

# 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**

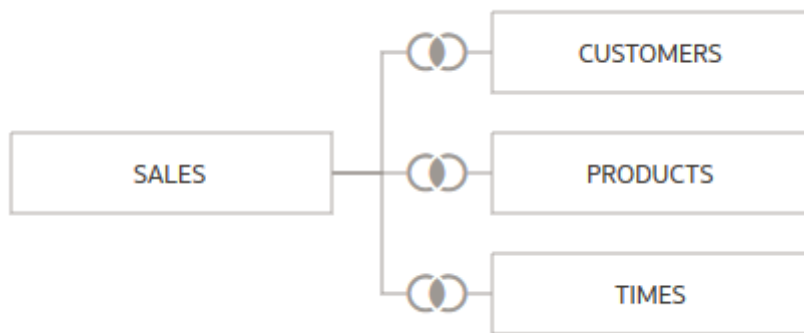
- **Password**
- **Service Name**

## 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.

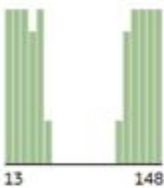
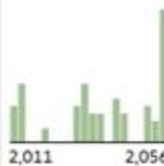
1. Click the **CUSTOMERS** table tab. In the CUSTOMERS table use the horizontal scroll bar to view the columns.


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

Formatting <input type="checkbox"/> Edit Definition  Create Workbook					
<b>CUST_ID</b>	<b>CUST_FIRST_N...</b>	<b>CUST_LAST_NA...</b>	<b>CUST_GENDER</b>	<b>CUST_YEAR_OF...</b>	<b>CUST_MARITAL</b>
	Gavin Cara Charli Bert Biddy Blanche Carrol Dakota Gideon Hali	Barkley Mukhin Abeles Brooks Crisp Cumm... Emme... Greenley Keith Alden	M F	1,946 1,951 1,949 1,948 1,953 1,954 1,956 1,961 1,968 1,957	single married Married NeverM Divorc. Separ. divorced widow Widowed Missin...
A CUST_ID	A CUST_FIRST ...	A CUST_LAST_ ...	A CUST_GENDER	# CUST_YEAR_...	A CUST_MARIT
22499	Ray	Litefoote	M	1,961	
25836	Barrett	Brooks	M	1,950	
8136	Becky	German	F	1,937	
21292	Hali	Emmerson	M	1,979	
17000	Mabel	Remler	F	1,959	married
9154	Ramsay	Mozzes	M	1,974	single
12773	Ronni	Lotto	M	1,943	
4191	Orilla	Riffken	F	1,962	single
102505	Rich	Abeles	M	1,976	NeverM
35578	Ina	Cummins	F	1,949	single
30196	Haywood	Murray	M	1,962	
35839	Rebba	Lyon	F	1,969	single

- Click **Edit Definition**.
- In Edit Definition, click **Remove All**.
- 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**
- Click **Add Selected**, and then click **OK**. Click **Save** .
- 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.

PROD_ID	PROD_NAME	PROD_DESC	PROD_SUBCATEGORY	PROD_SUBCATE...
	This column contains 98.61% unique values.	This column contains 98.61% unique values.	<ul style="list-style-type: none"> <li>Accessories</li> <li>Y Box Games</li> <li>Recordable CDs</li> <li>CD-ROM</li> <li>Documentation</li> <li>Camera Batte...</li> <li>Recordable D...</li> <li>Printer Supplies</li> <li>Camera Media</li> <li>Bulk Pack Dis...</li> </ul>	
A PROD_ID	A PROD_NAME	A PROD_DESC	A PROD_SUBCATEGORY	# PROD_SUBCA...
13	SMP Telephoto Digital Camera	SMP 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

- Click **Edit Definition**.
- In Edit Definition, click **Remove All**. Hold down the **Ctrl** key and select the following:
  - PROD\_ID**
  - PROD\_CATEGORY**
  - PROD\_NAME**
  - PROD\_SUBCATEGORY**
- 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.

- Click the **SALES** table tab.
- In **SALES** table, click **Edit Definition**.
- From the New Data Indicator list, select **TIME\_ID**, and then click **OK**.

The screenshot shows a configuration window for a dataset named 'SALES'. On the left, under 'Select Columns', the 'SALES' dataset is expanded, showing selected columns: PROD\_ID, CUST\_ID, TIME\_ID, CHANNEL\_ID, and PROMO\_ID. On the right, the configuration details are as follows:

Name SALES	
Description	External Embedded dataset for datamodel
Connection	
Data Access	Automatic Caching
New Data Indicator	TIME_ID
Created On	Yesterday
Modified On	Yesterday
Refreshed	Yesterday

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**.

Customer ...

+

Add Data - Customer Sales

Dataset

Customer Sales

Select...

Description

When Run

☐

Prompt to select Dataset

Columns


All (58)

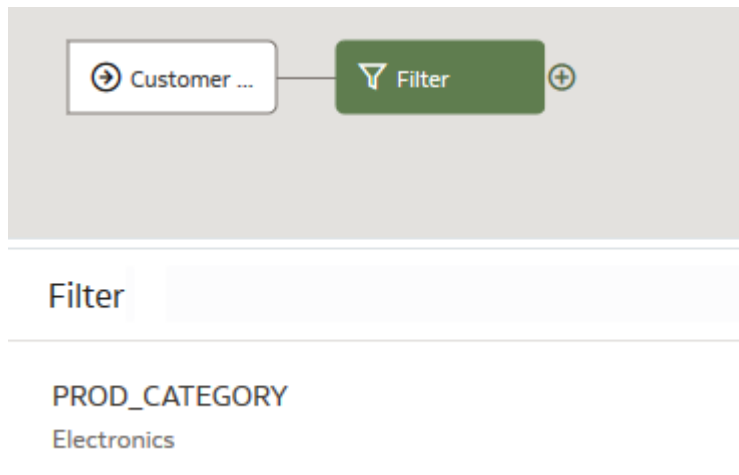
Selections (58)


Q

<input checked="" type="checkbox"/>	Folder	Name
<input checked="" type="checkbox"/>	PRODUCTS	PROD ID
<input checked="" type="checkbox"/>	PRODUCTS	PROD CATEGORY
<input checked="" type="checkbox"/>	PRODUCTS	PROD NAME
<input checked="" type="checkbox"/>	PRODUCTS	PROD SUBCATEGORY
<input checked="" type="checkbox"/>	CUSTOMERS	CUST ID
<input checked="" type="checkbox"/>	CUSTOMERS	CUST CITY
<input checked="" type="checkbox"/>	CUSTOMERS	CUST EMAIL

99	PROD_ID	ab	PROD_CATEGORY	ab	PROD_NAME	ab	PROD_SUBCA...
37			Peripherals and Accessories		Envoy External 8X CD-ROM		CD-ROM
37			Peripherals and Accessories		Envoy External 8X CD-ROM		CD-ROM

- On the Customer Sales node, click **Add a step** , and then click **Filter**.
- 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**.

Customer ...

Filter

Save Data

Save Dataset

Dataset

Electronics Sales

Dataset

Table

SALES

Description

Save data to

Database Connection

Connec...

Oracle DB

Table

SALES

When run

Add new data to existing d...

When Run

☐ Prompt to specify Dataset

Columns

Name

PROD\_ID

PROD\_CATEGORY

PROD\_NAME

PROD\_SUBCATEGORY

CUST\_ID

CUST\_CITY

CUST\_EMAIL

CUST\_FIRST\_NAME

CUST\_GENDER


CUST\_LAST\_NAME

CUST\_POSTAL\_CODE

CUST\_STATE\_PROVINCE

CUST\_STREET\_ADDRESS


TIME\_ID

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 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**.

### Schedule

Object	Sales Revenue		
Activity	Run Data Flow		
Name	<input type="text" value="Sales Revenue_1"/>		
Start	<input type="text" value="09/16/21"/> 	Time	<input type="text" value="03:07 PM"/> 
Repeat	<input type="text" value="Weekly"/> ▼	End	<input type="text" value="09/17/21"/> 
<div><input type="checkbox"/> Monday</div> <div><input type="checkbox"/> Tuesday</div> <div><input type="checkbox"/> Wednesday</div> <div><input checked="" type="checkbox"/> Thursday</div> <div><input type="checkbox"/> Friday</div> <div><input type="checkbox"/> Saturday</div> <div><input type="checkbox"/> Sunday</div>			
		<input type="button" value="Cancel"/>	<input type="button" value="OK"/>