Databricks

Databricks is a cloud-based data engineering tool that provides a collaborative environment for data scientists, data engineers, and machine learning engineers. Databricks is built on top of Apache Spark, which is a distributed computing framework that is designed for big data processing. Databricks provides a web-based user interface that makes it easy to work with Spark clusters and perform various data analytics tasks.

1. Go to databricks community edition. And you are already user sign in else signup.
2. Go to create and select cluster and create cluster
3. Again go to create and select Notebook and create notebook
4. Create a DataFrame from a Databricks dataset
   %python
   diamonds =
   spark.read.csv("/databricks-datasets/Rdatasets/data-001/csv/ggplot2/diamonds.csv", header="true", inferSchema="true")
5. Manipulate the data and displays the results
   %python
   from pyspark.sql.functions import avg
   display(diamonds.select("color","price").groupBy("color").agg(avg("price")).sort("color"))