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

© Copyright IBM Corporation, 2024

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

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.

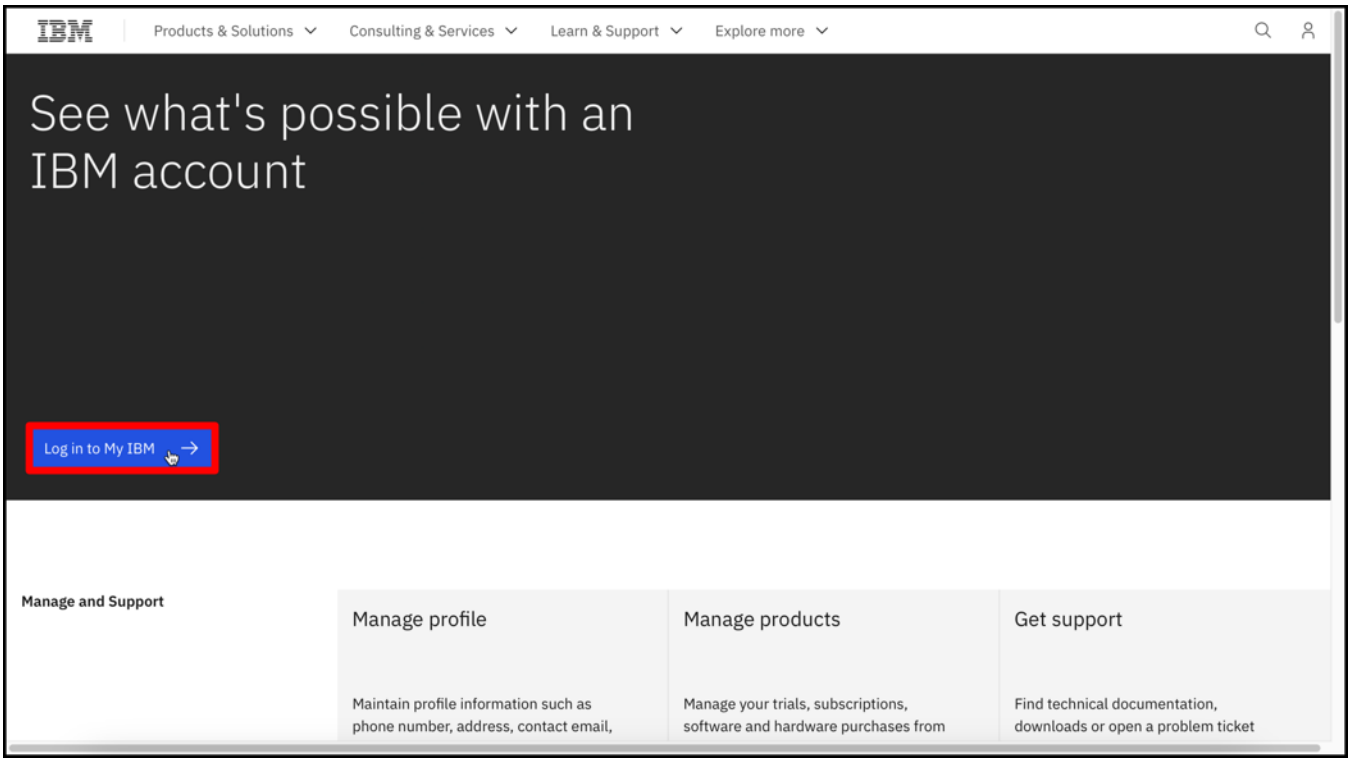
II. Provisioning a Db2 on Cloud Lite plan database on the IBM Cloud

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.

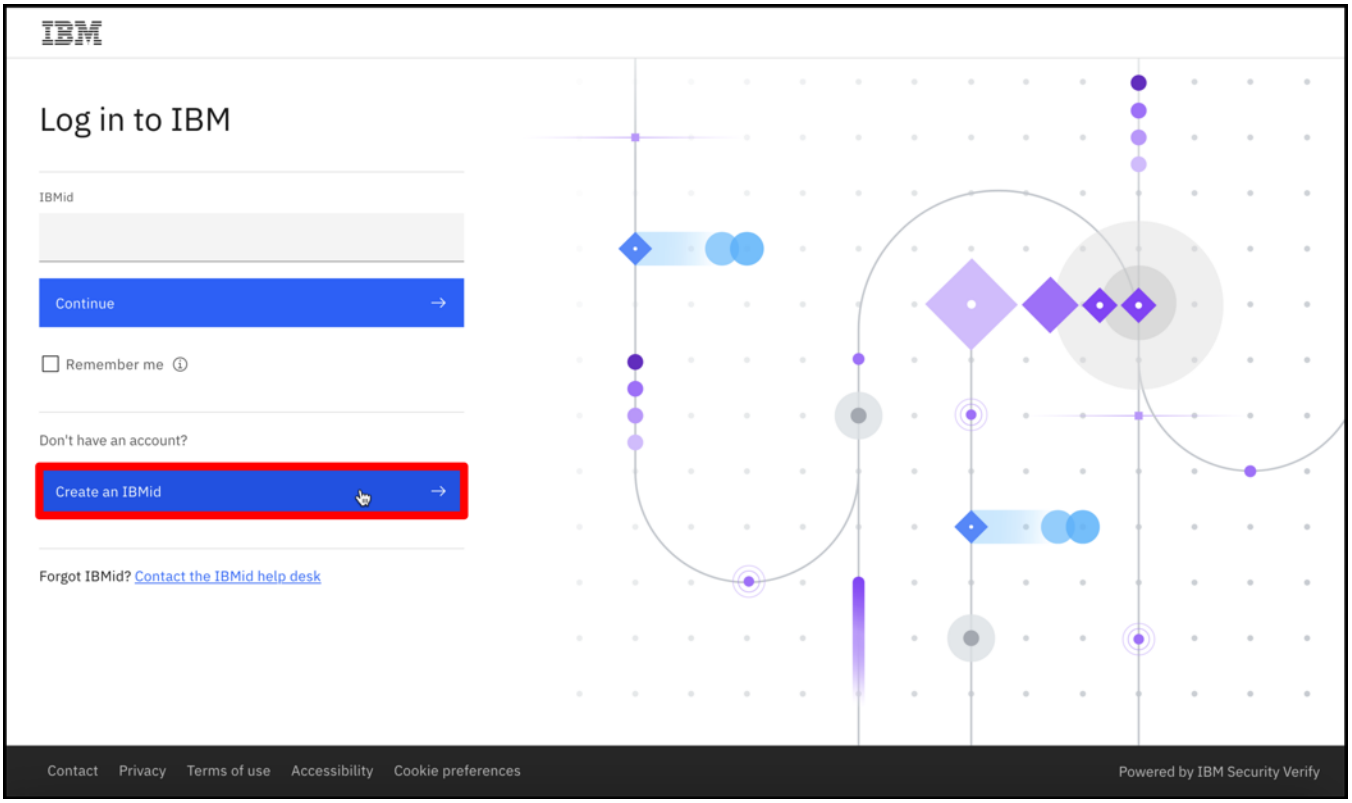
To provision a Db2 on Cloud "Lite" plan database, you must have an IBMid and an IBM Cloud account, both of which can be created free of charge. ***If you already have an IBMid and an IBM Cloud account, proceed directly to Step 3.***

1. Create an IBM account.

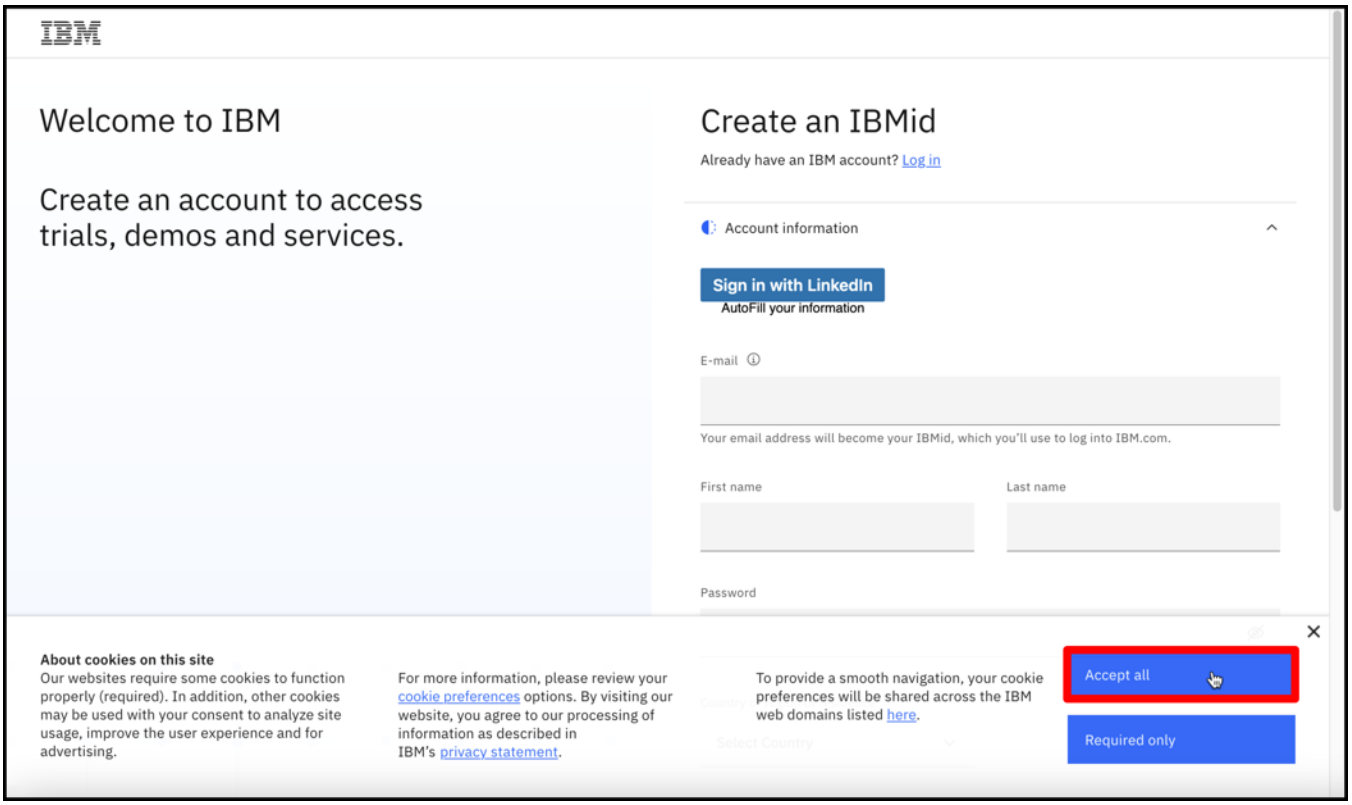
____ 1. Open a web browser and go to the *IBM Account* web site (https://www.ibm.com/account/us-en/). Once there, select the **Log in to My IBM** button.



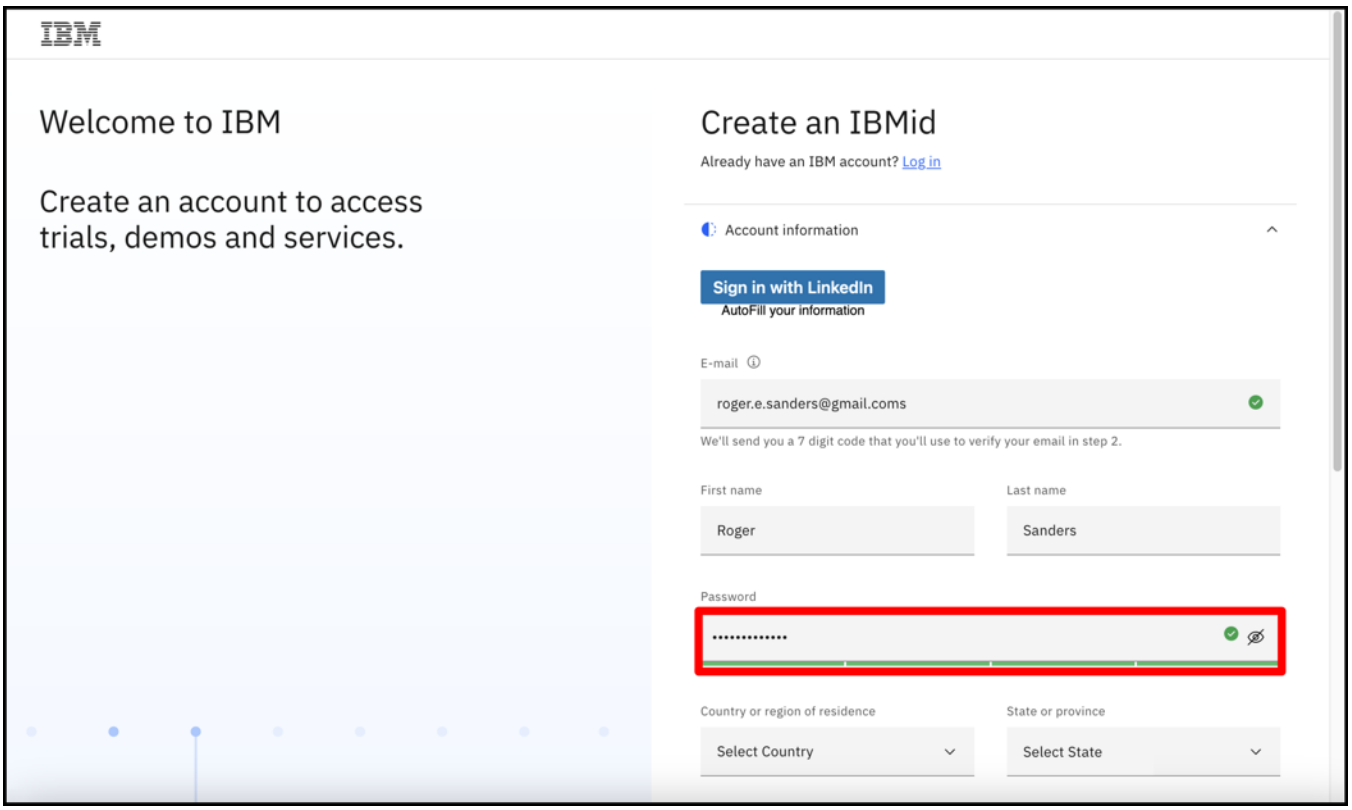
____ 2. When the *Log in to IBM* window is displayed, locate the **Create an IBMid** button and click on it.



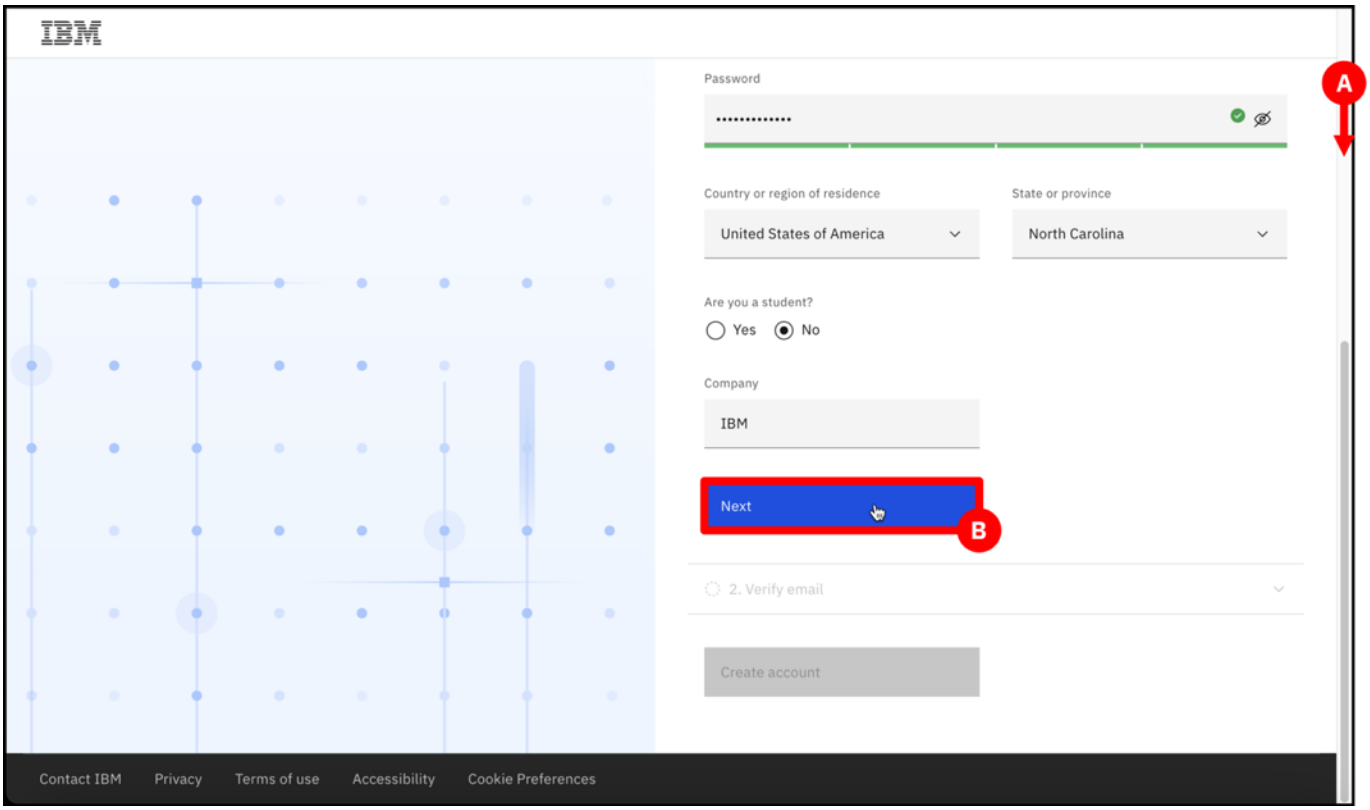
____ 3. When the *Create an IBM account* screen is displayed, you may be asked to indicate how you want to let IBM use cookies. If that is the case, click on the **Accept all** or the **Required only** button shown in the popup window displayed at the bottom of the screen. (In this example, the **Accept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



____ 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.

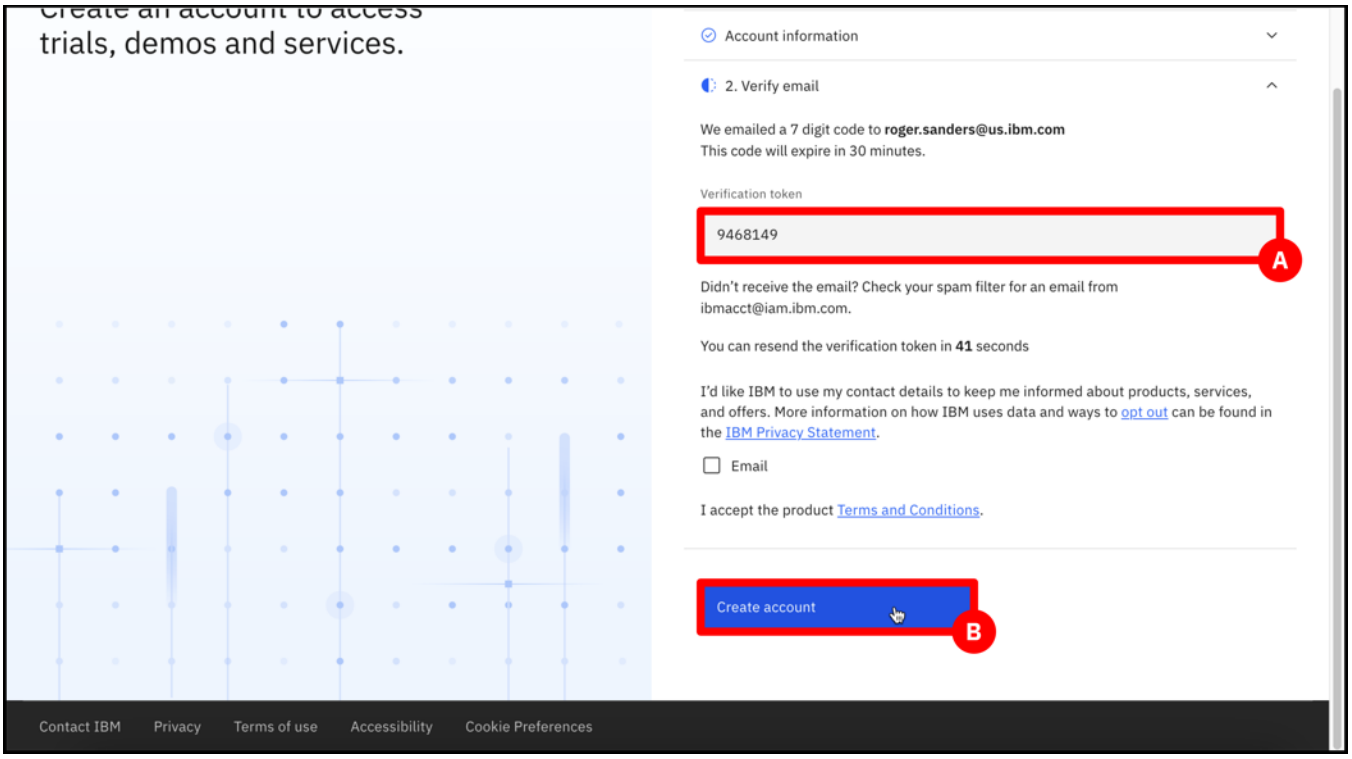


____ 5. **(A)** Scroll down and provide the appropriate information in the **Country or region of residence**, **State or province**, and **Company** fields. Also indicate whether you are a student by selecting the appropriate radio button (**Yes** or **No**) under the **Are you a student?** heading.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

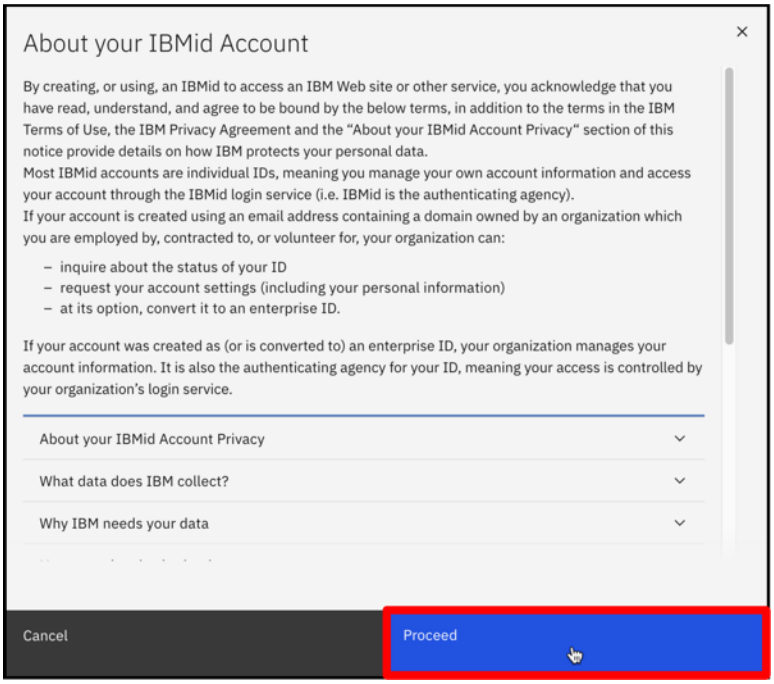


This will cause the **Account information** section of the screen to be collapsed, and the **Verify email** section to be expanded. A 7-digit confirmation code will then be sent to the email address you provided. (The email will come from *IBM Security* and the subject will be *New User Registration*.) 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.



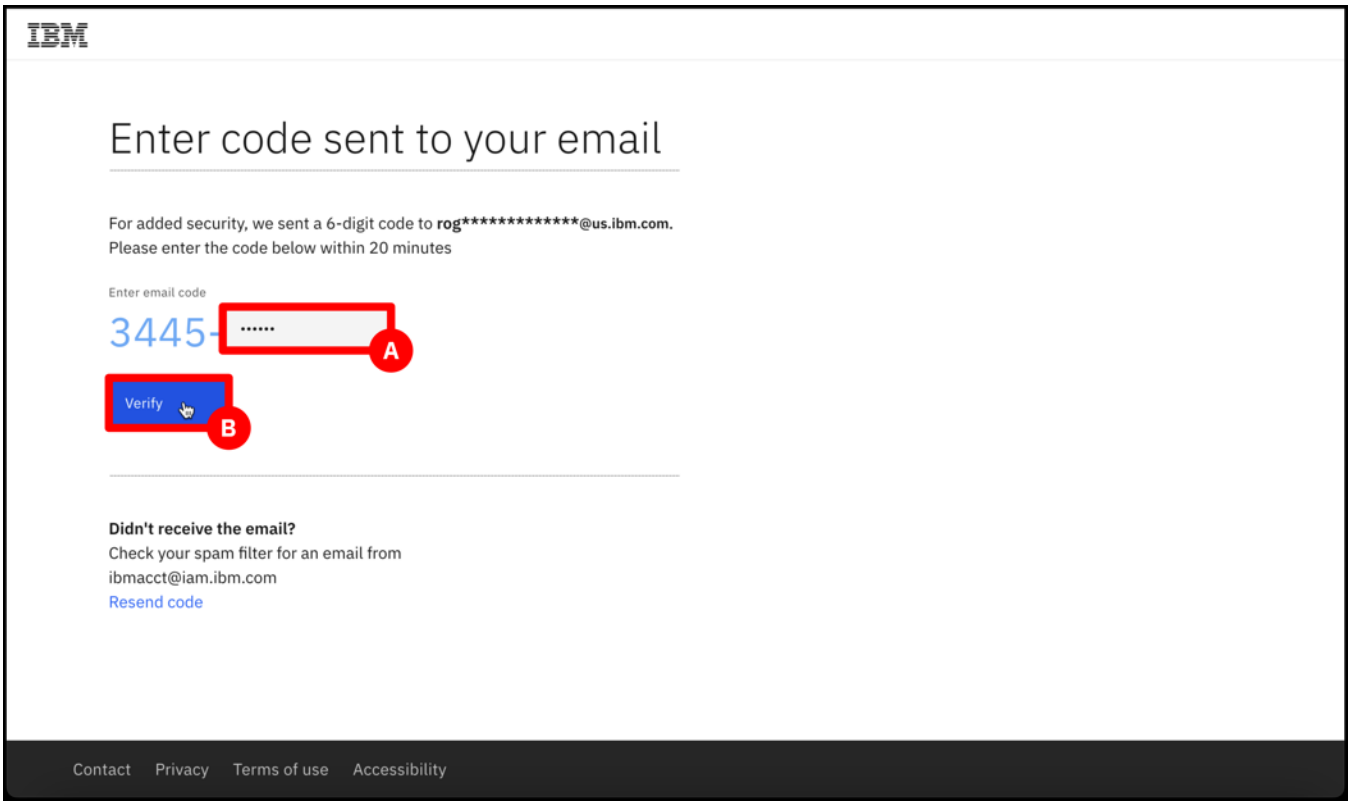
____ 7. When the *About your IBMid Account* popup appears, read through the terms and conditions of your IBM account. Then, when you are ready to continue, click the **Proceed** button.



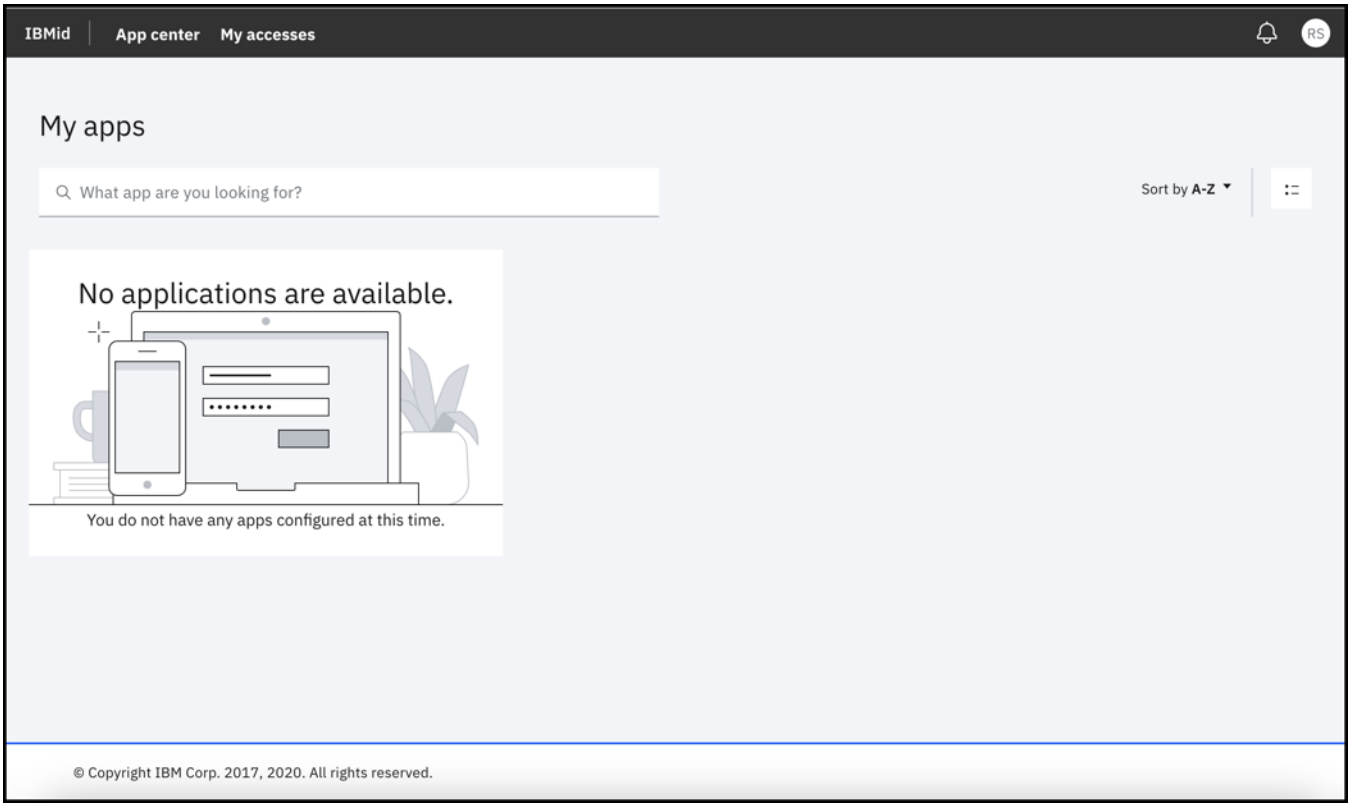
Once this is done, you may see the following message on the screen: "**Please wait while we create your IBMid and set up the service subscription.**"

____ 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.



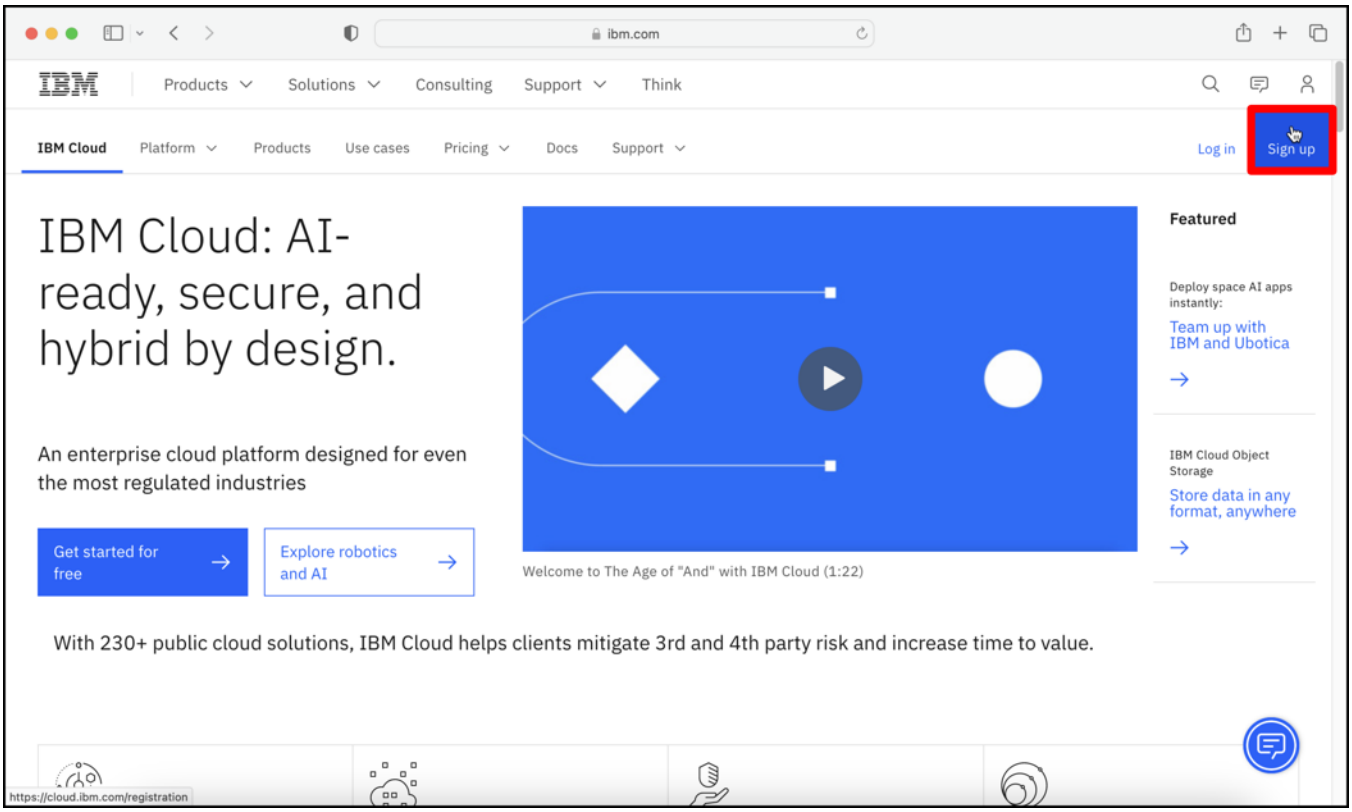
Once the email code you provided has been verified, you should see a screen that looks something like this:



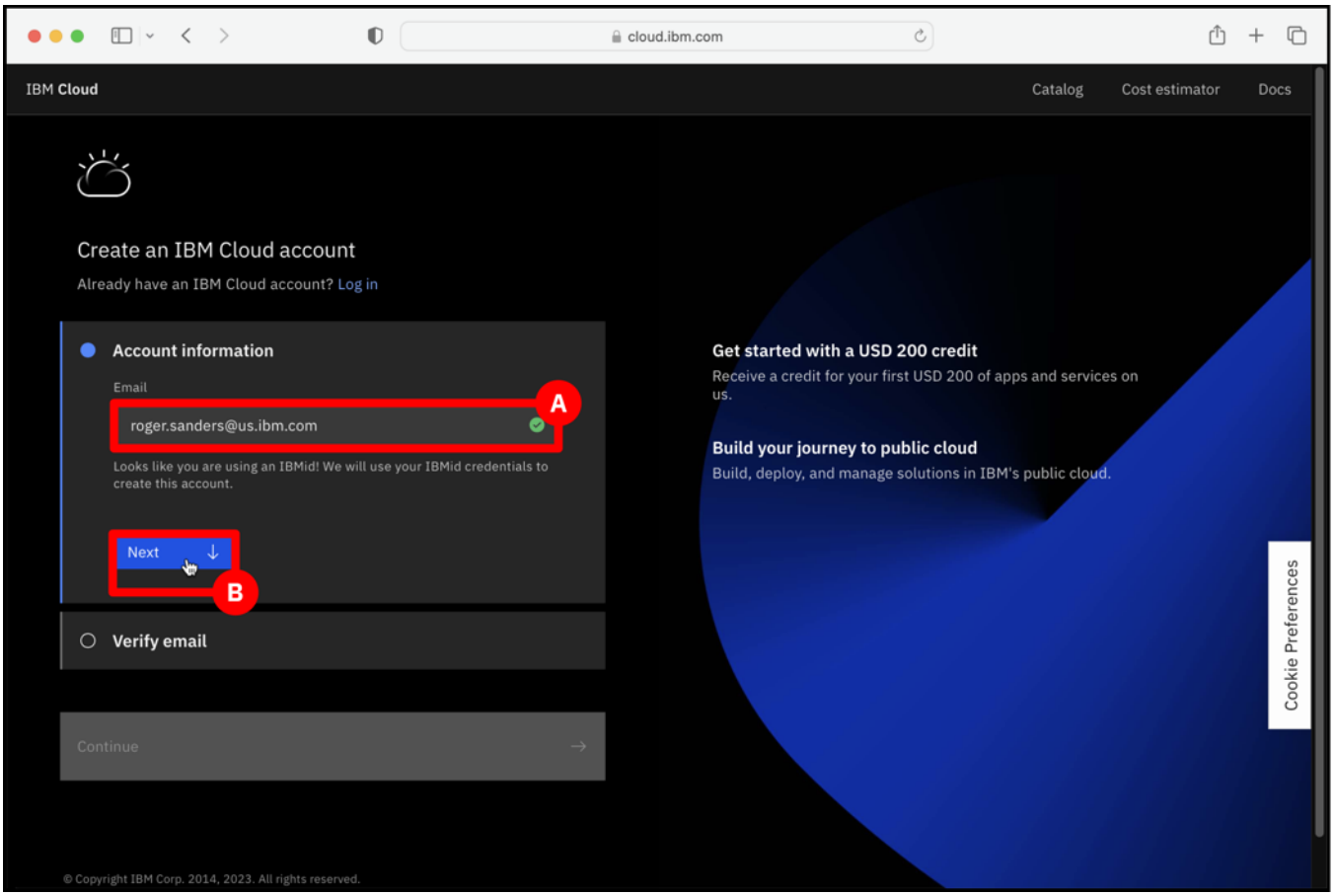
Congratulations! You now have an **IBMid**, which you will now use to create an IBM Cloud account (and later, to create an IBM Cloud Pak for Data as a Service account).

2. Create an IBM Cloud account.

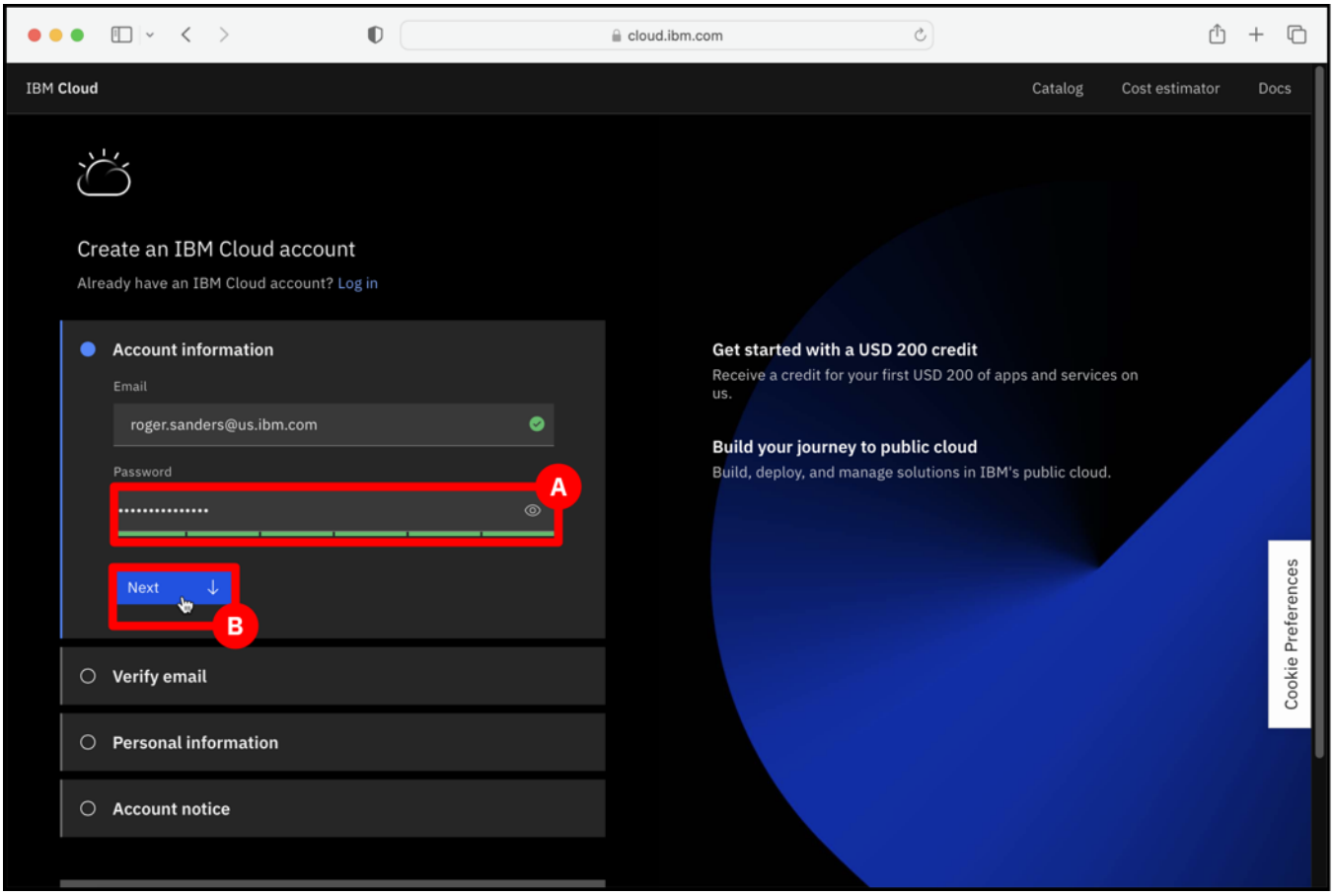
____ 1. Open a web browser and go to the *IBM Cloud* web site (https://www.ibm.com/cloud). Once there, select the **Sign up** button located in the top right corner of the screen.



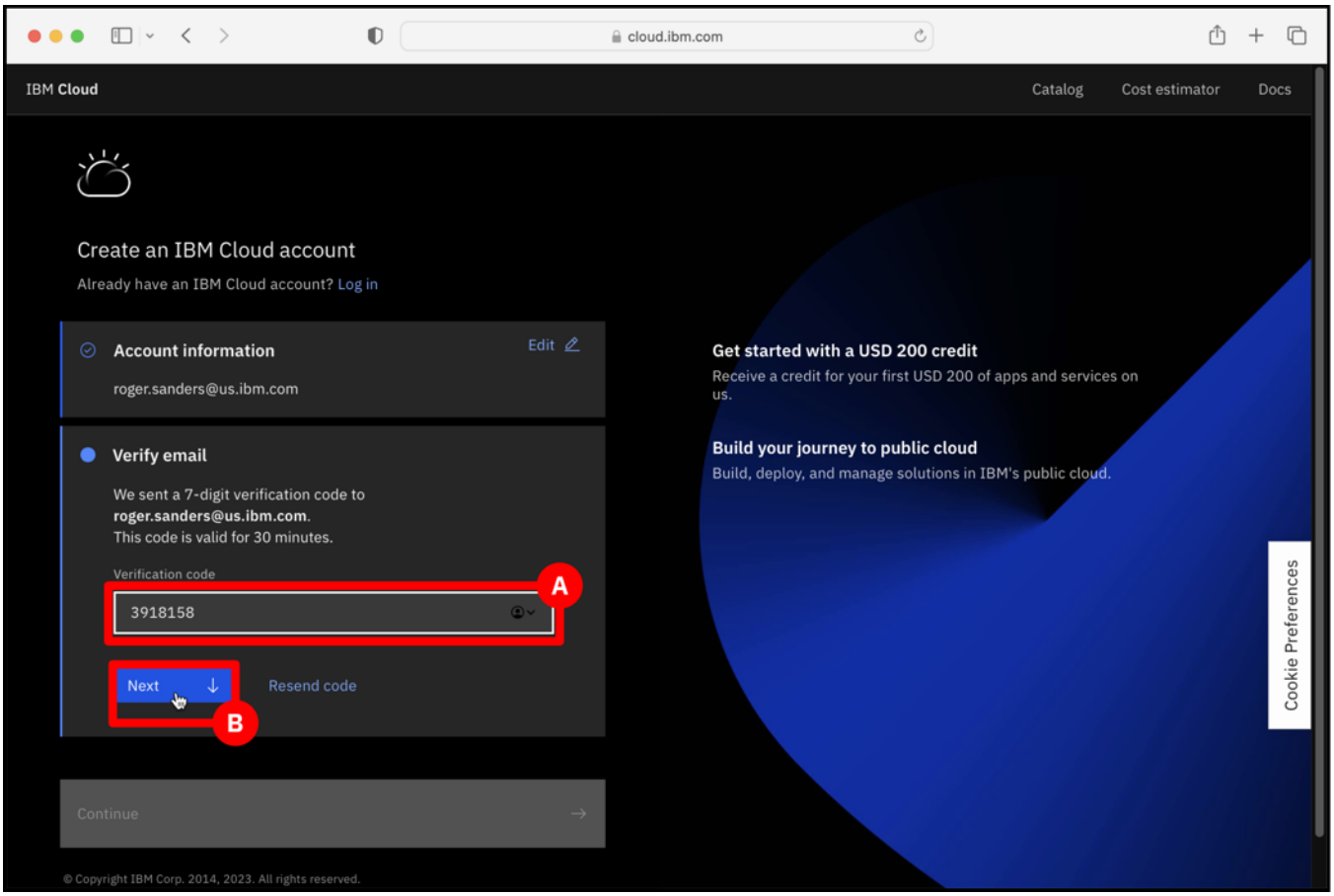
____ 2. When the *Create an IBM Cloud account* screen is displayed, **(A)** enter the email address you used to create your IBMid in the **Email** field (which is located inside the **Account information** box). **(B)** Then, click the **Next** button.



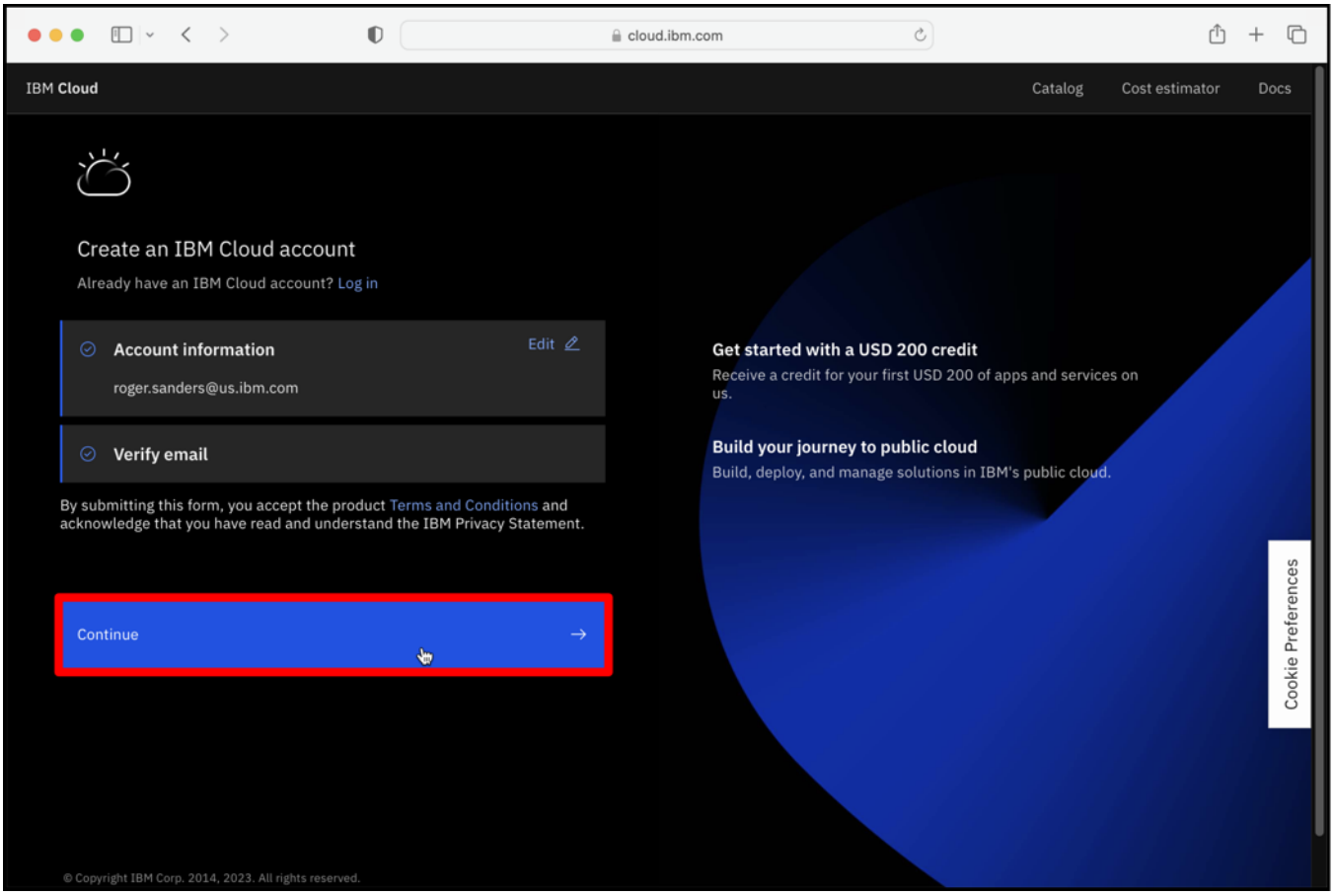
____ 3. When the **Password** field appears, **(A)** enter the password associated with the email address (IBMid) provided in the previous step and **(B)** click the **Next** button. This will cause a 7-digit verification code to be sent to the email address specified. (The email will come from *IBM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



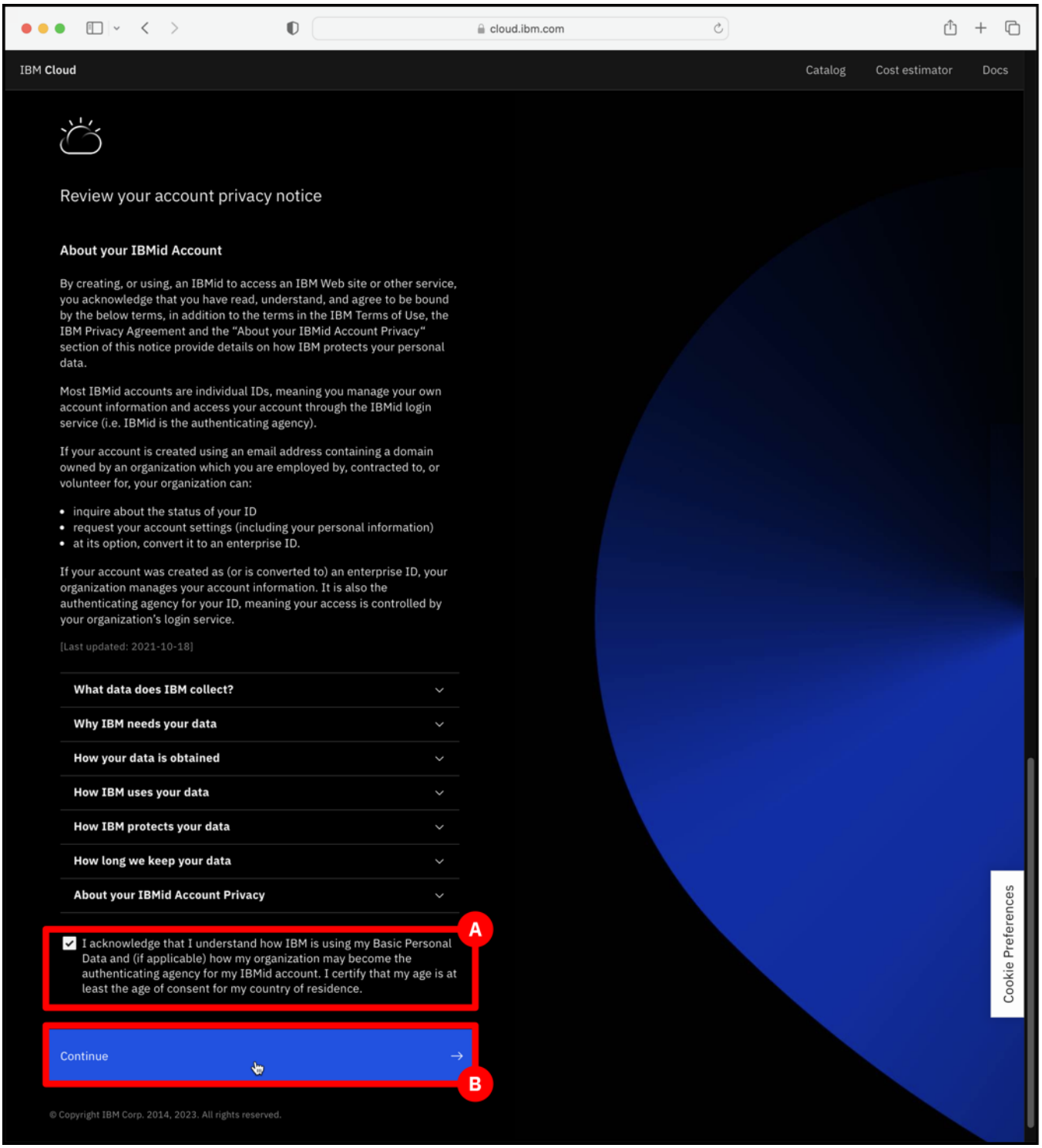
____ 4. Once you receive an email containing a verification code, **(A)** enter the code in the **Verification code** field and **(B)** click the **Next** button. *(Note: The verification code is only valid for 30 minutes. If, after 5 minutes, you don't see an email containing the code in your inbox,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

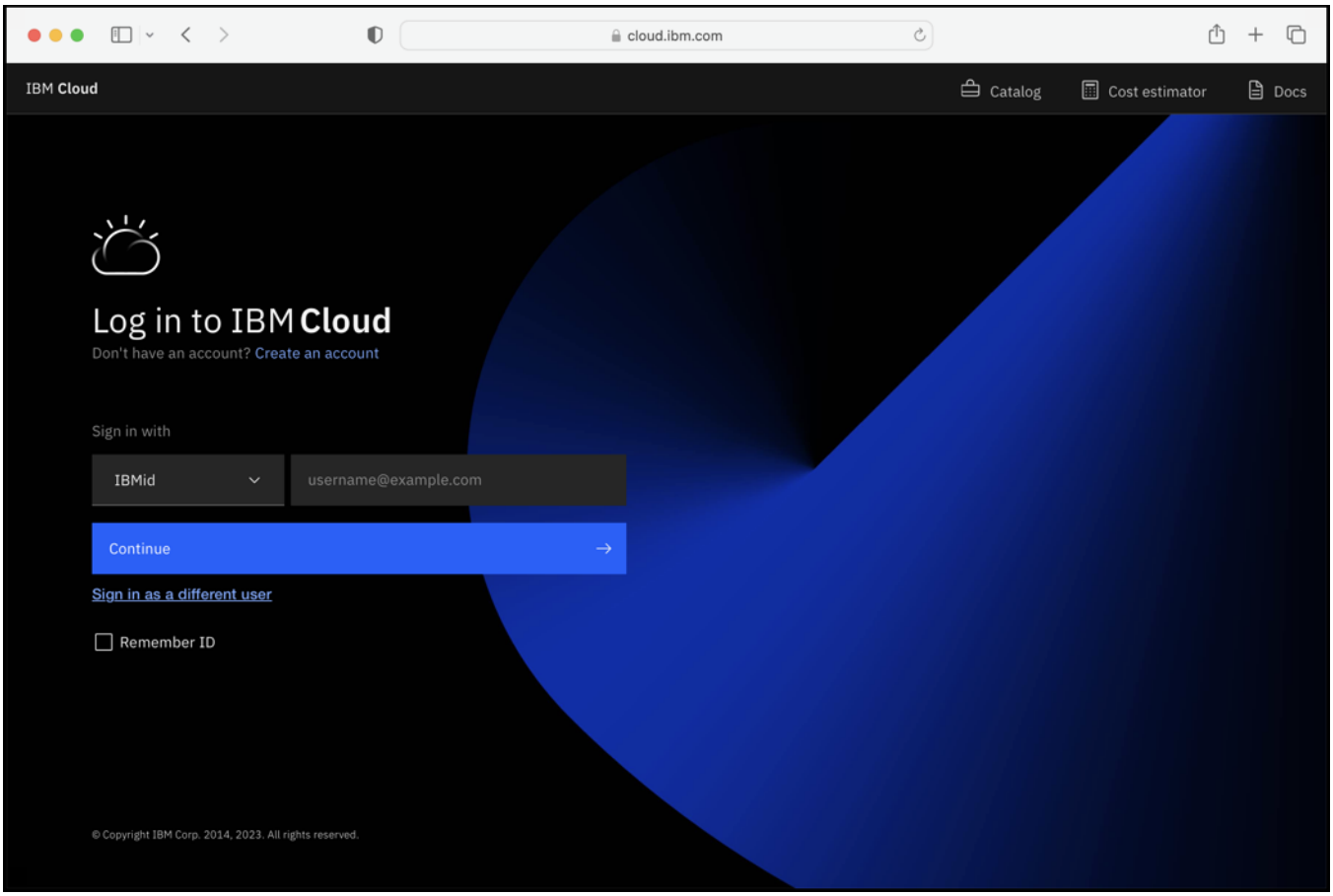


____ 6. When the *Review your account privacy notice* screen is displayed, read through the terms and conditions of your IBMid Account and then **(A)** click the **I acknowledge that I understand how IBM is using my Basic Personal Data ...** checkbox. **(B)** Click the **Continue** button to continue.

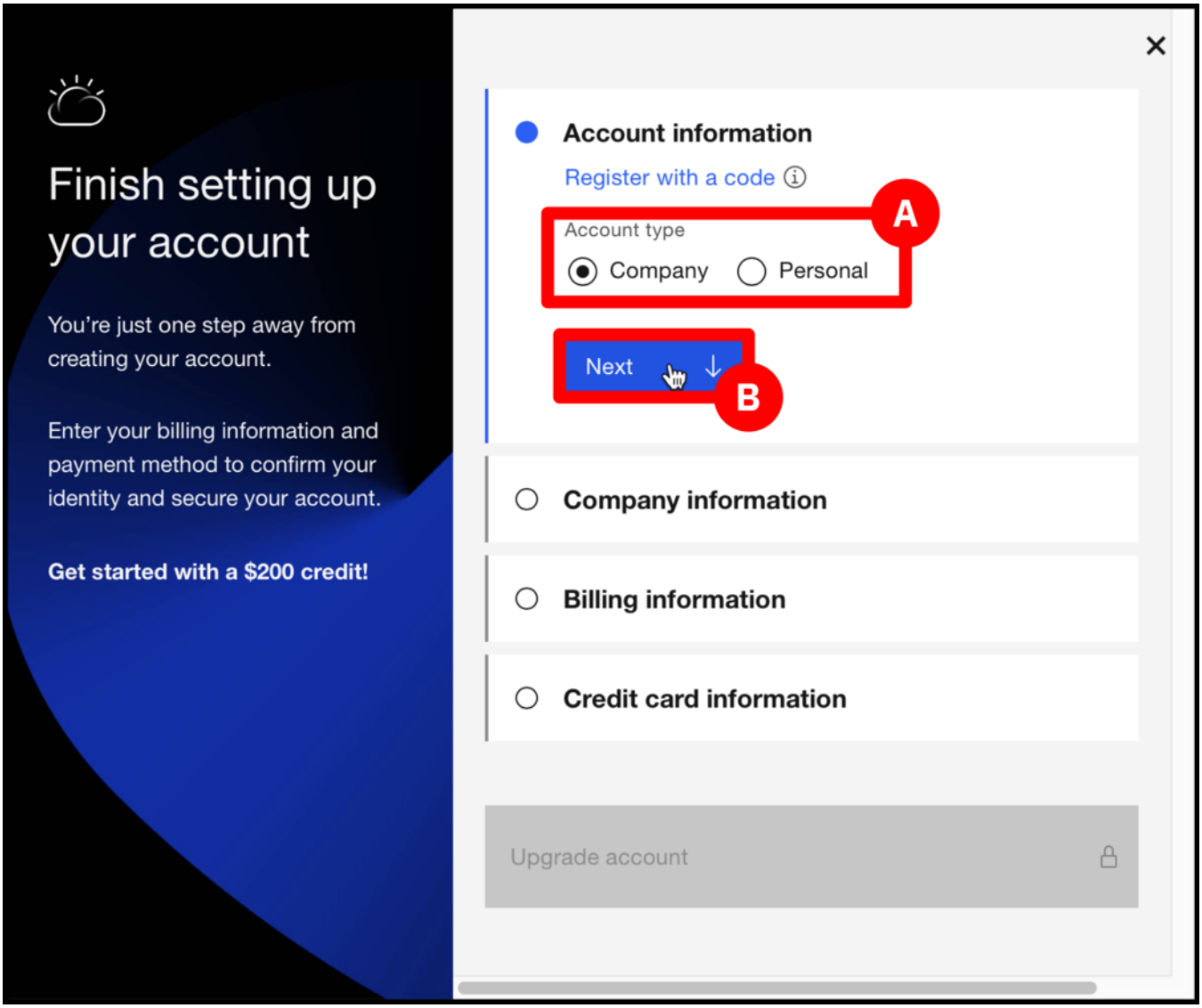


The **Continue** button should be replaced with the message "*Creating your account ...*"

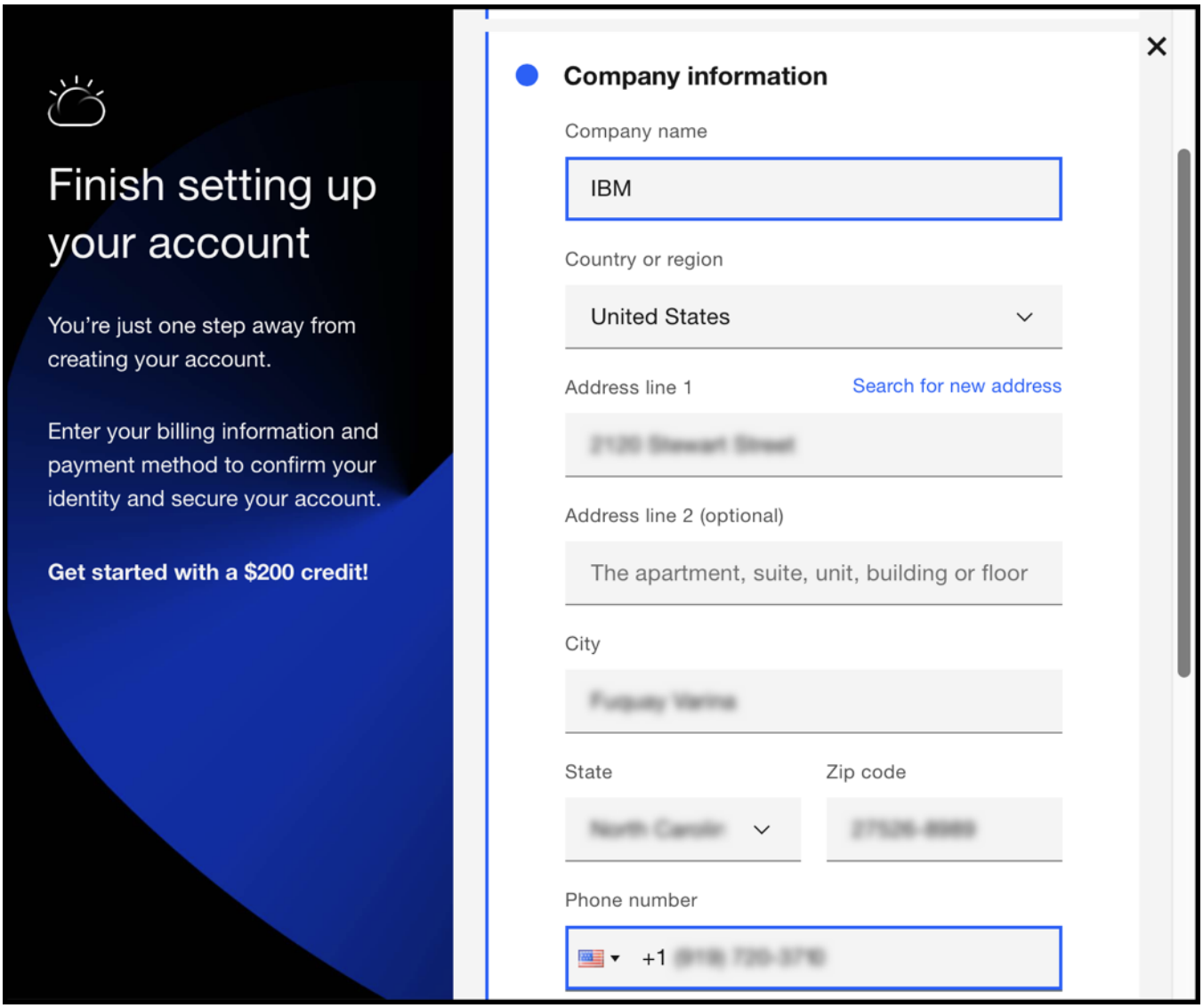
____ 7. Once your IBM Cloud account has been created, you will be presented with a *Log in to the IBM Cloud* screen that looks something like this:



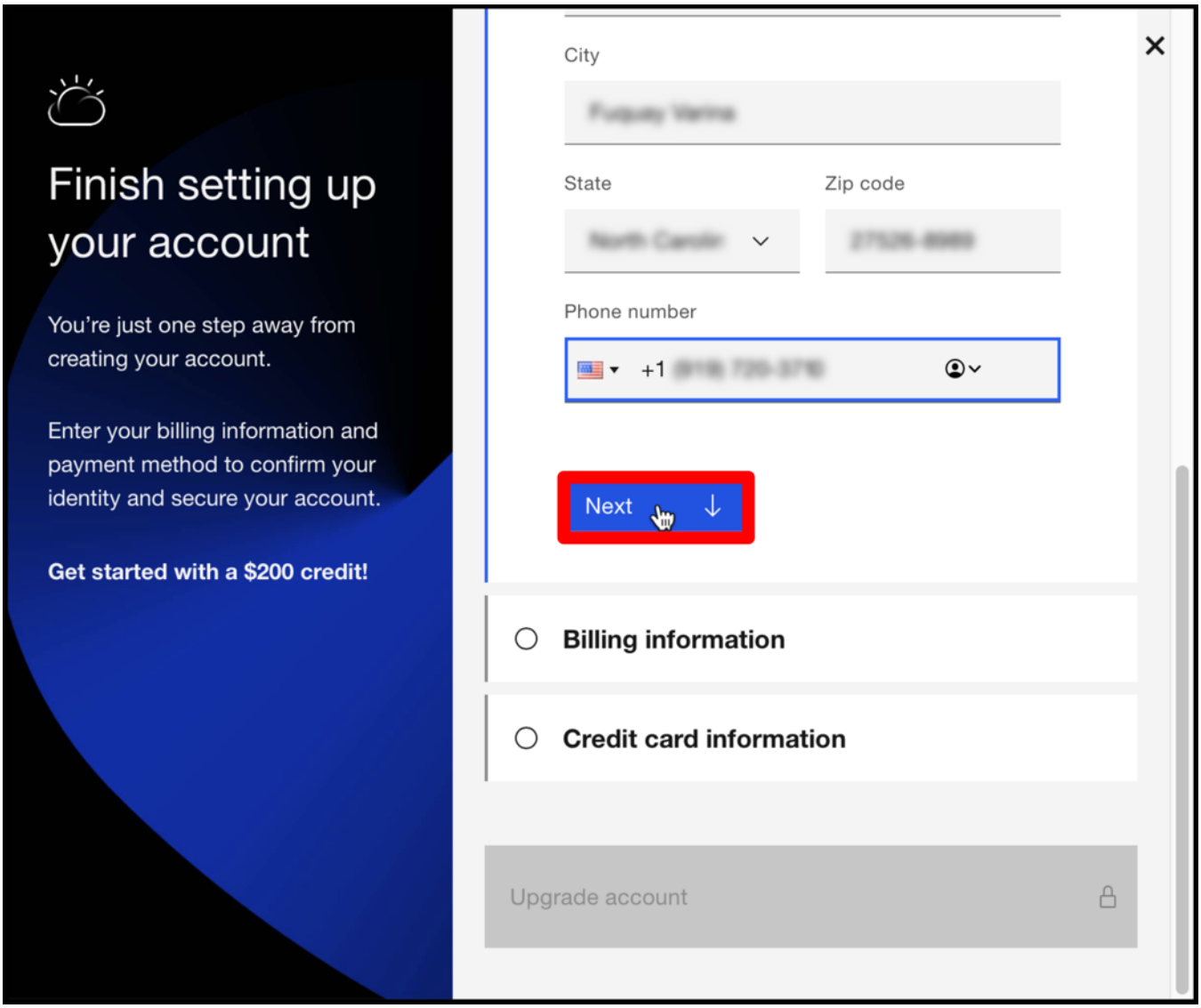
____ 8. Log in to your IBM Cloud account using the IBMid and password you provided earlier. If this is your first time logging into your account, a popup window will appear and you will be asked to provide additional information needed to activate the account. Start by **(A)** clicking the **Company**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



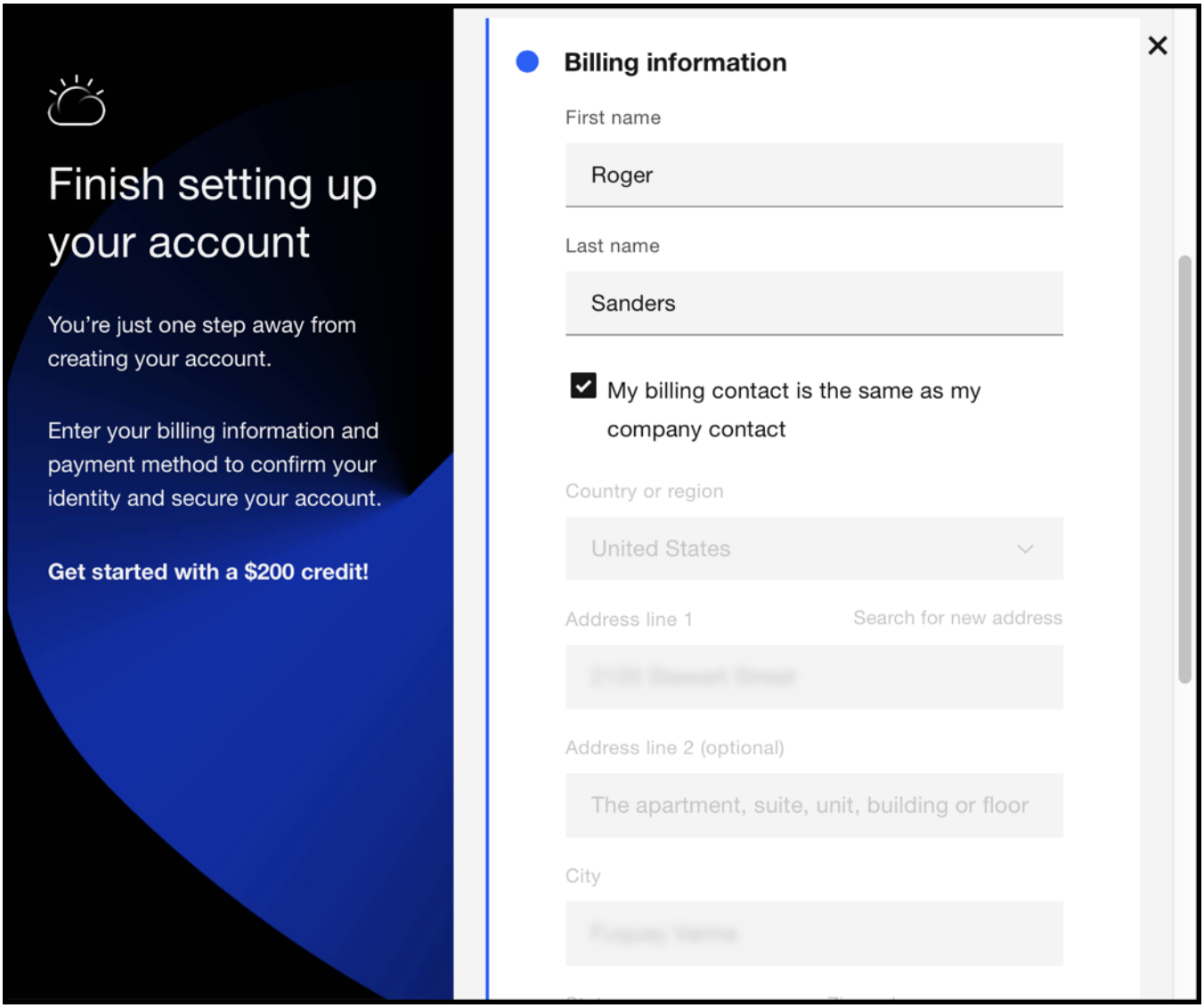
____ 9. If you indicated that your IBM Cloud account should be treated as a **Company** account, supply the appropriate **Company information** (i.e., **Company name** and company contact information) in the fields provided.



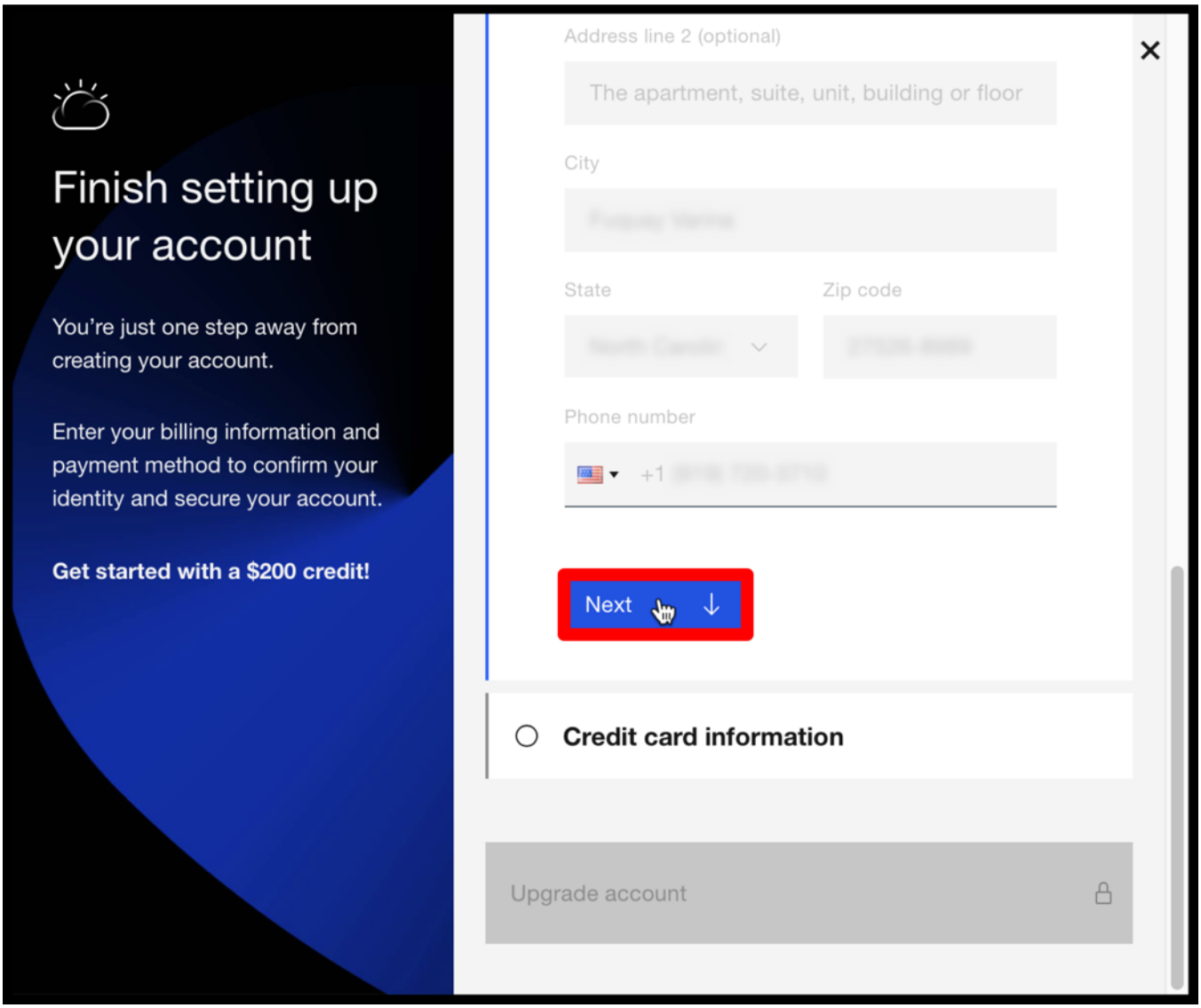
____ 10. After you have provided all the **Company information** requested (assuming this is a Company account), click the **Next** button.



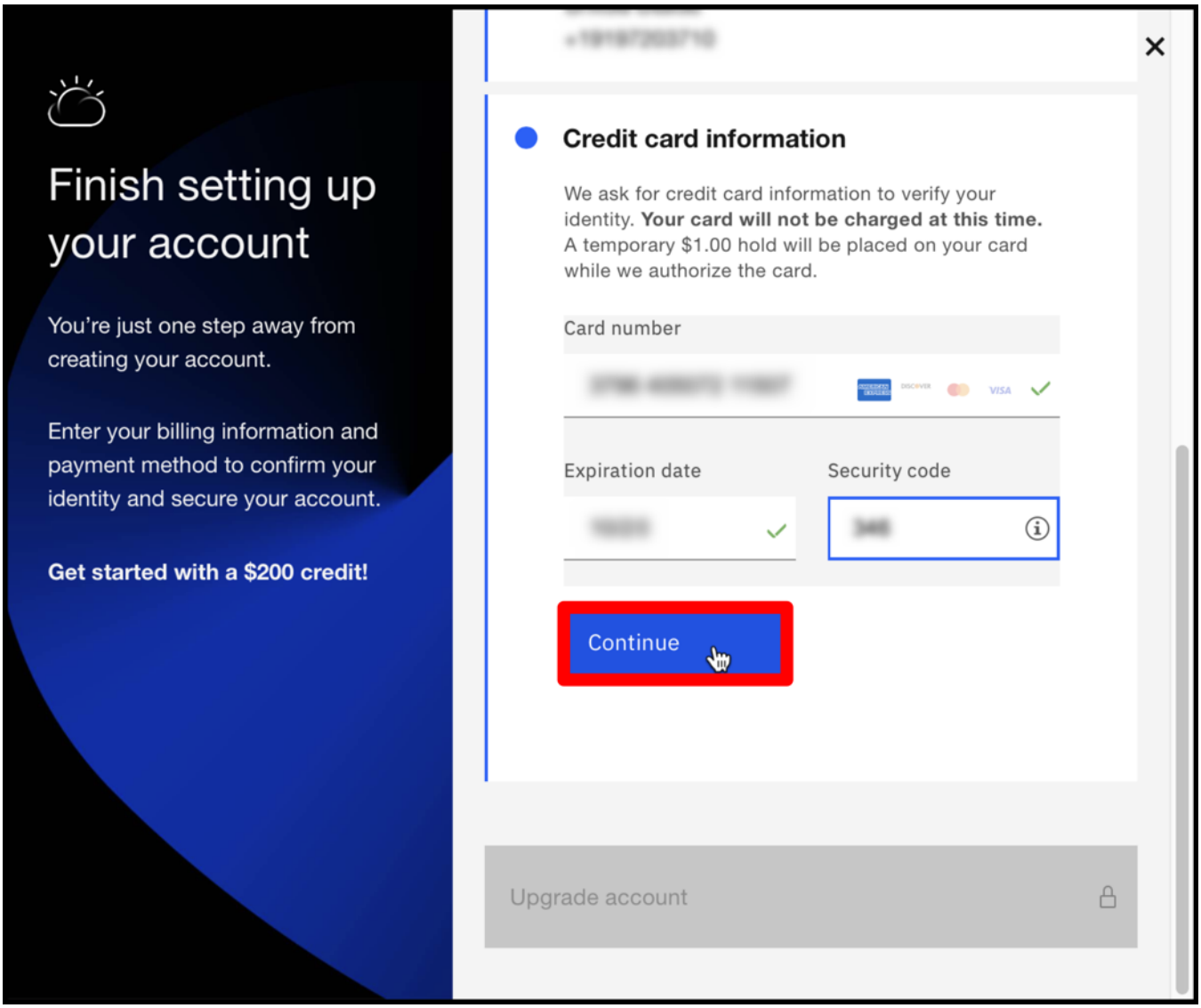
____ 11. Now provide the appropriate **Billing information.** If this is a **Company** account and you are the billing contact, click the **My billing contact is the same as my company contact** checkbox. Otherwise, provide the appropriate **Billing information** (i.e., contact information for the individual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



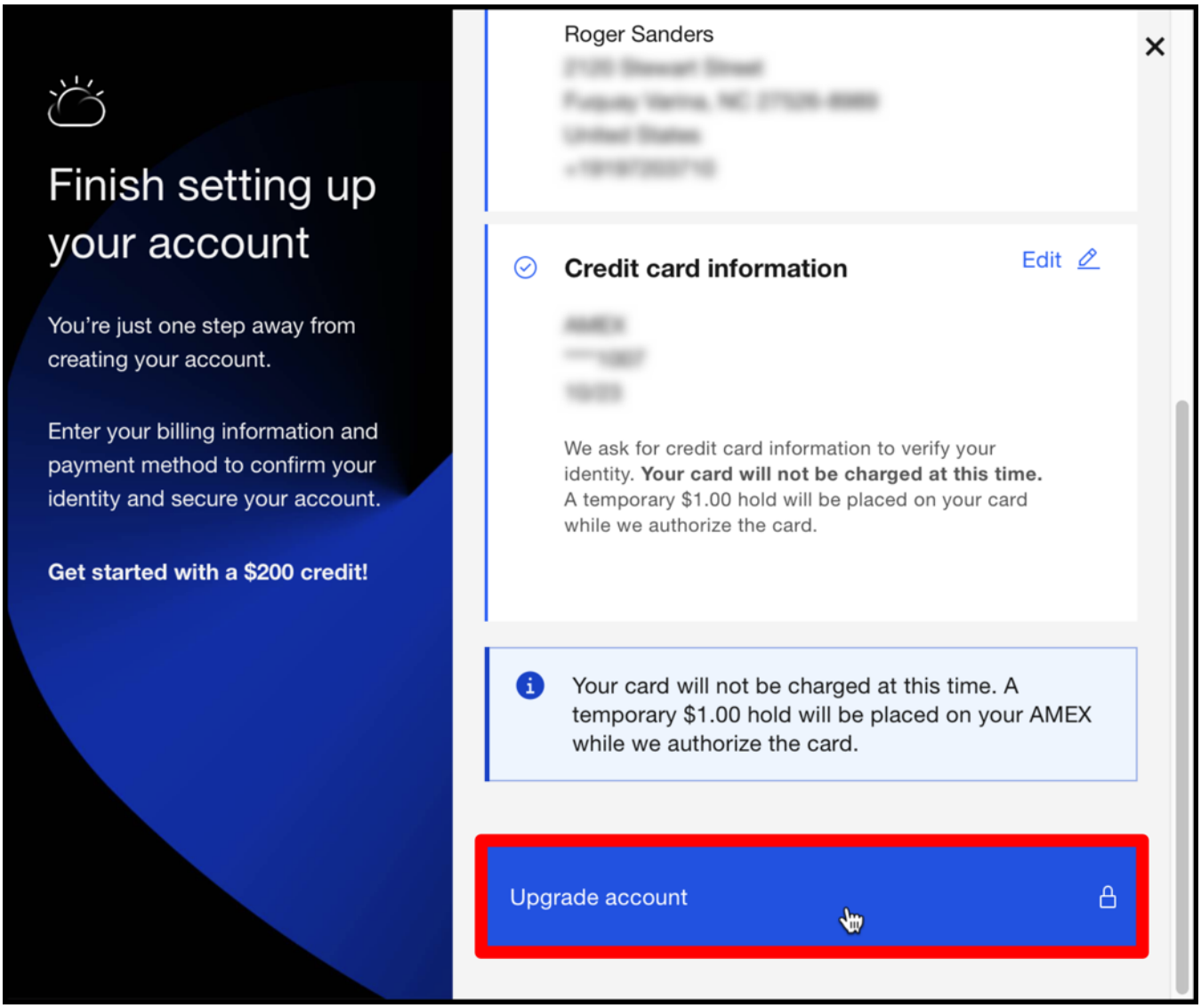
____ 12. Once you have confirmed the appropriate **Billing information** has been provided, click the **Next** button.



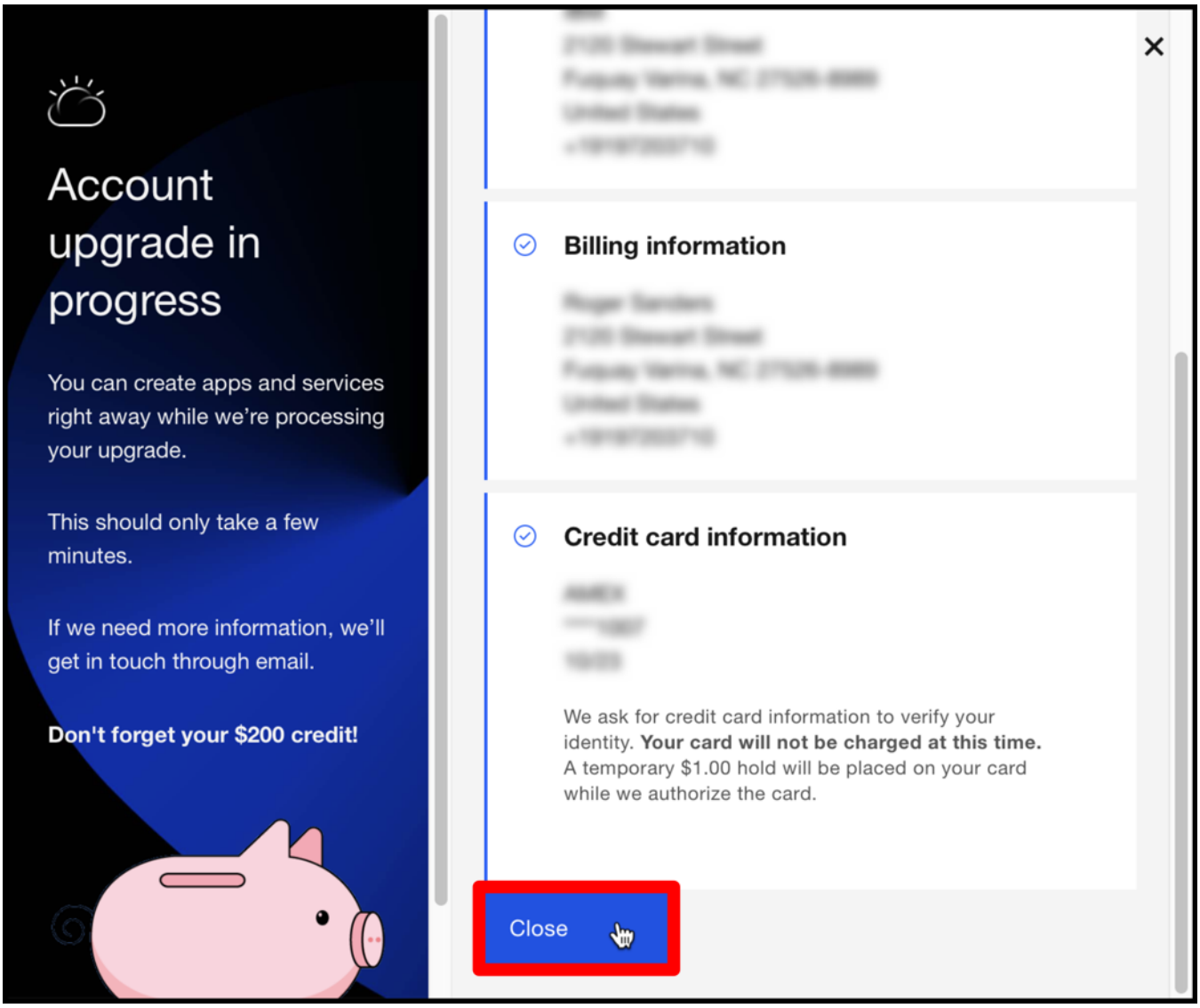
____ 13. Now, enter the appropriate **Credit card information** (i.e., **Card number**, **Expiration date**, and **Security code**) in the fields provided. It is important to note that ***this information is only used to verify your identity (therefore, the name on the credit card used MUST match the 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 ...**".)



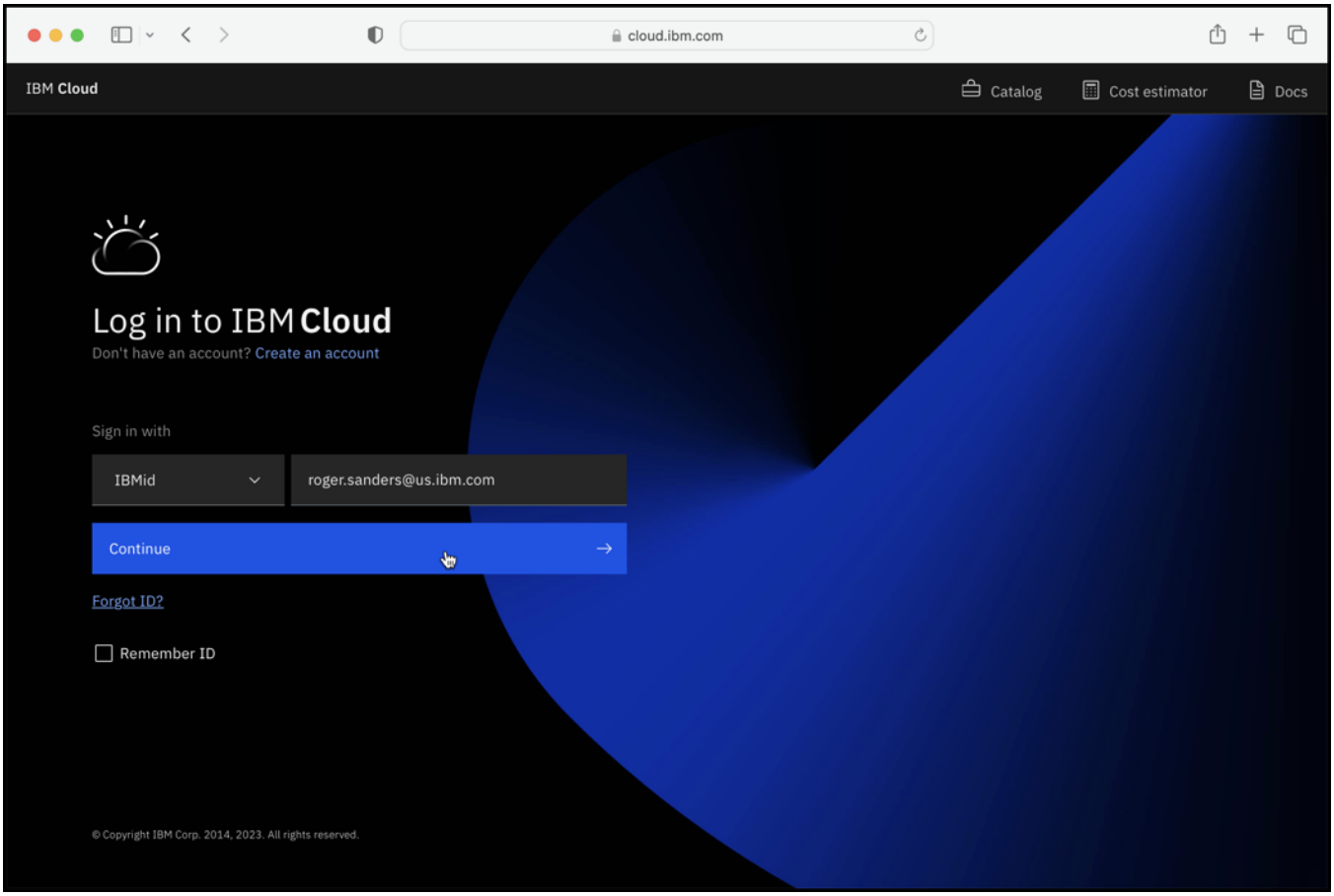
____ 15. When the credit card authorization process is complete, a **Close** button will be displayed at the bottom of the popup window. Click on this button. (When your IBM Cloud account is ready, you should receive an email from *IBM Cloud* saying "*Welcome to IBM Cloud! Let's get started.*")



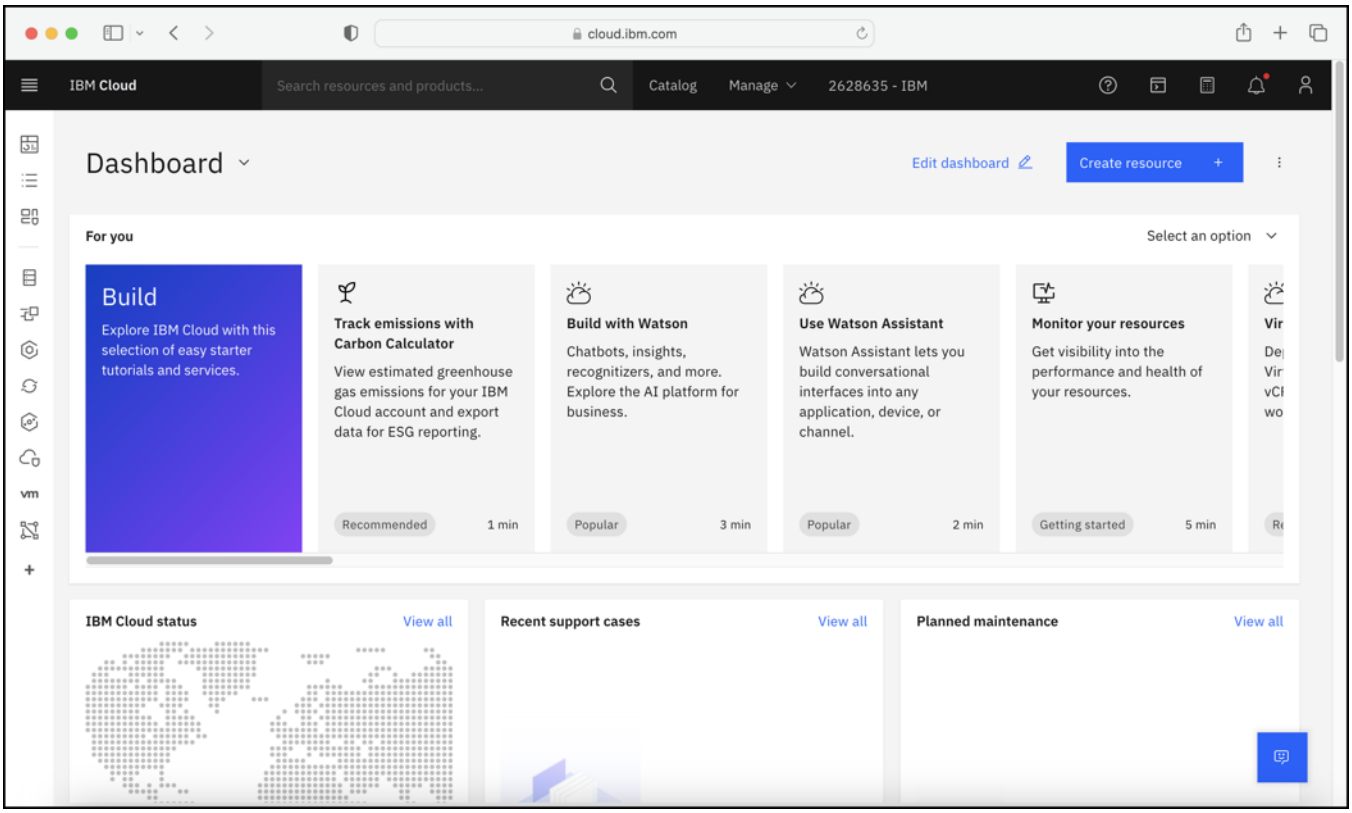
Congratulations! You now have an active **IBM Cloud** account that can be used to perform the rest of the exercises in this workshop.

3. Provision a Db2 on Cloud Lite plan database

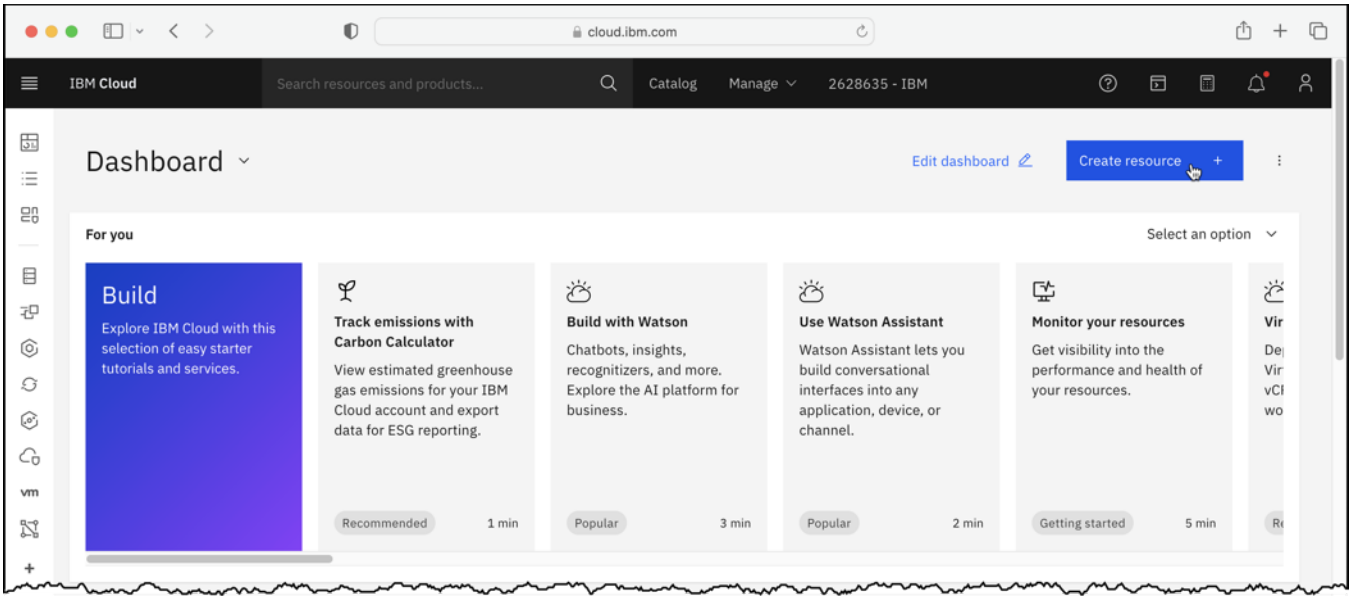
____ 1. Open a web browser and go to the *IBM Cloud* login web site (https://cloud.ibm.com/login). Once there, ensure the **Sign in with** dropdown list contains the value **IBMid** and enter the email address associated with your IBMid in the field immediately beside the dropdown list. When finished, 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.*)



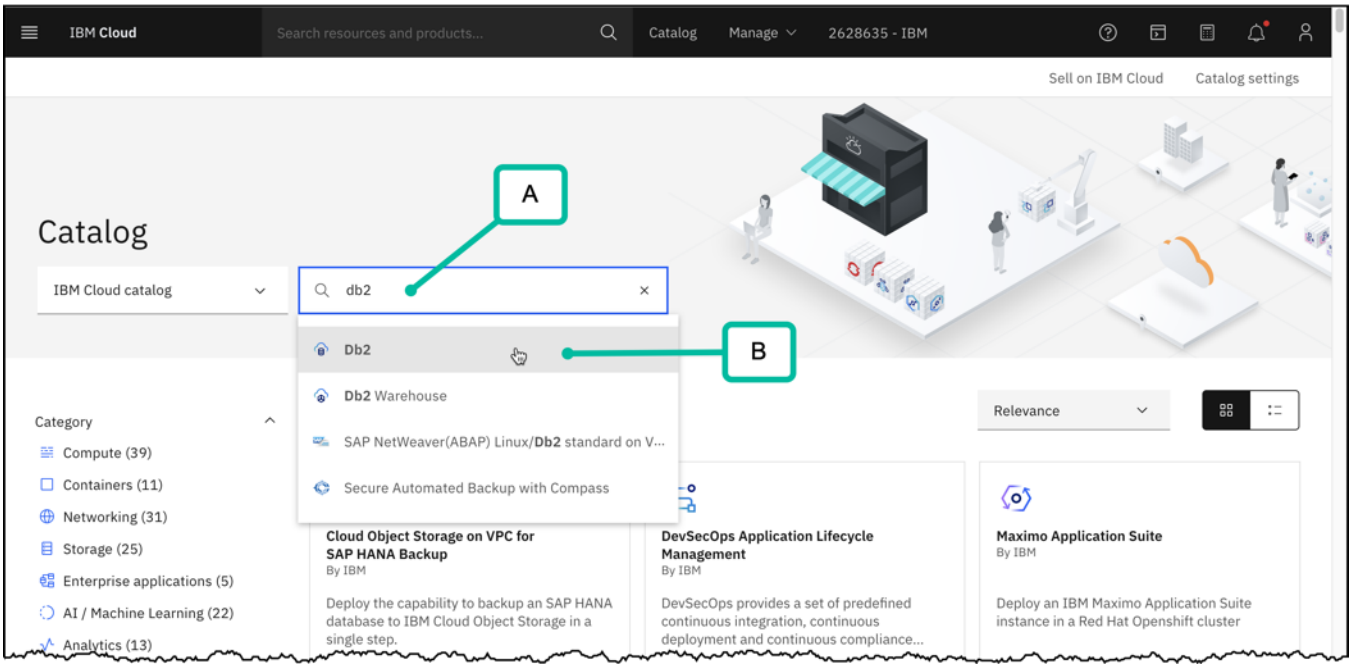
____ 2. If prompted, enter the password associated with the IBMid provided. Once you are logged in, you should see a screen that looks something like this:



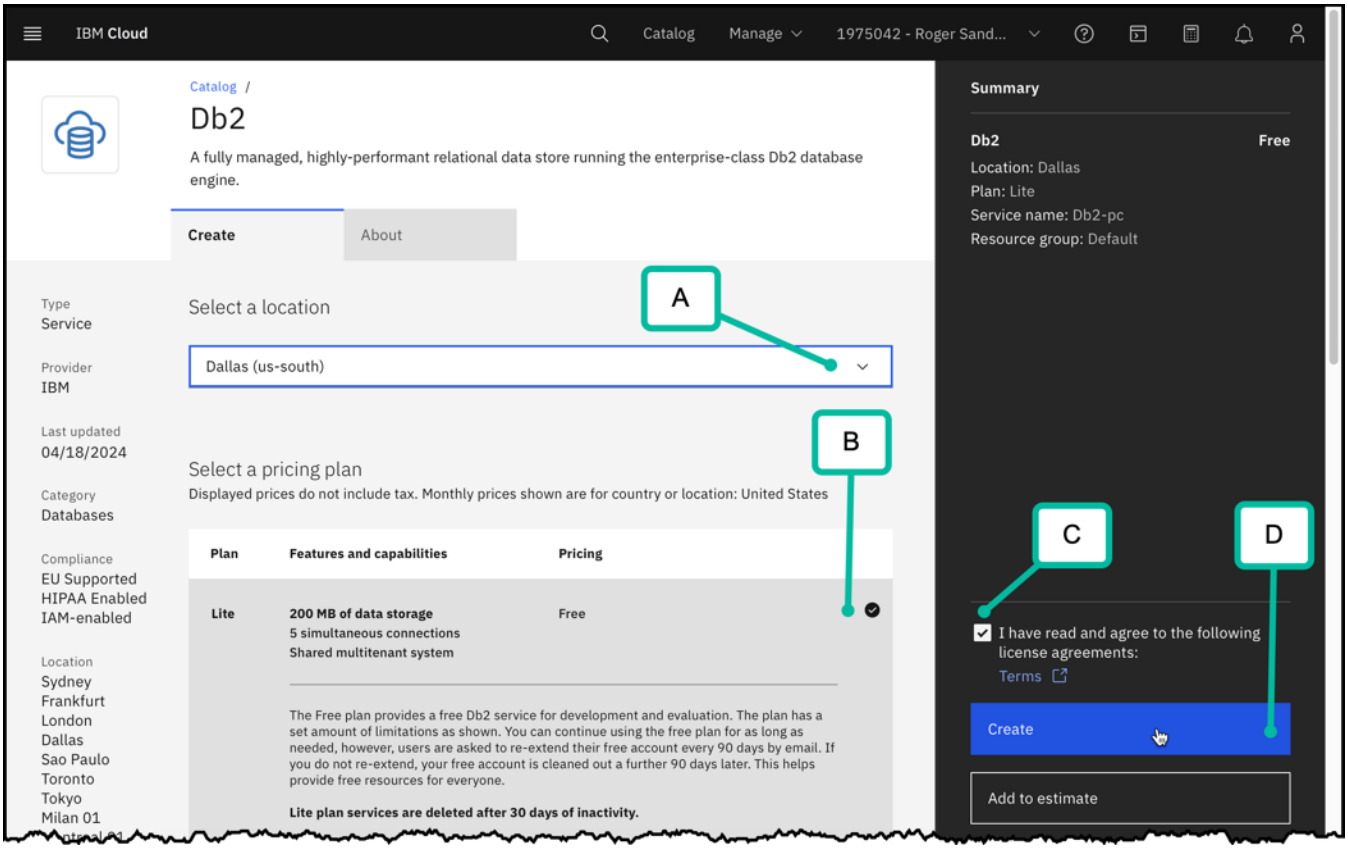
____ 3. Click on the **Create resource** button located at the top right corner of the screen. This will cause the IBM Cloud *Catalog* screen to be displayed.



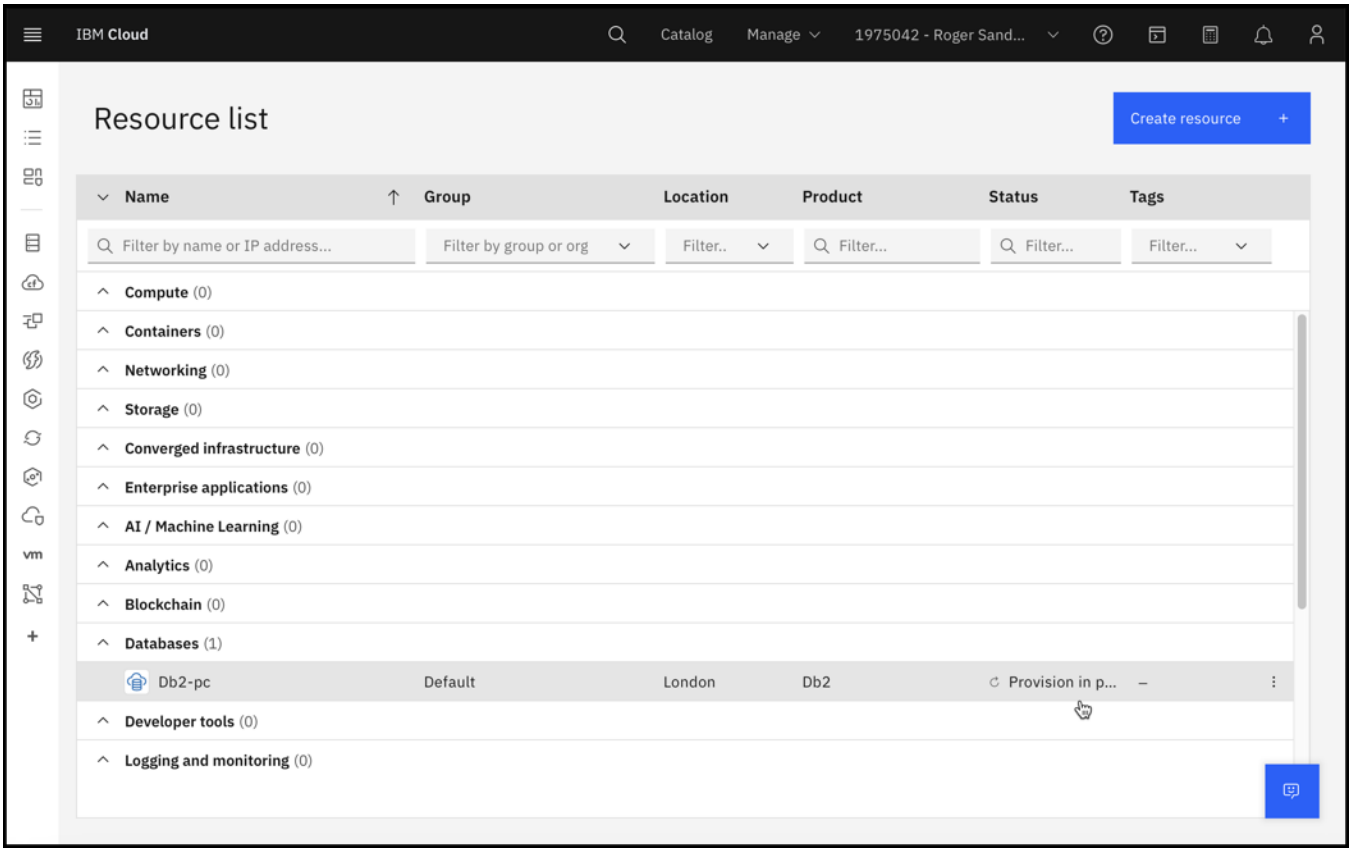
____ 4. When the IBM Cloud *Catalog* screen appears, **[A]** type the characters “db2” in the **Search** field located in the top left corner of the screen (i.e., the entry field with the magnifying glass icon on the left.) This will cause a list of the Db2 resources available on IBM Cloud to be displayed.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



____ 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.



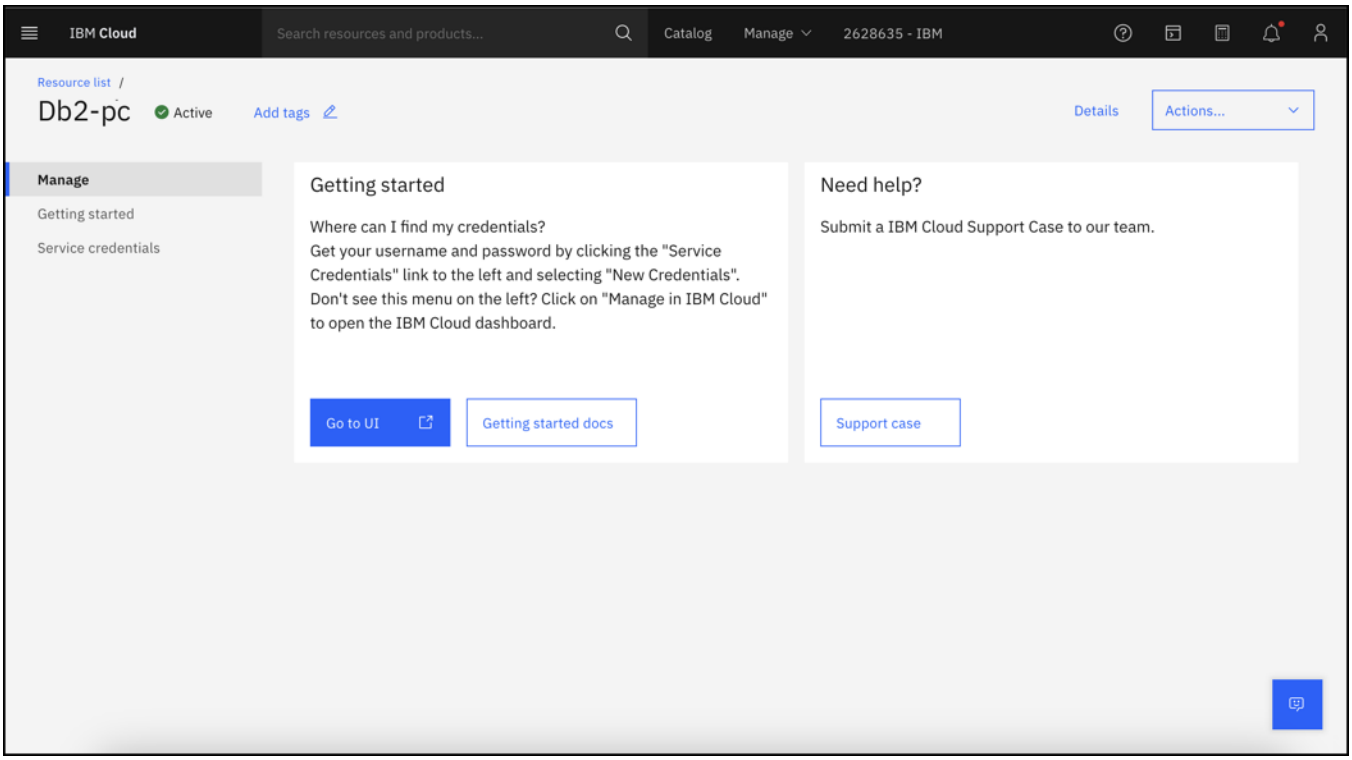
____ 6. When the **Create** button on the *Db2* resource screen is pressed, this screen will disappear and the message "**Creating ...** " will be displayed in its place. Shortly afterwards, the *Db2* resource screen will be replaced with an IBM Cloud *Resource list* screen. This screen contains a list 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**."



____ 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.

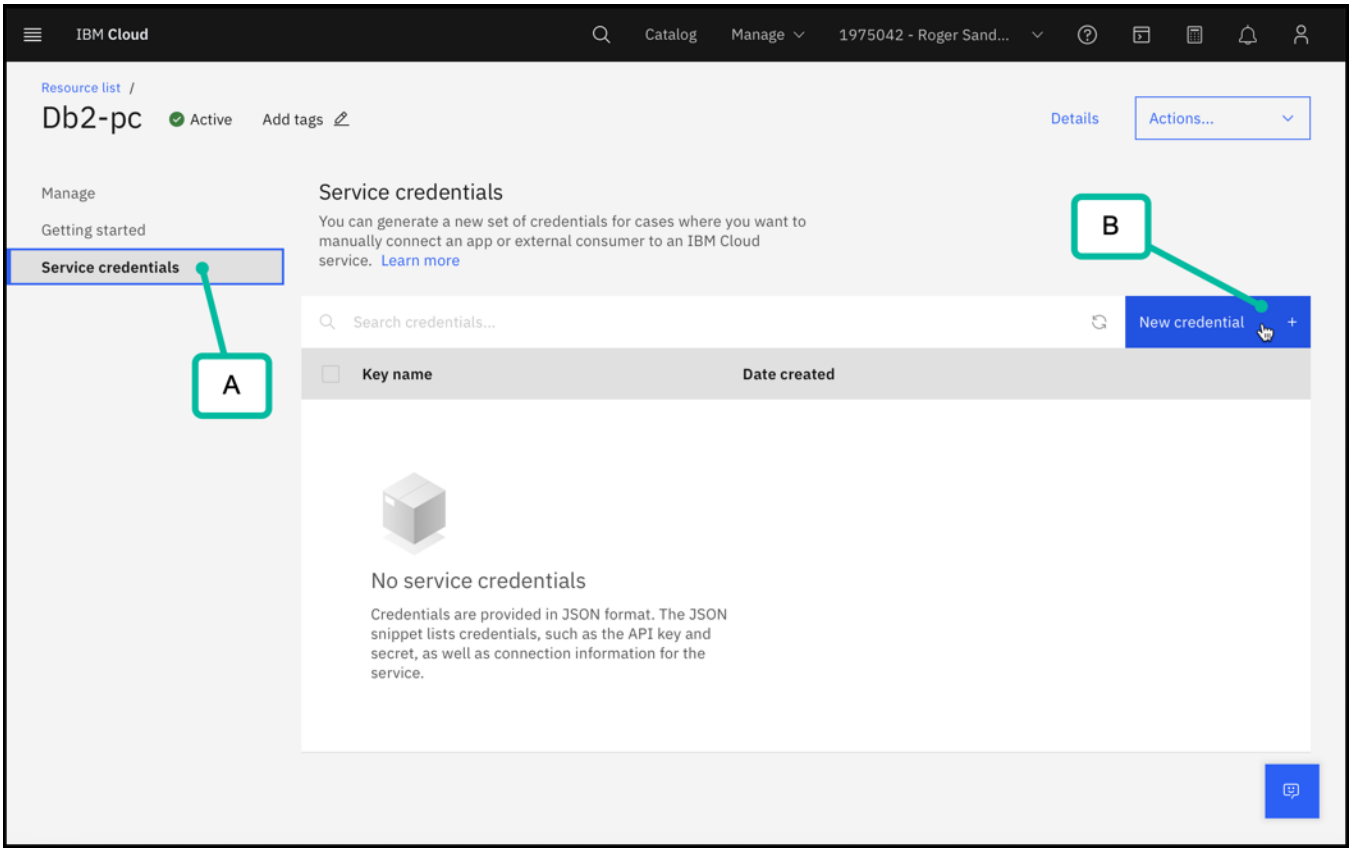
____ 8. Access the newly created Db2 on Cloud database by clicking on the database ID shown in the **Resource list**. (In this example, the database ID is *Db2-pc*. As we will see later, the name of this database is *bludb*.)

____ 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)*.

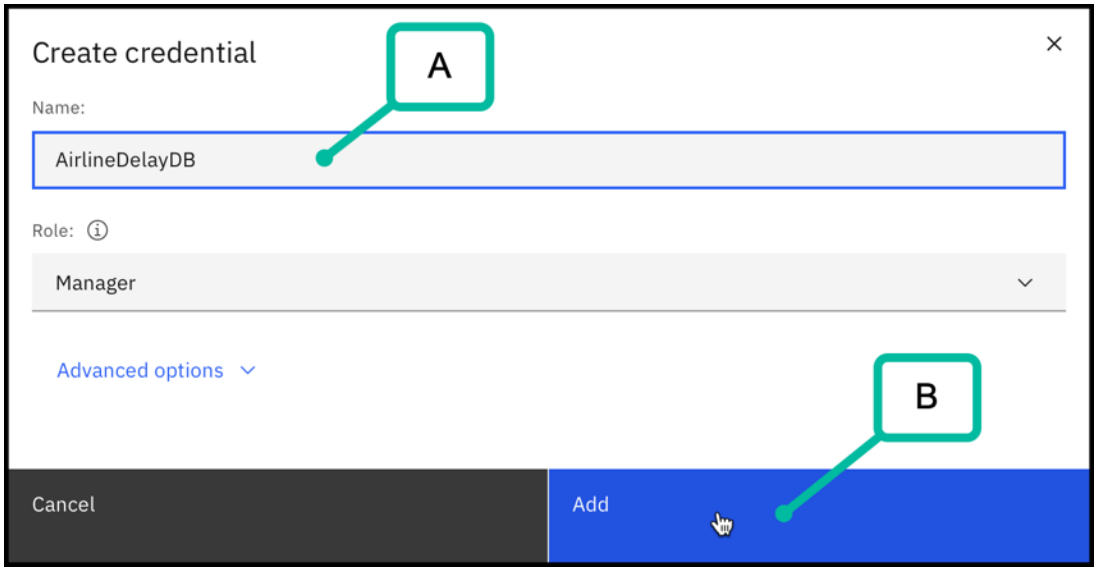


4. Create a service credential for the Db2 on Cloud database.

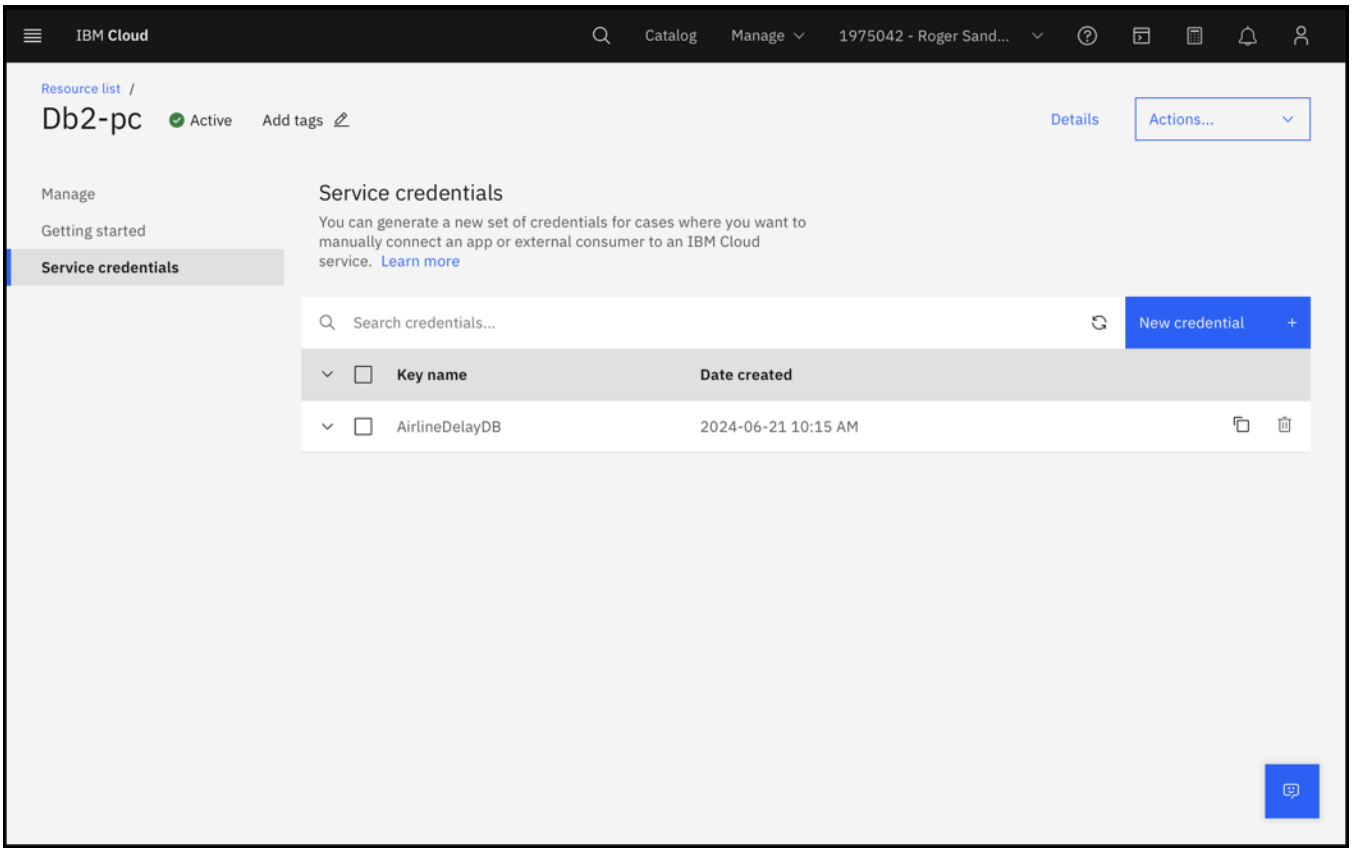
____ 1. To work with an IBM Cloud resource from another application (for example, a Jupyter Notebook), one or more service credentials for that resource must exist. As this screen illustrates, when a Db2 on Cloud database is first provisioned, it has **No service credentials**. To create a service credential 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.



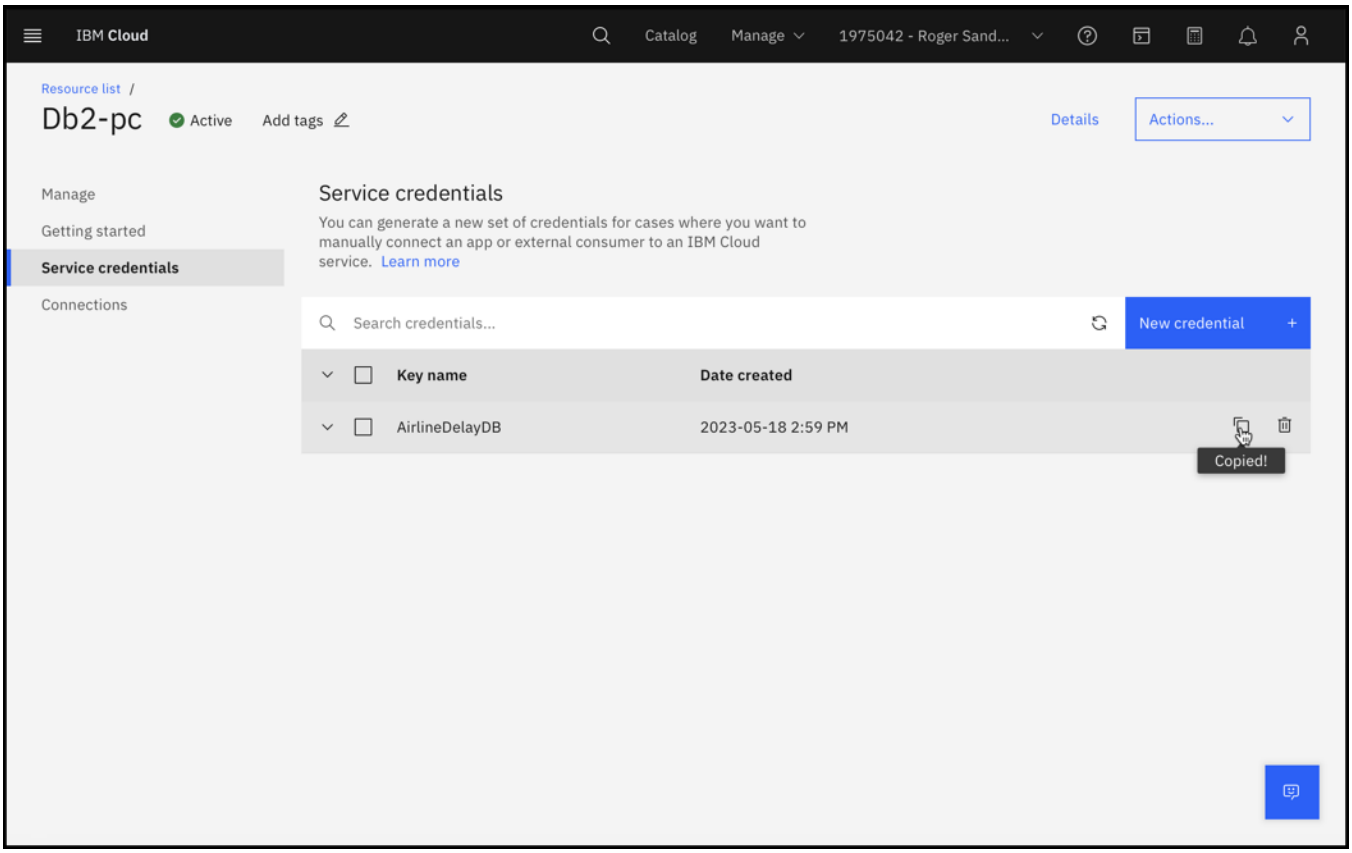
____ 2. When the *Create credential* popup window is displayed, **[A]** enter a name for the new service credential in the **Name:** field. (The default name will be *Service credentials-1*, but for this exercise, we're going to change the name to *AirlineDelayDB*). **[B]** Then, click the **Add** button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

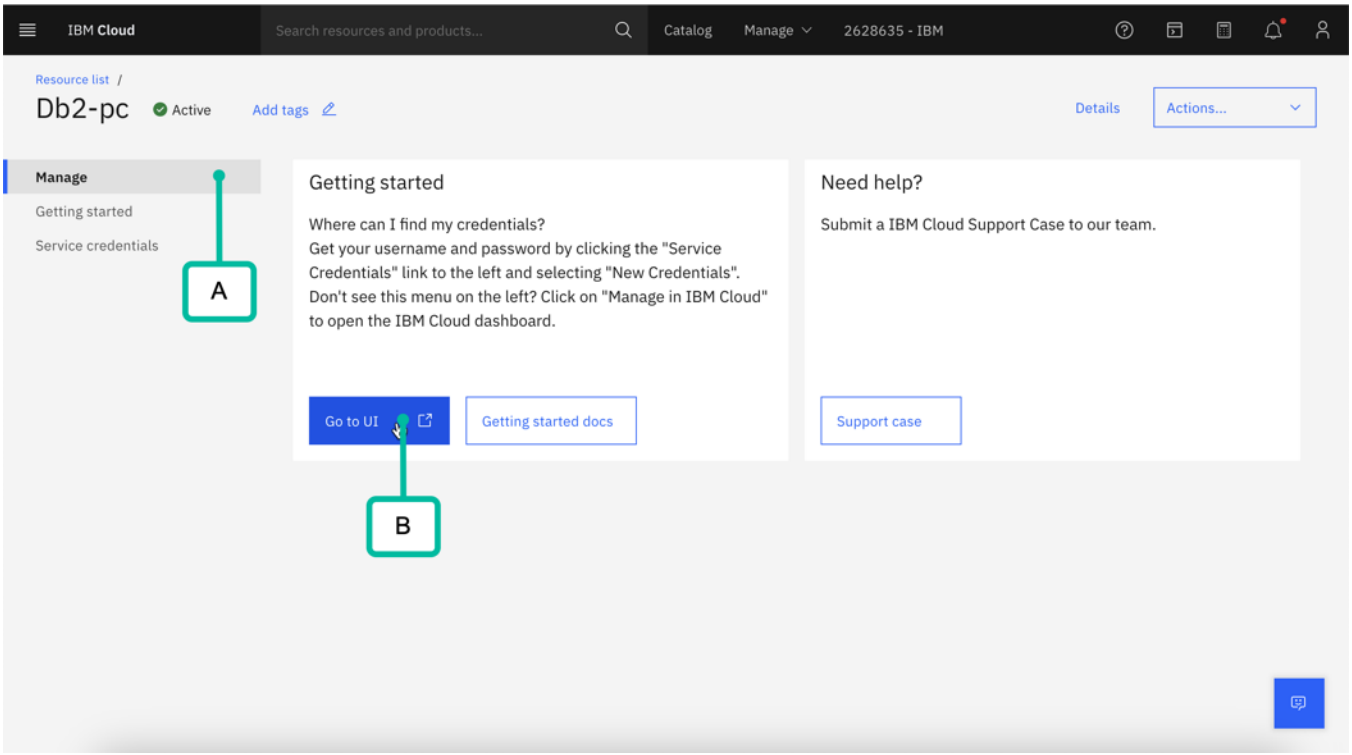


____ 4. Click the “**Copy to clipboard**” icon shown on the right-hand side of the listing for the service credential just created. Then, paste the information copied into a text file – for this workshop, use a file named **AirlineDelayDB_Credentials.txt** – and save it for later.

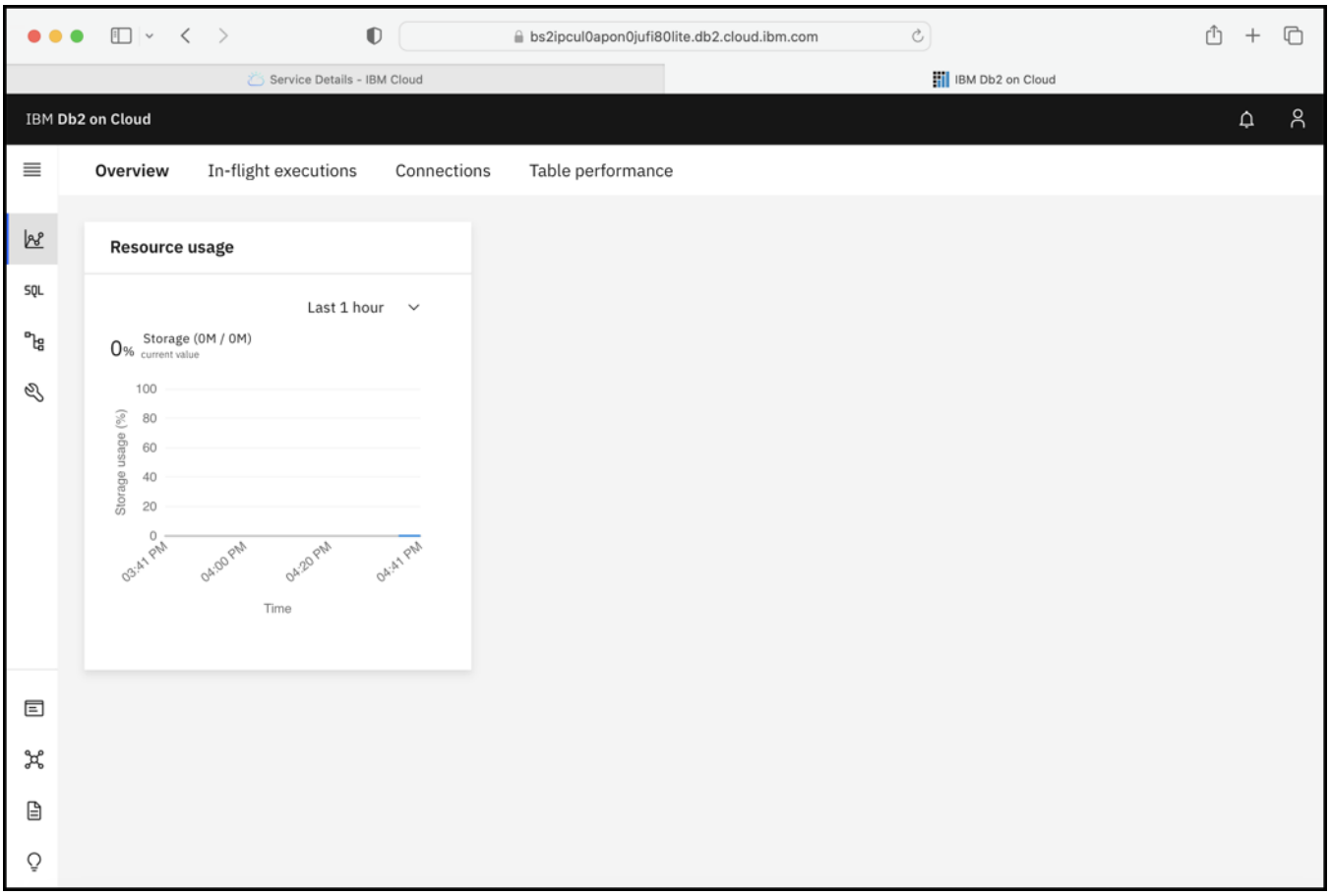


5. Query the system catalog using the Db2 on Cloud User Interface (UI).

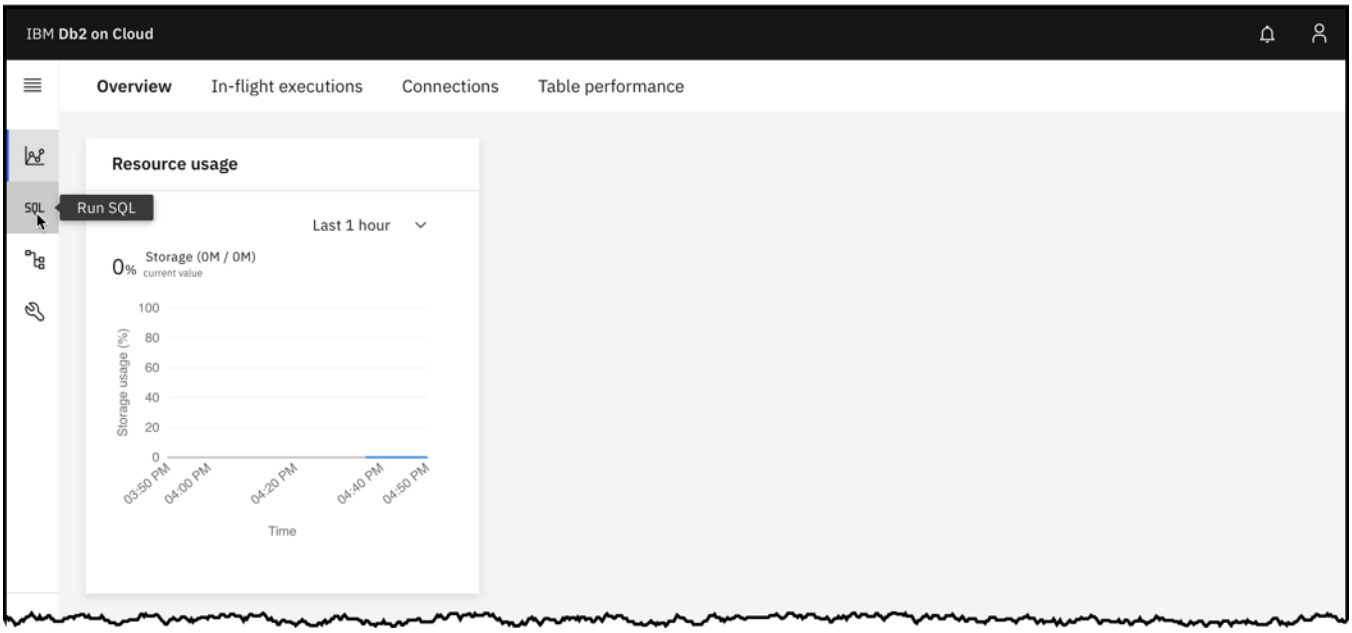
____ 1. To work with the Db2 on Cloud database using the Db2 on Cloud User Interface, **[A]** click on the **Manage** item in the menu shown in the top left corner of the screen. **[B]** Then, click the **Go to UI** button.



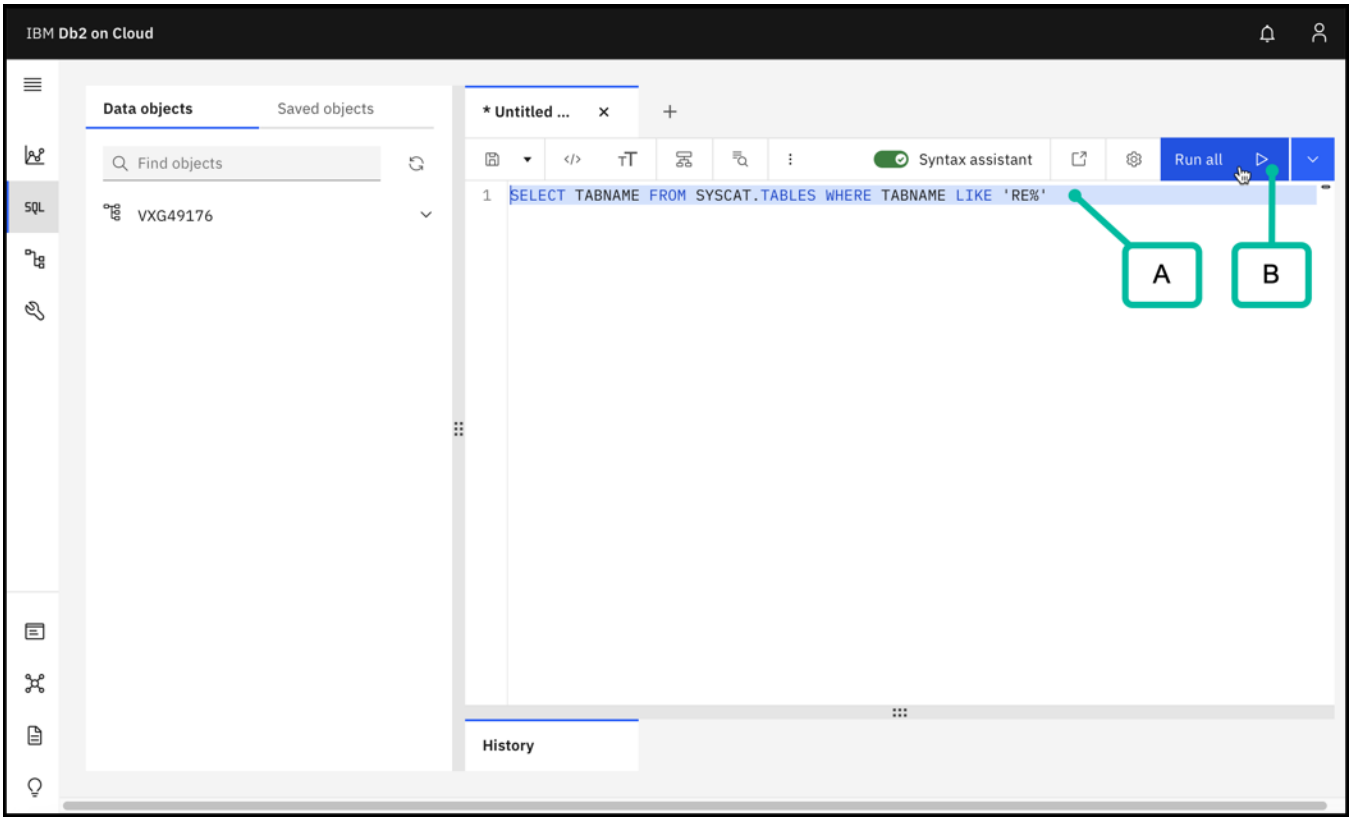
____ 2. When the **Go to UI** button is selected, a new web browser tab named **IBM Db2 on Cloud** is opened. This tab normally consists of an **Overview** screen, which is the default starting point for the Db2 on Cloud User Interface.



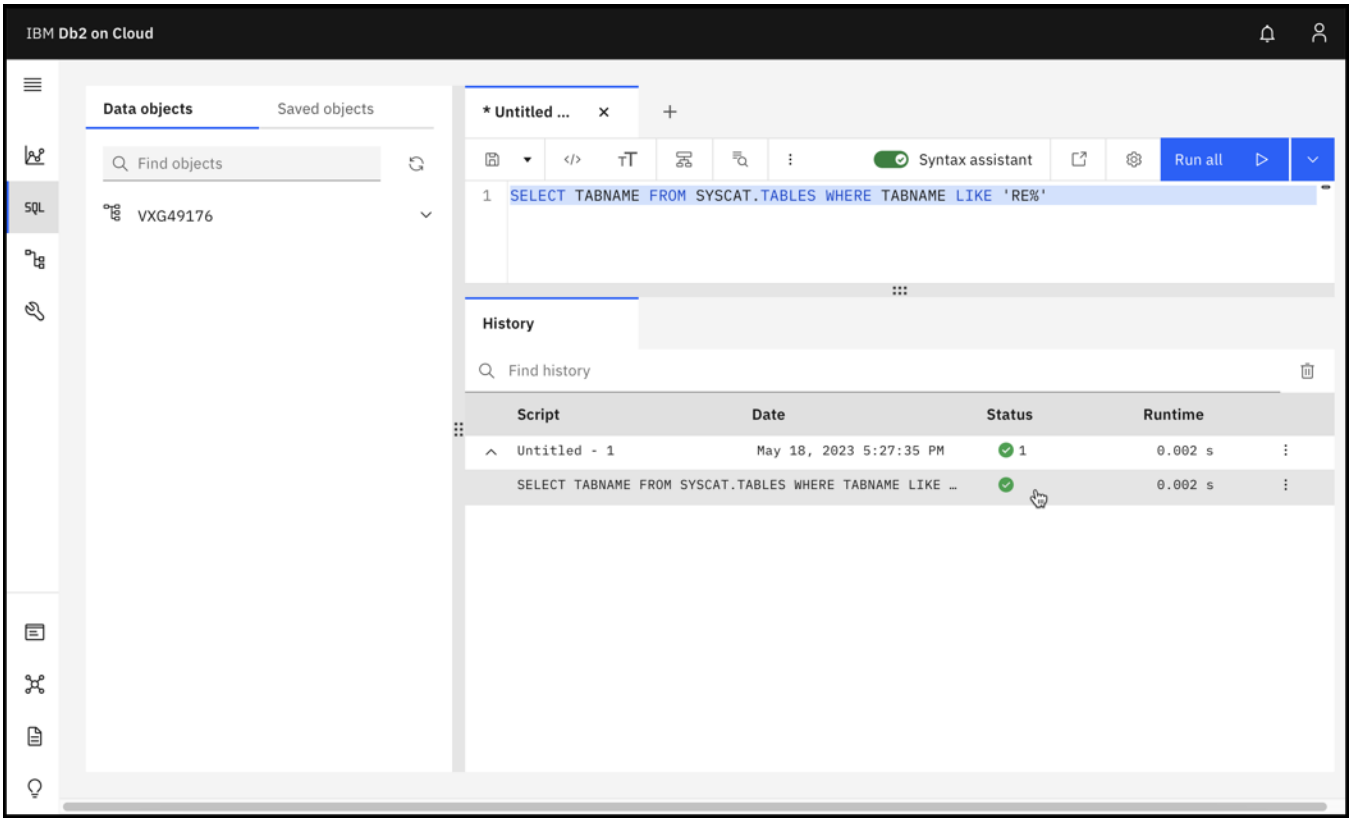
____ 3. Switch to the *Run SQL* screen by clicking the **SQL** icon found in the menu shown on the left-hand side of the screen.



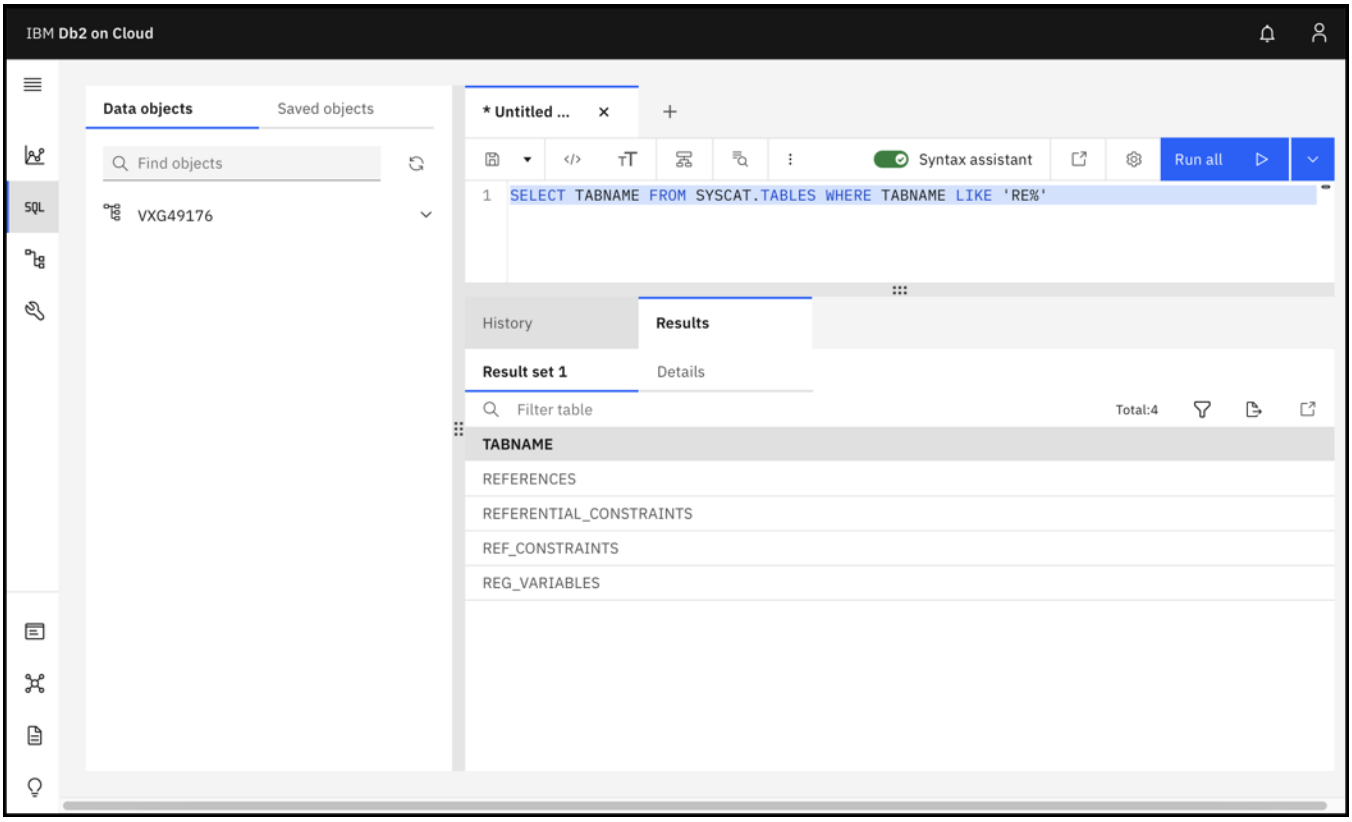
____ 4. **[A]** Enter the SQL statement **SELECT TABNAME FROM SYSCAT.TABLES WHERE TABNAME LIKE 'RE%'** in the ***Untitled - 1** portion of the *Run SQL* screen. **[B]** Then, click the **Run all** button to execute the statement. (When executed, this statement will query the database's system catalog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



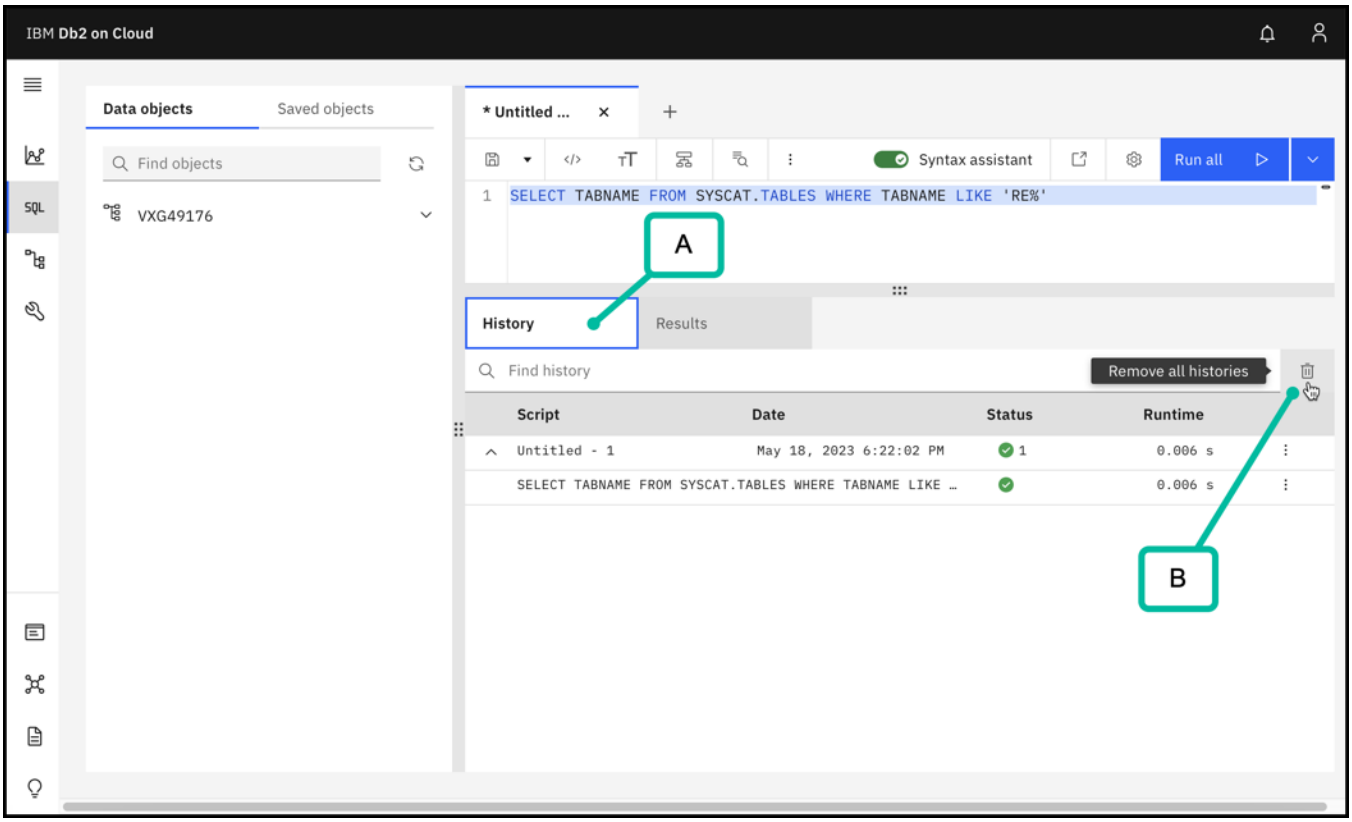
____ 5. When the SQL statement entered in the previous step is executed, a record about its execution will be recorded in a history log and a summary of the record will be displayed in the **History** window/tab located at the bottom of the screen. If the SQL statement executed successfully, the **Status** 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.)



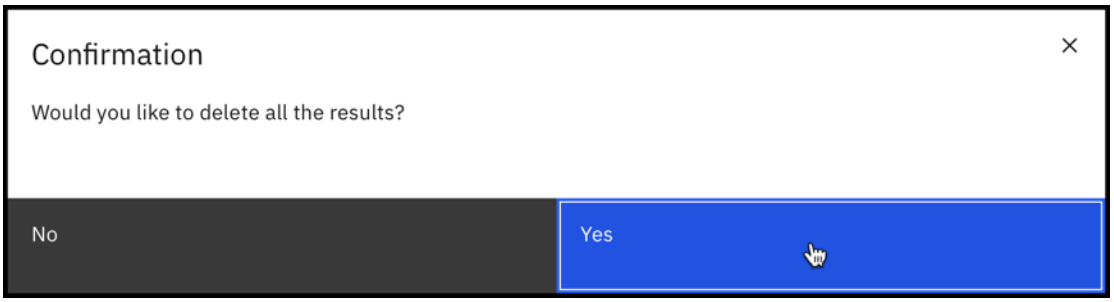
____ 6. When you click on the history record for the SQL statement just executed, you will be presented with a new tab named **Results** that contains a table showing the contents of result set produced when the statement was executed, together with the number of records (rows) returned in the result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



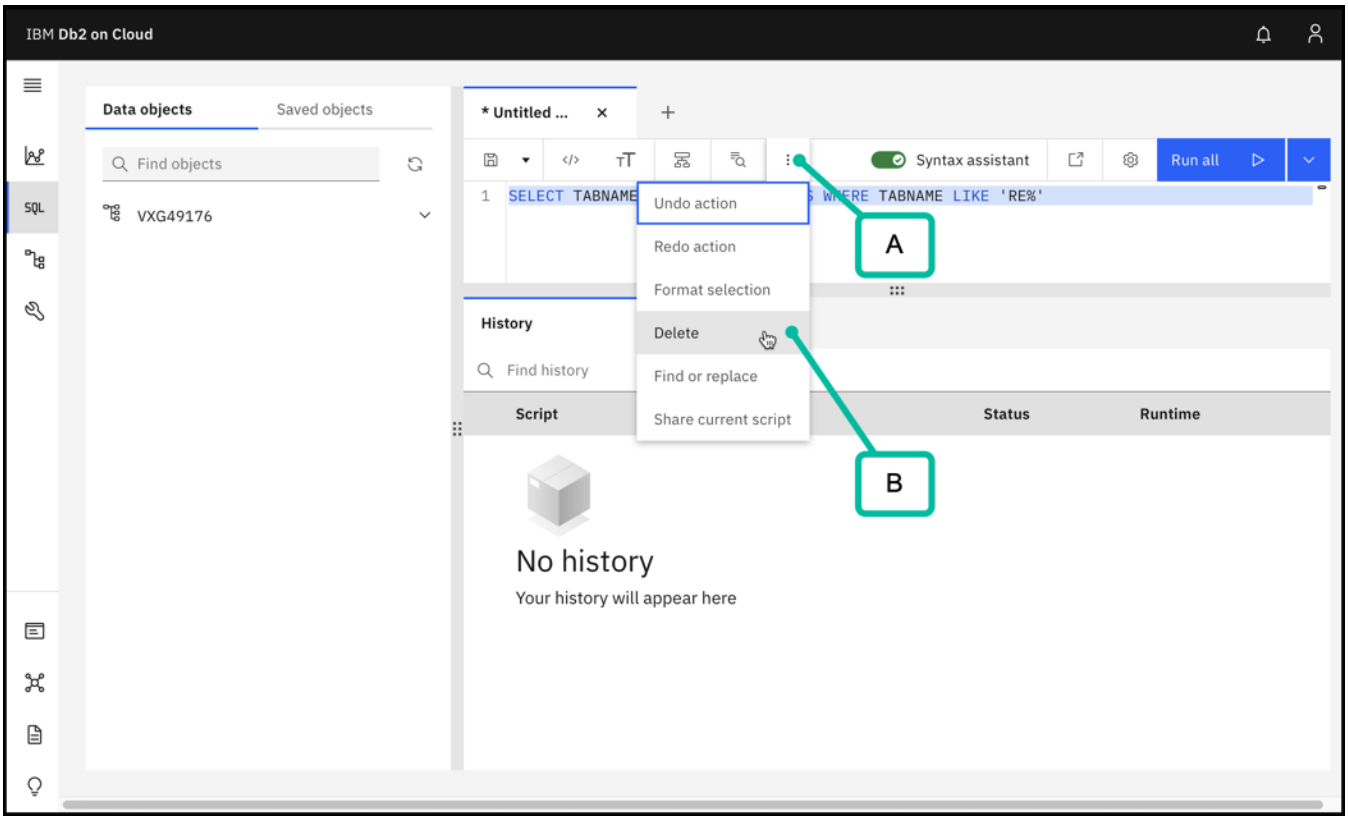
____ 7. Clear the results and the history log by **[A]** clicking on the **History** tab (to return to the history log display), and then **[B]** clicking on the trash can icon located at the top left corner of the history log table.



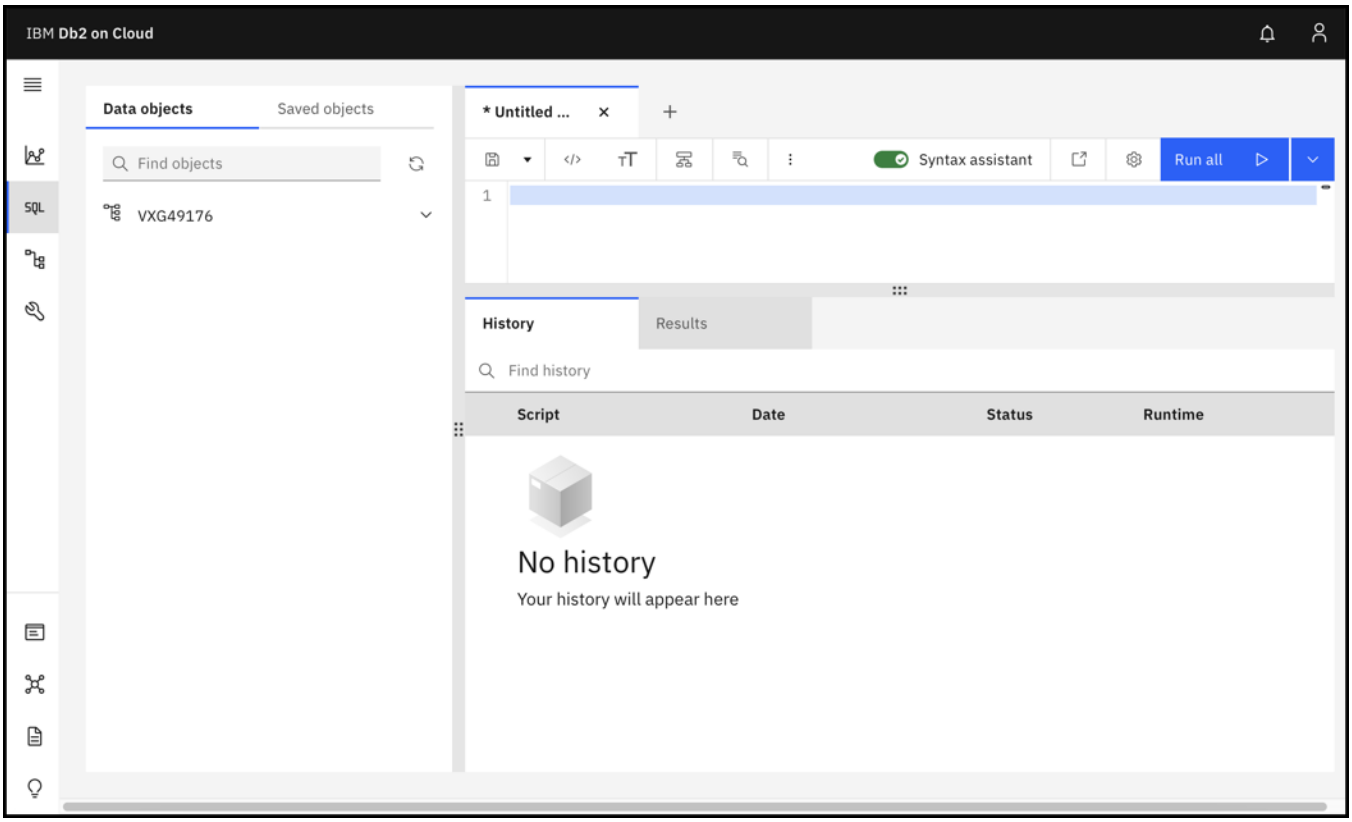
____ 8. When asked to confirm your decision to delete all results, click the **Yes** button.



____ 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.



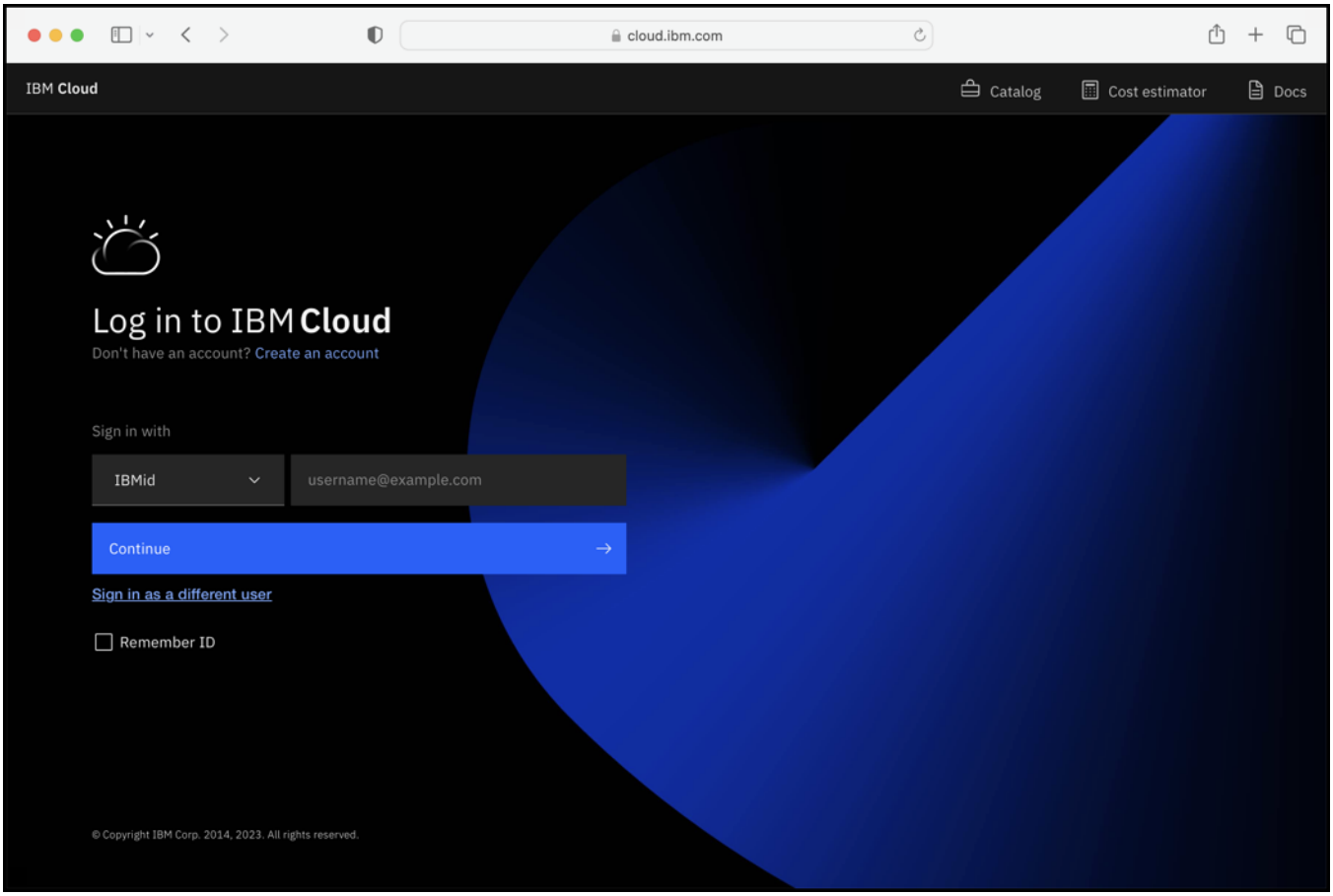
Congratulations! You have just created a Db2 on Cloud database and you have seen how to interact with that database using the Db2 on Cloud User Interface. We will use this database with other exercises in this workshop.

III. Building and populating a Db2 on Cloud database table

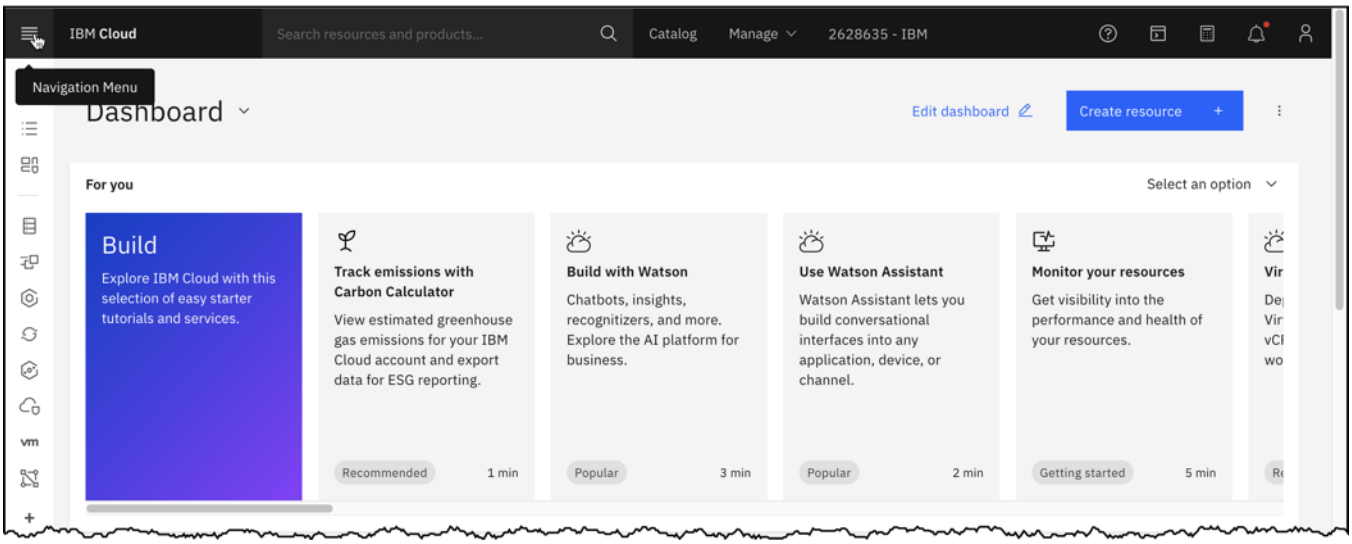
When an IBM Db2 on Cloud database is first created, it only contains the system catalog (which we looked at earlier.) For the database to be useful, it must contain at least one user-defined table. So, in this section, we will create a user-defined table and populate it with data stored in a Comma Separated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

1. Create a table named AIRLINE_DELAY_CAUSE.

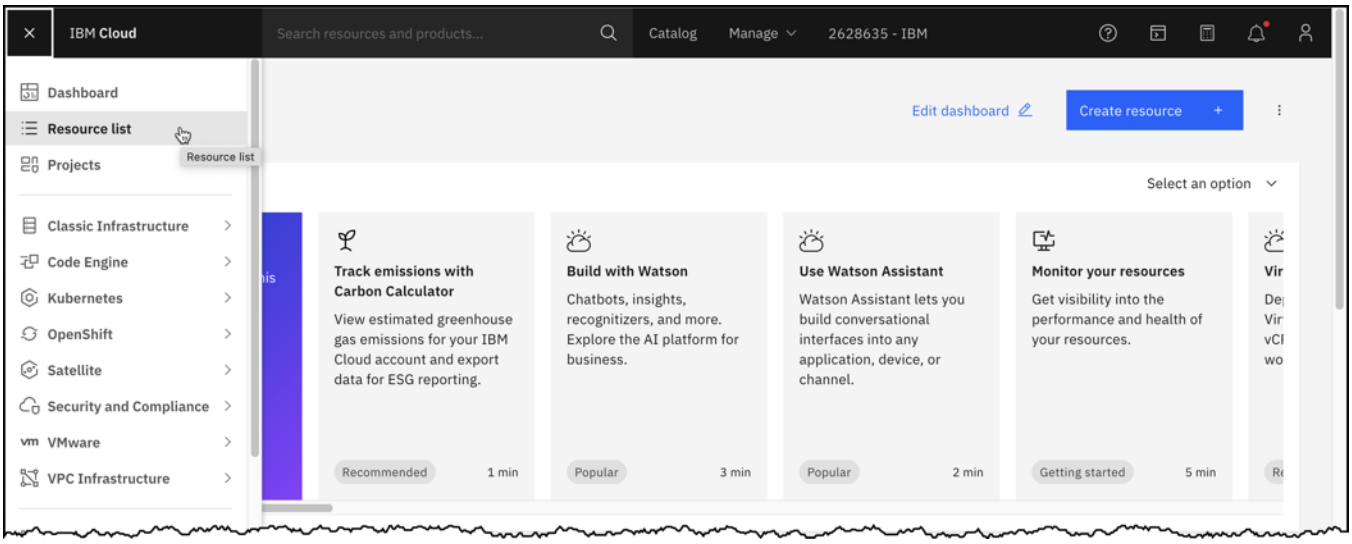
____ 1. Open a web browser and go to the *Log in to the IBM Cloud* web site (https://cloud.ibm.com/login). You should be presented with a screen that looks something like this:



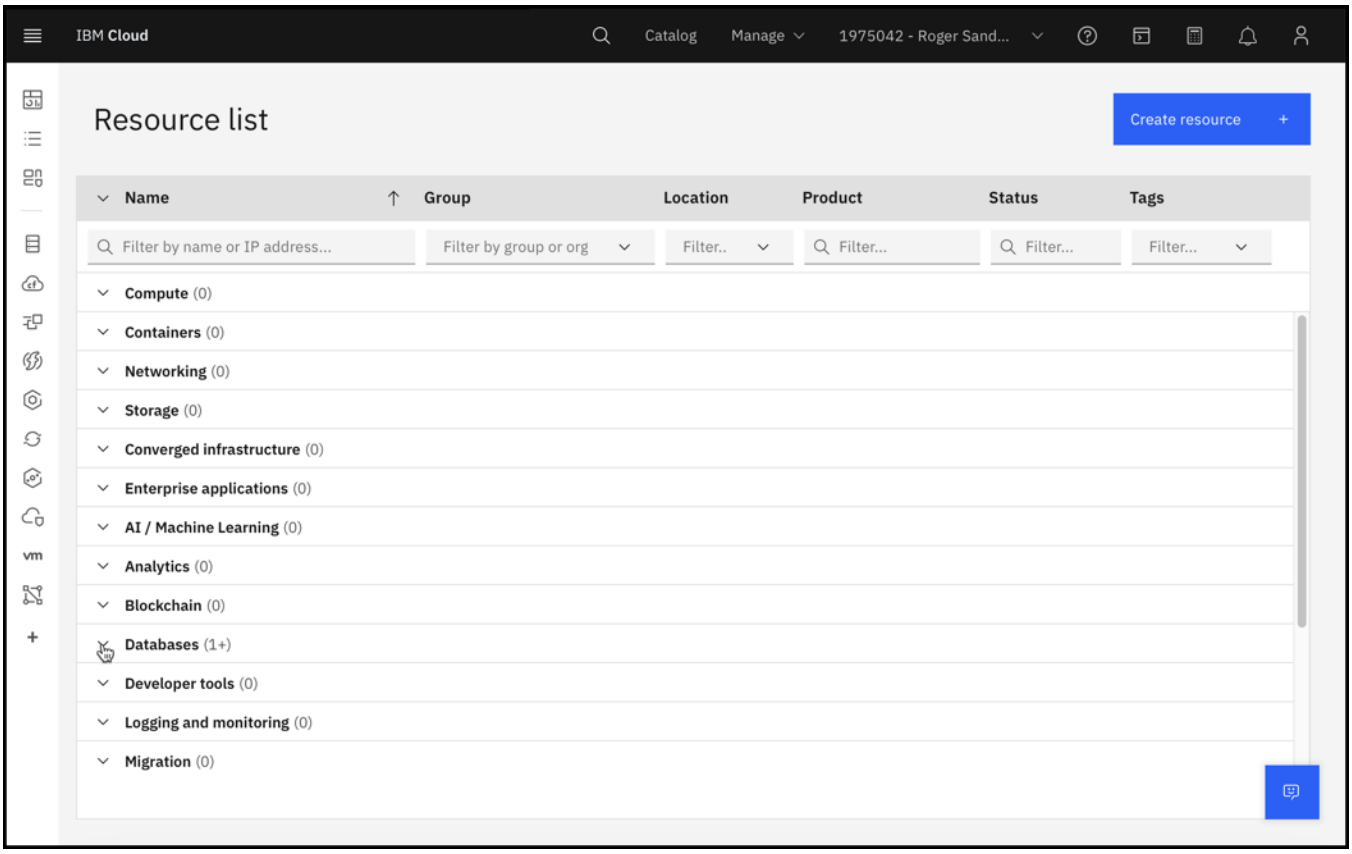
____ 2. Log in to IBM Cloud using your IBMid and password. If prompted, log in to IBM as well. Once you are logged in and the *IBM Cloud dashboard* appears, click the hamburger icon (also referred to as the triple bar or trigram symbol) located in the top left corner of the screen. This will open the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

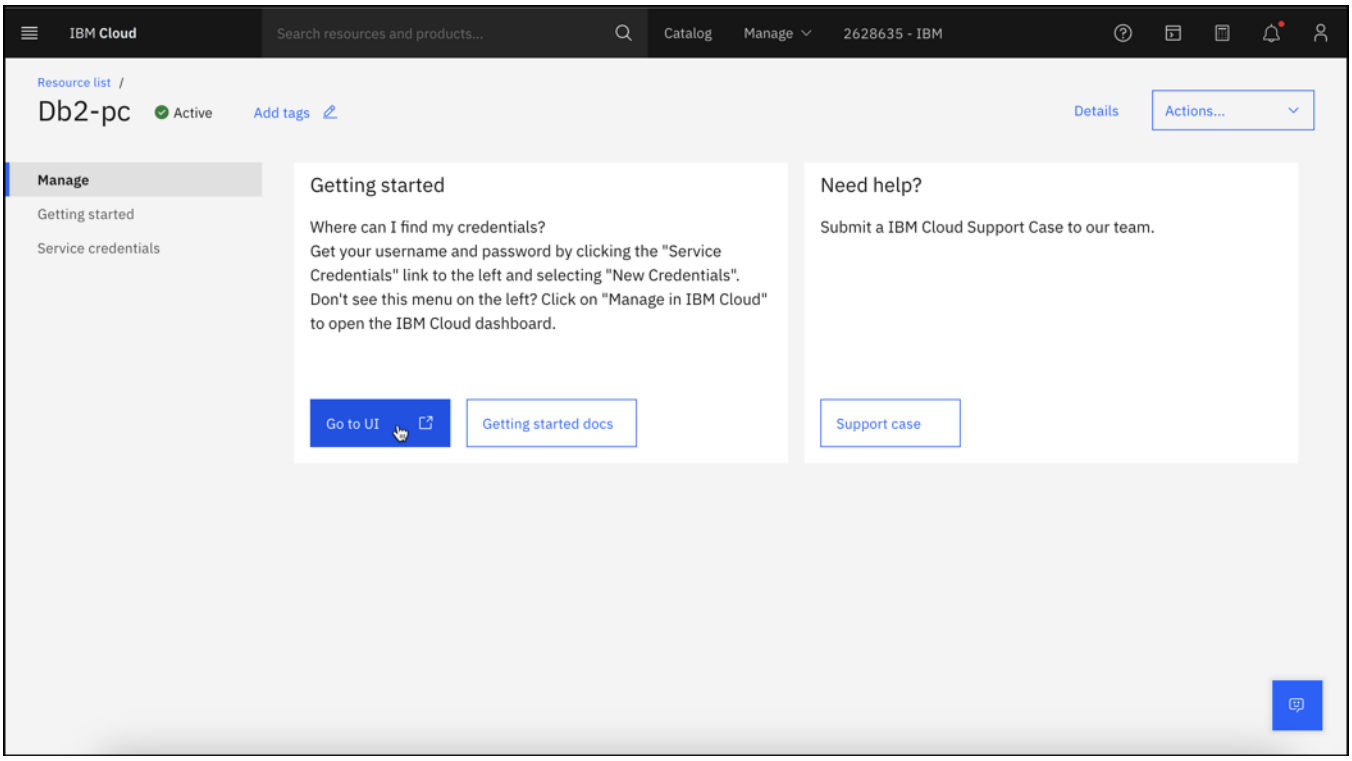


____ 4. Locate the **Databases** item in the resource list shown and click on the **V** shown to the left of it. You should see the ID for the Db2 on Cloud database that was created earlier.

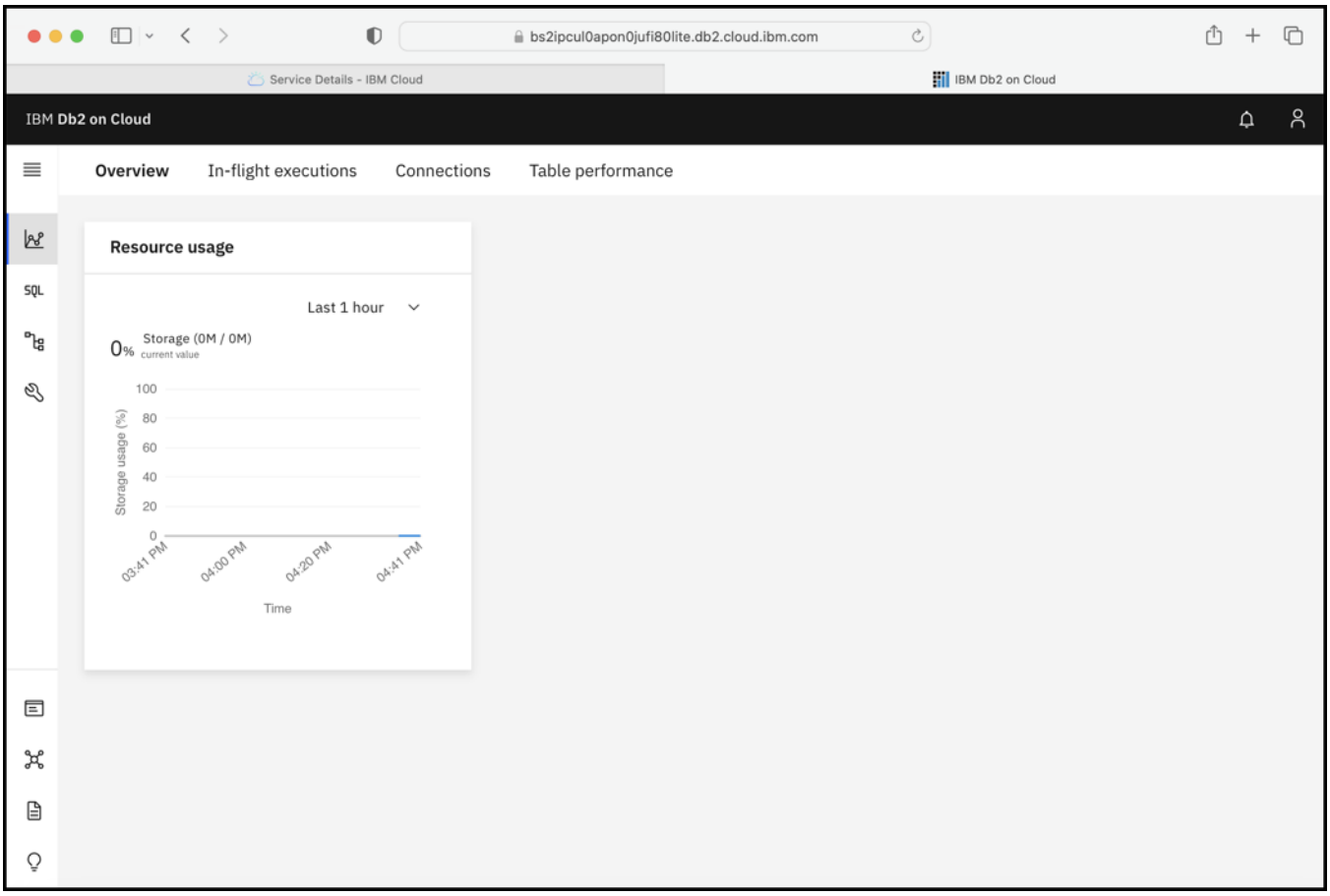


____ 5. Access the Db2 on Cloud database created earlier by clicking on the database ID shown in the **Resource list**. (In this example, the resource is the Db2 on Cloud database that has the ID *Db2-pc*).

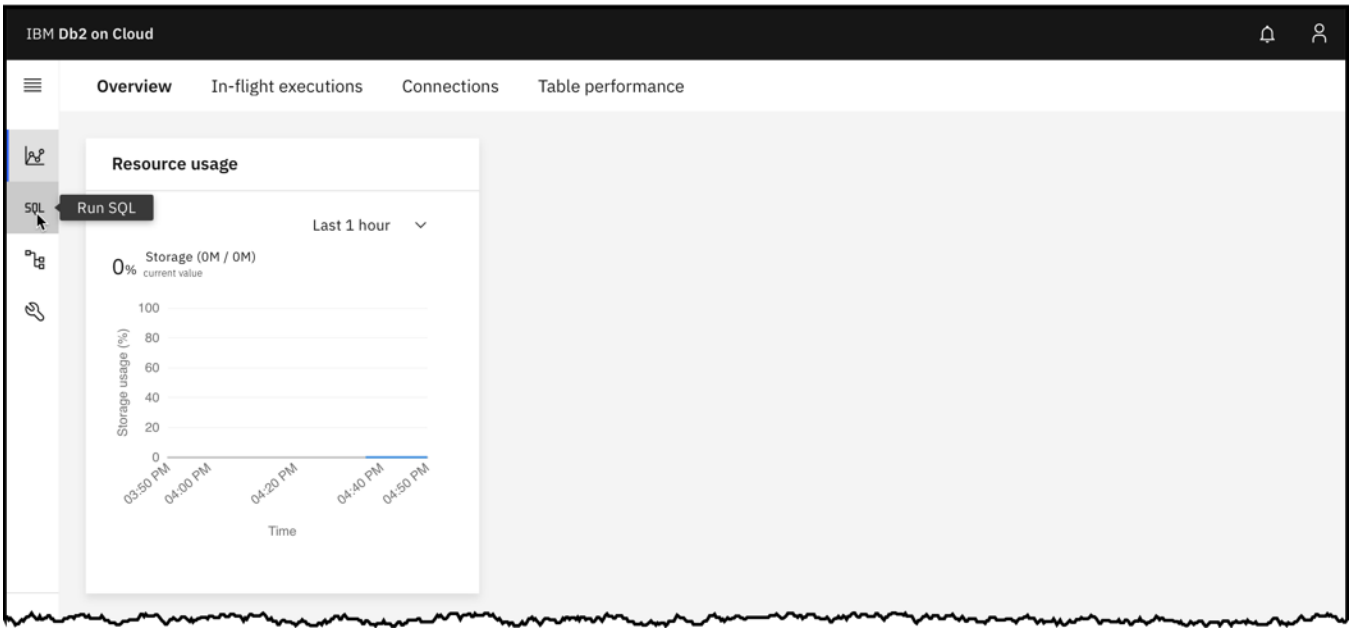
____ 6. As we saw earlier, clicking on a resource ID in the **Resource list** will cause the *Resource list* screen to be replaced with a high-level “management” screen for the resource selected. When the “management” screen for the Db2 on Cloud database resource selected appears, click the **Go to UI**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



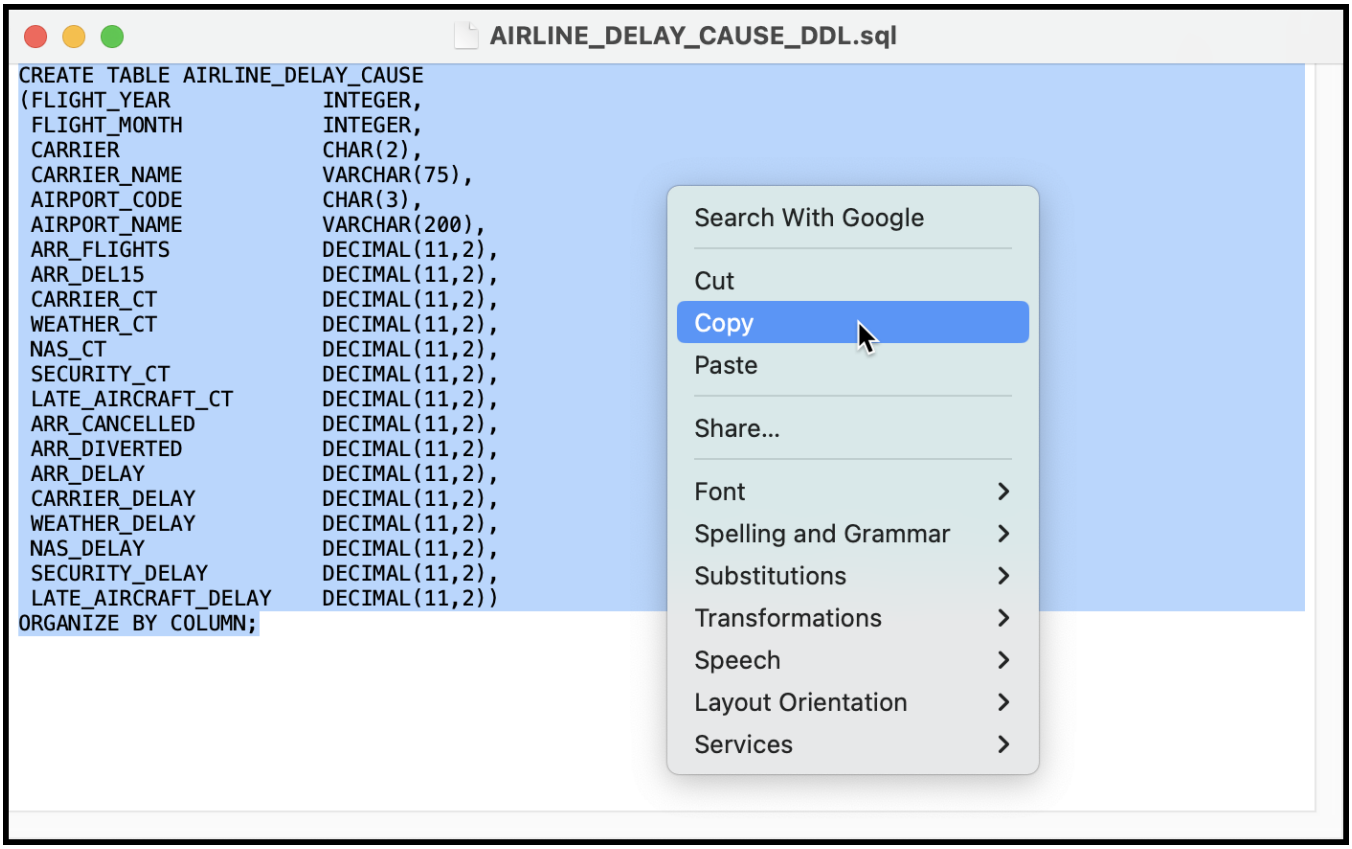
____ 7. Clicking on the **Go to UI** button will cause a new web browser tab named **IBM Db2 on Cloud** to be opened. This tab usually contains an **Overview** screen, which is the default starting point for the Db2 on Cloud User Interface.



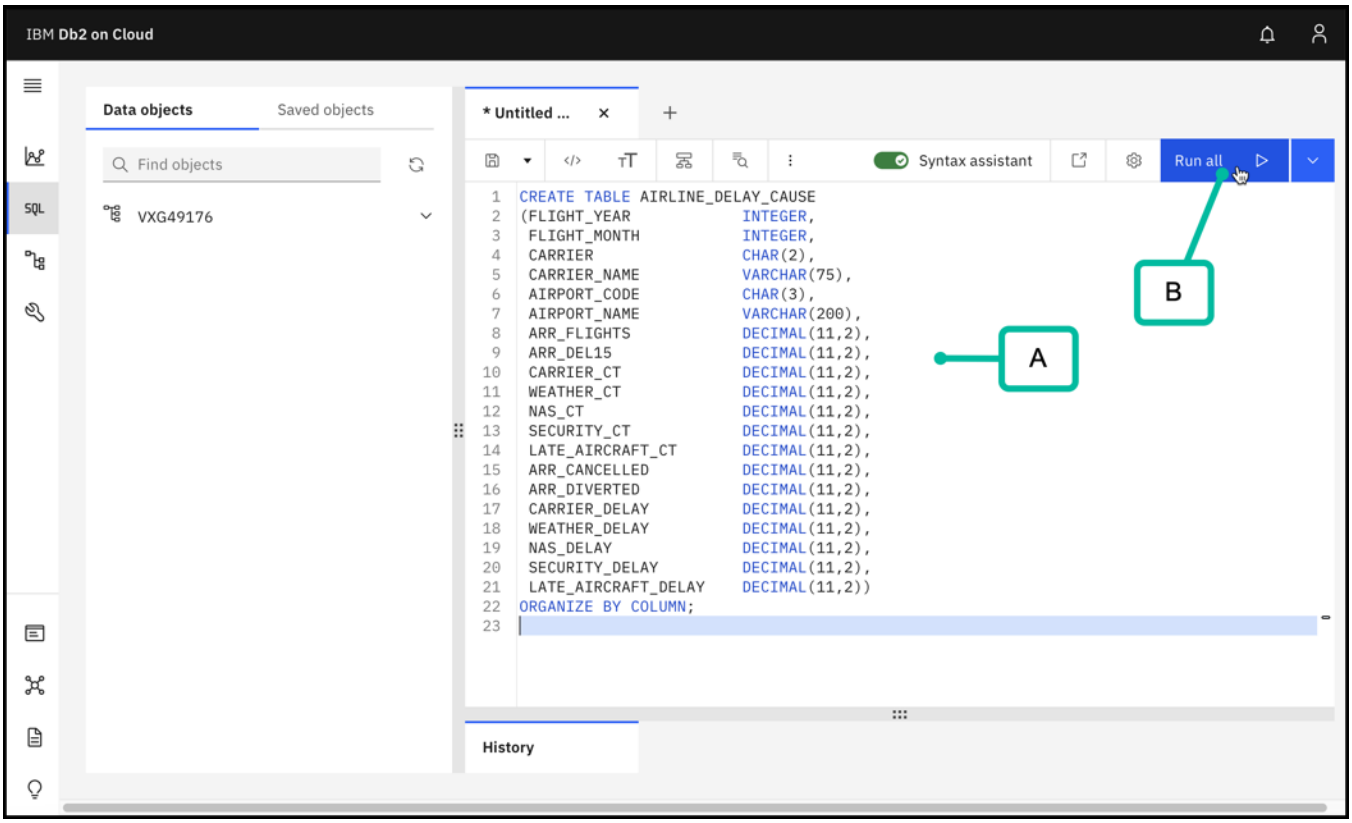
____ 8. Switch to the *Run SQL* screen by clicking the **SQL** icon found in the menu shown on the left-hand side of the screen.



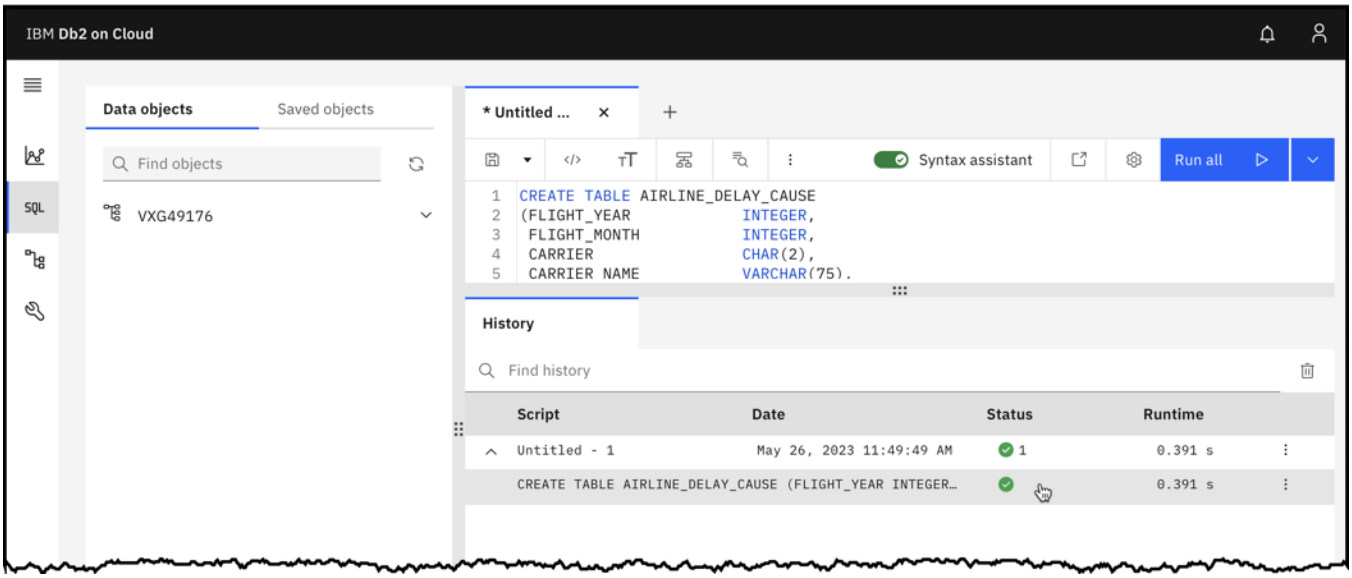
____ 9. Minimize your web browser and open the file named **AIRLINE_DELAY_CAUSE_DDL.sql** with a text editor. Then, select everything found in the file and copy it to the clipboard. (In this example, the file was opened with the **TextEdit** application on a MacBook Pro. After everything in the file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



____ 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.

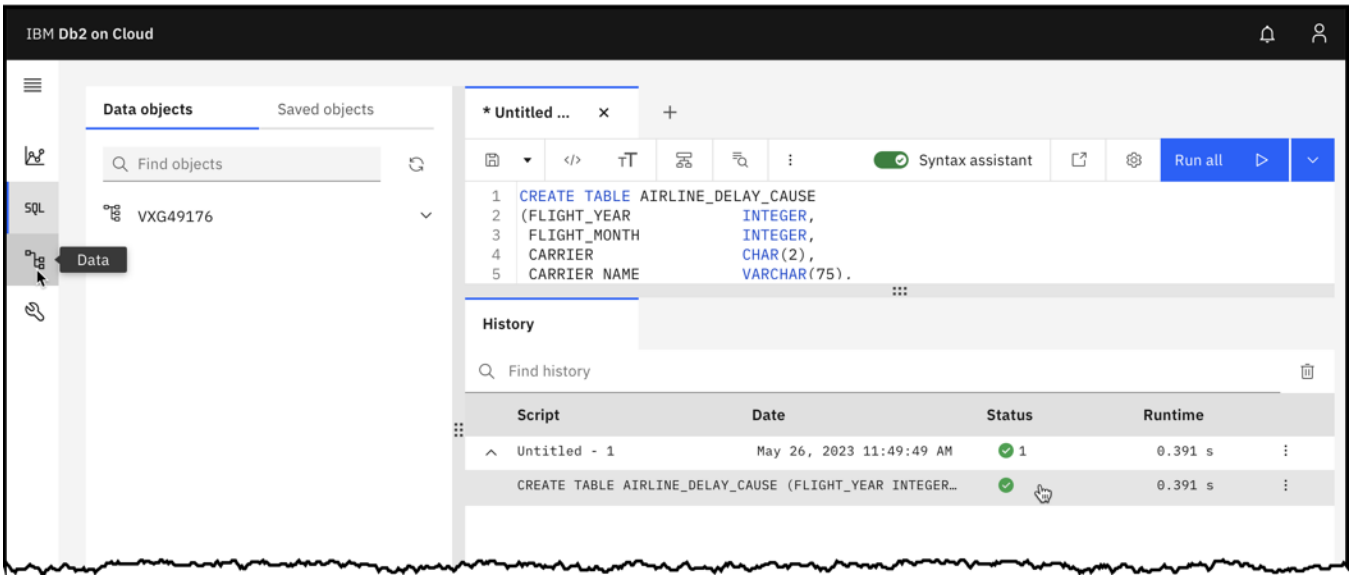


____ 11. Check the **History** window/tab located at the bottom of the screen. If the **Status** column for the record containing the text "**CREATE TABLE AIRLINE_DELAY_CAUSE …**" has a green circle with a check mark in it, a table named **AIRLINE_DELAY_CAUSE** was successfully created. If an error occurred, 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**.)

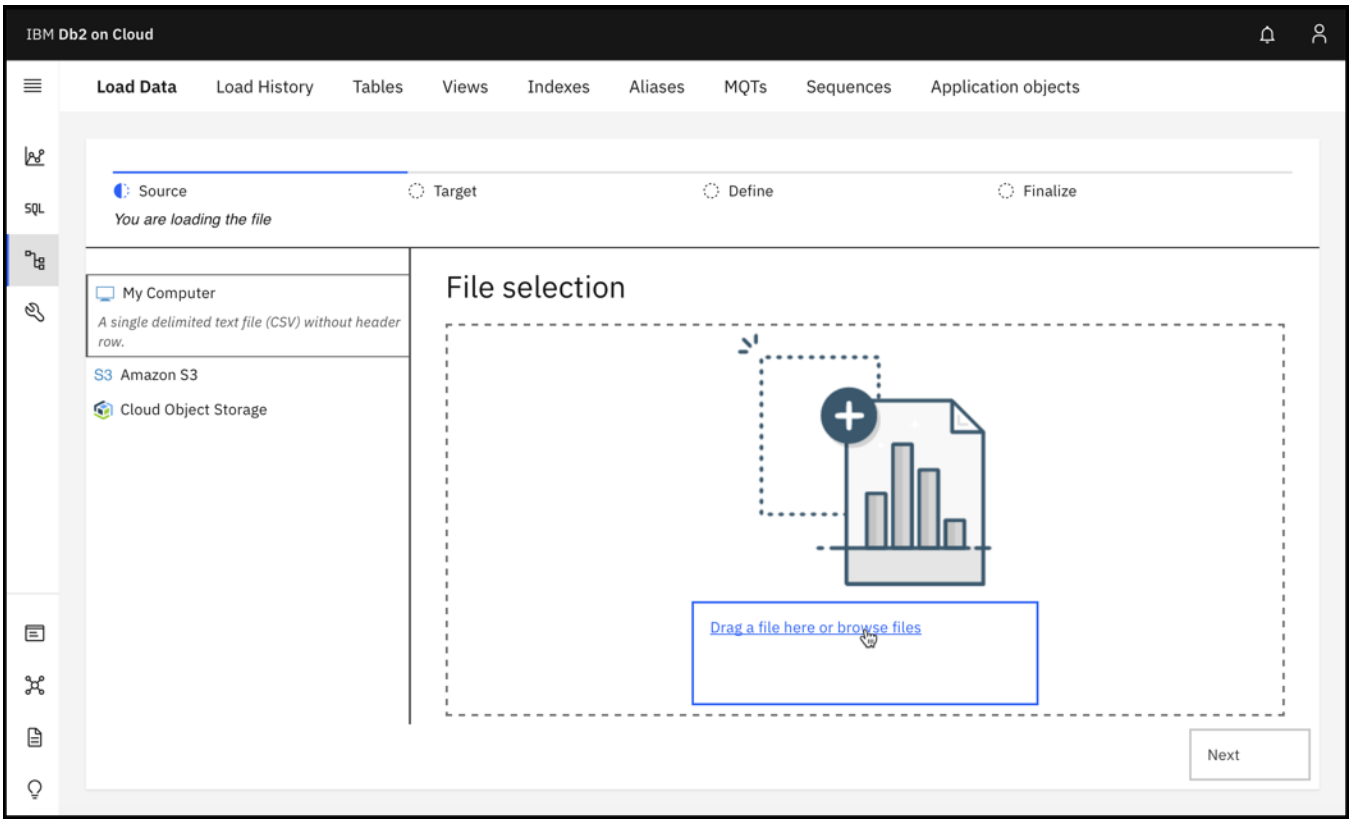


2. Populate the AIRLINE_DELAY_CAUSE table.

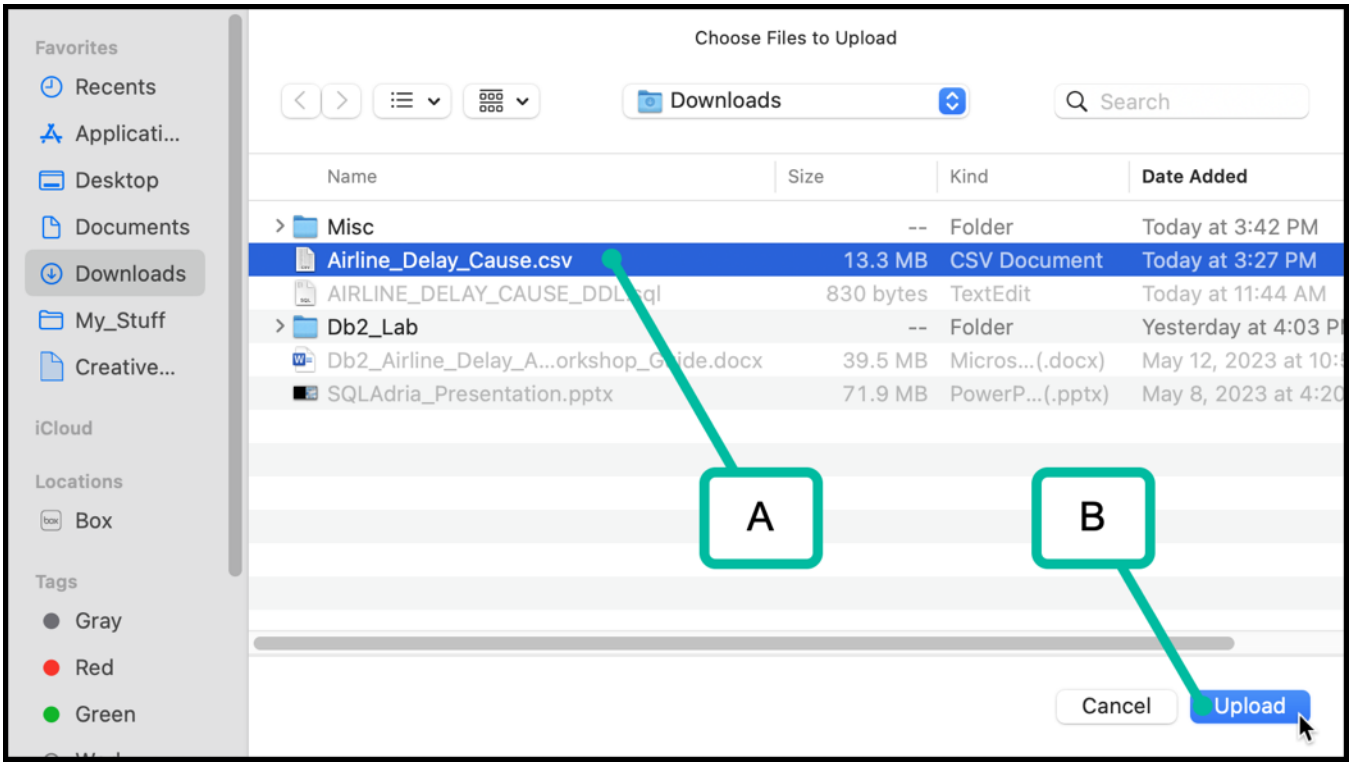
____ 1. Switch to the *Data* screen by clicking the **Data** icon (located immediately below the **SQL** icon) found in the menu shown on the left-hand side of the screen.



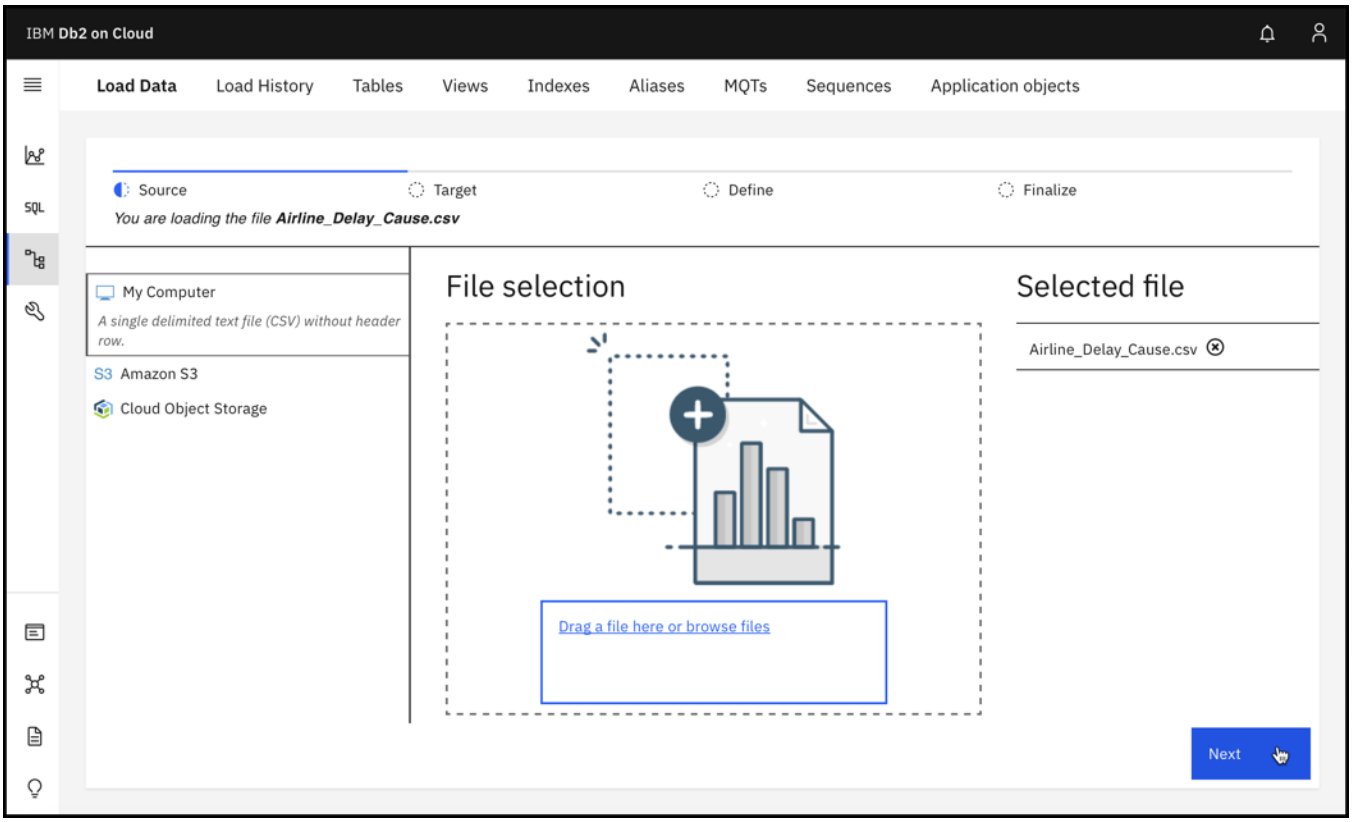
____ 2. When you go to the *Data* screen, the first thing you see is the *Load Data* window. From this screen, click on the **Drag a file here or browse files** link to open a *Finder* window on Mac or a *File Explorer* window on Windows.



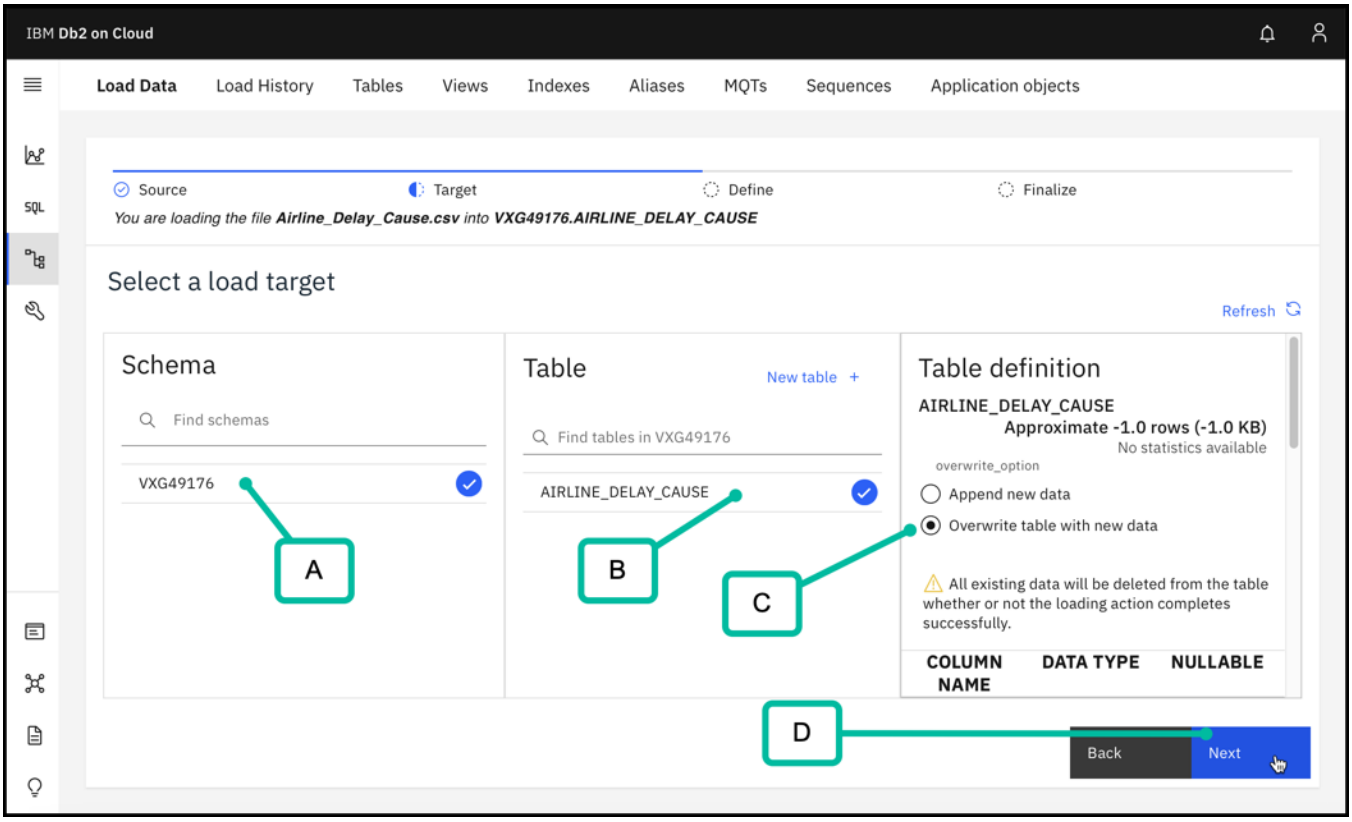
____ 3. **[A]** Select the file named **Airline_Delay_Cause.csv** and **[B]** click the **Upload** button.



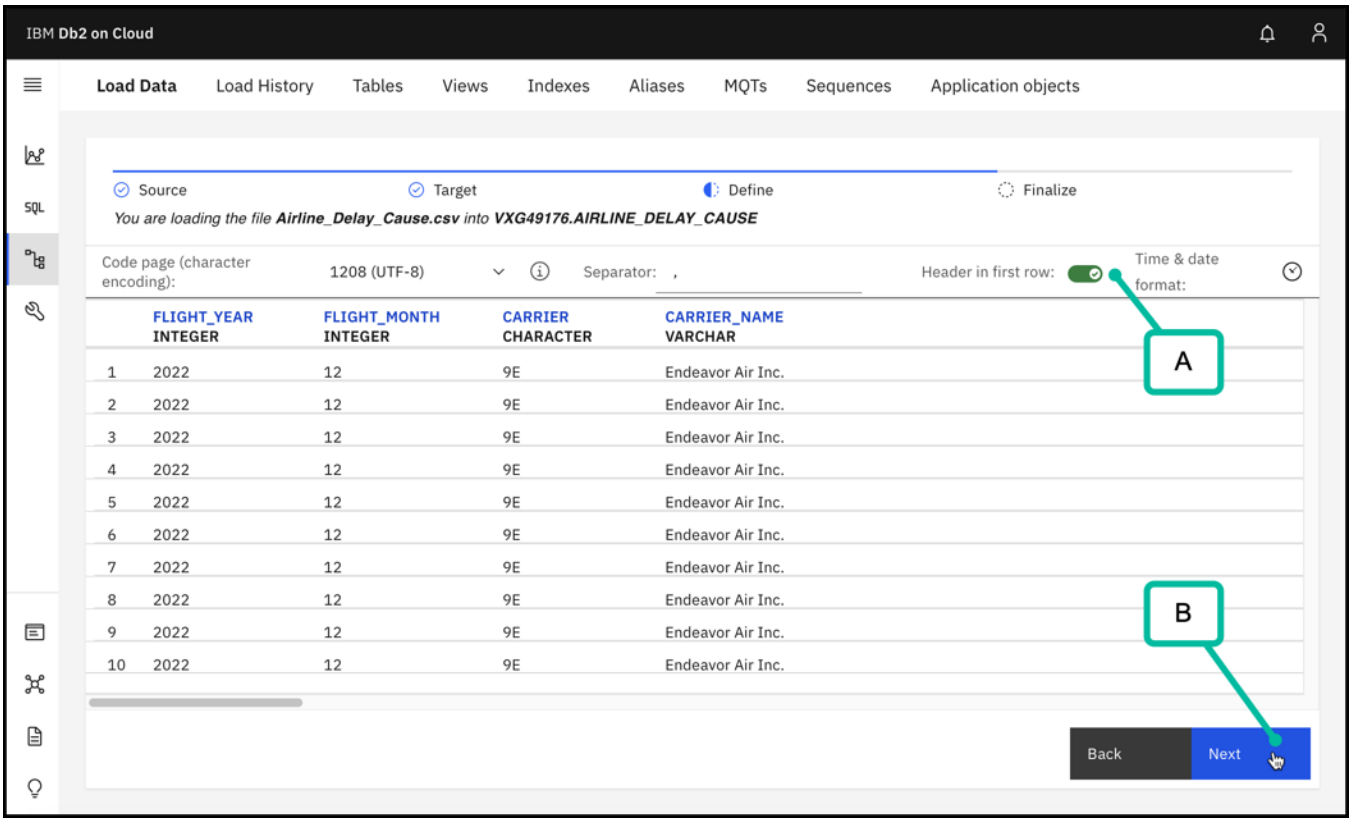
____ 4. Verify that the file name **Airline_Delay_Cause.csv** appears under the **Selected file** heading on the *Load Data* screen. Then, click the **Next** button located in the bottom right-hand corner of the screen.



____ 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.

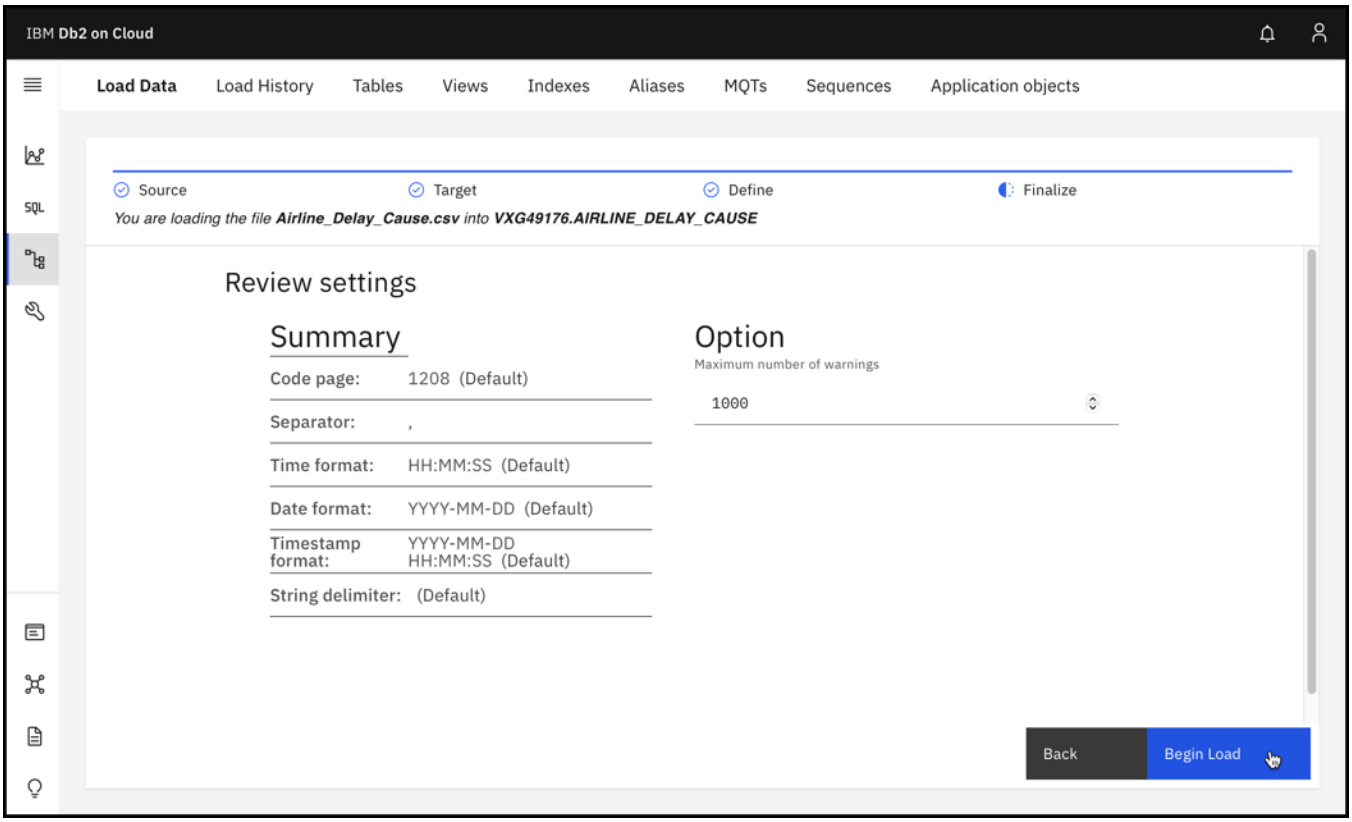


____ 6. Verify that the **Airline_Delay_Cause.csv** file used contains valid data. Then, **[A]** make sure the **Header in first row** switch is on (green) and **[B]** click the **Next** button.

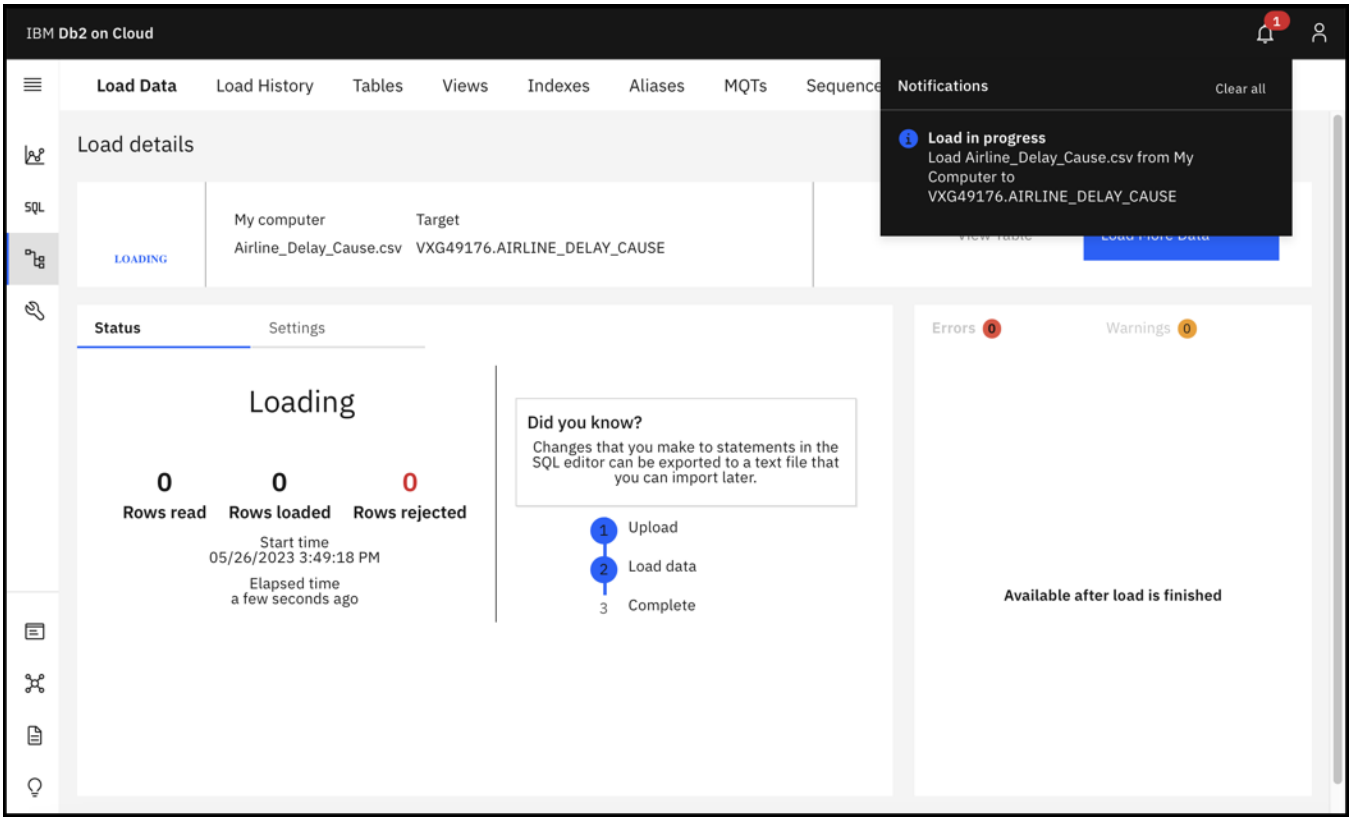


	FLIGHT_YEAR INTEGER	FLIGHT_MONTH INTEGER	CARRIER CHARACTER	CARRIER_NAME VARCHAR
1	2022	12	9E	Endeavor Air Inc.
2	2022	12	9E	Endeavor Air Inc.
3	2022	12	9E	Endeavor Air Inc.
4	2022	12	9E	Endeavor Air Inc.
5	2022	12	9E	Endeavor Air Inc.
6	2022	12	9E	Endeavor Air Inc.
7	2022	12	9E	Endeavor Air Inc.
8	2022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

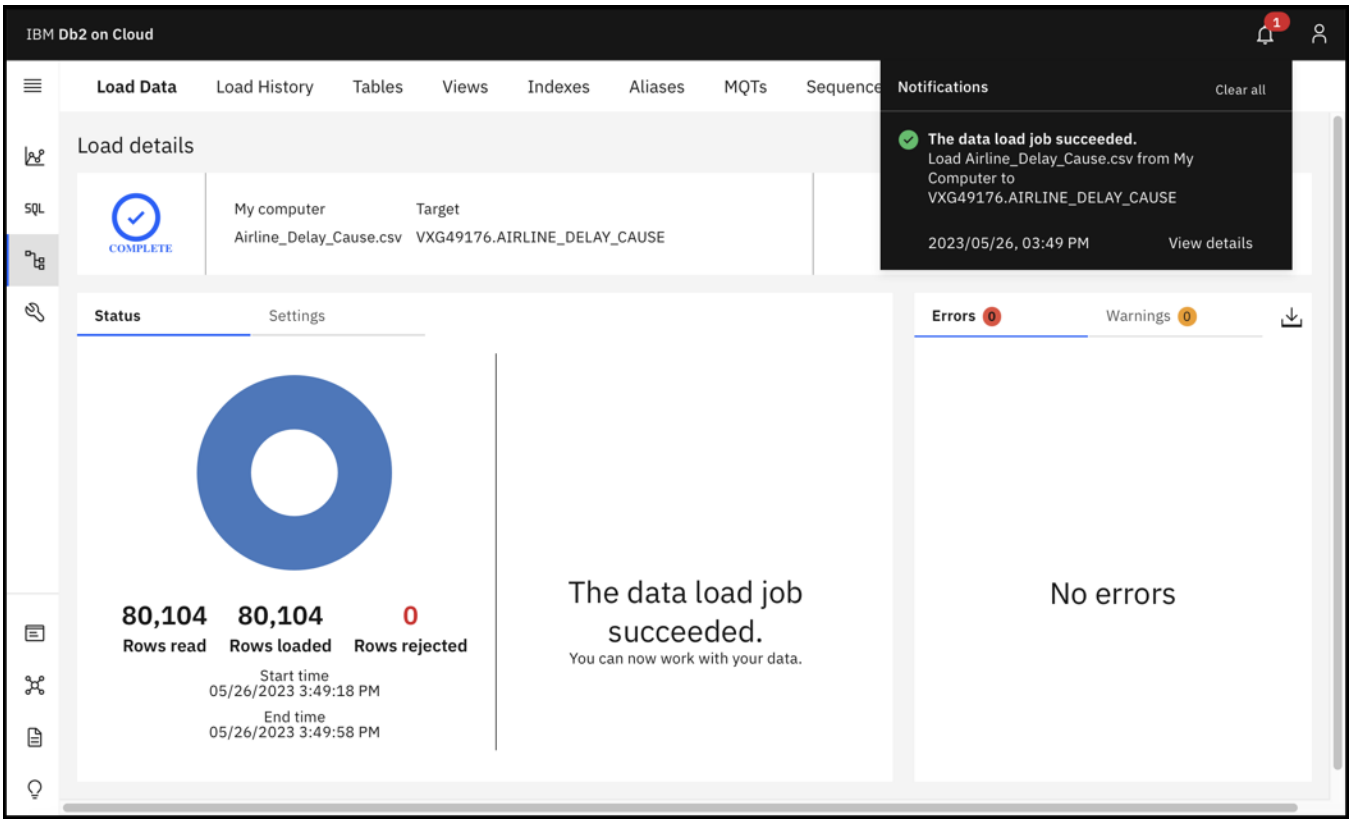
____ 7. Look over the *Review settings* screen to ensure everything is correct. Then, start the data load process by clicking the **Begin Load** button.



____ 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).



____ 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 succeed** (or **failed**).



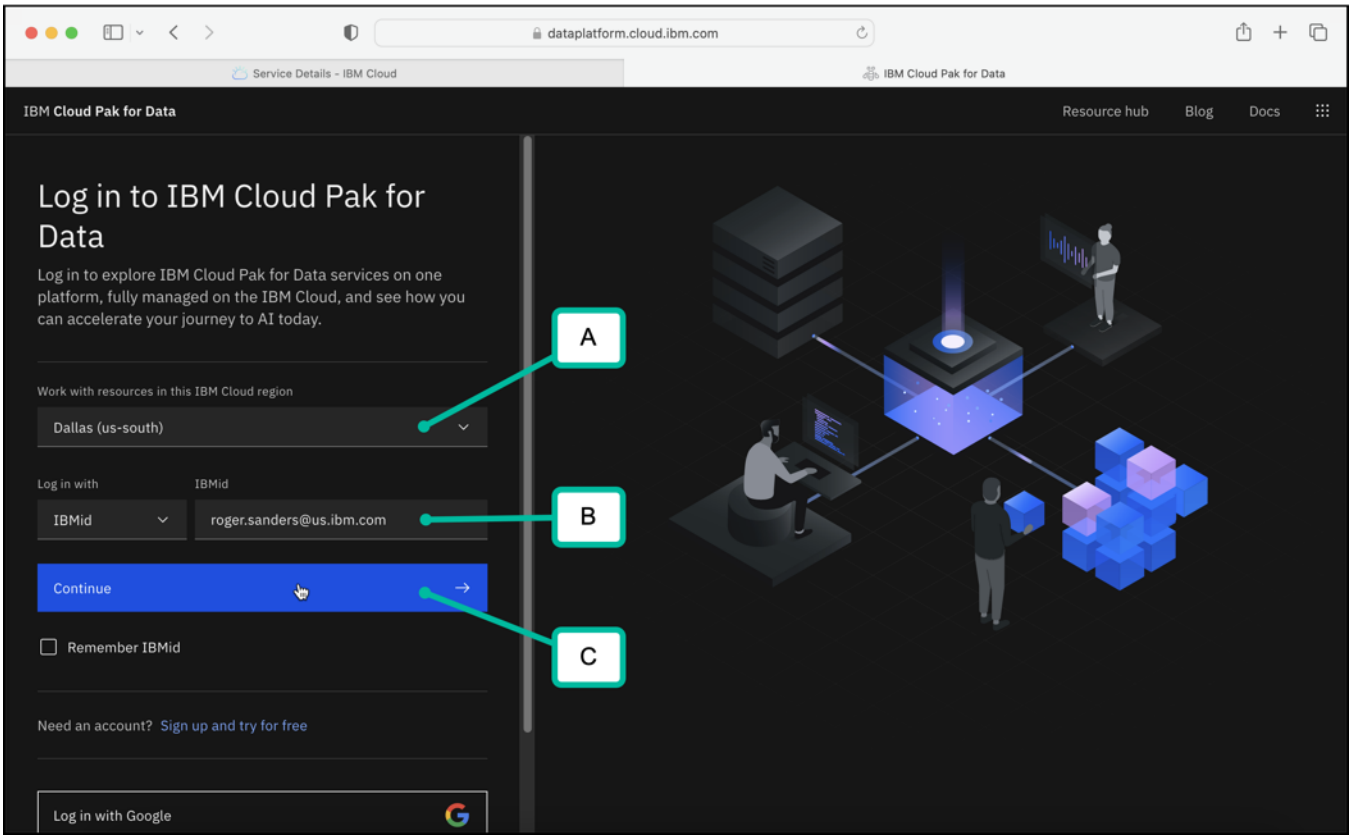
Congratulations! You have just created and populated a table named **AIRLINE_DELAY_CAUSE** in a Db2 on Cloud database. We will use the data in this table in another exercise later.

IV. Using IBM Watson Studio to run a Jupyter Notebook

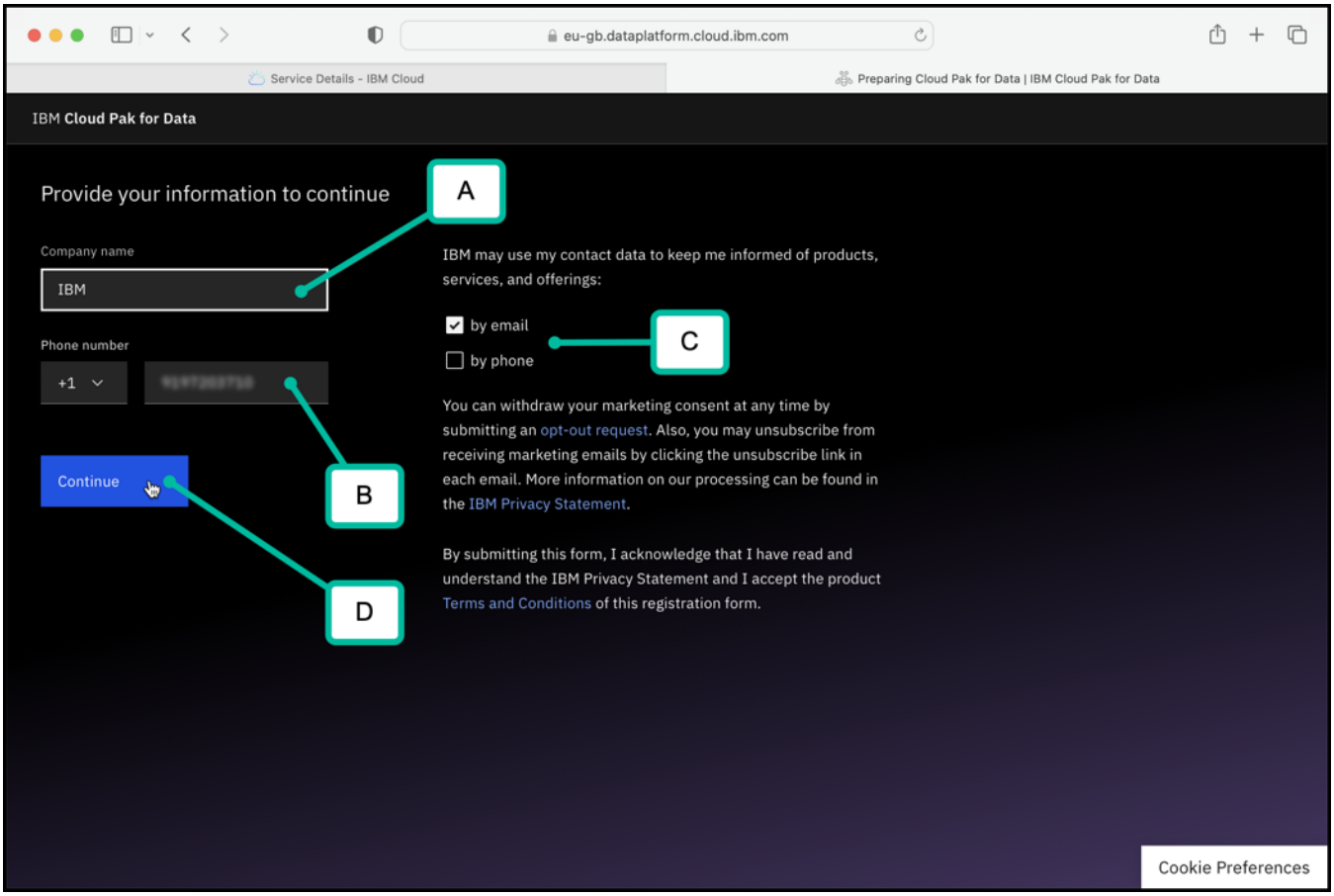
The free version of IBM Watson Studio (offered with IBM Cloud Pak for Data as a Service) will be used to execute the Jupyter Notebooks that are needed to complete the remaining exercises in this workshop. So, in this section you will set up a free IBM Cloud Pak for Data as a Service account and then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

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

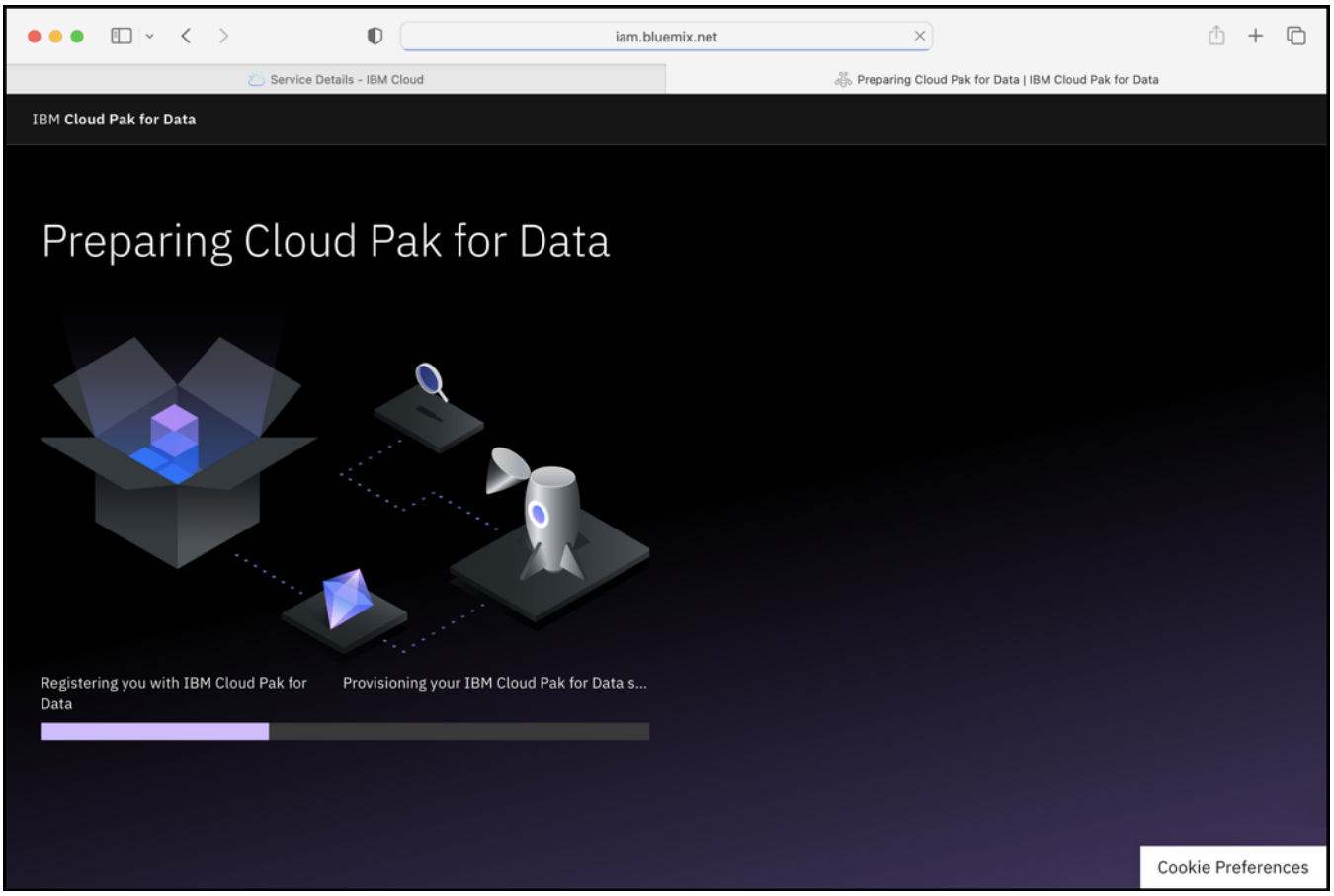
____ 1. Open a web browser and go to the *Log in to BM Cloud Pak for Data* web site (https://dataplatform.cloud.ibm.com/login). Once there, **[A]** click on the **Work with resources in this IBM Cloud region** field to display a dropdown list of locations where Cloud Pak for Data as a Service can be 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.



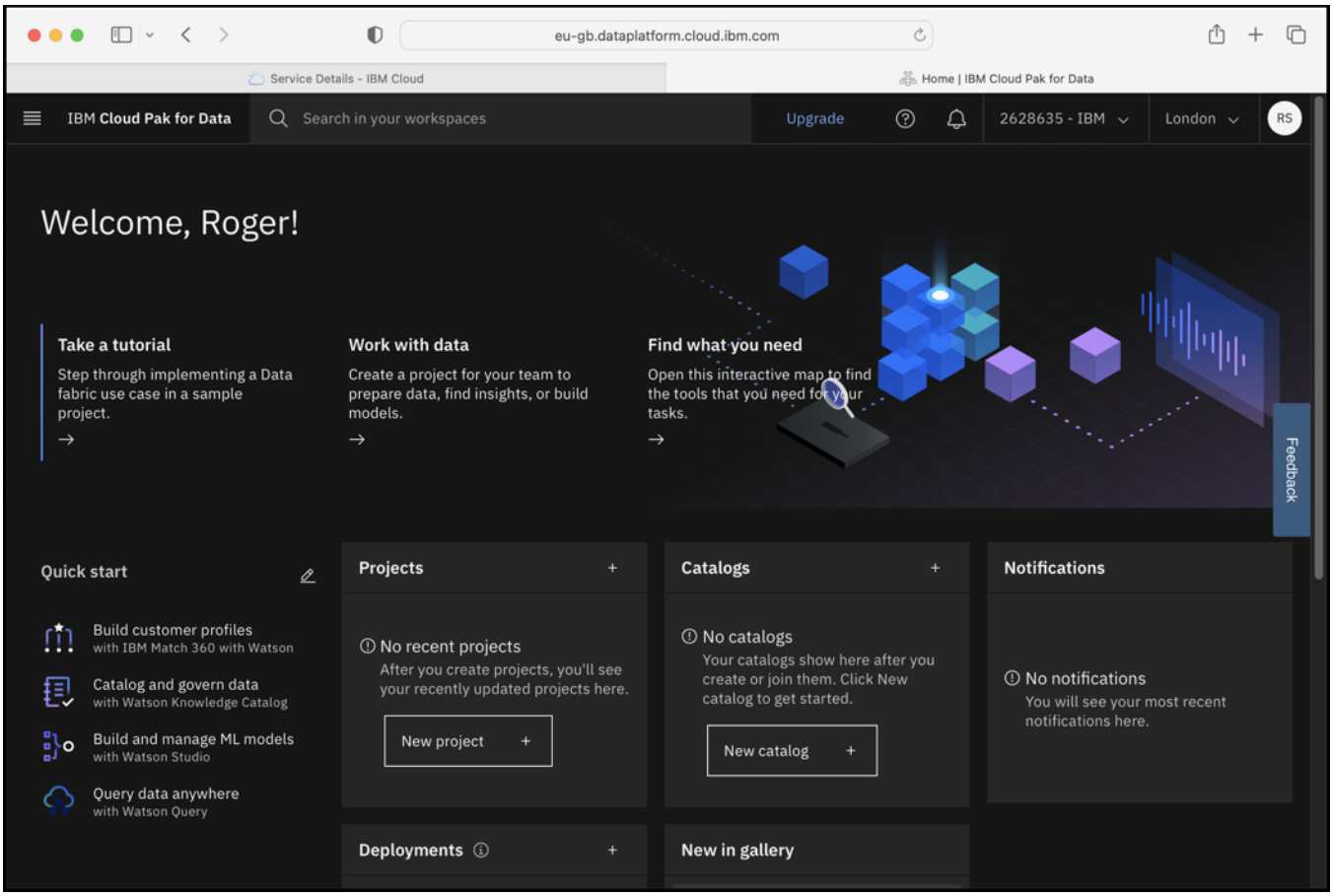
____ 2. When the *Provide your information to continue* screen appears, **[A]** enter your company name in the **Company name** field and **[B]** your telephone number in the **Phone number** field. (Click on the **Phone number** dropdown list and select the correct country code if your country code 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.



____ 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:



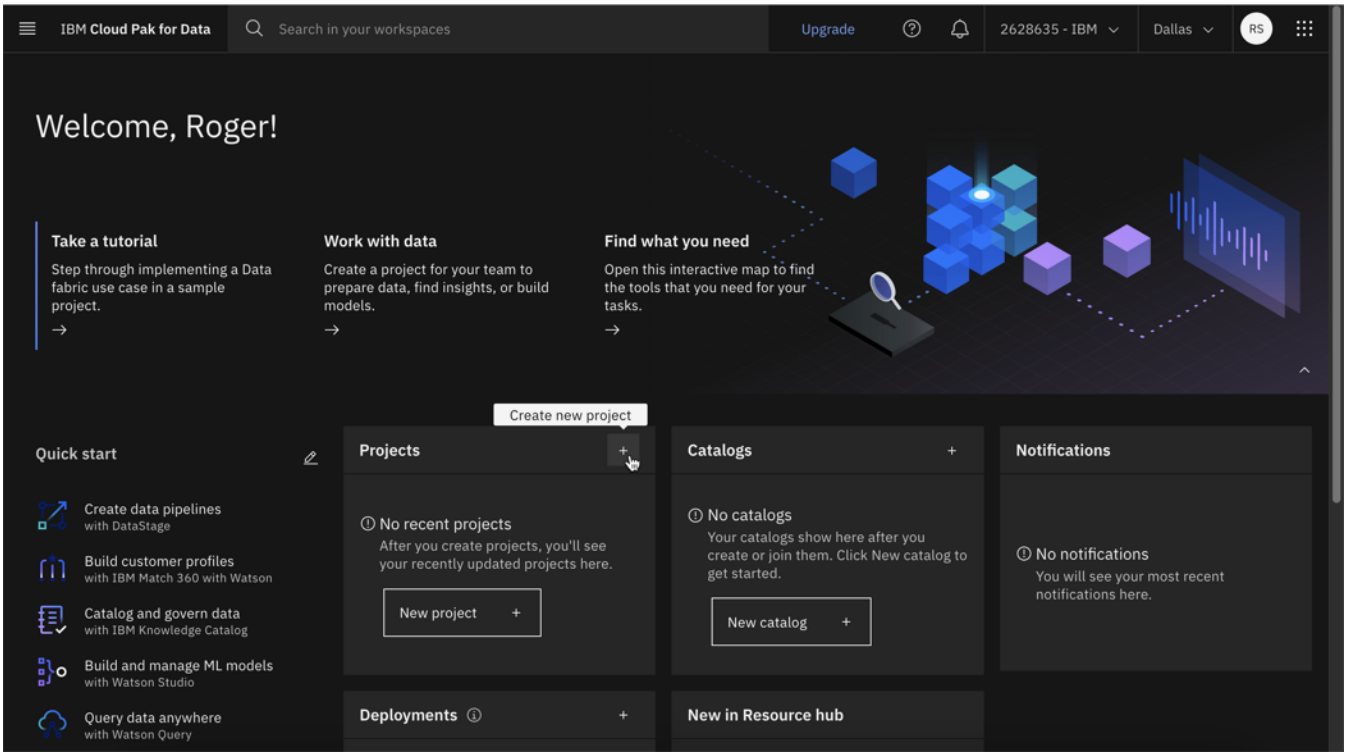
____ 4. Once your IBM Cloud Pak for Data as a Service environment has been created, you should see a *Welcome* screen that looks something like this:



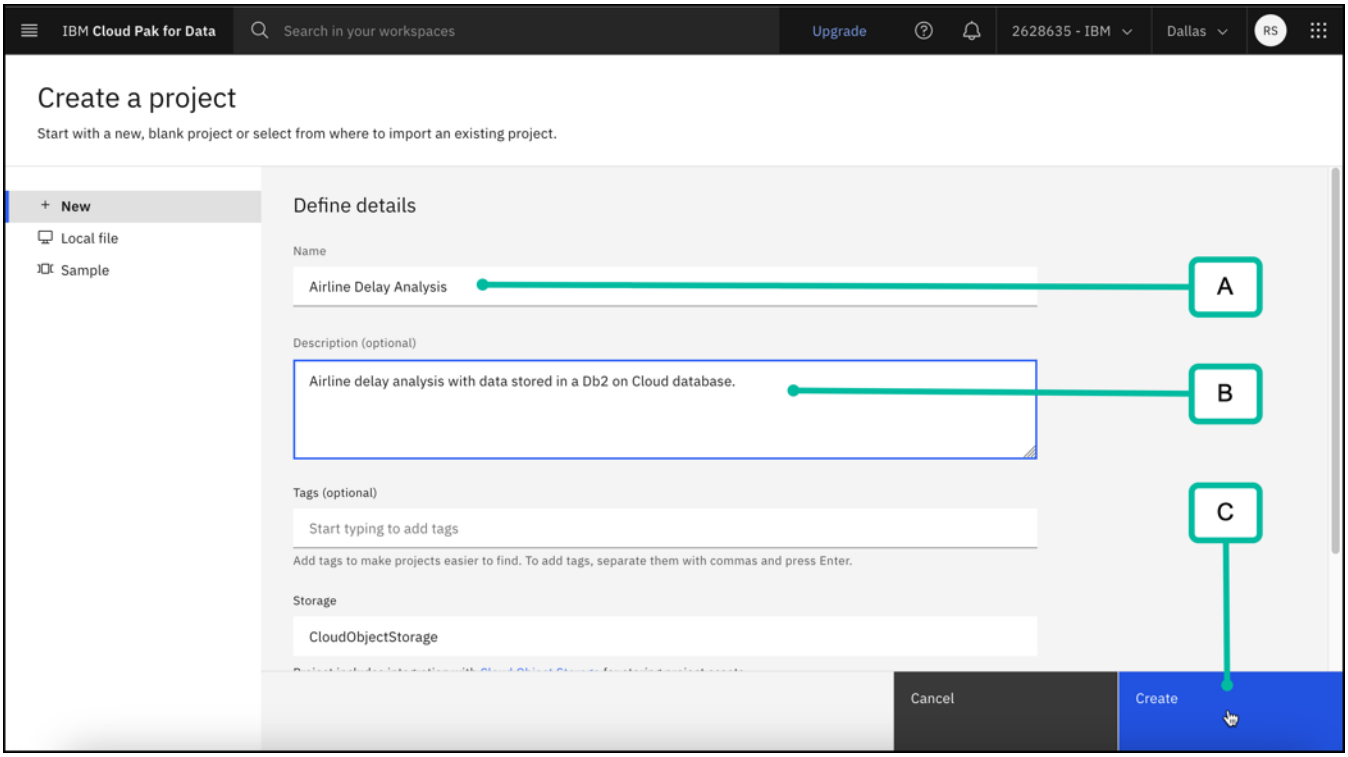
Congratulations! You now have an active **IBM Cloud Pak for Data as a Service** account that can be used to perform the rest of the exercises in this workshop.

2. Create a new project.

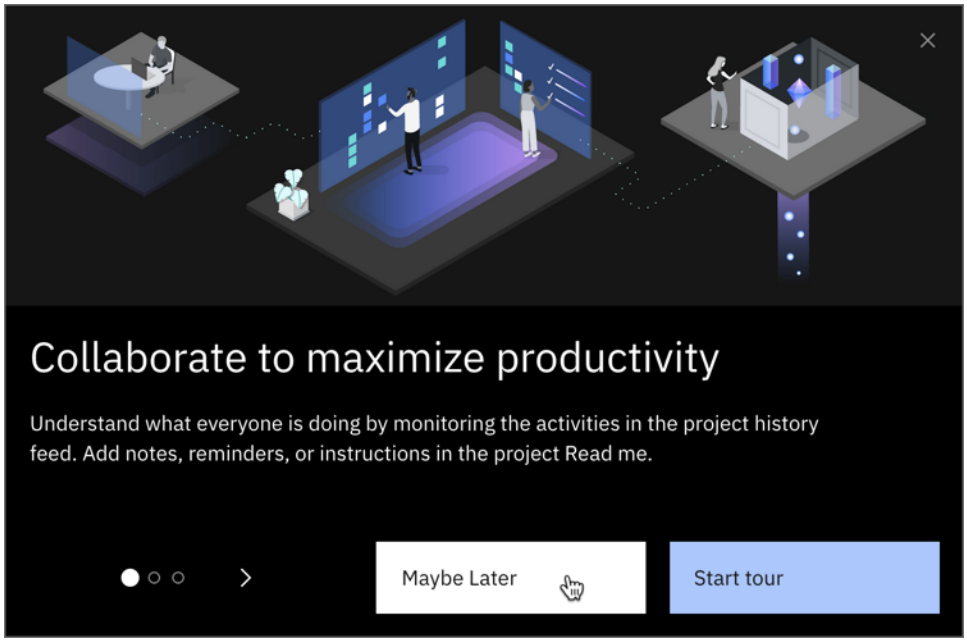
____ 1. Create a new project by clicking on the **+** icon located in the top right-hand corner of the **Projects** box. This will cause the IBM Cloud Pak for Data as a Service *Welcome* screen to be replaced with a *Create a project* screen.



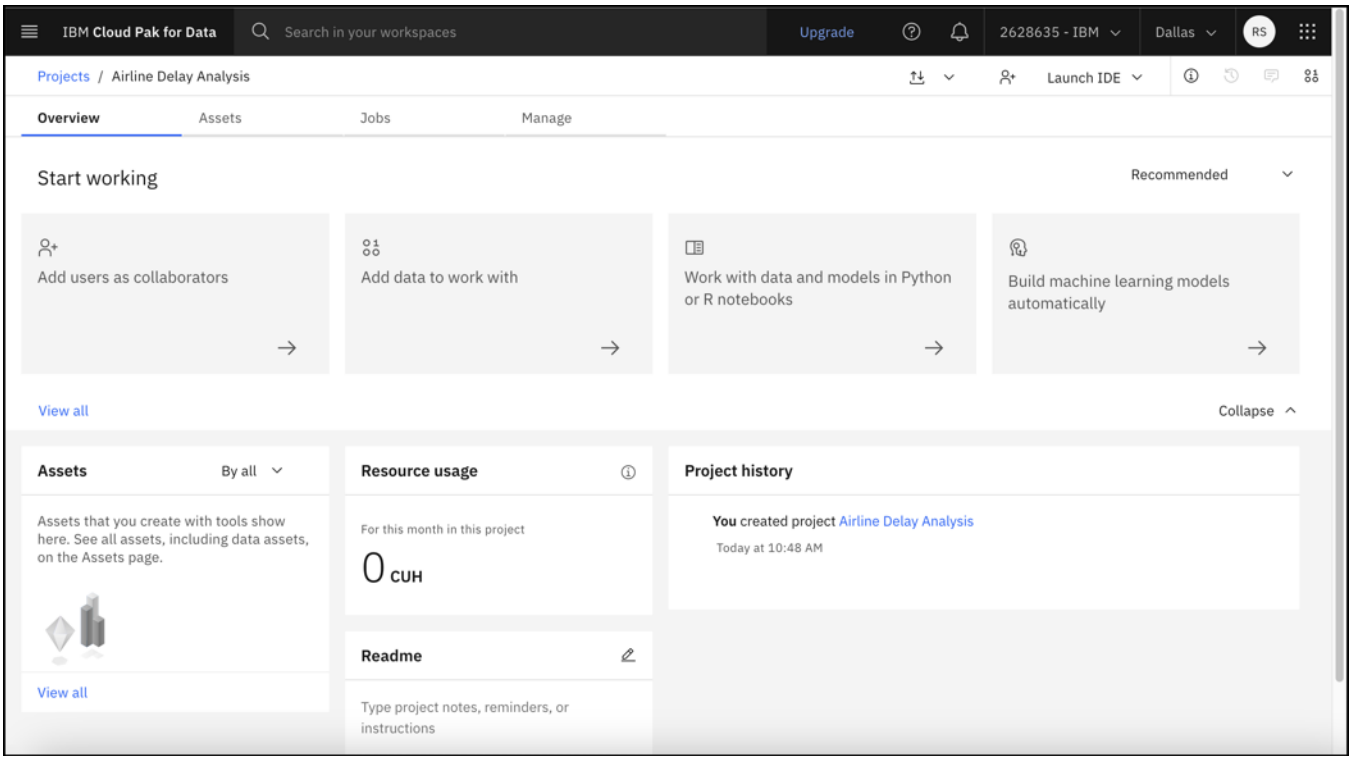
____ 2. When the *Create a project* screen is displayed, **[A]** enter a project name in the **Name** field. Then, **[B]** enter a project description in the **Description** field (*optional*). In this example, the project name "*Airline Delay Analysis"* and the description "*Airline delay analysis with 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.

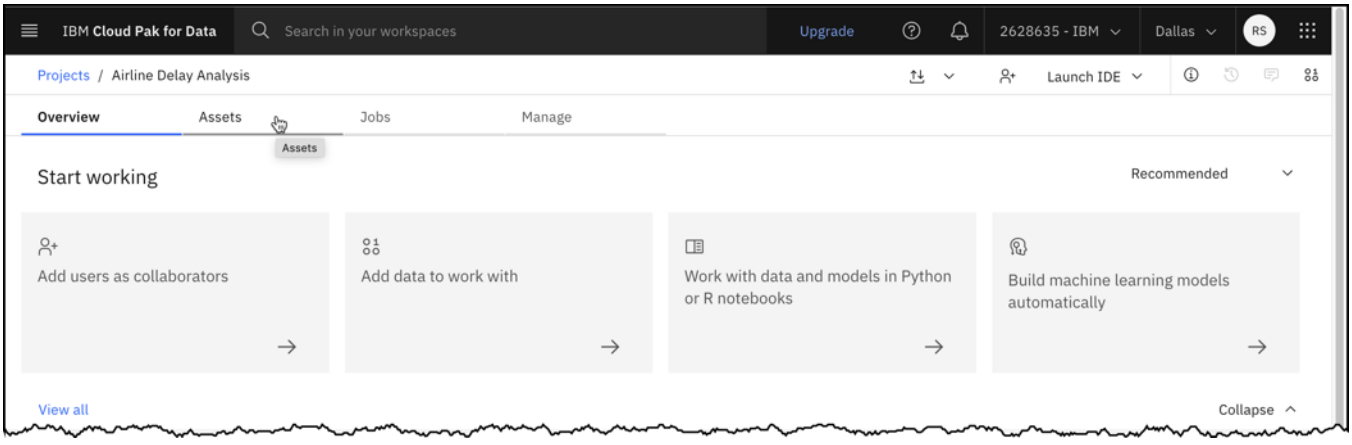


____ 4. Once the new project has been created, you should see a *Project Overview* screen that looks something like this:

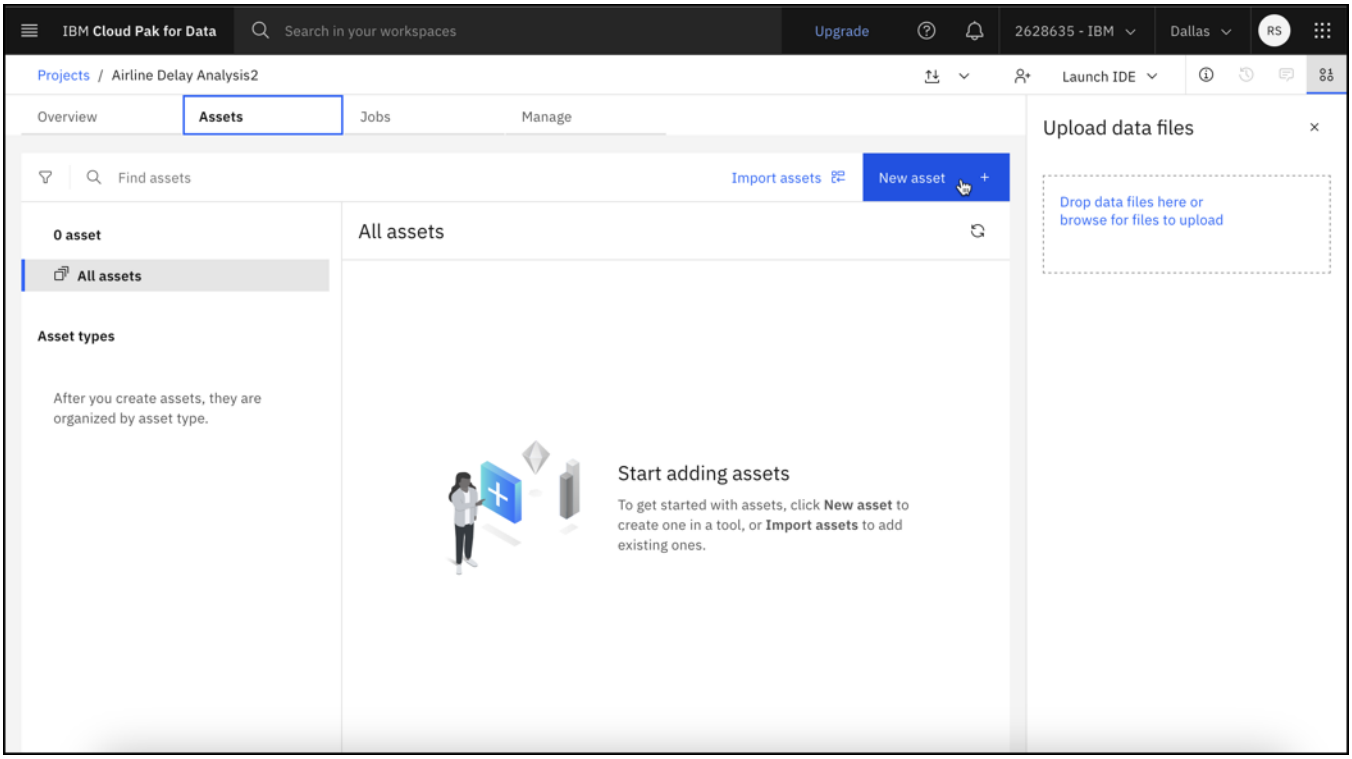


3. Add a Jupyter Notebook to the project.

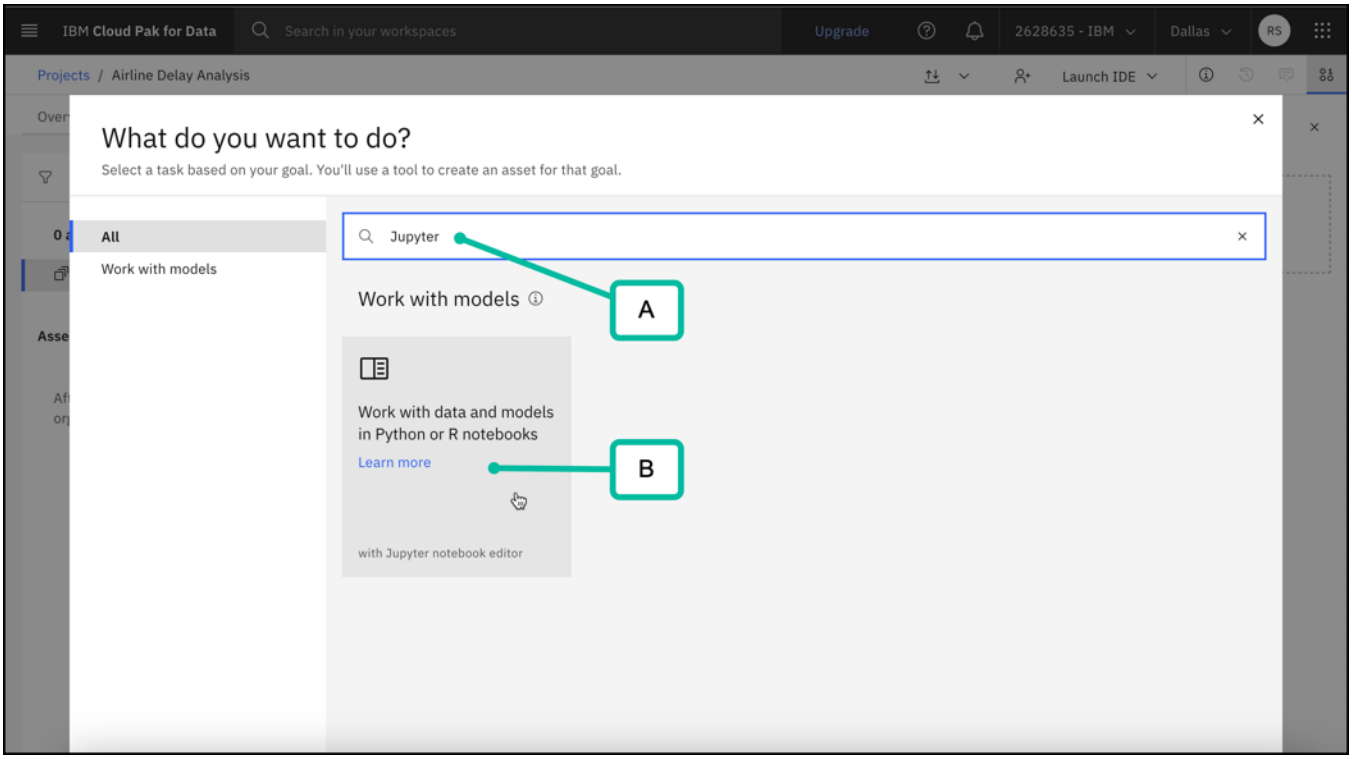
____ 1. From the *Project Overview* screen, click on the **Assets** tab.



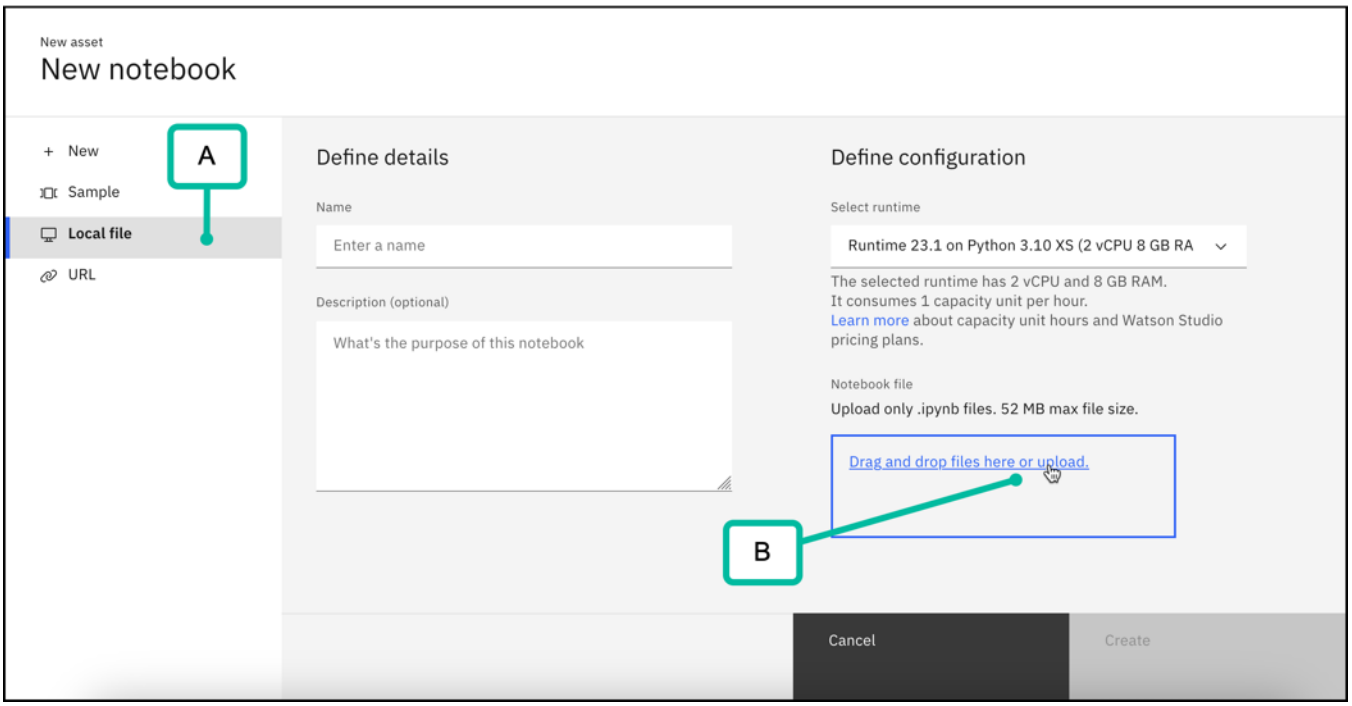
____ 2. When the *Assets* screen is displayed, click on the **New asset +** button.



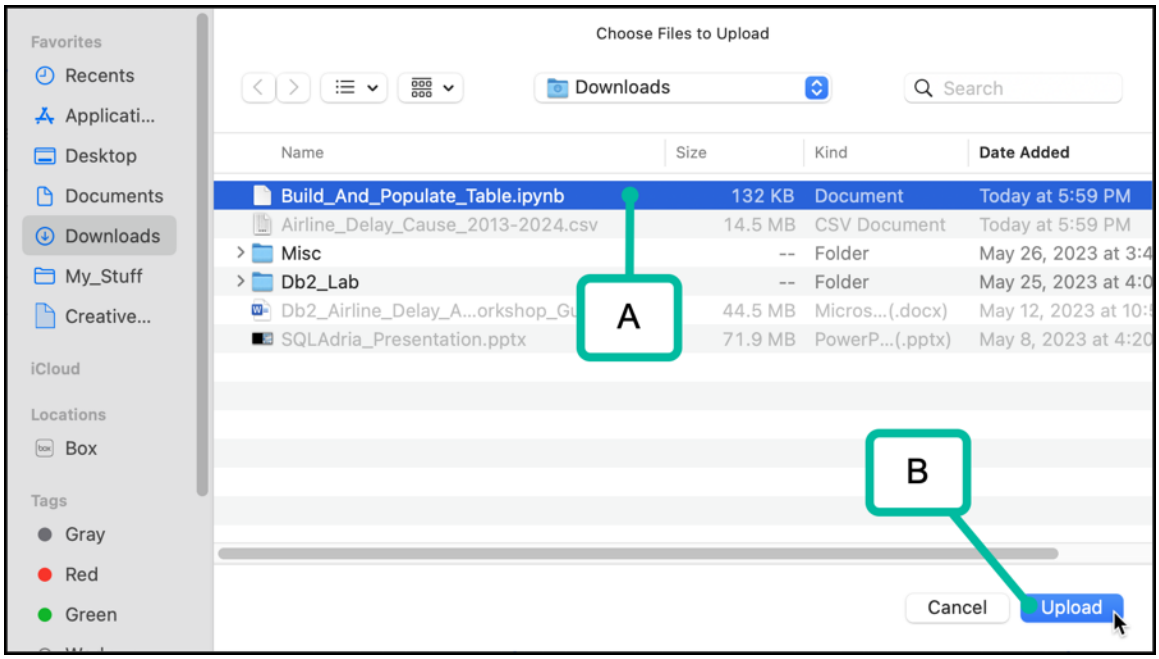
____ 3. From the *What do you want to do* popup window, **[A]** enter the text “Jupyter” in the **Search** entry field (i.e., the entry field with the magnifying glass icon on the left). Then, **[B]** click on the **Work with data and models in Python or R notebooks** tile displayed. This will cause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



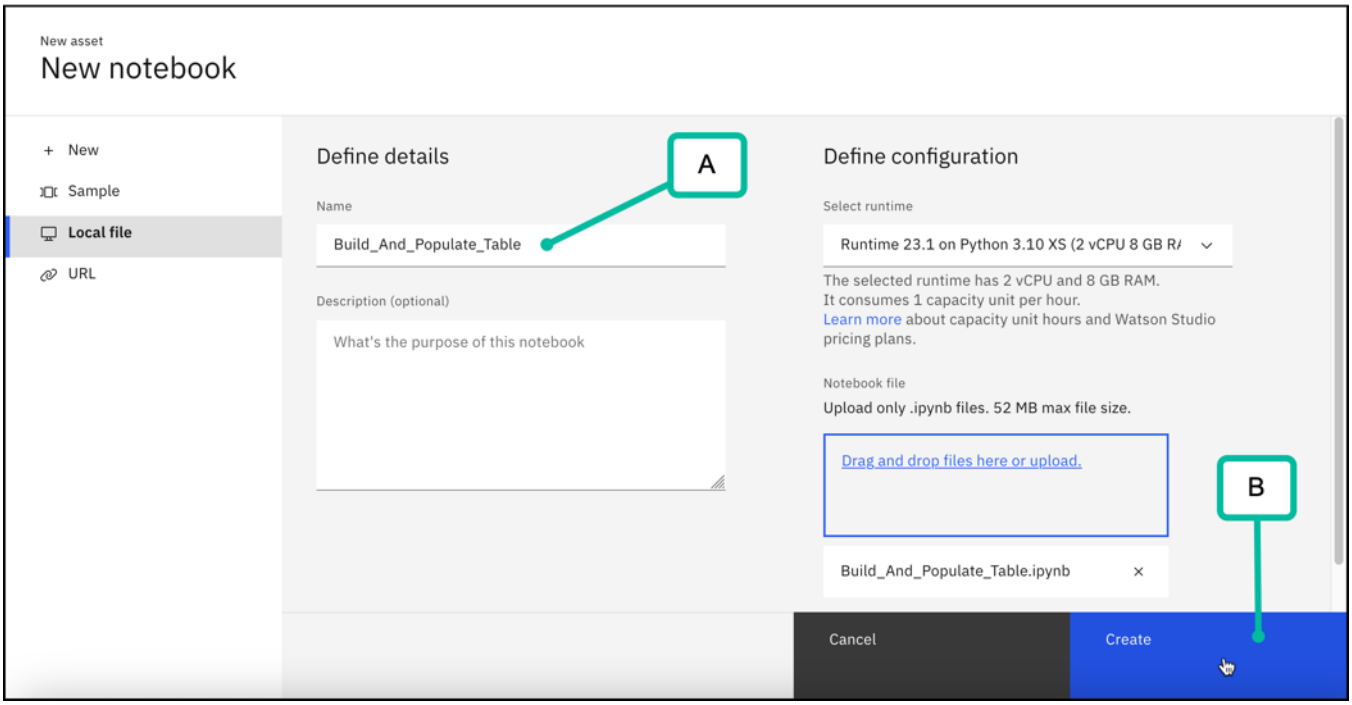
____ 4. From the *New notebook* screen, **[A]** click on the **Local file** item in the menu shown on the left-hand side of the screen. Then, **[B]** click on the **Drag and drop files here or upload** link to open a *Finder* window on Mac or a *File Explorer* window on Windows.



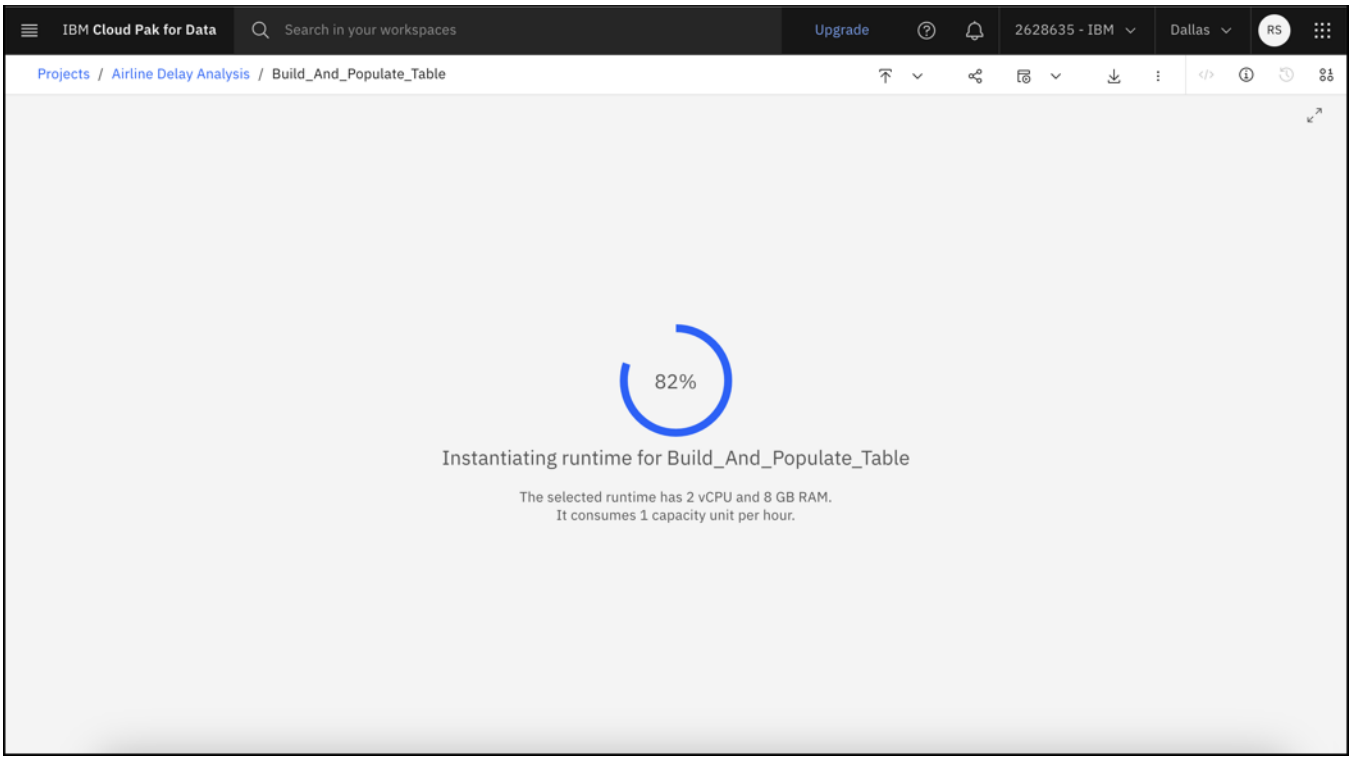
____ 5. **[A]** Select the file named **Build_And_Populate_Table.ipynb** and **[B]** click the **Upload** button.



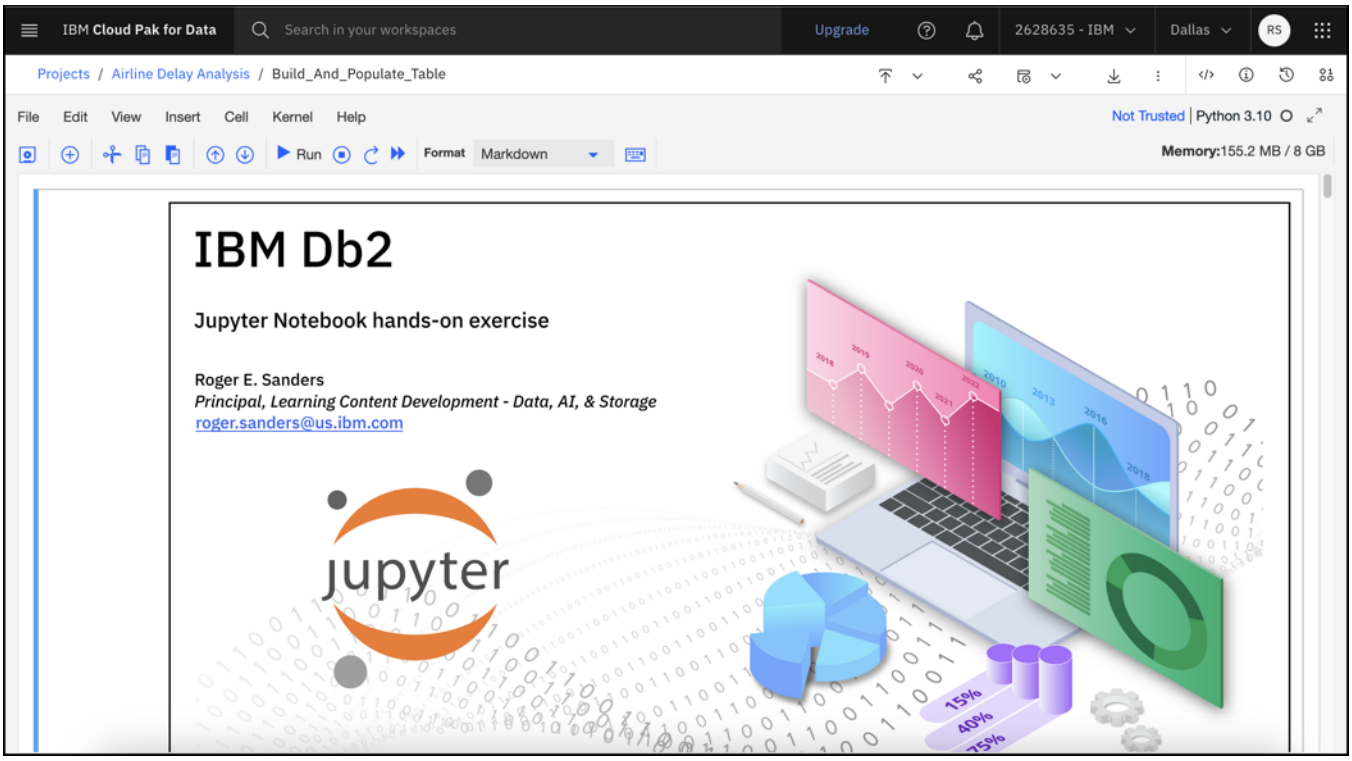
____ 6. **[A]** Confirm that the **Name** field on the *New notebook* screen contains the name of the file just selected, minus the file extension (**Build_And_Populate_Table**). Then, **[B]** click the **Create** button located at the bottom of the screen.



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

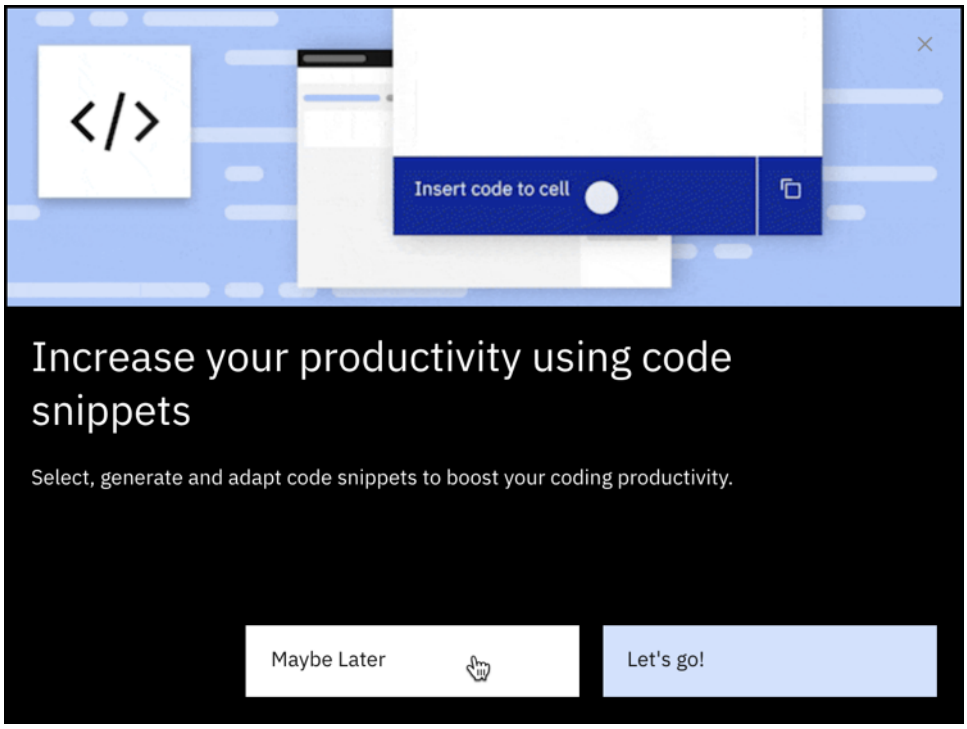


____ 8. Once the runtime for the Jupyter Notebook that was loaded has been instantiated, the *Instantiating runtime for* [*NotebookName*] screen should be replaced with a screen that looks similar to this:



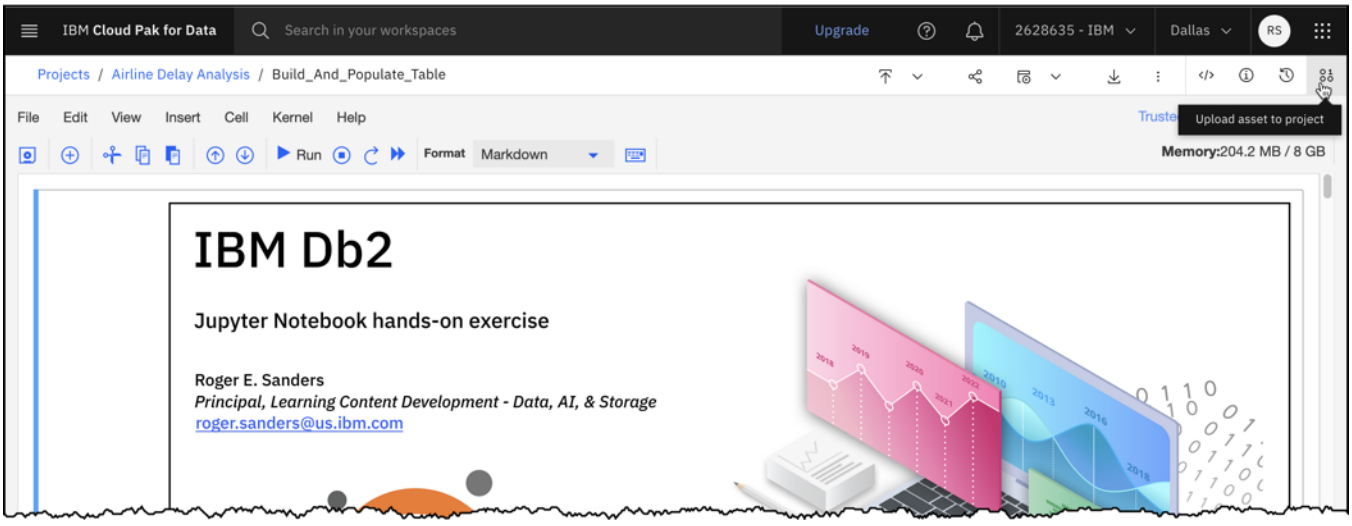
The Jupyter Notebook is now ready for editing (which we will do shortly).

____ 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.)

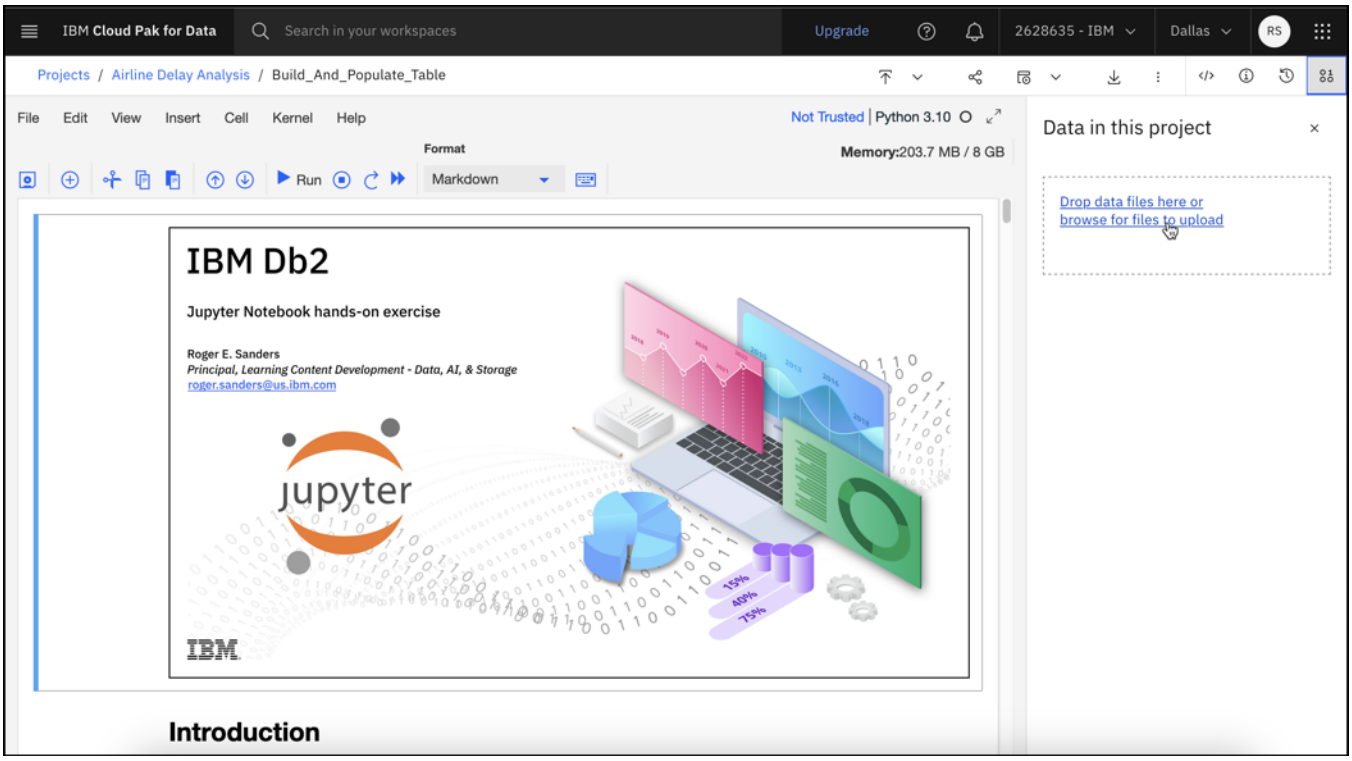


4. Add a data file to the project.

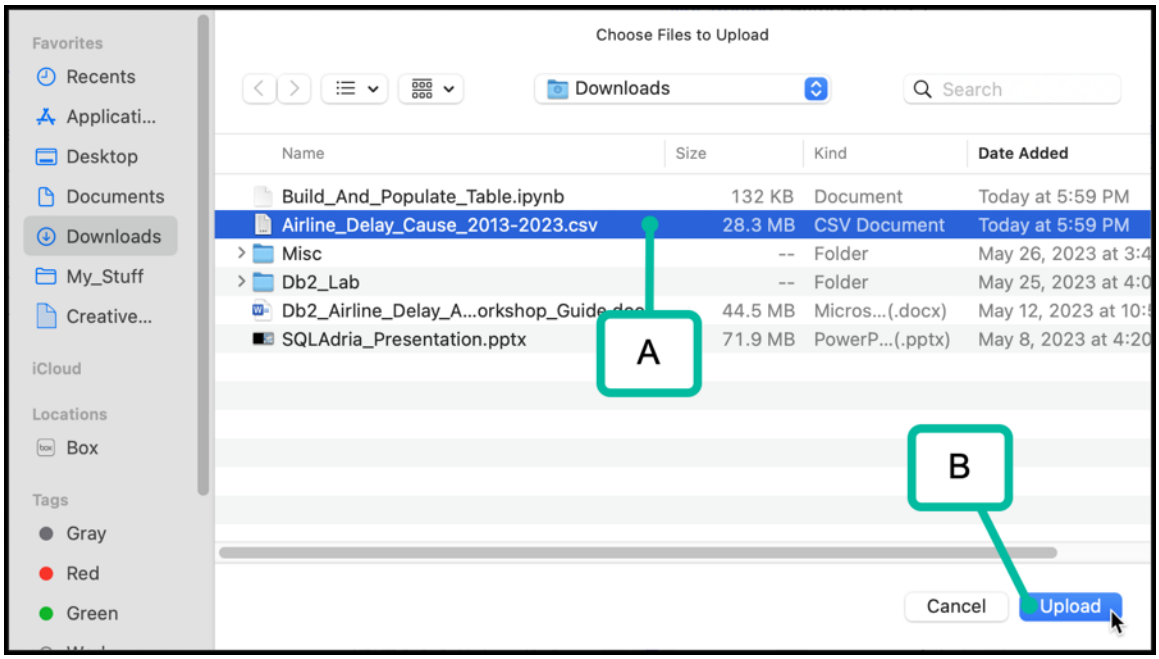
____ 1. As the name implies, the **Build_And_Populate_Table** Jupyter Notebook just instantiated is designed to create a new table in the Db2 on Cloud database created earlier and populate it with data. But, before the notebook can be used to do that, a data source must be added to our project that the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



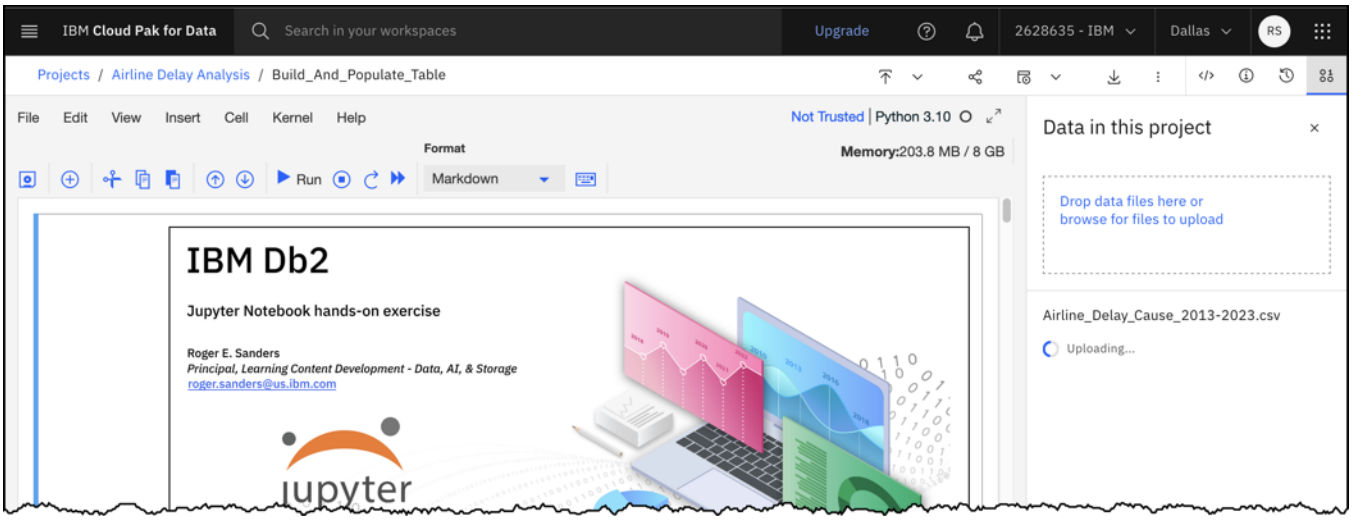
____ 2. When the *Data in this project* slideout window appears, click on the **Drop data files here or browse for files to upload** link to open a *Finder* window on Mac or a *File Explorer* window on Windows.



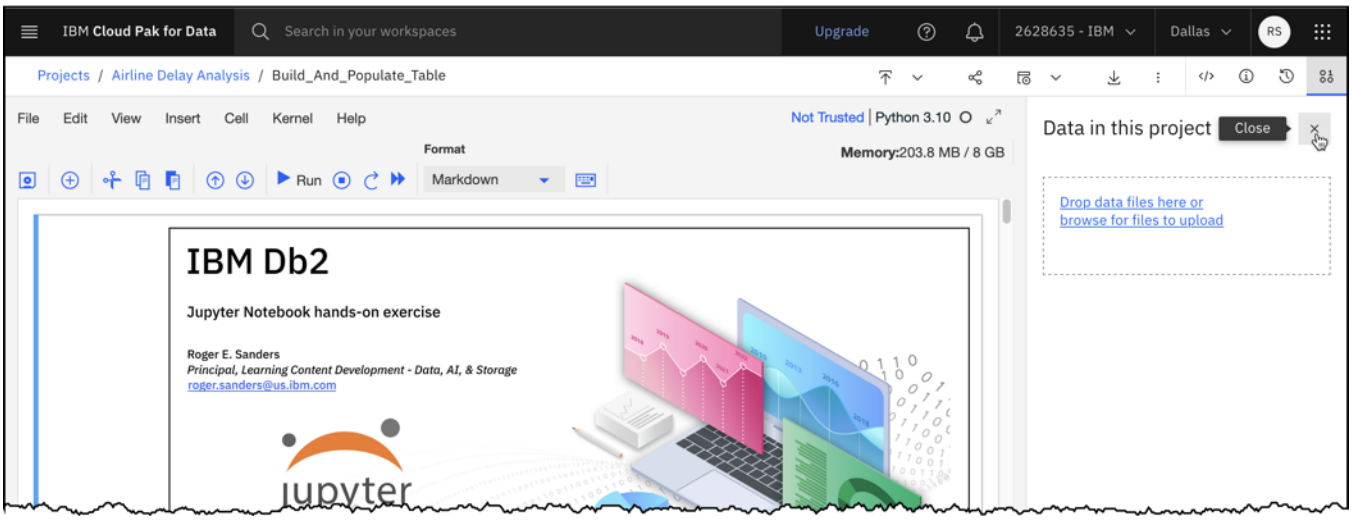
____ 3. **[A]** Select the file named **Airline_Delay_Cause_2013-2023.csv** and **[B]** click the **Upload** button.



____ 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:

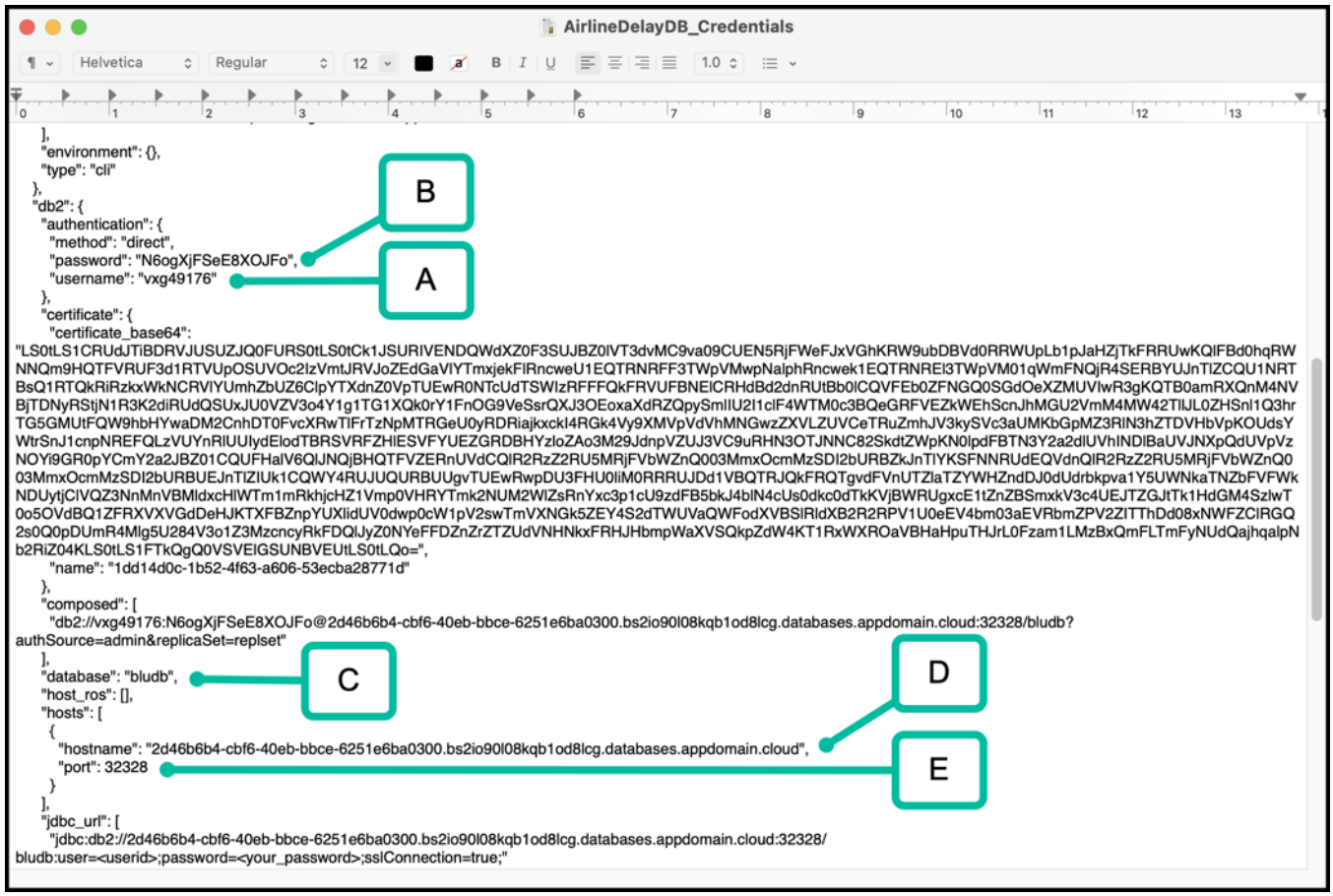
[A] User name (**“username”**)

[B] User password (**“password”**)

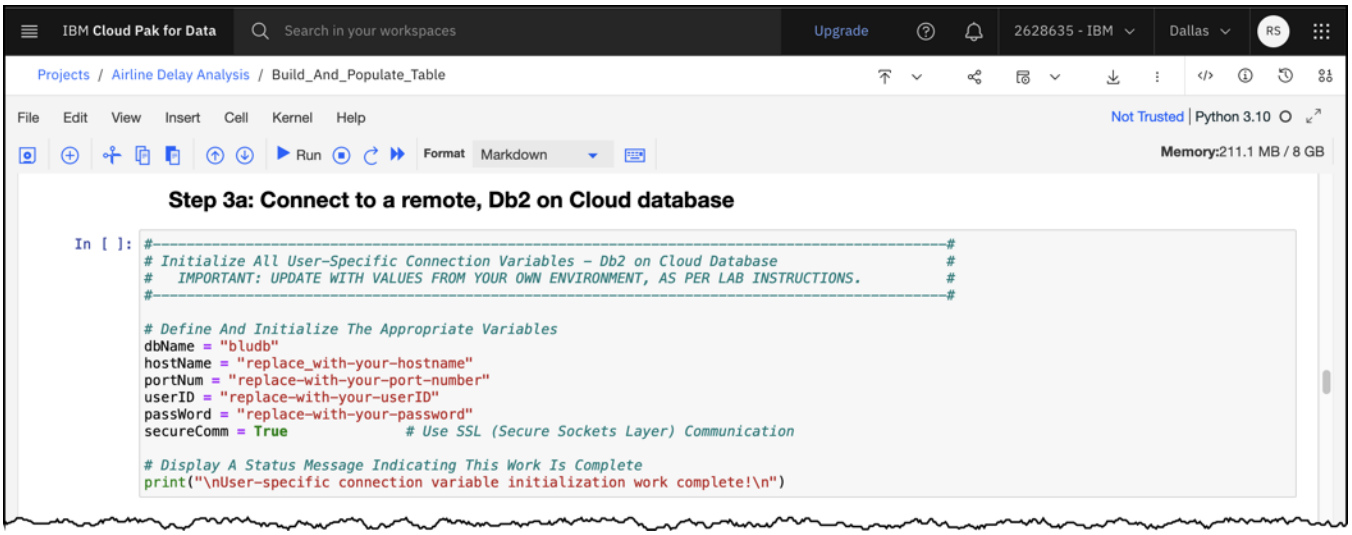
[C] Database name (**“database”**)
*NOTE: The database name is <u>always</u> **bludb** (BLUDB) for Db2 on Cloud databases.*

[D] Host name (**“hostname”**)

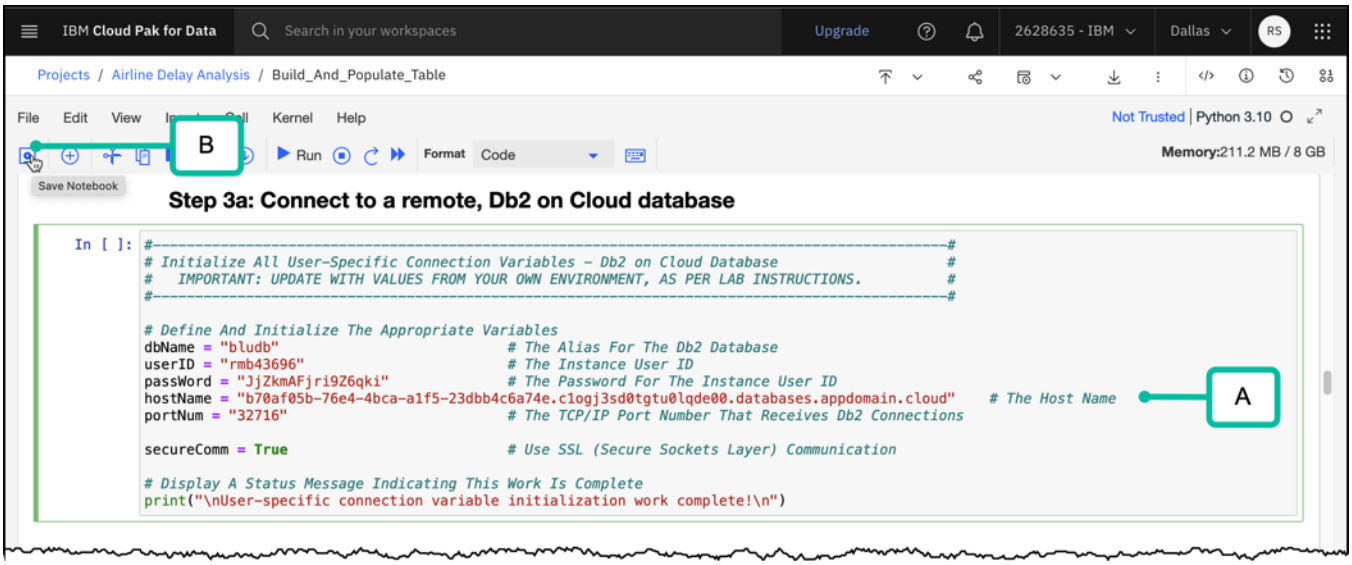
[E] Port number (**“port”**)



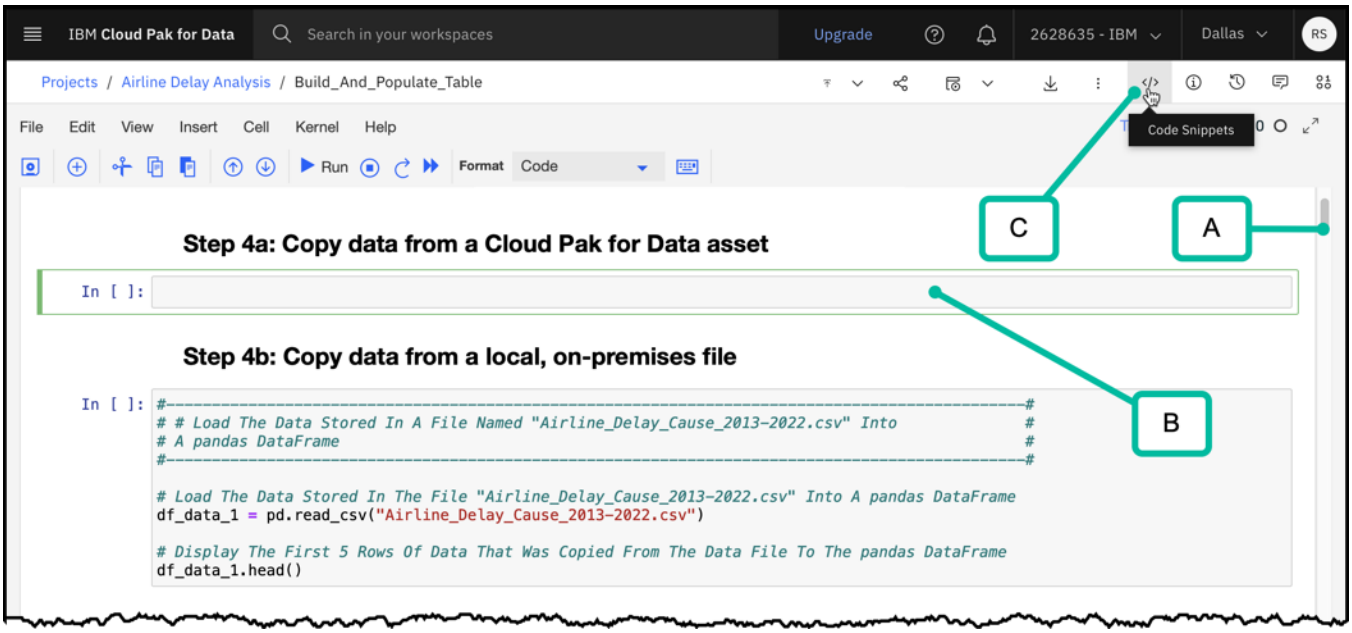
____ 2. In the Jupyter Notebook, scroll down until you see the code section under the heading **Step 3a: Connect to a remote, Db2 on Cloud database**.



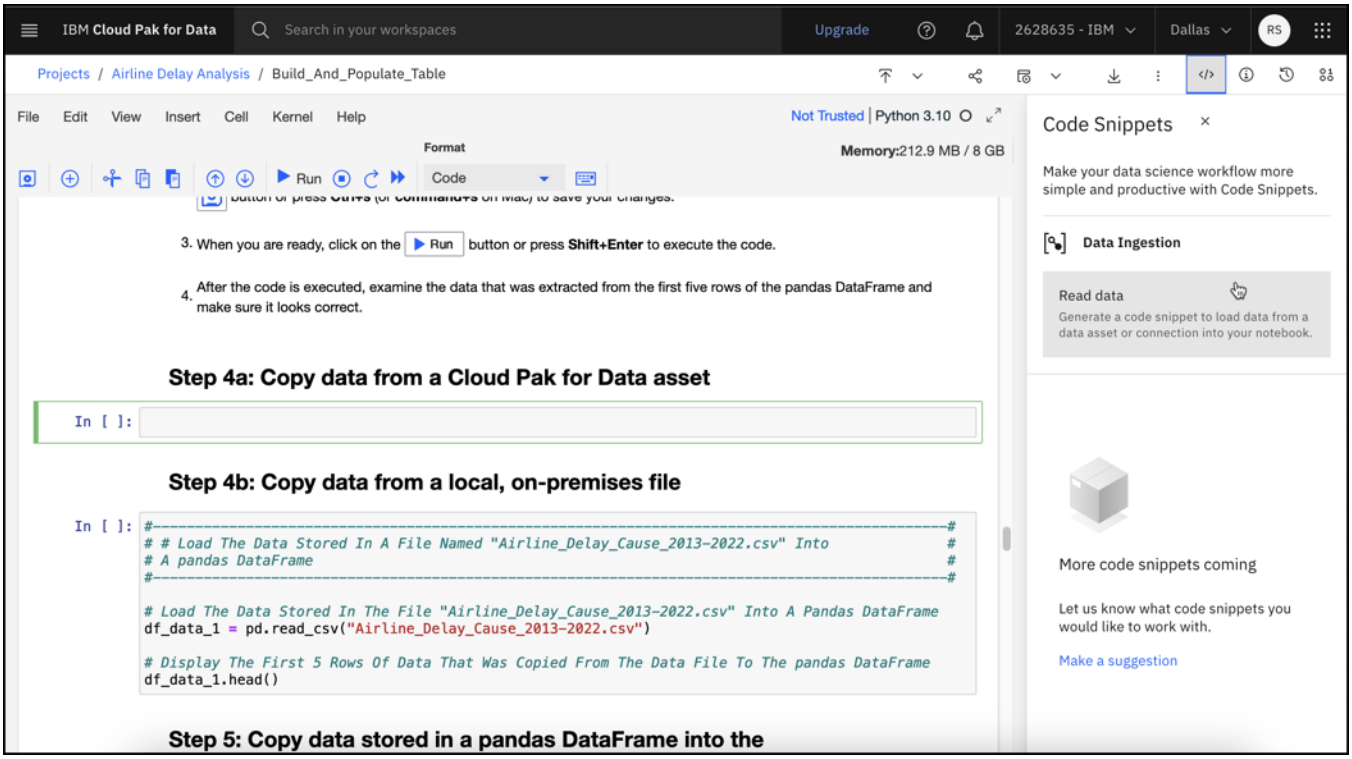
____ 3. **[A]** Click on this section to make it the current section. Then, copy the **hostname** *value* you found in the service credential file (**AirlineDelayDB_Credentials.txt**) to the clipboard and paste it over the text "**replace-with-your-hostname**" found in this code section. (*Copy the value 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.)



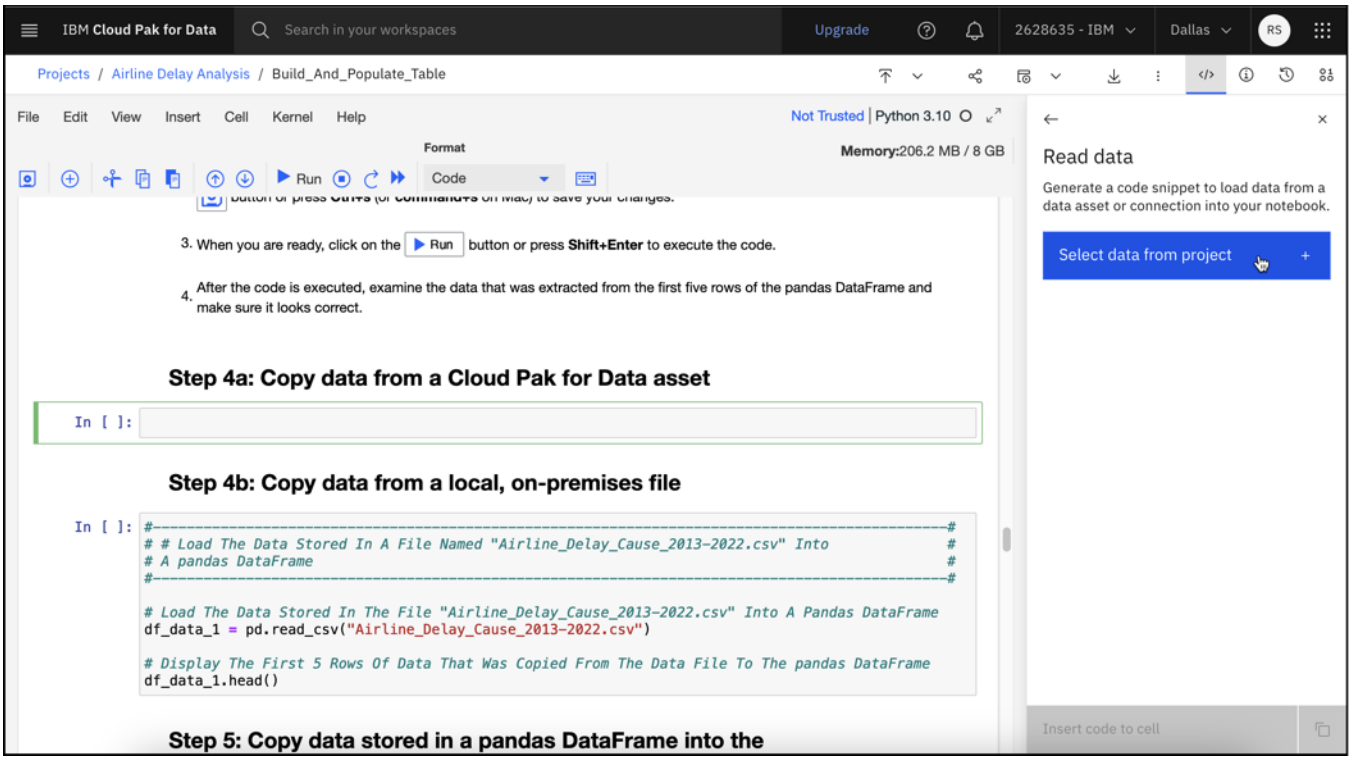
____ 4. **[A]** Scroll down until you see the code section under the heading **Step 4a: Copy data from a Cloud Pak for Data asset**. Next, **[B]** click on this section to make it the current section. Finally, **[C]** click on the **</> (Code Snippets)** icon located near the top right corner of the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



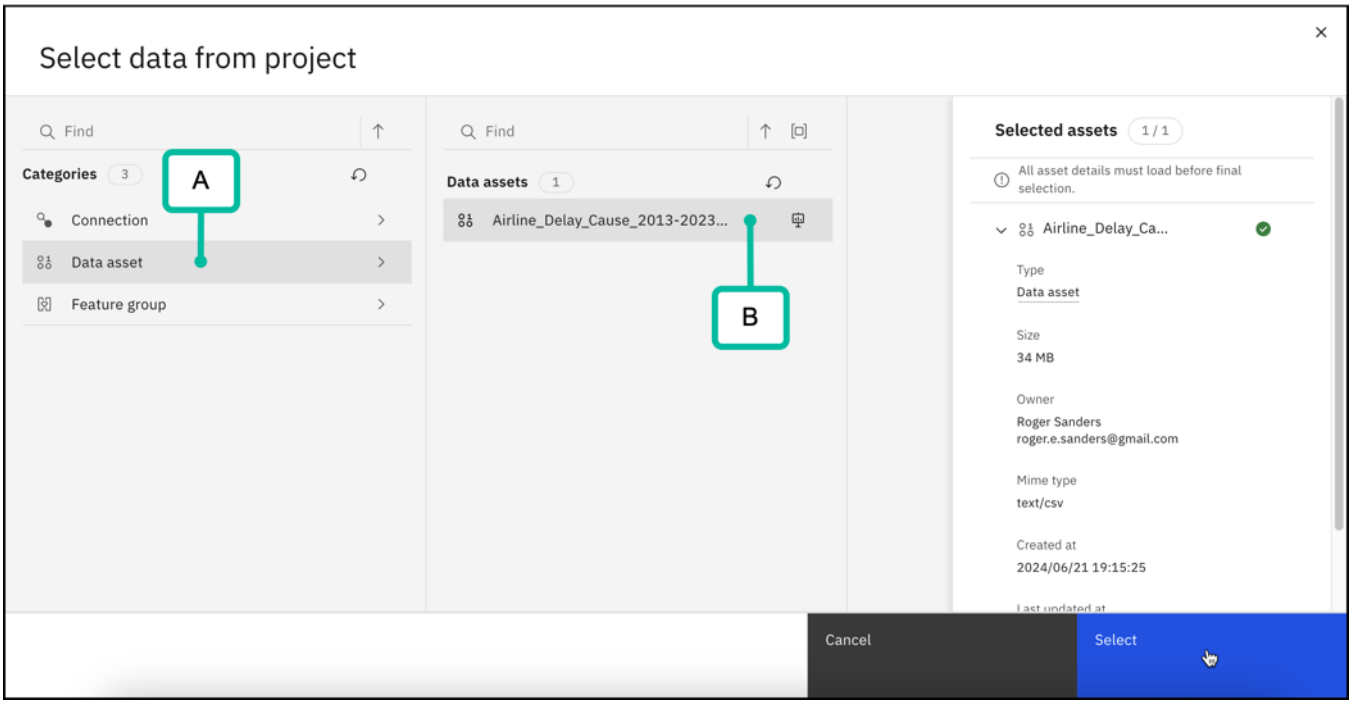
____ 5. When the *Code Snippets* slideout window appears on the right-hand side of the screen, locate the **Read data** tile (under the heading **Data Ingestion**) and click on it.



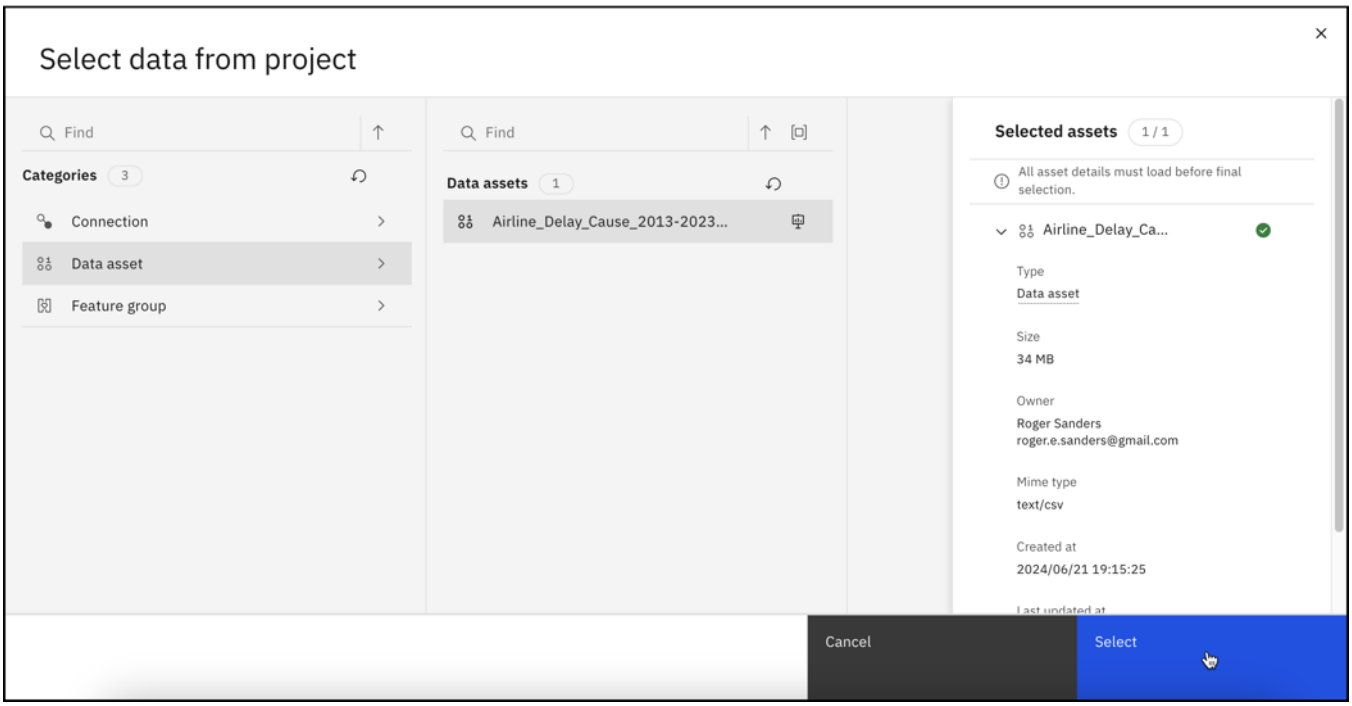
____ 6. When the *Code Snippets* slideout window is replaced with the *Read data* window, click on the **Select data from project** button. This will cause the *Select data from project* popup window to open.



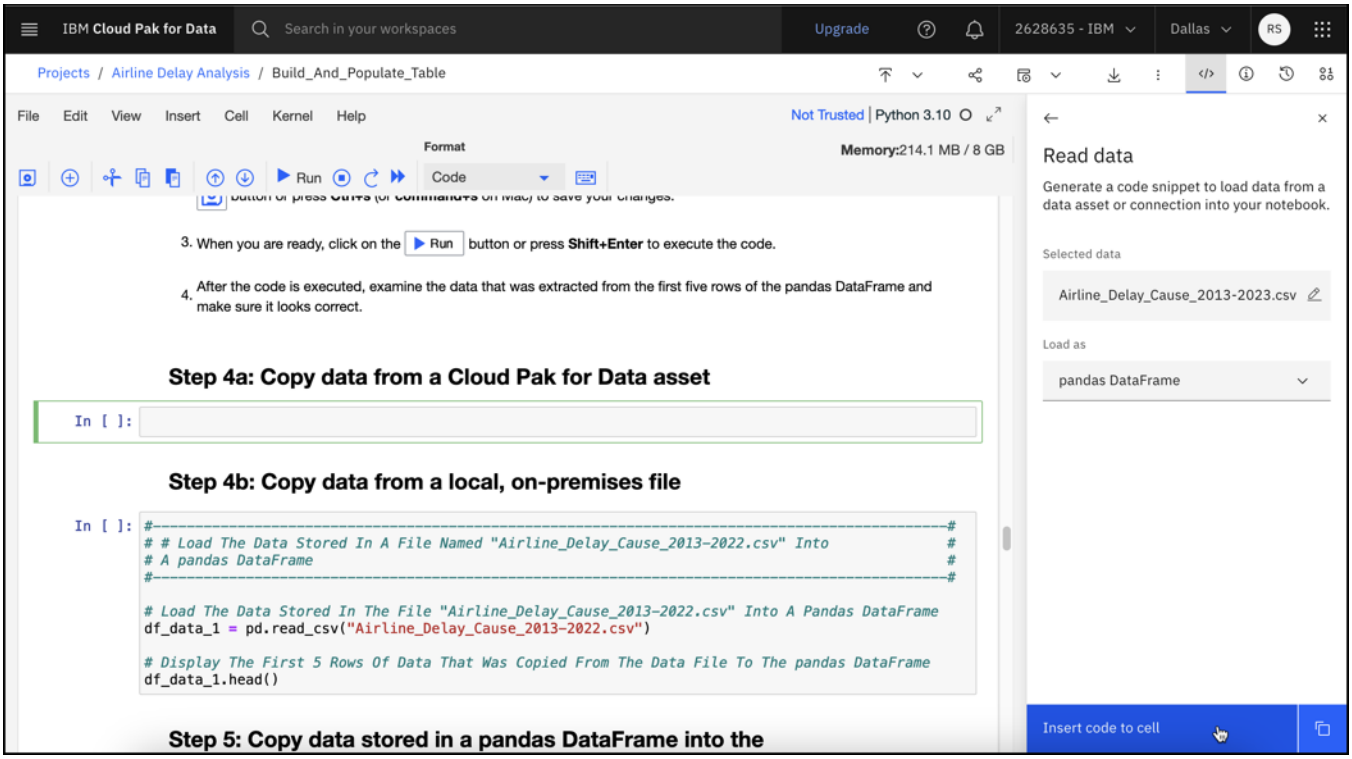
____ 7. When the *Select data from project* popup window appears, **[A]** locate the **Data asset** item in the **Categories** list and click on it. This will cause a list of all **Data assets** that have been defined for the **Airline Delay Analysis** project to be displayed. When this list appears,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



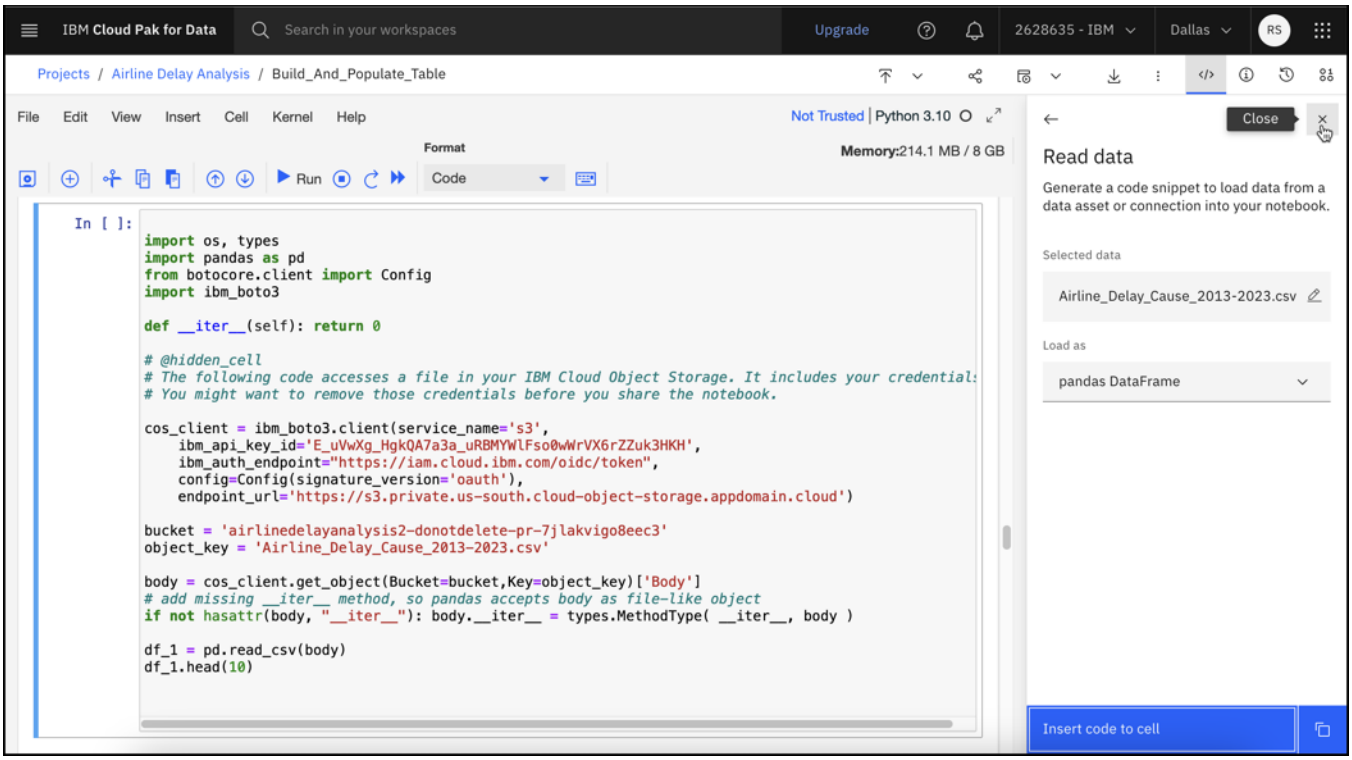
____ 8. Review the details provided about the data asset selected (shown under the **Selected assets** heading) and confirm the correct data asset was chosen. Then, click the **Select** button located in the bottom right corner of the window. This will close the popup window and return control to the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



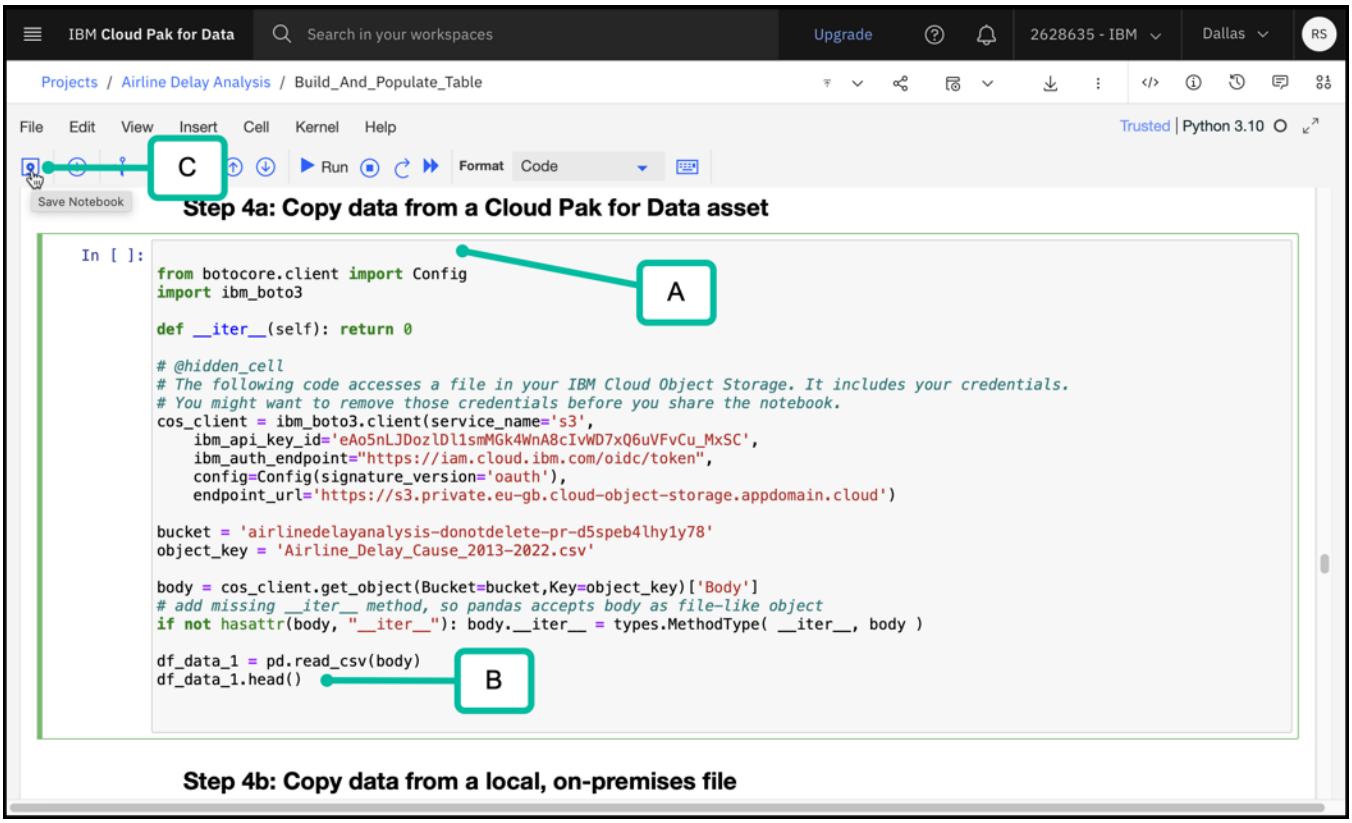
____ 10. Once the appropriate code has been generated, close the *Read data* slideout window by clicking on the **X** (Close) button located in the top right-hand corner of the window.



____ 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.

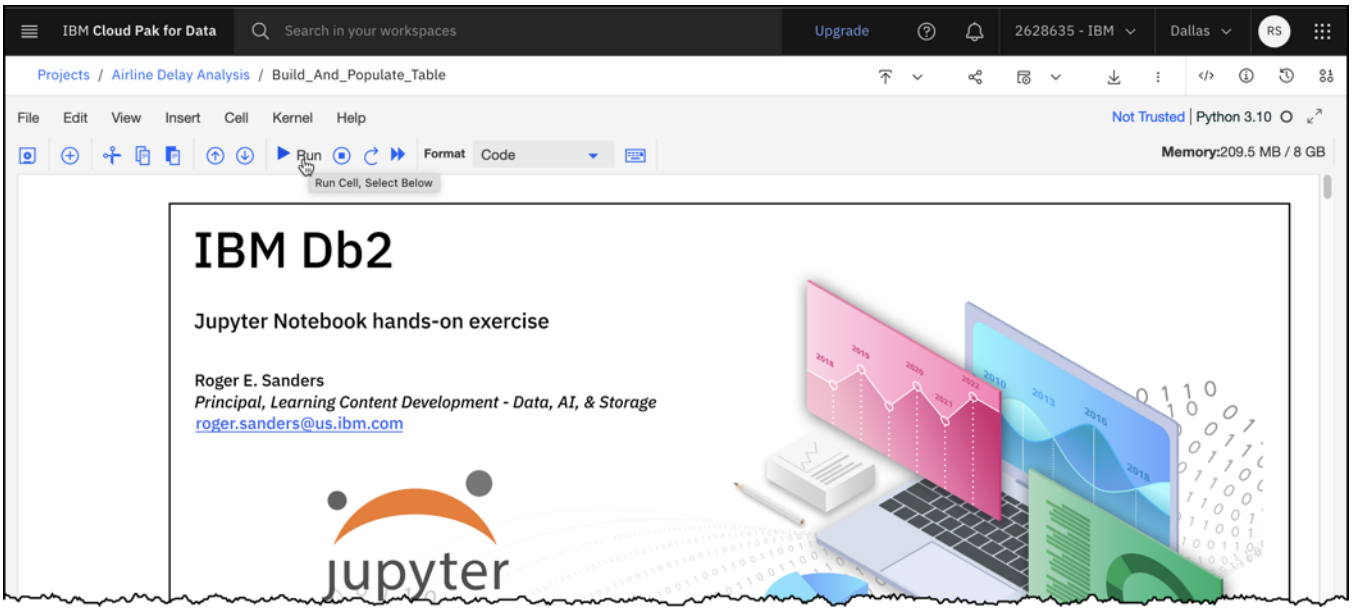


Congratulations! You are now ready to work your way through the Jupyter Notebook.

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.

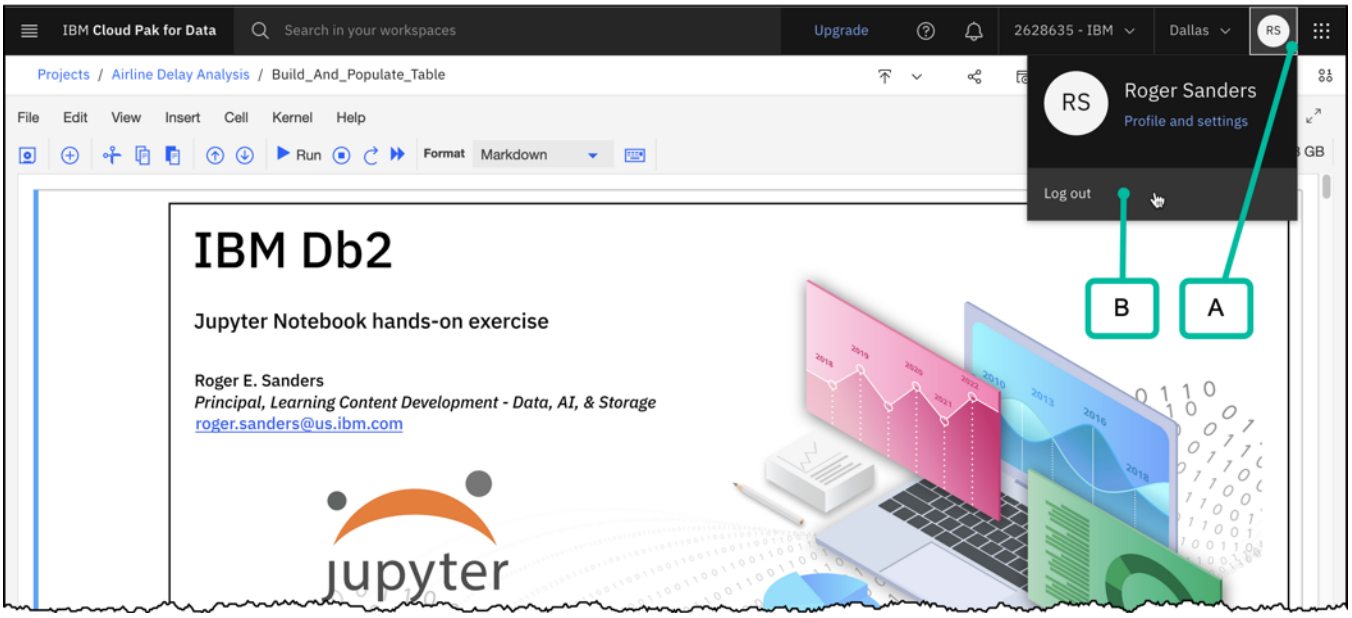


Workshop Tip

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.



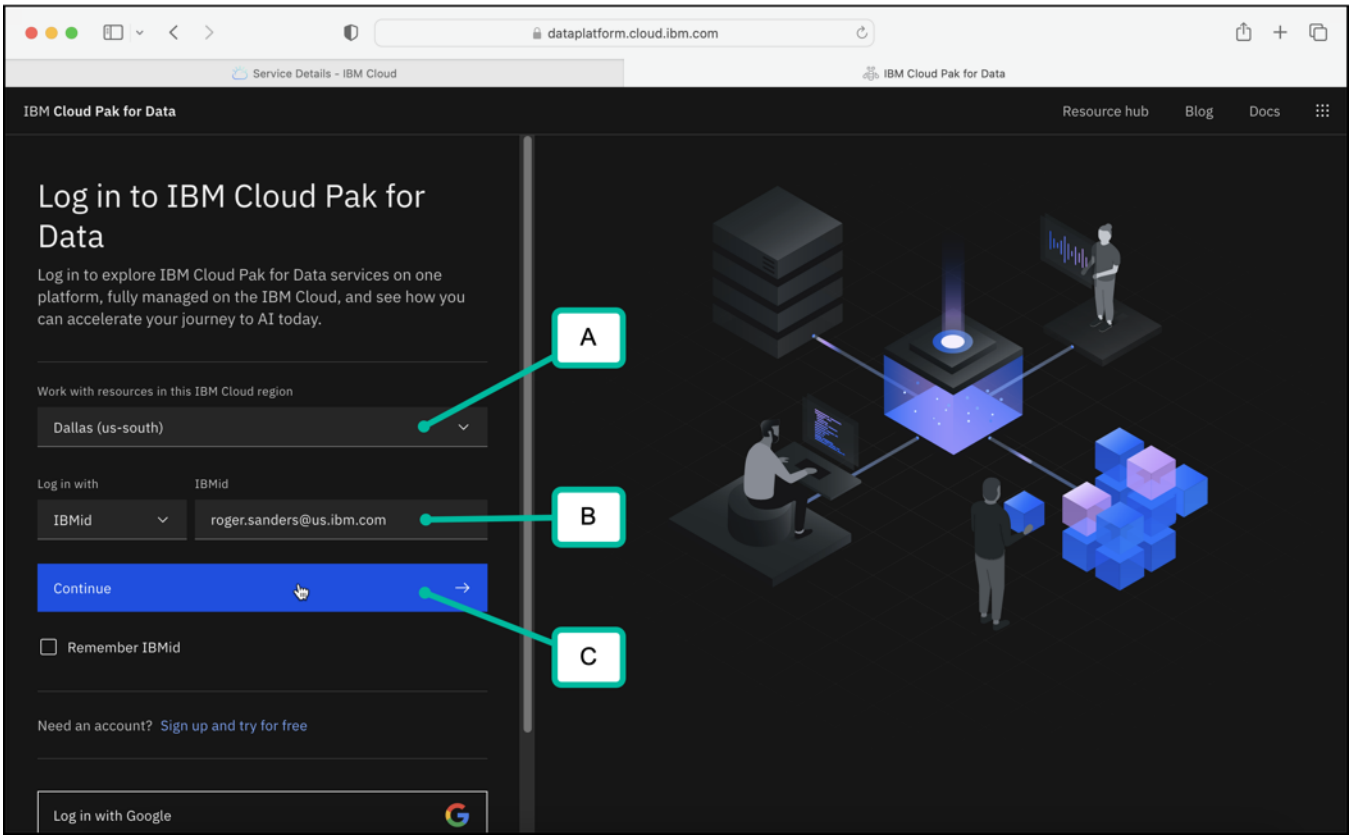
Congratulations! You have just used IBM Watson Studio (in IBM Cloud Pak for Data as a Service) and Jupyter Notebook to work with a Db2 on Cloud database.

V. Analyzing the data

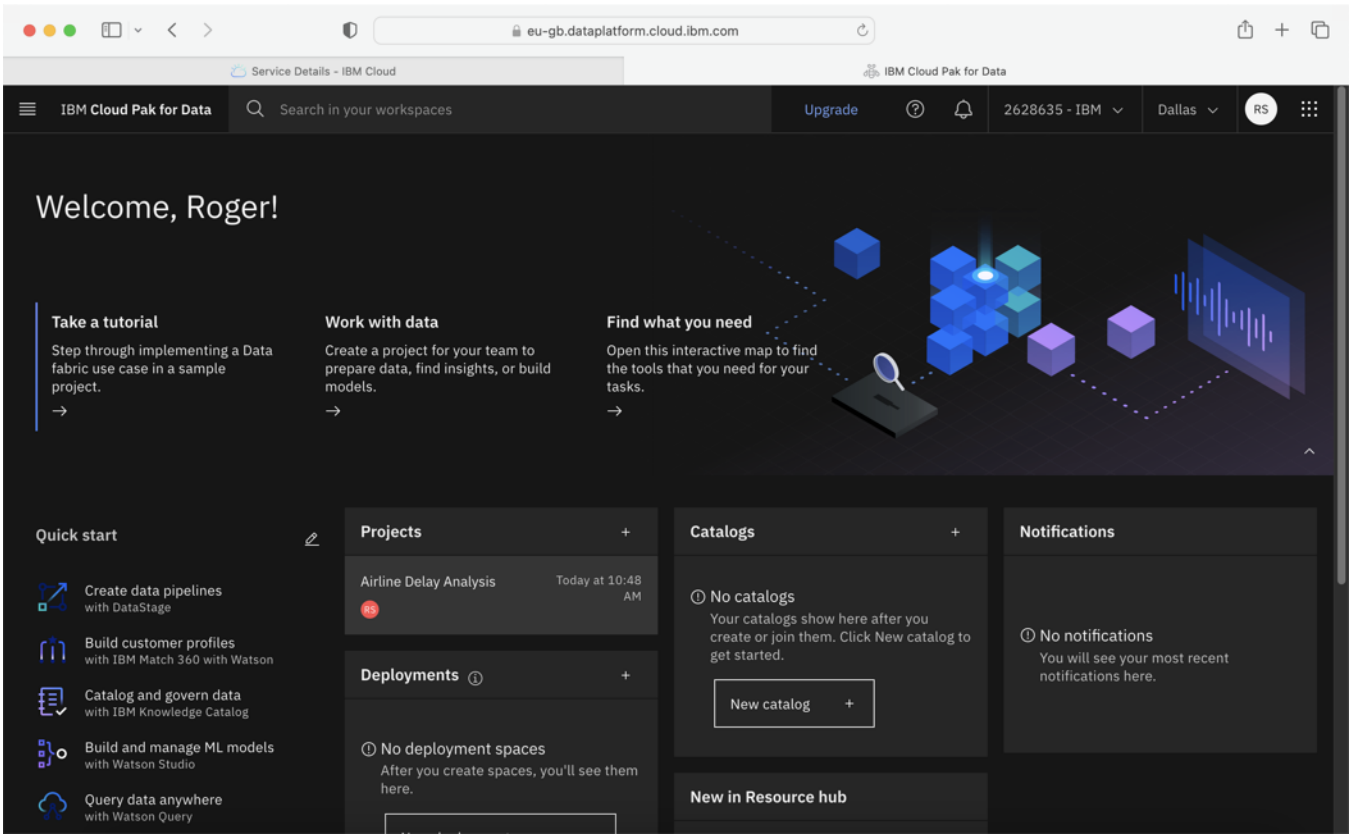
In this section of the workshop, you will use another Jupyter Notebook to analyze the airline flight delay data that was loaded into the Db2 on Cloud database. You will also use the Db2 Magic Commands to establish a database connection and retrieve data from the AIRLINE_DELAY_CAUSE and AIRLINE_DELAY_CAUSE_TOTAL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

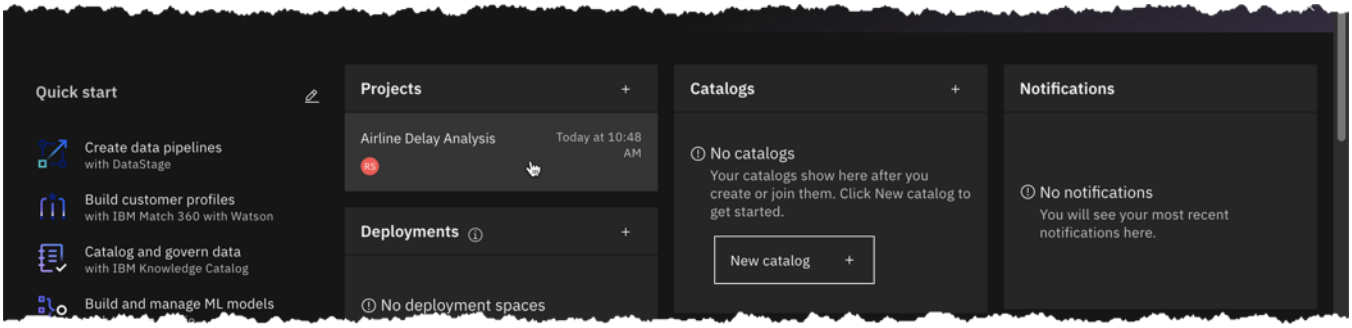


____ 3. Once you have successfully logged in, you should be presented with an IBM Cloud Pak for Data as a Service *Welcome* screen that looks something like this:

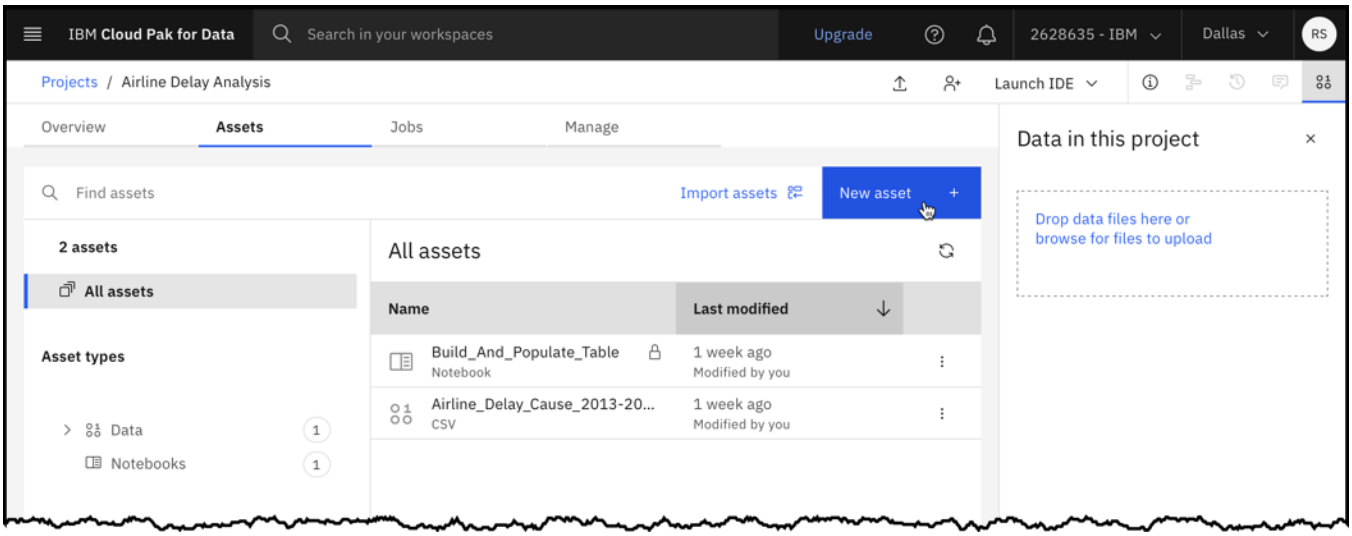


2. Add a new Jupyter Notebook to the project created earlier.

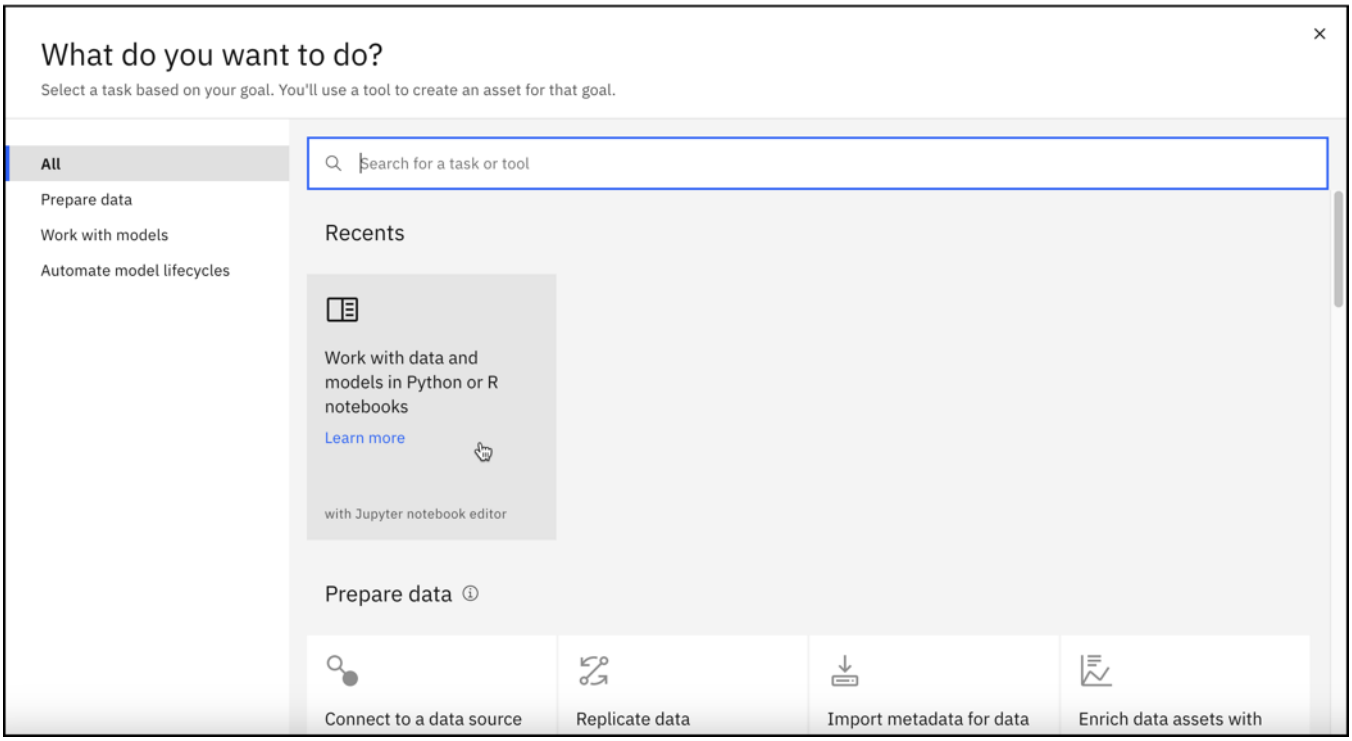
____ 1. When the IBM Cloud Pak for Data as a Service *Welcome* screen appears, locate the project in the **Projects** box that was created earlier (in this example, the project was named "*Airline Delay Analysis"*) and click on it. This should cause the *Welcome* screen to be replaced with an *Assets* 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.)



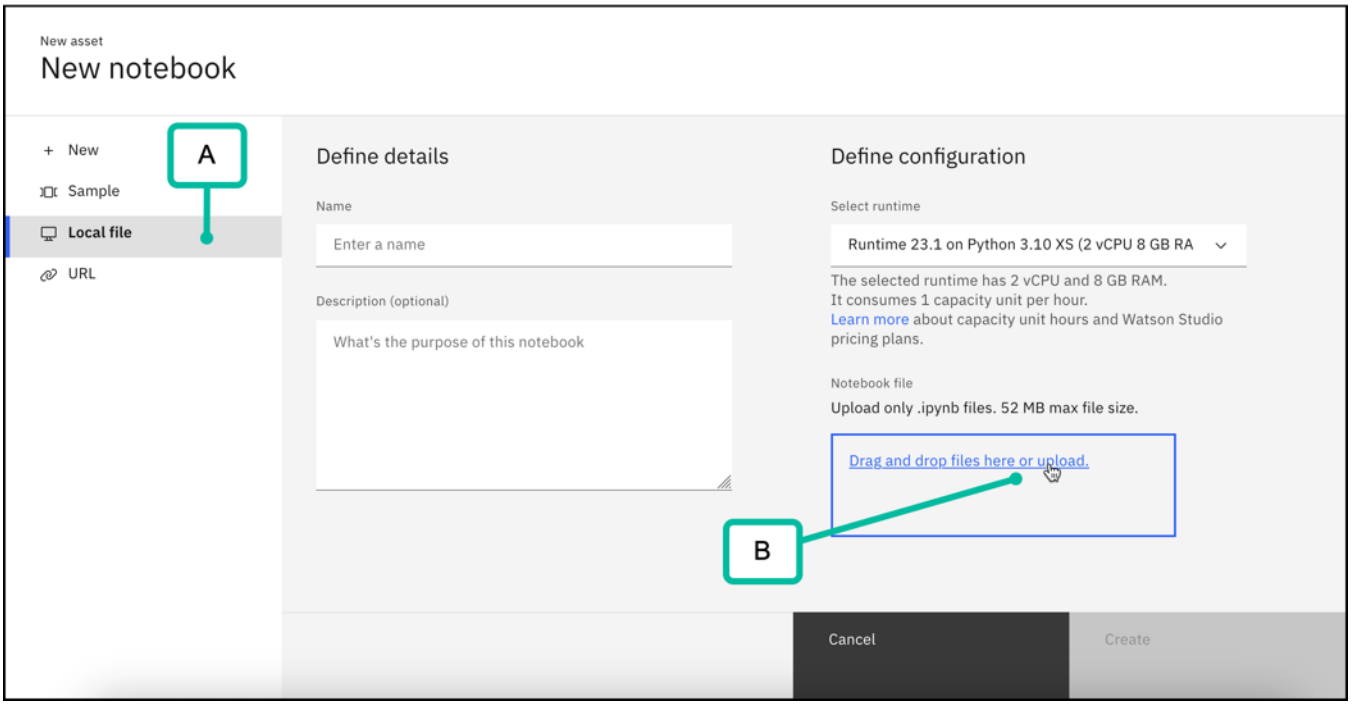
____ 2. When the *Assets* screen is displayed, click on the **New asset** button.



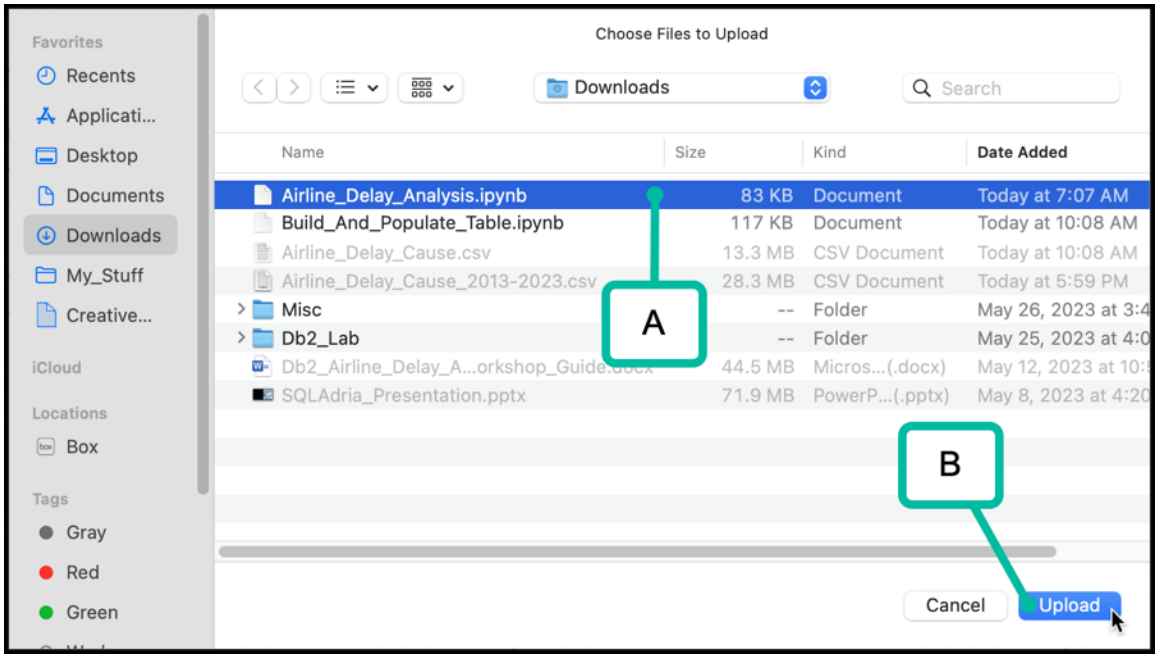
____ 3. When the *What do you want to do* popup window appears, click on the **Work with data and models in Python or R notebooks** tile. This will cause the *What do you want to do* screen to be replaced with the *New notebook screen.*



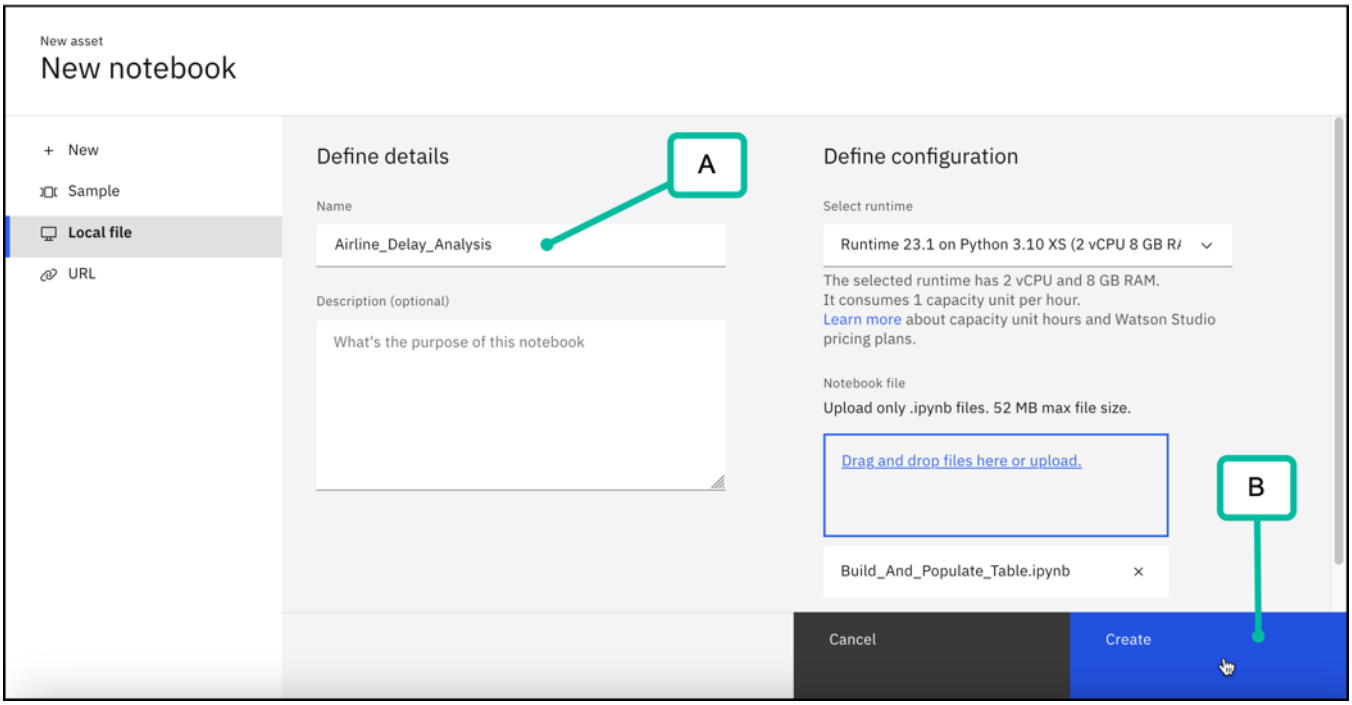
____ 4. From the *New notebook* screen, **[A]** click on the **Local file** item in the menu shown on the left-hand side of the screen. Then, **[B]** click on the **Drag and drop files here or upload** link to open a *Finder* window on Mac or a *File Explorer* window on Windows.



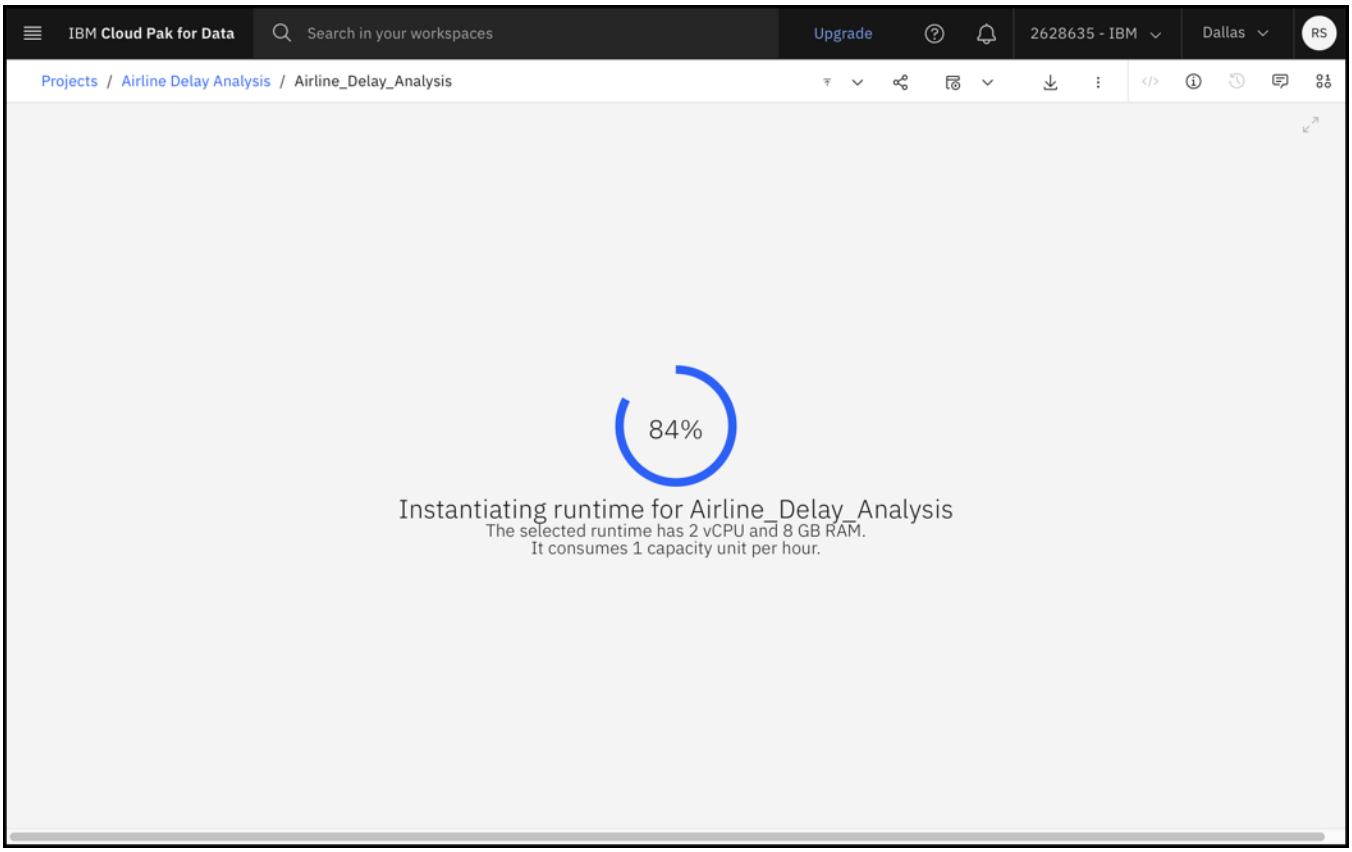
____ 5. **[A]** Select the file named **Airline_Delay_Analysis.ipynb** and **[B]** click the **Upload** button.



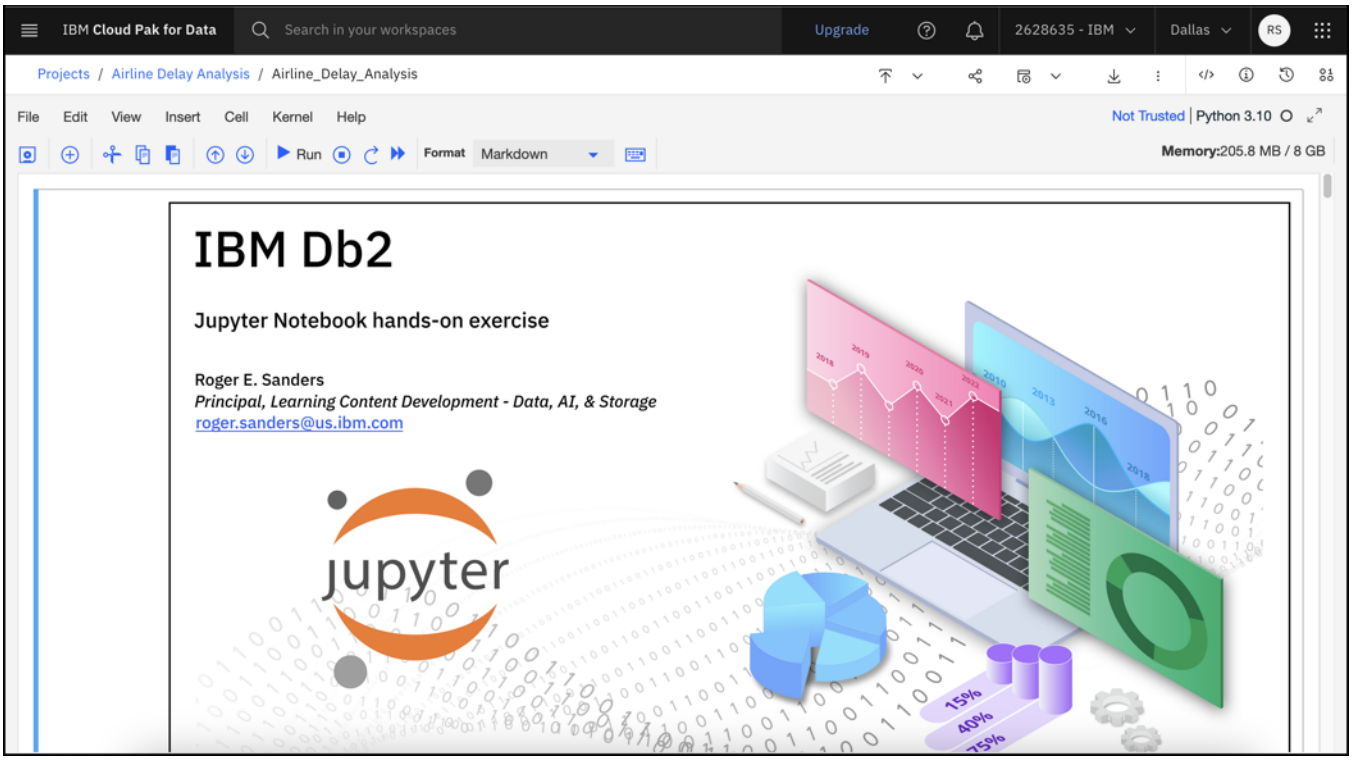
____ 6. **[A]** Confirm that the **Name** field on the *New notebook* screen contains the name of the file just selected, minus the file extension (**Airline_Delay_Analysis**). Then, **[B]** click the **Create** button shown at the bottom of the screen.



____ 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:

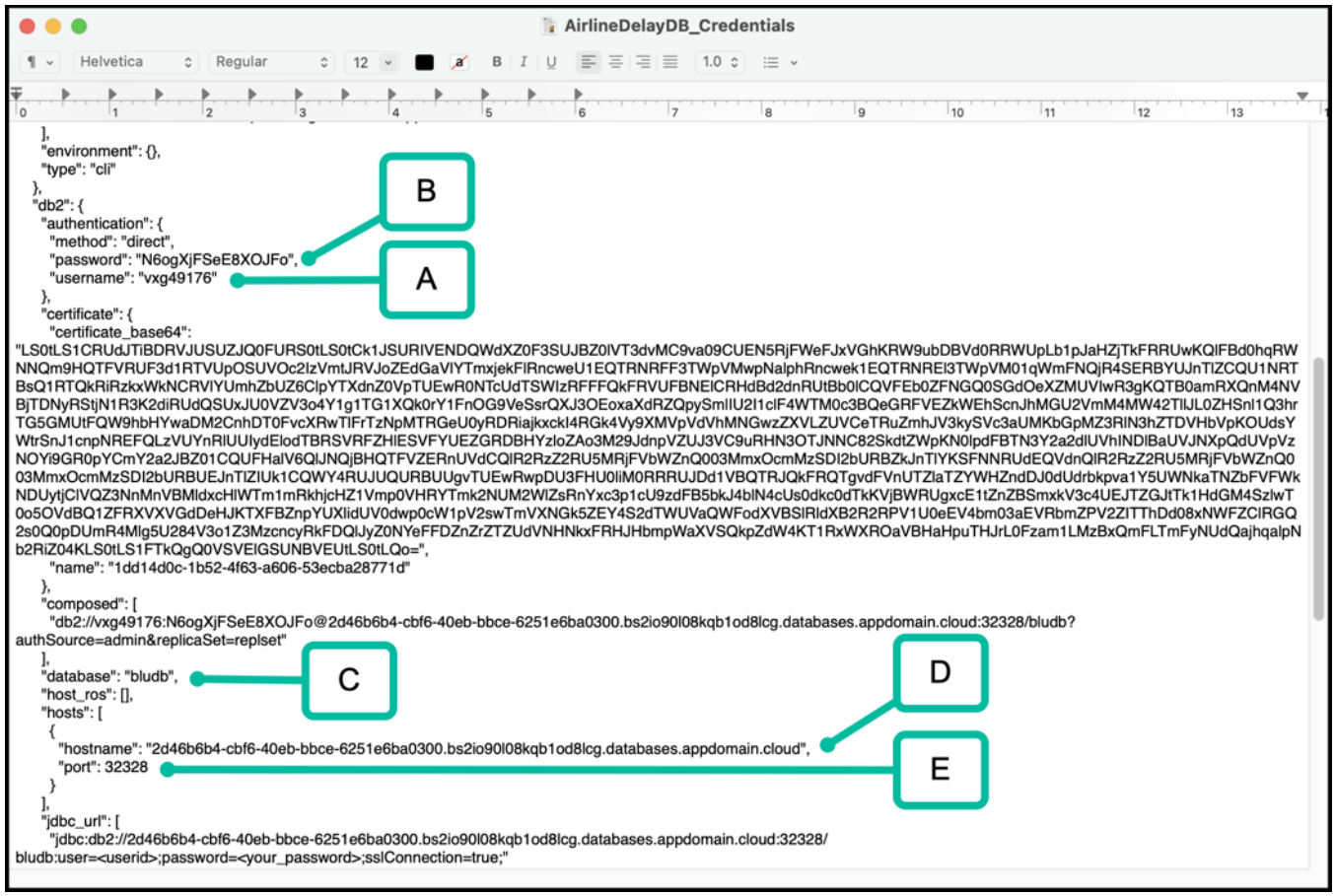
[A] User name (**"username"**)

[B] User password (**"password"**)

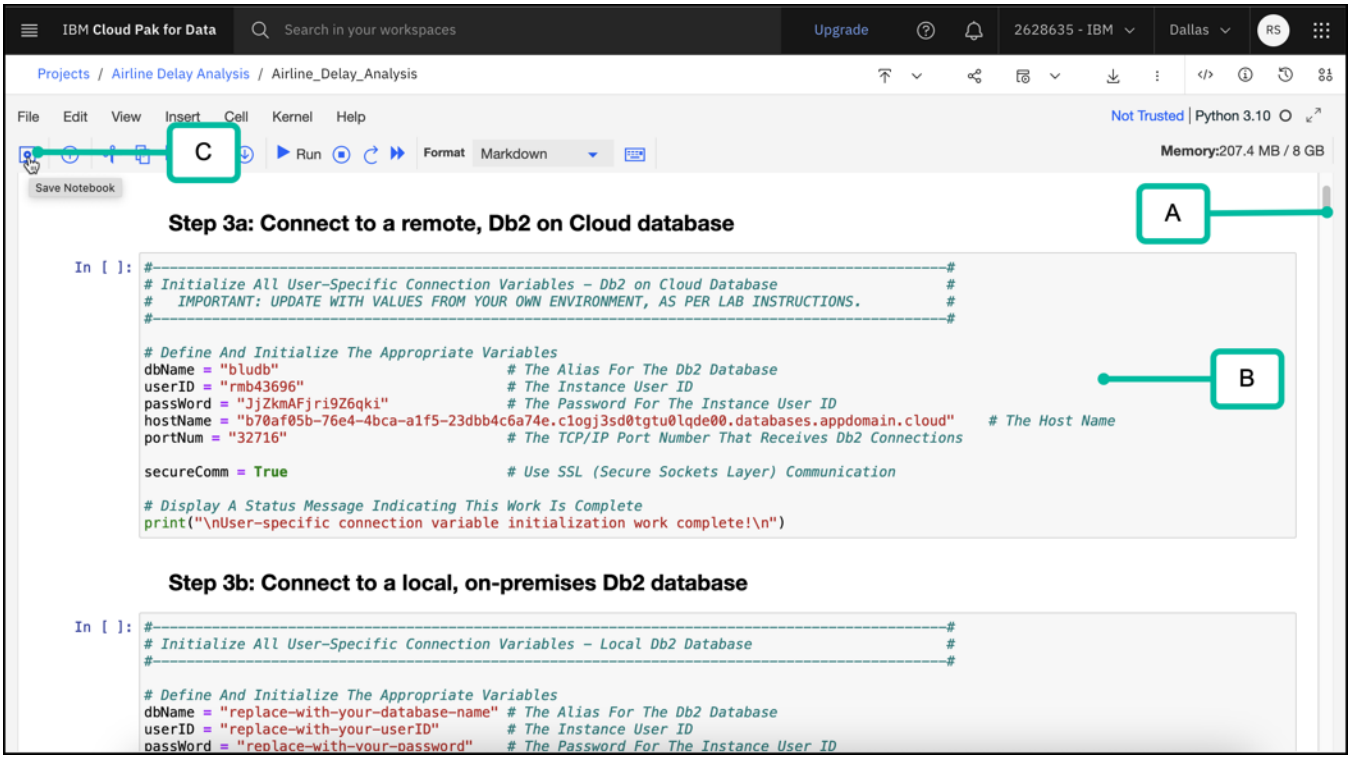
[C] Database name (**"database"**)
*NOTE: The database name is always **bludb** (BLUDB) for Db2 on Cloud databases.*

[D] Host name (**"hostname"**)

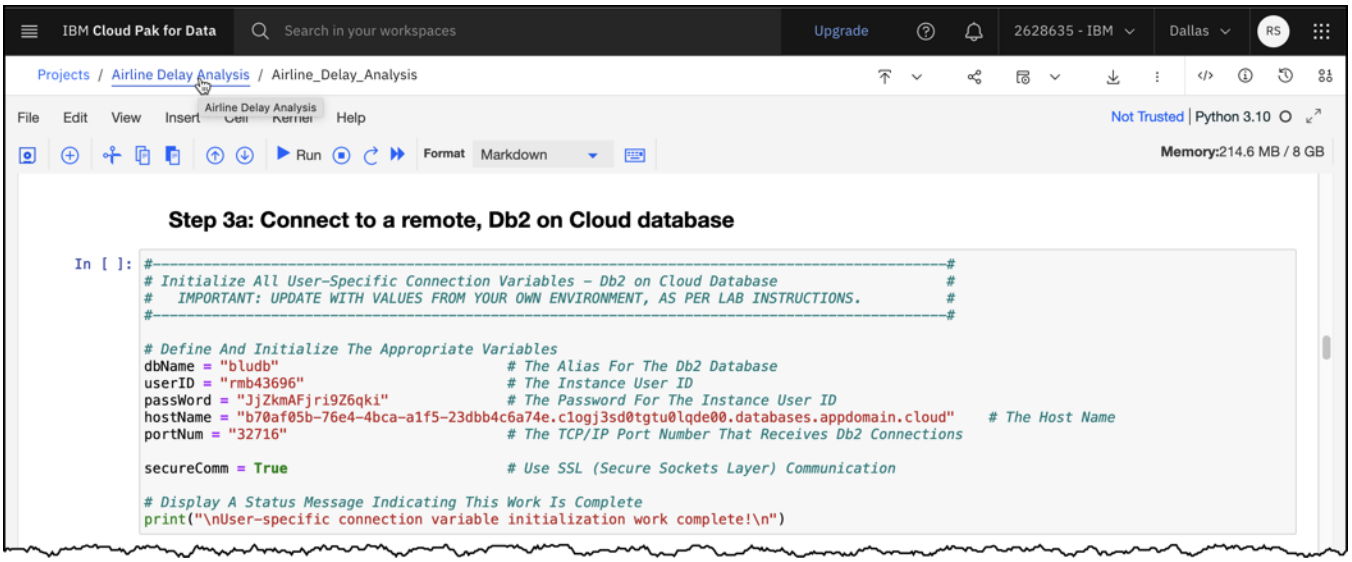
[E] Port number (**"port"**)



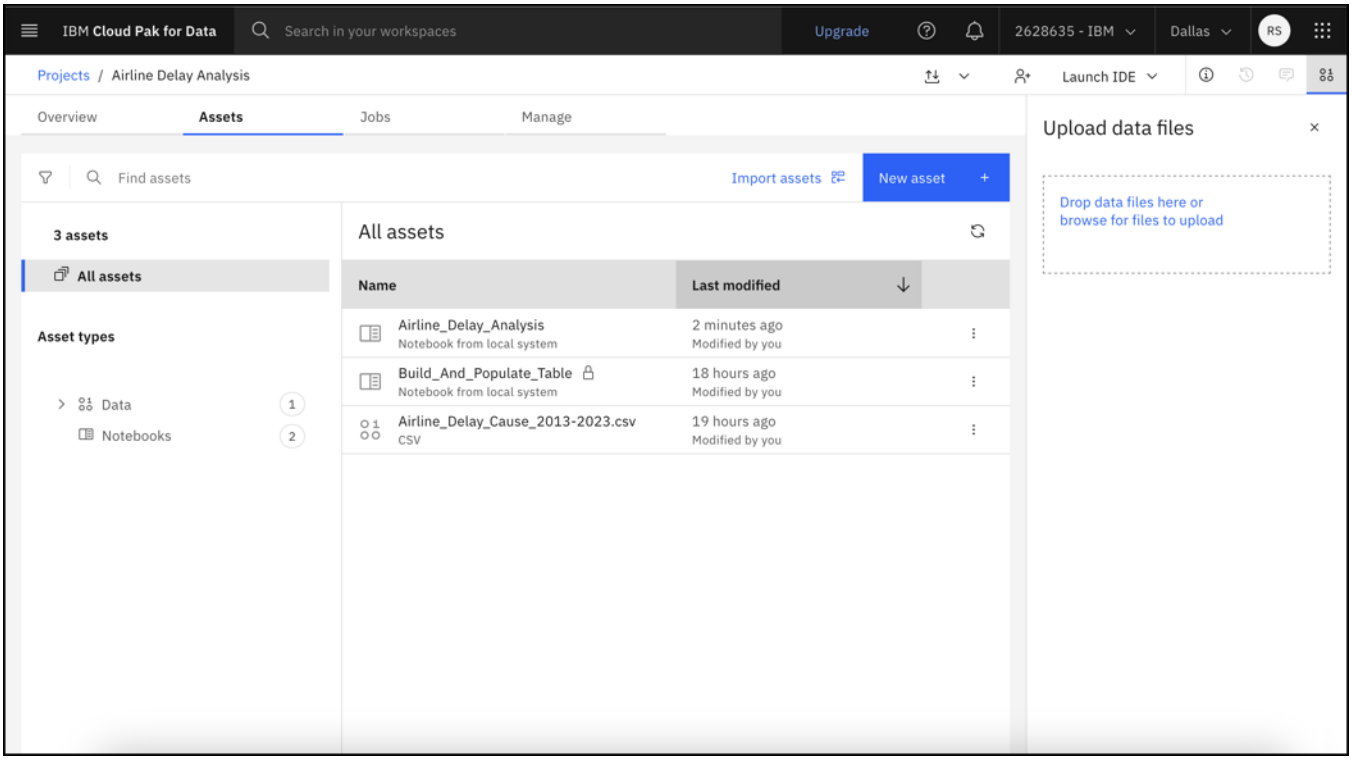
____ 2. In the Jupyter Notebook that was just added to the project (**Airline_Delay_Analysis.ipynb**), **[A]** scroll down until you see the code section under the heading **Step 3a: Connect to a remote, Db2 on Cloud database**. Then, **[B]** click on this section to make it the current section. Next, 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.)



____ 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.



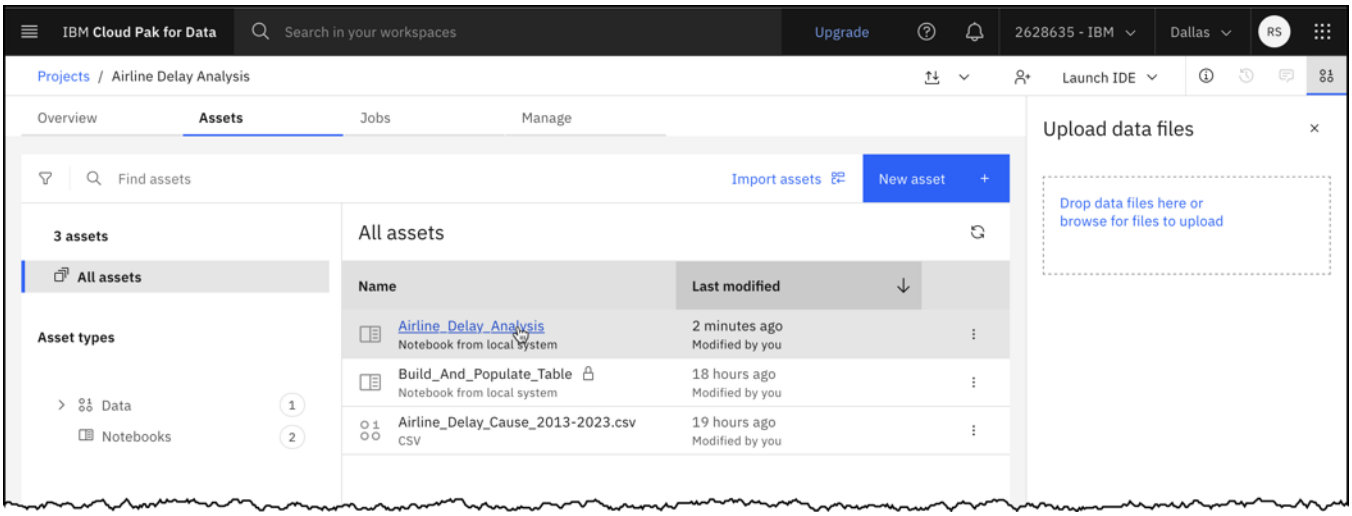
____ 5. When the *Assets* screen for the **Airline Delay Analysis** project is displayed again, the **Airline_Delay_Analysis** Jupyter Notebook should be in the project assets list.



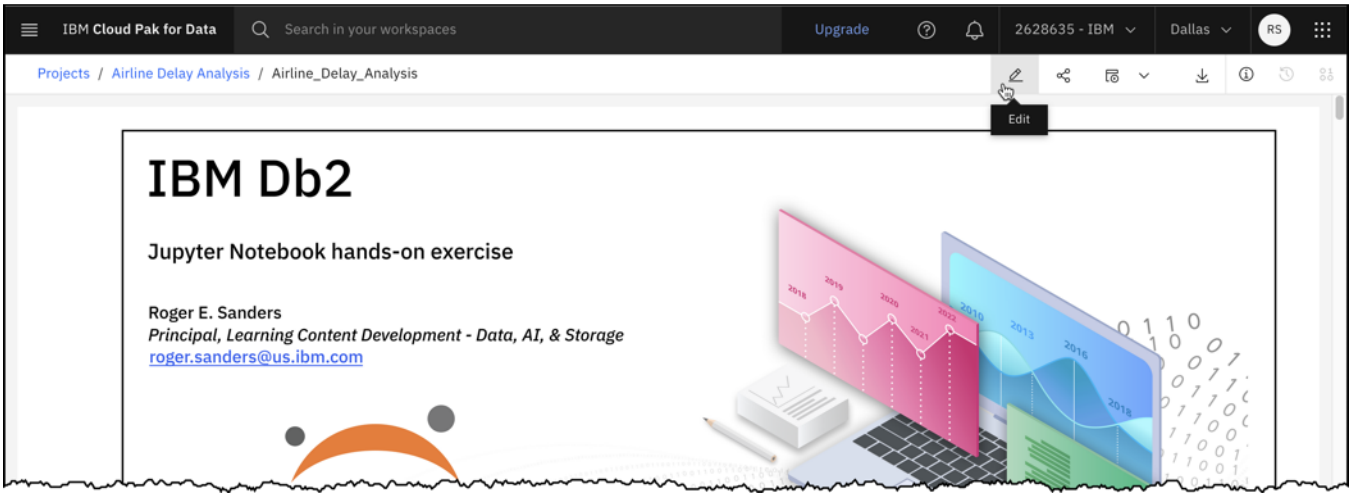
Congratulations! You are now ready to begin working your way through the **Airline_Delay_Analysis** Jupyter Notebook.

3. Run the Airline_Delay_Analysis Jupyter Notebook

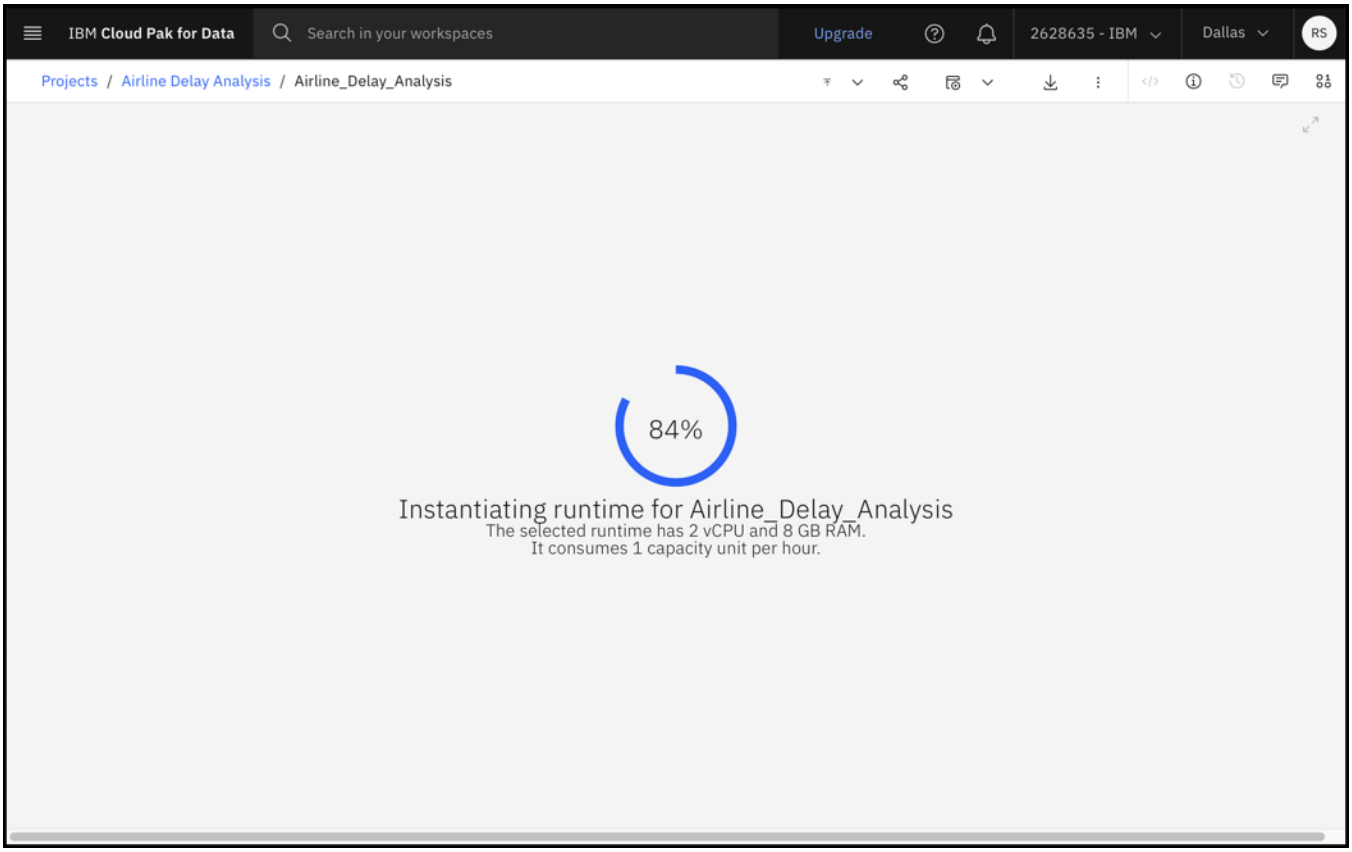
____ 1. If you are not already there, navigate to the *Assets* screen for the **Airline Delay Analysis** project, and click on the link for the Jupyter Notebook asset named **Airline_Delay_Analysis**.



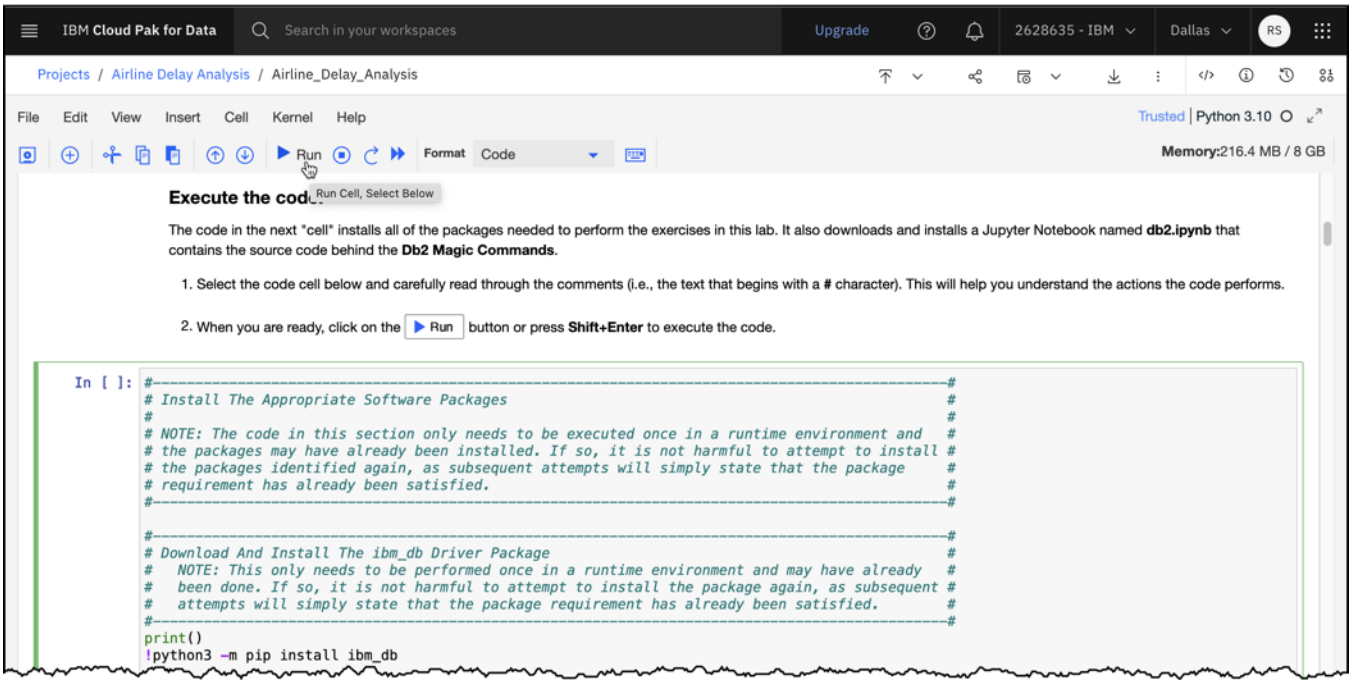
____ 2. When the Jupyter Notebook opens, you may notice that the Jupyter Notebook menu and icons are not present. That's because the notebook has not been instantiated (i.e., a kernel for the Jupyter Notebook has not been started). Instantiate the **Airline_Delay_Analysis** Jupyter Notebook by clicking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



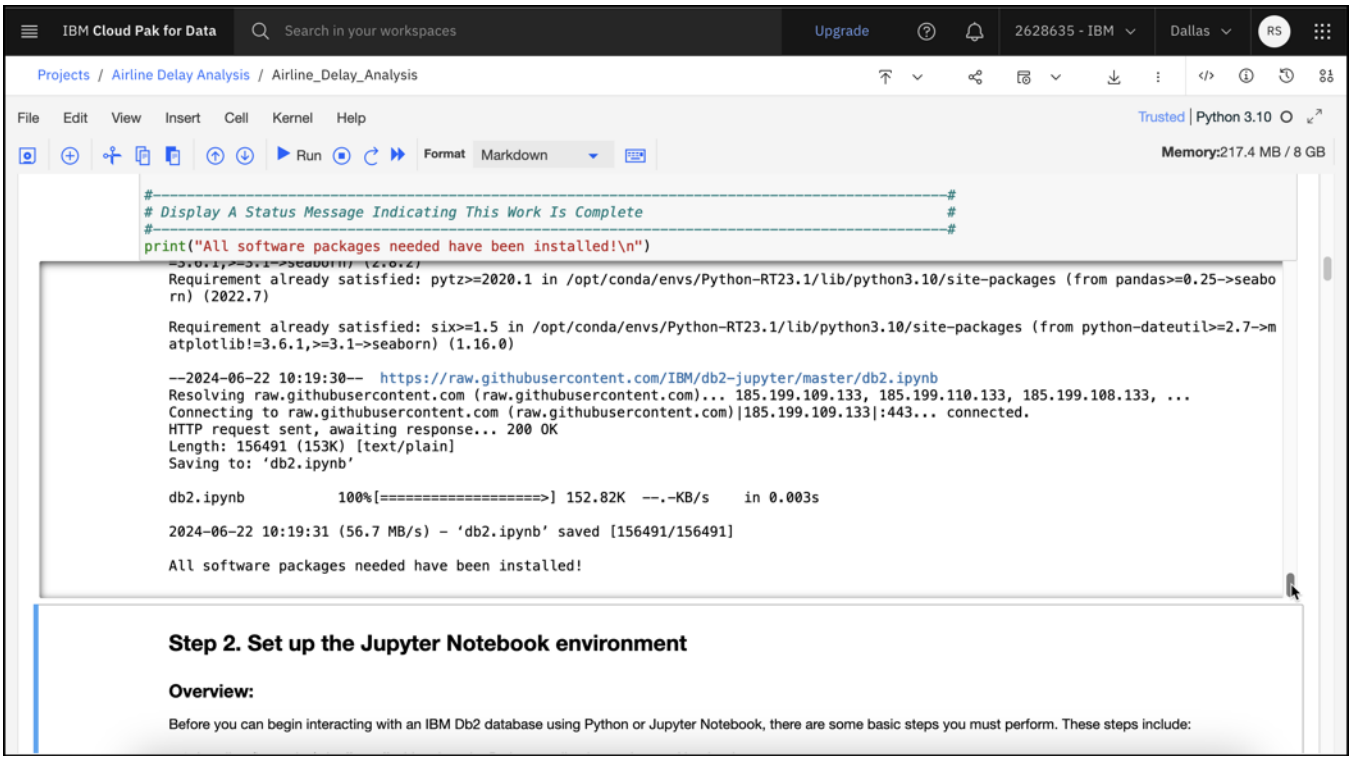
____ 3. While the Jupyter Notebook is being instantiated, you should see a screen that looks something like this:



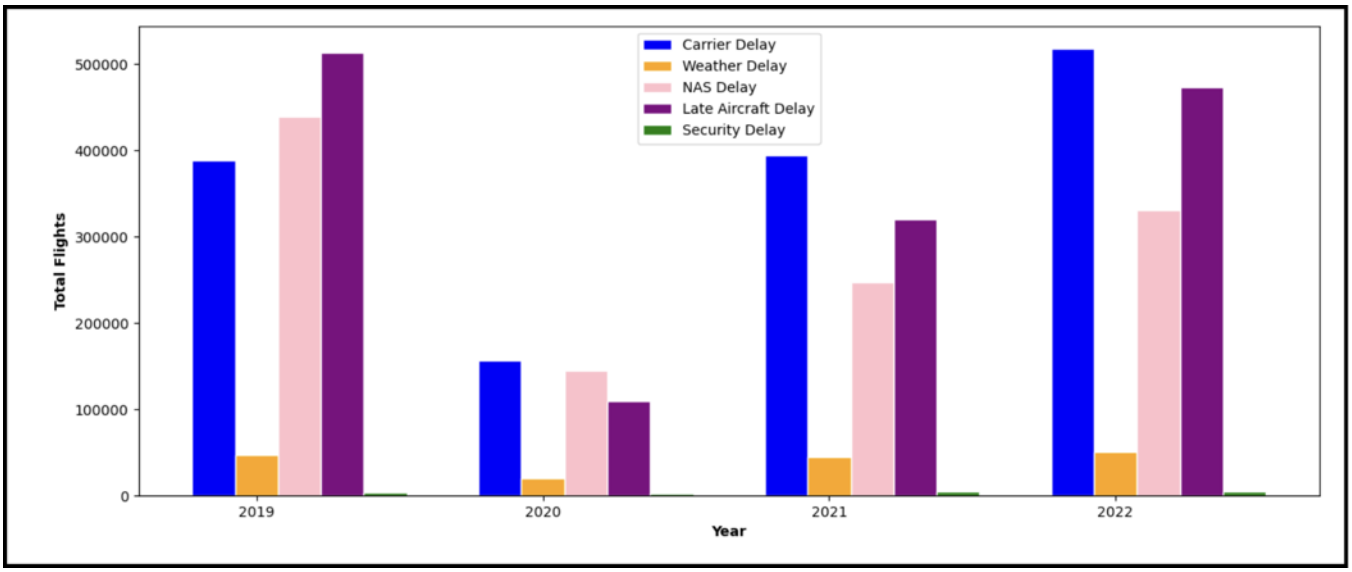
____ 4. Once the **Airline_Delay_Analysis** Jupyter Notebook has been instantiated, work your way through it, one cell at a time, reading the documentation found in markup cells and executing the source code found in code cells. The notebook is well documented and easy to work through. As a reminder,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



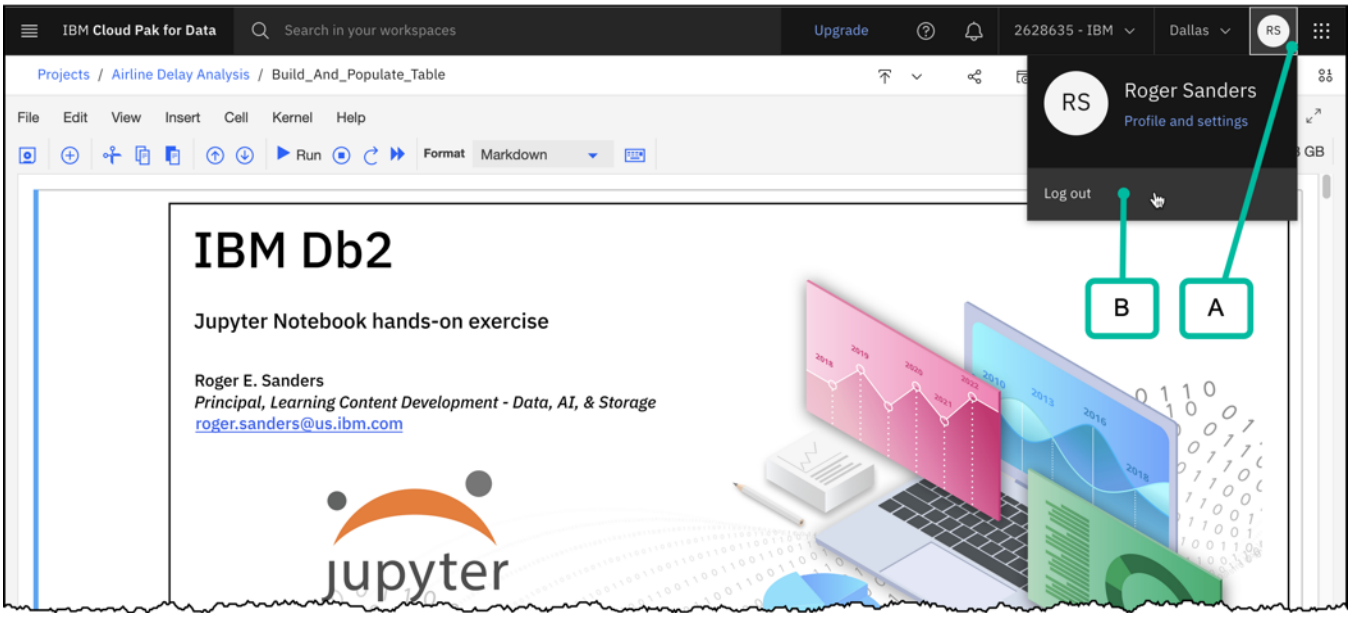
____ 5. You will find that with this Jupyter Notebook, some of the code cells generate a significant amount of output. When that is the case, use the scroll bars that are provided with the output window to view all the information produced.



____ 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.