IBM Data Integration

Hands-on Lab Guide

Amin Abou-Gallala
Americas Data Technical Specialist

Bob Reno
WW Data Technical Specialist

DataStage Core Lab

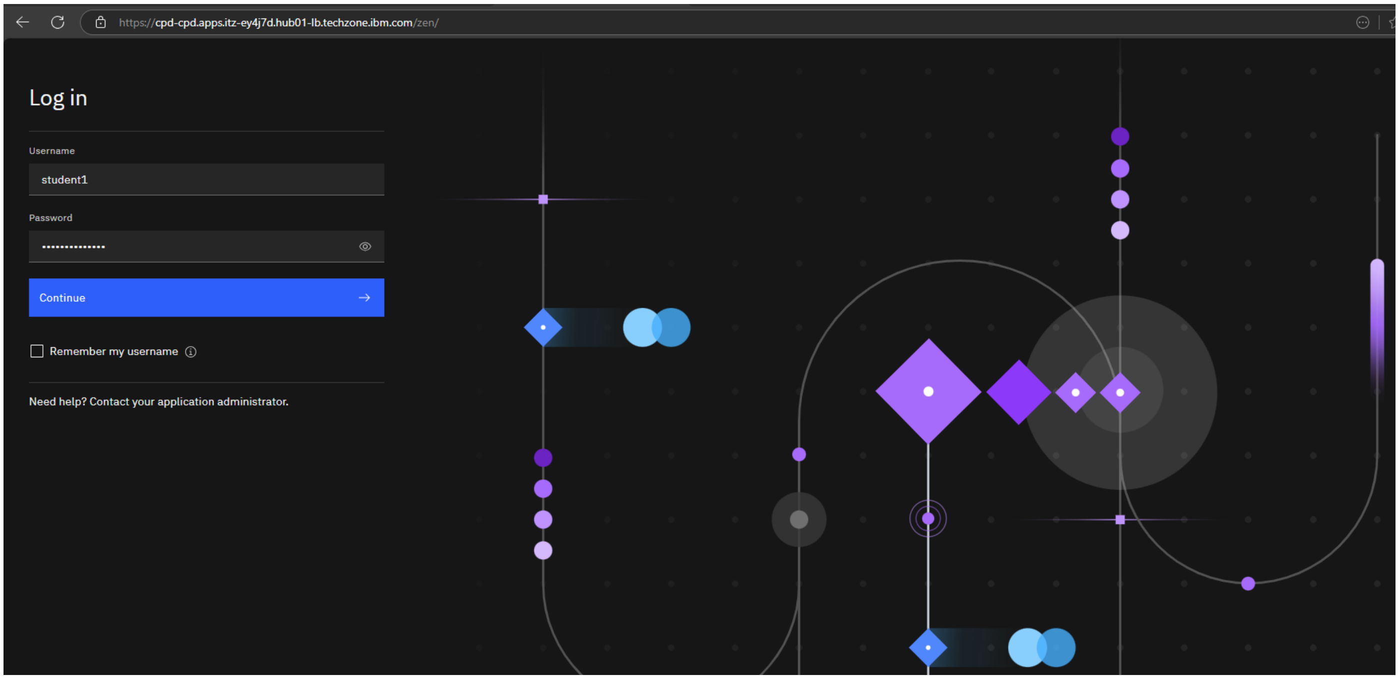
Environment Access

Use the cluster link provided by your instructor

1. Access the **shared Cloud Pak for Data cluster 1:** https://ibm.biz/BdnLHM

2. Choose a student id: **student1 – student20** as username

3. Enter **IDtechxchange** for the password.



https://cpd-cpd.apps.itz-ey4j7d.hub01-lb.techzone.ibm.com/zen/

Log in

Username

student1

Password

•••••••••••••••

Continue →

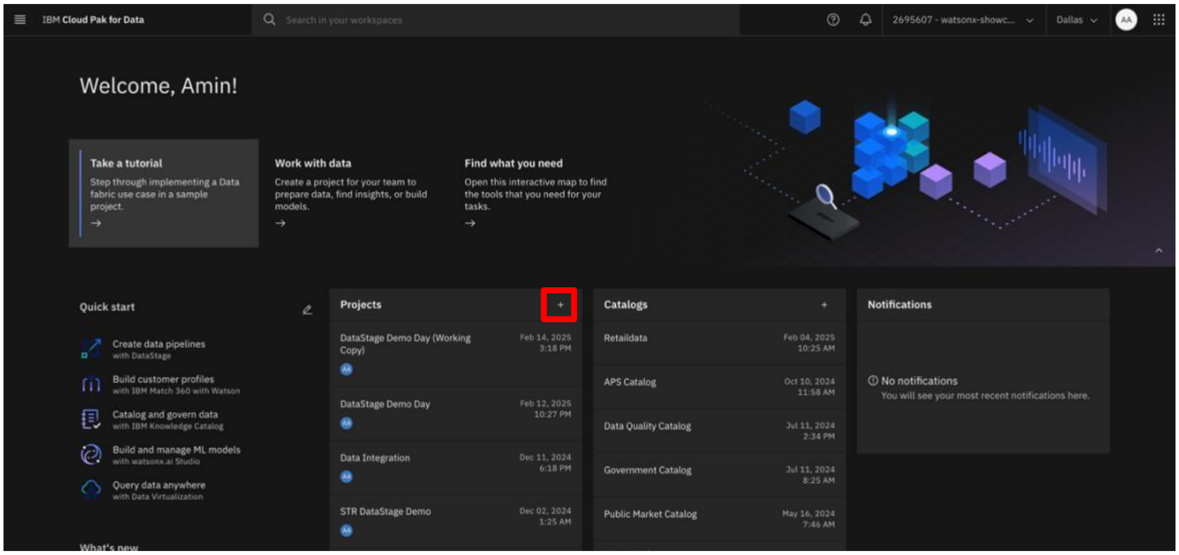
□ Remember my username ⓘ

Need help? Contact your application administrator.

Lab Setup (Building the flow)

To begin, download the **DataStage Lab Project ZIP file** from Box.

To get started, head to the shared SW cluster above homepage.
Navigate to **projects -> all projects**. Create a new project from file.



Welcome, Amin!

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.
→

Quick start

- 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 watsonx.ai Studio
- Query data anywhere with Data Virtualization

What's new

Projects +	
DataStage Demo Day (Working Copy)	Feb 14, 2025 3:18 PM
DataStage Demo Day	Feb 12, 2025 10:27 PM
Data Integration	Dec 11, 2024 6:18 PM
STR DataStage Demo	Dec 02, 2024 1:25 AM

Catalogs +	
Retaildata	Feb 04, 2025 10:25 AM
APS Catalog	Oct 10, 2024 11:58 AM
Data Quality Catalog	Jul 11, 2024 2:34 PM
Government Catalog	Jul 11, 2024 8:25 AM
Public Market Catalog	May 16, 2024 7:46 AM

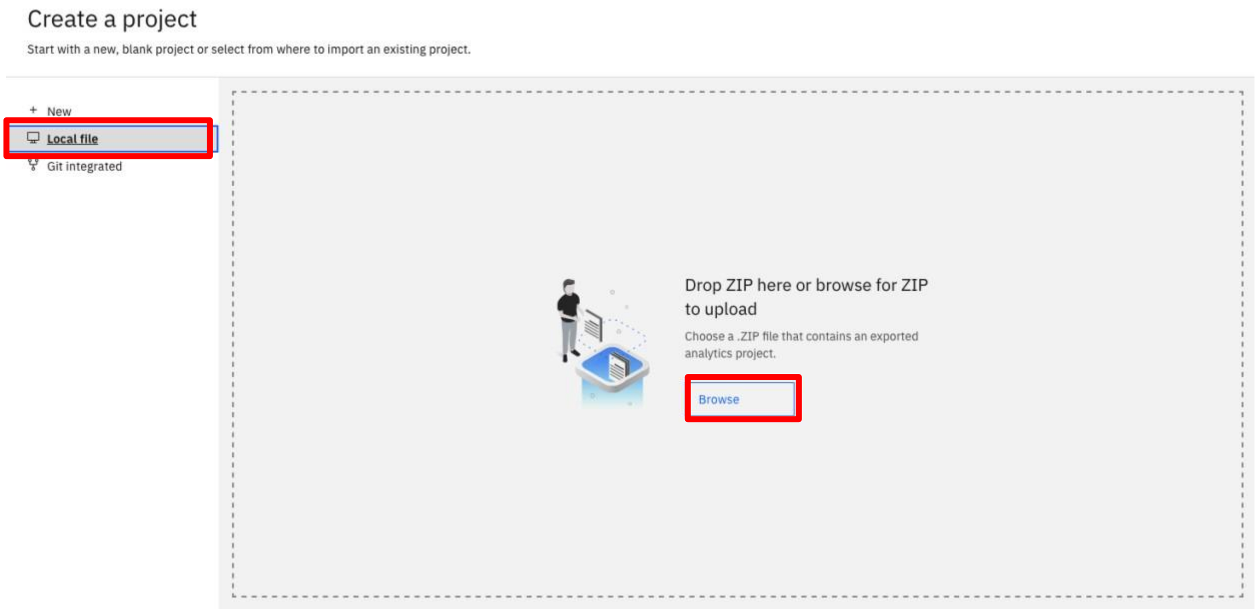
Notifications

ⓘ No notifications
You will see your most recent notifications here.

Explainer - DataStage Projects

A **project** is a collaborative workspace in Cloud Pak for Data where you work with data and other assets to accomplish a particular goal. In this case, we are using projects to transform and integrate data with DataStage.

Inside the **Create a Project** view, select **Local file** and upload the downloaded ZIP file.



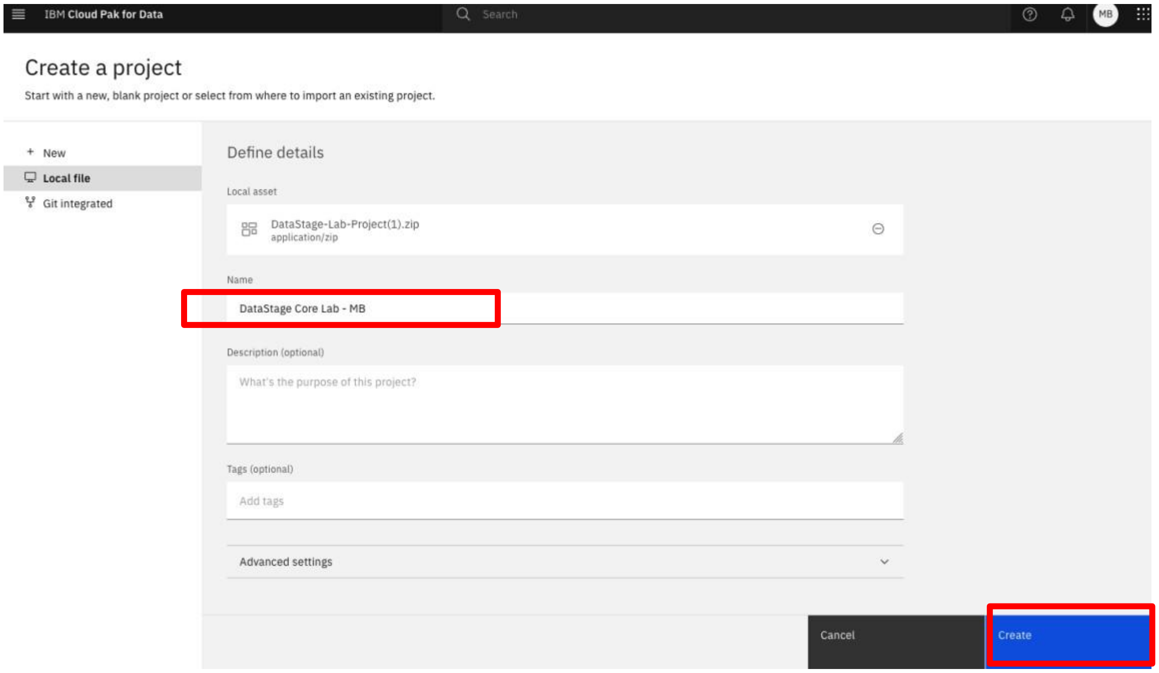
Create a project

Start with a new, blank project or select from where to import an existing project.

Drop ZIP here or browse for ZIP to upload

Choose a .ZIP file that contains an exported analytics project.

Once the DataStage Core Lab Project ZIP file is uploaded, provide a unique name for your project (i.e. **DataStage Core Lab _*Your Initials***)



Create a project

Start with a new, blank project or select from where to import an existing project.

Define details

Local asset: DataStage-Lab-Project(1).zip application/zip

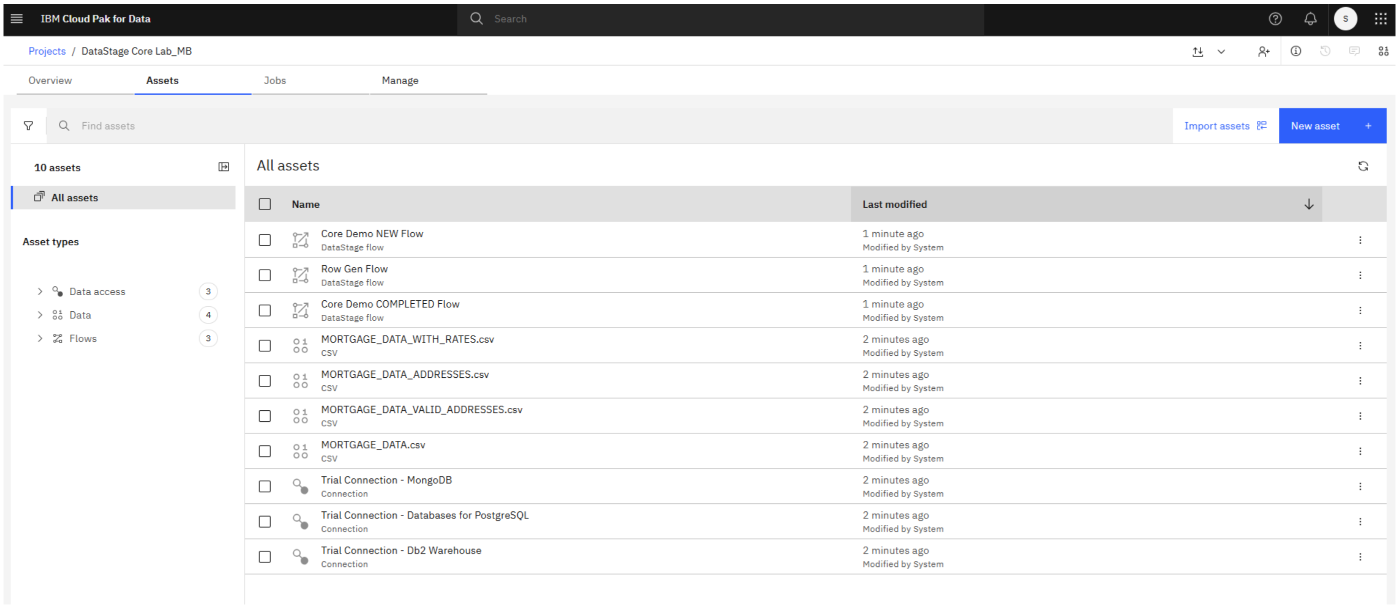
Name: DataStage Core Lab - MB

Description (optional): What's the purpose of this project?

Tags (optional): Add tags

Click the **Create** button to create the project.

Tip: Once inside the Assets view, sorting the assets by **Name** in ascending order will assist you in navigating the contents inside the Project



All assets

Name	Last modified
Core Demo NEW Flow DataStage flow	1 minute ago Modified by System
Row Gen Flow DataStage flow	1 minute ago Modified by System
Core Demo COMPLETED Flow DataStage flow	1 minute ago Modified by System
MORTGAGE_DATA_WITH_RATES.csv CSV	2 minutes ago Modified by System
MORTGAGE_DATA_ADDRESSES.csv CSV	2 minutes ago Modified by System
MORTGAGE_DATA_VALID_ADDRESSES.csv CSV	2 minutes ago Modified by System
MORTGAGE_DATA.csv CSV	2 minutes ago Modified by System
Trial Connection - MongoDB Connection	2 minutes ago Modified by System
Trial Connection - Databases for PostgreSQL Connection	2 minutes ago Modified by System
Trial Connection - Db2 Warehouse Connection	2 minutes ago Modified by System

Explainer - DataStage Flows

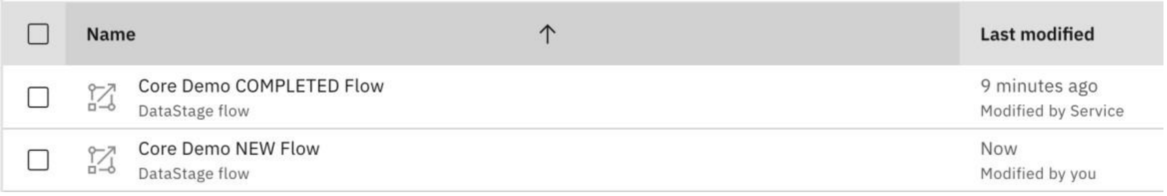
DataStage **flows** are the design-time assets that contain data integration logic.

The basic building blocks of a **flow** are:

- Data sources that read data
- Stages that transform the data
- Data targets that write data
- Links that connect the sources, stages, and targets

This Core Lab section is comprised of two DataStage flows in the project,
a **NEW** Flow and a **COMPLETED** Flow.

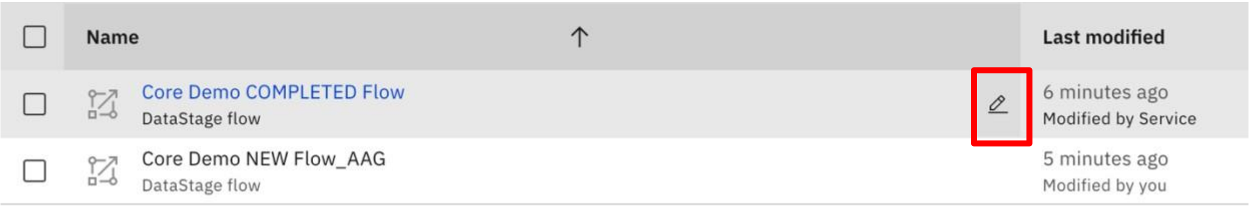
- The **NEW** Flow is the starting point for building the core capabilities of DataStage in the Core Lab.
- The **COMPLETED** Flow is the final output of the Core Lab, allowing you to have a reference example of the completed flow.



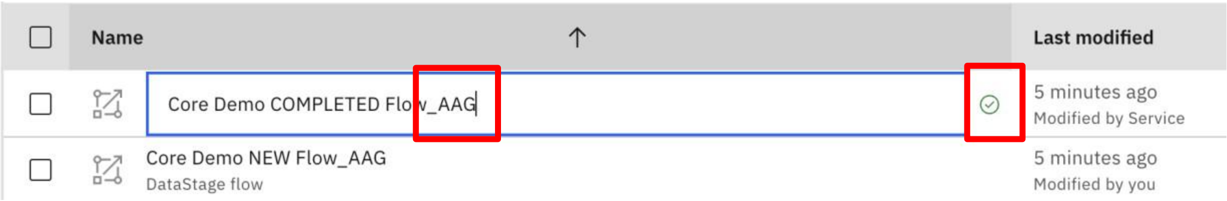
Name	Last modified
Core Demo COMPLETED Flow DataStage flow	9 minutes ago Modified by Service
Core Demo NEW Flow DataStage flow	Now Modified by you

Jump to Top

Rename these DataStage flows to include your initials.
This will help in identifying your assets.



□	Name ↑	Last modified
□	Core Demo COMPLETED Flow DataStage flow	6 minutes ago Modified by Service
□	Core Demo NEW Flow_AAG DataStage flow	5 minutes ago Modified by you



□	Name ↑	Last modified
□	Core Demo COMPLETED Flow_AAG	5 minutes ago Modified by Service
□	Core Demo NEW Flow_AAG DataStage flow	5 minutes ago Modified by you

Click into the NEW Flow to get started building your DataStage job!

Task 1: Run an existing DataStage Flow

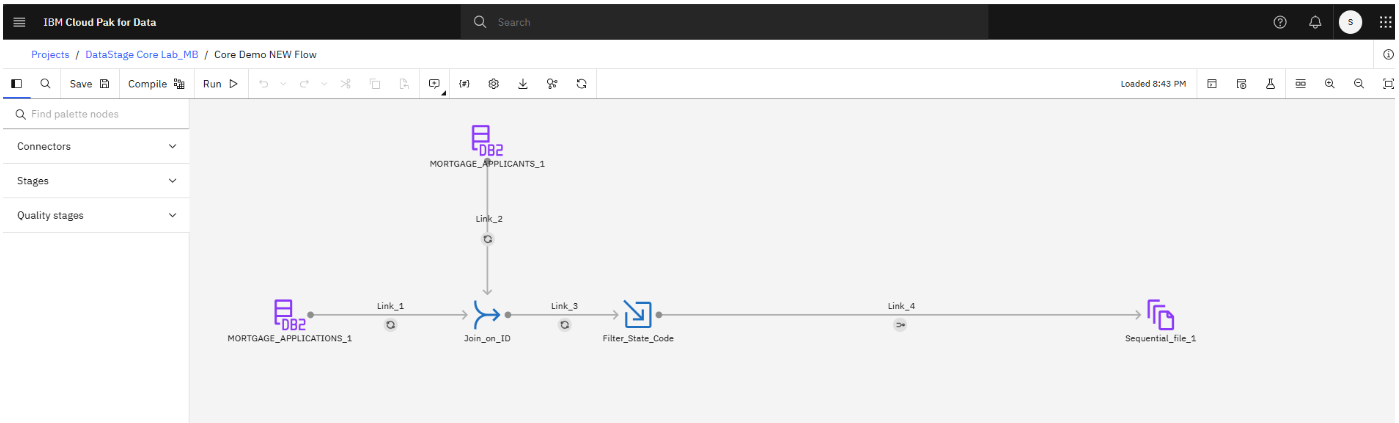
Let's with a basic DataStage flow that joins the mortgage applicants and mortgage applications data sets, then outputs that result to a CSV file in the project. Follow these steps to run the DataStage flow:

1. Start in your **DataStage Core Lab <*Your Initials*>** project.
   If you don't have the project open, follow these steps:
   a. From the Navigation menu Navigation menu, choose **Projects > View all projects**.
   b. Open the Data integration project.

2. Click the Assets tab to see all the assets in the project.

3. Click Flows > DataStage flows.

4. Click the **Core Demo NEW Flow_<your initials>** flow in the list to open it. This flow joins the *Mortgage Applicants* and *Mortgage Applications* tables that are stored in Db2 Warehouse, filters the data to those records from the State of California, and creates a sequential file in CSV format as the output.



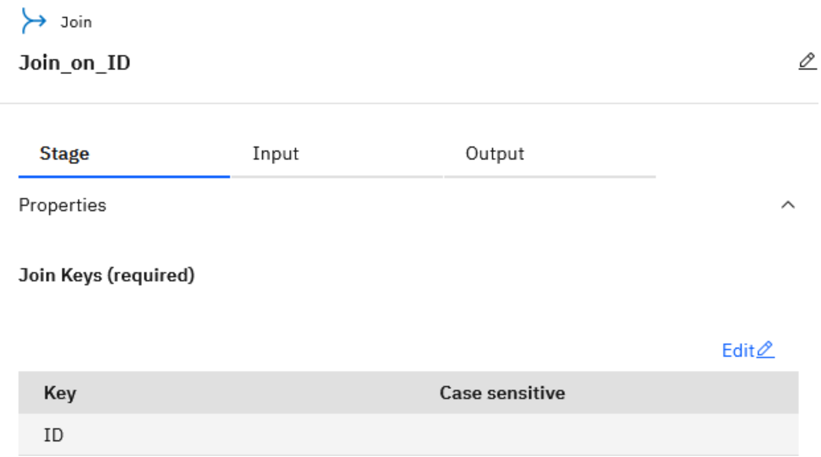
5. Click the **zoom in** icon zoom in and **zoom out** icon on the toolbar to set your preferred view of the canvas.

6. Double-click **MORTGAGE_APPLICATIONS_1** node to view the settings.

   a. Expand the **Properties** section.
   b. Scroll down to the bottom, then click **Preview data**. This data set includes information that is captured on a mortgage application.
   c. Click **Close**.

7. Double-click **MORTGAGE_APPLICANTS_1** node to view the settings.
   a. Expand the **Properties** section.
   b. Scroll down, and click **Preview data**. This data set includes information about mortgage applicants who applied for a loan.
   c. Optional: Visualize the data.
      i. Click the **Chart** panel.
      ii. In the *Columns to visualize* list, select **STATE**.
      iii. Click **Visualize data** to see a pie chart showing the distribution of the data by state.
      iv. Click the **Treemap** icon to see the same data in a treemap chart.
   d. Click **Close**.

8. Double-click **Join_on_ID** node to view the settings.
   a. Expand the **Properties** section.
   b. Note that the join key is the ID column.



Key	Case sensitive
ID	

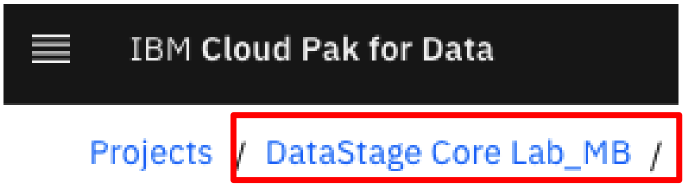
c. Click Cancel to close the settings.

9. Click the **Logs** icon ⊡ on the toolbar so you can watch the flow's progress.

10. Click **Compile**, and then click **Run**. Alternatively, you can click **Run** which compiles and then runs the DataStage flow. The run can take about one minute to complete.

11. View the logs. You can use the total rows and rows/sec for each step in the flow to visually verify that the filter is working as expected.

12. When the run completes successfully, click **DataStage Core Lab _ < *Your Initials*>** (your project name may differ) in the navigation trail to return to the project.

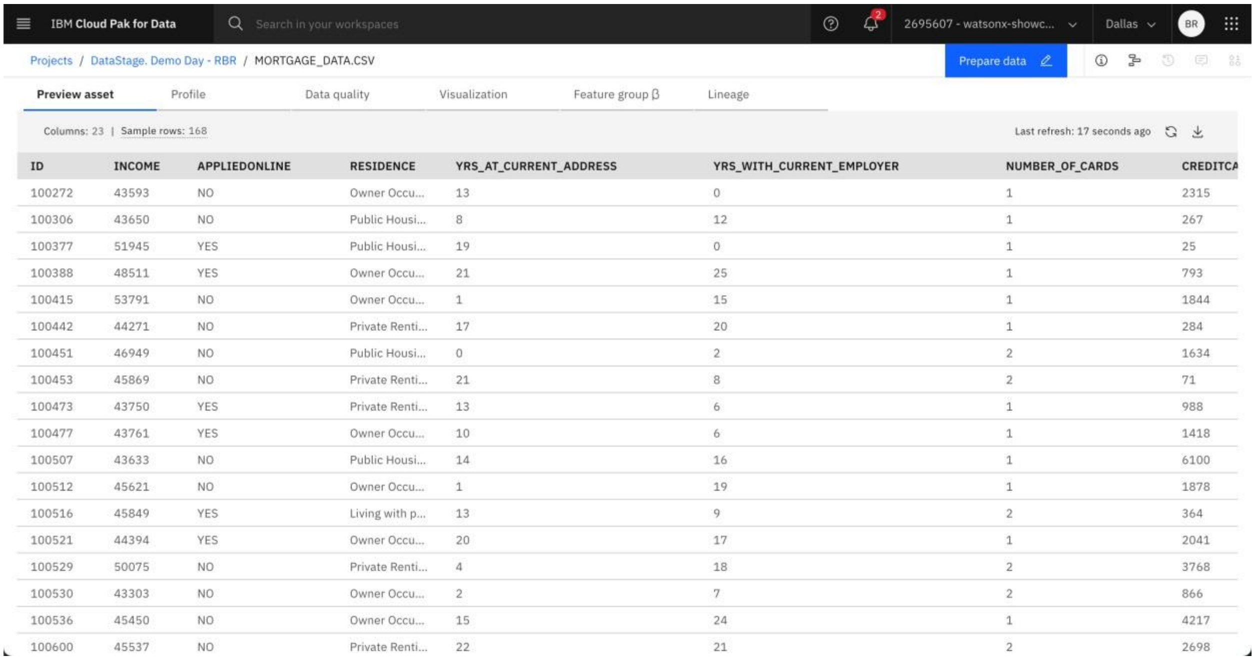


13. On the *Assets* tab, click **Data > Data assets**.

14. Open the **MORTGAGE_DATA.CSV file**. You can see that this file contains the columns from both the mortgage applicants and mortgage applications data sets.

Check your progress

The following image shows resulting CSV file. The next task is to edit the DataStage flow.



Projects / DataStage. Demo Day - RBR / MORTGAGE_DATA.CSV

Columns: 23 | Sample rows: 168

Last refresh: 17 seconds ago

ID	INCOME	APPLIEDONLINE	RESIDENCE	YRS_AT_CURRENT_ADDRESS	YRS_WITH_CURRENT_EMPLOYER	NUMBER_OF_CARDS	CREDITCA
100272	43593	NO	Owner Occu...	13	0	1	2315
100306	43650	NO	Public Housi...	8	12	1	267
100377	51945	YES	Public Housi...	19	0	1	25
100388	48511	YES	Owner Occu...	21	25	1	793
100415	53791	NO	Owner Occu...	1	15	1	1844
100442	44271	NO	Private Renti...	17	20	1	284
100451	46949	NO	Public Housi...	0	2	2	1634
100453	45869	NO	Private Renti...	21	8	2	71
100473	43750	YES	Private Renti...	13	6	1	988
100477	43761	YES	Owner Occu...	10	6	1	1418
100507	43633	NO	Public Housi...	14	16	1	6100
100512	45621	NO	Owner Occu...	1	19	1	1878
100516	45849	YES	Living with p...	13	9	2	364
100521	44394	YES	Owner Occu...	20	17	1	2041
100529	50075	NO	Private Renti...	4	18	2	3768
100530	43303	NO	Owner Occu...	2	7	2	866
100536	45450	NO	Owner Occu...	15	24	1	4217
100600	45537	NO	Private Renti...	22	21	2	2698

Overview: Edit the DataStage Flow

Now that you joined the mortgage applicant and application data, you are ready to edit the DataStage flow to:

- Task 2: Specify a key column for the Join stage.
- Task 3: Add credit score data from a PostgreSQL database.
- Task 4: Add a Join stage to join the credit score data with the applicant and application data.
- Task 5: Add a Transformer stage to calculate total debt.
- Task 6: Add interest rate data from a MongoDB database.
- Task 7: Add a Lookup stage to look up interest rates for applicants based on their credit scores and Golden Bank's daily interest rate ranges.

Explainer - DataStage Stages

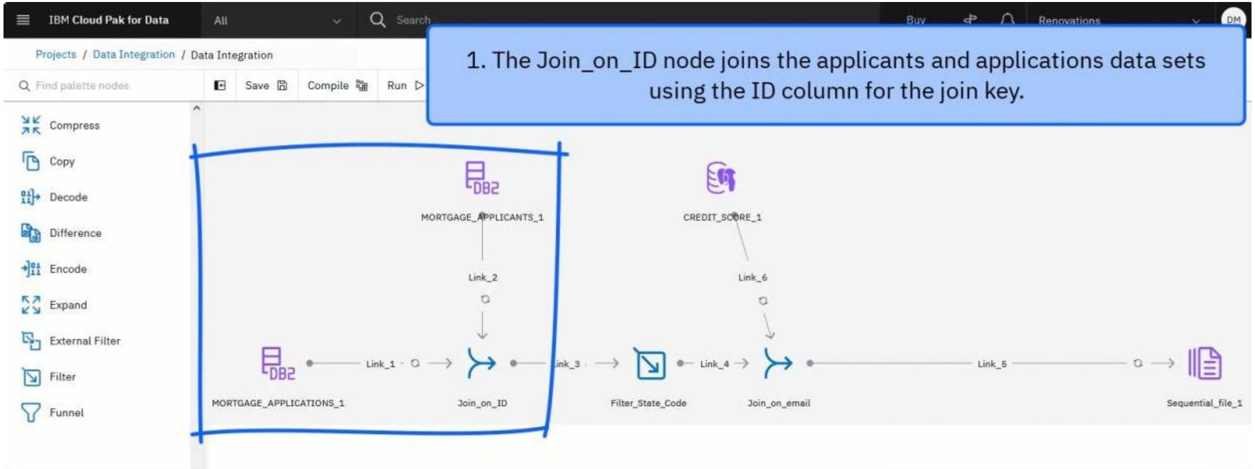
A DataStage flow consists of **stages** that are linked together, which describe the flow of data from a data source to a data target. A stage describes a data source, a processing step, or a target system. The stage also defines the processing logic that moves the data from the input links to the output links.

A stage usually has at least one data input or one data output. However, some stages can accept more than one data input, and output to more than one stage. The following documentation link lists the available stages and gives details on their functions.

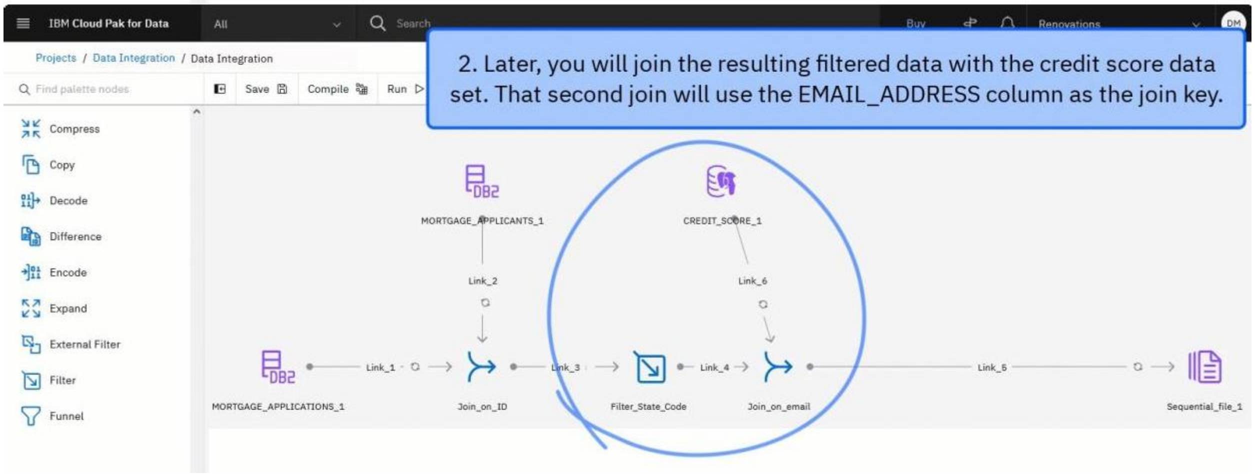
Task 2: Specify the key column for the Join stage

Identifying a key column indicates to DataStage that column contains unique values. The **Join_on_ID** node joins the mortgage applicants and mortgage applications data sets using the ID column for the join key. The next phase is to join the resulting data set with the credit score data. Later, you will join the resulting filtered data with the credit score data set. The second join will use the EMAIL_ADDRESS column as the join key. In this task, you edit the DataStage flow to specify the EMAIL_ADDRESS column as the key column for the resulting data set when it is joined with the credit score data.

The following images provide a visual representation as an alternative to the description of the two join nodes.



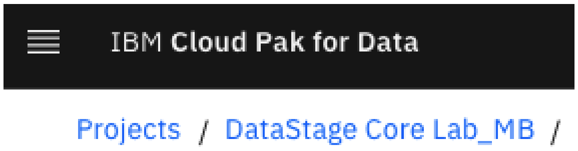
1. The Join_on_ID node joins the applicants and applications data sets using the ID column for the join key.



2. Later, you will join the resulting filtered data with the credit score data set. That second join will use the EMAIL_ADDRESS column as the join key.

Follow these steps to change the Join node settings:

Click **DataStage Core Lab_ <Your Initials> (your project name may differ)** in the navigation trail to return to the project.


IBM Cloud Pak for Data

Projects / DataStage Core Lab_MB /

1. On the **Assets** tab, click **Flows > DataStage flows**.

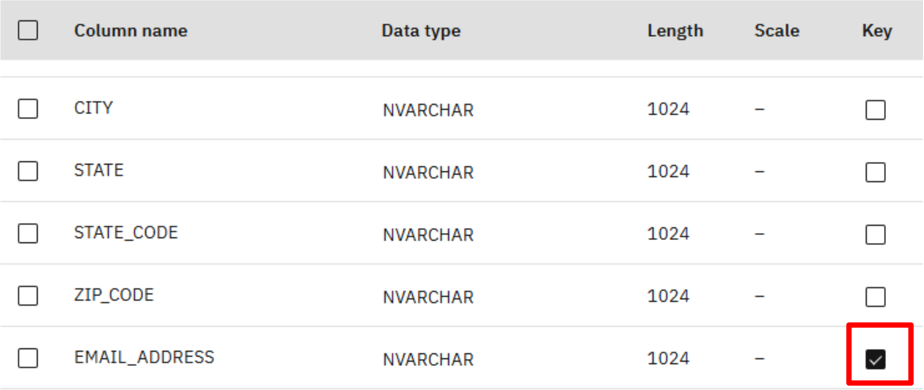
2. Open the **Core Demo NEW Flow_<your initials>** flow.

3. Double-click the **Join_on_ID** node to edit the settings.

4. Click the **Output** tab, and expand the **Columns** section to see a list of the columns in the joined data set.

5. Click **Edit**.

6. For the *EMAIL_ADDRESS* column name, select **Key**.



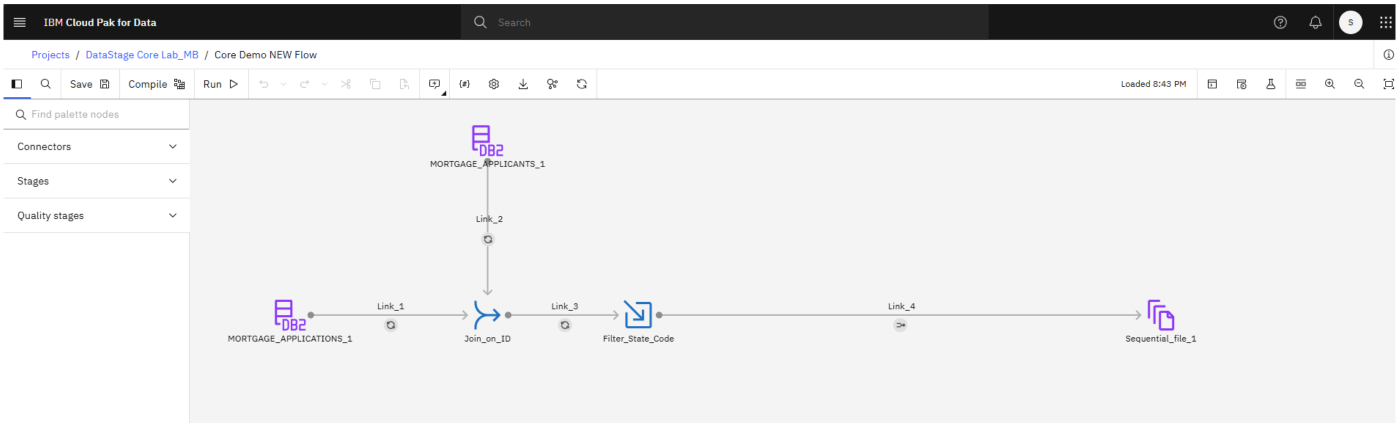
□	Column name	Data type	Length	Scale	Key
□	CITY	NVARCHAR	1024	–	□
□	STATE	NVARCHAR	1024	–	□
□	STATE_CODE	NVARCHAR	1024	–	□
□	ZIP_CODE	NVARCHAR	1024	–	□
□	EMAIL_ADDRESS	NVARCHAR	1024	–	☑

7. Click **Apply and return** to return to the *Join_on_ID* node settings.

8. Click **Save** to save the *Join_on_ID* node settings.

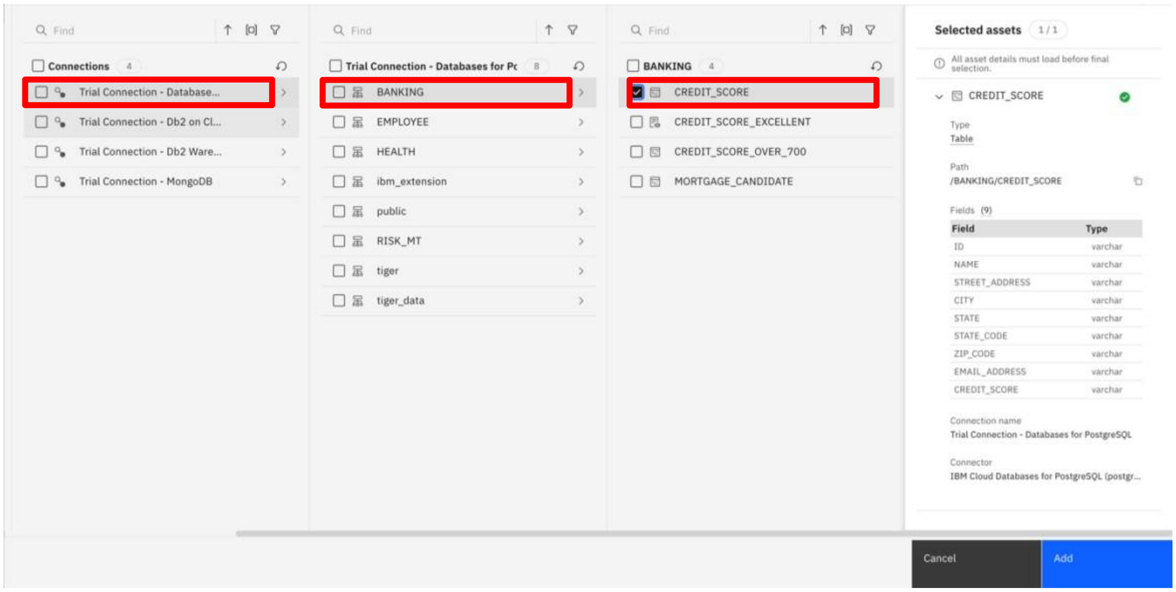
Check your progress

The following image shows the DataStage flow with the edited Join_on_id stage. Now that you identified the EMAIL_ADDRESS column as the key column, you can add the **PostgreSQL** data containing the applicants credit scores.



Explainer - Asset Browser for DataStage

The Asset browser is used to search for connections and assets and add them to your DataStage flows. When you open the asset browser from the palette, you can use it to browse connectors, DataStage subflows, and data assets (.csv, .txt, .xls., .xlsx, .xml, .json files). The data can then be previewed before being onboarded to the DataStage flow – this feature saves developers time searching for the right data and is an important benefit.



Selected assets 1 / 1

All asset details must load before final selection.

CREDIT_SCORE

Type
Table

Path
/BANKING/CREDIT_SCORE

Fields (9)

Field	Type
ID	varchar
NAME	varchar
STREET_ADDRESS	varchar
CITY	varchar
STATE	varchar
STATE_CODE	varchar
ZIP_CODE	varchar
EMAIL_ADDRESS	varchar
CREDIT_SCORE	varchar

Connection name
Trial Connection - Databases for PostgreSQL

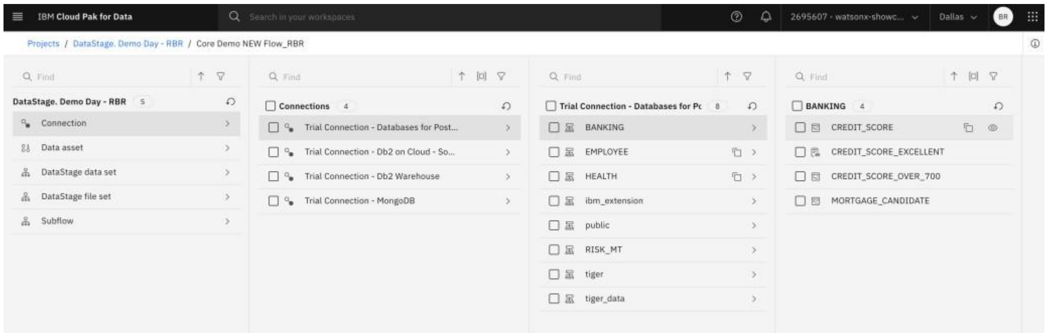
Connector
IBM Cloud Databases for PostgreSQL (postgr...

Task 3: Add credit score data from a PostgreSQL database

Follow these steps to add the credit score data that is stored in a PostgreSQL database to the DataStage flow:

1. In the node palette, expand the **Connectors** section.
2. Drag the **Asset browser** connector to the canvas beside the *MORTGAGE_APPLICANTS_1* node.
3. Locate the asset by selecting **Connection > Trial Connection - Databases for PostgreSQL > BANKING > CREDIT_SCORE**.

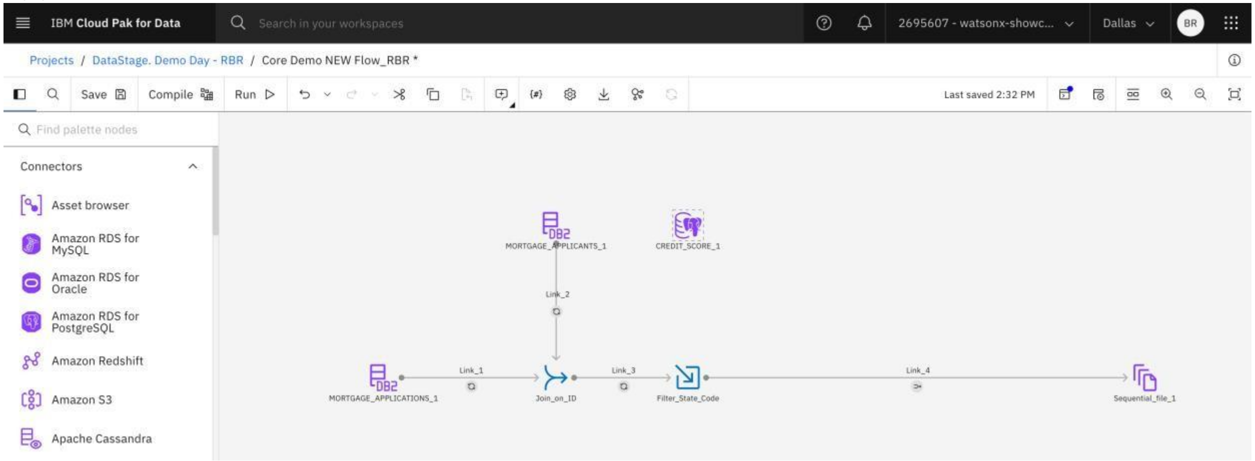
Note: Click the connection or schema name instead of the checkbox to expand the connection and schema.


IBM Cloud Pak for Data
Search in your workspaces
2695607 - watsonx-showc...
Dallas
BR
Projects / DataStage. Demo Day - RBR / Core Demo NEW Flow_RBR
Find
DataStage. Demo Day - RBR 5
Connection
Data asset
DataStage data set
DataStage file set
Subflow
Find
□ Connections 4
□ Trial Connection - Databases for Post...
□ Trial Connection - Db2 on Cloud - So...
□ Trial Connection - Db2 Warehouse
□ Trial Connection - MongoDB
Find
□ Trial Connection - Databases for Pc 8
□ BANKING
□ EMPLOYEE
□ HEALTH
□ ibm_extension
□ public
□ RISK_MT
□ tiger
□ tiger_data
Find
□ BANKING 4
□ CREDIT_SCORE
□ CREDIT_SCORE_EXCELLENT
□ CREDIT_SCORE_OVER_700
□ MORTGAGE_CANDIDATE

4. Click the **Preview** icon ◎ to preview the credit score data for each applicant.
5. Click **Add**.

Check your progress

The following image shows the DataStage flow with the credit score asset added. Now that you added the credit score data to the canvas, you need to join the applicant, application, and credit score data.



Task 4: Add a Join stage to join the credit score data with the applicant and application data

Follow these steps to add another Join stage to join the filtered mortgage application and mortgage applicant joined data with the credit score data in the DataStage flow:

1. In the node palette, expand the **Stages** section.

2. Drag the **Join** stage on to the canvas, and drop the node on the link line between the *Filter_State_Code* and *Sequential_file_1* nodes.

3. Hover over the **CREDIT_SCORE_1** connector to see the arrow. Connect the arrow to the **Join** stage.

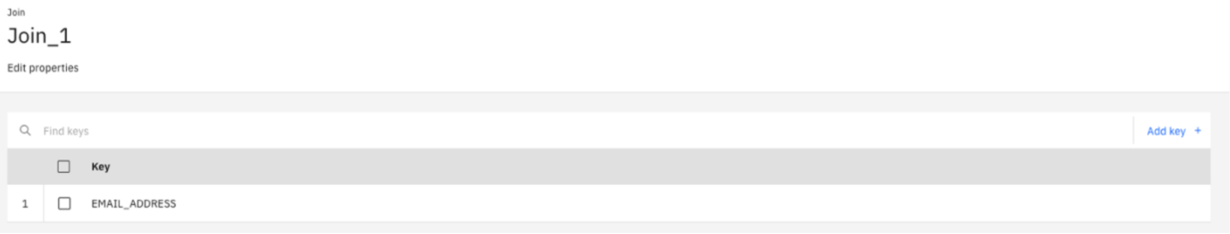
4. Double-click the **CREDIT_SCORE_1** node to edit the settings.

   a. Click the **Output** tab, and expand the **Columns** section to see a list of the columns in the joined data set.
   b. Click **Edit**.
   c. For the *EMAIL_ADDRESS* and *CREDIT_SCORE* column names, select **Key**.
   d. Click **Apply and return** to return to the *CREDIT_SCORE_1* node settings.
   e. Click **Save** to save the *CREDIT_SCORE_1* node settings.

5. Double-click the **Join_1** node to edit the settings.

   a. Expand the **Properties** section.
   b. Click **Add key**.

      i. Click **Add key** again.
      ii. Select **EMAIL_ADDRESS** from the list of possible keys.
      iii. Click **Apply**.



Join

Join_1

Edit properties

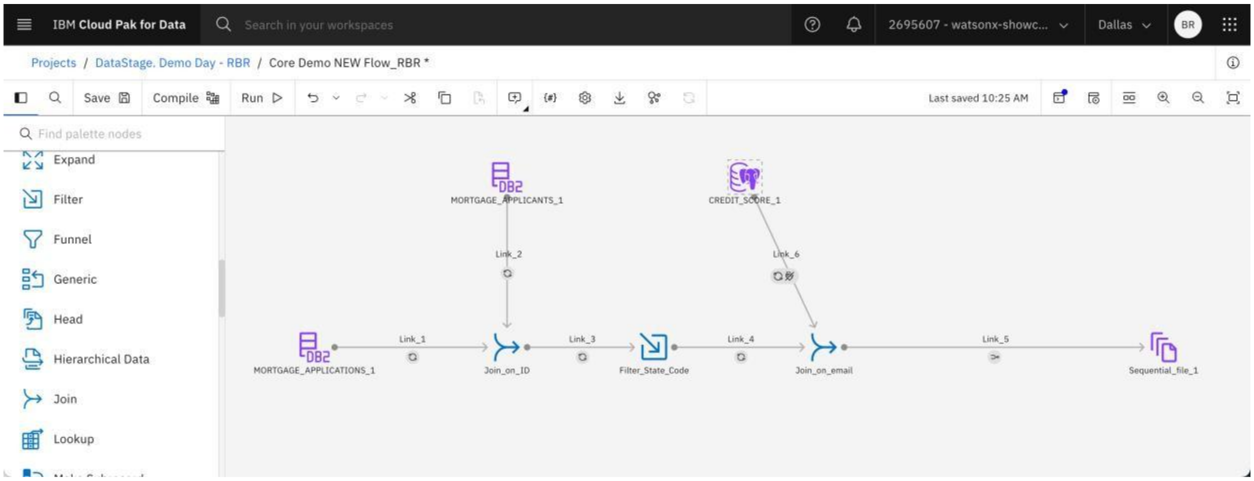
Find keys

	☐	Key
1	☐	EMAIL_ADDRESS

   c. Click **Apply and return** to return to the *Join_1* node settings.
   d. Change the *Join_1* node name to `Join_on_email`.
   e. Click **Save** to save the Join_1 node settings.

Check your progress

The following image shows the DataStage flow with a second Join stage added. Now that you joined the application, applicant, and credit score data, you need to add a Transformer stage to calculate each applicant's total debt.



Task 5: Add a Transformer stage to calculate total debt

Follow these steps to add a Transformer stage that creates a new column by summing the LOAN_AMOUNT and CREDITCARD_DEBT columns:

1. In the *Stages* section, drag the **Transformer** stage on to the canvas, and drop the node on the link line between the ***Join_on_email*** and ***Sequential_file_1*** nodes.

2. Double-click the **Transformer** node to edit the settings.

3. Click the **Output** tab.

    a. Click **Add column**.
    b. Scroll down in the list of columns to see the new column.
    c. Name the column **`TOTAL_DEBT`**.
    d. Click the **Edit** icon ✎in the row's ***Derivation*** column.
    e. Click the Calculator icon 🖩 in the ***Derivation*** column to open the expression builder.
    f. Search for **`LOAN_AMOUNT`**, and double-click the column name to add it to the expression. Note that the link number is appended to the column name.
    g. Type a plus sign **`+`**.
    h. Search for **`CREDITCARD_DEBT`**, and then double-click the column name to add it to the expression. Note that the link number is appended to the column name.
    i. Verify that the final expression is **`Link_7.LOAN_AMOUNT + Link_7.CREDITCARD_DEBT`**.

       Note: Your link number may be different.

    j. Click **Apply and return** to return to the ***Transformer*** page. If you see an error, change the column **Data type to nvarchar.**

    k. For the ***CREDIT_SCORE*** column name, scroll to the right and select Key.

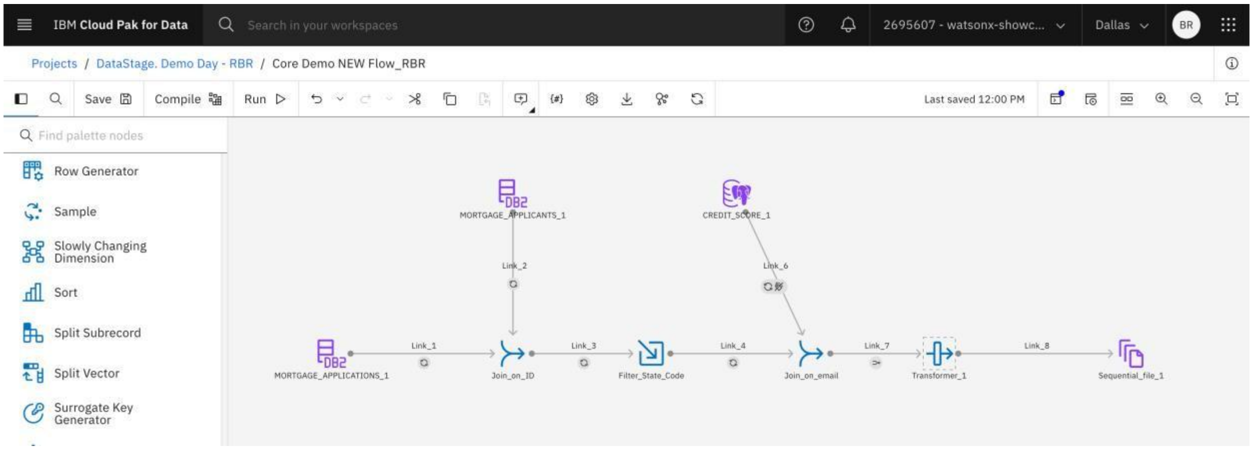
4. Click the **Stage** tab.

    a. Select the **Advanced** page.
    b. Change the ***Execution mode*** to **Sequential**.

5. Click **Save and return** to return to the canvas.

Check your progress

The following image shows the DataStage flow with the Transformer stage added. Now that you calculated each applicant's total debt, you need to add the table of interest rates to offer based on credit score ranges.



Task 6: Add interest rate data from a MongoDB database

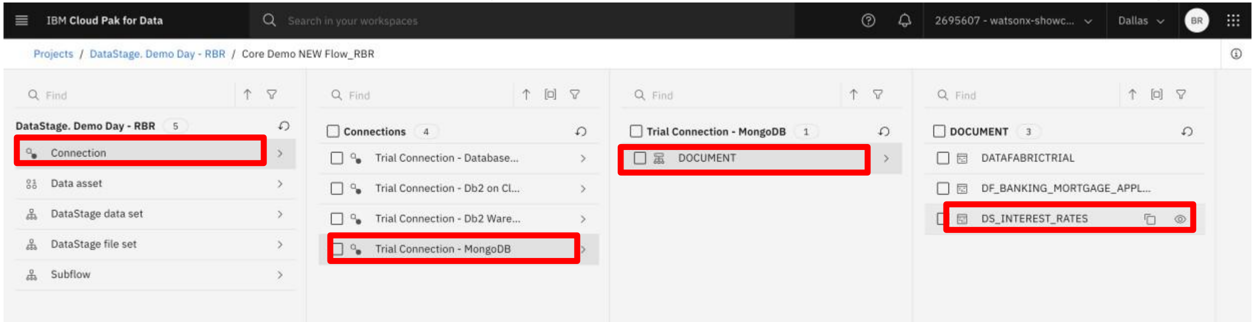
Follow these steps to include the interest rates in the flow by adding a data asset connector to a MongoDB database:

1. In the node palette, expand the **Connectors** section.

2. Drag the **Asset browser** connector on to the canvas beside the ***CREDIT_SCORE_1*** node.

3. Locate the asset by selecting **Connection > Trial Connection - Mongo DB > DOCUMENT > DS_INTEREST_RATES**.

4. Click the Preview icon 👁 to preview interest rates for each credit score range.

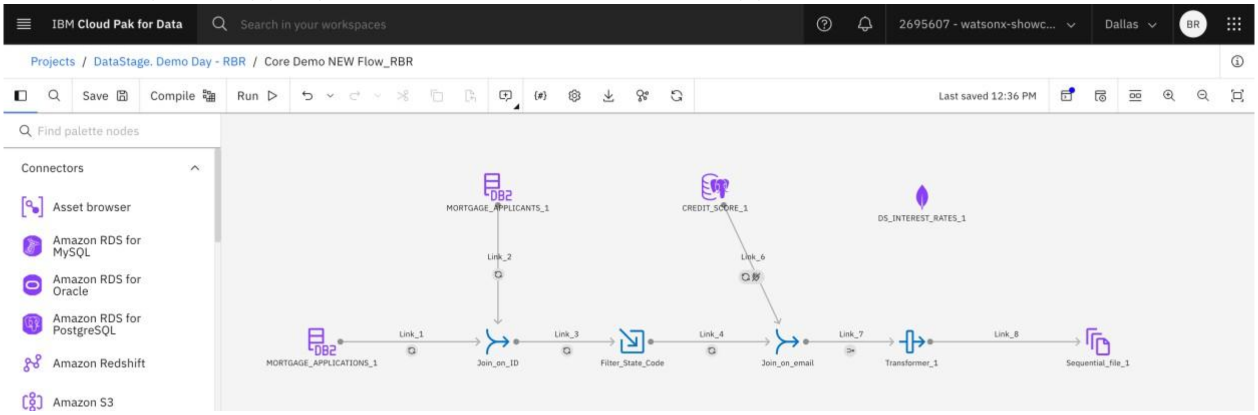

IBM Cloud Pak for Data
Search in your workspaces
2695607 - watsonx-showc...
Dallas
BR
Projects / DataStage. Demo Day - RBR / Core Demo NEW Flow_RBR
Find
DataStage. Demo Day - RBR 5
Connection
Data asset
DataStage data set
DataStage file set
Subflow
Find
Connections 4
Trial Connection - Database...
Trial Connection - Db2 on Cl...
Trial Connection - Db2 Ware...
Trial Connection - MongoDB
Find
Trial Connection - MongoDB 1
DOCUMENT
Find
DOCUMENT 3
DATAFABRICTRIAL
DF_BANKING_MORTGAGE_APPL...
DS_INTEREST_RATES

You can use the values in the STARTING_LIMIT and ENDING_LIMIT columns to look up the appropriate interest rate based on the applicant's credit score. The ID column is not needed, so you will delete that column in the next step.

5. Click **Add**.

Check your progress

The following image shows the DataStage flow with the interest rates data asset added from the MongoDB external source. Now that you added the interest rates table, you can look up the appropriate interest rate for each applicant.



Task 7: Add a Lookup stage to look up interest rates for applicants

Based on each applicant's credit score, you want to look up the appropriate interest rate. Follow these steps to add a Lookup stage and specify the range for starting and ending credit score limits for each interest rate:

1. In the ***Stages*** section, drag the **Lookup** stage on to the canvas, and drop the node on the link line between the ***Transformer_1*** and ***Sequential_file_1*** nodes.

2. Connect the ***DS_INTEREST_RATES_1*** connector to the ***Lookup_1*** stage.

3. Double-click the ***DS_INTEREST_RATES_1*** node to edit the settings.

4. Click the **Output** tab.

   a. Expand the **Columns** section, and click **Edit**.
   b. Select the **_ID** column.

c. Click the **Delete** icon 🗑 to delete the ***_ID*** column.
d. Click **Apply and return** to return to the ***DS_INTEREST_RATES_1*** node settings.
e. Click **Save** to save the changes to the ***DS_INTEREST_RATES_1*** node.

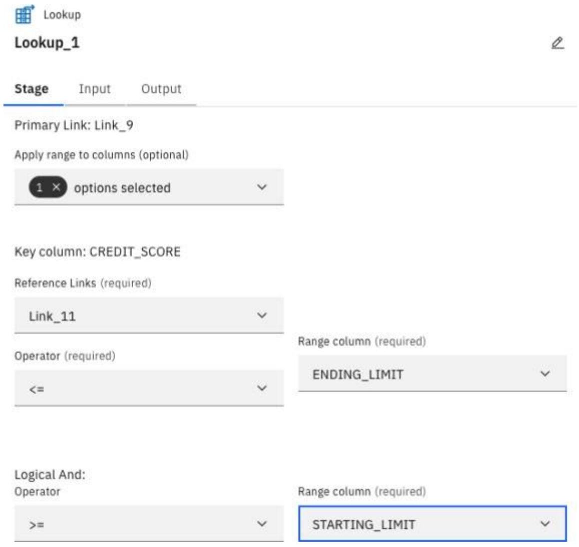
5. Double-click the ***Lookup_1*** node to edit the settings.

6. Expand the **Properties** section.

a. For the ***Apply range to columns*** field, select **CREDIT_SCORE**. The ***Reference Links, Operator***, and ***Range*** column fields display.
b. For the ***Reference Links***, select **Link_9**.

Note: Your link number may be different.

c. For the first ***Operator***, select **◀ =**.
d. For the first ***Range column***, select **ENDING_LIMIT**.
e. For the second ***Operator***, select **>=**.
f. For the second ***Range column***, select **STARTING_LIMIT**.



Lookup

Lookup_1

Primary Link: Link_9

Apply range to columns (optional): 1 × options selected

Key column: CREDIT_SCORE

Reference Links (required): Link_11

Operator (required): <=

Range column (required): ENDING_LIMIT

Logical And:
Operator: >=

Range column (required): STARTING_LIMIT

7. Click the **Output** tab.

a. Expand the **Columns** section, and click **Edit**.
b. Select the **STARTING_LIMIT** and **ENDING_LIMIT** columns.
c. Click the **Delete** icon 🗑 to delete these unnecessary **STARTING_LIMIT** and **ENDING_LIMIT** columns.

d. Click **Apply and return** to return to the ***Lookup_1*** node settings.
e. Click **Save** to save the changes to the Lookup_1 node.

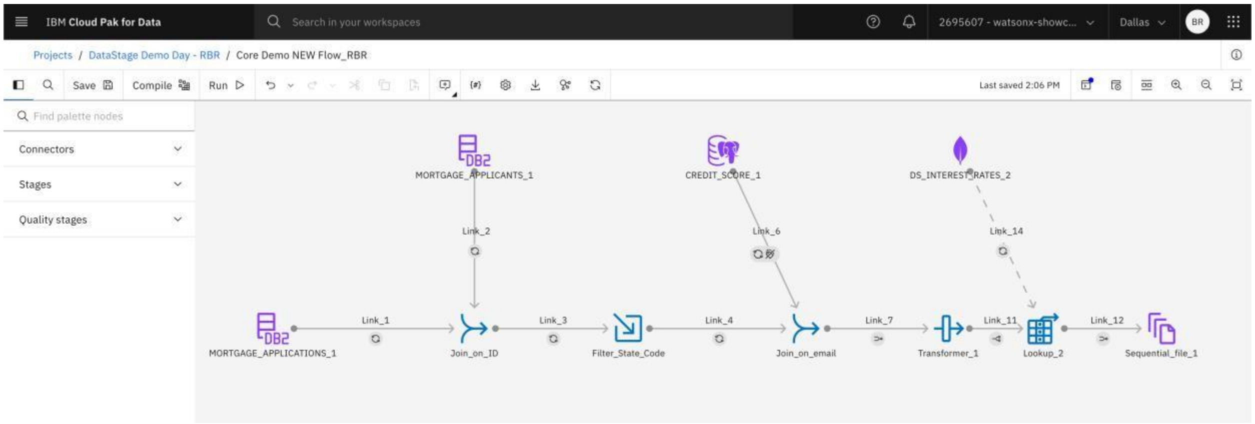
Explainer **-** Column metadata change propagation

When you add/remove columns or change a column's metadata, **Column metadata change propagation** automatically propagates these changes downstream. For example, when the **STARTING_LIMIT** and **ENDING_LIMIT** columns were deleted, these changes are propagated to the output Sequential File automatically, so those columns will not be seen as a part of the input.

Note: Changes made **upstream** do not apply to a column once you modify its metadata. If you delete a column, modifying the column in a later stage will not add the column back.

Check your progress

The following image shows that the DataStage flow with the Lookup stage added. The DataStage flow is now complete. The last task before running the flow is to specify the name for the output file.



Task 8: Edit the Sequential file node and run the DataStage flow

Follow these steps to edit the Sequential file node to create a final output file as a data asset in the project, and then compile and run the DataStage flow:

1. Double-click the ***Sequential_file_1*** node to edit the settings.

2. Click the **Input** tab.

3. Expand the **Properties** section.

4. For the Target File, copy and paste `MORTGAGE_APPLICANTS_INTEREST_RATES.CSV` for the file name.

5. Select **Create data asset**.

6. For the First line is column names field, select **True**.

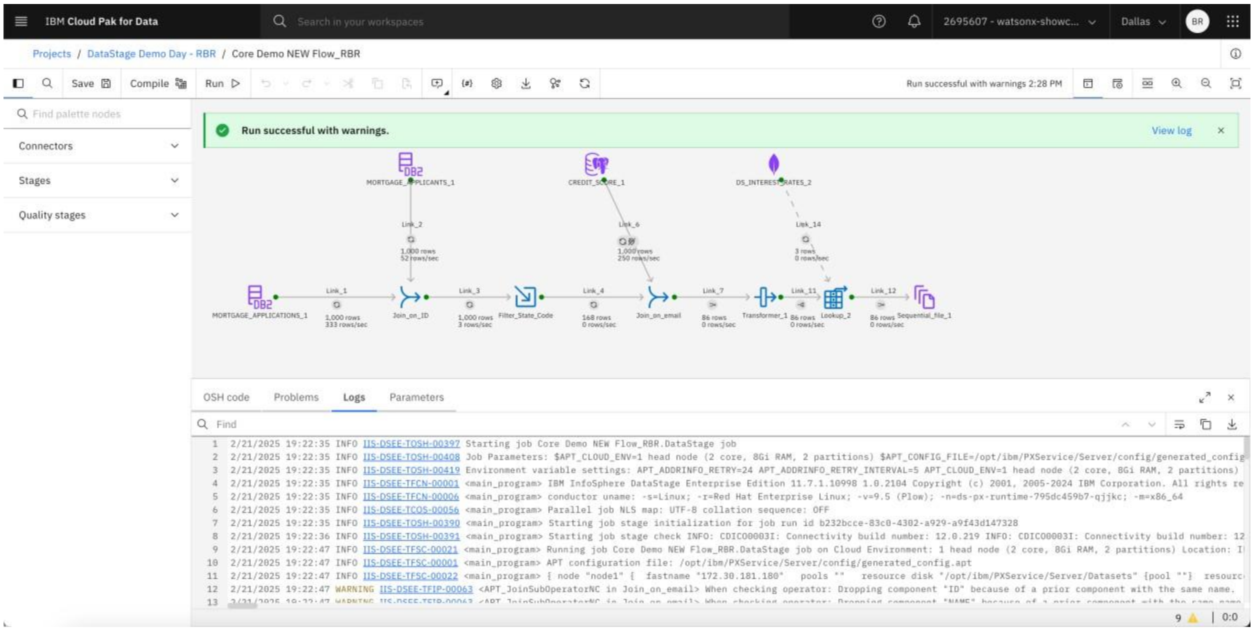
7. Click **Save**.

8. Click **Run** which compiles and then runs the DataStage flow. The job takes about 1 minute to complete.

9. Click **Logs** on the toolbar to watch the flow's progress. It is normal to see warnings during the run, and then you see that the flow ran successfully.

Check your progress

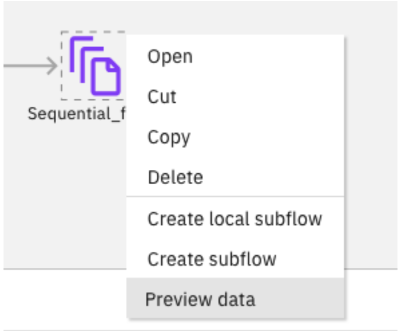
The following image shows that the DataStage flow ran successfully!



Run successful with warnings.

```
2/21/2025 19:22:35 INFO IIS-DSEE-TOSH-00397 Starting job Core Demo NEW Flow_RBR.DataStage job
2/21/2025 19:22:35 INFO IIS-DSEE-TOSH-00408 Job Parameters: $APT_CLOUD_ENV=1 head node (2 core, 8Gi RAM, 2 partitions) $APT_CONFIG_FILE=/opt/ibm/PXService/Server/config/generated_config
2/21/2025 19:22:35 INFO IIS-DSEE-TOSH-00419 Environment variable settings: APT_ADDRINFO_RETRY=24 APT_ADDRINFO_RETRY_INTERVAL=5 APT_CLOUD_ENV=1 head node (2 core, 8Gi RAM, 2 partitions)
2/21/2025 19:22:35 INFO IIS-DSEE-TFCN-00001 <main_program> IBM InfoSphere DataStage Enterprise Edition 11.7.1.10998 1.0.2104 Copyright (c) 2001, 2005-2024 IBM Corporation. All rights re
2/21/2025 19:22:35 INFO IIS-DSEE-TFCN-00006 <main_program> conductor uname: -s=Linux; -r=Red Hat Enterprise Linux; -v=9.5 (Plow); -n=ds-px-runtime-795dc459b7-qjjkc; -m=x86_64
2/21/2025 19:22:35 INFO IIS-DSEE-TCOS-00056 <main_program> Parallel job NLS map: UTF-8 collation sequence: OFF
2/21/2025 19:22:35 INFO IIS-DSEE-TOSH-00390 <main_program> Starting job stage initialization for job run id b232bcce-83c0-4302-a929-a9f43d147328
2/21/2025 19:22:36 INFO IIS-DSEE-TOSH-00391 <main_program> Starting job stage check INFO: CDIC00003I: Connectivity build number: 12.0.219 INFO: CDIC00003I: Connectivity build number: 12
2/21/2025 19:22:47 INFO IIS-DSEE-TFSC-00021 <main_program> Running job Core Demo NEW Flow_RBR.DataStage job on Cloud Environment: 1 head node (2 core, 8Gi RAM, 2 partitions) Location: I
2/21/2025 19:22:47 INFO IIS-DSEE-TFSC-00001 <main_program> APT configuration file: /opt/ibm/PXService/Server/config/generated_config.apt
2/21/2025 19:22:47 INFO IIS-DSEE-TFSC-00022 <main_program> { node "node1" {  fastname "172.30.181.180"   pools ""   resource disk "/opt/ibm/PXService/Server/Datasets" {pool ""}  resourc
2/21/2025 19:22:47 WARNING IIS-DSEE-TFIP-00063 <APT_JoinSubOperatorNC in Join_on_email> When checking operator: Dropping component "ID" because of a prior component with the same name.
2/21/2025 19:22:47 WARNING IIS-DSEE-TFIP-00063 <APT_JoinSubOperatorNC in Join_on_email> When checking operator: Dropping component "NAME" because of a prior component with the same name
```

10. Hover over the sequential file then click the ellipses to preview the data. The CSV file includes all applicants and their interest rates.



Conclusion and Next Steps

This section covered the core functionalities of DataStage on Cloud Pak for Data, including how to create projects and modify an existing flow to provide a personalized interest rate for each mortgage applicant. Key concepts and features of DataStage's modernized developer experience were covered – including the Asset Browser and Column Metadata Change Propagation, which are new, key features in DataStage on Cloud Pak for Data.

This concludes the lab guide.