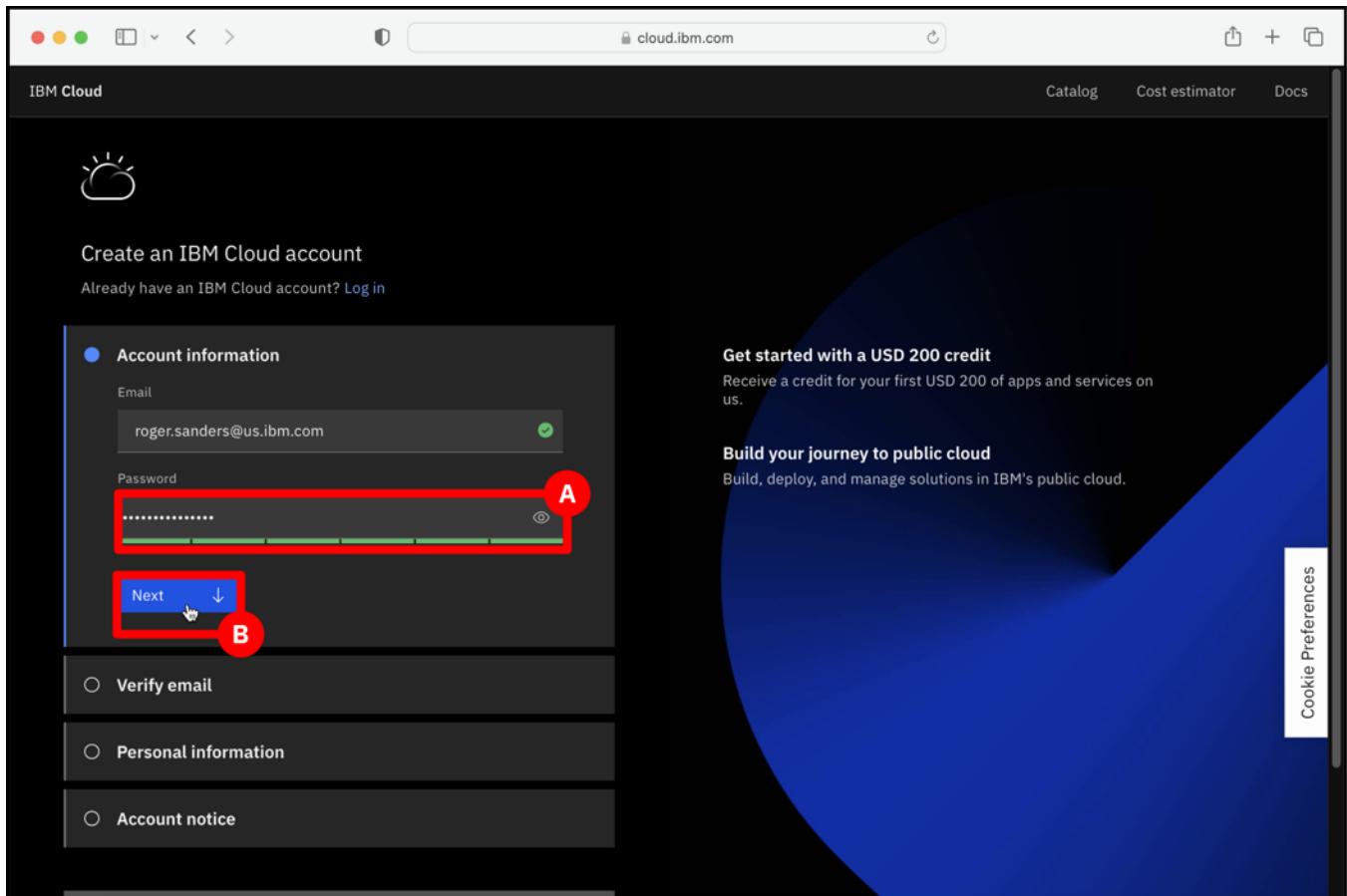
1. Open a web browser and go to the *IBM Cloud* web site (<https://www.ibm.com/cloud>). Once there, select the **Sign up** button located in the top right corner of the screen.



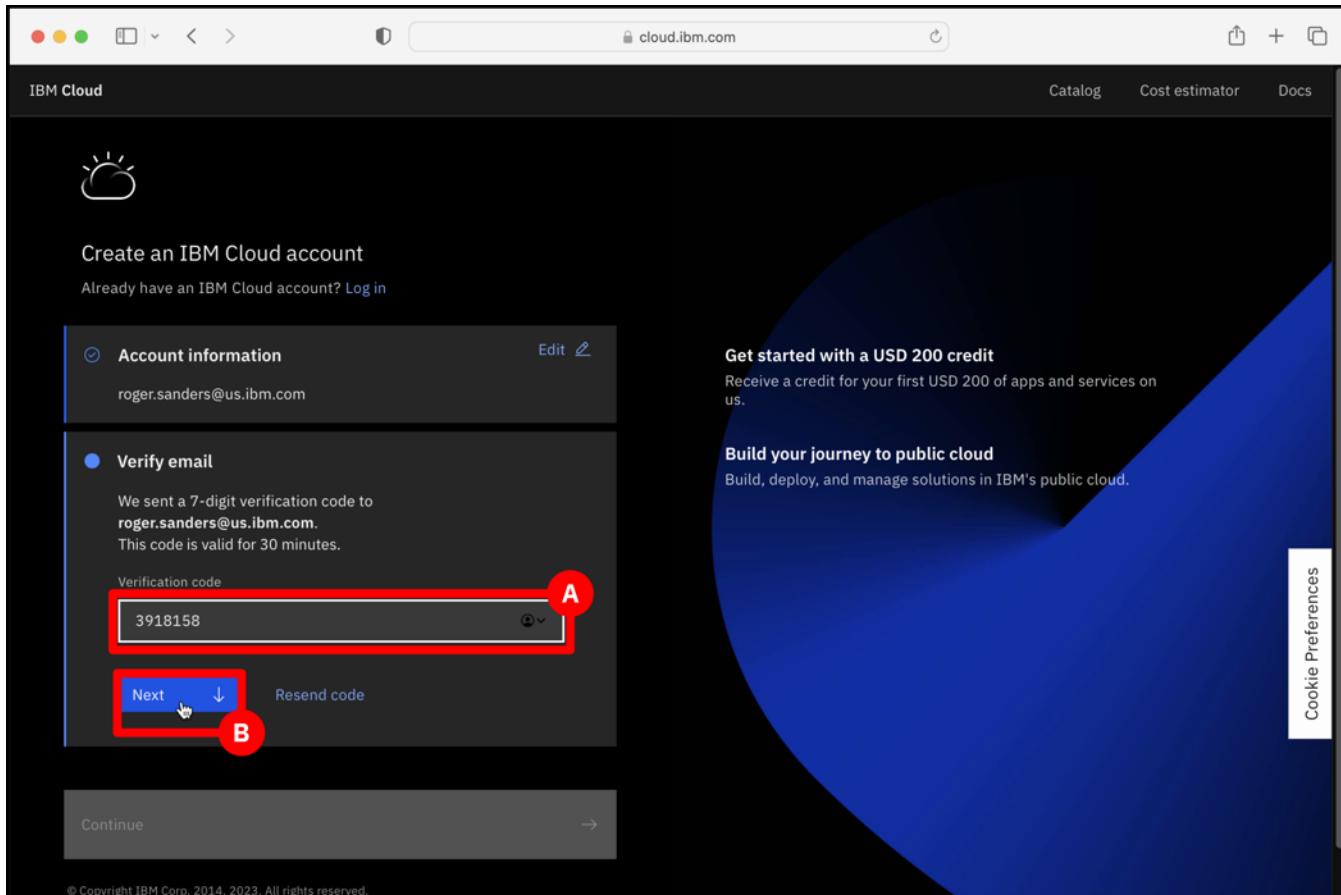
_____ 2. When the *Create an IBM Cloud account* screen is displayed, (A) enter the email address you used to create your IBMid in the **Email** field (which is located inside the **Account information** box). (B) Then, click the **Next** button.



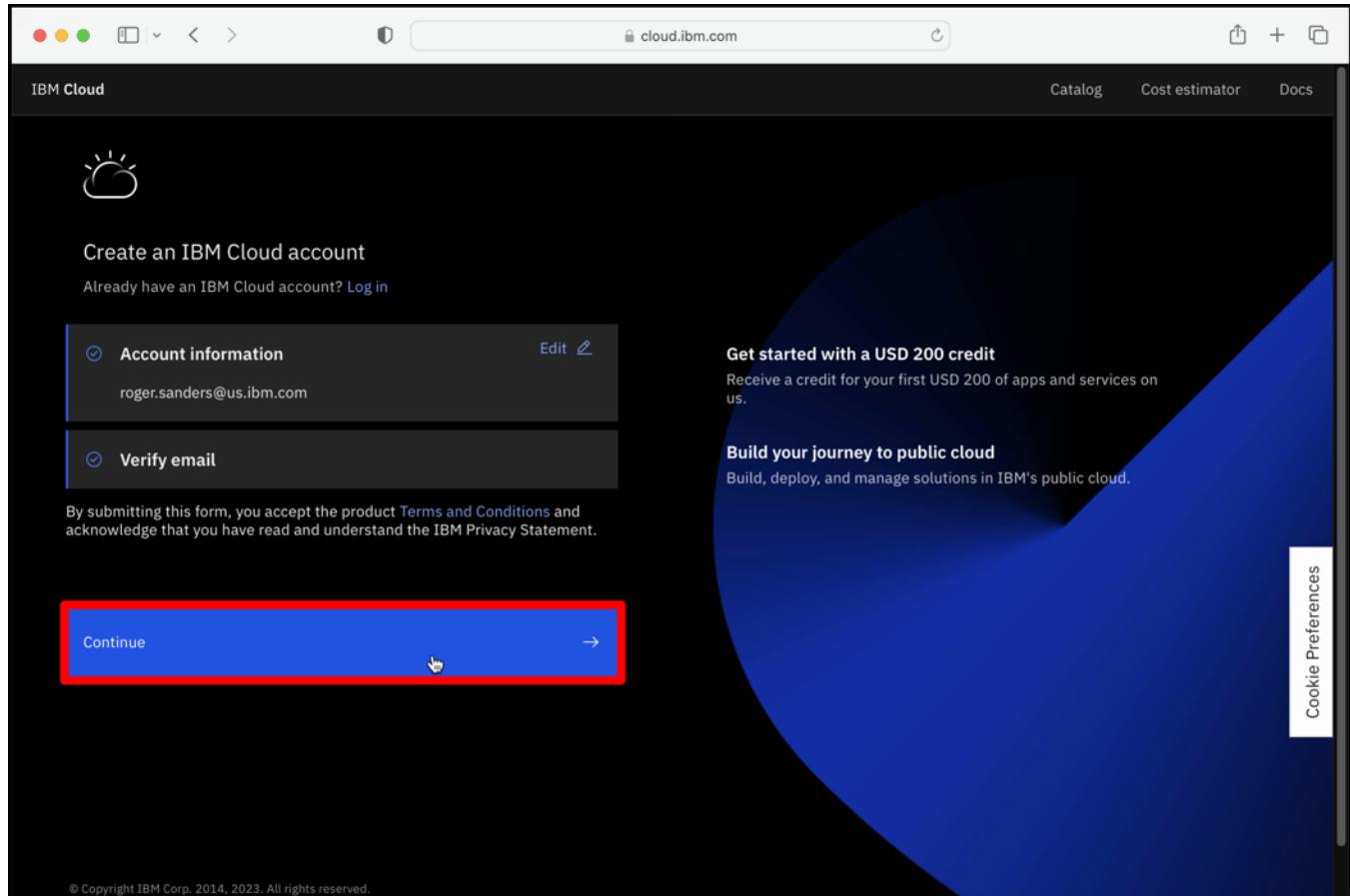
_____ 3. When the **Password** field appears, (A) enter the password associated with the email address (IBMid) provided in the previous step and (B) click the **Next** button. This will cause a 7-digit verification code to be sent to the email address specified. (The email will come from *IBM Cloud* and the subject will be *Your IBM Cloud verification code*.)



____ 4. Once you receive an email containing a verification code, (A) enter the code in the **Verification code** field and (B) click the **Next** button. (Note: The verification code is only valid for 30 minutes. If, after 5 minutes, you don't see an email containing the code in your inbox, check your spam/junk email folder.)



____ 5. After the verification code you entered in the **Verification code** field is authenticated, the **Continue** button located at the bottom of the screen will be enabled. (Until then, this button is disabled.) When the **Continue** button is enabled, click it. This will cause the current screen to be replaced with the *Review your account privacy notice* screen.



____ 6. When the *Review your account privacy notice* screen is displayed, read through the terms and conditions of your IBMid Account and then **(A)** click the **I acknowledge that I understand how IBM is using my Basic Personal Data ...** checkbox. **(B)** Click the **Continue** button to continue.

The screenshot shows the 'Review your account privacy notice' page on the IBM Cloud website. The page includes sections like 'About your IBMid Account', 'What data does IBM collect?', and 'About your IBMid Account Privacy'. A red box labeled 'A' highlights the checkbox for acknowledging data use. A red box labeled 'B' highlights the 'Continue' button at the bottom.

Review your account privacy notice

About your IBMid Account

By creating, or using, an IBMid to access an IBM Web site or other service, you acknowledge that you have read, understand, and agree to be bound by the below terms, in addition to the terms in the IBM Terms of Use, the IBM Privacy Agreement and the "About your IBMid Account Privacy" section of this notice provide details on how IBM protects your personal data.

Most IBMid accounts are individual IDs, meaning you manage your own account information and access your account through the IBMid login service (i.e. IBMid is the authenticating agency).

If your account is created using an email address containing a domain owned by an organization which you are employed by, contracted to, or volunteer for, your organization can:

- inquire about the status of your ID
- request your account settings (including your personal information)
- at its option, convert it to an enterprise ID.

If your account was created as (or is converted to) an enterprise ID, your organization manages your account information. It is also the authenticating agency for your ID, meaning your access is controlled by your organization's login service.

[Last updated: 2021-10-18]

What data does IBM collect?

Why IBM needs your data

How your data is obtained

How IBM uses your data

How IBM protects your data

How long we keep your data

About your IBMid Account Privacy

I acknowledge that I understand how IBM is using my Basic Personal Data and (if applicable) how my organization may become the authenticating agency for my IBMid account. I certify that my age is at least the age of consent for my country of residence.

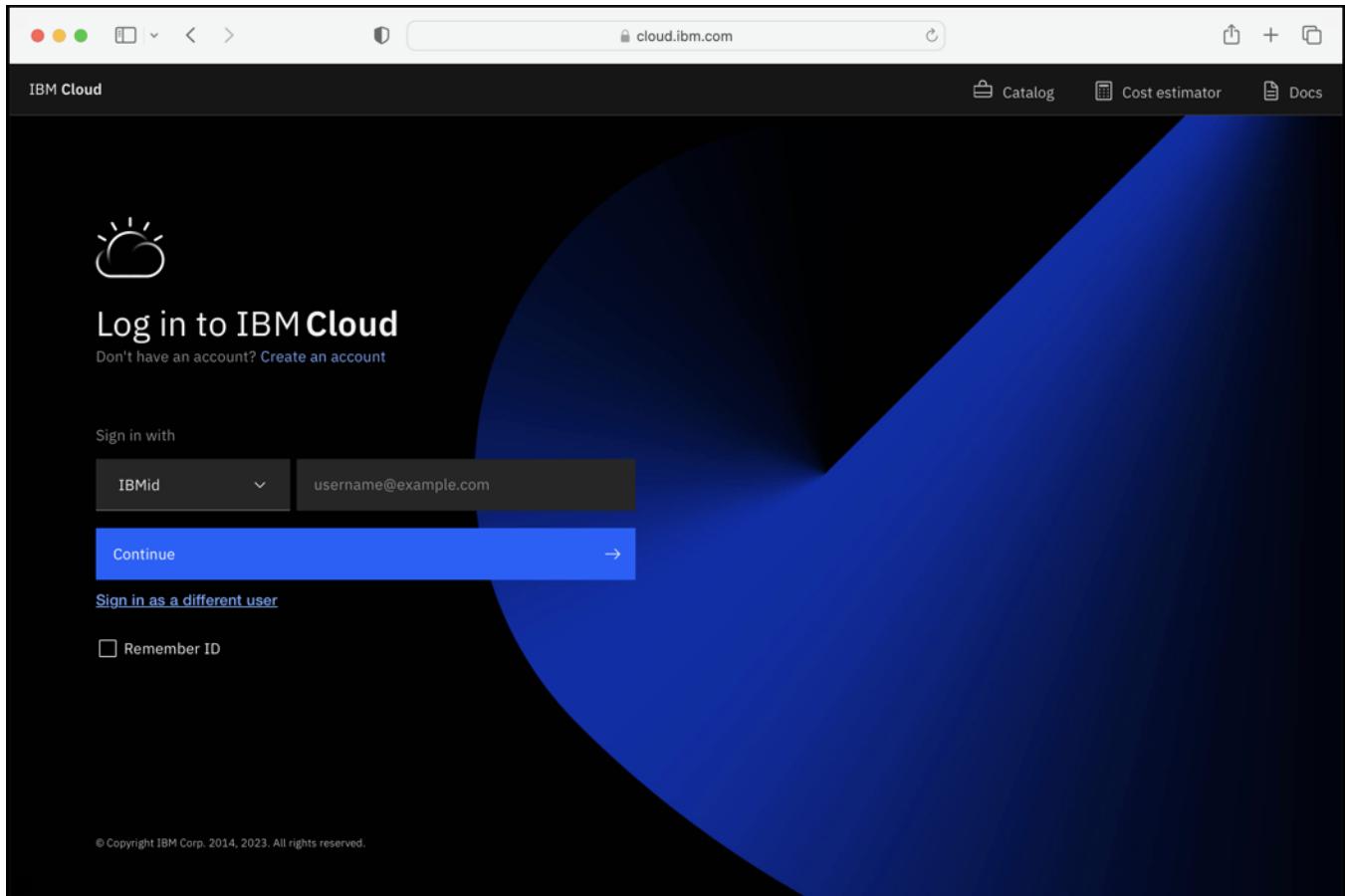
Continue →

© Copyright IBM Corp. 2014, 2023. All rights reserved.

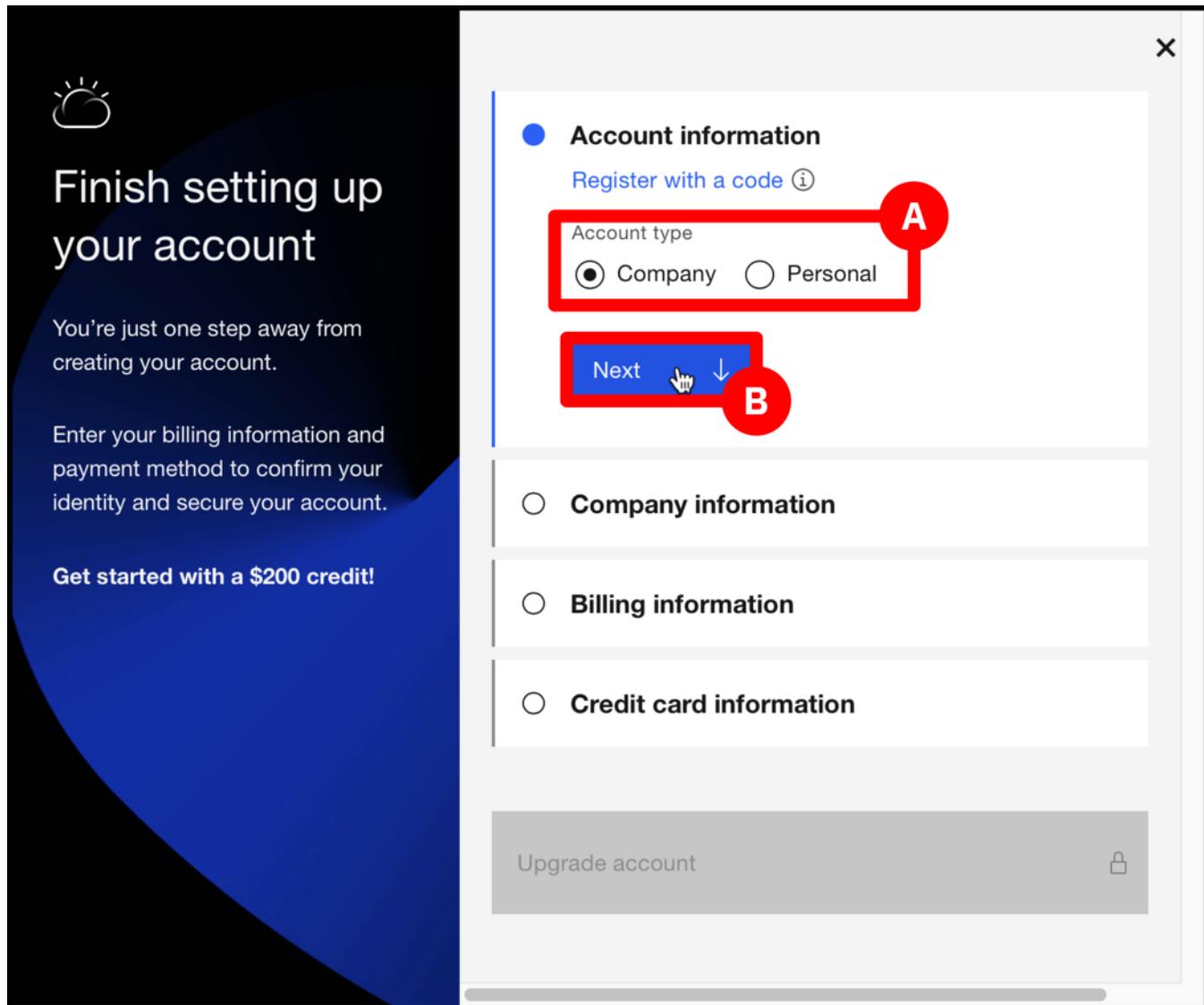
Cookie Preferences

The **Continue** button should be replaced with the message “*Creating your account ...*”

7. Once your IBM Cloud account has been created, you will be presented with a *Log in to the IBM Cloud* screen that looks something like this:



_____ 8. Log in to your IBM Cloud account using the IBMid and password you provided earlier. If this is your first time logging into your account, a popup window will appear and you will be asked to provide additional information needed to activate the account. Start by (A) clicking the **Company** or **Personal** radio-button to specify the **Account type**. After you have made your selection, (B) click the **Next** button.



____ 9. If you indicated that your IBM Cloud account should be treated as a **Company** account, supply the appropriate **Company information** (i.e., **Company name** and company contact information) in the fields provided.

The screenshot shows a dark-themed account setup interface. On the left, a sidebar displays a cloud icon, the text "Finish setting up your account", a message about being one step away from account creation, and a promotional offer for \$200 credit. The main area is titled "Company information" and contains the following fields:

- Company name: IBM
- Country or region: United States
- Address line 1: 2100 Stewart Street
- Address line 2 (optional): The apartment, suite, unit, building or floor
- City: Foreign Words
- State: North Carolina
- Zip code: 27526-0000
- Phone number: +1 919 700-0799

____ 10. After you have provided all the **Company information** requested (assuming this is a Company account), click the **Next** button.

The screenshot shows a user interface for account setup. On the left, a dark blue sidebar features a cloud icon at the top, followed by the text "Finish setting up your account", "You're just one step away from creating your account.", "Enter your billing information and payment method to confirm your identity and secure your account.", and "Get started with a \$200 credit!". The main area is titled "City" with a placeholder "Fresno, CA". Below it are "State" (North Carolina) and "Zip code" (37026-0000). The "Phone number" field contains "+1 (800) 729-2776" with a dropdown for country code and a verify button. A large red box highlights the "Next" button. Below the phone number, two radio buttons are shown: "Billing information" and "Credit card information". At the bottom is a grey "Upgrade account" button with a lock icon.

City

Fresno, CA

State Zip code

North Carolina 37026-0000

Phone number

+1 (800) 729-2776

Next

Billing information

Credit card information

Upgrade account

11. Now provide the appropriate **Billing information**. If this is a **Company** account and you are the billing contact, click the **My billing contact is the same as my company contact** checkbox. Otherwise, provide the appropriate **Billing information** (i.e., contact information for the individual responsible for paying any charges incurred) in the fields shown.

Finish setting up your account

You're just one step away from creating your account.

Enter your billing information and payment method to confirm your identity and secure your account.

Get started with a \$200 credit!

Billing information

First name
Roger

Last name
Sanders

My billing contact is the same as my company contact

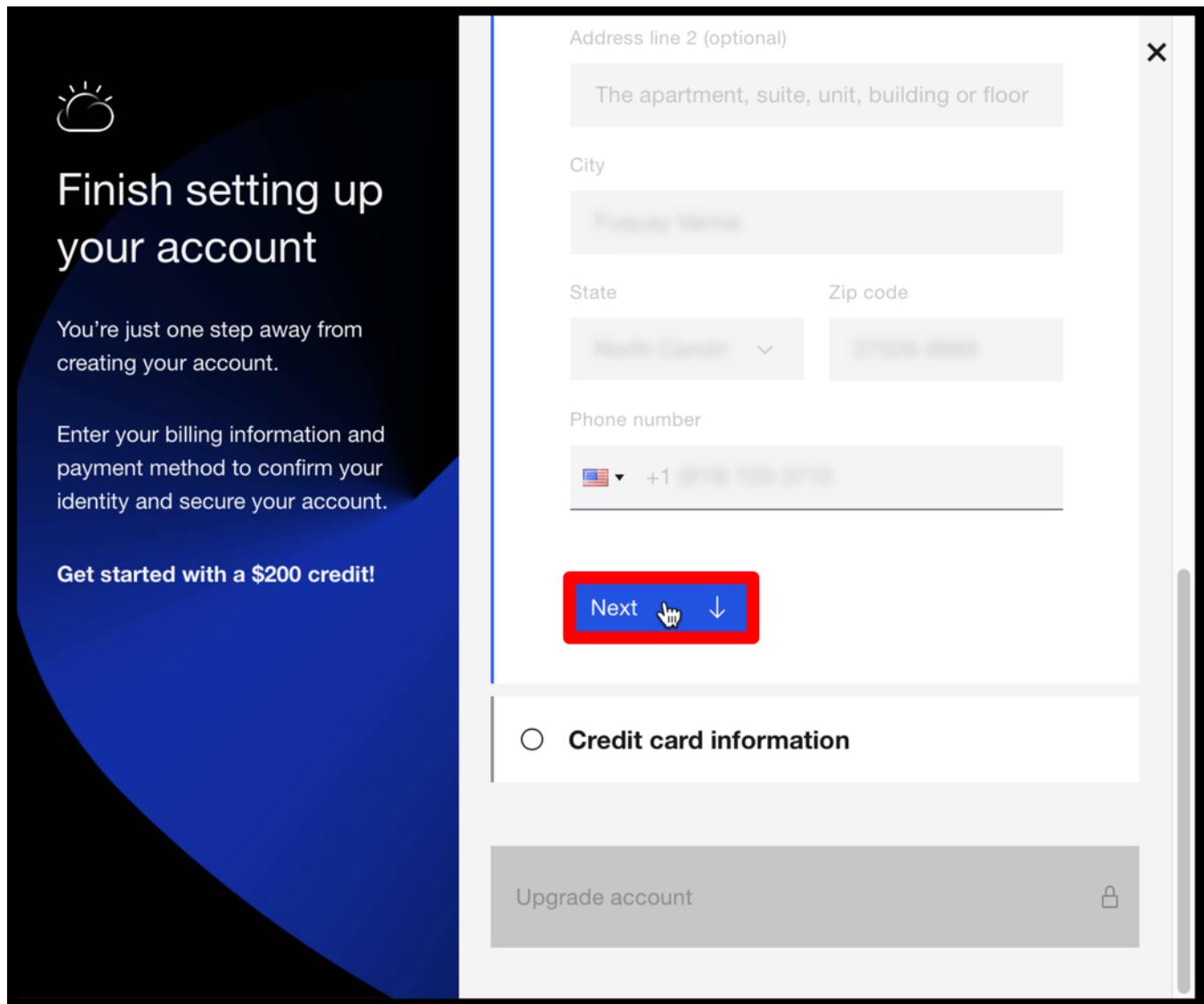
Country or region
United States

Address line 1
[redacted]

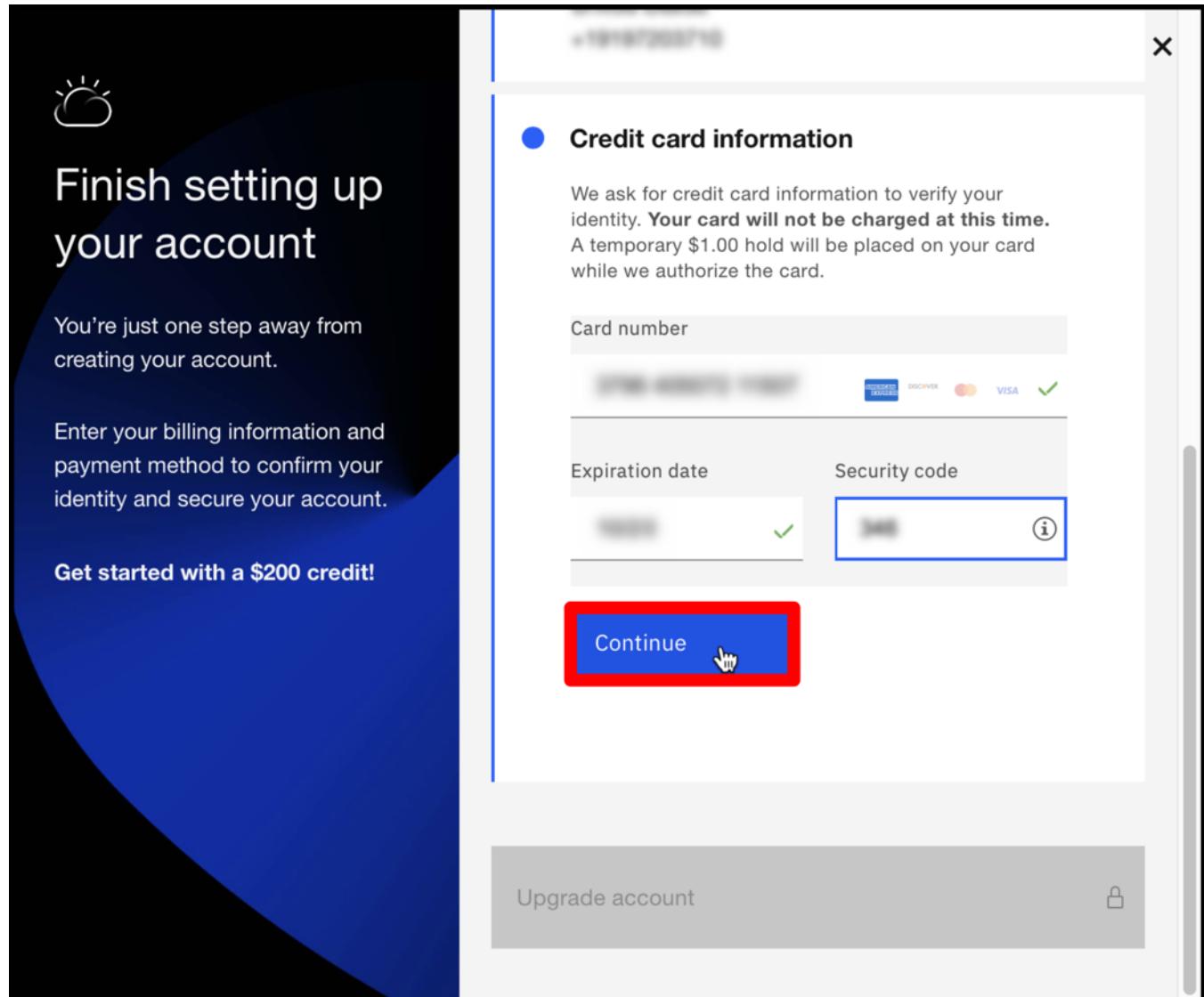
Address line 2 (optional)
The apartment, suite, unit, building or floor

City
[redacted]

12. Once you have confirmed the appropriate **Billing information** has been provided, click the **Next** button.



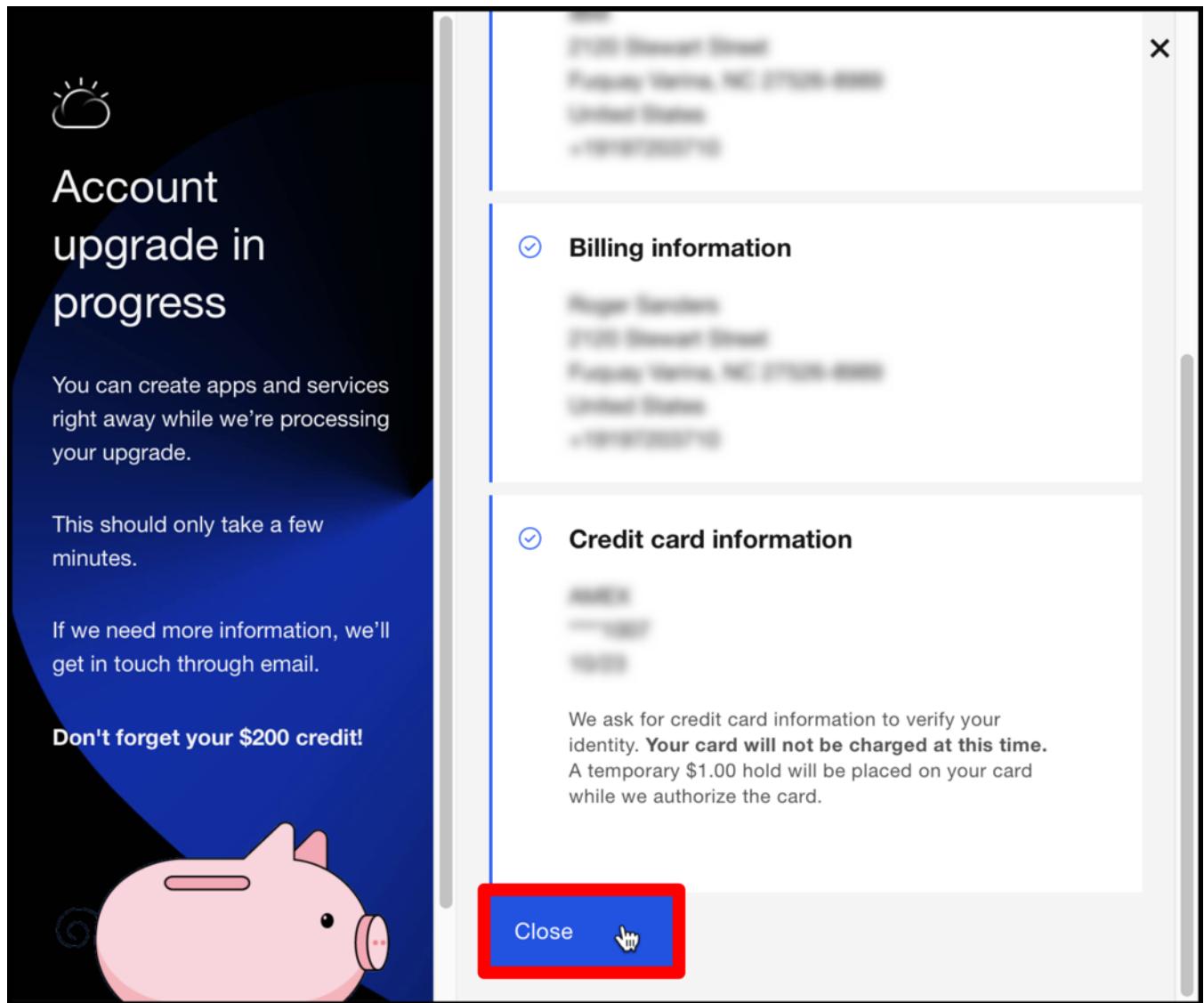
13. Now, enter the appropriate **Credit card information** (i.e., **Card number**, **Expiration date**, and **Security code**) in the fields provided. It is important to note that **this information is only used to verify your identity (therefore, the name on the credit card used MUST match the name provided for the billing contact)**. Your card will **NOT** be charged unless you decide to purchase a billable IBM Cloud service. **Because this workshop uses the no-charge IBM Db2 “Lite” service and the IBM Watson Studio “Lite” service, NO charges will be incurred.** After you have provided the credit card information requested, click the **Continue** button. This will cause the **Upgrade account** button located at the bottom of the screen to be enabled.



14. Once the **Upgrade account** button is enabled, click it. This will cause a temporary \$1.00 US hold to be placed on the credit card provided for authorization purposes. (This will also cause the **Upgrade account** button to be replaced with a message saying, “**Authorizing your credit card ...**”.)

The screenshot shows a user interface for account setup. On the left, a dark sidebar with a cloud icon contains the text "Finish setting up your account", "You're just one step away from creating your account.", "Enter your billing information and payment method to confirm your identity and secure your account.", and "Get started with a \$200 credit!". The main area has a header "Roger Sanders" with an "X" icon. A section titled "Credit card information" with an "Edit" link is shown. Below it, a note states: "We ask for credit card information to verify your identity. Your card will not be charged at this time. A temporary \$1.00 hold will be placed on your card while we authorize the card." A callout box contains the same note. At the bottom is a blue button labeled "Upgrade account" with a lock icon, which is highlighted with a red rectangle.

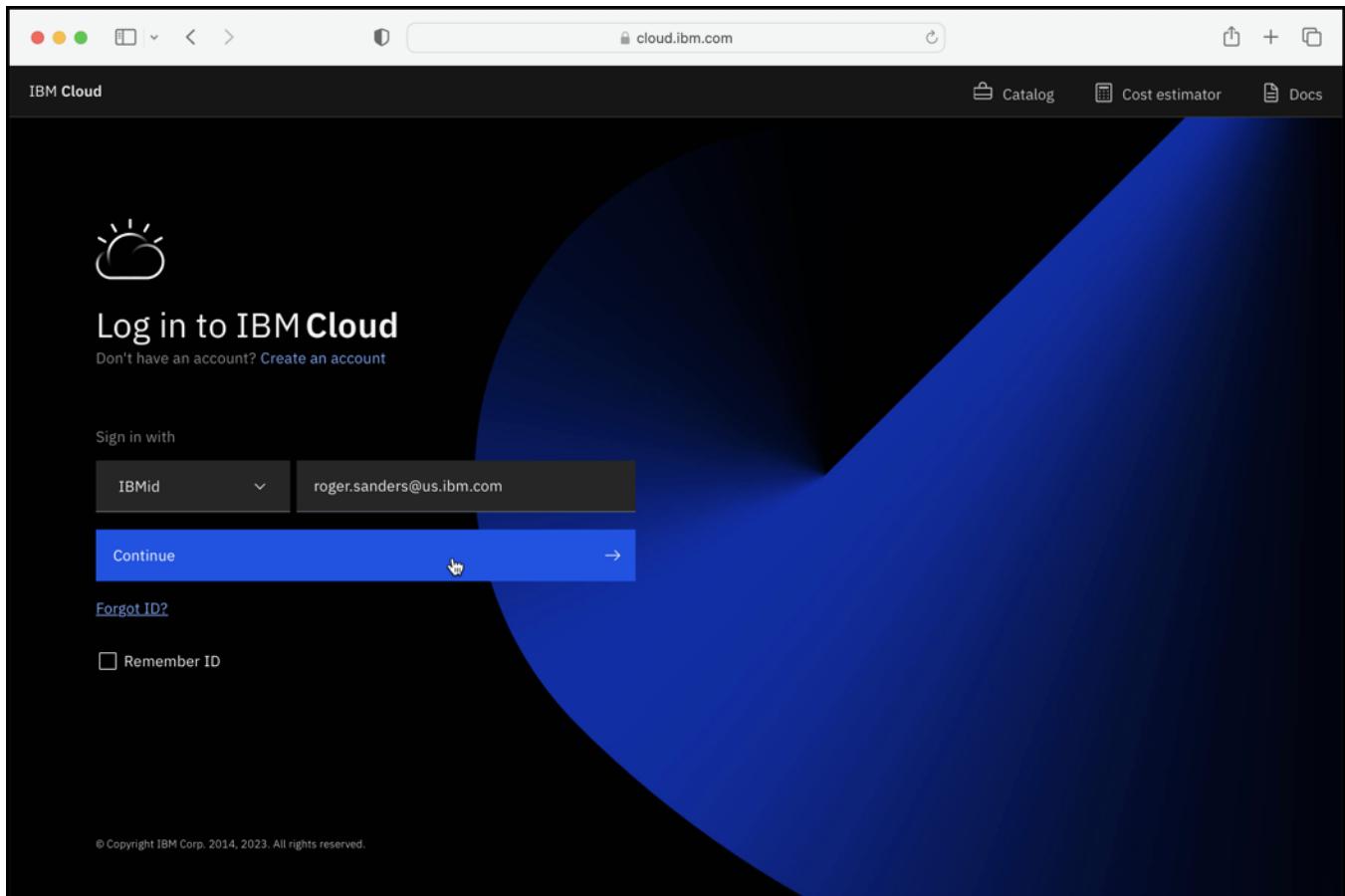
15. When the credit card authorization process is complete, a **Close** button will be displayed at the bottom of the popup window. Click on this button. (When your IBM Cloud account is ready, you should receive an email from *IBM Cloud* saying “Welcome to IBM Cloud! Let’s get started.”)



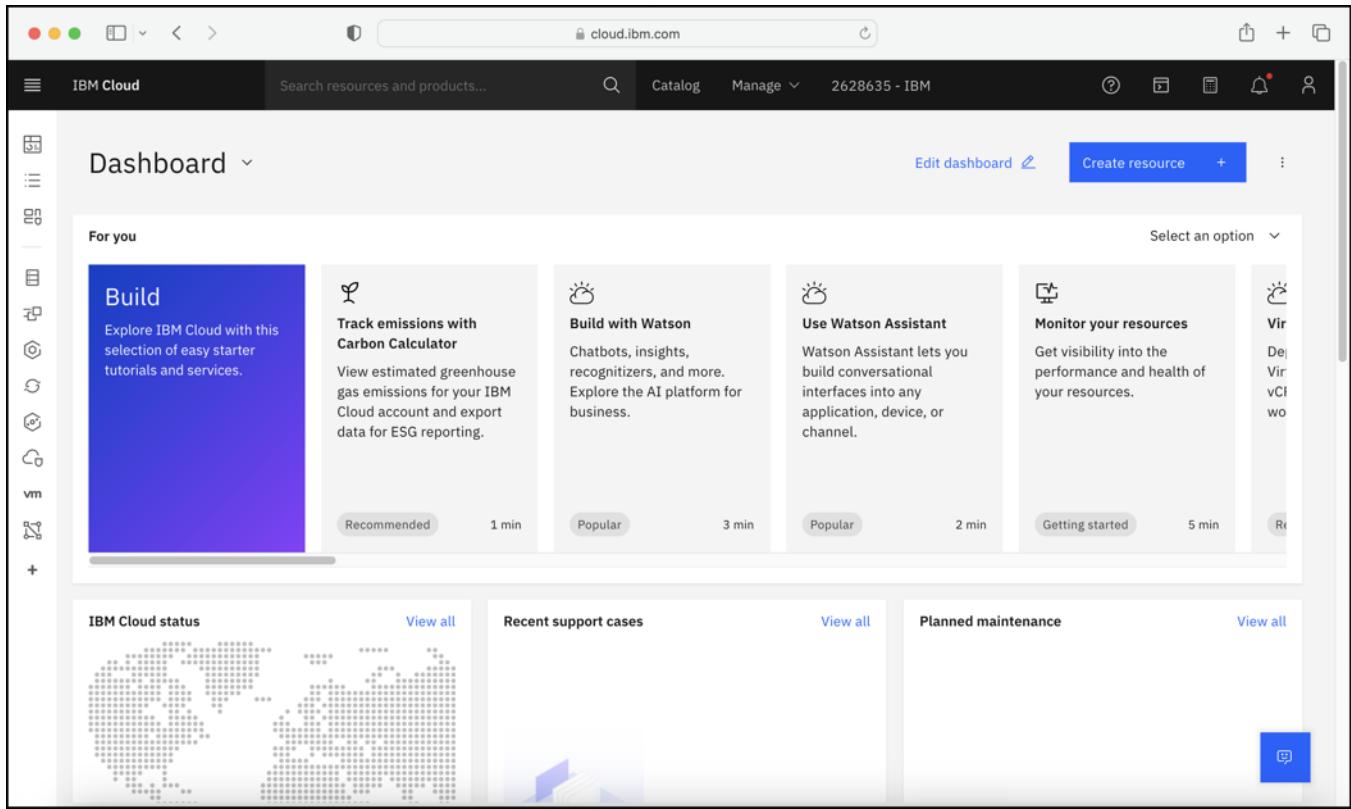
Congratulations! You now have an active **IBM Cloud** account that can be used to perform the rest of the exercises in this workshop.

3. Provision a Db2 on Cloud Lite plan database

____ 1. Open a web browser and go to the *IBM Cloud* login web site (<https://cloud.ibm.com/login>). Once there, ensure the **Sign in with** dropdown list contains the value **IBMid** and enter the email address associated with your IBMid in the field immediately beside the dropdown list. When finished, click the **Continue** button. (Note: If you do not have a valid IBMid and an IBM Cloud account, complete steps 1 and 2 of this exercise before continuing.)

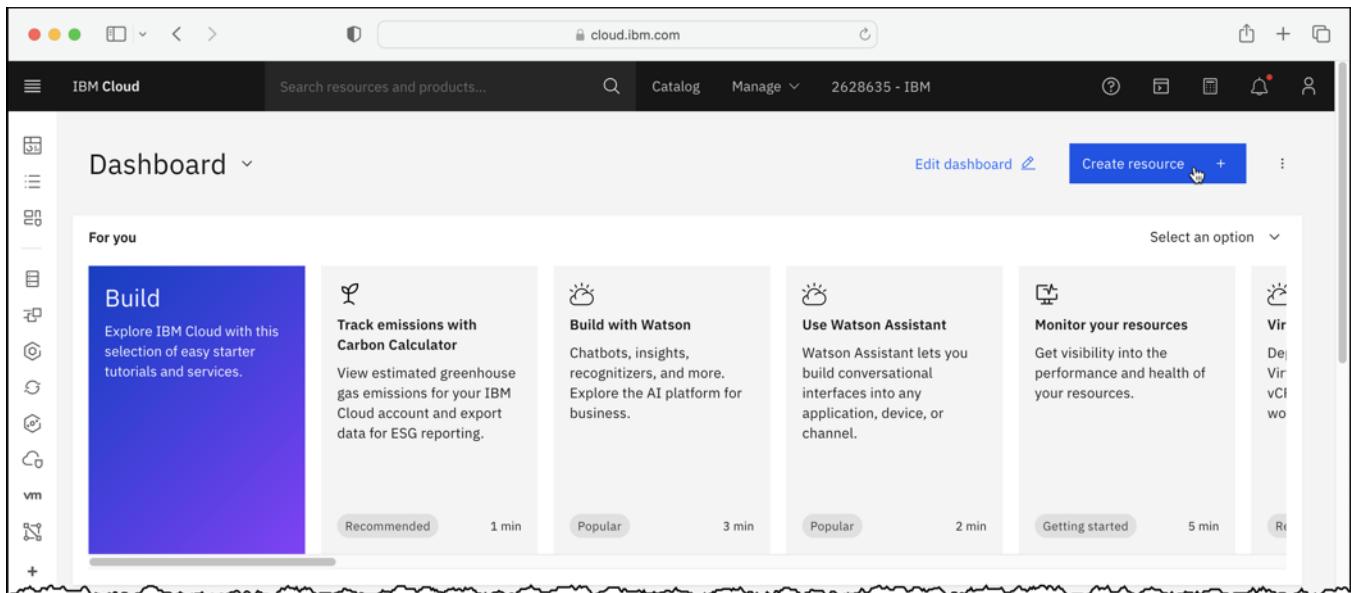


____ 2. If prompted, enter the password associated with the IBMid provided. Once you are logged in, you should see a screen that looks something like this:



The screenshot shows the IBM Cloud dashboard. At the top, there's a navigation bar with 'IBM Cloud' and a search bar. Below the navigation is a 'Dashboard' section with a 'For you' sidebar containing icons for 'Build', 'Track emissions with Carbon Calculator', 'Build with Watson', 'Use Watson Assistant', 'Monitor your resources', and 'Vir'. The main area has cards for 'IBM Cloud status' (a map of the world), 'Recent support cases', and 'Planned maintenance'.

____ 3. Click on the **Create resource** button located at the top right corner of the screen. This will cause the IBM Cloud Catalog screen to be displayed.



This screenshot is identical to the one above, but the 'Create resource' button in the top right corner is highlighted with a cursor, indicating it is the target for the next click.

____ 4. When the IBM Cloud Catalog screen appears, [A] type the characters “db2” in the Search field located in the top left corner of the screen (i.e., the entry field with the magnifying glass icon on the left.) This will cause a list of the Db2 resources available on IBM Cloud to be displayed. Then, [B] click on the item **Db2** in this list. This will cause the *Db2* resource screen to be displayed.

The screenshot shows the IBM Cloud Catalog interface. At the top, there is a search bar with the placeholder "Search resources and products..." and a magnifying glass icon. To the right of the search bar are buttons for "Catalog", "Manage", and "2628635 - IBM". Further right are icons for help, file, settings, notifications, and user profile. Below the search bar, the word "Catalog" is displayed. A sidebar on the left lists categories: "Category" (Compute, Containers, Networking, Storage, Enterprise applications, AI / Machine Learning, Analytics), "IBM Cloud catalog" (Compute, Containers, Networking, Storage, Enterprise applications, AI / Machine Learning, Analytics), and "Relevance" with sorting options. The main area shows a 3D isometric illustration of a data center with servers, databases, and people. In the search results, the term "db2" has been typed into the search bar, and a dropdown menu appears with "Db2" highlighted. A green callout labeled "A" points to the search bar, and another labeled "B" points to the "Db2" item in the dropdown. Other items in the dropdown include "Db2 Warehouse", "SAP NetWeaver(ABAP) Linux/Db2 standard on V...", and "Secure Automated Backup with Compass". Below the search results, there are three card-like entries: "Cloud Object Storage on VPC for SAP HANA Backup" by IBM, "DevSecOps Application Lifecycle Management" by IBM, and "Maximo Application Suite" by IBM.

5. On the *Db2* resource screen, [A] use the **Select a location** dropdown list to choose the location where the *Db2* on Cloud database is to be created. For this workshop, either the **Dallas (us-south)** or the **London (eu-gb)** location must be chosen – that's because these are the only locations that offer the *Db2* on Cloud “Lite” (free) plan. [B] Next, choose the **Lite** plan from the **Select a pricing plan** list shown on the screen. [C] Once you have made these selections, indicate that you have read and agree to the license agreements for using *Db2* on Cloud by clicking on the **I have read and agree ...** checkbox. This will enable the **Create** button. (The license agreements can be reviewed by clicking on the **Terms** link located under this checkbox.) [D] When ready, click the **Create** button.

Catalog /

Db2

A fully managed, highly-performant relational data store running the enterprise-class Db2 database engine.

Create **About**

Type: Service
Provider: IBM
Last updated: 04/18/2024
Category: Databases
Compliance: EU Supported, HIPAA Enabled, IAM-enabled
Location: Sydney, Frankfurt, London, Dallas, São Paulo, Toronto, Tokyo, Milan 01

Select a location: **Dallas (us-south)**

Select a pricing plan:
Displayed prices do not include tax. Monthly prices shown are for country or location: United States

Plan	Features and capabilities	Pricing
Lite	200 MB of data storage 5 simultaneous connections Shared multitenant system	Free

The Free plan provides a free *Db2* service for development and evaluation. The plan has a set amount of limitations as shown. You can continue using the free plan for as long as needed, however, users are asked to re-extend their free account every 90 days by email. If you do not re-extend, your free account is cleaned out a further 90 days later. This helps provide free resources for everyone.

Lite plan services are deleted after 30 days of inactivity.

Summary

Db2 **Free**

Location: Dallas
Plan: Lite
Service name: Db2-**pc**
Resource group: Default

I have read and agree to the following license agreements:
[Terms](#)

Create

_____ 6. When the **Create** button on the *Db2* resource screen is pressed, this screen will disappear and the message “**Creating ...**” will be displayed in its place. Shortly afterwards, the *Db2* resource screen will be replaced with an IBM Cloud *Resource list* screen. This screen contains a list of every resource/service that has been provisioned under your IBM Cloud account. If you expand the **Databases** item in this list, the *Db2* on Cloud database just provisioned can be seen. Notice that the **Status** column for this database contains the value “**Provision in progress.**”

The screenshot shows the IBM Cloud Resource list interface. At the top, there is a navigation bar with icons for Home, Catalog, Manage, and a user profile. Below the navigation bar is a search bar and a "Create resource" button. The main area is titled "Resource list" and contains a table with columns: Name, Group, Location, Product, Status, and Tags. The "Status" column for the first row (Db2-pc) shows "Provision in p..." with a tooltip indicating "Provision in progress". The "Name" column lists various service categories: Compute (0), Containers (0), Networking (0), Storage (0), Converged infrastructure (0), Enterprise applications (0), AI / Machine Learning (0), Analytics (0), Blockchain (0), and Databases (1). The "Databases" row is expanded, showing sub-items: Db2-pc, Developer tools (0), and Logging and monitoring (0).

_____ 7. Refresh the web browser screen periodically until the **Status** column for the new *Db2* on Cloud database contains the value “**Active.**” At this point, your database is ready for use.

This screenshot is identical to the one above, showing the IBM Cloud Resource list. However, the "Status" column for the Db2-pc database now displays a green checkmark icon followed by the word "Active", indicating that the provisioning process has completed successfully.

_____ 8. Access the newly created Db2 on Cloud database by clicking on the database ID shown in the **Resource list**. (In this example, the database ID is *Db2-pc*. As we will see later, the name of this database is *bludb*.)

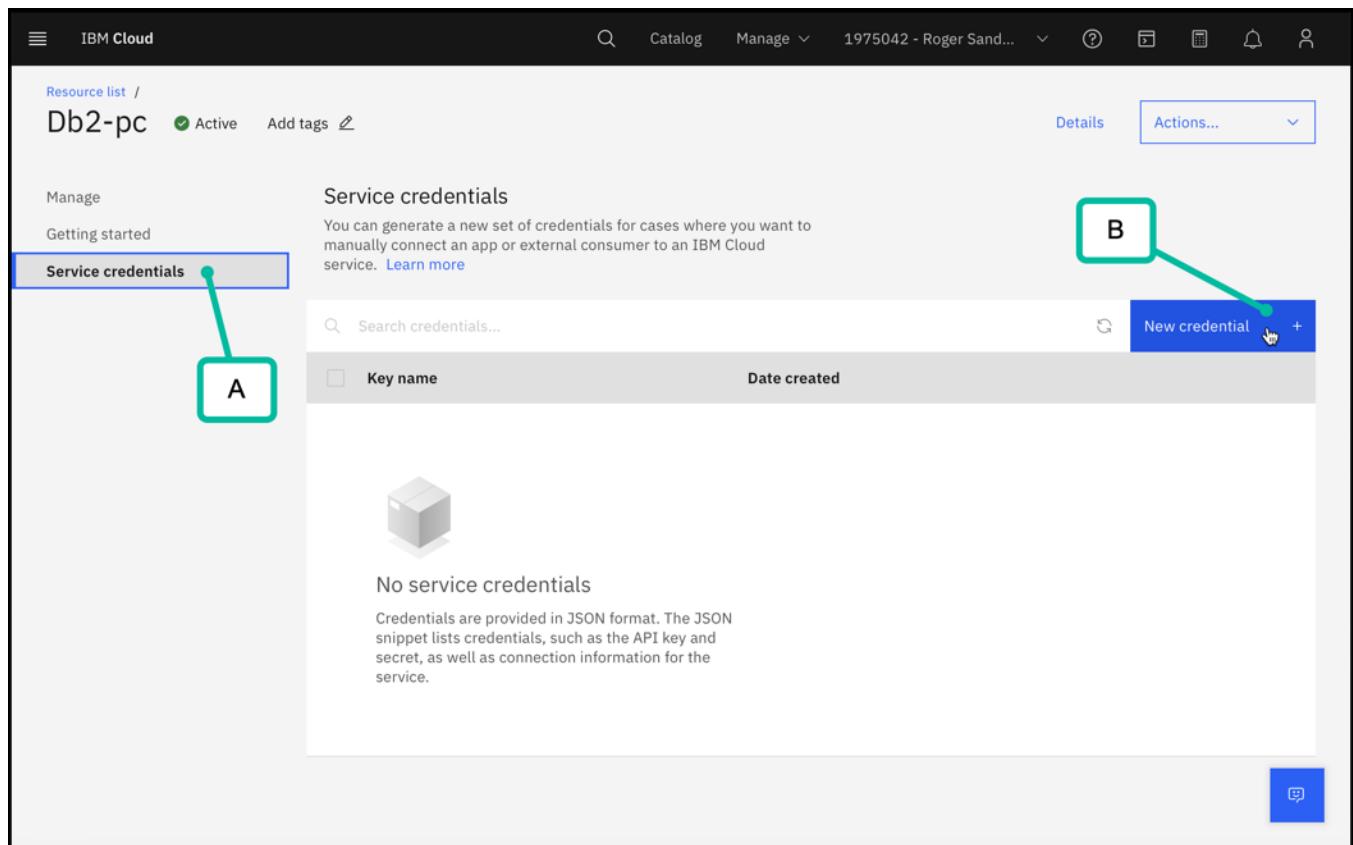


_____ 9. Clicking on a resource ID in the **Resource list** will cause the screen to be replaced with a high-level “management” screen for the resource selected. (In this example, the resource selected is the Db2 on Cloud database with the ID *Db2-pc*). If the resource chosen was a Db2 on Cloud database, once this “management” screen appears, you can begin working directly with the database, [but only through the Db2 on Cloud User Interface \(UI\)](#).

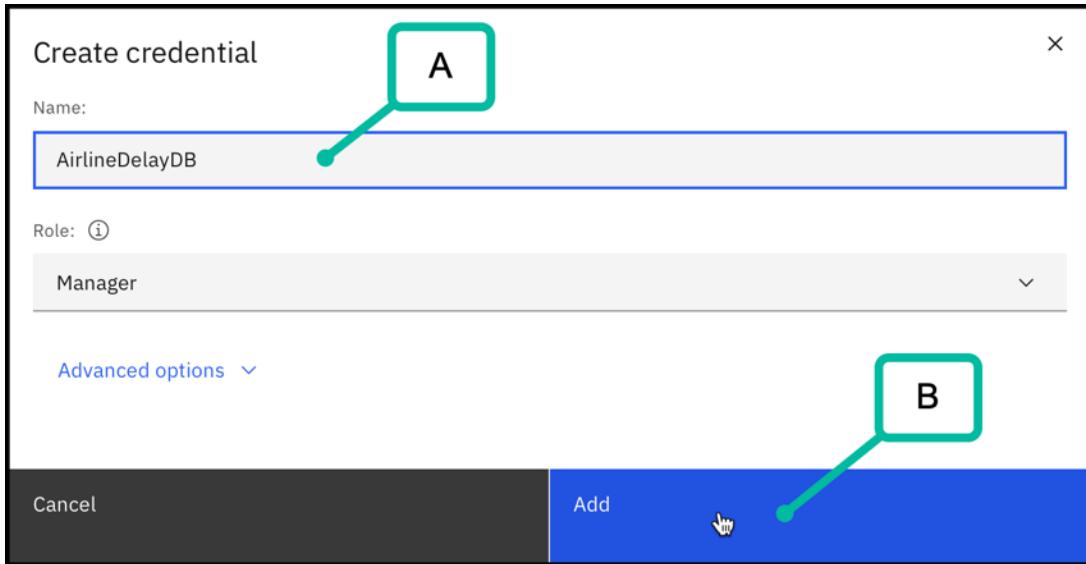
A screenshot of the 'Db2-PC' management page. At the top, there's a header with 'IBM Cloud', a search bar, and various navigation links. Below the header, the resource ID 'Db2-PC' is displayed along with its status ('Active') and a 'Add tags' button. On the left, a sidebar titled 'Manage' has sections for 'Getting started' and 'Service credentials'. The main content area is divided into three columns: 'Getting started' (with a 'Go to UI' button), 'Need help?' (with a 'Support case' button), and 'Getting started docs' (with a link). The entire page has a light gray background with blue highlights for buttons and links.

4. Create a service credential for the Db2 on Cloud database.

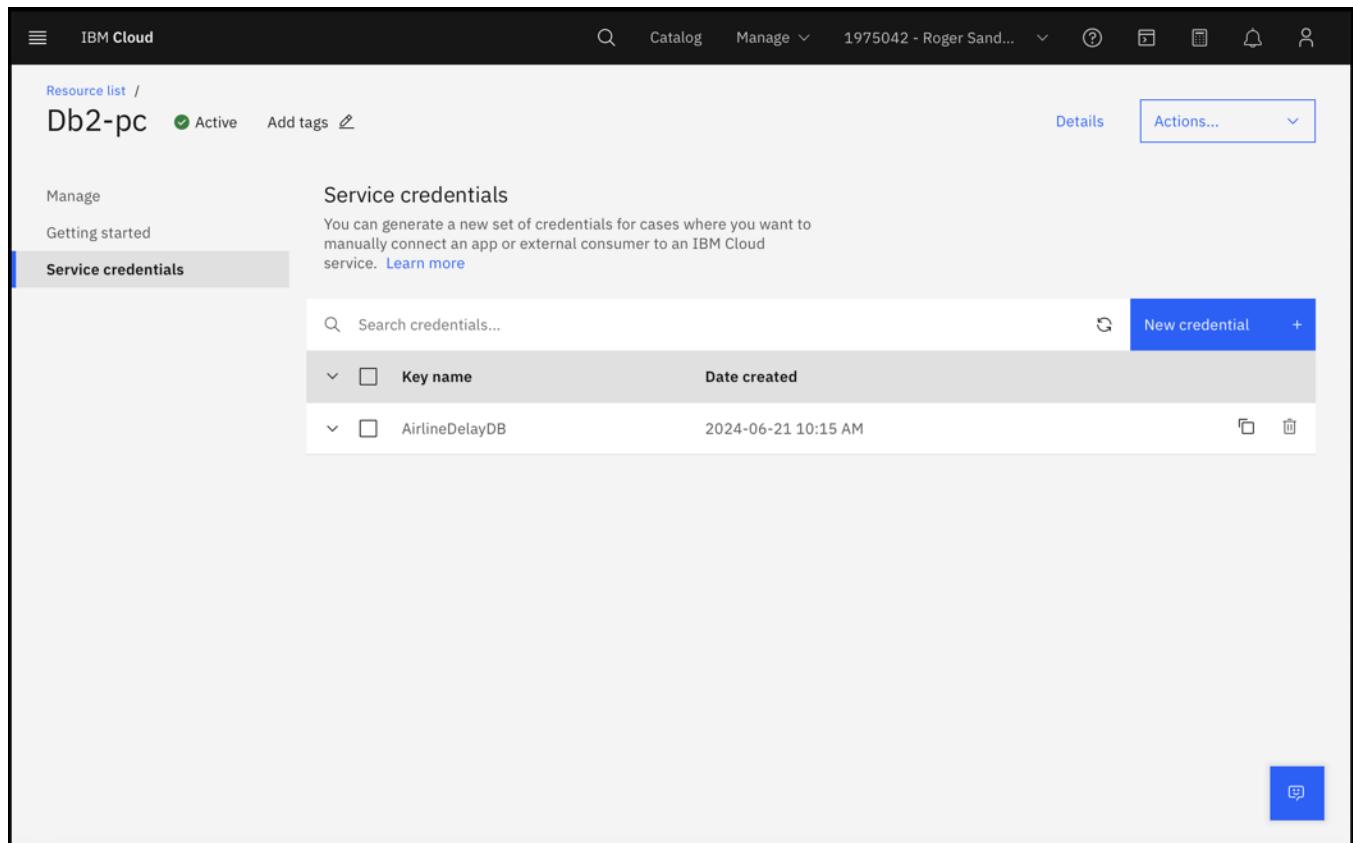
1. To work with an IBM Cloud resource from another application (for example, a Jupyter Notebook), one or more service credentials for that resource must exist. As this screen illustrates, when a Db2 on Cloud database is first provisioned, it has **No service credentials**. To create a service credential for the database, [A] click on the **Service credentials** item in the menu shown in the top left corner of the screen. [B] Then, click the **New credential** button located on the right. This will cause the *Create credential* popup window to be displayed.



2. When the *Create credential* popup window is displayed, [A] enter a name for the new service credential in the **Name:** field. (The default name will be *Service credentials-1*, but for this exercise, we're going to change the name to *AirlineDelayDB*). [B] Then, click the **Add** button located at the bottom right corner of the window. (Keep the **Role:** of *Manager*.)



_____ 3. When the *Create credential* popup window closes, the new service credential will be shown in the list of service credentials available.



The screenshot shows the IBM Cloud Service Credentials page for a resource named 'Db2-pc'. The 'Service credentials' tab is selected. A table lists one credential entry:

Key name	Date created	Actions
AirlineDelayDB	2024-06-21 10:15 AM	<input type="button" value="Edit"/> <input type="button" value="Delete"/>

At the top right of the table, there is a blue button labeled 'New credential' with a '+' sign.

_____ 4. Click the “**Copy to clipboard**” icon shown on the right-hand side of the listing for the service credential just created. Then, paste the information copied into a text file – for this workshop, use a file named **AirlineDelayDB_Credentials.txt** – and save it for later.

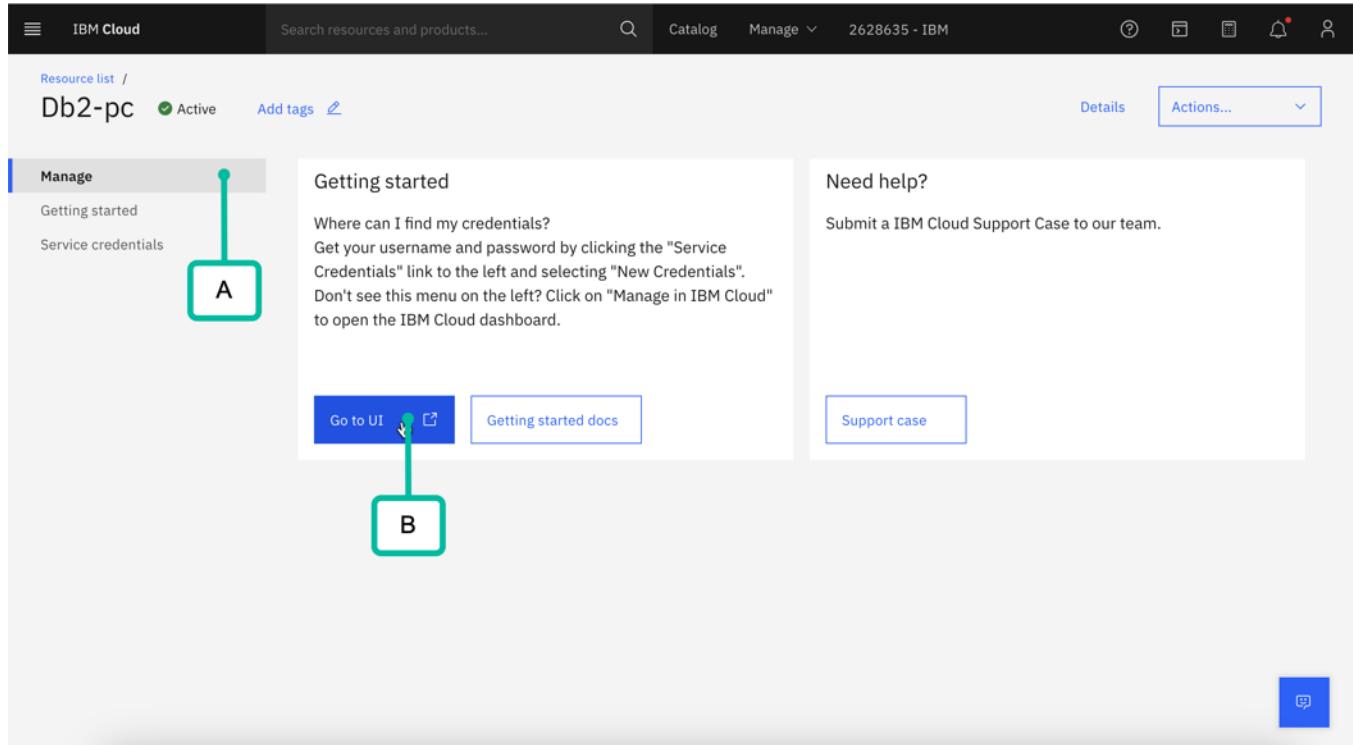
The screenshot shows the IBM Cloud Service Credentials page for a resource named "Db2-pc". The "Service credentials" tab is selected. A table lists a single credential entry:

Key name	Date created
AirlineDelayDB	2023-05-18 2:59 PM

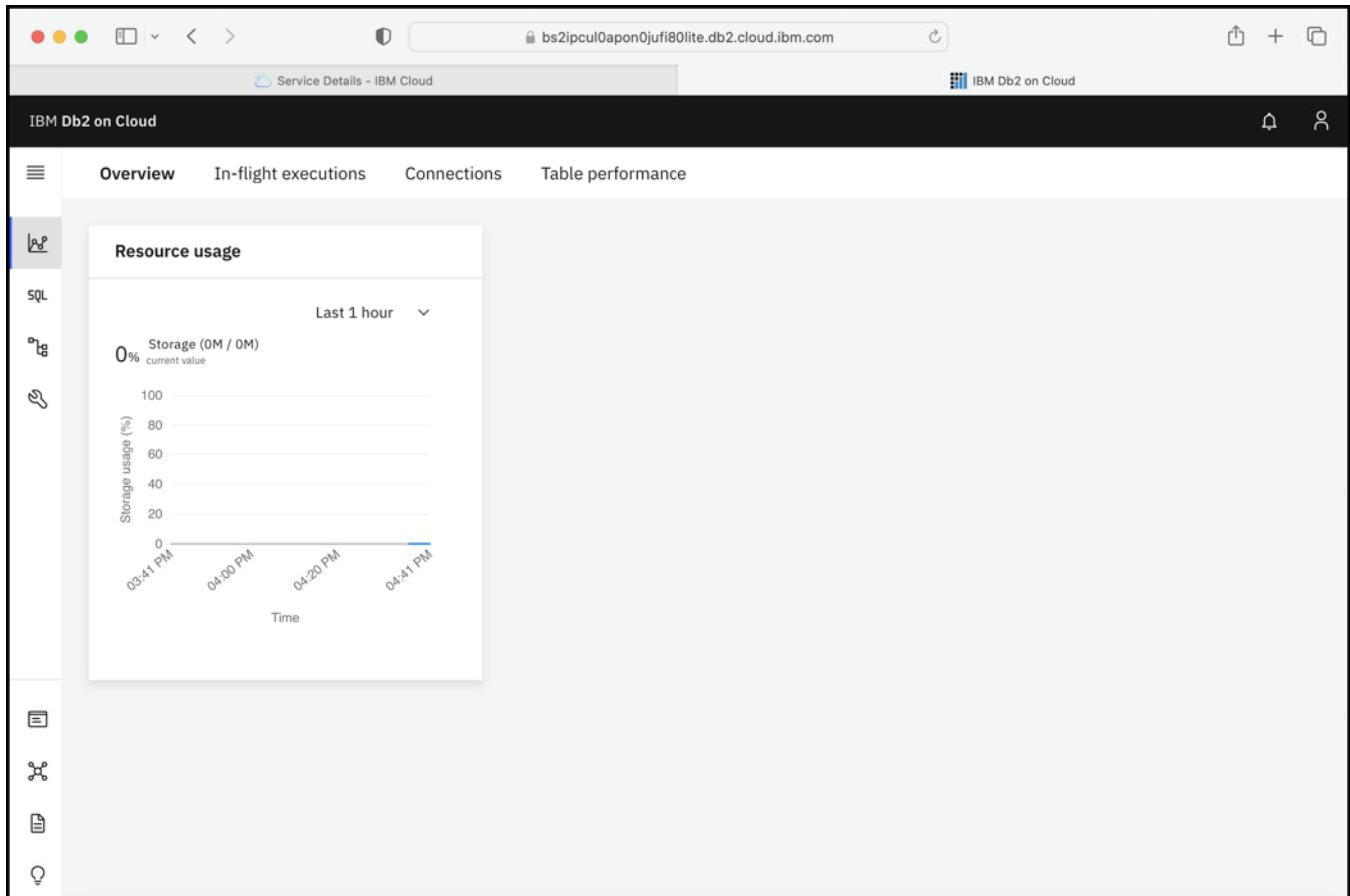
To the right of the table, there are three icons: a copy icon (highlighted with a blue box), a delete icon, and a refresh icon. A small "Copied!" message is displayed near the copy icon. The top navigation bar includes links for Catalog, Manage, and Details, along with various account and notification icons.

5. Query the system catalog using the Db2 on Cloud User Interface (UI).

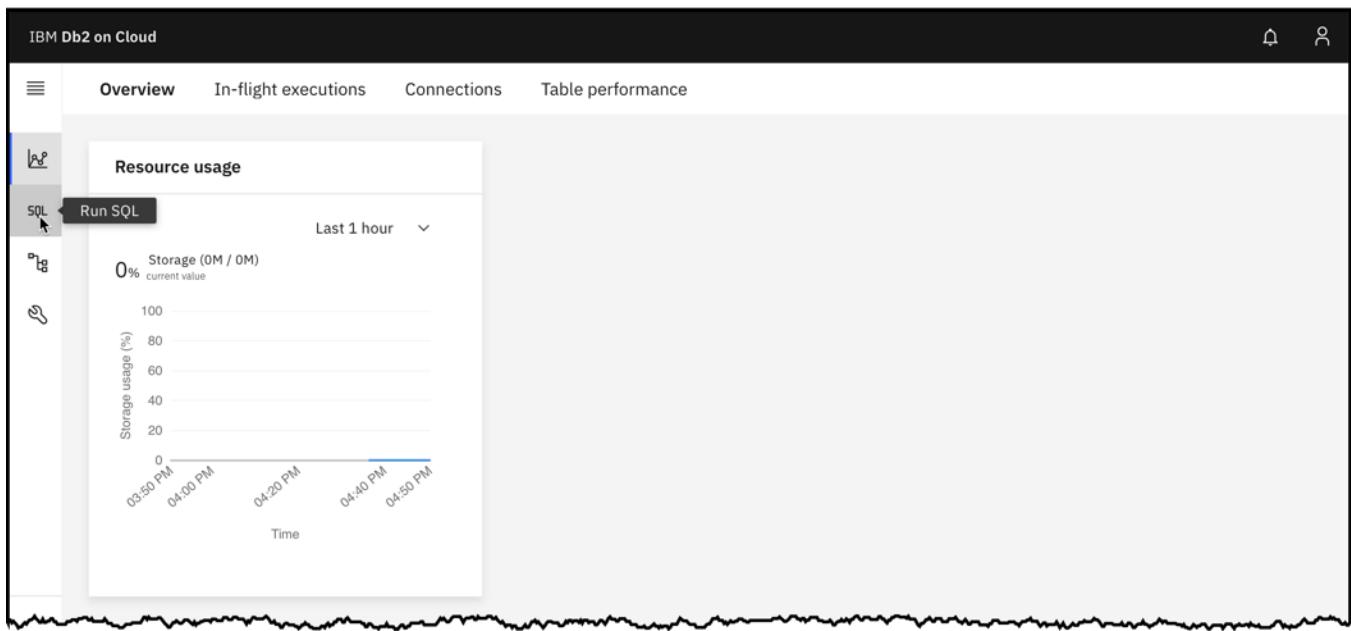
1. To work with the Db2 on Cloud database using the Db2 on Cloud User Interface, [A] click on the **Manage** item in the menu shown in the top left corner of the screen. [B] Then, click the **Go to UI** button.



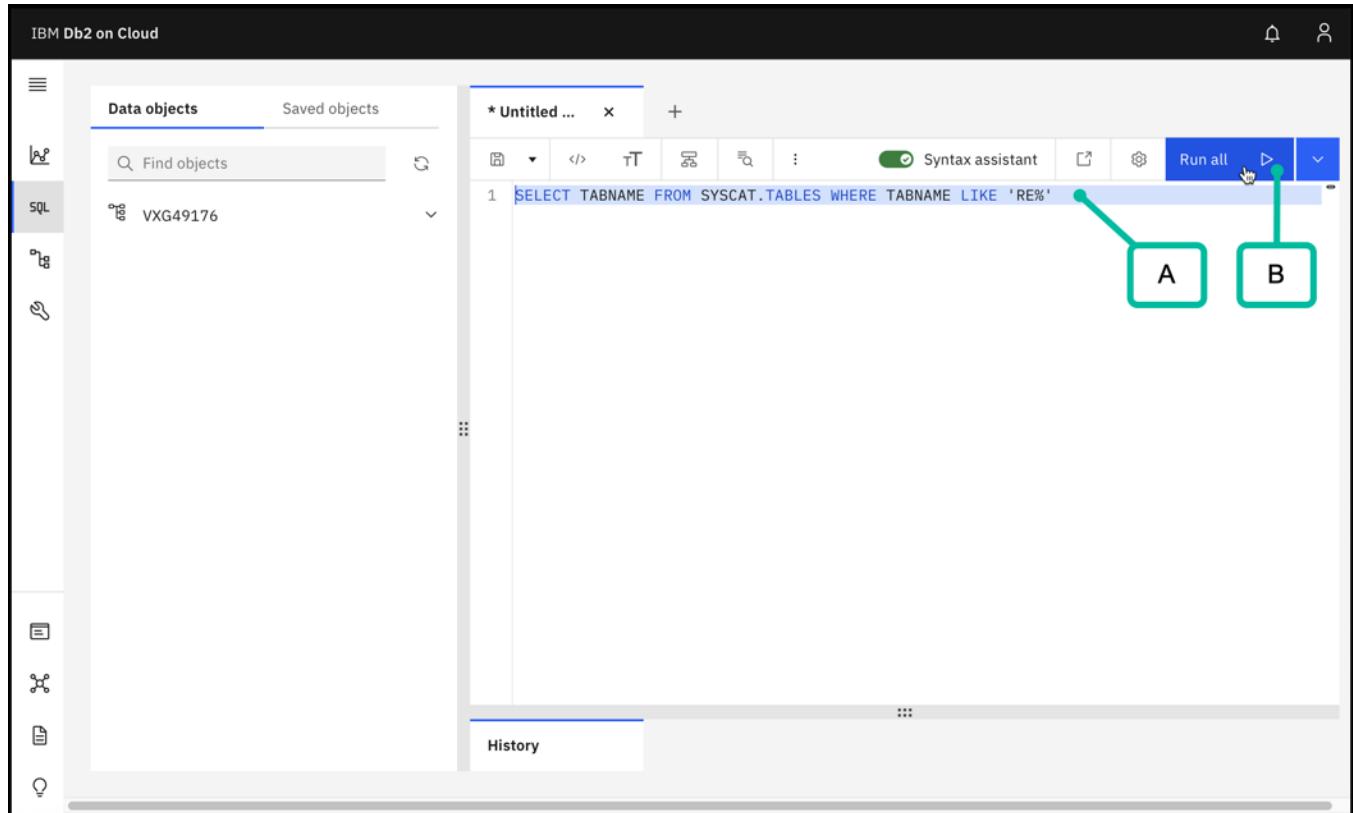
2. When the **Go to UI** button is selected, a new web browser tab named **IBM Db2 on Cloud** is opened. This tab normally consists of an **Overview** screen, which is the default starting point for the Db2 on Cloud User Interface.



3. Switch to the *Run SQL* screen by clicking the **SQL** icon found in the menu shown on the left-hand side of the screen.

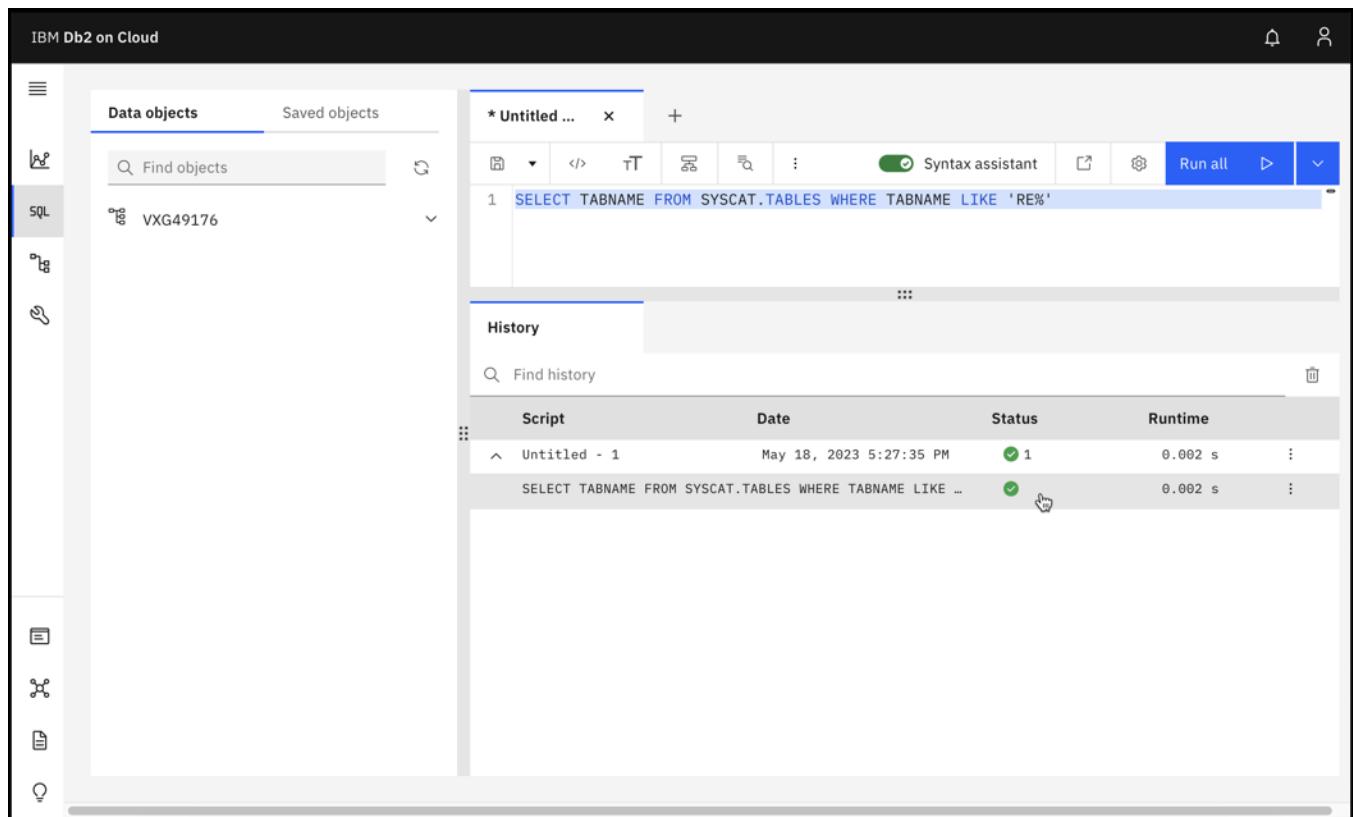


4. [A] Enter the SQL statement **SELECT TABNAME FROM SYSCAT.TABLES WHERE TABNAME LIKE 'RE%'** in the *Untitled - 1 portion of the *Run SQL* screen. [B] Then, click the **Run all** button to execute the statement. (When executed, this statement will query the database's system catalog and return a list of tables whose name begins with the letters "RE".)



The screenshot shows the IBM Db2 on Cloud interface. On the left is a sidebar with icons for Data objects, Saved objects, Find objects, SQL, History, and other options. The main area is a SQL editor titled '*Untitled ...'. The SQL statement `SELECT TABNAME FROM SYSCAT.TABLES WHERE TABNAME LIKE 'RE%'` is entered in the editor. The editor has a toolbar with various icons and a status bar at the bottom. Two green callout boxes are overlaid on the screen: box 'A' points to the SQL statement, and box 'B' points to the 'Run all' button in the toolbar.

5. When the SQL statement entered in the previous step is executed, a record about its execution will be recorded in a history log and a summary of the record will be displayed in the **History** window/tab located at the bottom of the screen. If the SQL statement executed successfully, the **Status** column will contain a green circle with a check mark in it; if the statement couldn't be executed, a red circle with a slash in it will appear instead. Find the record for the statement just executed and click on it. (You can find this record by looking for the text of the SQL statement in the **Script** column of the table displayed in the **History** window.)



The screenshot shows the IBM Db2 on Cloud interface. On the left, there's a sidebar with icons for Data objects, Saved objects, Find objects, SQL, and other database management functions. The main area has a central editor window titled 'Untitled ...' containing the SQL query: 'SELECT TABNAME FROM SYSCAT.TABLES WHERE TABNAME LIKE 'RE%''. Below the editor is a 'History' tab. A table lists the history of executed scripts:

Script	Date	Status	Runtime
Untitled - 1 SELECT TABNAME FROM SYSCAT.TABLES WHERE TABNAME LIKE ...	May 18, 2023 5:27:35 PM	✓ 1	0.002 s

6. When you click on the history record for the SQL statement just executed, you will be presented with a new tab named **Results** that contains a table showing the contents of result set produced when the statement was executed, together with the number of records (rows) returned in the result set. (In this example, the text “**Total:4**” indicates that four records were returned.)

The screenshot shows the IBM Db2 on Cloud interface. On the left, there's a sidebar with icons for Data objects, Saved objects, Find objects, SQL (which is selected), History, Details, and other options. The main area has a toolbar with various icons. A SQL editor window titled “* Untitled ...” contains the following code:

```
1 SELECT TABNAME FROM SYSCAT.TABLES WHERE TABNAME LIKE 'RE%'
```

Below the editor, there are tabs for History and Results. The Results tab is active, showing a table with one column labeled “TABNAME”. The data in the table is:

TABNAME
REFERENCES
REFERENTIAL_CONSTRAINTS
REF_CONSTRAINTS
REG_VARIABLES

In the bottom right corner of the results pane, it says “Total:4”.

____ 7. Clear the results and the history log by [A] clicking on the **History** tab (to return to the history log display), and then [B] clicking on the trash can icon located at the top left corner of the history log table.

The screenshot shows the IBM Db2 on Cloud interface. On the left, there's a sidebar with icons for Data objects, Saved objects, Find objects, and a dropdown for the current database (VXG49176). The main area has tabs for History and Results, with History currently selected. A green box labeled 'A' highlights the History tab. Below it is a table of history logs:

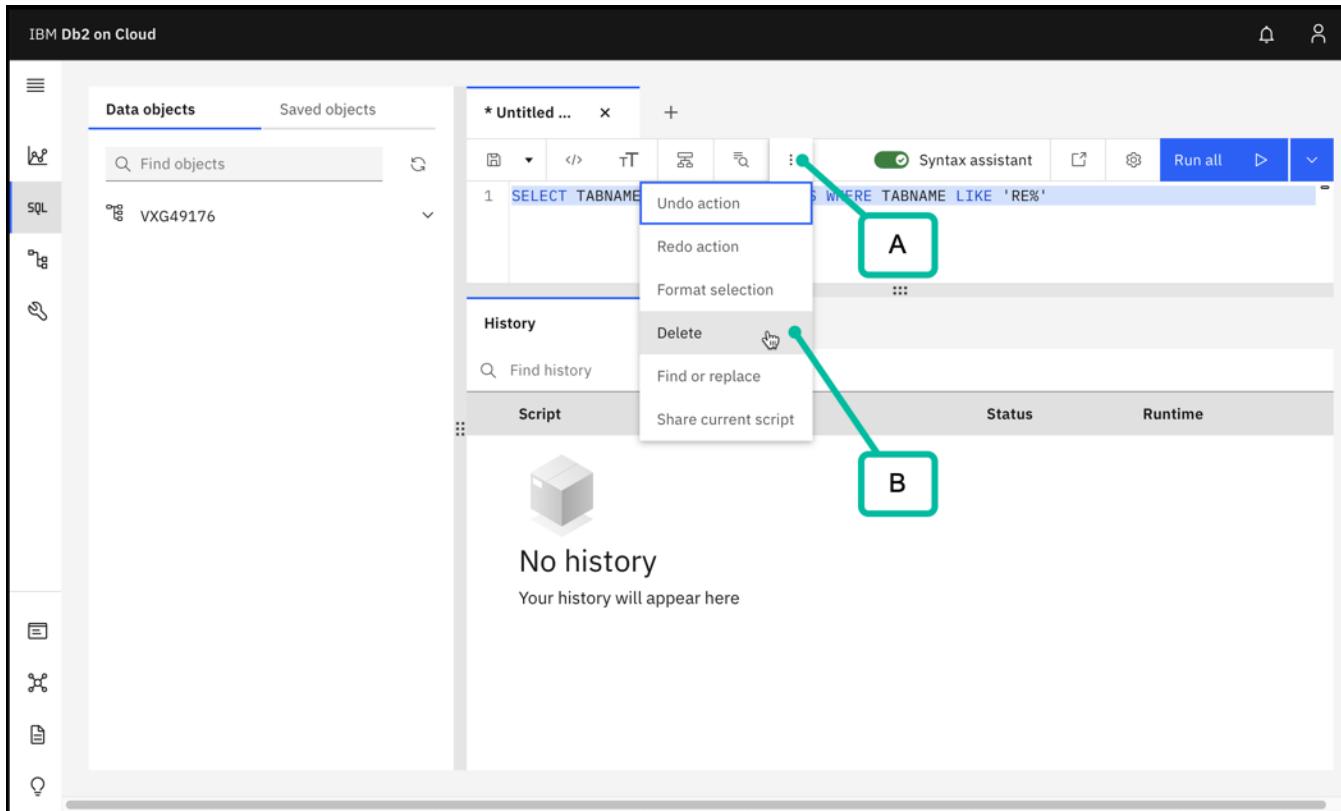
Script	Date	Status	Runtime
Untitled - 1	May 18, 2023 6:22:02 PM	✓ 1	0.006 s
SELECT TABNAME FROM SYSCAT.TABLES WHERE TABNAME LIKE ...		✓	0.006 s

A green box labeled 'B' highlights the trash can icon in the top right corner of the history log table. The status bar at the bottom shows 'Run all' and a progress bar.

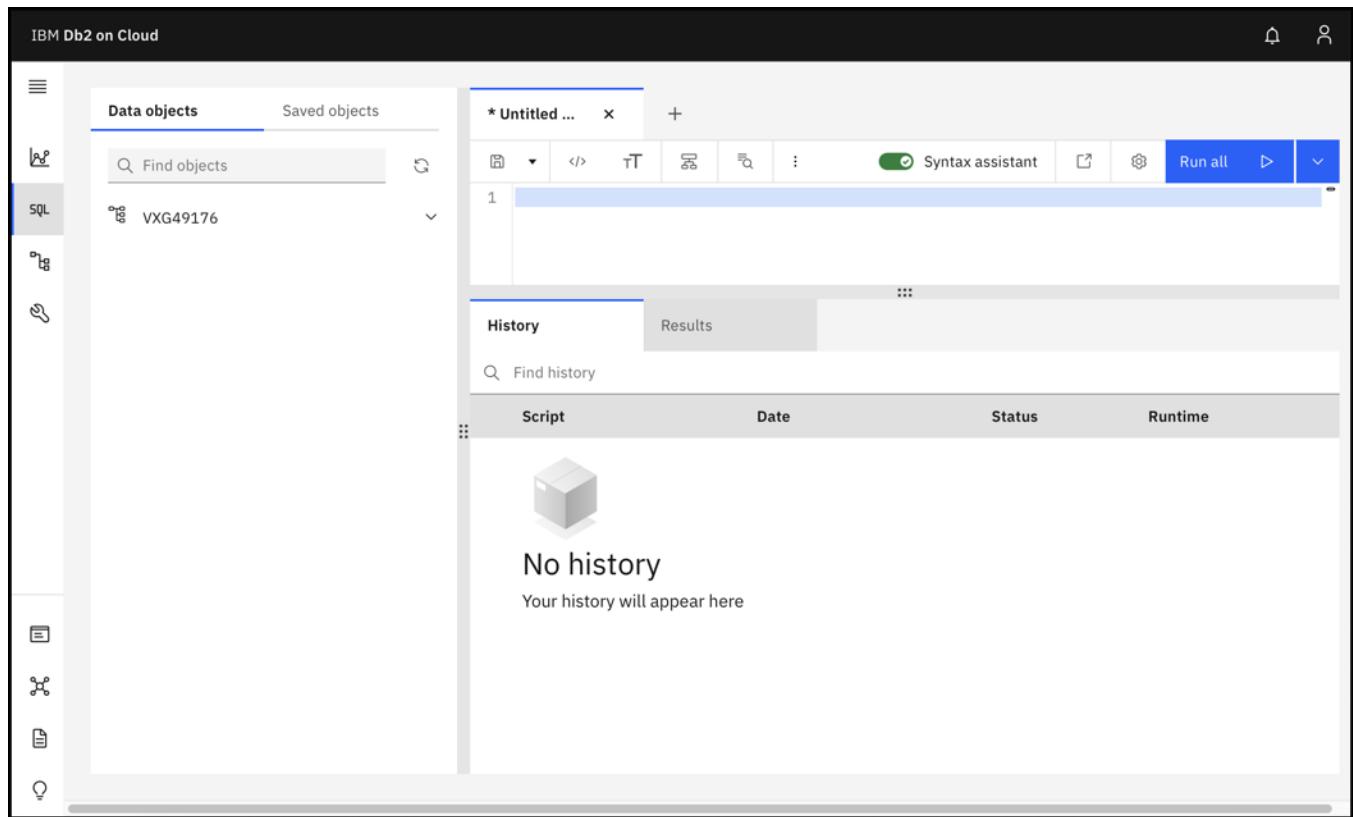
____ 8. When asked to confirm your decision to delete all results, click the **Yes** button.



9. Finally, clear the **Untitled-1** section of the *Run SQL* screen by [A] clicking on the **More** icon (the icon consisting of three vertical dots), [B] and then clicking on the **Delete** item shown in the drop-down menu presented.



10. When finished, the *Run SQL* screen should look similar to the way it did when the Db2 on Cloud User Interface was started.



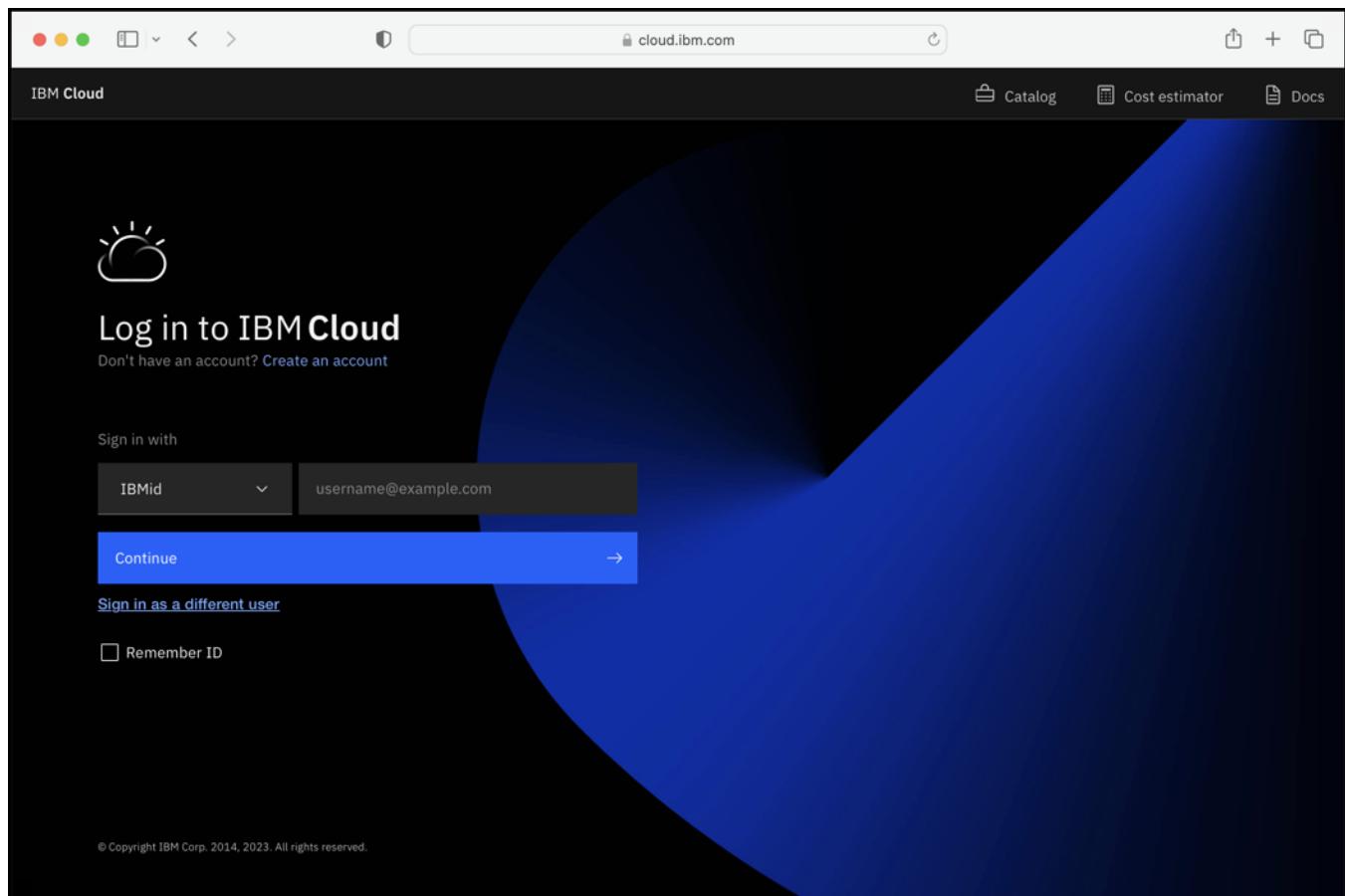
Congratulations! You have just created a Db2 on Cloud database and you have seen how to interact with that database using the Db2 on Cloud User Interface. We will use this database with other exercises in this workshop.

III. Building and populating a Db2 on Cloud database table

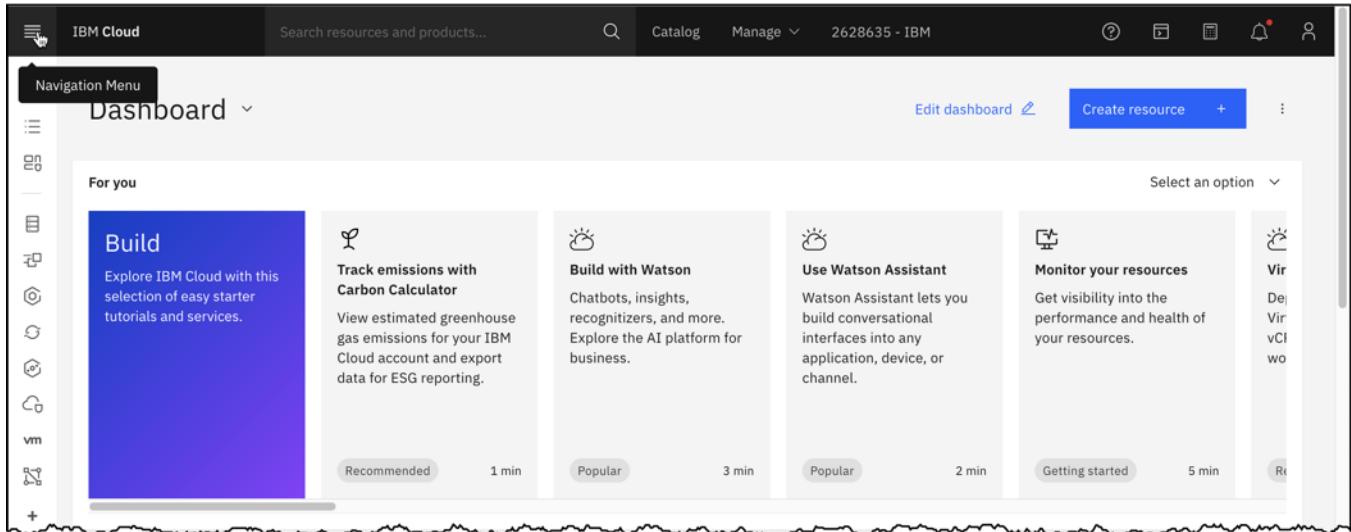
When an IBM Db2 on Cloud database is first created, it only contains the system catalog (which we looked at earlier.) For the database to be useful, it must contain at least one user-defined table. So, in this section, we will create a user-defined table and populate it with data stored in a Comma Separated Values (CSV)-formatted file.

1. Create a table named AIRLINE_DELAY_CAUSE.

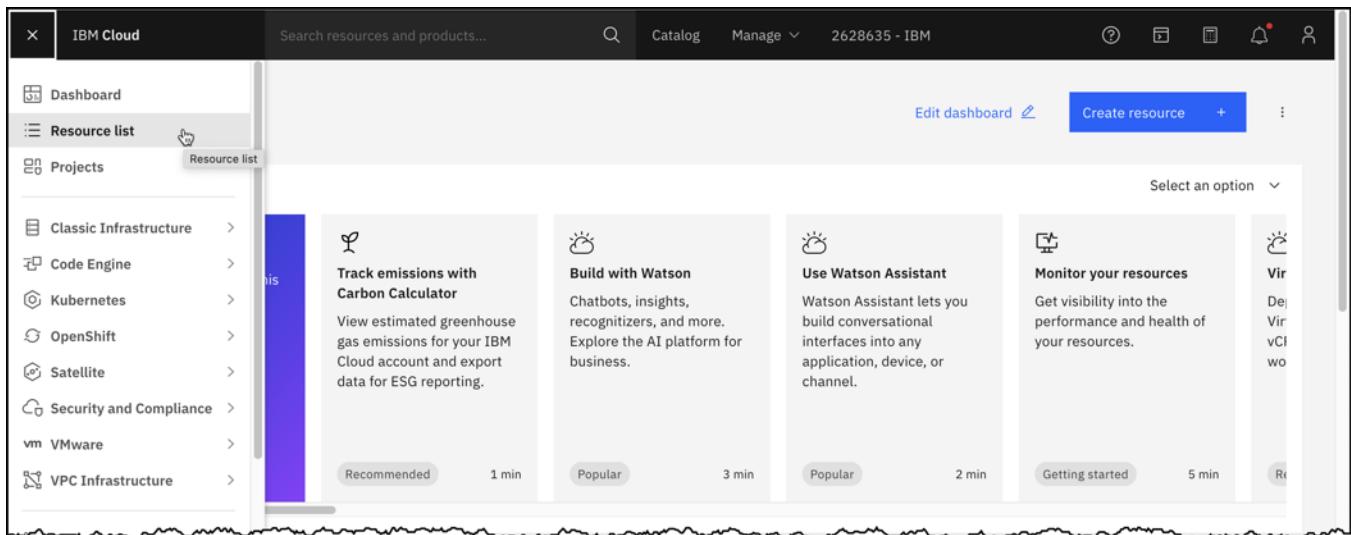
1. Open a web browser and go to the *Log in to the IBM Cloud* web site (<https://cloud.ibm.com/login>). You should be presented with a screen that looks something like this:



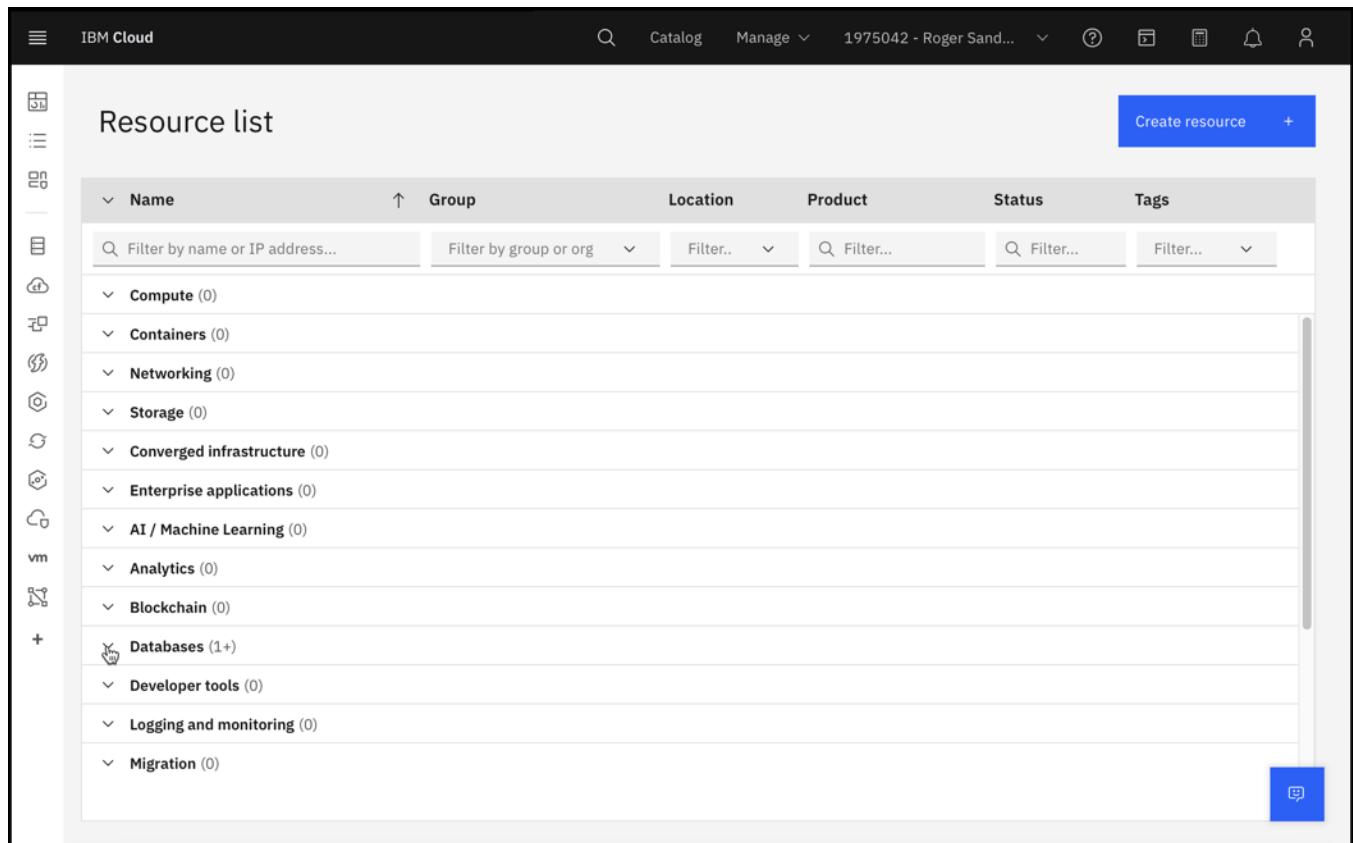
____ 2. Log in to IBM Cloud using your IBMid and password. If prompted, log in to IBM as well. Once you are logged in and the *IBM Cloud dashboard* appears, click the hamburger icon (also referred to as the triple bar or trigram symbol) located in the top left corner of the screen. This will open the *Navigation Menu*.



____ 3. Locate the **Resource list** item in the *Navigation Menu* and click on it. This will take you to the *Resource list* screen, where you can view all the resources that have been created in your IBM Cloud account.



4. Locate the **Databases** item in the resource list shown and click on the **V** shown to the left of it. You should see the ID for the Db2 on Cloud database that was created earlier.



The screenshot shows the IBM Cloud Resource list interface. On the left, there's a sidebar with icons for Compute, Containers, Networking, Storage, Converged infrastructure, Enterprise applications, AI / Machine Learning, Analytics, Blockchain, and a plus sign for more. Below this is a list of categories: Databases (1+), Developer tools (0), Logging and monitoring (0), and Migration (0). The 'Databases' category is expanded, showing a single entry: 'Db2-rc'. At the top right, there's a 'Create resource' button and a '+' icon. The main area has search and filter fields for Name, Group, Location, Product, Status, and Tags.

5. Access the Db2 on Cloud database created earlier by clicking on the database ID shown in the **Resource list**. (In this example, the resource is the Db2 on Cloud database that has the ID *Db2-pc*).

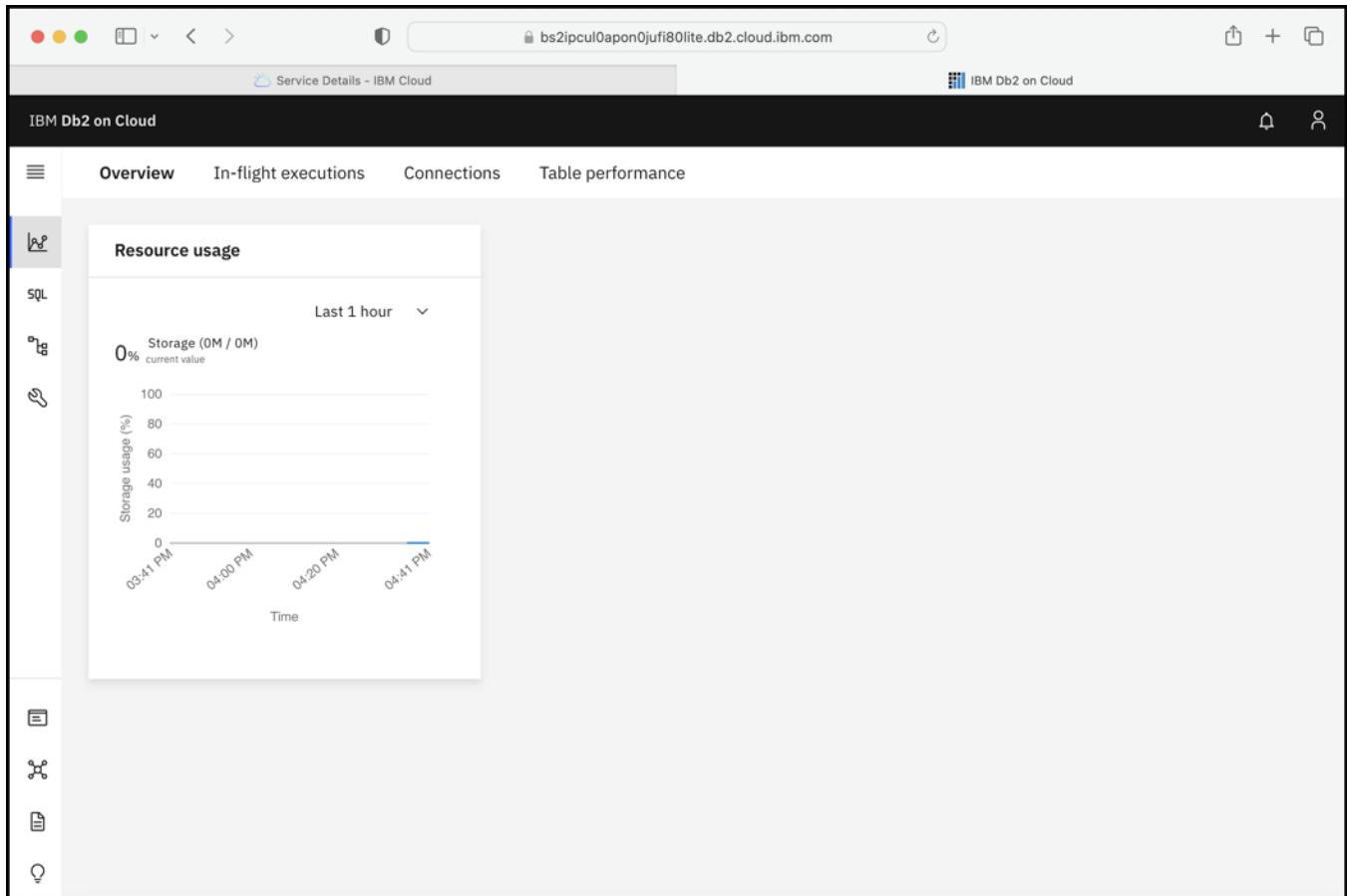


This screenshot shows a detailed view of the 'Db2-rc' database resource. It includes tabs for Default, London, and Db2, with the Db2 tab selected. The status is marked as Active. A blue box highlights the 'Db2-rc' ID. The sidebar on the left is identical to the one in the previous screenshot, showing the expanded 'Databases' section with 'Db2-rc' listed.

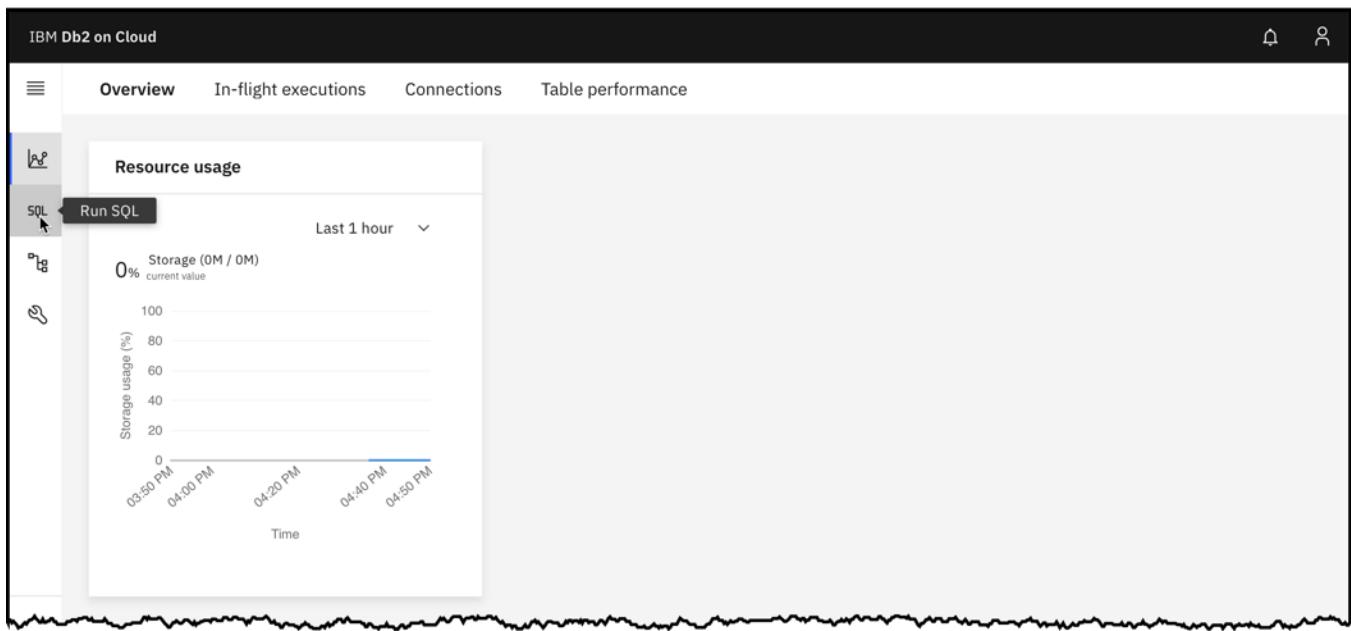
6. As we saw earlier, clicking on a resource ID in the **Resource list** will cause the *Resource list* screen to be replaced with a high-level “management” screen for the resource selected. When the “management” screen for the Db2 on Cloud database resource selected appears, click the **Go to UI** button.

The screenshot shows the IBM Cloud Resource Management interface. At the top, there's a navigation bar with 'IBM Cloud', a search bar, and various management icons. Below the header, the URL 'Resource list /' and the resource name 'Db2-pc' are displayed, along with status 'Active' and a 'Add tags' button. On the right, there are 'Details' and 'Actions...' buttons. The main content area is divided into sections: 'Manage' (selected), 'Getting started', and 'Need help?'. The 'Getting started' section contains instructions for finding credentials and navigating the service. It features a prominent blue 'Go to UI' button with a cursor icon hovering over it, and other links like 'Getting started docs' and 'Support case'. The 'Need help?' section has a link to submit a support case. A small blue feedback icon is located in the bottom right corner of the main content area.

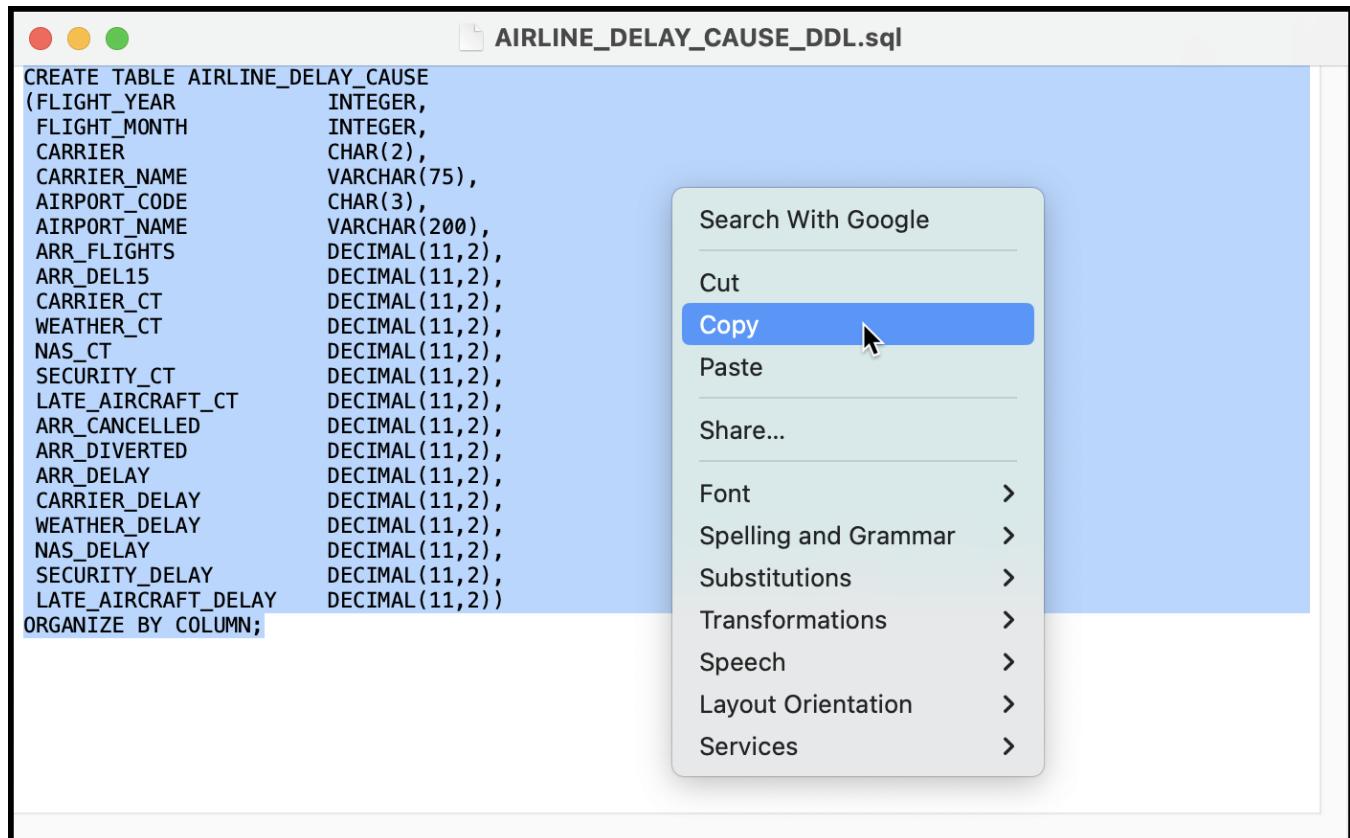
7. Clicking on the **Go to UI** button will cause a new web browser tab named **IBM Db2 on Cloud** to be opened. This tab usually contains an **Overview** screen, which is the default starting point for the Db2 on Cloud User Interface.



8. Switch to the *Run SQL* screen by clicking the **SQL** icon found in the menu shown on the left-hand side of the screen.



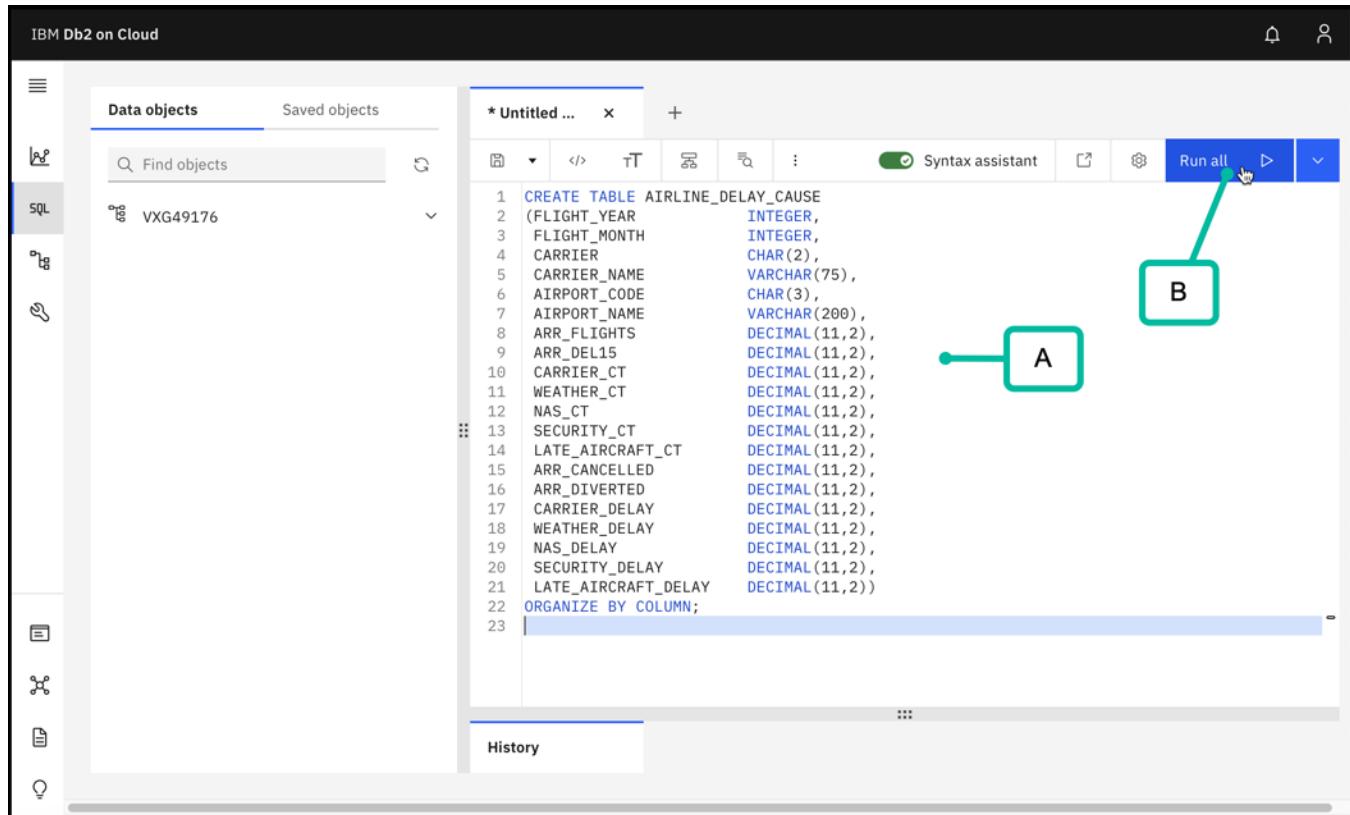
____ 9. Minimize your web browser and open the file named **AIRLINE_DELAY_CAUSE_DDL.sql** with a text editor. Then, select everything found in the file and copy it to the clipboard. (In this example, the file was opened with the **TextEdit** application on a MacBook Pro. After everything in the file was selected, a menu was opened by clicking the right button on the mouse. Then, the menu item **Copy** was selected.)



The screenshot shows a Mac OS X context menu for a file named "AIRLINE_DELAY_CAUSE_DDL.sql". The menu includes options like "Search With Google", "Cut", "Copy" (which is highlighted with a blue background), "Paste", "Share...", "Font", "Spelling and Grammar", "Substitutions", "Transformations", "Speech", "Layout Orientation", and "Services". The "Copy" option is the second item in the list.

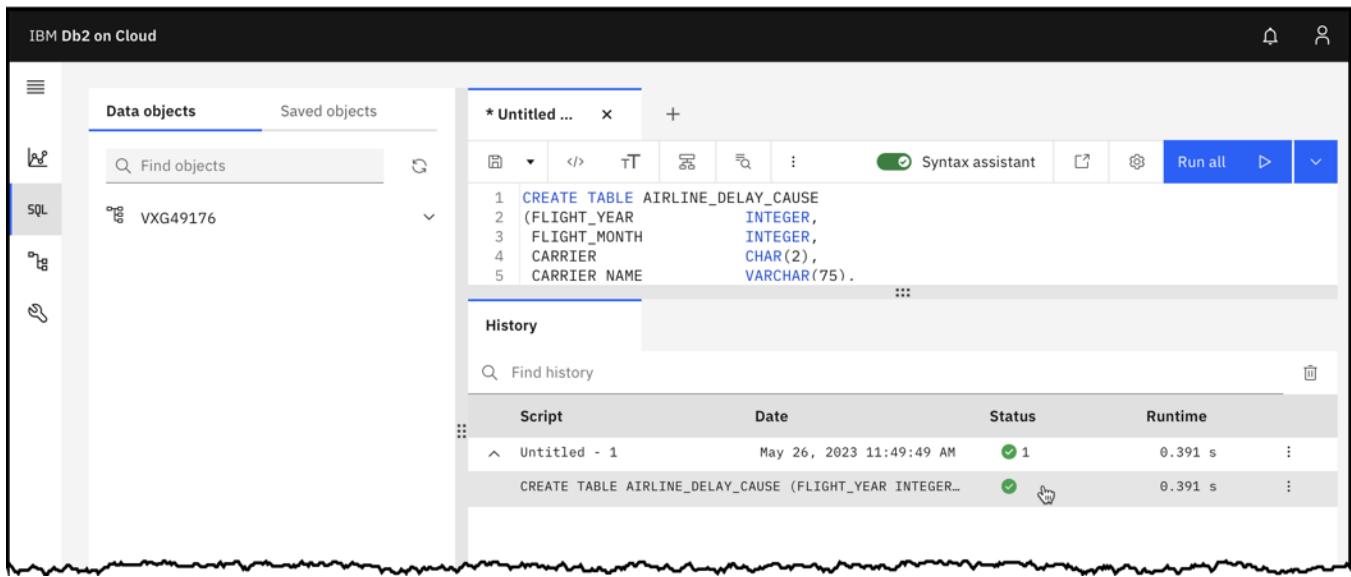
```
CREATE TABLE AIRLINE_DELAY_CAUSE
(FLIGHT_YEAR          INTEGER,
 FLIGHT_MONTH         INTEGER,
 CARRIER              CHAR(2),
 CARRIER_NAME         VARCHAR(75),
 AIRPORT_CODE         CHAR(3),
 AIRPORT_NAME         VARCHAR(200),
 ARR_FLIGHTS          DECIMAL(11,2),
 ARR_DEL15             DECIMAL(11,2),
 CARRIER_CT           DECIMAL(11,2),
 WEATHER_CT           DECIMAL(11,2),
 NAS_CT               DECIMAL(11,2),
 SECURITY_CT           DECIMAL(11,2),
 LATE_AIRCRAFT_CT     DECIMAL(11,2),
 ARR_CANCELLED        DECIMAL(11,2),
 ARR_DIVERTED         DECIMAL(11,2),
 ARR_DELAY             DECIMAL(11,2),
 CARRIER_DELAY         DECIMAL(11,2),
 WEATHER_DELAY         DECIMAL(11,2),
 NAS_DELAY             DECIMAL(11,2),
 SECURITY_DELAY        DECIMAL(11,2),
 LATE_AIRCRAFT_DELAY   DECIMAL(11,2))
ORGANIZE BY COLUMN;
```

10. Restore your web browser and [A] use the “hot keys” **command-V** on Mac or **Ctrl-V** on Windows to paste the data copied to the clipboard into the *Untitled - 1 window of the *Run SQL* screen. Then, [B] click the **Run all** button to execute the statement just pasted into the window.



```
1 CREATE TABLE AIRLINE_DELAY_CAUSE
2 (FLIGHT_YEAR
3 INTEGER,
4 FLIGHT_MONTH
5 INTEGER,
6 CARRIER
7 CHAR(2),
8 CARRIER_NAME
9 VARCHAR(75),
10 AIRPORT_CODE
11 CHAR(3),
12 AIRPORT_NAME
13 VARCHAR(200),
14 ARR_DEL15
15 DECIMAL(11,2),
16 CARRIER_CT
17 DECIMAL(11,2),
18 WEATHER_CT
19 DECIMAL(11,2),
20 NAS_CT
21 SECURITY_CT
22 DECIMAL(11,2),
23 LATE_AIRCRAFT_CT
24 DECIMAL(11,2),
25 ARR_CANCELLED
26 DECIMAL(11,2),
27 ARR_DIVERTED
28 DECIMAL(11,2),
29 CARRIER_DELAY
30 DECIMAL(11,2),
31 WEATHER_DELAY
32 DECIMAL(11,2),
33 NAS_DELAY
34 DECIMAL(11,2),
35 SECURITY_DELAY
36 DECIMAL(11,2),
37 LATE_AIRCRAFT_DELAY
38 DECIMAL(11,2))
39 ORGANIZE BY COLUMN;
```

____ 11. Check the **History** window/tab located at the bottom of the screen. If the **Status** column for the record containing the text “**CREATE TABLE AIRLINE_DELAY_CAUSE ...**” has a green circle with a check mark in it, a table named **AIRLINE_DELAY_CAUSE** was successfully created. If an error occurred, clear the contents of the ***Untitled - 1** window and redo steps 9 and 10 (make sure you copy every line in the file **AIRLINE_DELAY_CAUSE_DDL.sql.**)



The screenshot shows the IBM Db2 on Cloud interface. On the left, there's a sidebar with icons for Data objects, Saved objects, Find objects, SQL, Data, and a magnifying glass. The SQL icon is currently selected. The main area has a title bar * Untitled ... and contains a SQL script window with the following code:

```

1 CREATE TABLE AIRLINE_DELAY_CAUSE
2 (FLIGHT_YEAR      INTEGER,
3  FLIGHT_MONTH     INTEGER,
4  CARRIER          CHAR(2),
5  CARRIER_NAME     VARCHAR(75).

```

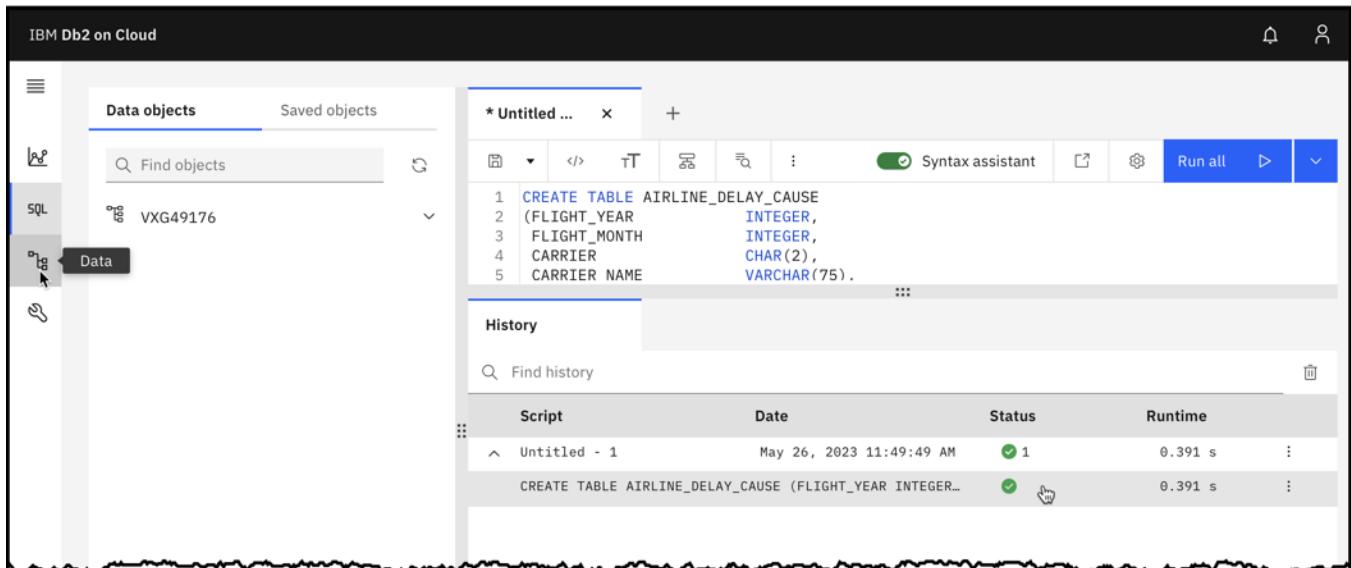
Below the script is a History section with a table:

Script	Date	Status	Runtime
Untitled - 1	May 26, 2023 11:49:49 AM	✓ 1	0.391 s
CREATE TABLE AIRLINE_DELAY_CAUSE (FLIGHT_YEAR INTEGER...		✓	0.391 s

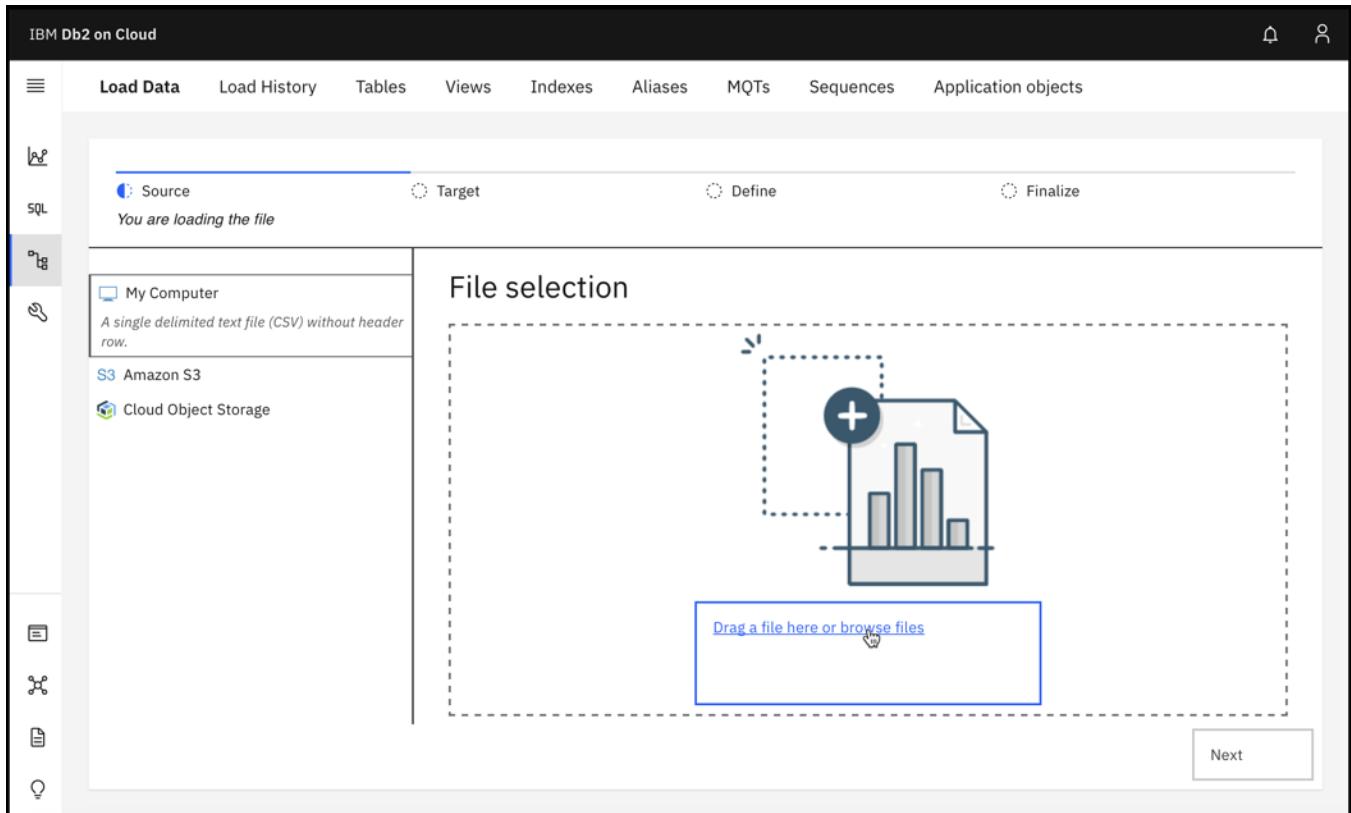
The status column for the last row shows a green checkmark and the number 1, indicating success.

2. Populate the AIRLINE_DELAY_CAUSE table.

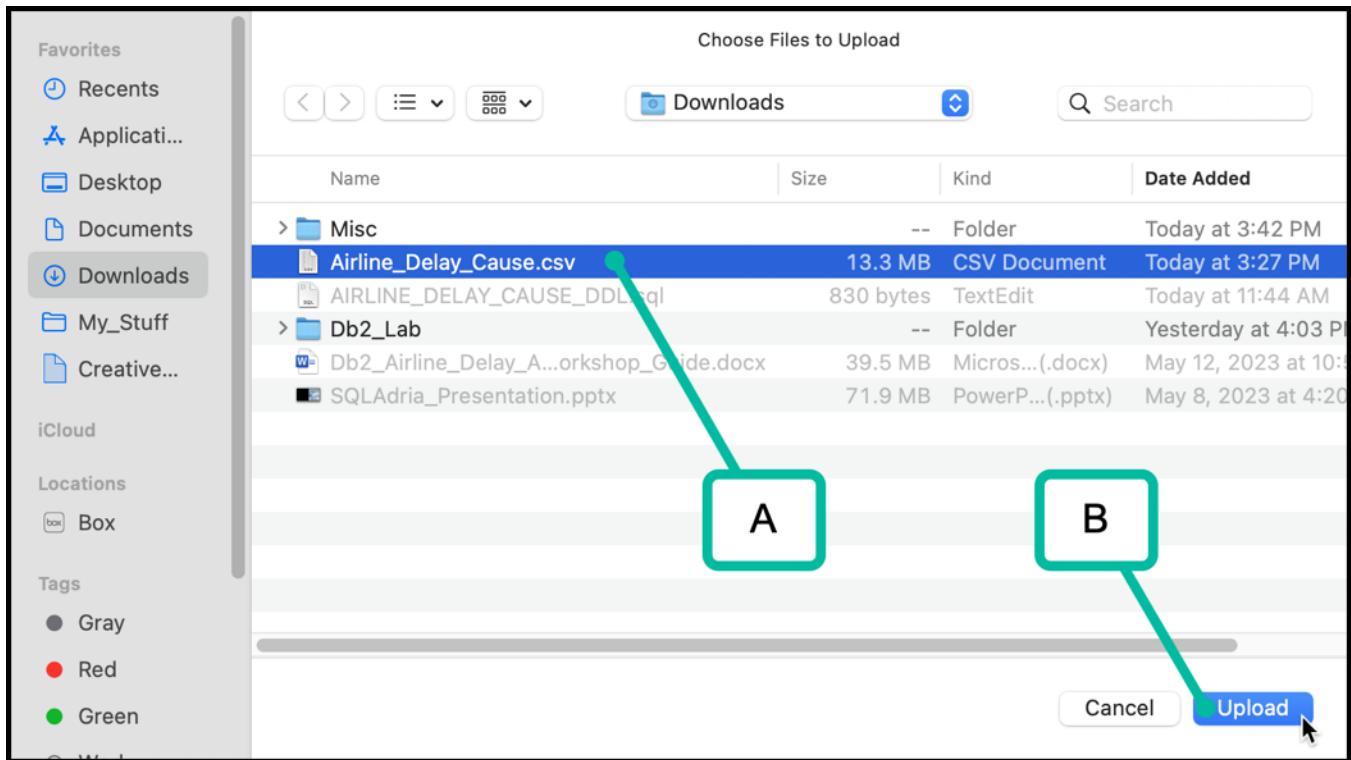
____ 1. Switch to the *Data* screen by clicking the **Data** icon (located immediately below the **SQL** icon) found in the menu shown on the left-hand side of the screen.



____ 2. When you go to the *Data* screen, the first thing you see is the *Load Data* window. From this screen, click on the **Drag a file here or browse files** link to open a *Finder* window on Mac or a *File Explorer* window on Windows.



_____ 3. [A] Select the file named **Airline_Delay_Cause.csv** and [B] click the **Upload** button.



4. Verify that the file name **Airline_Delay_Cause.csv** appears under the **Selected file** heading on the *Load Data* screen. Then, click the **Next** button located in the bottom right-hand corner of the screen.

The screenshot shows the 'Load Data' screen in the IBM Db2 on Cloud interface. The top navigation bar includes 'Load Data', 'Load History', 'Tables', 'Views', 'Indexes', 'Aliases', 'MQTs', 'Sequences', and 'Application objects'. On the left, there's a sidebar with icons for 'SQL', 'Tables', 'Views', 'Indexes', 'Aliases', 'MQTs', 'Sequences', and 'Application objects'. The main area has tabs for 'Source' (selected), 'Target', 'Define', and 'Finalize'. Below the tabs, it says 'You are loading the file **Airline_Delay_Cause.csv**'. A 'File selection' section contains a 'My Computer' folder icon with the text 'A single delimited text file (CSV) without header row.' and an 'Amazon S3' icon with 'Cloud Object Storage' text. To the right, a 'Selected file' section shows the file 'Airline_Delay_Cause.csv' with a delete icon. At the bottom right is a blue 'Next' button with a hand cursor icon.

____ 5. From the *Select a load target* screen, [A] click on the *Schema* name associated with your IBMid. (In this example, the schema name is **VXG49176**.) Then, [B] click on the **AIRLINE_DELAY_CAUSE** table shown in the *Table* list. Next, [C] select the **Overwrite table with new data** radio button. Finally, [D] click the **Next** button shown in the bottom right-hand corner of the screen.

The screenshot shows the 'Load Data' interface in IBM Db2 on Cloud. The top navigation bar includes 'Load Data', 'Load History', 'Tables', 'Views', 'Indexes', 'Aliases', 'MQTs', 'Sequences', and 'Application objects'. On the left, there's a sidebar with icons for 'SQL', 'Tables', 'Views', 'Indexes', 'Sequences', and 'Application objects'. The main area is titled 'Select a load target'. It has three main sections: 'Schema' (containing 'VXG49176'), 'Table' (containing 'AIRLINE_DELAY_CAUSE'), and 'Table definition'. In the 'Table definition' section, the 'Overwrite table with new data' radio button is selected. At the bottom, there are 'Back' and 'Next' buttons, with 'Next' being highlighted.

6. Verify that the **Airline_Delay_Cause.csv** file used contains valid data. Then, [A] make sure the **Header in first row** switch is on (green) and [B] click the **Next** button.

The screenshot shows the IBM Db2 on Cloud interface for loading data. The top navigation bar includes 'Load Data', 'Load History', 'Tables', 'Views', 'Indexes', 'Aliases', 'MQTs', 'Sequences', and 'Application objects'. On the left, there's a sidebar with icons for 'SQL', 'Tables', 'Views', 'Indexes', 'Aliases', 'MQTs', 'Sequences', and 'Application objects'.

The main area shows a configuration for loading 'Airline_Delay_Cause.csv' into the table 'VXG49176.AIRLINE_DELAY_CAUSE'. The 'Source' tab is selected. The target table structure is defined as:

FLIGHT_YEAR INTEGER	FLIGHT_MONTH INTEGER	CARRIER CHARACTER	CARRIER_NAME VARCHAR
1 2022	12	9E	Endeavor Air Inc.
2 2022	12	9E	Endeavor Air Inc.
3 2022	12	9E	Endeavor Air Inc.
4 2022	12	9E	Endeavor Air Inc.
5 2022	12	9E	Endeavor Air Inc.
6 2022	12	9E	Endeavor Air Inc.
7 2022	12	9E	Endeavor Air Inc.
8 2022	12	9E	Endeavor Air Inc.
9 2022	12	9E	Endeavor Air Inc.
10 2022	12	9E	Endeavor Air Inc.

Below the table structure, the 'Header in first row' switch is turned on (green), and the 'Time & date format:' dropdown is set to '(none)'. A green callout labeled 'A' points to this switch. A green callout labeled 'B' points to the 'Next' button at the bottom right of the screen.

7. Look over the *Review settings* screen to ensure everything is correct. Then, start the data load process by clicking the **Begin Load** button.

The screenshot shows the 'Review settings' page of the IBM Db2 on Cloud interface. At the top, there are tabs for 'Load Data', 'Load History', 'Tables', 'Views', 'Indexes', 'Aliases', 'MQTs', 'Sequences', and 'Application objects'. Below the tabs, there are four radio buttons: 'Source' (selected), 'Target', 'Define', and 'Finalize'. A message states: 'You are loading the file **Airline_Delay_Cause.csv** into **VXG49176.AIRLINE_DELAY_CAUSE**'. On the left, there's a sidebar with icons for 'SQL', 'Tables', 'Views', 'Indexes', 'Aliases', 'MQTs', 'Sequences', and 'Application objects'. The main area is divided into two sections: 'Summary' and 'Option'. The 'Summary' section contains the following settings:

Code page:	1208 (Default)
Separator:	,
Time format:	HH:MM:SS (Default)
Date format:	YYYY-MM-DD (Default)
Timestamp format:	YYYY-MM-DD HH:MM:SS (Default)
String delimiter:	(Default)

The 'Option' section has a dropdown for 'Maximum number of warnings' set to 1000. At the bottom right, there are 'Back' and 'Begin Load' buttons, with 'Begin Load' being highlighted.

_____ 8. When the **Begin Load** button is selected, the *Review settings* screen will be replaced by the *Load details* screen, and you will receive a notification that there is a **Load in progress**. This notification will include information about where data is coming from (in this example, the **Airline_Delay_Cause.csv** file stored on **My Computer**) and where it is going (the **VXG49176.AIRLINE_DELAY_CAUSE** table).

The screenshot shows the IBM Db2 on Cloud interface with the 'Load Data' tab selected. In the top right corner, there is a notifications badge with a red '1'. The main area displays 'Load details' for a task named 'LOADING'. It shows 'My computer' as the source for 'Airline_Delay_Cause.csv' and 'Target' as 'VXG49176.AIRLINE_DELAY_CAUSE'. Below this, the 'Status' tab is active, showing a 'Loading' section with counts: 0 Rows read, 0 Rows loaded, and 0 Rows rejected. It also shows the start time as '05/26/2023 3:49:18 PM' and an elapsed time of 'a few seconds ago'. To the right, a 'Did you know?' box provides information about exporting SQL editor changes. A vertical timeline on the right indicates the process: 1. Upload (blue dot), 2. Load data (blue dot), and 3. Complete (blue dot). At the bottom right, it says 'Available after load is finished'. Navigation icons for tables, views, indexes, aliases, MQTs, sequences, and notifications are visible along the left and top edges.

9. Once the load operation is completed, the *Load details* screen will be updated to reflect the number of rows read from the source file, the number of rows loaded into the table specified, and the number of rows rejected. You will also receive a notification that **The data load job succeeded** (or failed).

The screenshot shows the 'Load details' page in the IBM Db2 on Cloud interface. On the left, there's a sidebar with icons for Load Data, Load History, Tables, Views, Indexes, Aliases, MQTs, Sequences, Notifications, and Clear all. The main area has tabs for Status (selected) and Settings. In the Status section, there's a large blue circular progress indicator. Below it, the status is shown as 'COMPLETE'. The 'My computer' section lists 'Airline_Delay_Cause.csv' as the source and 'VXG49176.AIRLINE_DELAY_CAUSE' as the target. The 'Rows' section shows '80,104' for both read and loaded rows, and '0' for rejected rows. Below this, the start time is listed as '05/26/2023 3:49:18 PM' and the end time as '05/26/2023 3:49:58 PM'. To the right, a 'Notifications' panel shows a green checkmark icon with the message 'The data load job succeeded. Load Airline_Delay_Cause.csv from My Computer to VXG49176.AIRLINE_DELAY_CAUSE' and the timestamp '2023/05/26, 03:49 PM'. It also includes 'View details' and 'Errors 0' and 'Warnings 0' buttons.

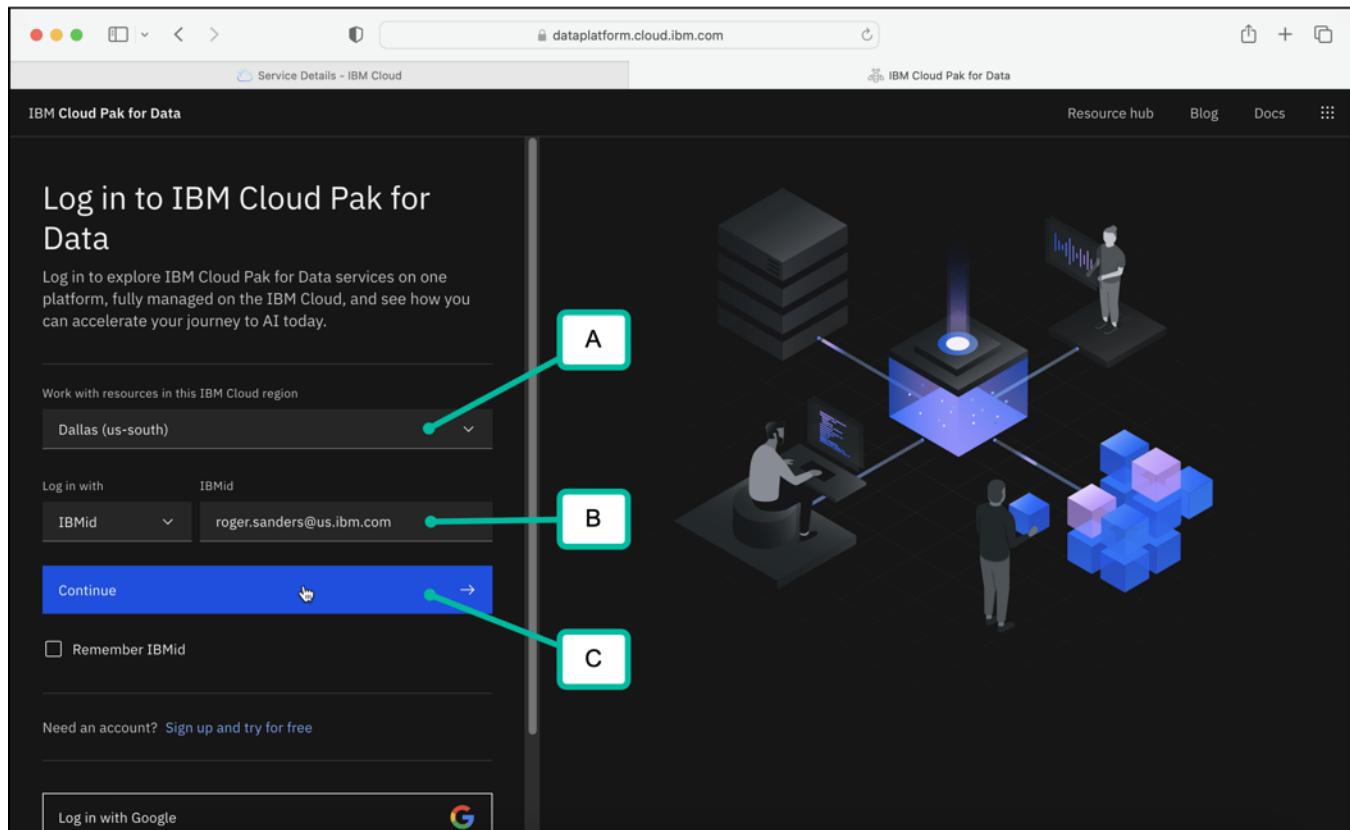
Congratulations! You have just created and populated a table named **AIRLINE_DELAY_CAUSE** in a Db2 on Cloud database. We will use the data in this table in another exercise later.

IV. Using IBM Watson Studio to run a Jupyter Notebook

The free version of IBM Watson Studio (offered with IBM Cloud Pak for Data as a Service) will be used to execute the Jupyter Notebooks that are needed to complete the remaining exercises in this workshop. So, in this section you will set up a free IBM Cloud Pak for Data as a Service account and then use that account to run a simple Jupyter Notebook.

1. Create an IBM Cloud Pak for Data as a Service account.

1. Open a web browser and go to the *Log in to IBM Cloud Pak for Data* web site (<https://dataplateform.cloud.ibm.com/login>). Once there, [A] click on the **Work with resources in this IBM Cloud region** field to display a dropdown list of locations where Cloud Pak for Data as a Service can be deployed and choose the location desired. (In this example, the **Dallas** location was selected.) Next, make sure the **Log in with** field contains the value **IDBId**, and then [B] enter the email address you used to create your IBMid in the **IBMid** field (located beside the **Log in with** field). **IMPORTANT: Please use the same IBMid that was used earlier to create an IBM Cloud account and provision a Db2 on Cloud Lite plan database.** When finished, [C] click the **Continue** button.



2. When the *Provide your information to continue* screen appears, [A] enter your company name in the **Company name** field and [B] your telephone number in the **Phone number** field. (Click on the **Phone number** dropdown list and select the correct country code if your country code is not **+1**.) Then, [C] indicate how you would like for IBM to contact you by selecting the **by email** and/or the **by phone** checkboxes. Finally, [D] click the **Continue** button.

eu-gb.dataplatform.cloud.ibm.com

Service Details - IBM Cloud

Preparing Cloud Pak for Data | IBM Cloud Pak for Data

IBM Cloud Pak for Data

Provide your information to continue

Company name

IBM

Phone number

+1 8887268750

IBM may use my contact data to keep me informed of products, services, and offerings:

by email

by phone

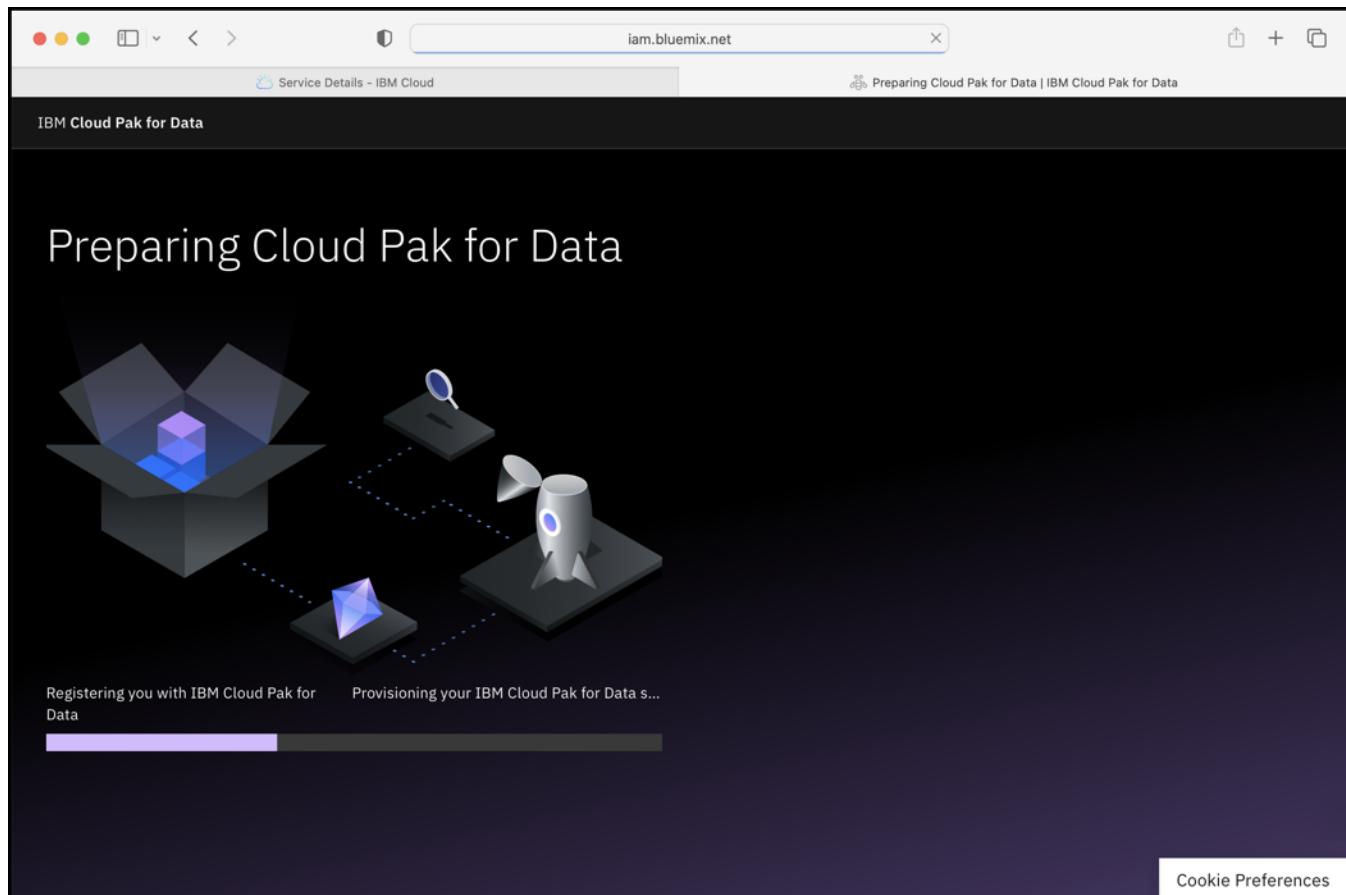
You can withdraw your marketing consent at any time by submitting an [opt-out request](#). Also, you may unsubscribe from receiving marketing emails by clicking the unsubscribe link in each email. More information on our processing can be found in the [IBM Privacy Statement](#).

By submitting this form, I acknowledge that I have read and understand the [IBM Privacy Statement](#) and I accept the product [Terms and Conditions](#) of this registration form.

Continue

Cookie Preferences

_____ 3. While you are being registered with IBM Cloud Pak for Data and your IBM Cloud Pak for Data as a Service environment is being created, you should see a *Preparing Cloud Pak for Data* screen similar to this:



4. Once your IBM Cloud Pak for Data as a Service environment has been created, you should see a *Welcome* screen that looks something like this:

The screenshot shows the 'Home | IBM Cloud Pak for Data' interface. At the top, there's a navigation bar with tabs for 'IBM Cloud Pak for Data' (selected), 'Service Details - IBM Cloud', 'Upgrade', a help icon, a notification bell, account information (2628635 - IBM, London), and a 'RS' button. Below the header is a dark-themed main area. On the left, a 'Welcome, Roger!' message is displayed. To the right is a 3D-style graphic of floating 3D cubes. Below the graphic are three sections: 'Take a tutorial', 'Work with data', and 'Find what you need'. The 'Work with data' section includes a 'Projects' card with 'No recent projects' and a 'New project' button, a 'Catalogs' card with 'No catalogs' and a 'New catalog' button, and a 'Notifications' card with 'No notifications' and a descriptive message. On the far left, there's a 'Quick start' sidebar with four items: 'Build customer profiles with IBM Match 360 with Watson', 'Catalog and govern data with Watson Knowledge Catalog', 'Build and manage ML models with Watson Studio', and 'Query data anywhere with Watson Query'. A vertical 'Feedback' button is located on the right edge of the main content area.

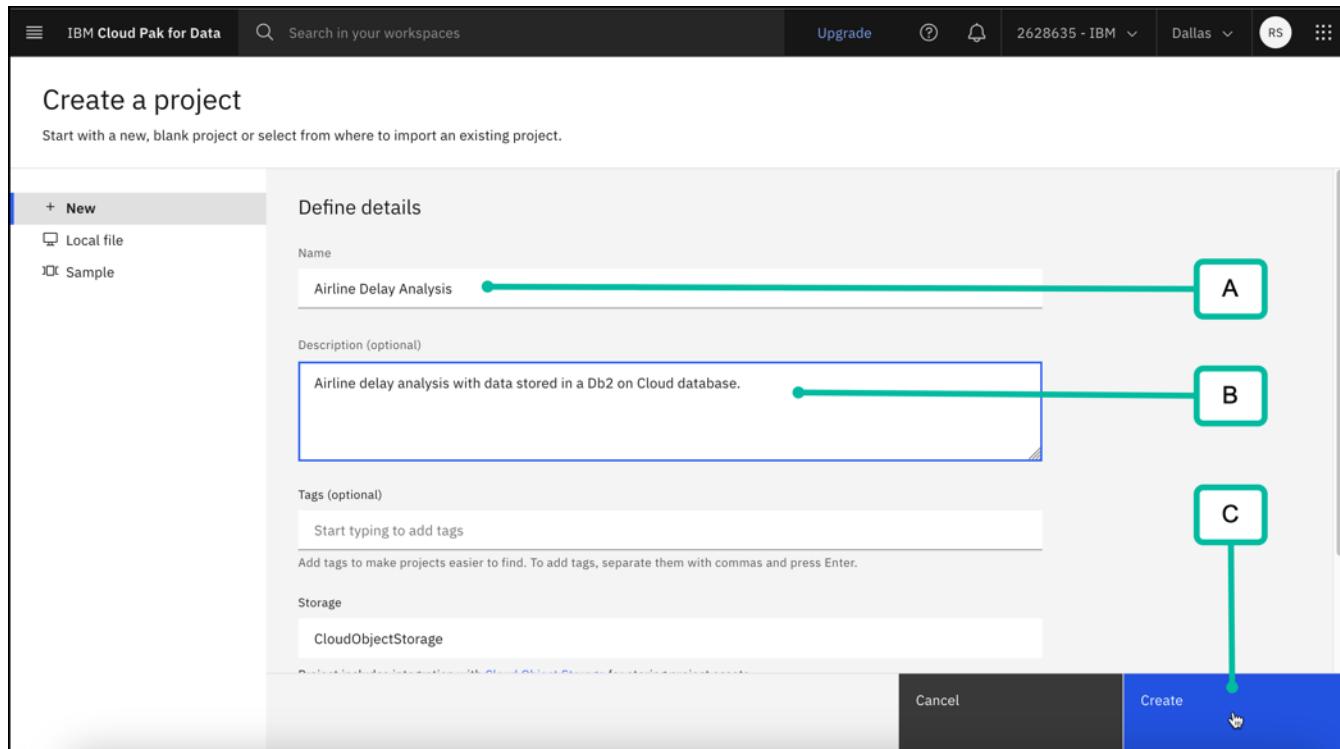
Congratulations! You now have an active **IBM Cloud Pak for Data as a Service** account that can be used to perform the rest of the exercises in this workshop.

2. Create a new project.

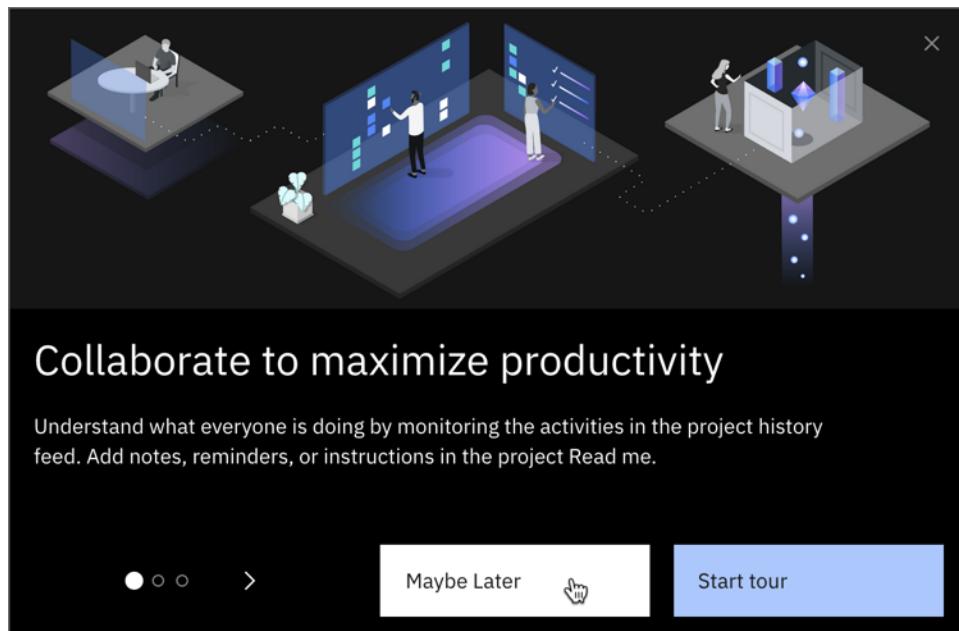
1. Create a new project by clicking on the + icon located in the top right-hand corner of the **Projects** box. This will cause the IBM Cloud Pak for Data as a Service *Welcome* screen to be replaced with a *Create a project* screen.

The screenshot shows the IBM Cloud Pak for Data 'Welcome' screen. At the top, there's a navigation bar with 'IBM Cloud Pak for Data', a search bar, 'Upgrade', a help icon, a notifications icon, '2628635 - IBM', 'Dallas', and a user profile icon. Below the header, it says 'Welcome, Roger!'. There are three main sections on the left: 'Take a tutorial', 'Work with data', and 'Find what you need'. The 'Work with data' section has a sub-section 'Create a project for your team to prepare data, find insights, or build models.' with a '→' arrow. The 'Find what you need' section features an interactive map with blue and purple cubes and a magnifying glass icon. In the center, there's a 'Create new project' button above a 'Projects' box. The 'Projects' box shows 'No recent projects' and a 'New project +' button. To the right are 'Catalogs' (No catalogs, New catalog +) and 'Notifications' (No notifications). On the far left, there's a 'Quick start' sidebar with five items: 'Create data pipelines with DataStage', 'Build customer profiles with IBM Match 360 with Watson', 'Catalog and govern data with IBM Knowledge Catalog', 'Build and manage ML models with Watson Studio', and 'Query data anywhere with Watson Query'.

____ 2. When the *Create a project* screen is displayed, [A] enter a project name in the **Name** field. Then, [B] enter a project description in the **Description (optional)** field. In this example, the project name “*Airline Delay Analysis*” and the description “*Airline delay analysis with data stored in a Db2 on Cloud database.*” were used. Finally, [C] click the **Create** button.



____ 3. While the project is being created, you may be presented with a *Collaborate to maximize productivity* popup window. If so, click the **Maybe Later** button to close it.



____ 4. Once the new project has been created, you should see a *Project Overview* screen that looks something like this:

The screenshot shows the 'Project Overview' screen for the 'Airline Delay Analysis' project. The top navigation bar includes the IBM Cloud Pak for Data logo, a search bar, upgrade and help buttons, and account information (2628635 - IBM, Dallas). The main content area features a 'Start working' section with four cards: 'Add users as collaborators', 'Add data to work with', 'Work with data and models in Python or R notebooks', and 'Build machine learning models automatically'. Below this is a 'View all' button and a 'Collapse' button. The 'Assets' tab is selected, showing a summary of assets, resource usage (0 CUH), and a project history section indicating the project was created today at 10:48 AM.

3. Add a Jupyter Notebook to the project.

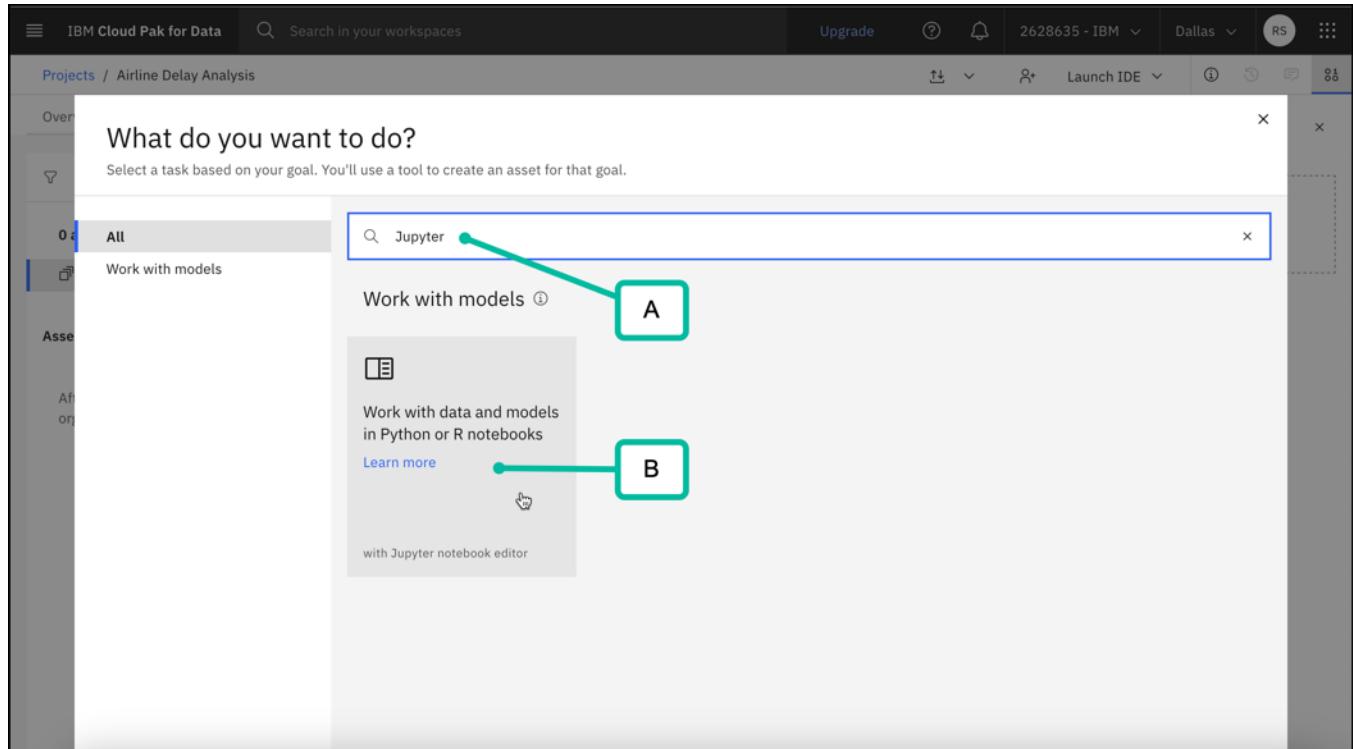
____ 1. From the *Project Overview* screen, click on the **Assets** tab.

This screenshot is identical to the one above, showing the 'Project Overview' screen for the 'Airline Delay Analysis' project. The difference is that the 'Assets' tab is now highlighted with a mouse cursor, indicating it is the active tab.

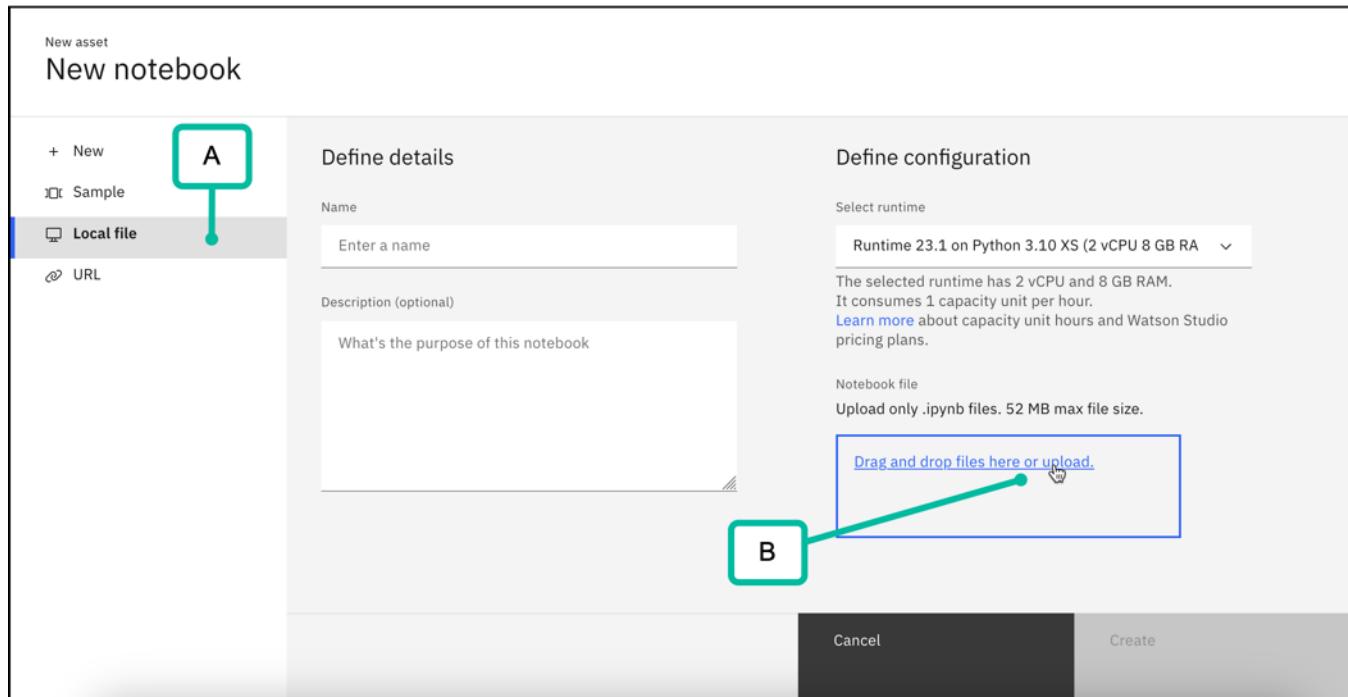
_____ 2. When the Assets screen is displayed, click on the **New asset** + button.

The screenshot shows the 'Assets' tab selected in the IBM Cloud Pak for Data interface. A prominent blue button labeled 'New asset' with a '+' icon is highlighted with a red box. To the right, a modal window titled 'Upload data files' contains a dashed box for dropping files or browsing for upload. The main area displays '0 assets' and a section for 'Asset types' with a note about organizing assets by type. A central callout 'Start adding assets' provides instructions to click 'New asset' or 'Import assets'.

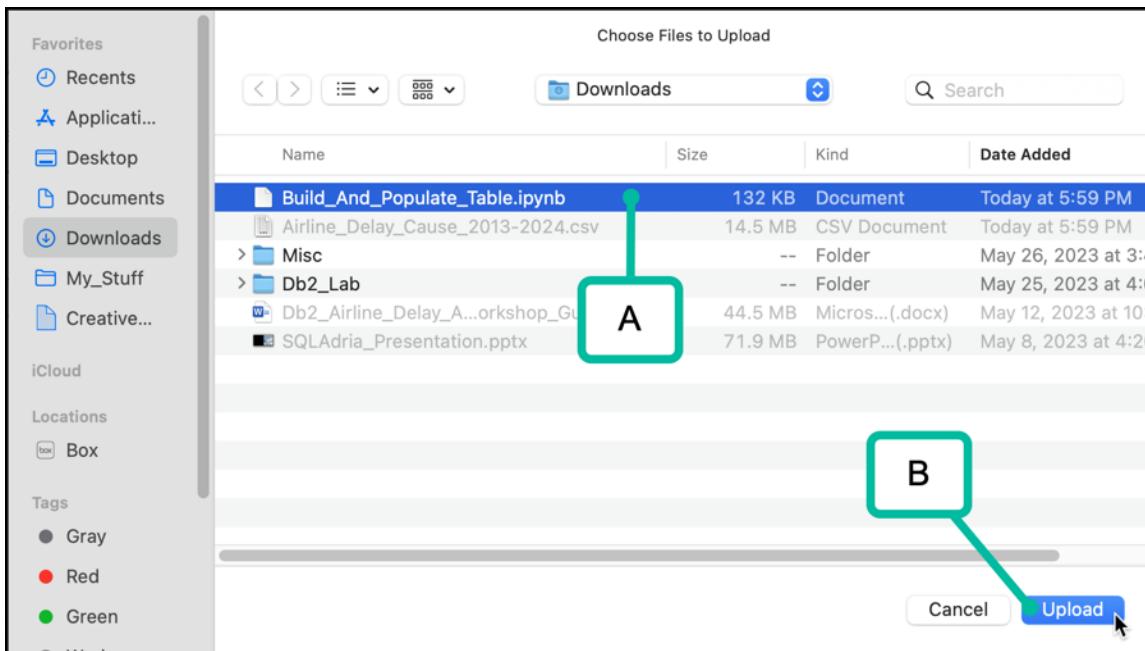
3. From the *What do you want to do* popup window, [A] enter the text “Jupyter” in the **Search** entry field (i.e., the entry field with the magnifying glass icon on the left). Then, [B] click on the **Work with data and models in Python or R notebooks** tile displayed. This will cause the *What do you want to do* screen to be replaced with the *New notebook screen*.



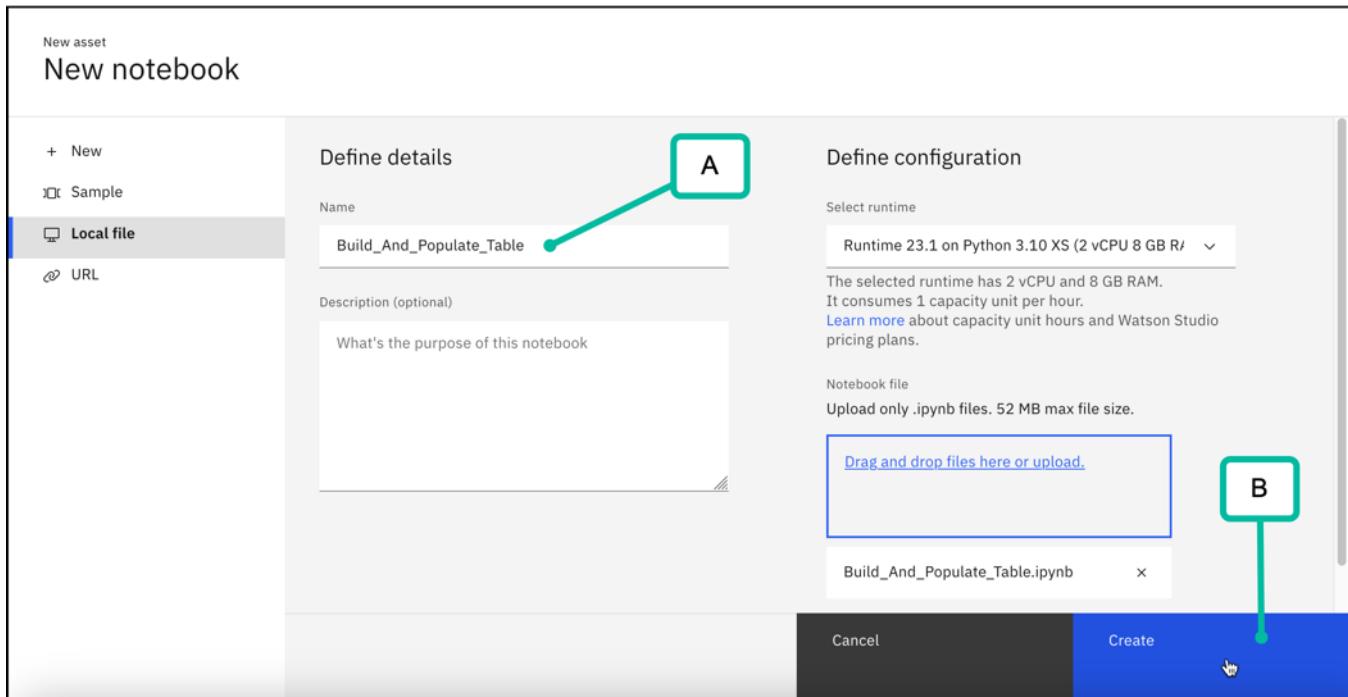
____ 4. From the *New notebook* screen, [A] click on the **Local file** item in the menu shown on the left-hand side of the screen. Then, [B] click on the **Drag and drop files here or upload** link to open a *Finder* window on Mac or a *File Explorer* window on Windows.



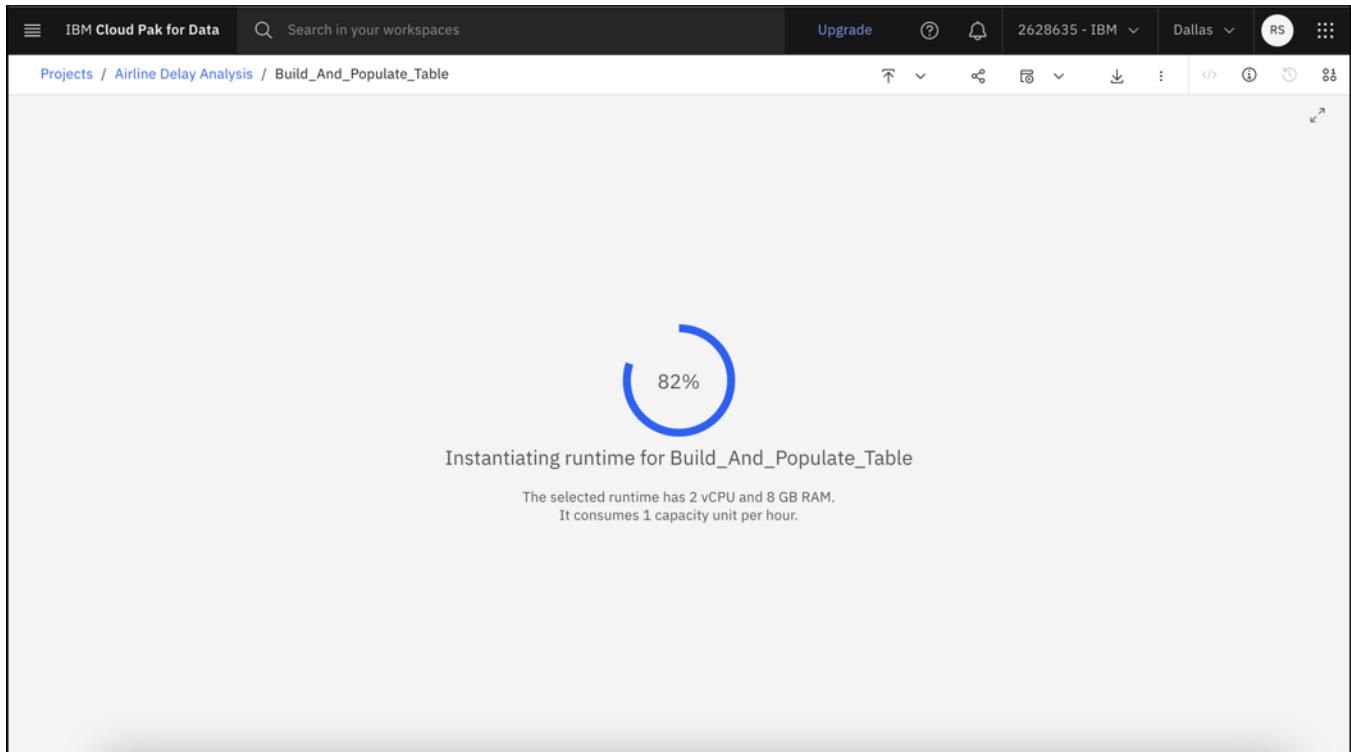
____ 5. [A] Select the file named **Build_And_Populate_Table.ipynb** and [B] click the **Upload** button.



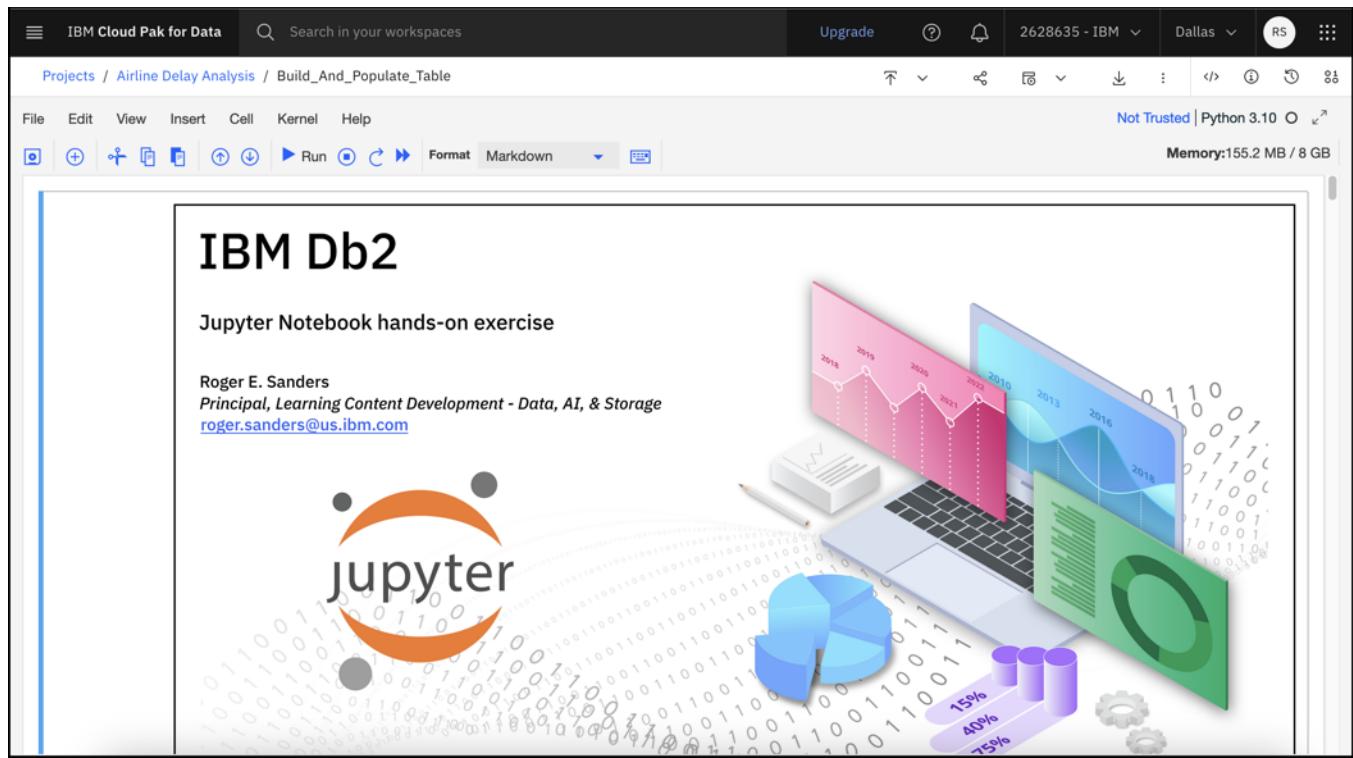
____ 6. [A] Confirm that the **Name** field on the *New notebook* screen contains the name of the file just selected, minus the file extension (**Build_And_Populate_Table**). Then, [B] click the **Create** button located at the bottom of the screen.



7. The *New notebook* screen should be replaced with an *Instantiating runtime for [NotebookName]* screen that looks something like this:

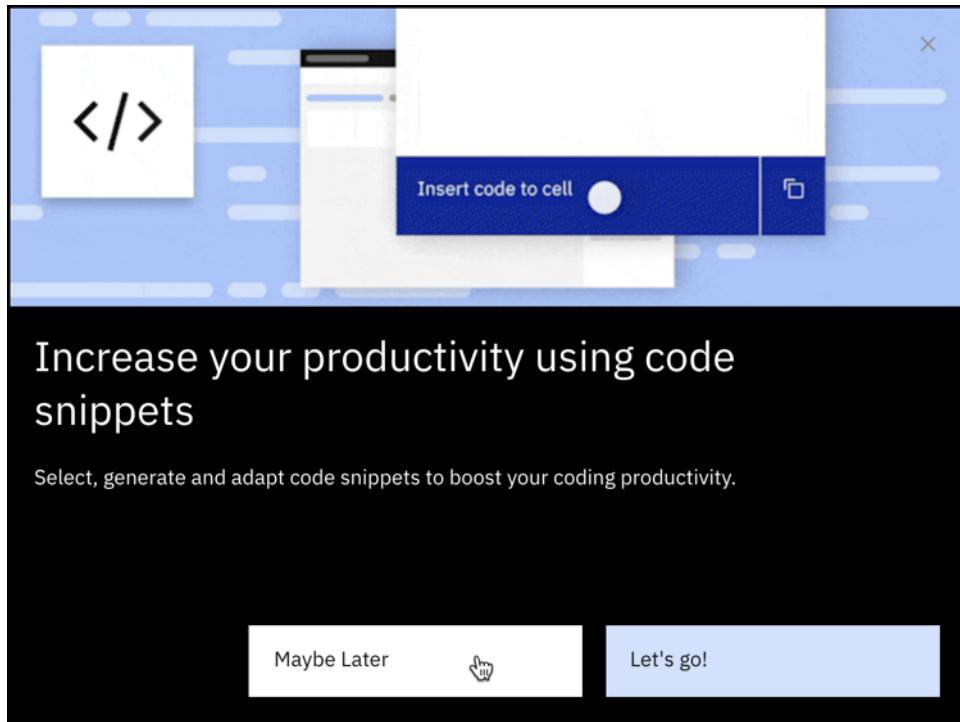


_____ 8. Once the runtime for the Jupyter Notebook that was loaded has been instantiated, the *Instantiating runtime for [NotebookName]* screen should be replaced with a screen that looks similar to this:



The Jupyter Notebook is now ready for editing (which we will do shortly).

_____ 9. Once the Jupyter Notebook is instantiated, you may be presented with an *Increase your productivity using code snippets* popup window. If so, click the **Maybe Later** button to close it. (We will be working with code snippets later.)



4. Add a data file to the project.

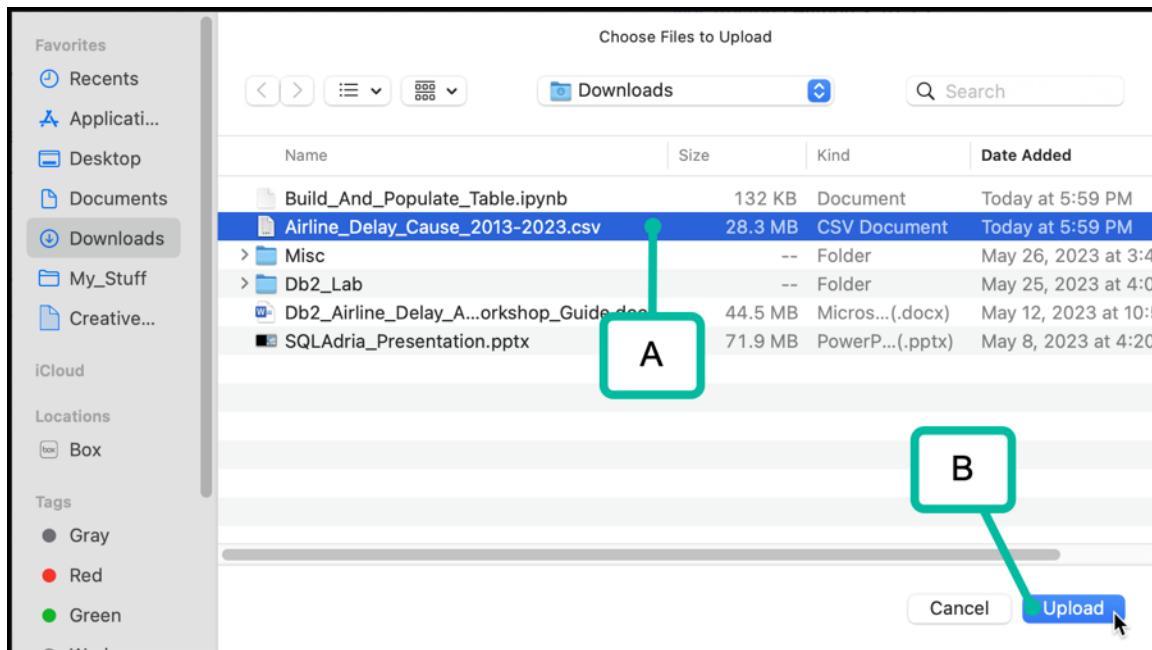
_____ 1. As the name implies, the **Build_And_Populate_Table** Jupyter Notebook just instantiated is designed to create a new table in the Db2 on Cloud database created earlier and populate it with data. But, before the notebook can be used to do that, a data source must be added to our project that the notebook can use. To do this, click on the **Upload asset to project** icon located in the top right-hand corner of the screen.

A screenshot of the IBM Cloud Pak for Data interface. At the top, there is a navigation bar with 'IBM Cloud Pak for Data', a search bar, and various icons. The main area shows a Jupyter Notebook titled 'IBM Db2'. The notebook content includes a section titled 'Jupyter Notebook hands-on exercise' and author information: 'Roger E. Sanders Principal, Learning Content Development - Data, AI, & Storage roger.sanders@us.ibm.com'. In the top right corner of the notebook area, there is a small icon labeled 'Upload asset to project' with a hand cursor icon hovering over it. The bottom right corner of the screen features the IBM logo.

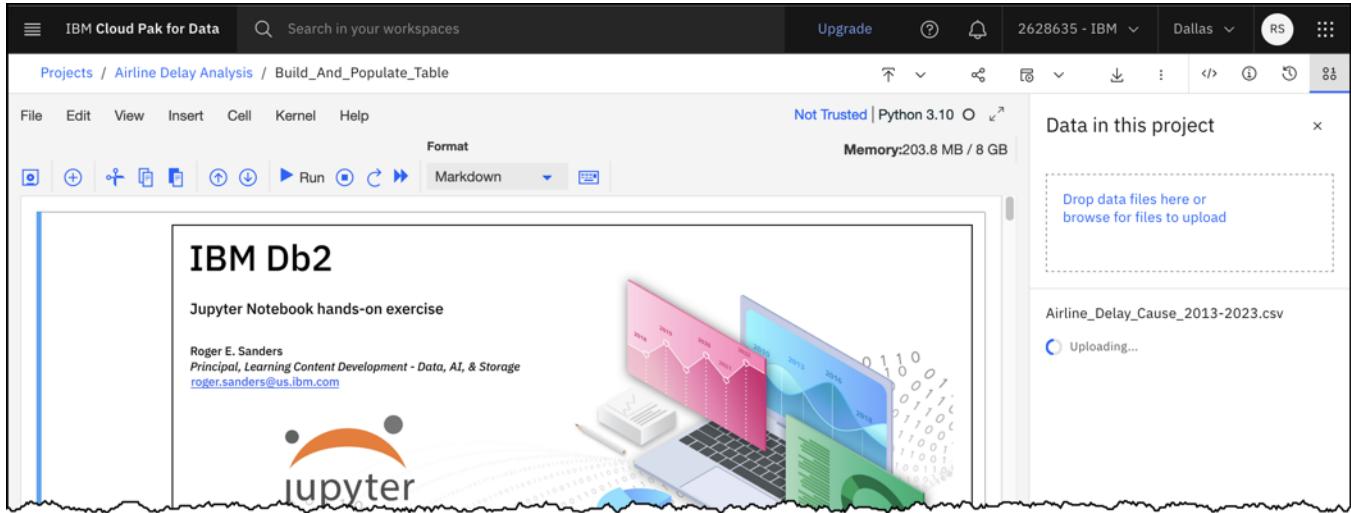
____ 2. When the *Data in this project* slideout window appears, click on the **Drop data files here or browse for files to upload** link to open a *Finder* window on Mac or a *File Explorer* window on Windows.

The screenshot shows the IBM Cloud Pak for Data interface. At the top, there's a navigation bar with 'IBM Cloud Pak for Data', a search bar, and various system status indicators like 'Upgrade', 'Not Trusted | Python 3.10', and 'Memory: 203.7 MB / 8 GB'. Below the navigation bar is a menu bar with 'File', 'Edit', 'View', 'Insert', 'Cell', 'Kernel', and 'Help'. A toolbar follows with icons for file operations like Open, Save, and Run. The main content area displays a Jupyter Notebook titled 'IBM Db2' with a sub-section 'Jupyter Notebook hands-on exercise' by Roger E. Sanders. The notebook features a large graphic of a laptop displaying charts and binary code. To the right of the notebook is a slideout window titled 'Data in this project' containing a dashed box with the text 'Drop data files here or browse for files to upload'.

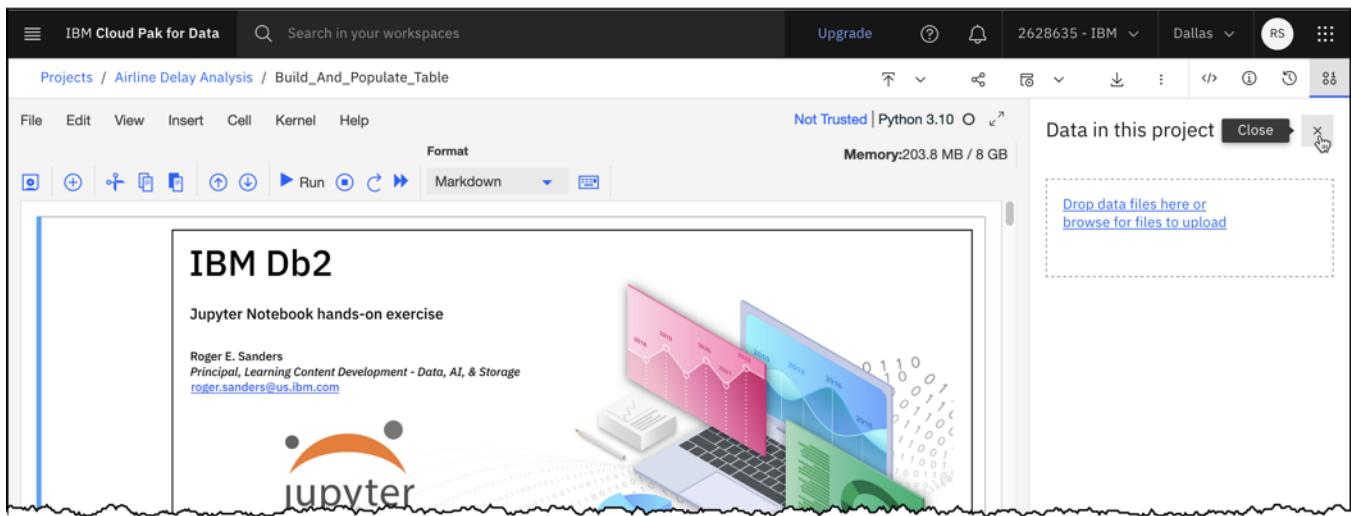
____ 3. [A] Select the file named **Airline_Delay_Cause_2013-2023.csv** and [B] click the **Upload** button.



____ 4. Once the appropriate data file has been selected, the *Data in this project* slideout window should look something like this:



____ 5. After the data file has been successfully added to the project, close the *Data in this project* slideout window by clicking on the X (Close) button located in the top right-hand corner of the window.



Congratulations! You now have a Jupyter Notebook and a data asset in your **Airline Data Analysis** project. However, we need to make a couple of important alterations to the Jupyter Notebook before it can be run.

5. Prepare the Jupyter Notebook for execution.

____ 1. Earlier, we learned that to work with an IBM Cloud resource from an application (such as a Jupyter Notebook), a service credential for that resource must exist. So, right after we created a Db2 on Cloud 'Lite' plan database, we created a service credential for that database (named **AirlineDelayDB**) and we copied the information associated with that service credential into a text file (named **AirlineDelayDB_Credentials.txt**). Open the file that was saved earlier and locate the following information:

[A] User name (“**username**”)

[B] User password (“**password**”)

[C] Database name (“**database**”)

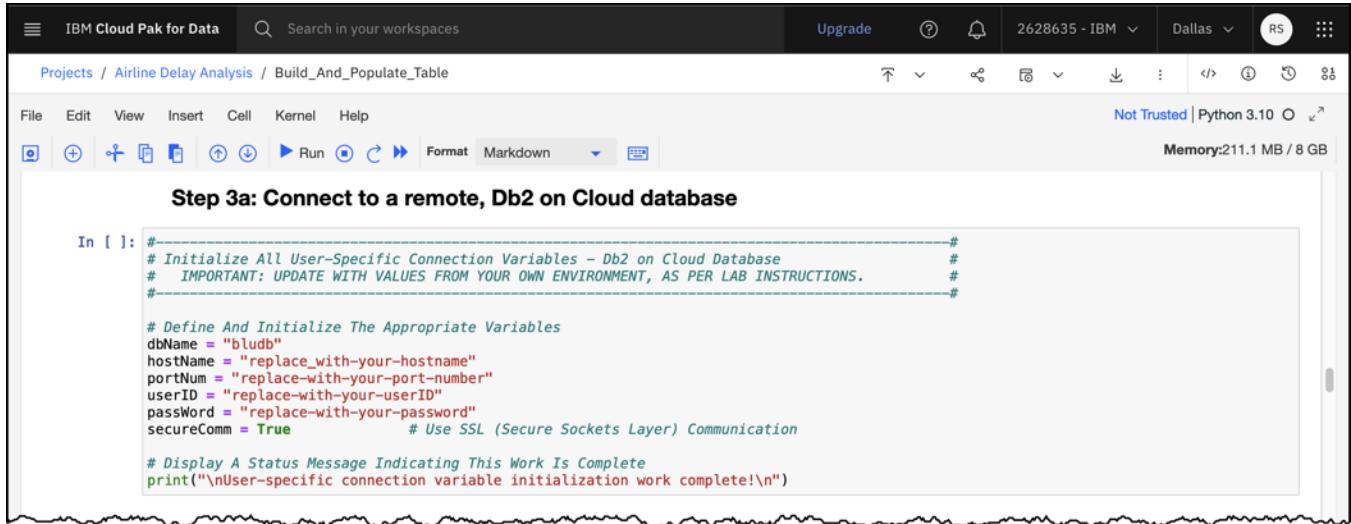
NOTE: The database name is always **bludb** (BLUDB) for Db2 on Cloud databases.

[D] Host name (“**hostname**”)

[E] Port number (“**port**”)

```
1. "environment": {},  
  "type": "cli"  
},  
"db2": {  
  "authentication": {  
    "method": "direct",  
    "password": "N6ogxJFSeE8XOJFo",  
    "username": "vxg49176"  
  },  
  "certificate": {  
    "certificate_base64": "  
LS0tLS1CRUdjTiBDRVJUSUZJQ0FURS0tLs0tCk1JSURIVENDQWdXZ0F3SUJBZ0lVT3dvMC9va09CUEN5RjFWfJxVGHKRW9ubDBv0RRWUpLb1pJaHZTkFRRUwKQIFBd0hqRW  
NNQm9HQTFVRUF3d1RTVUpOSUVOCo2lzVmJRVJoZEdGaViYTmxjekFIrncweU1EQTRNRF3TWpVMwpNalphRncwek1EQTRNREI3TWpVM01qWmFnQjR4SERBYUJnTzCQU1NRT  
BsQ1RTQkRfRzkWkNCRVYUmhZbUZ6CpTyTXdnZ0VpTUewR0NTcUdTzWlzfRFFFQkFRVFBNICRH0Bd2dnRutBb0lCQVFEb0zFNGQSGd0eXZMUVlwR3gKQTBoamRXQnM4NV  
BjTDNyRSt1R3K2diRUdQSUsJU0VZV3o4Y1g1TG1XQk0Y1FrOG9VeSsrQXJ3OEoxaxDrZQpySmllU21clf4WTM0c3BQeGRFVEZkWehScnJhMGU2VmM4MW42Tljl0Zhsn1Q3h  
TG5GMUxFQW9hbhYwaDM2CnhDT0FvcXRwTfRtzNpMTRGeU0yRDRiajkxckl4RGk4V9XMVpVdVhMNtgZXLZUVCe1TruZmhJv3kySVc3aUMkbGpMz3RIN3hZTDVHbVpKOudsY  
WtrSnj1cnpNREFQLzVUYnRIUlydElodTBRSVRFZHIESFYUEZGRDBHYzloZAo3M29JdpVZUJ3VC9uRHN3OTJNNC82SktzWpKn0lpFBTN3Y2a2diUvhINDIBaUVJNxPQdUvpVz  
NOY9GR0pYCmY2aJBZ01CQUFhalV6CQJNQjBHQTfVZERnUvdCQIR2z2RU5MjFVbWZnQ003MmxOcmMzSDl2bURBZkJnTYKSFNNUdEQVdnQlr2Rzz2RU5MjFVbWZnQ0  
03MmxOcmMzSDl2bURBUEJnTzIuk1CQWY4RUJUQURBUUgvyTEwRwpDU3FHU0UlmORRRUJDd1VBQTRJQkFR0TgvdfVnUTZlaTZyWHZndDj0dJdrbkpva1Y5UWNkaTNzbFVFWk  
NDUytjCIVQZ3NnMnVmBMDxChWtM1mRkhjcHZ1Vmp0VHRYTmk2NUM2WlzsRnYxc3p1cU9zdFB5bkJ4bIn4cUs0dkc0dTkVjBWRUgxcE1ZnZBSmxkV3c4UEJTZGjtTk1HdGM4SzlwT  
0o5OvdBQ1ZFfRXVXVGdDehJKTXFBZnpYUXlidUV0dwpoCw1pV2swTmVXNGk5ZEY4S2dTWUvaQWFodXVBSIRidXB2R2RPV1UoeEV4bm03aEVrbmZPV2ZITThDd08xNWfZCIRGQ  
2s0QopDUmR4Mlg5U284V3o1Z3MzcncyRkFDQJyZONyeffFDZnZrTZUdVNHNkxFRHJHbmpWaXVSQkpZdW4KT1RxWXROaVBHaHpuTHJrL0Fzam1LMzBxQmFLTmFyNUdQajhqlapN  
b2RiZ04KLs0tLs1FTkQgQoVsEIGSUNBVEUts0tLQo=",  
  "name": "1dd14d0c-1b52-4f63-a606-53ecba28771d"  
},  
  "composed": [  
    "db2://vxg49176:N6ogxJFSeE8XOJFo@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2i090l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb?  
authSource=admin&replicaSet=replicaSet"  
],  
  "database": "bludb",  
  "host_ress": [],  
  "hosts": [  
    {  
      "hostname": "2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2i090l08kqb1od8lcg.databases.appdomain.cloud",  
      "port": 32328  
    }  
  ],  
  "jdbc_url": [  
    "jdbc:db2://2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2i090l08kqb1od8lcg.databases.appdomain.cloud:32328/  
bludb:userid;password=your_password;sslConnection=true;"  
  ]
```

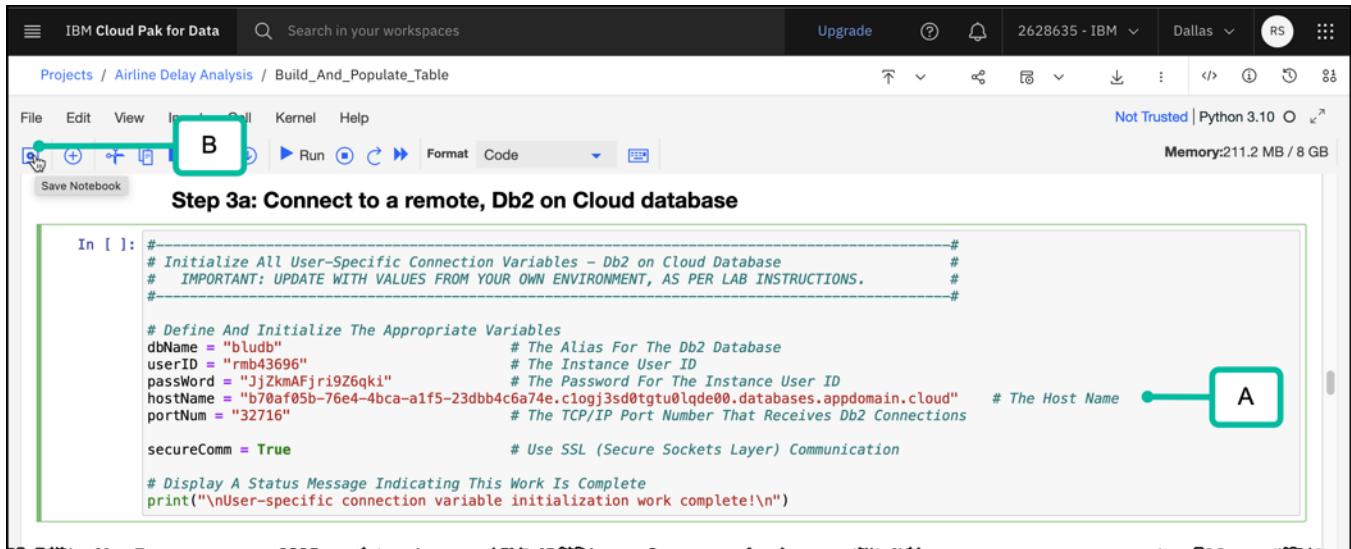
_____ 2. In the Jupyter Notebook, scroll down until you see the code section under the heading **Step 3a: Connect to a remote, Db2 on Cloud database.**



The screenshot shows the IBM Cloud Pak for Data Jupyter Notebook interface. The title bar reads "IBM Cloud Pak for Data". The top menu includes "File", "Edit", "View", "Insert", "Cell", "Kernel", and "Help". The status bar indicates "Not Trusted | Python 3.10" and "Memory:211.1 MB / 8 GB". The main area displays a code cell titled "Step 3a: Connect to a remote, Db2 on Cloud database". The code initializes connection variables for a Db2 on Cloud database, including db_name, host_name, port_num, user_id, and pass_word, and prints a completion message.

```
In [ ]: #  
# Initialize All User-Specific Connection Variables - Db2 on Cloud Database  
# IMPORTANT: UPDATE WITH VALUES FROM YOUR OWN ENVIRONMENT, AS PER LAB INSTRUCTIONS.  
  
# Define And Initialize The Appropriate Variables  
dbName = "bludb"  
hostName = "replace_with-your-hostname"  
portNum = "replace-with-your-port-number"  
userID = "replace-with-your-userID"  
passWord = "replace-with-your-password"  
secureComm = True          # Use SSL (Secure Sockets Layer) Communication  
  
# Display A Status Message Indicating This Work Is Complete  
print("\nUser-specific connection variable initialization work complete!\n")
```

_____ 3. [A] Click on this section to make it the current section. Then, copy the **hostname** value you found in the service credential file (**AirlineDelayDB_Credentials.txt**) to the clipboard and paste it over the text “**replace-with-your-hostname**” found in this code section. (*Copy the value beside the **hostname** keyword in the service credential file and do NOT remove the quotation marks!*) Repeat this process for the **port** (“**replace-with-your-port-number**”), **username** (“**replace-with-your-userID**”), and **password** “**replace-with-your-password**” keywords. When you are finished, [B] click on the **Save Notebook** icon to make the changes made permanent. (NOTE: You can use the “hot keys” **command-C** on Mac or **Ctrl-C** on Windows to copy data to the clipboard; you can use the “hot keys” **command-V** on Mac or **Ctrl-V** on Windows to paste the copied data.)



The screenshot shows the same Jupyter Notebook interface as above, but with annotations. A green box labeled "B" highlights the "Save Notebook" button in the toolbar. A green box labeled "A" highlights the "hostName" variable in the code, which is set to "b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tqtu0lqde00.databases.apdomain.cloud". A blue arrow points from the "A" label to the "hostName" variable.

```
In [ ]: #  
# Initialize All User-Specific Connection Variables - Db2 on Cloud Database  
# IMPORTANT: UPDATE WITH VALUES FROM YOUR OWN ENVIRONMENT, AS PER LAB INSTRUCTIONS.  
  
# Define And Initialize The Appropriate Variables  
dbName = "bludb"          # The Alias For The Db2 Database  
userID = "rmb43696"        # The Instance User ID  
passWord = "JjZkmAFjr19Z6qki" # The Password For The Instance User ID  
hostName = "b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tqtu0lqde00.databases.apdomain.cloud" # The Host Name  
portNum = "32716"          # The TCP/IP Port Number That Receives Db2 Connections  
  
secureComm = True           # Use SSL (Secure Sockets Layer) Communication  
  
# Display A Status Message Indicating This Work Is Complete  
print("\nUser-specific connection variable initialization work complete!\n")
```

____ 4. [A] Scroll down until you see the code section under the heading **Step 4a: Copy data from a Cloud Pak for Data asset**. Next, [B] click on this section to make it the current section. Finally, [C] click on the </> (Code Snippets) icon located near the top right corner of the screen.

The screenshot shows the IBM Cloud Pak for Data interface with the following elements:

- Header:** IBM Cloud Pak for Data, Search in your workspaces, Upgrade, Help, 2628635 - IBM, Dallas, RS.
- Toolbar:** File, Edit, View, Insert, Cell, Kernel, Help, Run, Format, Code.
- Section Headers:** Step 4a: Copy data from a Cloud Pak for Data asset, Step 4b: Copy data from a local, on-premises file.
- Code Block (In []):**

```
# # Load The Data Stored In A File Named "Airline_Delay_Cause_2013-2022.csv" Into
# A pandas DataFrame
#
# Load The Data Stored In The File "Airline_Delay_Cause_2013-2022.csv" Into A pandas DataFrame
df_data_1 = pd.read_csv("Airline_Delay_Cause_2013-2022.csv")
#
# Display The First 5 Rows Of Data That Was Copied From The Data File To The pandas DataFrame
df_data_1.head()
```
- Code Snippets Icon:** A circular icon with '</>' and 'Code Snippets' text.
- Annotations:**
 - [A] Points to the 'Code Snippets' icon in the top right.
 - [B] Points to the code block in the 'Step 4b' section.
 - [C] Points to the 'Code Snippets' icon in the top right.

____ 5. When the *Code Snippets* slideout window appears on the right-hand side of the screen, locate the **Read data** tile (under the heading **Data Ingestion**) and click on it.

The screenshot shows the IBM Cloud Pak for Data interface with the 'Code Snippets' slideout window open on the right side:

- Header:** IBM Cloud Pak for Data, Search in your workspaces, Upgrade, Help, Not Trusted | Python 3.10, Memory: 212.9 MB / 8 GB.
- Toolbar:** File, Edit, View, Insert, Cell, Kernel, Help, Run, Format, Code.
- Code Snippets Slideout:**
 - Code Snippets:** Make your data science workflow more simple and productive with Code Snippets.
 - Data Ingestion:**
 - Read data:** Generate a code snippet to load data from a data asset or connection into your notebook.
 - More code snippets coming:**
 - Let us know what code snippets you would like to work with:**
 - Make a suggestion:**
- Section Headers:** Step 4a: Copy data from a Cloud Pak for Data asset, Step 4b: Copy data from a local, on-premises file.
- Code Block (In []):**

```
# # Load The Data Stored In A File Named "Airline_Delay_Cause_2013-2022.csv" Into
# A pandas DataFrame
#
# Load The Data Stored In The File "Airline_Delay_Cause_2013-2022.csv" Into A Pandas DataFrame
df_data_1 = pd.read_csv("Airline_Delay_Cause_2013-2022.csv")
#
# Display The First 5 Rows Of Data That Was Copied From The Data File To The pandas DataFrame
df_data_1.head()
```
- Section Header:** Step 5: Copy data stored in a pandas DataFrame into the

6. When the *Code Snippets* slideout window is replaced with the *Read data* window, click on the **Select data from project** button. This will cause the *Select data from project* popup window to open.

The screenshot shows the IBM Cloud Pak for Data interface. The top navigation bar includes 'IBM Cloud Pak for Data', a search bar, and various system status indicators like 'Upgrade', 'Not Trusted | Python 3.10', 'Memory:206.2 MB / 8 GB', and location 'Dallas'. A slideout window titled 'Read data' is open on the right, with the sub-titile 'Select data from project' highlighted in blue. The main workspace contains code snippets for steps 3 and 4, and a placeholder for step 5.

Step 4a: Copy data from a Cloud Pak for Data asset

In []:

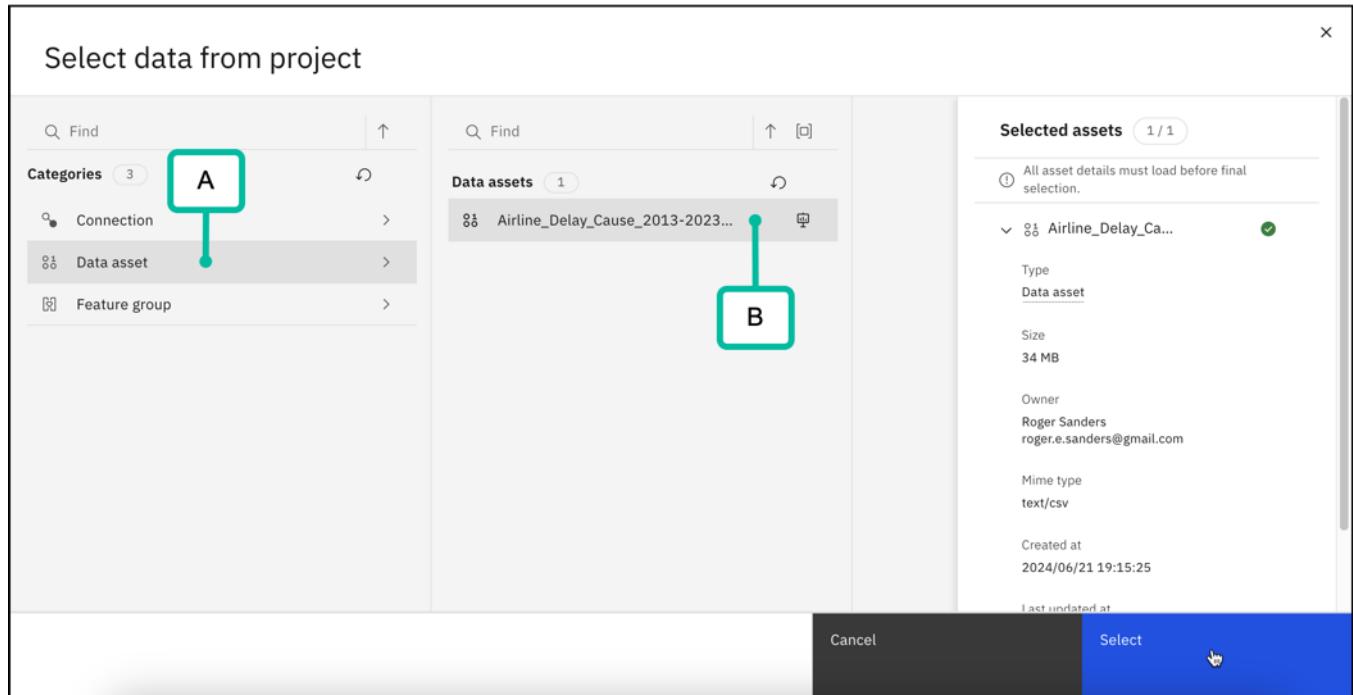
Step 4b: Copy data from a local, on-premises file

In []:

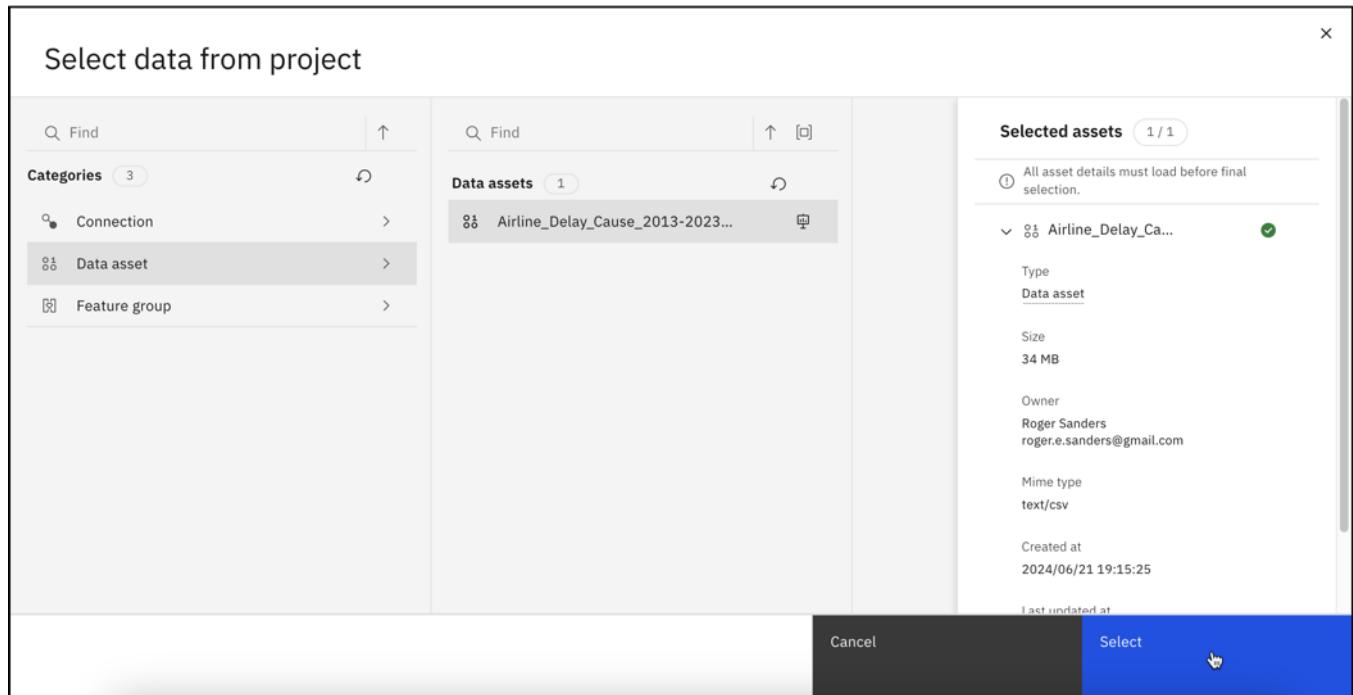
```
# # Load The Data Stored In A File Named "Airline_Delay_Cause_2013-2022.csv" Into
# A pandas DataFrame
#
# Load The Data Stored In The File "Airline_Delay_Cause_2013-2022.csv" Into A Pandas DataFrame
df_data_1 = pd.read_csv("Airline_Delay_Cause_2013-2022.csv")
#
# Display The First 5 Rows Of Data That Was Copied From The Data File To The pandas DataFrame
df_data_1.head()
```

Step 5: Copy data stored in a pandas DataFrame into the

7. When the *Select data from project* popup window appears, [A] locate the **Data asset** item in the **Categories** list and click on it. This will cause a list of all **Data assets** that have been defined for the **Airline Delay Analysis** project to be displayed. When this list appears, [B] locate the **Airline_Delay_Cause_2013-2023.csv** item and click on it to select it. (This should be the only item in the list.)



_____ 8. Review the details provided about the data asset selected (shown under the **Selected assets** heading) and confirm the correct data asset was chosen. Then, click the **Select** button located in the bottom right corner of the window. This will close the popup window and return control to the *Read data* slideout window.



9. The data asset just selected should be displayed under the **Selected data** heading and the **Load as** dropdown/field should contain the value **pandas DataFrame**. (If that's not the case, click on the **Load as** dropdown and select **pandas DataFrame** from the list provided). Click the **Insert code to cell** button shown at the bottom of the window. This will cause the code that is needed to access the asset chosen to be generated in the code section under the heading **Step 4a: Copy data from a Cloud Pak for Data asset**.

The screenshot shows a Jupyter Notebook interface within the IBM Cloud Pak for Data environment. The top navigation bar includes 'IBM Cloud Pak for Data', a search bar, and various system status indicators. The main workspace displays two sections:

- Step 4a: Copy data from a Cloud Pak for Data asset**: A code cell labeled 'In []:' containing the instruction '3. When you are ready, click on the **Run** button or press Shift+Enter to execute the code.' Below it, another instruction says '4. After the code is executed, examine the data that was extracted from the first five rows of the pandas DataFrame and make sure it looks correct.'
- Step 4b: Copy data from a local, on-premises file**: A code cell labeled 'In []:' containing Python code to load a CSV file into a pandas DataFrame and display its first five rows.

On the right side of the interface, there is a sidebar titled 'Read data' which lists the selected data asset as 'Airline_Delay_Cause_2013-2023.csv' and the 'Load as' type as 'pandas DataFrame'. At the bottom right of the workspace, there is a blue button labeled 'Insert code to cell' with a hand cursor icon.

```

In [ ]: # Load The Data Stored In A File Named "Airline_Delay_Cause_2013-2022.csv" Into A pandas DataFrame
# Load The Data Stored In The File "Airline_Delay_Cause_2013-2022.csv" Into A Pandas DataFrame
df_data_1 = pd.read_csv("Airline_Delay_Cause_2013-2022.csv")
# Display The First 5 Rows Of Data That Was Copied From The Data File To The pandas DataFrame
df_data_1.head()

```

10. Once the appropriate code has been generated, close the *Read data* slideout window by clicking on the X (Close) button located in the top right-hand corner of the window.

The screenshot shows a Jupyter-style notebook interface in the IBM Cloud Pak for Data workspace. The title bar indicates the workspace is 'Airline Delay Analysis / Build_And_Populate_Table'. The menu bar includes File, Edit, View, Insert, Cell, Kernel, and Help. The toolbar includes icons for Run, Code, and various cell types. The status bar shows 'Not Trusted | Python 3.10' and 'Memory: 214.1 MB / 8 GB'. The main area contains a code cell (In []):

```
import os, types
import pandas as pd
from botocore.client import Config
import ibm_boto3

def __iter__(self): return 0

# @hidden_cell
# The following code accesses a file in your IBM Cloud Object Storage. It includes your credential.
# You might want to remove those credentials before you share the notebook.

cos_client = ibm_boto3.client(service_name='s3',
    ibm_api_key_id='E_uVwXg_HgkQA73a_urBMYWLfs0WWrVX6rZZuk3HKh',
    ibm_auth_endpoint="https://iam.cloud.ibm.com/oidc/token",
    config=Config(signature_version='oauth'),
    endpoint_url='https://s3.private.us-south.cloud-object-storage.appdomain.cloud')

bucket = 'airlinedelayanalysis2-donotdelete-pr-7jlakvigo8eec3'
object_key = 'Airline_Delay_Cause_2013-2023.csv'

body = cos_client.get_object(Bucket=bucket,Key=object_key)['Body']
# add missing __iter__ method, so pandas accepts body as file-like object
if not hasattr(body, "__iter__"): body.__iter__ = types.MethodType( __iter__, body )

df_1 = pd.read_csv(body)
df_1.head(10)
```

To the right of the code cell is a slideout window titled 'Read data'. It contains instructions: 'Generate a code snippet to load data from a data asset or connection into your notebook.' Below this are sections for 'Selected data' (showing 'Airline_Delay_Cause_2013-2023.csv') and 'Load as' (set to 'pandas DataFrame'). A 'Close' button is visible at the top right of the slideout window. At the bottom of the notebook interface is a blue button labeled 'Insert code to cell'.

____ 11. Edit the code that was generated and stored in the code section under the heading **Step 4a: Copy data from a Cloud Pak for Data asset** as follows:

____ 11a. [A] Remove the lines “`import os, types`” and “`import pandas as pd`”. (These modules are imported in the first code cell in the Jupyter Notebook so there’s no need to import them again here.)

____ 11b. [B] Make sure the two references to the variable `df_x` are set to `df_data_1`. (If you look at the screenshot shown in the previous step, you’ll notice that a variable named `df_1` was generated. The variable name `df_data_1` is referenced in other code cells, so it must be used here.) When you are finished, [C] click on the **Save Notebook** icon located in the top right-hand corner of the screen to make the changes made permanent.

The screenshot shows a Jupyter Notebook interface in a browser window. The title bar says "IBM Cloud Pak for Data". The main area has a toolbar with "File", "Edit", "View", "Insert", "Cell", "Kernel", "Help", "Format", "Code", and "Save Notebook". A "Save Notebook" button is highlighted with a red box and labeled "C". Below the toolbar, the notebook title is "Step 4a: Copy data from a Cloud Pak for Data asset". The code cell contains Python code for reading a CSV file from an S3 bucket using botocore. A red box highlights the "Run" button and the code line "df_data_1 = pd.read_csv(body)" and its reference "df_data_1.head()", both labeled "B". Another red box highlights the "Save Notebook" button labeled "C". The status bar at the bottom says "Trusted | Python 3.10". The notebook title for the next section is "Step 4b: Copy data from a local, on-premises file".

```
In [ ]:
from botocore.client import Config
import ibm_boto3

def __iter__(self): return 0

# @hidden_cell
# The following code accesses a file in your IBM Cloud Object Storage. It includes your credentials.
# You might want to remove those credentials before you share the notebook.
cos_client = ibm_boto3.client(service_name='s3',
    ibm_api_key_id='eAo5nLJDozLD1smMGk4WnA8cIvWD7xQ6uVFvCu_MxSC',
    ibm_auth_endpoint="https://iam.cloud.ibm.com/oidc/token",
    config=Config(signature_version='oauth'),
    endpoint_url='https://s3.private.eu-gb.cloud-object-storage.appdomain.cloud')

bucket = 'airlinedelayanalysis-donotdelete-pr-d5speb4lhy78'
object_key = 'Airline_Delay_Cause_2013-2022.csv'

body = cos_client.get_object(Bucket=bucket,Key=object_key)['Body']
# add missing __iter__ method, so pandas accepts body as file-like object
if not hasattr(body, "__iter__"): body.__iter__ = types.MethodType( __iter__, body )

df_data_1 = pd.read_csv(body)
df_data_1.head()
```

Congratulations! You are now ready to work your way through the Jupyter Notebook.

6. Run the Jupyter Notebook

____ 1. Scroll back to the beginning of the Jupyter Notebook and work your way through each cell, one cell at a time. The notebook is clearly documented and easy to understand.

IMPORTANT: Do NOT execute Section 3, Step 1: Remove the AIRLINE_DELAY_CAUSE_TOTAL table created earlier! This table will be needed for the next part of this workshop. Although this step is not meant to be used, it is included in the notebook to show you how to drop a Db2 on Cloud database table when it is no longer needed.

IBM Cloud Pak for Data

Search in your workspaces

Upgrade

2628635 - IBM

Dallas

IBM

Projects / Airline Delay Analysis / Build_And_Populate_Table

File Edit View Insert Cell Kernel Help

Not Trusted | Python 3.10

Memory: 209.5 MB / 8 GB

Run Cell, Select Below

IBM Db2

Jupyter Notebook hands-on exercise

Roger E. Sanders
Principal, Learning Content Development - Data, AI, & Storage
roger.sanders@us.ibm.com

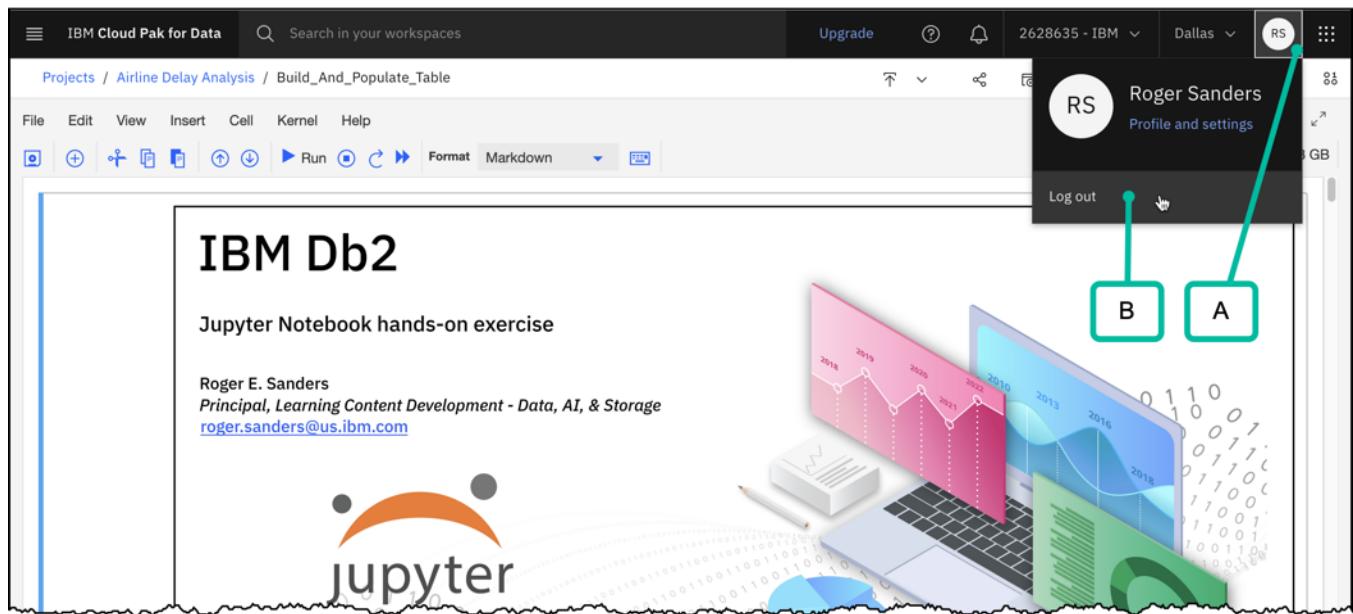
jupyter

Workshop Tip

If your Jupyter Notebook appears to be non-responsive, you can “reboot” it and start over from the beginning by selecting **Kernel** from the menu bar shown at the top of the screen and choosing the option **Restart and Clear Output** from the dropdown menu presented.

This will cause a confirmation popup window to be displayed, and you must confirm the operation before the Jupyter Notebook kernel can be restarted.

_____ 2. When you are finished, log out of IBM Cloud Pak for Data as a Service by [A] clicking on your user icon located in the top right-hand corner of the screen and then [B] clicking on the **Log out** item in the menu presented.



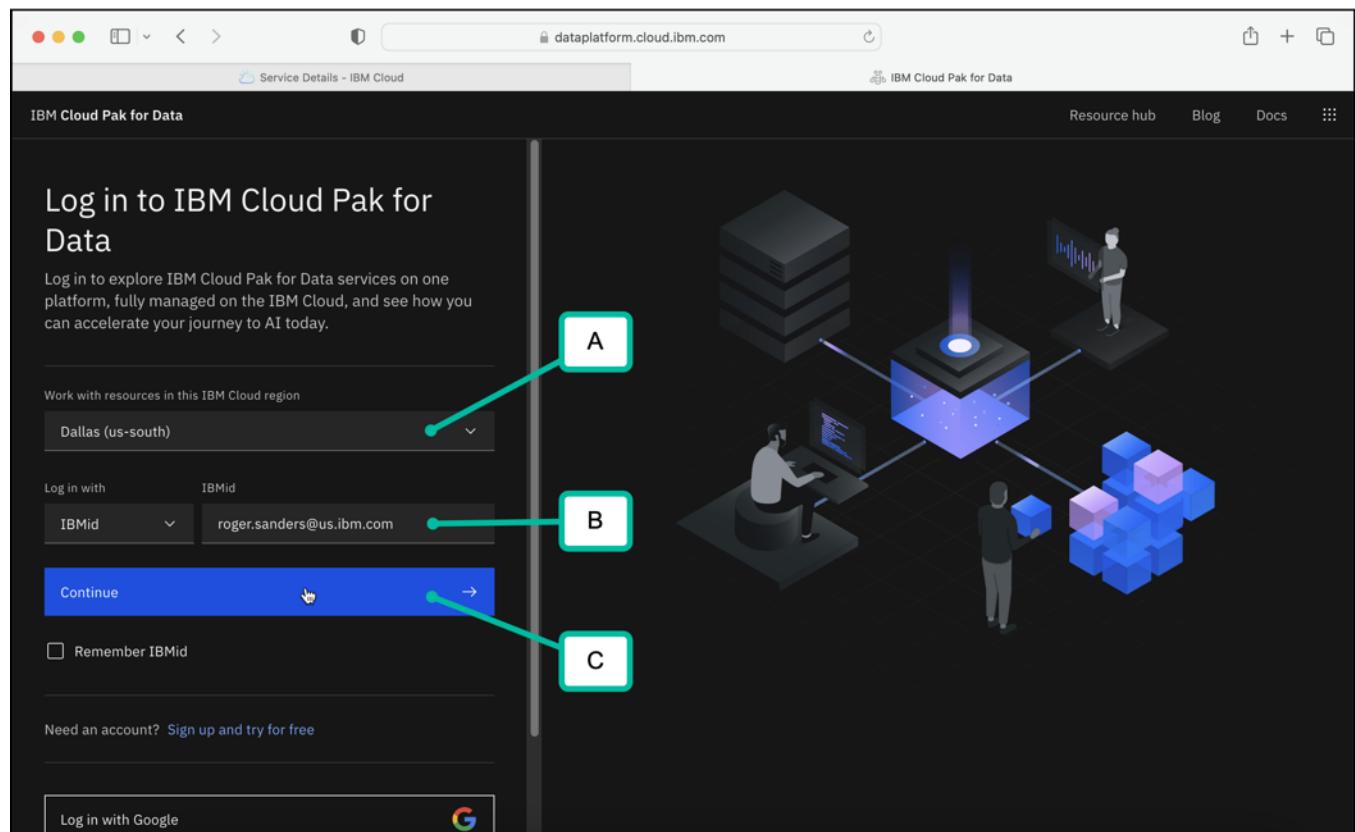
Congratulations! You have just used IBM Watson Studio (in IBM Cloud Pak for Data as a Service) and Jupyter Notebook to work with a Db2 on Cloud database.

V. Analyzing the data

In this section of the workshop, you will use another Jupyter Notebook to analyze the airline flight delay data that was loaded into the Db2 on Cloud database. You will also use the Db2 Magic Commands to establish a database connection and retrieve data from the AIRLINE_DELAY_CAUSE and AIRLINE_DELAY_CAUSE_TOTAL tables.

1. Log in to your IBM Cloud Pak for Data as a Service account.

1. Open a web browser and go to the *Log in to BM Cloud Pak for Data* web site (<https://dataplatform.cloud.ibm.com/login>). Once there, [A] click on the **Work with resources in this IBM Cloud region** field to display a dropdown list of locations where Cloud Pak for Data as a Service can be deployed and choose the location desired. (In this example, the **Dallas** location was selected.) Next, make sure the **Log in with** field contains the value **IDBId**, and then [B] enter the email address you used to create your IBMid in the **IBMid** field (located beside the **Log in with** field). **IMPORTANT: Please use the same IBMid that was used earlier to create an IBM Cloud account and provision a Db2 on Cloud Lite plan database.** When finished, [C] click the **Continue** button.



____ 3. Once you have successfully logged in, you should be presented with an IBM Cloud Pak for Data as a Service *Welcome* screen that looks something like this:

The screenshot shows the IBM Cloud Pak for Data service details page. At the top, there's a navigation bar with icons for back, forward, search, and user profile. The URL is eu-gb.dataplatform.cloud.ibm.com. The main header says "Service Details - IBM Cloud" and "IBM Cloud Pak for Data". On the right, there are buttons for "Upgrade", a help icon, a bell icon, a workspace dropdown (2628635 - IBM), a location dropdown (Dallas), and a "RS" button.

The main content area has a dark background with a 3D cube visualization. It features three main call-to-action sections:

- Take a tutorial**: Step through implementing a Data fabric use case in a sample project. →
- Work with data**: Create a project for your team to prepare data, find insights, or build models. →
- Find what you need**: Open this interactive map to find the tools that you need for your tasks. →

Below these sections is a "Quick start" sidebar with five items:

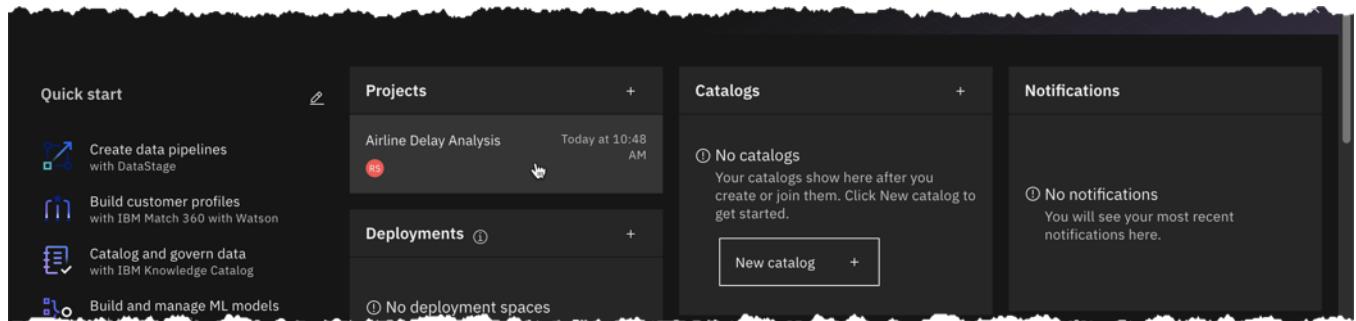
- Create data pipelines with DataStage
- Build customer profiles with IBM Match 360 with Watson
- Catalog and govern data with IBM Knowledge Catalog
- Build and manage ML models with Watson Studio
- Query data anywhere with Watson Query

The main dashboard area shows:

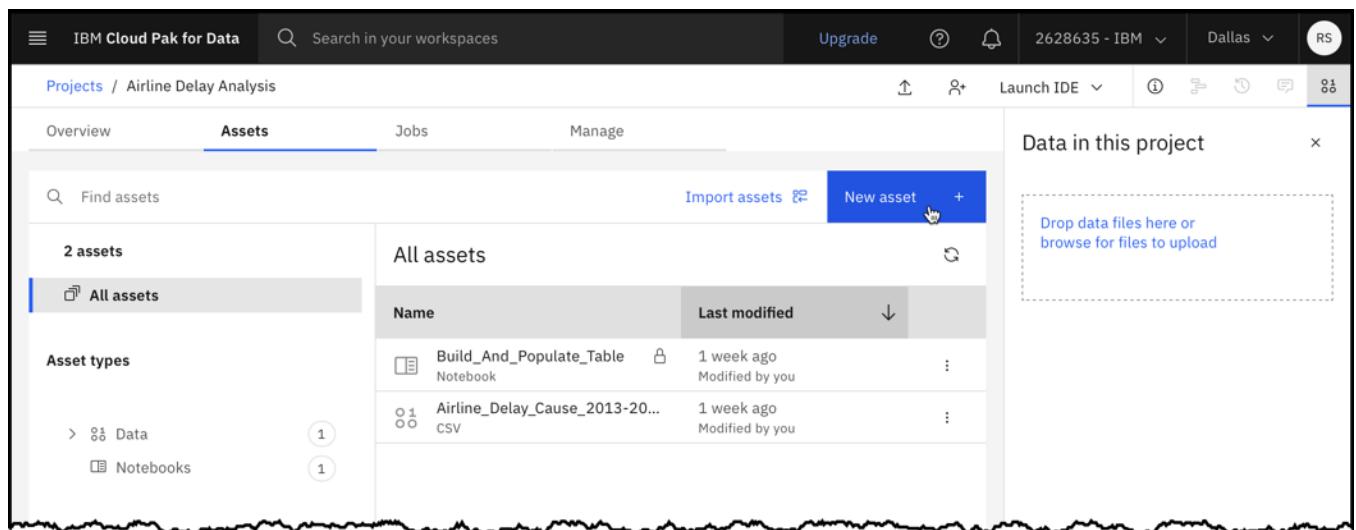
- Projects**: Airline Delay Analysis (Today at 10:48 AM)
- Deployments**: No deployment spaces
- Catalogs**: No catalogs. A "New catalog" button is available.
- Notifications**: No notifications. A message says "You will see your most recent notifications here."
- New in Resource hub**: (empty)

2. Add a new Jupyter Notebook to the project created earlier.

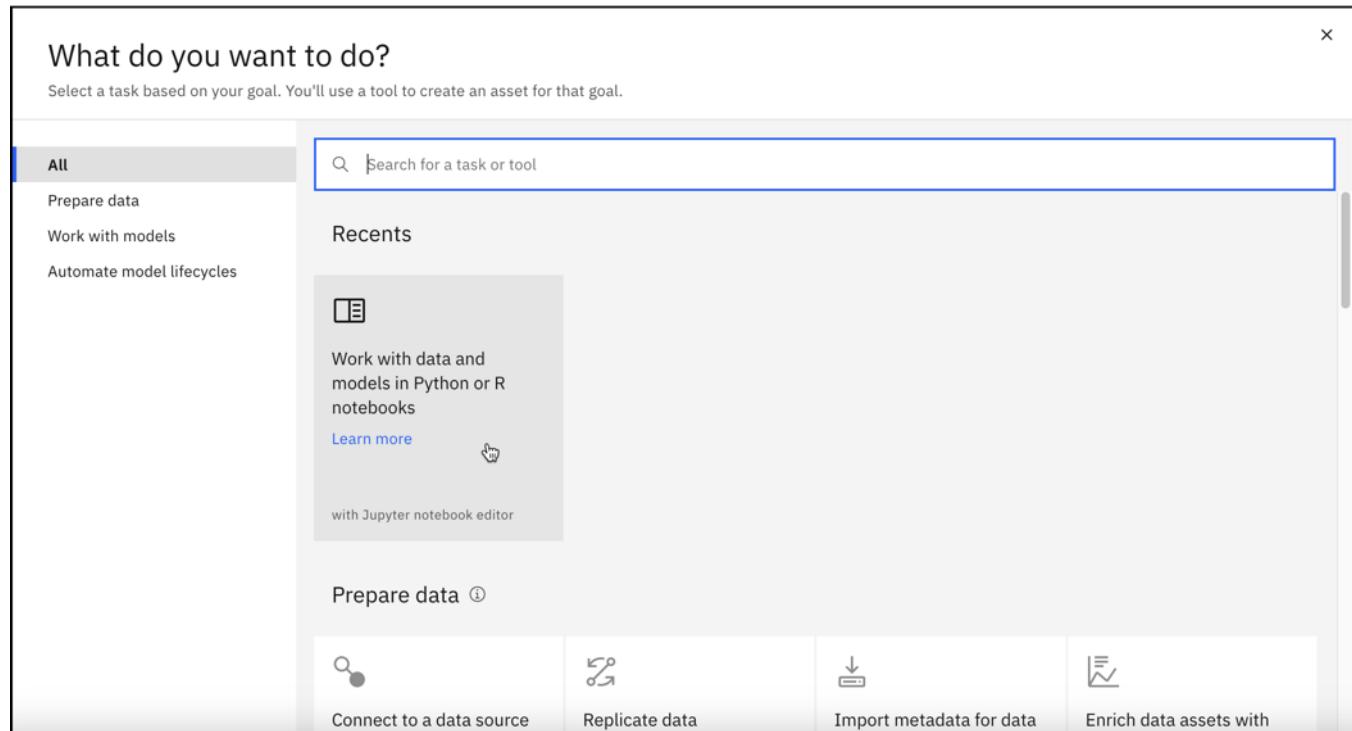
1. When the IBM Cloud Pak for Data as a Service *Welcome* screen appears, locate the project in the **Projects** box that was created earlier (in this example, the project was named “*Airline Delay Analysis*”) and click on it. This should cause the *Welcome* screen to be replaced with an *Assets* screen that shows the assets that have been created for the project selected. (If an *Overview* screen is presented instead, click on the **Assets** tab at the top of the screen to switch to the *Assets* screen.)



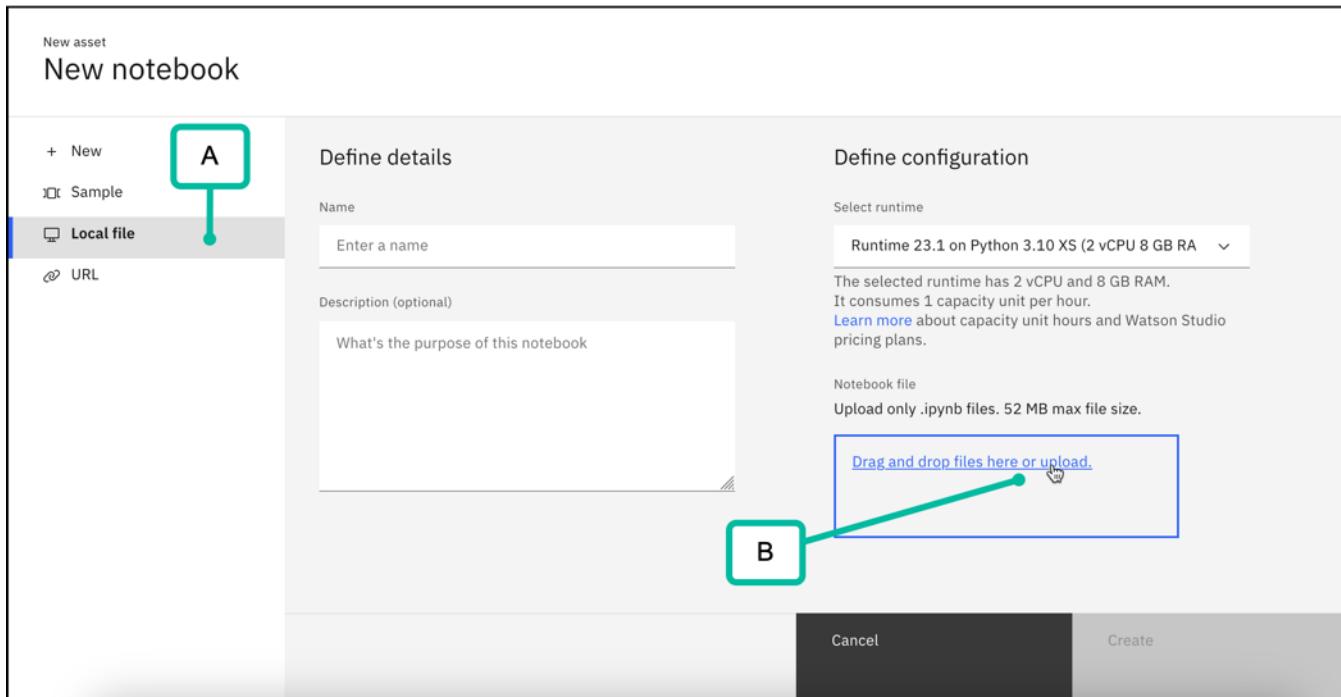
2. When the *Assets* screen is displayed, click on the **New asset** button.



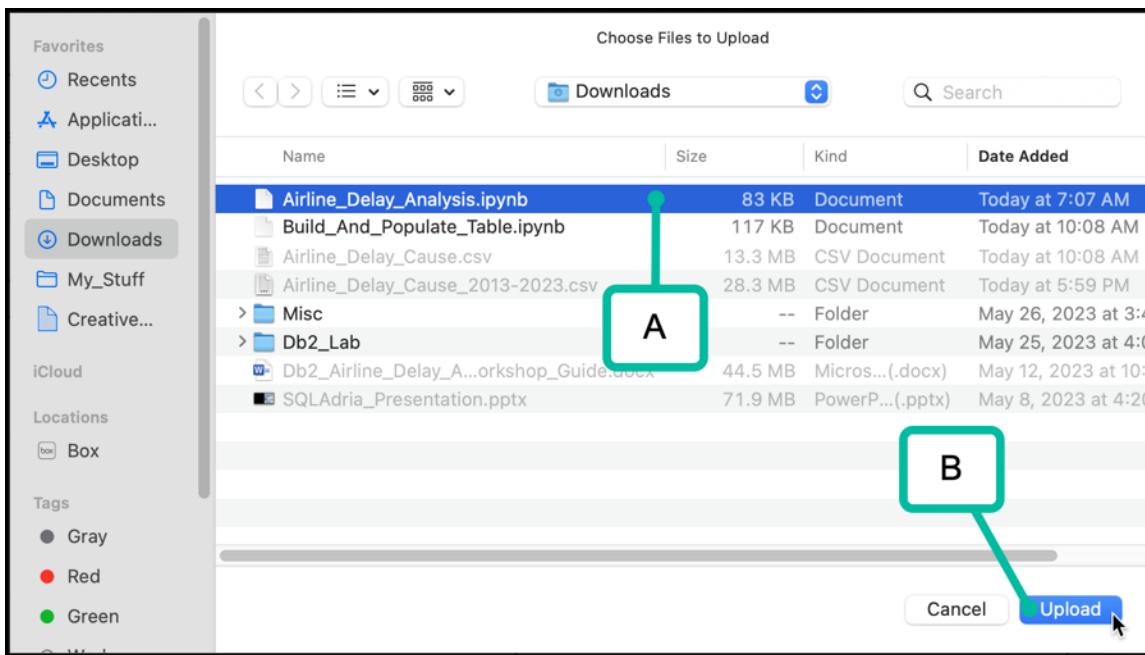
_____ 3. When the *What do you want to do* popup window appears, click on the **Work with data and models in Python or R notebooks** tile. This will cause the *What do you want to do* screen to be replaced with the *New notebook screen*.



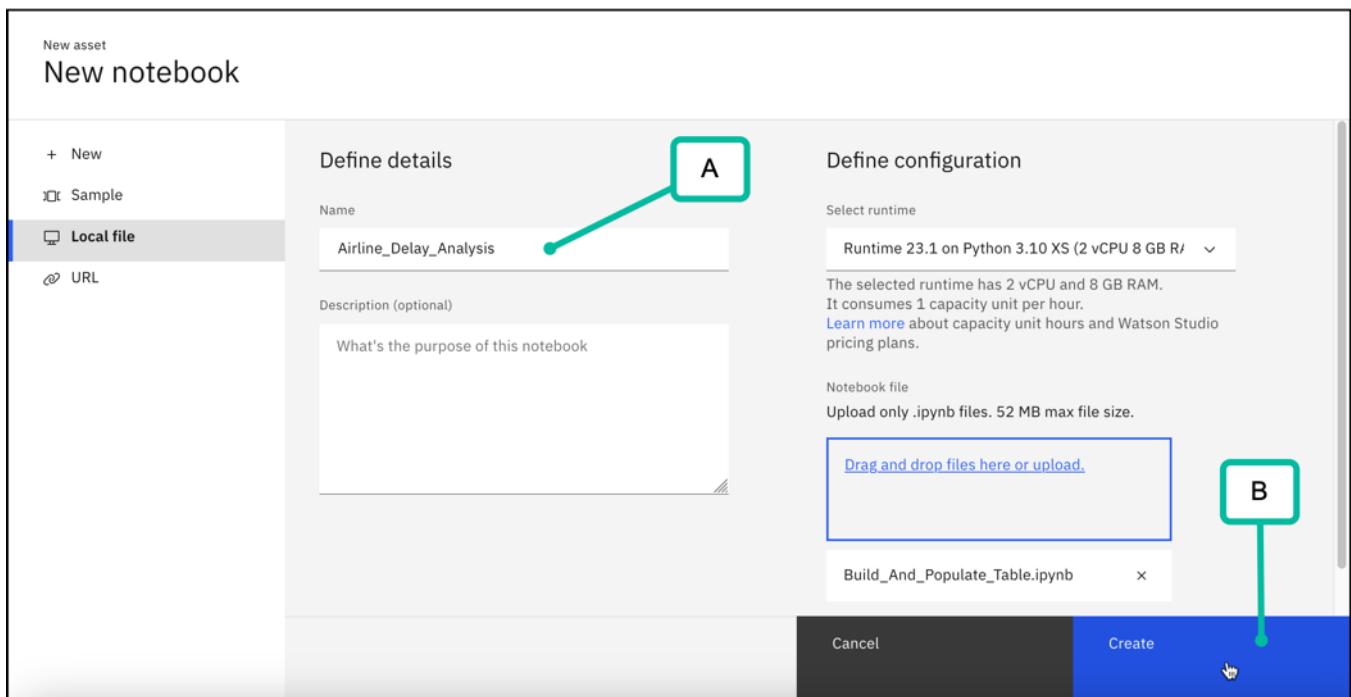
4. From the *New notebook* screen, [A] click on the **Local file** item in the menu shown on the left-hand side of the screen. Then, [B] click on the **Drag and drop files here or upload** link to open a *Finder* window on Mac or a *File Explorer* window on Windows.



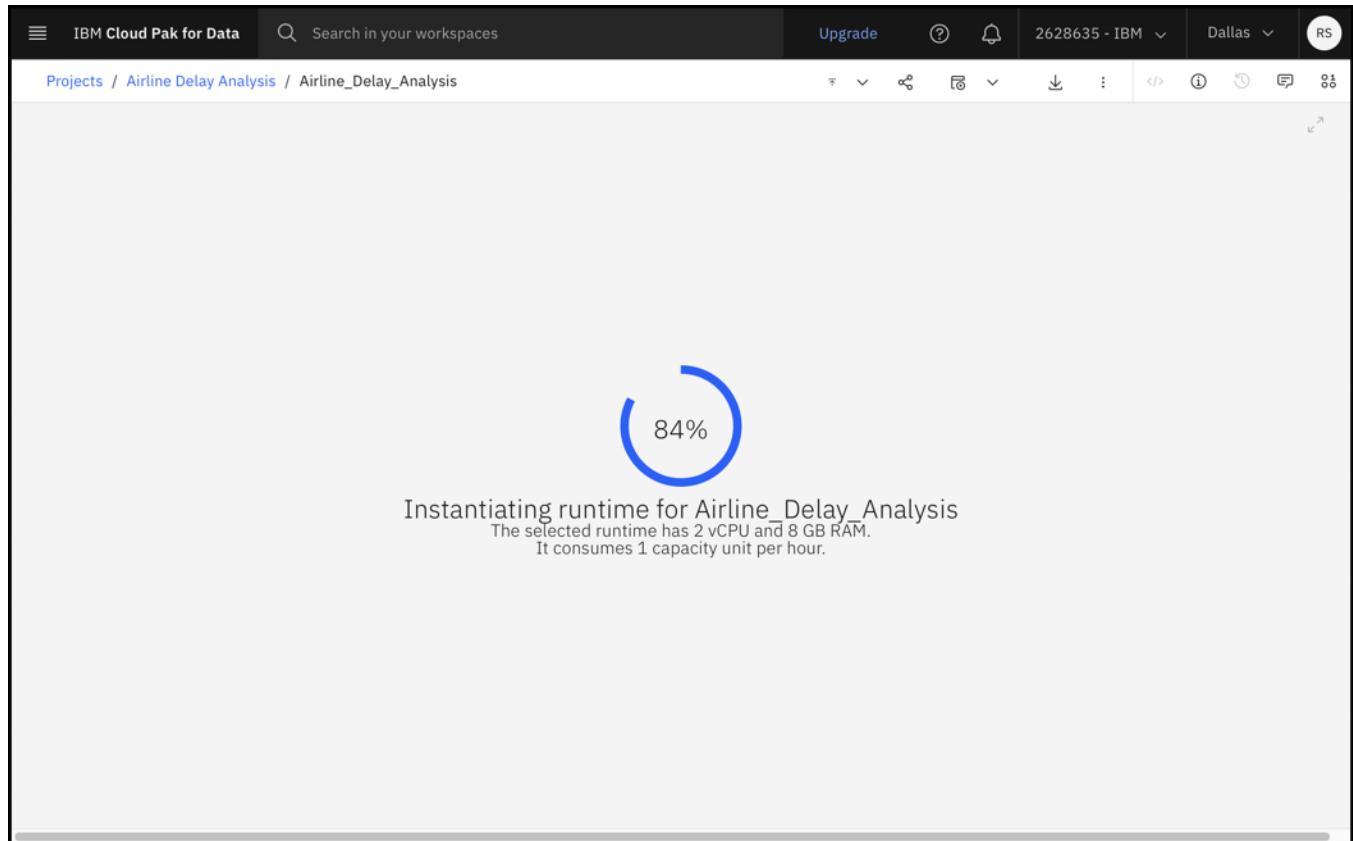
_____ 5. [A] Select the file named **Airline_Delay_Analysis.ipynb** and [B] click the **Upload** button.



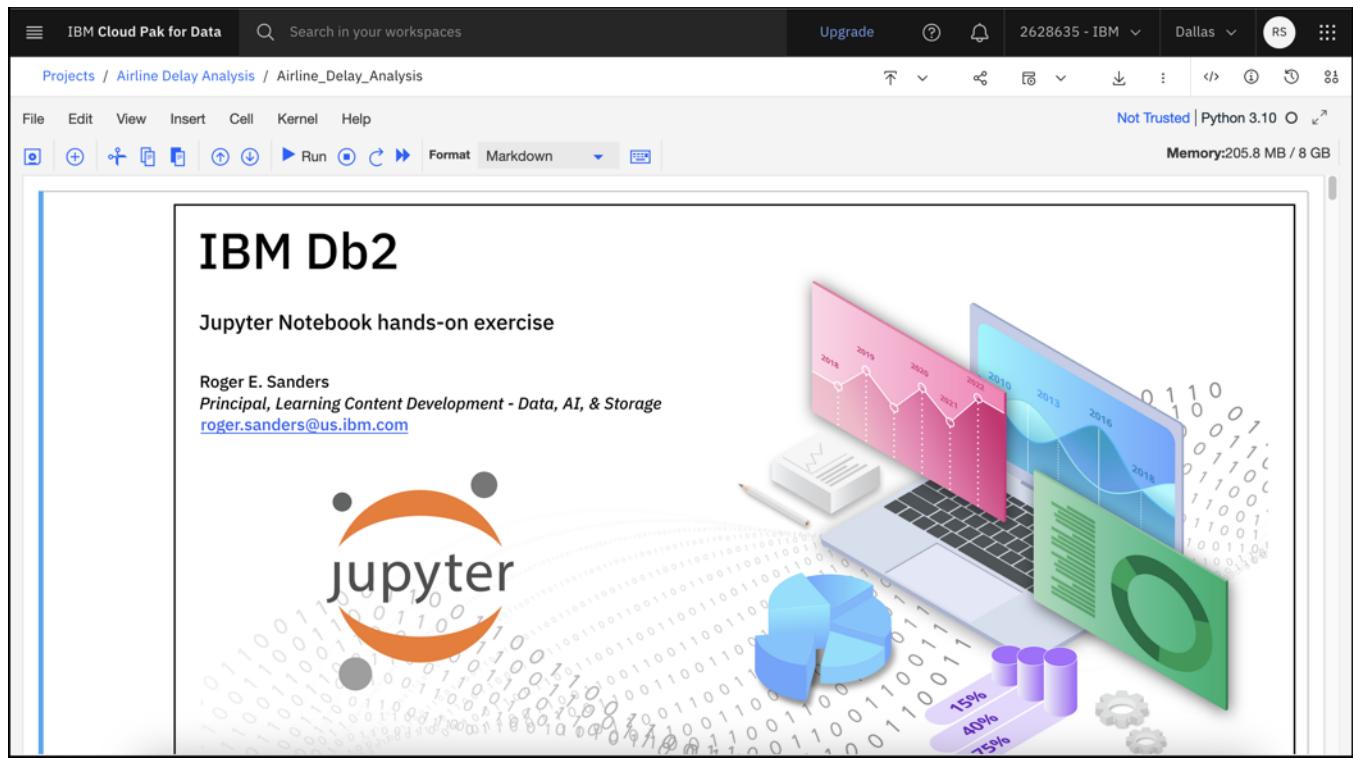
_____ 6. [A] Confirm that the **Name** field on the *New notebook* screen contains the name of the file just selected, minus the file extension (**Airline_Delay_Analysis**). Then, [B] click the **Create** button shown at the bottom of the screen.



7. The *New notebook* screen should be replaced with an *Instantiating runtime for [NotebookName]* screen that looks something like this:



_____ 8. Once the runtime for the Jupyter Notebook that was loaded has been instantiated, the *Instantiating runtime for [NotebookName]* screen should be replaced with a screen that looks similar to this:



The Jupyter Notebook is now ready for editing.

Prepare the Jupyter Notebook for execution.

1. Earlier, we learned that to work with an IBM Cloud resource from an application (such as a Jupyter Notebook), a service credential for that resource must exist. So, right after we created a Db2 on Cloud ‘Lite’ plan database, we created a service credential for that database (named **AirlineDelayDB**) and we copied the information associated with that service credential into a text file (named **AirlineDelayDB_Credentials.txt**). Open the file that was saved earlier and locate the following information:

- [A] User name (“**username**”)
- [B] User password (“**password**”)
- [C] Database name (“**database**”)
- NOTE: The database name is always **bludb** (BLUDB) for Db2 on Cloud databases.
- [D] Host name (“**hostname**”)
- [E] Port number (“**port**”)

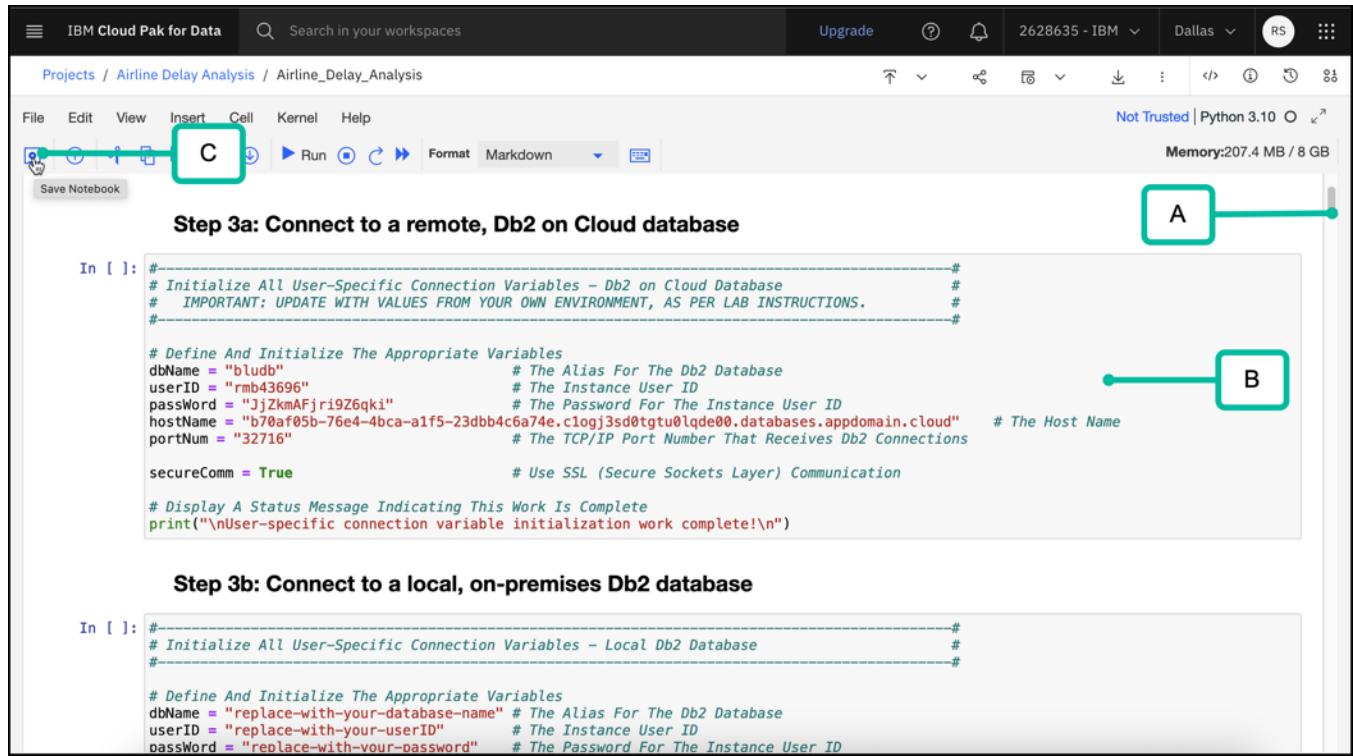
The diagram illustrates the mapping of text file fields to JSON objects in the `AirlineDelayDB_Credentials.txt` file. The file content is a JSON object with several properties. Arrows point from specific fields in the text to their corresponding JSON keys:

- Field [B] "User password ("password")" points to the `"password": "N6ogXjFSeE8XOJFo"` key in the `db2` object.
- Field [C] "Database name ("database")" points to the `"database": "bludb"` key in the `hosts` array.
- Field [D] "Host name ("hostname")" points to the `"hostname": "2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb?authSource=admin&replicaSet=repset"` key in the `hosts` array.
- Field [E] "Port number ("port")" points to the `"port": 32328` key in the `hosts` array.

```
File content (AirlineDelayDB_Credentials.txt):
```

```
{
  "environment": {},
  "type": "cli",
  "db2": {
    "authentication": {
      "method": "direct",
      "password": "N6ogXjFSeE8XOJFo",
      "username": "xvg49176"
    },
    "certificate": {
      "certicate_base64": "LS0tLS1CRUdjTiBDRVJUSUZQ0FURS0tLS0tCk1JSURIVENDQWdXZ0F3SUJBZ0iT3dvMC9va09CUEN5RjFWefJxVGhKRW9ubDBVd0RRWuPb1pJaHZjTkFRRUwKQIFBd0hqRWNNQm9HQTFVRUF3d1RTVUpOSUV0c2lzVmJRVJ0ZEEdGaVIYTrmxjekFIRncwe1EQTRNRFF3TWpVMwpNalpHnrcwe1EQTRNRREi3TWpVM01qWmFNQjR4SERBYUJnTlZCQU1NRTBsQ1RTQkRiRzkwNCRVYUmhZbuZ6ClpYTxdnZ0VpTUewR0NTcUdTSWizRFFFQkFRVUFBNIECRHBDzdnRUTb0lCQVFEb0ZFNQGQSGd0eXZMUVlwR3gKQTBoamRXQnM4NVBJTDNyRStjN1R3K2diRUjQSUxJU0VZv3o4Y1g1TG1XQk0rY1FnOG9VeSsrQXJ3OEoxaxdRZQOpSmIU21clF4WTM0c3BQeGRFVEZkWEhScnJhMGU2VmM4MW42TlJL0ZHSn1Q3hrTG5GMUTFQW9hbhYwaDM2CnhDT0FvcXRwTlFrTzNpMTRGeU0yRDRiajkxckl4RGk4V9XMVpVdVhMNGwzZXLZUVCeTruZmhJv3kySVc3aUMKbGpM23RIN3hZTDVHbvpKOUDsYWrtrNcpNREFQlJyvNRIUlyElodTBRSVRFZHIeSFYUEzGRDBHYzloZaoM29jdnpVZUJ3C9uRHn3OTJNCC82SkdtZWpK0lpdFBTN3Y2a2dIUvhNDIBaUVJNxPQdUvPzNOV9GRopYCMv2aJBz01CQUFHalV6QJNQBHQTFVZEFrnUvdCQIR2RzZ2RU5M0jFVbWZnQ003MmxOcmMzSDl2bURBZkJnTYKSFNNUdEQVdnQIR2RzZ2RU5M0jFVbWZnQ003MmxOcmMzSDl2bURBUEJnTlZluk1CQWY4RUJUQURBUUgvTUewRwpDz3FU0liM0RRRUJDd1V0BQTRJQkFRQTgvdfVnUTZlaTZyWHZndDJ0dudrbkpva1Y5UWNikaTNZbFVFWkNDUyJCVQZ3NnMnVBMldcxHlWTm1mRkhjcH21Vmp0VHRYtmk2NUM2WpDz3FU0liM0RRRUJDd1V0BQTRJQkFRQTgvdfVnUTZlaTZyWHZndDJ0dudrbkpva1Y5UWNikaTNZbFVFWk0050vdBq1ZFRXVXVGdDehJKTXFBZnpYUXlidUV0dwpoW1pV2swTmVXNGk5ZEY4S2dTWWUvaQWFodXVBsIRidxB2R2RPV1UoeEV4bm03aEVRbmZPV2ZITThDd08xNWFCZIRg2s0QopDmUrF4Mlg5U284V3o123MzcnyRkFDQlJyZ0NYeFFDZnZrTZUdVNHNkxFRHJHbmpWaXVSQkpZdW4KT1RxWXROaVBHaHpuTHJrL0Fzam1LMzBxQmFLtmFyNuDQajhqlpNb2RiZ04KLS0tFTkQgQ0VSVEIGSUNBVEUtlS0lQo=",
      "name": "1dd14d0c-1b52-4f63-a606-53ecba28771d"
    },
    "composed": [
      "db2://xg49176:N6ogXjFSeE8XOJFo@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb?"
    ],
    "hostSource": "admin&replicaSet=repset"
  },
  "hosts": [
    {
      "hostname": "2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud",
      "port": 32328
    }
  ],
  "jdbc_url": [
    "jdbc:db2://2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb:user=<userid>;password=<your_password>;sslConnection=true;"
  ]
}
```

____ 2. In the Jupyter Notebook that was just added to the project (**Airline_Delay_Analysis.ipynb**), [A] scroll down until you see the code section under the heading **Step 3a: Connect to a remote, Db2 on Cloud database**. Then, [B] click on this section to make it the current section. Next, copy the **hostname** value you found in the service credential file (**AirlineDelayDB_Credentials.txt**) to the clipboard and paste it over the text “**replace-with-your-hostname**” found in this code section. (*Copy the value beside the **hostname** keyword in the service credential file and do NOT remove the quotation marks!*) Repeat this process for the **port** (“**replace-with-your-port-number**”), **username** (“**replace-with-your-userID**”), and **password** “**replace-with-your-password**” keywords. When you are finished, [C] click on the **Save Notebook** icon to make the changes made permanent. (NOTE: You can use the “hot keys” **command-C** on Mac or **Ctrl-C** on Windows to copy data to the clipboard; you can use the “hot keys” **command-V** on Mac or **Ctrl-V** on Windows to paste the copied data.)



```

IBM Cloud Pak for Data Search in your workspaces Upgrade ? 2628635 - IBM Dallas RS
Projects / Airline Delay Analysis / Airline_Delay_Analysis Not Trusted | Python 3.10 Memory:207.4 MB / 8 GB
File Edit View Insert Cell Kernel Help Run Format Markdown Save Notebook
Step 3a: Connect to a remote, Db2 on Cloud database
In [ ]: # Initialize All User-Specific Connection Variables - Db2 on Cloud Database
# IMPORTANT: UPDATE WITH VALUES FROM YOUR OWN ENVIRONMENT, AS PER LAB INSTRUCTIONS.
#
# Define And Initialize The Appropriate Variables
dBName = "bludb" # The Alias For The Db2 Database
userID = "rmb43696" # The Instance User ID
passWord = "JjZkmAFjr19Z6qki" # The Password For The Instance User ID
hostName = "b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lgde00.databases.appdomain.cloud" # The Host Name
portNum = "32716" # The TCP/IP Port Number That Receives Db2 Connections
secureComm = True # Use SSL (Secure Sockets Layer) Communication
#
# Display A Status Message Indicating This Work Is Complete
print("\nUser-specific connection variable initialization work complete!\n")

```

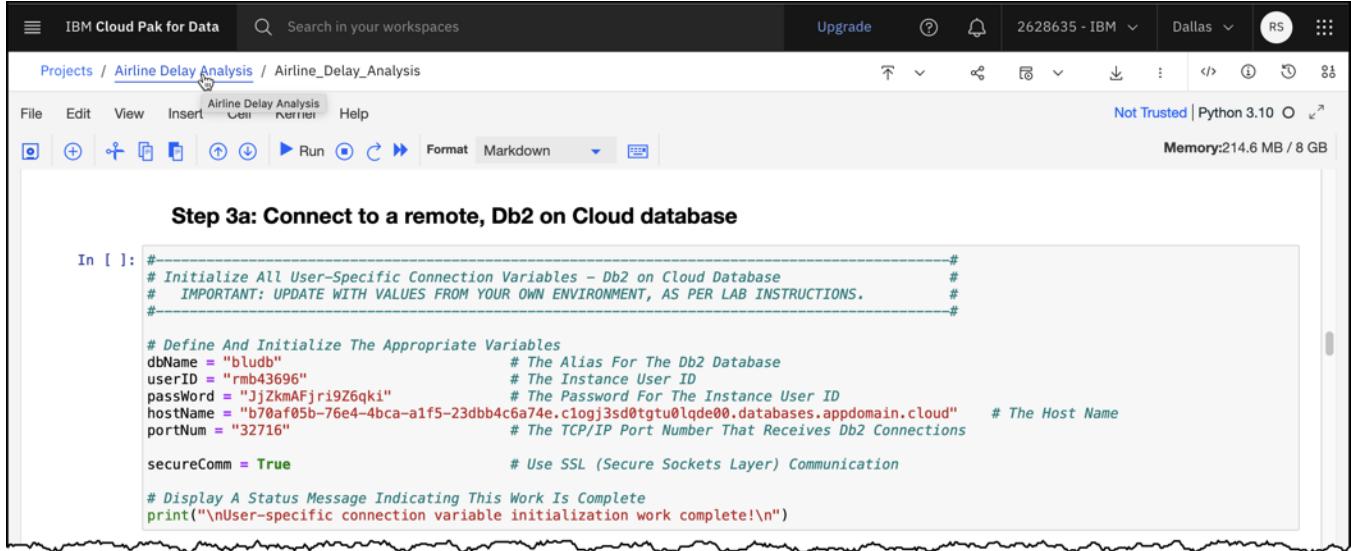


```

Step 3b: Connect to a local, on-premises Db2 database
In [ ]: # Initialize All User-Specific Connection Variables - Local Db2 Database
#
# Define And Initialize The Appropriate Variables
dBName = "replace-with-your-database-name" # The Alias For The Db2 Database
userID = "replace-with-your-userID" # The Instance User ID
passWord = "replace-with-your-password" # The Password For The Instance User ID

```

____ 4. Exit the Jupyter Notebook and return to the project by clicking on the **Airline Delay Analysis** project breadcrumb located at the top of the screen. This will cause the Assets screen for the project to be displayed.



The screenshot shows the IBM Cloud Pak for Data Jupyter Notebook interface. The title bar indicates the project is "Airline Delay Analysis". The notebook cell contains Python code for connecting to a Db2 on Cloud database, including variables for alias, user ID, password, host name, port number, and secure communication settings. A status message is printed at the end of the code.

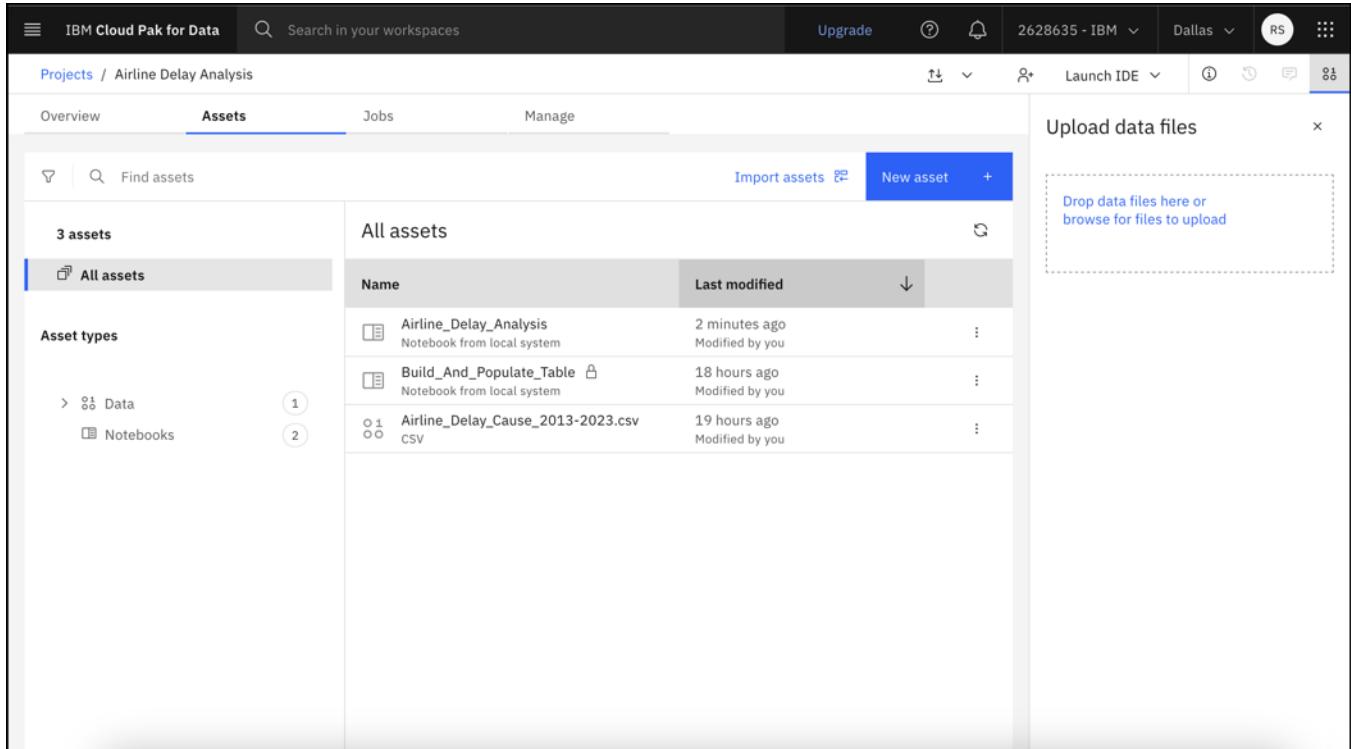
```
In [ ]: # Initialize All User-Specific Connection Variables - Db2 on Cloud Database
# IMPORTANT: UPDATE WITH VALUES FROM YOUR OWN ENVIRONMENT, AS PER LAB INSTRUCTIONS.

# Define And Initialize The Appropriate Variables
dBName = "bludb"                                # The Alias For The Db2 Database
userID = "rbm43696"                               # The Instance User ID
password = "JjZkmAFjri9Z6qki"                     # The Password For The Instance User ID
hostName = "b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.clogj3s0tgtu0lade00.databases.apppdomain.cloud"    # The Host Name
portNum = "32716"                                  # The TCP/IP Port Number That Receives Db2 Connections

secureComm = True                                    # Use SSL (Secure Sockets Layer) Communication

# Display A Status Message Indicating This Work Is Complete
print("\nUser-specific connection variable initialization work complete!\n")
```

____ 5. When the Assets screen for the **Airline Delay Analysis** project is displayed again, the **Airline_Delay_Analysis** Jupyter Notebook should be in the project assets list.



The screenshot shows the IBM Cloud Pak for Data Assets screen for the "Airline Delay Analysis" project. The "Assets" tab is selected. On the left, there's a sidebar for "Asset types" with "Data" and "Notebooks" sections. The main area lists "All assets" with three items: "Airline_Delay_Analysis", "Build_And_Populate_Table", and "Airline_Delay_Cause_2013-2023.csv". A "New asset" button is visible at the top right of the asset list table. To the right, there's a "Upload data files" section with a placeholder for file uploads.

Name	Last modified	Actions
Airline_Delay_Analysis	2 minutes ago Modified by you	⋮
Build_And_Populate_Table	18 hours ago Modified by you	⋮
Airline_Delay_Cause_2013-2023.csv	19 hours ago Modified by you	⋮

Congratulations! You are now ready to begin working your way through the **Airline_Delay_Analysis** Jupyter Notebook.

3. Run the Airline_Delay_Analysis Jupyter Notebook

1. If you are not already there, navigate to the **Assets** screen for the **Airline Delay Analysis** project, and click on the link for the Jupyter Notebook asset named **Airline_Delay_Analysis**.

The screenshot shows the 'Assets' tab selected in the top navigation bar of the IBM Cloud Pak for Data interface. On the left, there's a sidebar with '3 assets' and categories for 'Data' (1 item) and 'Notebooks' (2 items). The main area displays a table titled 'All assets' with columns for 'Name' and 'Last modified'. Three items are listed:

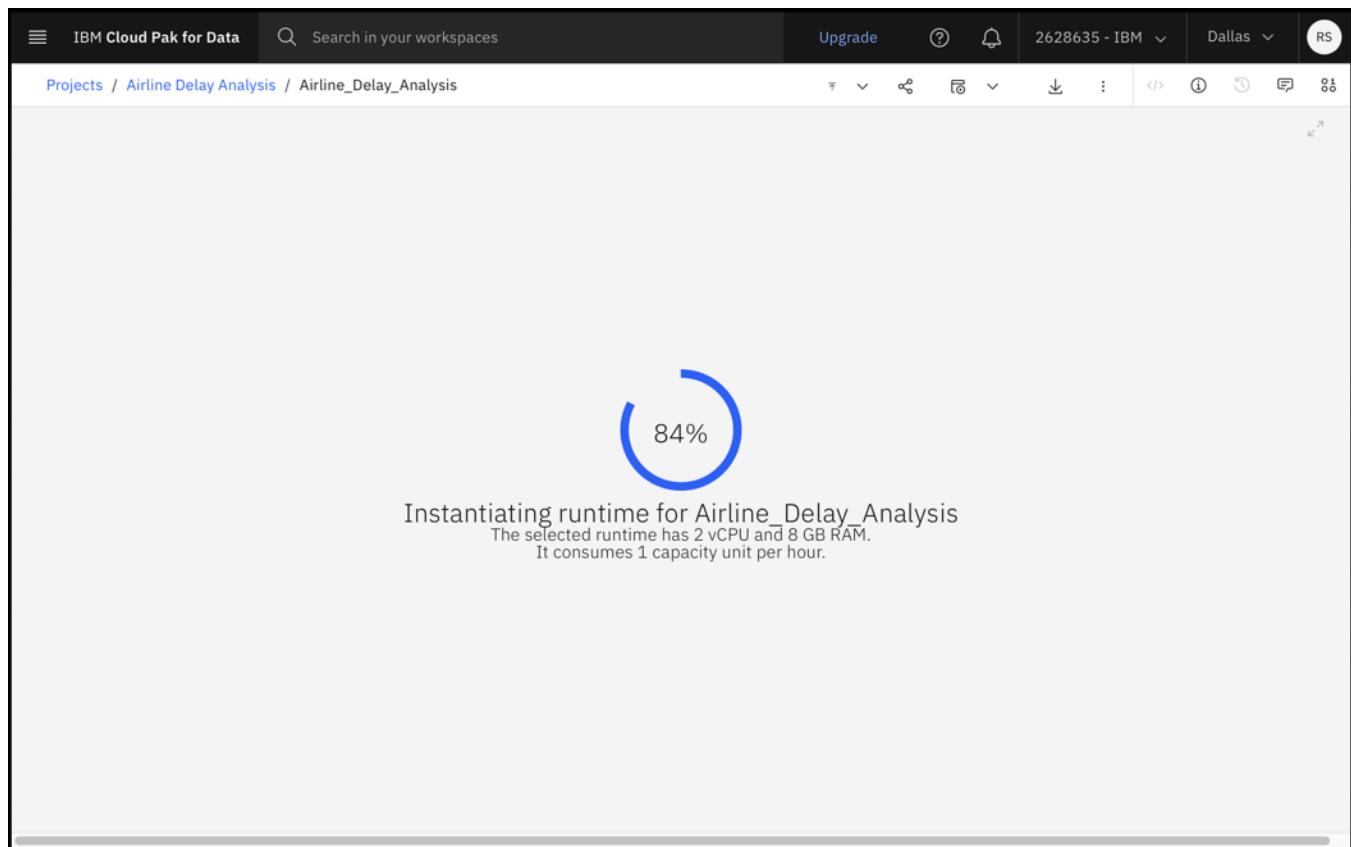
Name	Last modified
Airline_Delay_Analysis Notebook from local system	2 minutes ago Modified by you
Build_And_Populate_Table Notebook from local system	18 hours ago Modified by you
Airline_Delay_Cause_2013-2023.csv CSV	19 hours ago Modified by you

An 'Upload data files' panel is open on the right, with a message 'Drop data files here or browse for files to upload'.

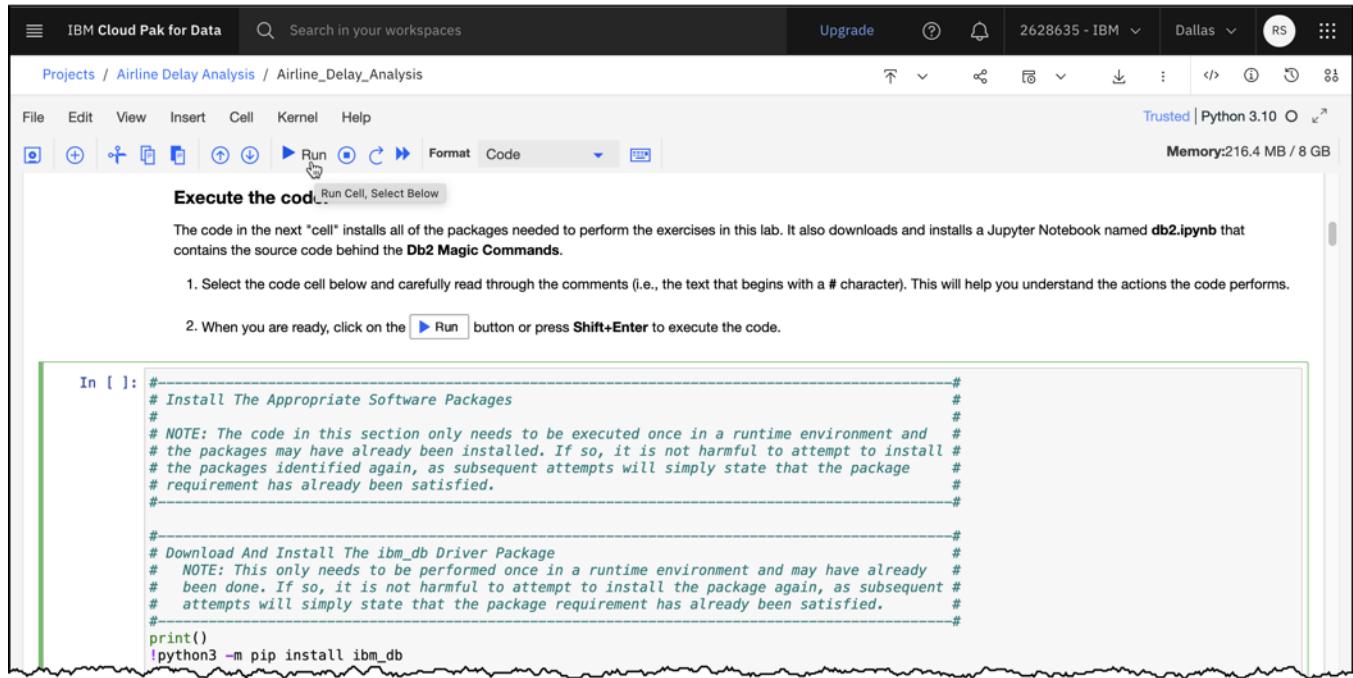
2. When the Jupyter Notebook opens, you may notice that the Jupyter Notebook menu and icons are not present. That's because the notebook has not been instantiated (i.e., a kernel for the Jupyter Notebook has not been started). Instantiate the **Airline_Delay_Analysis** Jupyter Notebook by clicking on the **Edit** (pencil) icon located at the top of the screen.

The screenshot shows the Jupyter Notebook interface for the 'Airline_Delay_Analysis' notebook. At the top, the title 'Airline_Delay_Analysis' is visible, and the 'Edit' button is highlighted with a cursor. The notebook content includes a title 'IBM Db2', a subtitle 'Jupyter Notebook hands-on exercise', author information 'Roger E. Sanders Principal, Learning Content Development - Data, AI, & Storage roger.sanders@us.ibm.com', and a decorative graphic featuring a keyboard, a pencil, and a 3D chart.

_____ 3. While the Jupyter Notebook is being instantiated, you should see a screen that looks something like this:



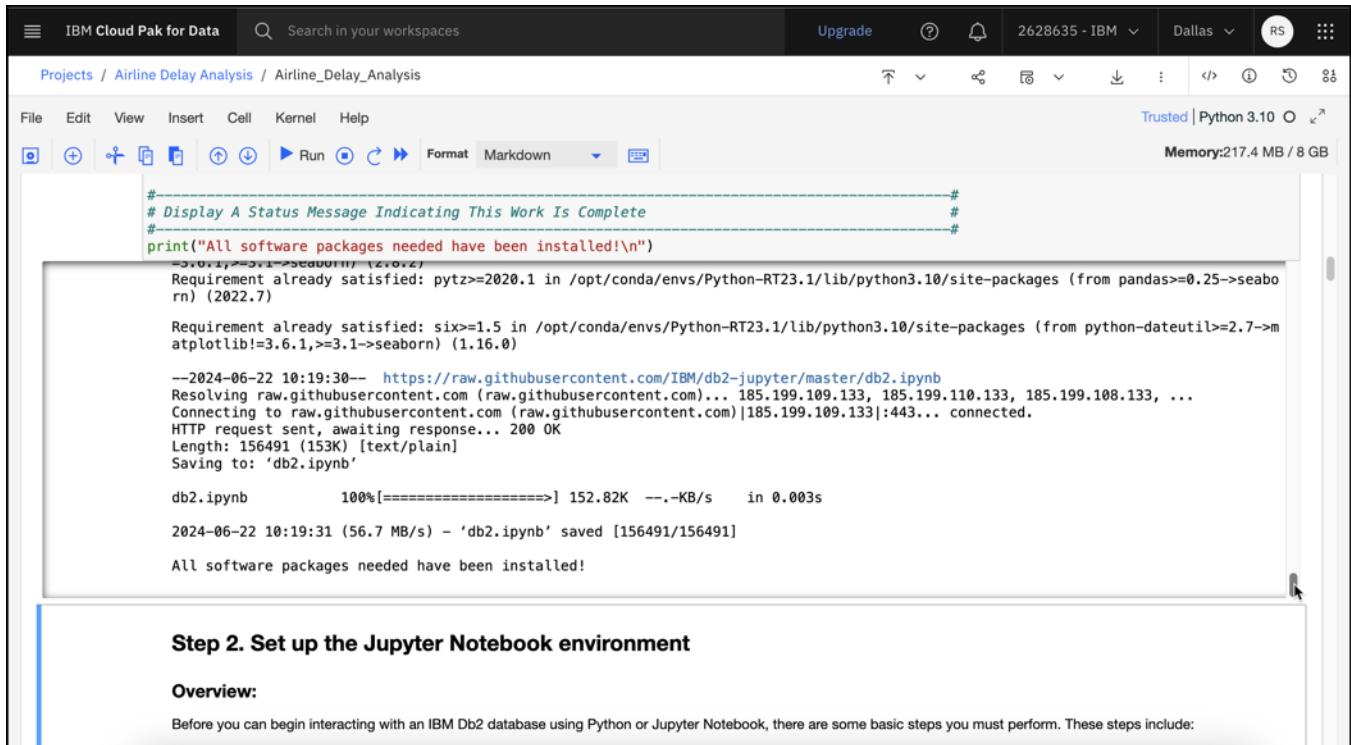
4. Once the **Airline_Delay_Analysis** Jupyter Notebook has been instantiated, work your way through it, one cell at a time, reading the documentation found in markup cells and executing the source code found in code cells. The notebook is well documented and easy to work through. As a reminder, to execute a code cell, simply click on the cell to select it and then click on the ► Run icon.



The screenshot shows the IBM Cloud Pak for Data Jupyter Notebook interface. The top navigation bar includes 'IBM Cloud Pak for Data', a search bar, 'Upgrade', a help icon, '2628635 - IBM', 'Dallas', and a user profile icon. The main area displays a code cell with the following content:

```
In [ ]: #  
# Install The Appropriate Software Packages  
#  
# NOTE: The code in this section only needs to be executed once in a runtime environment and #  
# the packages may have already been installed. If so, it is not harmful to attempt to install #  
# the packages identified again, as subsequent attempts will simply state that the package #  
# requirement has already been satisfied.  
#-----#  
# Download And Install The ibm_db Driver Package  
# NOTE: This only needs to be performed once in a runtime environment and may have already #  
# been done. If so, it is not harmful to attempt to install the package again, as subsequent #  
# attempts will simply state that the package requirement has already been satisfied.  
#-----#  
print()  
!python3 -m pip install ibm_db
```

____ 5. You will find that with this Jupyter Notebook, some of the code cells generate a significant amount of output. When that is the case, use the scroll bars that are provided with the output window to view all the information produced.



The screenshot shows a Jupyter Notebook interface within the IBM Cloud Pak for Data workspace. The notebook title is "Airline Delay Analysis". The code cell contains a single line of Python code:

```
print("All software packages needed have been installed!")
```

. The output pane displays the results of the code execution, including dependency resolution logs and the final message "All software packages needed have been installed!".

```
#-----#
# Display A Status Message Indicating This Work Is Complete #
#-----#
print("All software packages needed have been installed!")

Requirement already satisfied: pytz>=2020.1 in /opt/conda/envs/Python-RT23.1/lib/python3.10/site-packages (from pandas>=0.25->seaborn) (2022.7)

Requirement already satisfied: six>=1.5 in /opt/conda/envs/Python-RT23.1/lib/python3.10/site-packages (from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.1->seaborn) (1.16.0)

--2024-06-22 10:19:30-- https://raw.githubusercontent.com/IBM/db2-jupyter/master/db2.ipynb
Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.189.133, 185.199.110.133, 185.199.108.133, ...
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|185.199.109.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 156491 (153K) [text/plain]
Saving to: 'db2.ipynb'

db2.ipynb      100%[=====] 152.82K  --.-KB/s   in 0.003s

2024-06-22 10:19:31 (56.7 MB/s) - 'db2.ipynb' saved [156491/156491]

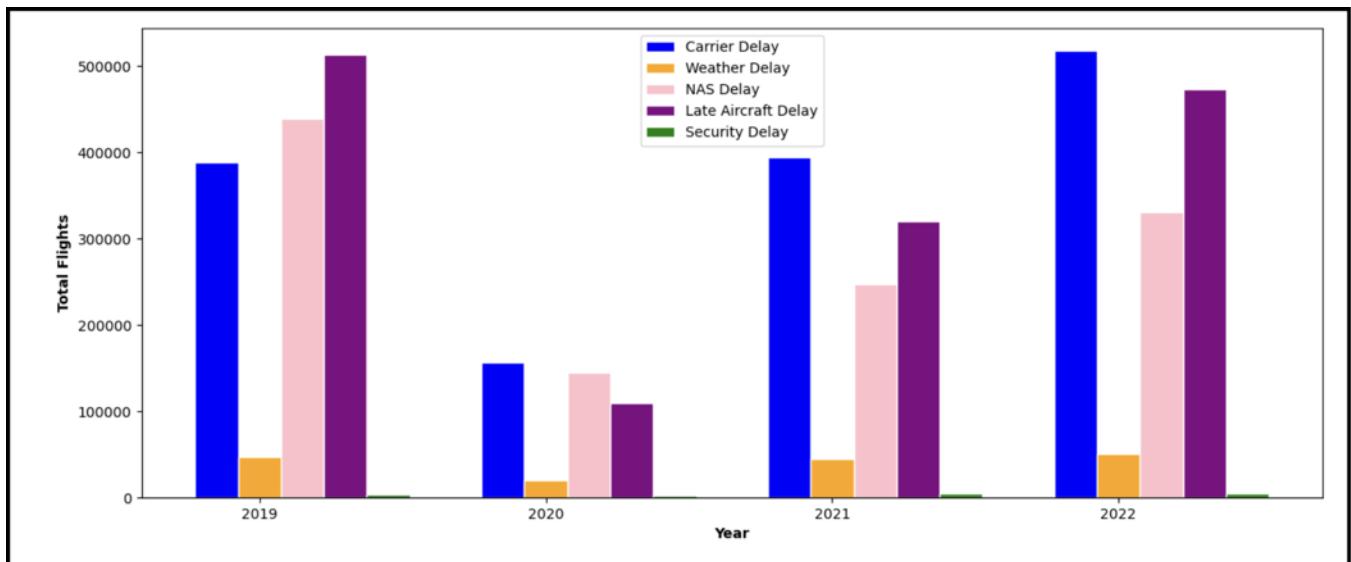
All software packages needed have been installed!
```

Step 2. Set up the Jupyter Notebook environment

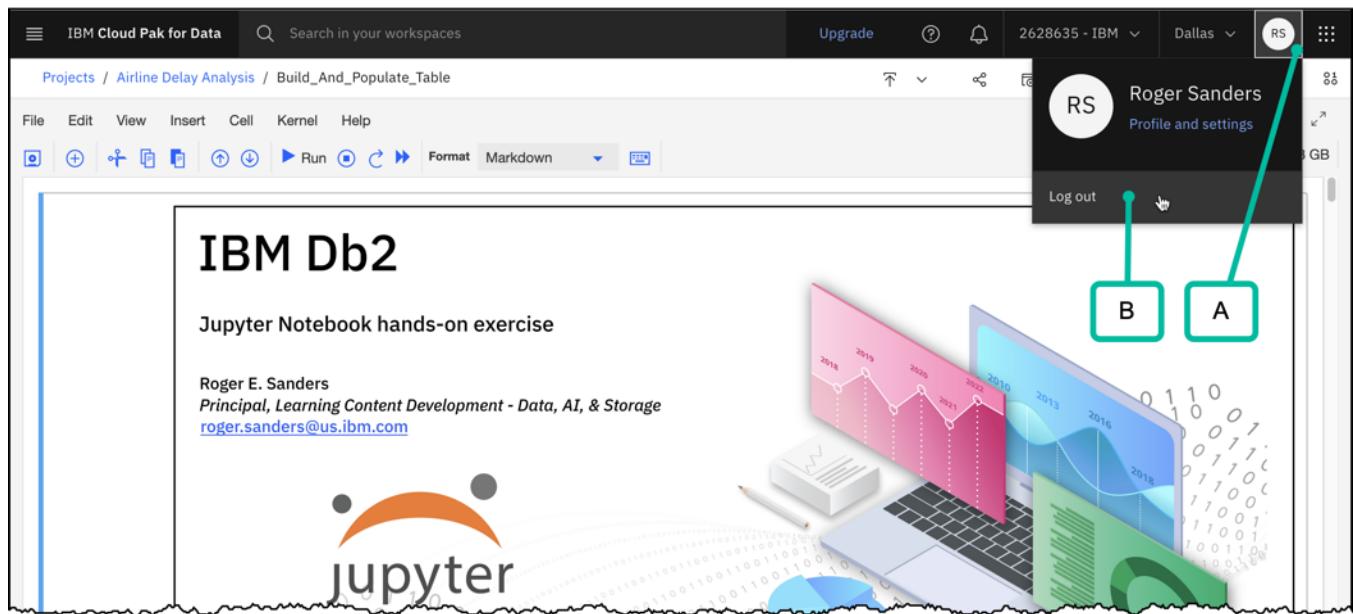
Overview:

Before you can begin interacting with an IBM Db2 database using Python or Jupyter Notebook, there are some basic steps you must perform. These steps include:

____ 6. Pay close attention to the output produced by the code cells that generate graphs. These graphs are designed to help you visualize the data analysis that is being performed.



7. When you are finished, log out of IBM Cloud Pak for Data as a Service by [A] clicking on your user icon located in the top right-hand corner of the screen and then [B] clicking on the **Log out** item in the menu presented.



Congratulations! You have just used Watson Studio (running in IBM Cloud Pak for Data as a Service) and a Jupyter Notebook to analyze data stored in a Db2 on Cloud database.

VI. Thank you!

Thank you for participating in this workshop! I hope you found it both enjoyable and informative.