Implement Incremental Processing in a Data Flow

Before you Begin

This lab shows you how to implement incremental processing in a data flow with a dataset created from a connection.

Background

You can use incremental processing in your data flow to add the latest data available from the connected data source to your dataset. When your data flow runs on a schedule, incremental processing enables updating the dataset between scheduled runs. In this tutorial, you learn how to specify a new data indicator column in the dataset to enable incremental processing and how to set parameters in the data flow to update the dataset.

Incremental processing is only available with datasets created from a connection.

What Do You Need?

- Access to Oracle Analytics
- Ability to connect a relational data source such as Oracle Autonomous Data Warehouse or Oracle Database
- Access to the Oracle sample SH schema to perform the steps in this lab, see Installing Sample Schemas

Create a Connection

This tutorial uses an Oracle Database connection to an instance with the SH schema. In this section, use these steps to create a connection to the data source.

If you already have a connection, you can skip to the next section.

1. Sign in to Oracle Analytics.
2. On the Home page, click **Create**, and then click **Connection**.
3. In Create Connection - Select Connection Type, click your database connection type.

   This example uses an Oracle Database connection type. Your connection variables depend on the selected database connection type.
4. In Create Connection when using an Oracle Database, enter the following values, and then click **Save**:
   - **Connection Name**: for example, MyOracleDB
   - **Host**
   - **Port**
   - **User Name**

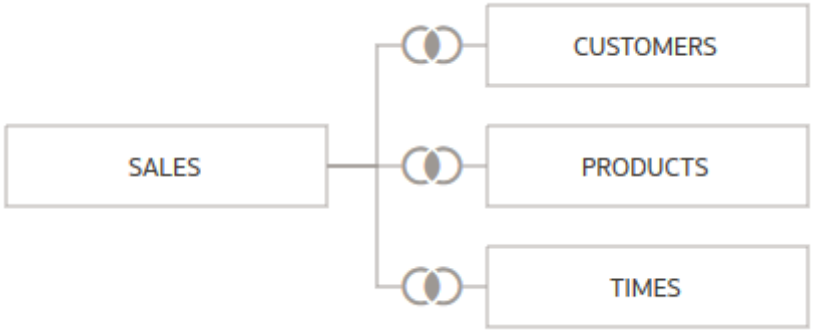
- **Password**
- **Service Name**

Create a Dataset

In this section, you create a dataset from the connection. In the next section, you use the dataset in a data flow.

1. On the Home page, click **Create**, select **Dataset**, and then click the database connection containing the SH schema.
2. In the Connections page, expand the **SH** schema.
3. Hold down the **Ctrl** key and click the **CUSTOMERS**, **PRODUCTS**, **SALES**, and **TIMES** tables, drag and then release the tables in the Join Diagram.

   Oracle Analytics automatically creates the joins using the relationships defined in the schema.



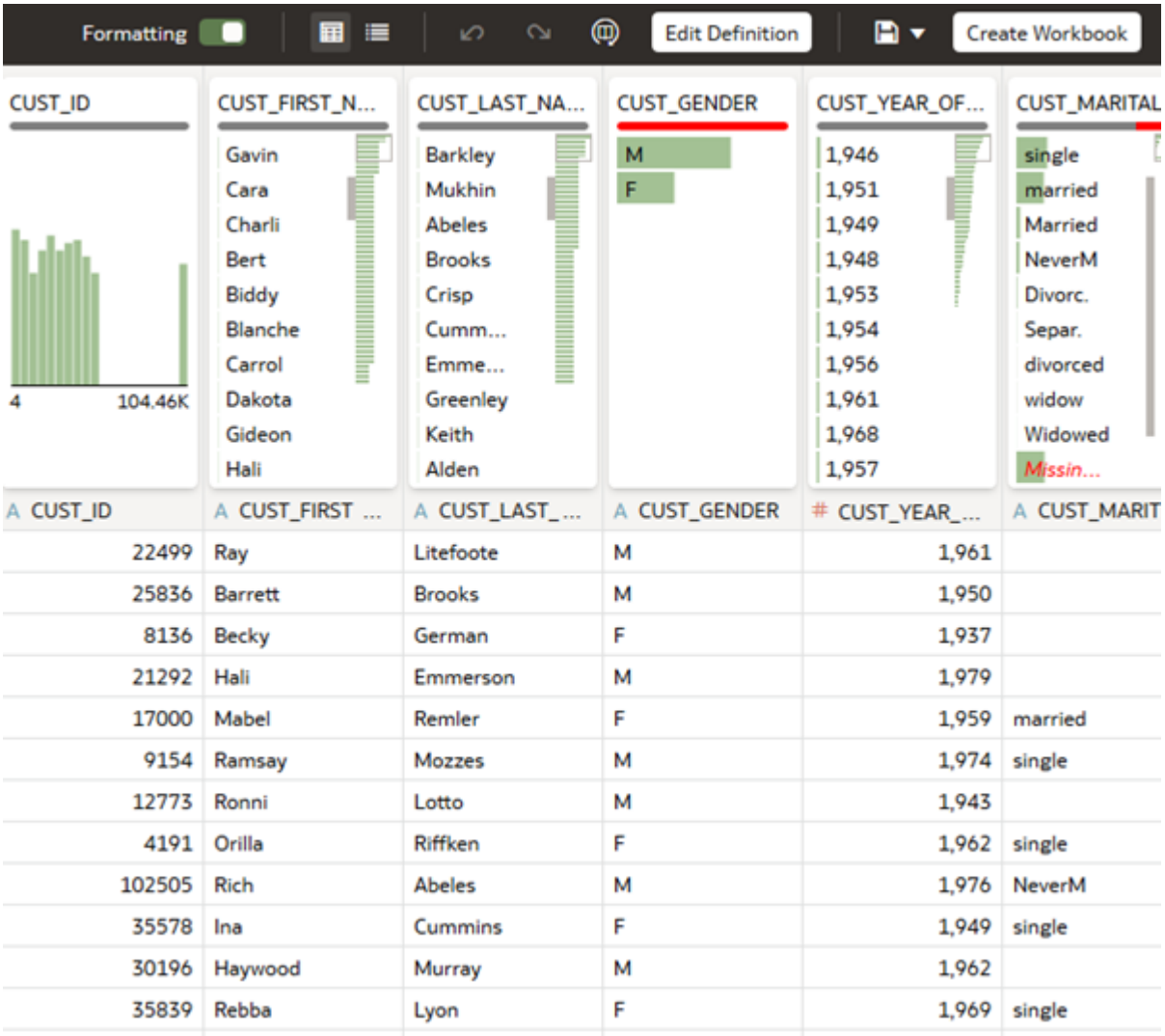
4. Click **Save** . In Save Dataset As, enter Customer Sales in **Name**, and then click **OK**.

Edit Table Definitions

In this section, you remove columns that aren't needed from the tables in the dataset.

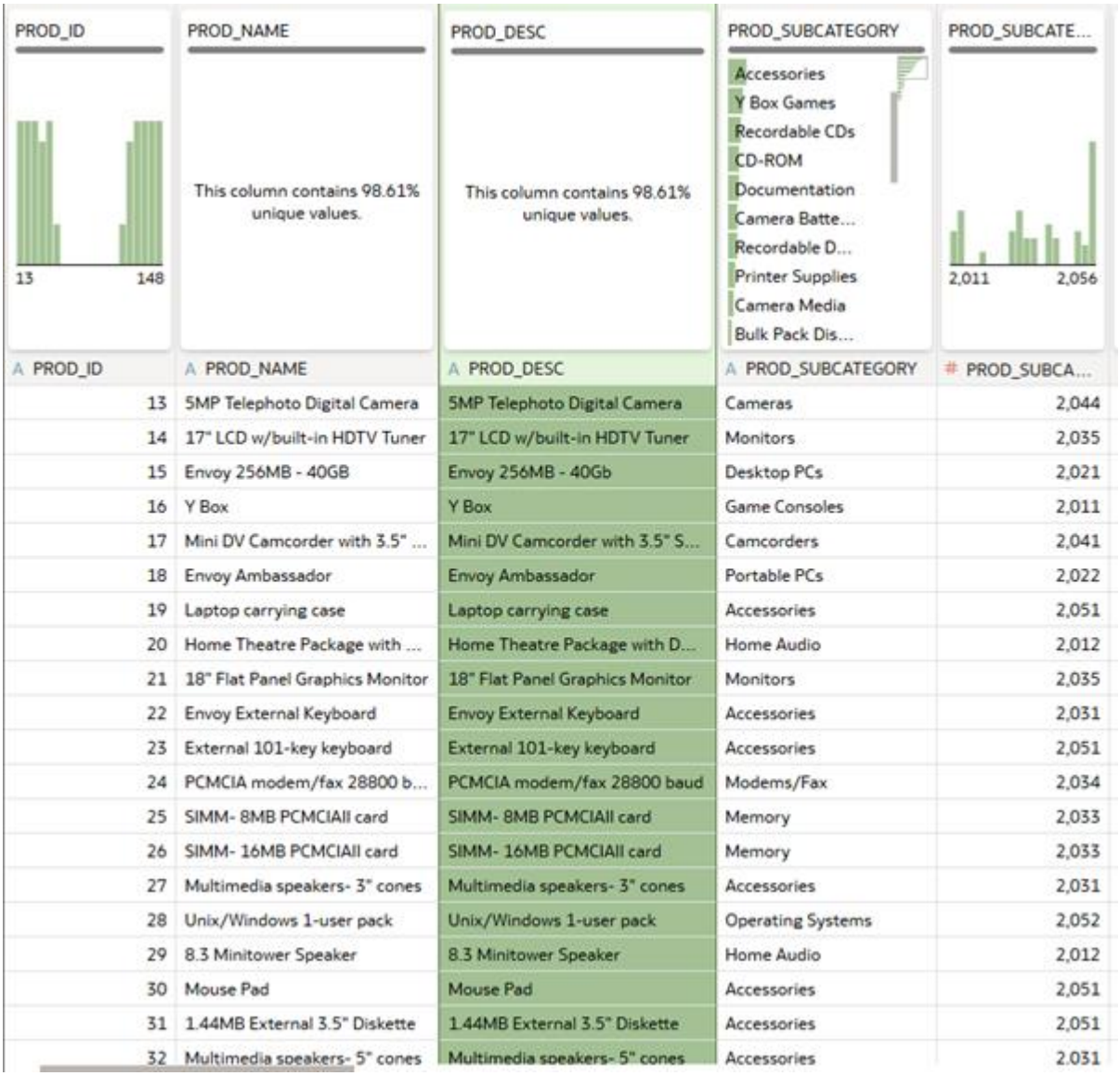
1. Click the **CUSTOMERS** table tab. In the CUSTOMERS table use the horizontal scroll bar to view the columns.

   The CUSTOMERS table contains 23 data elements. You don't need all of the columns in your dataset.



2. Click **Edit Definition**.
3. In Edit Definition, click **Remove All**.
4. Hold down the **Ctrl** key and select the following:
    - **CUST_ID**
    - **CUST_CITY**
    - **CUST_FIRST_NAME**
    - **CUST_LAST_NAME**
    - **CUST_GENDER**
    - **CUST_POSTAL_CODE**
    - **CUST_STATE_PROVINCE**
    - **CUST_STREET_ADDRESS**
5. Click **Add Selected**, and then click **OK**. Click **Save** .
6. Click the **PRODUCTS** table tab. Use the horizontal scroll bar to view the columns.

    The PRODUCTS table contains 22 data elements. You don't need all of the columns in your dataset.



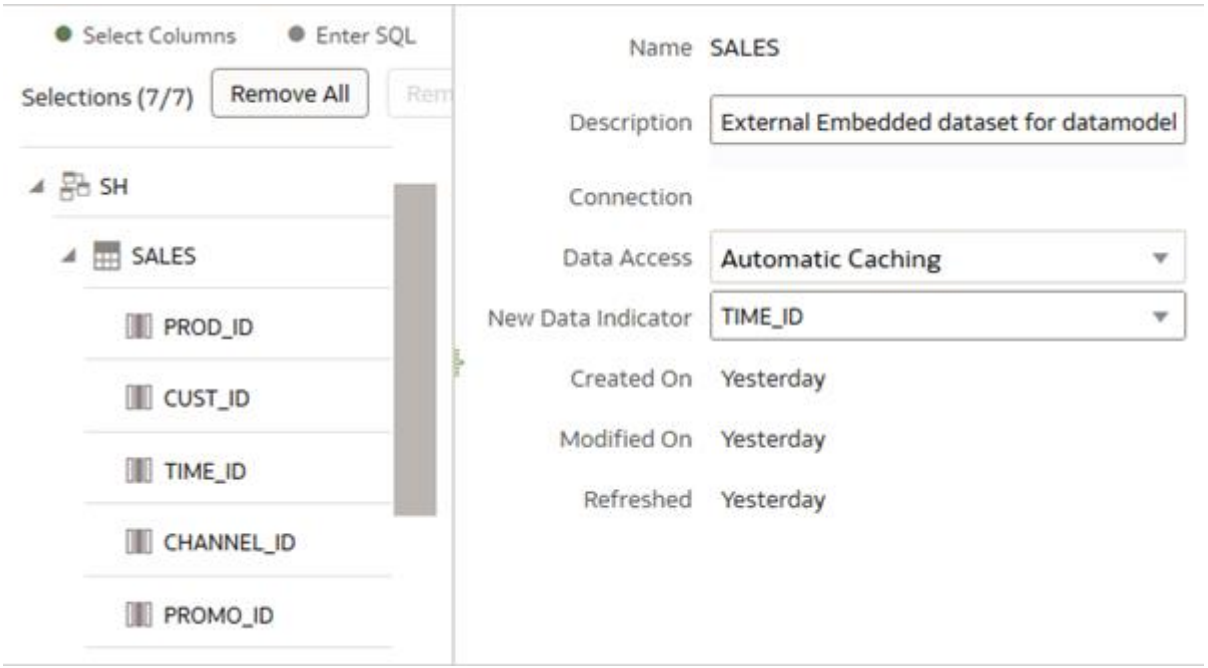
A PROD_ID	A PROD_NAME	A PROD_DESC	A PROD_SUBCATEGORY	# PROD_SUBCA...
13	5MP Telephoto Digital Camera	5MP Telephoto Digital Camera	Cameras	2,044
14	17" LCD w/built-in HDTV Tuner	17" LCD w/built-in HDTV Tuner	Monitors	2,035
15	Envoy 256MB - 40GB	Envoy 256MB - 40Gb	Desktop PCs	2,021
16	Y Box	Y Box	Game Consoles	2,011
17	Mini DV Camcorder with 3.5" ...	Mini DV Camcorder with 3.5" S...	Camcorders	2,041
18	Envoy Ambassador	Envoy Ambassador	Portable PCs	2,022
19	Laptop carrying case	Laptop carrying case	Accessories	2,051
20	Home Theatre Package with ...	Home Theatre Package with D...	Home Audio	2,012
21	18" Flat Panel Graphics Monitor	18" Flat Panel Graphics Monitor	Monitors	2,035
22	Envoy External Keyboard	Envoy External Keyboard	Accessories	2,031
23	External 101-key keyboard	External 101-key keyboard	Accessories	2,051
24	PCMCIA modem/fax 28800 b...	PCMCIA modem/fax 28800 baud	Modems/Fax	2,034
25	SIMM- 8MB PCMCIAII card	SIMM- 8MB PCMCIAII card	Memory	2,033
26	SIMM- 16MB PCMCIAII card	SIMM- 16MB PCMCIAII card	Memory	2,033
27	Multimedia speakers- 3" cones	Multimedia speakers- 3" cones	Accessories	2,031
28	Unix/Windows 1-user pack	Unix/Windows 1-user pack	Operating Systems	2,052
29	8.3 Minitower Speaker	8.3 Minitower Speaker	Home Audio	2,012
30	Mouse Pad	Mouse Pad	Accessories	2,051
31	1.44MB External 3.5" Diskette	1.44MB External 3.5" Diskette	Accessories	2,051
32	Multimedia speakers- 5" cones	Multimedia speakers- 5" cones	Accessories	2,031

7. Click **Edit Definition**.
8. In Edit Definition, click **Remove All**. Hold down the **Ctrl** key and select the following:
    - **PROD_ID**
    - **PROD_CATEGORY**
    - **PROD_NAME**
    - **PROD_SUBCATEGORY**
9. Click **Add Selected**, and then click **OK**. Click **Save** .

Specify New Data Indicator

In this section, you set the new data indicator property to update the dataset. In this example, when a sale occurs the transaction is listed with a time ID, making it a good new data indicator.

1. Click the **SALES** table tab.
2. In SALES table, click **Edit Definition**.
3. From the New Data Indicator list, select **TIME_ID**, and then click **OK**.

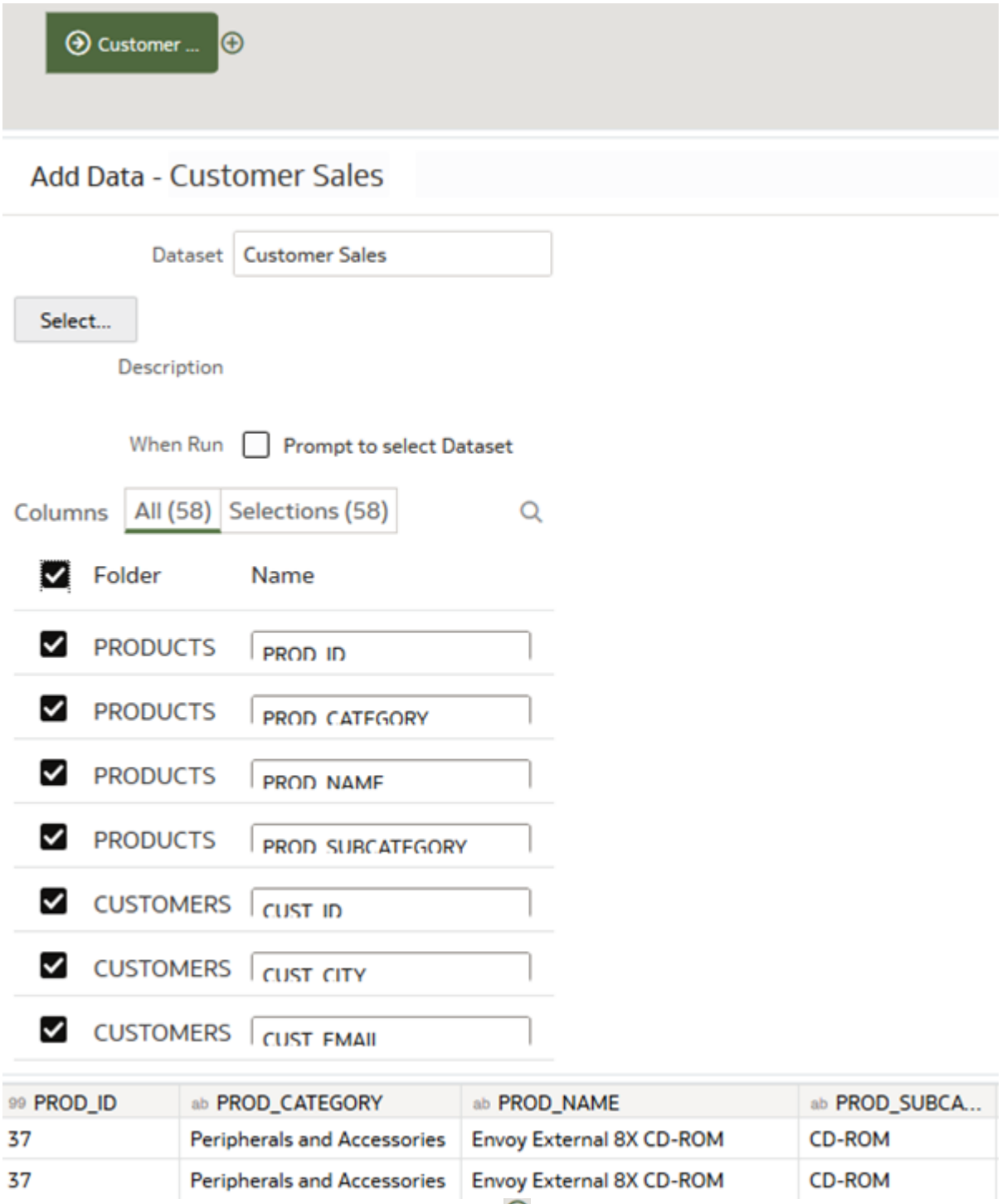


4. Click **Save**.
5. Click **Go back** .

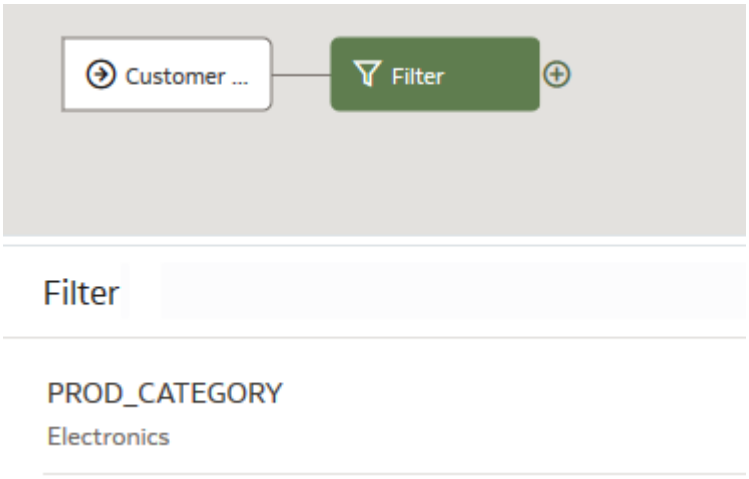
Create a Data Flow

In this section, you create a data flow with the Customer Sales dataset.

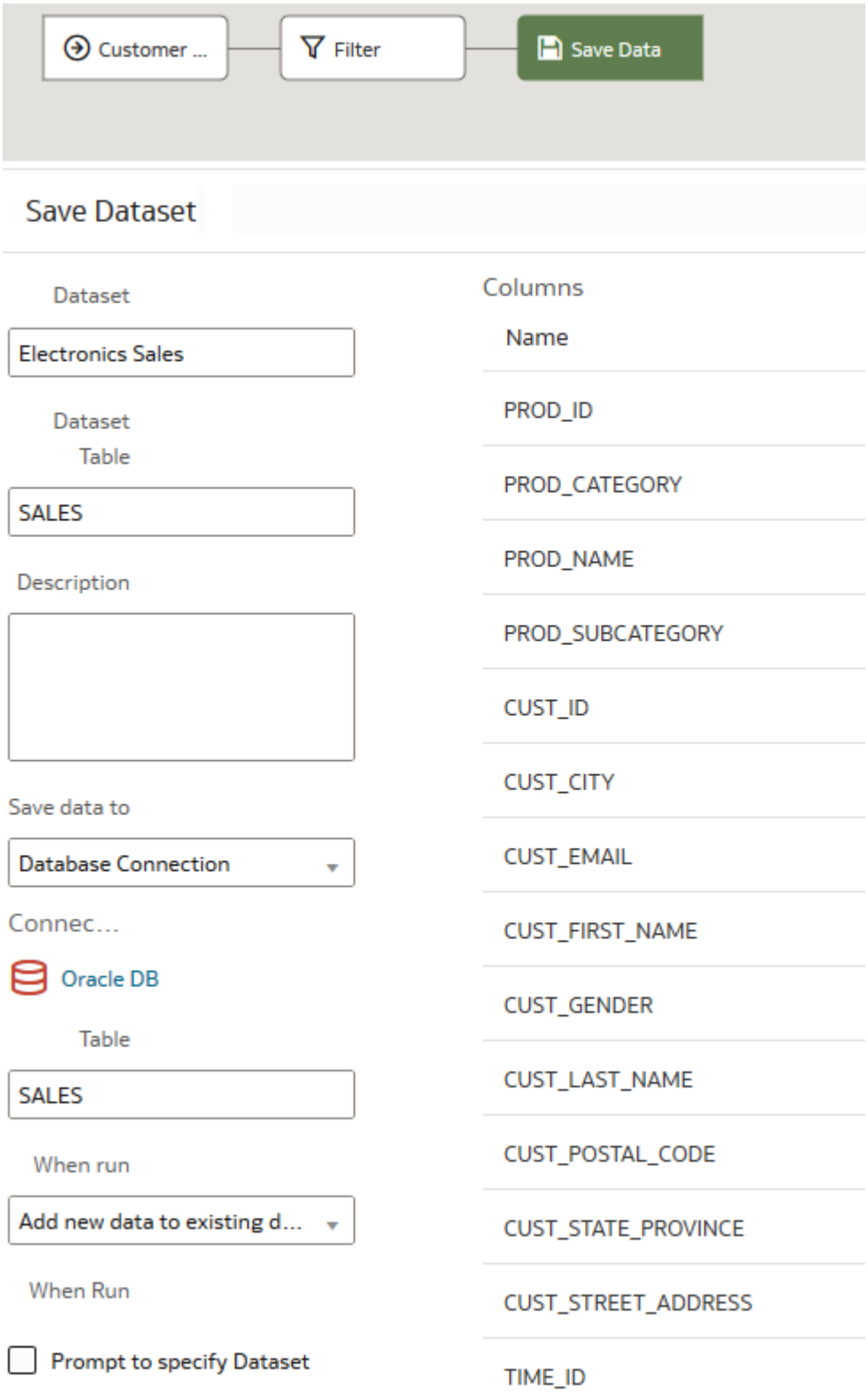
1. On the Home page, click **Create**, and then select **Data Flow**.
2. In Add Dataset, click **Customer Sales**, and then click **Add**.
3. In Add Data - Customer Sales, click **Folder**. In the Select All message, click **Yes**.



4. On the Customer Sales node, click **Add a step**, and then click **Filter**.
5. In Filter, click **Add Filter**. From the Available data list, click **PROD_CATEGORY**. From the PROD_CATEGORY list, click **Electronics**.



6. In the data flow, click **Add a step** on the Filter node. Select **Save Data**.
7. In Save Dataset, enter Electronics Sales.
8. From the Save Data to list, select **Database Connection**. Click **Database Connection**, and then click the connection containing your dataset.
9. In Table, enter SALES. From the When run list, select **Add new data to existing data.**



10. Click **Save**. In Save Data Flow As, enter Sales Revenues, and then click **OK**.
11. Click **Run Data Flow** ▶.

Schedule the Data Flow

Incremental processing runs when changes occur in the data source between data flow runs. This section shows you how to schedule a data flow.

1. On the Home page, click **Data**, enter enter Sales Revenues in the Search bar, and then press **Enter**.
2. Select your data flow, click **Actions menu** ⋮, and then select **New Schedule**.
3. In Schedule, enter a **Name** or keep the default name.
4. Click the calendar in **Start**, and then select a start date. Click the calendar in **End** to specify an ending date or leave **End** empty.
5. In **Time**, enter the hour and minutes of the start time. From the **Repeat** list, select a frequency for running the data flow, and then click **OK**.