

# Working with Data in Azure Databricks

You will learn to load data by using DBFS and manipulate it by using Spark Dataframes. Databricks File System (DBFS) is a distributed file system mounted into a Databricks workspace and available on Databricks clusters. DataFrames are the distributed collections of data allowing the processing of huge amounts of data.

## Prerequisites

Before starting this lab, complete the **Getting Started with Azure Databricks** lab to set up your Azure Databricks environment and import the data and notebooks you require.

## Work with Data in Azure Databricks

In this exercise, you will learn how to load and manipulate data inside the Azure Databricks environment.

1. In a web browser, open your Azure Databricks workspace.
2. If your cluster is not running, on the **Compute** page, select your cluster and use the ► **Start** button to start it
3. In the Azure Databricks Workspace, using the command bar on the left, select **Workspace**. Then select **Users**, and ► **your\_user\_name**. Then in the folder named **01 - Introduction to Azure Databricks**, open the **Working with data in Azure Databricks** notebook.
4. Attach the notebook to your cluster. Then read the notes in the notebook, running each code cell in turn.

## Clean-up

If you're finished working with Azure Databricks for now, in Azure Databricks workspace, on the **Compute** page, select your cluster and select ■ **Terminate** to shut it down. Otherwise, leave it running for the next exercise.