**Working with Databricks**

|  |
| --- |
| **Few notes about working with Databricks**  **Importing files to Databricks**   1. Click on "Tables" icon on left bar 2. Click on "Create Table" link at top 3. Select "File" option 4. Click on "Drop file or click here to upload" to upload file 5. Note down the path somewhere to refer to later   **Working with files on Databricsk**  Run the following lines of codes in Databricks notebook  // to see usage  dbutils.fs.help()  val folder = "/FileStore/tables"  display(dbutils.fs.ls(folder))  // making directory  dbutils.fs.mkdirs("/FileStore/<folder>")  // removing directory, last parameter is for recursive or not, default not  dbutils.fs.rm("/FileStore/<folder>", true)  // Databricks hosted dataset  display(dbutils.fs.ls("dbfs:/databricks-datasets"))  For more details see Databricks Overview -> Databricks File System  ===> The other option is to use the following command:  %fs ls /mnt/sf\_open\_data/fire\_incidents  // renaming a directory  dbutils.fs.mv("<source folder|file name>", "<target folder|file name>", true)  **Reading TSV Data**  **val dataDF = spark.sqlContext.read.csv("csv")**  **.option("header", "true")**  **.option("inferSchema", "true")**  **.option("delimiter", "\t")**  **.load("<path or file>")**  **Reading CSV Data**  **val dataDF = sqlContext.read.format("csv")**  **.option("header", "true")**  **.option("inferSchema", "true")**  **.load(s"<path or file>")**  Descriptions of various data sets hosted by Databricks is [here](https://docs.cloud.databricks.com/docs/latest/databricks_guide/03%20Data%20Sources/6%20Databricks%20Public%20Datasets/1%20DBFS%20Hosted%20Datasets.html).  **Existing Data Sets**  **=======================**  dbfs:/databricks-datasets/README.md  dbfs:/databricks-datasets/Rdatasets/  dbfs:/databricks-datasets/SPARK\_README.md  dbfs:/databricks-datasets/adult/  dbfs:/databricks-datasets/airlines/  dbfs:/databricks-datasets/amazon/  dbfs:/databricks-datasets/asa/  dbfs:/databricks-datasets/bikeSharing/  dbfs:/databricks-datasets/cs100/  dbfs:/databricks-datasets/cs110x/  dbfs:/databricks-datasets/cs190/  dbfs:/databricks-datasets/data.gov/  dbfs:/databricks-datasets/flights/  dbfs:/databricks-datasets/learning-spark/  dbfs:/databricks-datasets/mnist-digits/  dbfs:/databricks-datasets/news20.binary/  dbfs:/databricks-datasets/online\_retail/  dbfs:/databricks-datasets/power-plant/  dbfs:/databricks-datasets/sample\_logs/  dbfs:/databricks-datasets/samples/  dbfs:/databricks-datasets/sfo\_customer\_survey/  dbfs:/databricks-datasets/sms\_spam\_collection/  dbfs:/databricks-datasets/songs/  dbfs:/databricks-datasets/structured-streaming/  dbfs:/databricks-datasets/tpch/  dbfs:/databricks-datasets/wiki/  dbfs:/databricks-datasets/wikipedia-datasets/ |