

# Pentaho Data Import Integration

This article explains how to use Treasure Data with [Pentaho Data Integration](#) using [presto-jdbc driver](#). By combining Treasure Data with Pentaho, users can scale their existing Pentaho Data Integration environment to handle huge volumes of data.

Continue to the following topics:

- [Prerequisites](#)
- [Download Pentaho Data Integration \(Kettle\)](#)
- [Download JDBC Driver](#)
- [Copy JDBC Driver Jar to Pentaho Data Integration](#)
- [Create Treasure Data Database Connection](#)
- [Use Treasure Data Database as Table Input](#)
- [Tip: How can I use Pentaho through a proxy?](#)

## Prerequisites

- Basic knowledge of Treasure Data.



**Don't have time** to set up Pentaho + Treasure Data? Leverage our [Setup Consultation Service](#).

## Download Pentaho Data Integration (Kettle)

You can download Pentaho Data Integration (Kettle) from the link below. Version 8.2 was tested for this article

- [Pentaho Data Integration \(Kettle\)](#)

## Download JDBC Driver

You can download the driver from the link below. The driver is still in beta; any feedback is appreciated.

To work with Pentaho, you must use **td-jdbc-VERSION-jar-with-dependencies.jar** file.

- [JDBC Driver Download](#)



This driver works only with Treasure Data. It does not work with other environments such as your local Hadoop/Hive cluster.

## Copy JDBC Driver Jar to Pentaho Data Integration

Before starting Pentaho Data Integration, please copy the Treasure Data JDBC driver to the `lib` directory specified by Pentaho Data Integration.

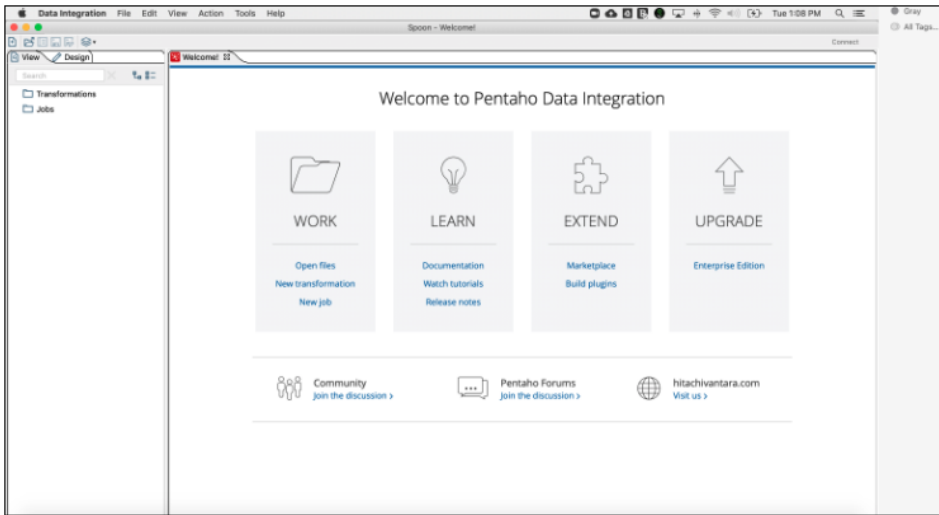
```
$ cp td-jdbc-VERSION-jar-with-dependencies.jar \  
<pentaho-data-integration-install>/data-integration/lib
```

## Create Treasure Data Database Connection

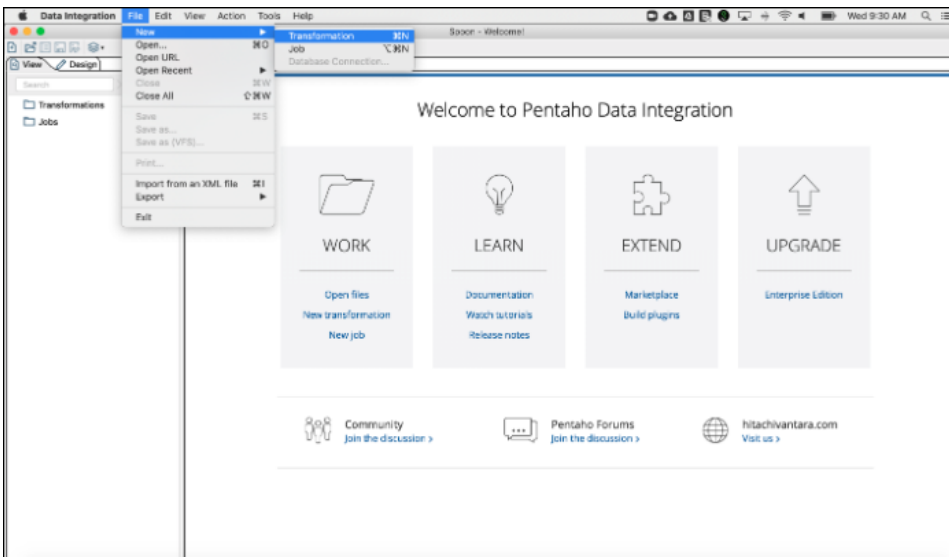
Connect Treasure Data to Pentaho Data Integration's database connection and make a new transformation. Please follow the procedure below.

### Create a New Transform

1. Open the Pentaho DI application.

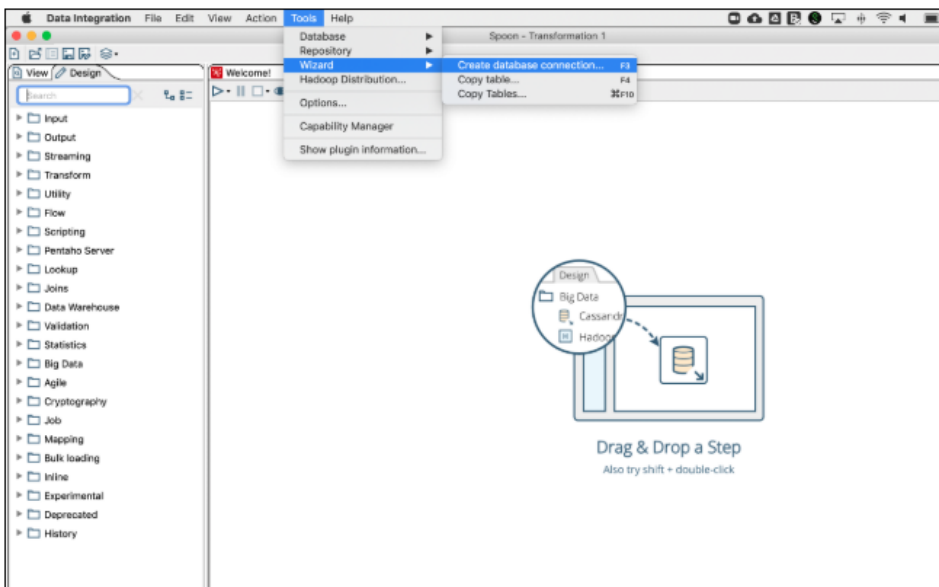


2. Select File > New > Transformation



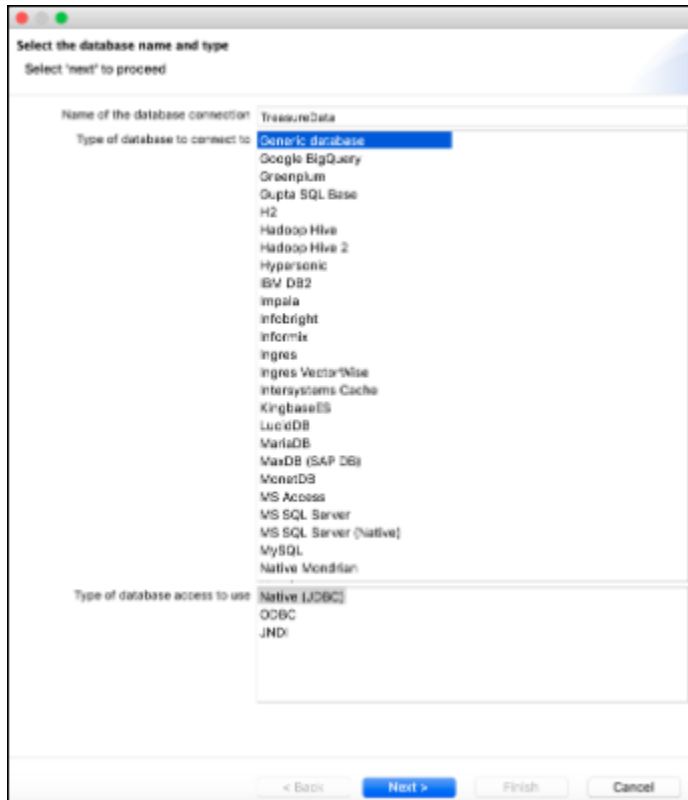
## Create New Database Connection

1. Navigate to Tools > Wizard > Create database connection.



2. Edit the details in the pop-up dialog.

- Name the database connection
- Select Generic Database for type of database
- Select Native (JDBC) for type of database access



3. Specify the connection's URL (view [API endpoints](#)) and the name of the driver class.

- Driver class is `com.treasuredata.jdbc.TreasureDataDriver`

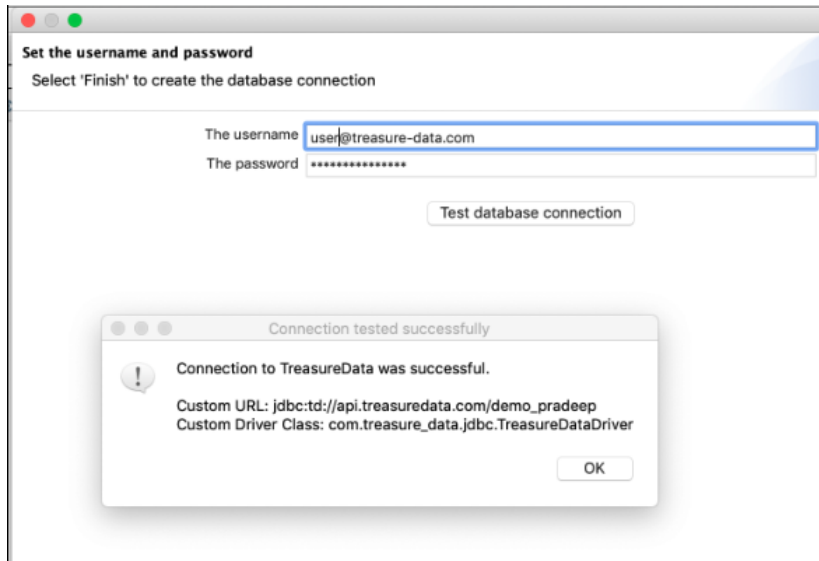


The URL can be specified in any of the following ways:

- `jdbc:td://api.treasuredata.com/<db_name>`
- use `jdbc:td://api.treasuredata.com/<db_name>;useSSL=True` if you want to enforce SSL
- You can choose the query engine [Hive](#) or [Presto](#) (default) by the parameter
- `jdbc:td://api.treasuredata.com/sample_db;useSSL=true;type=hive`
- `jdbc:td://api.treasuredata.com/sample_db;useSSL=true;type=presto`

4. Specify your username and password. Use your Treasure Data credentials for these fields (Your User Name is the email address used to register on Treasure Data).

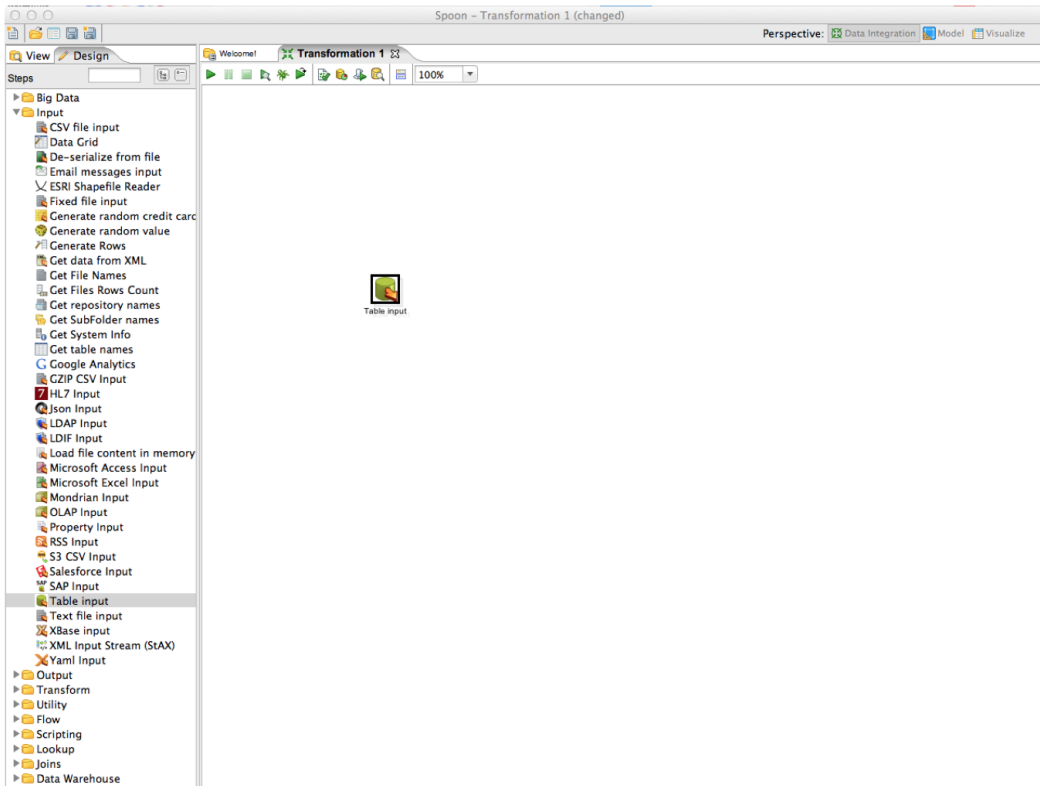
- Select Test database connection.



## Use Treasure Data Database as Table Input

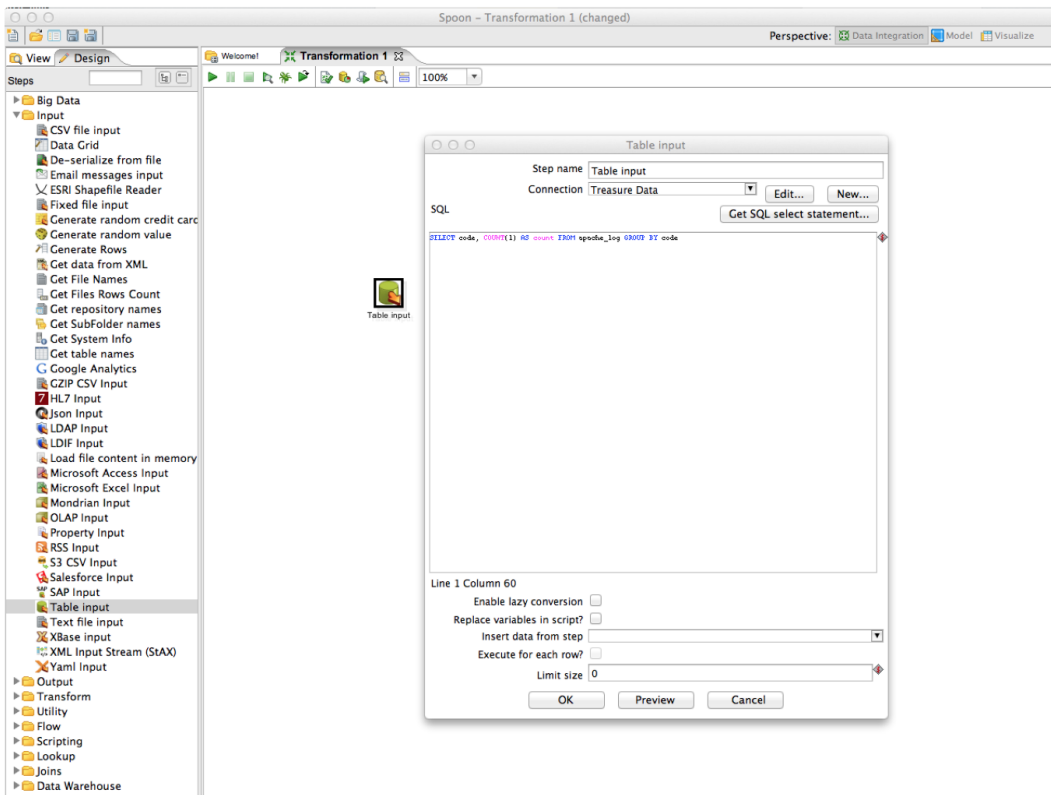
### Specify Table Input

1. Select **Table Input** from the transform's Input menu.
2. Drag and drop onto the workspace as follows:



### Edit the Table Input

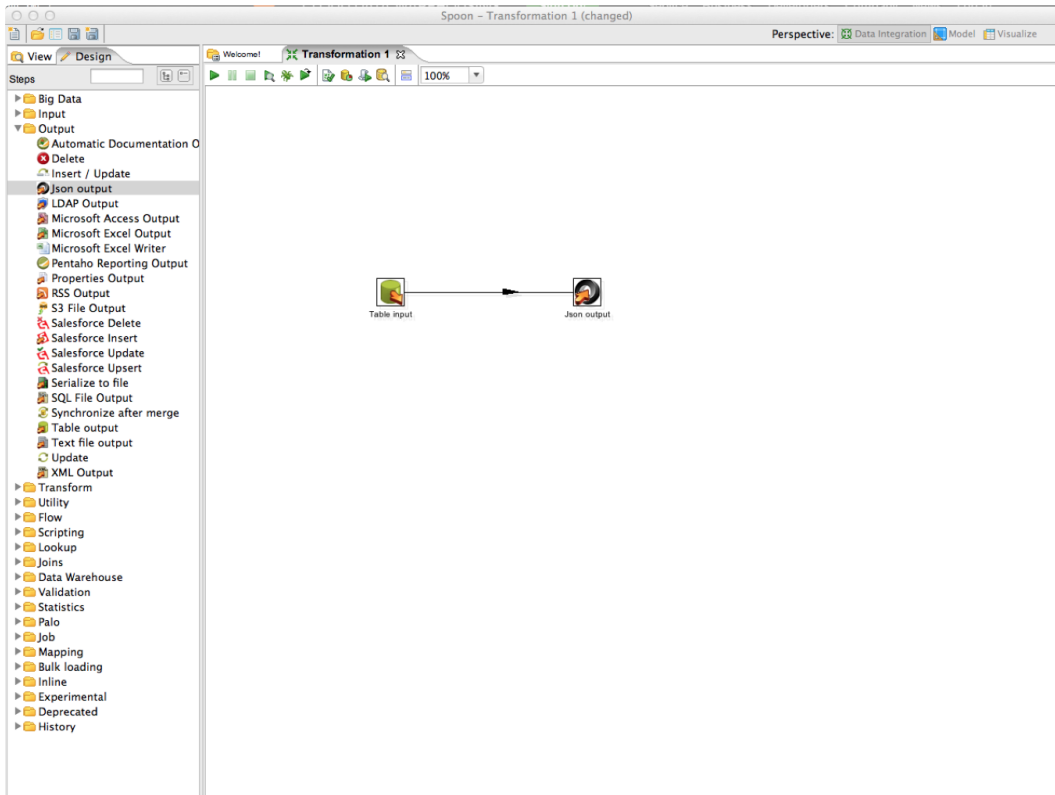
1. Right-click the Table input icon on the workspace.
2. Select **Edit** from the context menu.
3. Configure your Table input as follows:



The preceding query is reviewed in the [JasperSoft iReport with JDBC Driver](#) article.

## Confirm the Table Input

To confirm the behavior of the Table input, send your data on Treasure Data to JSON output as follows:



## Tip: How can I use Pentaho through a proxy?

Check 'start-pentaho.bat' or 'start-pentaho.sh', and add the following parameters to CATALINA\_OPTS option:

```
CATALINA_OPTS="-Dhttp.proxyHost=<proxy address> -Dhttp.proxyPort=<proxy port>"
```