

Training Guide

Get Started with Databricks and Tableau

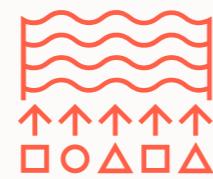
See and understand ALL your data
with Databricks and Tableau



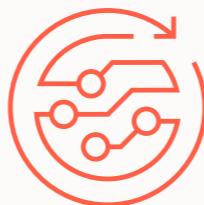
Overview

The Databricks Lakehouse provides one simple platform to handle all your data and AI use cases. It's built on an open and reliable data foundation that efficiently handles every data type and applies one common security and governance approach across all your data and cloud platforms.

Through the simple connection to Tableau, you can:



Use a lakehouse as your single source of data – from structured tabular data for visualizations to unstructured files like video and audio files for data science



Use the industry-leading Databricks data science platform on the same data lakehouse, and store results in the lakehouse to make it available to Tableau



Stream IoT and other data into the lakehouse and provide zero-latency dashboards in Tableau

Part 1

Here are the steps to get started:

01

In your Databricks Workspace, generate a **Personal Access Token** to allow Tableau secure access (if you don't have Databricks, you can start your [free trial here](#))

02

In your Databricks Workspace, create a **SQL Warehouse** for Tableau to connect to

If you wish to use **Tableau Cloud**:

03

Simply upload the included workbook titled **NYC_Taxi_Part 1.twb**

04

Watch the included video titled **Tableau Starter Kit.mov**, which includes a quick demonstration of setting up your connection after the steps above

If you wish to use **Tableau Desktop**:

05

Download and Install Tableau Desktop (2019.3 or above)

06

Download and Install Databricks ODBC Driver (2.6.15 or above)

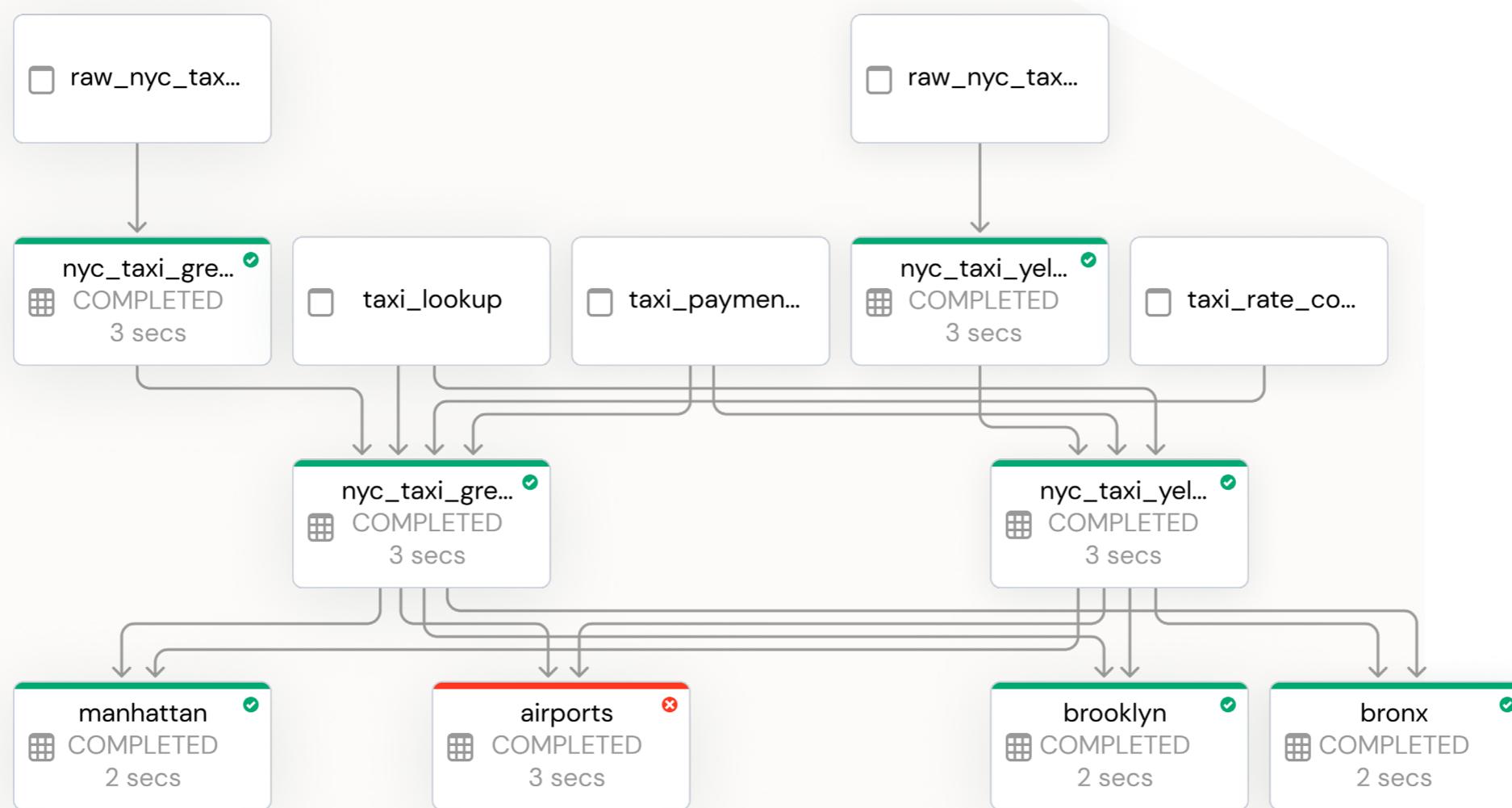
07

Download and open the Tableau Workbook included in the kit titled **NYC_Taxi_Part 1.twb**

08

Watch the included video titled **Tableau Starter Kit.mov**, which includes a quick demonstration of setting up your connection after the steps above

Notice that the sample data included in Databricks is only for January – February 2016. That is not too insightful. Let's get the latest data ourselves, and prepare it for our Tableau Dashboard! To make things interesting, we will leverage a shape file with custom taxi zones and create relationships with our yellow cab data.



Part 2

Ingest the latest taxi data from the [NYC Taxi and Limousine Commission](#), prepare it and create a table for Tableau:

01

Explore the site above and scroll down to the latest month, download the Parquet file for Yellow Taxi Trip Records

03

Import the included notebook titled **Prepare Yellow Cab Datadbc** using [these instructions](#)

02

Upload the Parquet file to DBFS using [these instructions](#)

04

Open the Tableau Workbook titled **NYC_Taxi_Part 2.twb** and edit the connection with [your warehouse and table](#)

05

You can download the Shape file [here](#). Once you connect to the file, Tableau will create a Geometry field for your polygons, which you can use to create a map.

06

There you have it! A dashboard with actual trip data that you can count on

Resources

Databricks Notebook: [Prepare Yellow Cab Datadbc](#)

Tableau Workbook Part 1: [NYC_Taxi_Part 1.twb](#)

Tableau Workbook Part 2: [NYC_Taxi_Part 2.twb](#)

About Databricks

Databricks is the lakehouse company. More than 7,000 organizations worldwide — including Comcast, Condé Nast, H&M and over 50% of the Fortune 500 — rely on the Databricks Lakehouse Platform to unify their data, analytics and AI. Databricks is headquartered in San Francisco, with offices around the globe. Founded by the original creators of Apache Spark™, Delta Lake and MLflow, Databricks is on a mission to help data teams solve the world's toughest problems. To learn more, follow Databricks on [Twitter](#), [LinkedIn](#) and [Facebook](#).

