

Using Databricks Tables

Contents:

- *Uses of Databricks Tables*
- *Limitations*
- *Before You Begin Using Databricks Tables*
- *Storing Data in Databricks Tables*
- *Reading from Databricks Tables*
- *Writing to Databricks Tables*
 - *Ad-hoc Publishing to Databricks Tables*

This section describes how you interact through the Trifacta® platform with your Azure Databricks Tables data.

- Databricks Tables enables interaction with flat-format files as schematized datasets.
- For more information, see <https://docs.microsoft.com/en-us/azure/databricks/data/tables>.

NOTE: Use of Azure Databricks Tables requires installation on Azure, integration with Azure Databricks, and an Azure Databricks connection. For more information, see *Configure for Azure Databricks*.

Uses of Databricks Tables

The Trifacta platform can use Databricks Tables for the following tasks:

1. Create datasets by reading from Databricks Tables tables.
2. Write data to Databricks Tables.

Table Type	Support	Notes
Databricks managed tables	Read /Write	
Delta tables	Read /Write	NOTE: Versioning and rollback of Delta tables is not supported within the Trifacta platform . The latest version is always used. You must use external tools to manage versioning and rollback.
External tables	Read /Write	NOTE: When writing to an external table the TRUNCATE and DROP publishing actions are not supported.
Databricks unmanaged tables	Read /Write	
Delta Tables (managed and unmanaged tables)	Read/Write	
Partitioned tables	Read	

The underlying format for Databricks Tables is Parquet.

Limitations

- Access to external Hive metastores is not supported.
- Ad-hoc publishing to Azure Databricks is not supported.
- Creation of datasets with custom SQL is not supported.

Before You Begin Using Databricks Tables

- **Databricks Tables deployment:** Your Trifacta administrator must enable use of Databricks Tables. See *Create Databricks Tables Connections*.
- **Databricks Personal Access Token:** You must acquire and save a Databricks Personal Access Token into your Trifacta account. For more information, see *Databricks Personal Access Token Page*.

Storing Data in Databricks Tables

NOTE: The Trifacta platform does not modify source data in Databricks Tables. Datasets sourced from Databricks Tables are read without modification from their source locations.

Reading from Databricks Tables

You can create a Trifacta dataset from a table or view stored in Databricks Tables.

- Read support is also available for Databricks Delta Lake.
- For more information, see *Databricks Tables Browser*.

For more information on how data types are imported from Databricks Tables, see *Databricks Tables Data Type Conversions*.

Writing to Databricks Tables

You can write data back to Databricks Tables using one of the following methods:

- Job results can be written directly to Databricks Tables as part of the normal job execution.
 - Data is written as a managed table to DBFS in Parquet format.
 - Create a new publishing action to write to Databricks Tables. See *Run Job Page*.
- For more information on how data is converted to Databricks Tables, see *Databricks Tables Data Type Conversions*.

Ad-hoc Publishing to Databricks Tables

Not supported.