JOB AID

Introduction to Pipeline Designer

Overview

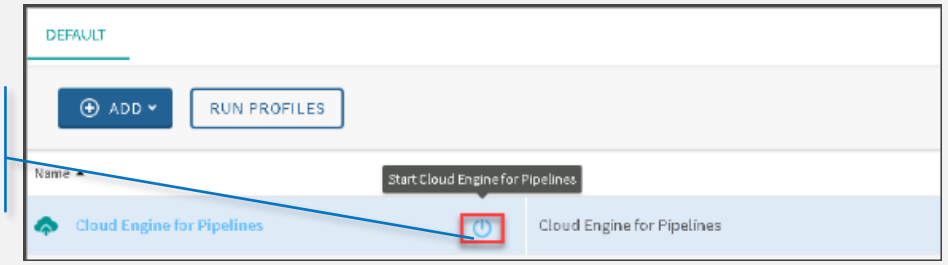
In this module, you learn how to create connections, datasets, and how to develop pipelines with Pipeline Designer to process data in Talend Cloud.

Key steps

1 Starting a Cloud Engine for Pipelines

Once connected to Talend Cloud Portal, using your cloud account (*<login name>@<domain name>*), you can start the Cloud Engine for Pipelines to run pipelines.

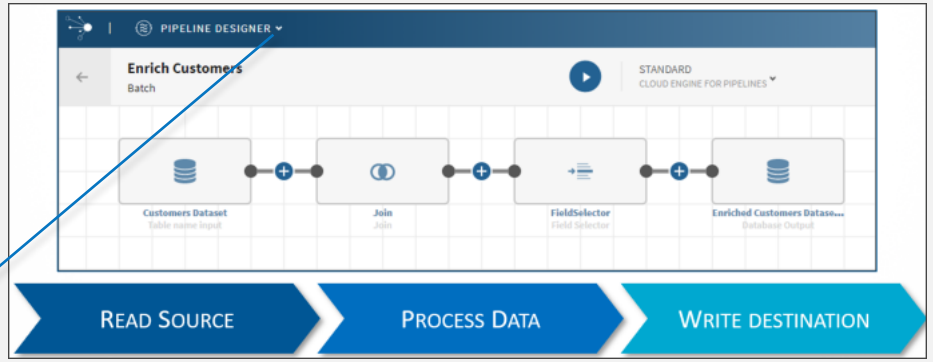
In Talend Management Console, click the **Engines** tab, then click the **Start Cloud Engine for Pipelines** button.



2 Starting the Pipeline Designer application

Pipeline Designer is a modern, lightweight integration tool. It provides a graphical web-based user interface in the cloud to create end-to-end pipelines to read, process, and write data.

Start Pipeline Designer by using the top menu of Talend Cloud Portal and selecting Pipeline Designer.

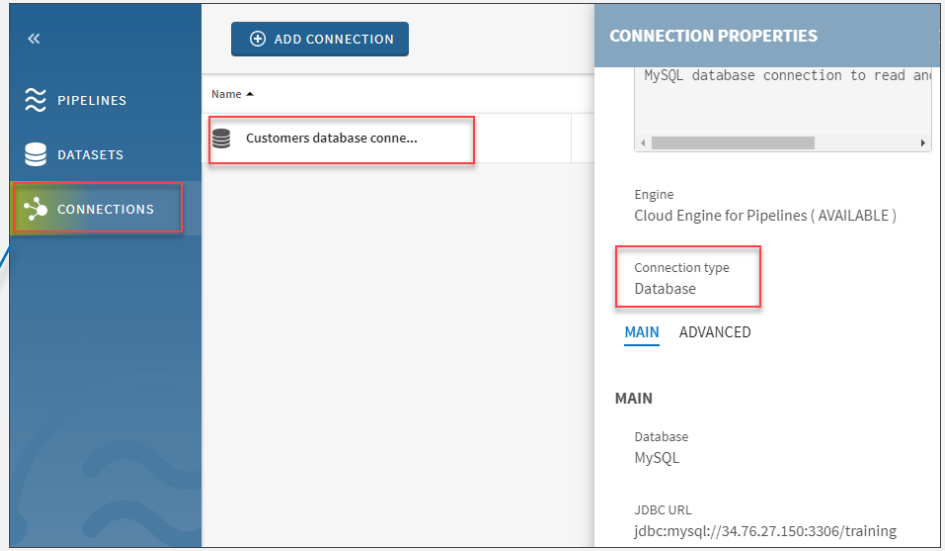


JOB AID

3 Creating connections and datasets

You can create different types of connections as a database (for example, MySQL, SQL Server, or Snowflake), Amazon S3, messages brokers like Kafka or Amazon Kinesis.

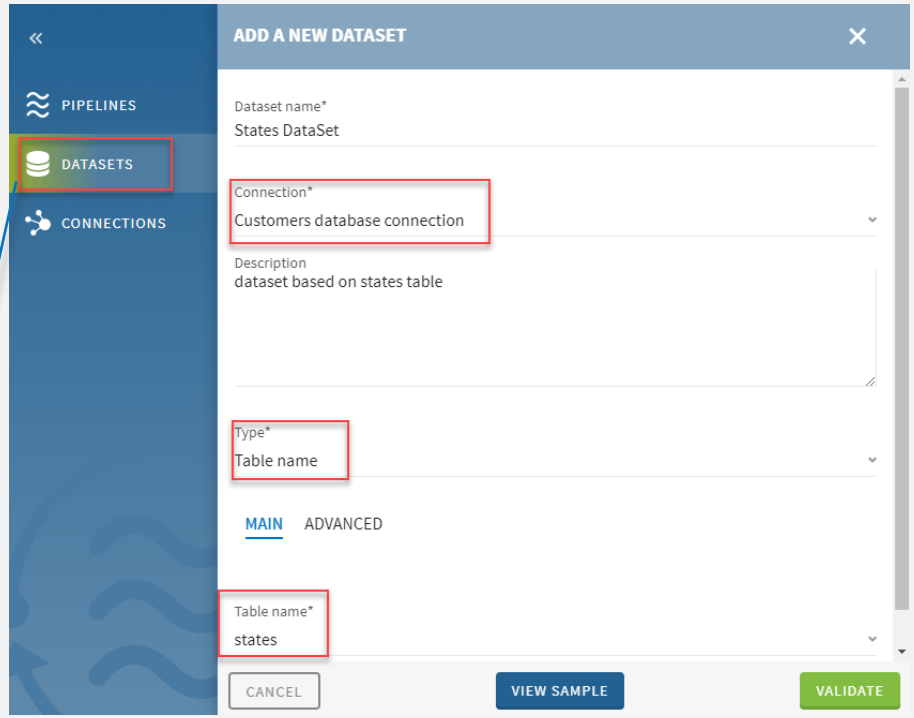
In this example, a database MySQL connection named **Customers database connection** is created in Pipeline Designer.



You can then define the dataset by using a defined connection. Datasets are collections of data.

They can be database tables, file names, topics (for example, Kafka), queues (for example, JMS), or file paths (for example, HDFS).

In this example, you can create a dataset with the connection **Customers database connection** and a table named **states**.



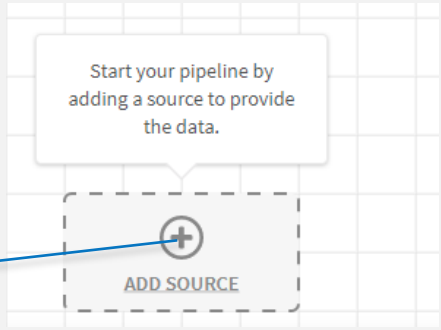
JOB AID

4 Developing a pipeline

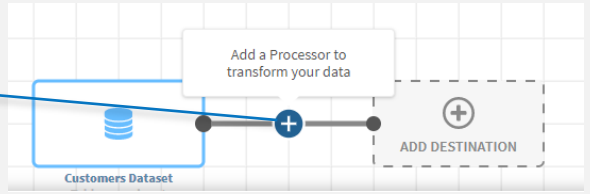
A pipeline reads input data from a dataset, processes the data using a processor, then writes the data to a dataset.

After you created a pipeline, by clicking the **PIPELINES** tab, developing a pipeline consists of:

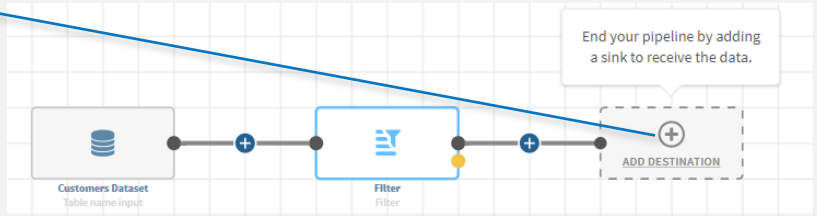
Adding a source and selecting a dataset to read the data



Adding a processor (for example, Filter, Join, or Aggregate) to process the data



Adding a destination by selecting a dataset to write data



Here is a sample of a pipeline with one dataset source, a Filter processor, and two dataset destinations.