Reuse a Data Flow in Oracle Analytics

Before you Begin

This lab shows you how to specify parameters in a data flow that prompt for the input data source, output dataset, or both before executing the data flow.

Background

You use data flows to apply transformations to datasets, add joins between datasets, add filters, remove unwanted columns, add derived measures and columns, and perform other operations such as sentiment analysis and time series forecast.

If you implement parameter prompts in your data flow, you can reuse the data flow without editing it directly. To use the same data flow, the input data source must use the same structure, but with different data records, as the data source originally used to create the data flow.

If you schedule a data flow, the prompts appears before the job runs.

What Do You Need?

- Access Oracle Analytics Cloud or Oracle Analytics Desktop
- Download the following spreadsheet files:
    - orders_5.xlsx
    - orders_6.xlsx
    - states.xlsx

Create Datasets

In this section, you create three datasets using the spreadsheets that you downloaded to your computer.

1. Sign in to Oracle Analytics Cloud.
2. On the Home page, click **Create**, and then click **Dataset**. In Create Dataset, click **Drop data file here or click to browse**. In File Upload, select orders_5.xlsx file, and then click **Open**.
3. In Create Dataset Table from orders_5.xlsx, click **OK**. Click **Save** 💾. In Save Dataset As, enter orders_5 in **Name**, and then click **OK**.
4. In the Join Diagram, use the scroll bar to review the columns. Click **Go back** ⬅ to return to the Home page.
5. On the Home page, click **Create**, and then click **Dataset**. In Create Dataset, click **Drop data file here or click to browse**. In File Upload, select orders_6.xlsx file, and then click **Open**.
6. In Create Dataset Table from orders_6.xlsx, click **OK**. Click **Save** 💾. In Save Dataset As, enter orders_6 in **Name**, and then click **OK**.

7. In the Join Diagram, use the scroll bar to review the columns. Click **Go back** ⬅.
8. On the Home page, click **Create**, and then click **Dataset**. In Create Dataset, click **Drop data file here or click to browse**. In File Upload, select states.xlsx file, and then click **Open**.
9. In Create Dataset Table from states.xlsx, click **OK**. Click **Save** 💾. In Save Dataset As, enter states in **Name**, and then click **OK**.
10. In the Join Diagram, use the scroll bar to review the columns. Click **Go back** ⬅.


states
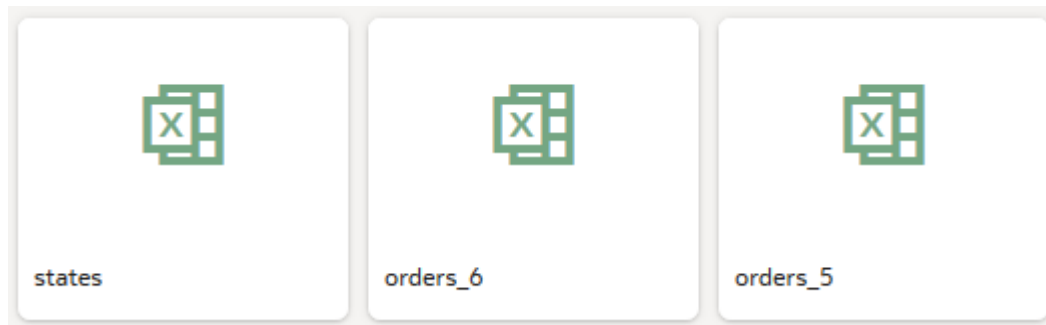
orders_6

orders_5

Create a Data Flow

1. On the Home page, click **Create**, and then click **Data Flow**. In Add Data, select **orders_5**, and then click **Add**.
2. Drag **Add Data** from the Data Flow Steps panel to **Add a step** ⊕ in the data flow editor.
3. In Add Data, click the **states** dataset, and then click **Add**.

   The datasets are automatically joined on the city column in each dataset.

Join

Keep rows

| ① Input 1 | Matching rows ▼ |
| ② Input 2 | Matching rows ▼ |

Match columns

Input 1		Operator	Input 2	
City	▼	= ▼	City	▼

4. Click **Add a step** ⊕ on the Join node, and then click **Rename Columns**.

5. In Rename Columns, scroll to **Customer Segment**, and delete **Customer** to change the column name to Segment.
6. Click **Add a step** ⊕ on the Rename Columns node, and then click **Merge Columns**.
7. In Merge Columns, enter City_State in New column name. Next to **Merge column**, click the hyperlink (Order Line ID), and then select **City**. Next to **With**, click the hyperlink (Order Line ID), and then select **State**. From the Delimiter list, select **Comma (,)**.

Merge Columns

New column name	City_State
Merge column	City
With	State
⊕ Column	
Delimiter	Comma (,) ▾
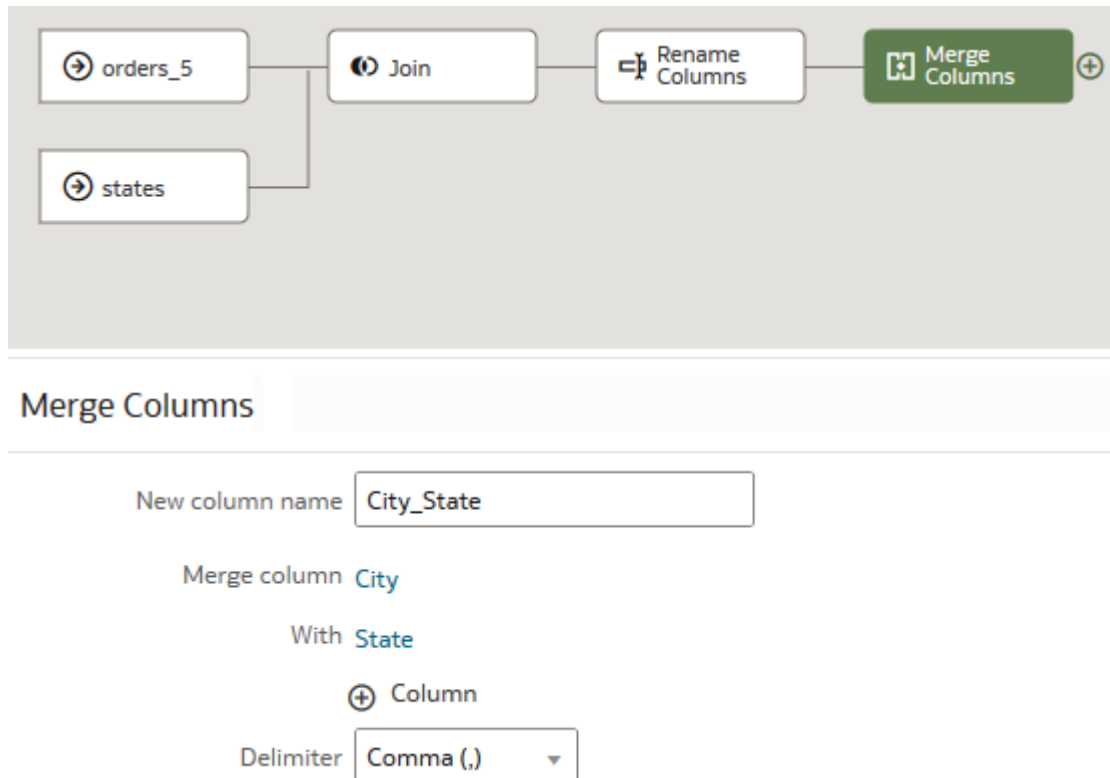
8. Click **Add a step** on the Merge Columns node, and then click **Save Data**. In Name, enter Sales Orders. Click **Save**, and then select **Save As**. In Save Data Flow As, enter Sales_Orders, and then click **OK**. Click **Save**.
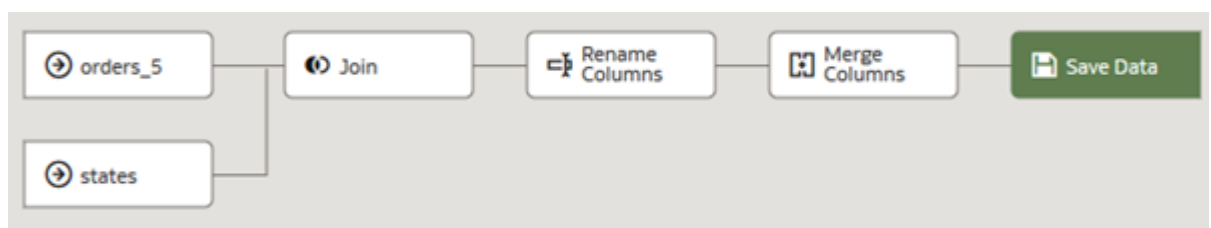
9. Click **Run Data Flow** to make sure that the data flow is valid.

Run the Data Flow with a Different Dataset

You can reuse a data flow with datasets that contain the same data as the original dataset. In this tutorial, the original dataset is orders_5.

1. In the Sales_Orders data flow, click the **orders_5 node**. In order_5 node details, click **When Run Prompt to select Dataset**. In Name, use the default value. In Prompt, enter Select the data source.

Add Data - orders_5

Dataset	orders_5	Select...

Description Uploaded from orders_5.xlsx.

When Run ☑ Prompt to select Dataset

Parameter

Name Target - orders_5

Prompt Select the data source

Columns All (20) Selections (20)

☑ Name

☑ Order Line ID

☑ Order ID

☑ Order Priority

☑ Customer ID

☑ Customer Name

☑ Customer Segment

☑ City

2. Click **Save**. Click **Run Data Flow**. In Dataflow Prompt, click the **orders_5** hyperlink.

Dataflow Prompt

This data flow contains parameters for which you can either use default values, or define alternate values below.

Sources

Select the data source

Target - orders_5   orders_5

Targets
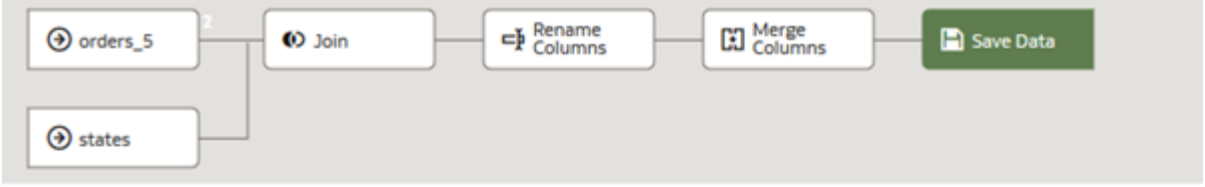
Cancel   OK

3. In Add Data, click **orders_6**, and then click **Add**. In Dataflow Prompt, click **OK**. The data flow executes with the selected dataset.

4. After the message, *Data Flow "Sales Orders" complete* appears, click **Go back** ⬅ to return to the Home page.

5. In the Home page, click **Data** in the search bar, enter Sales Orders, and the click **Search**.

6. In the *Sales Orders* dataset, click the **Actions menu** ⋮, and then select **Inspect**.

7. In the dataset page, click **Data Elements** to review the output from the data flow. Click **Close**.

Create a New Target from the Data Flow

In this section, you specify a prompt to create a new dataset when the data flow runs.

1. On the Home page, select the **Sales_Orders_DF**, click the **Actions menu** ⋮, and then click **Open**.
2. In the Sales Orders data flow, click the **Save Data** node. In Save Data node details, click **When Run Prompt to specify Dataset**. In Name, use the default value. In Prompt, enter the phrase Enter target name.

3. Click **Save**. Click **Run Data Flow**.
4. In Enter Target Name, enter My Sales, and then click **OK**.

Dataflow Prompt

This data flow contains parameters for which you can either use default values, or define alternate values below.

Sources

Select the data source

Target - orders_5   orders_5

Targets

Enter target name

Target - Sales Orders | My Sales |

Cancel   **OK**

5. After the message, *Data Flow "My Sales" complete* appears, click back ⬅ to return to the Home page.
6. In the Home page, click **Data** in the search bar, enter My Sales, and the click **Search**.
7. In the *My Sales* dataset, click the **Actions menu** ⋮, and then select **Inspect**.
8. In the dataset page, click **Data Elements** to review the output from the data flow.