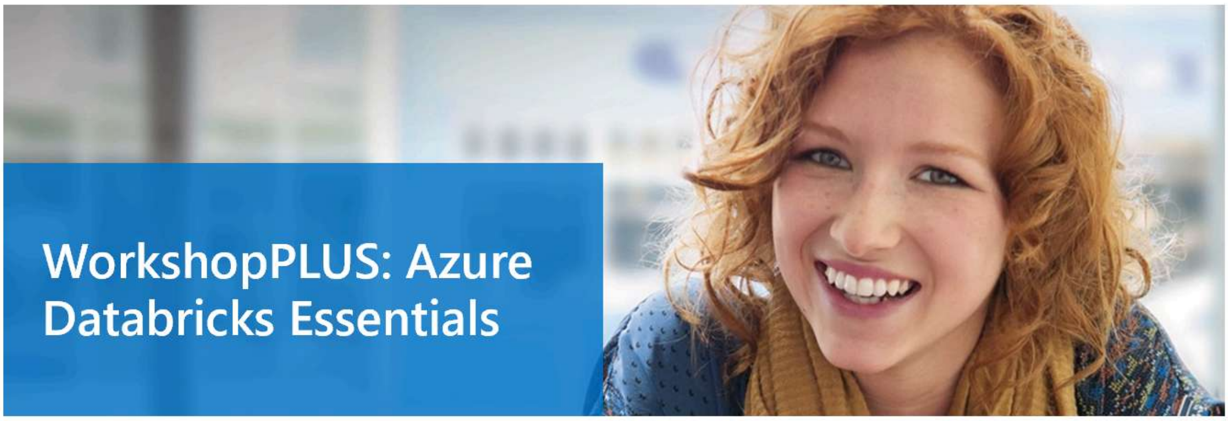

Working with Delta Tables

Template Version: 1.0

Introduction

- This Lab uses a Bike Rental dataset, and demonstrates:
  - How to create Delta table. Partition Delta table and perform batch insert operations.
  - Create delta table from a csv using tempview
  - Data exploration to find a good column for partitioning
  - Delta Table Partitioning
  - Time Travel
  - Optimize and Analyze
  - Databricks delta table Vacuum for data retention specially for timetravel to older versions

Estimated Time

30 minutes

Objectives

At the end of this lab, you will be able to:

- Create delta table and perform paritioning.
- Know how to perform Time Travel operations on Delta Table
- Know how to optimize and Analyze delta table
- Perform maintainance activity like Vacuum on Delta Table

Table of Contents

Lab: Working with Delta Tables

Tasks

1. Familiarize with Dataset information

   This lab contains two files, which would be used for creation of delta table :

   a. Days.csv
   b. Hour.csv

Both hour.csv and day.csv have the following fields, except hr which is not available in day.csv

```
- instant: record index
- dteday : date
- season : season (1:winter, 2:spring, 3:summer, 4:fall)
- yr : year (0: 2011, 1:2012)
- mnth : month ( 1 to 12)
- hr : hour (0 to 23)
- holiday : weather day is holiday or not (extracted from [Web Link])
- weekday : day of the week
- workingday : if day is neither weekend nor holiday is 1, otherwise is 0.
+ weathersit :
- 1: Clear, Few clouds, Partly cloudy, Partly cloudy
- 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist
- 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain +
- 4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog
- temp : Normalized temperature in Celsius. The values are derived via (t-t_
- atemp: Normalized feeling temperature in Celsius. The values are derived v
- hum: Normalized humidity. The values are divided to 100 (max)
- windspeed: Normalized wind speed. The values are divided to 67 (max)
- casual: count of casual users
- registered: count of registered users
- cnt: count of total rental bikes including both casual and registered
```

ref : https://archive.ics.uci.edu/ml/datasets/bike+sharing+dataset

2. Perform the lab steps

   Refer to M03_L02_Lab01.ipynb file for Databricks notebook code. Remember to import the notebook to your workspace and read the directions before executing the code

Exercise 2 has been completed