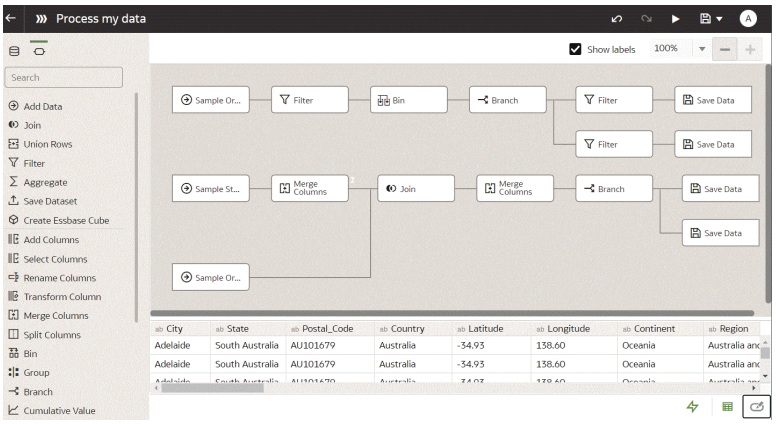

To build a data flow, you add steps. Each step performs a specific function, for example, add data, join tables, merge columns, transform data, save your data. Use the data flow editor to add and configure your steps. Each step is validated when you add or change it. When you've configured your data flow, you execute it to produce or update a dataset.

When you add your own columns or transform data, you can use a wide range of SQL operators (for example, BETWEEN, LIKE, IN), conditional expressions (for example, CASE), and functions (for example, Avg, Median, Percentile).

Database Support for Data Flows

With data flows you can curate data from datasets, subject areas, or database connections.

You can execute data flows individually or in a sequence. You can include multiple data sources in a data flow and specify how to join them.

You can save the output data from a data flow in either a dataset or in one of the supported database types. If you save data to a database, you can transform the data source by overwriting it with data from the data flow. The data source and data flow tables must be in the same database and have the same name. Before you start, create a connection to one of the supported database types.

What Steps Can I Use to Curate My Data?

45

You can curate your data by building data flow using steps to perform common transformations.

Add Columns

Add custom columns to your target dataset. For example, you might calculate the value of your stock by multiplying the number of units in a `UNITS` column by the sale price in a `RETAIL_PRICE` column (that is, `UNITS * RETAIL_PRICE`).

Add Data

Add data sources to your data flow. For example, if you're merging two datasets, you add both datasets to your data flow.

45

Aggregate

Create group totals by applying aggregate functions. For example, count, sum, or average.

Analyze Sentiment

Detect sentiment for a given text column. For example, you might analyze customer feedback to determine whether it's positive or negative. Sentiment analysis evaluates text based on words and phrases that indicate a positive, neutral, or negative emotion. Based on the outcome of the analysis, a new column contains Positive, Neutral, or Negative.

Bin

Assign data values into categories, such as high, low, or medium. For example, you might categorize values for `RISK` into three bins for low, medium, and high.

Branch

Creates multiple outputs from a data flow. For example, if you have sales transactions data based on country, you might save data for United States in the first branch and data for Canada in the second branch.

Create Essbase Cube

Create an Essbase cube from a spreadsheet or database.

Cumulative Value

Calculate cumulative totals such as moving aggregate or running aggregate.

Database Analytics

Perform advanced analysis and data mining analysis. For example, you can detect anomalies, cluster data, sample data, and perform affinity analysis. To display this step type in the data flow editor, you must connect to an Oracle database or Oracle Autonomous Data Warehouse (the analytics are computed in the database, not in Oracle Analytics). See Database Analytics Functions.

45

Filter

Select only the data that you're interested in. For example, you might create a filter to limit sales revenue data to the years 2020 through 2022.

Graph Analytics

Perform geo-spatial analysis, such as calculating the distance or the number of hops between two vertices. To display this step type in the data flow editor, you must connect to an Oracle database or Oracle Autonomous Data Warehouse (the analytics are computed in the database, not in Oracle Analytics). See Graph Analytics Functions.

Group

Categorize non-numeric data into groups that you define. For example, you might put orders for lines of business `Communication` and `Digital` into a group named `Technology`, and orders for `Games` and `Stream` into a group named `Entertainment`.

45

Join

Combine data from multiple data sources using a database join based on a common column. For example, you might join an `Orders` dataset to a `Customer_orders` dataset using a customer ID field.

Merge

Combine multiple columns into a single column. For example, you might merge the street address, street name, state, and ZIP code columns into one column.

Rename Columns

Change the name of a column to more meaningful. For example, you might change CELL to Contact Cell Number.

Save Data

Specify where to save the data generated by the data flow. You can save the data in a dataset in Oracle Analytics or in a database. You can also specify runtime parameters, or change the default dataset name.

45

Select Columns

Specify which columns to include or exclude in your data flow (the default is to include all data columns).

Split Columns

Extract data from within columns. For example, if a column contains `001011Black`, you might split this data into two separate columns, `001011` and `Black`.

Time Series Forecast

Calculate forecasted values based on historical data. A forecast takes a time column and a target column from a given dataset and calculates forecasted values for the target column.

Train <model type>

Train machine learning models using algorithms for numeric prediction, multi-classification, binary-classification and clustering. See Data Flow Steps for Training Machine Learning Models.

When you've trained a machine learning model, apply it to your data using the **Apply Model** step.

Transform Column

Change the format, structure, or values of data. For example, you might convert text to uppercase, trim leading and trailing spaces from data, or calculate a percentage increase in value.

45

Union Rows

Merge the rows of two data sources (known as a UNION command in SQL terminology).