

Perform advanced data transformation

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-perform-advanced-data-transformation-in-azure-databricks/blob/master/DBC/05.2-Advanced-ETL.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

To complete the labs, continue working in your Azure Databricks workspace and open the new 05.2-Advanced-ETL folder. You'll find Python and Scala in the folder. Choose the folder for the language you prefer to use, open the corresponding folder, and then open the notebook. Follow the instructions in the notebook until you've completed the entire notebook. Continue with the remaining notebooks in order:

1. 01-Course-Overview-and-Setup - This notebook gets you started with your Databricks workspace.
2. 02-Common-Transformations - In this notebook, you perform some common data transformation by using Spark built-in functions.
3. 03-User-Defined-Functions - In this notebook, you perform custom transformation by using UDFs.
4. 04-Advanced-UDFs - In this notebook, you use advanced UDFs to perform complex data transformations.
5. 05-Joins-and-Lookup-Tables - In this notebook, you learn how to use standard and broadcast joins for tables.
6. 06-Database-Writes - This notebook contains exercises to write data to target databases in parallel, storing the transformed data from your ETL job.
7. 07-Table-Management - In this notebook, you handle managed and unmanaged tables to optimize your data storage.
8. Custom-Transformations - This notebook is located in the Optional subfolder and includes a sample project for you to explore later.