

Create data pipelines by using Databricks Delta

Task 1. Clone the Databricks archive

1. From the Azure portal, go to your Databricks workspace and select Launch workspace.
2. In the left pane, select Workspace, select Users, and then select your username (the entry with the house icon).
3. In the pane that appears, select the downward-pointing chevron next to your name, and then select Import.
4. In the Import Notebooks pane, select URL, and paste in the following URL:

<https://github.com/MicrosoftDocs/mslearn-create-data-pipelines-using-databricks-delta/blob/master/DBC/06-Databricks-Delta.dbc?raw=true>

5. Select Import.
6. A folder named after the archive should appear. Select that folder. The folder contains one or more notebooks that you'll use in completing this lab.

Task 2. Complete the following notebooks

Continue with the remaining notebooks in order:

1. 01-Introducing-Delta: Get a brief overview of Databricks Delta and instructions for setting up your Databricks workspace for doing exercises.
2. 02-Create: Look into the problems with traditional data pipelines, and then resolve those issues by using Databricks Delta.
3. 03-Append: Add new records to your sample dataset in a Databricks Delta table.
4. 04-Upsert: Updated or insert records in an existing Databricks Delta table.
5. 05-Streaming: Get instructions about how to read and write streaming data by using Databricks Delta.
6. 06-Optimization: Apply Databricks Delta operations to optimize your data pipeline.
7. 07-Architecture: Work on your Databricks Delta architecture.
8. Data-Lake-Insights (located in the Optional subfolder): Find a sample project for you to explore later.