Lab 18 – Explore Databricks

1. Provision Azure Databricks Workspace

A screenshot of a computer

AI-generated content may be incorrect.

1. Create a cluster

A screenshot of a computer

AI-generated content may be incorrect.

1. Use Spark to analyze a data file

A screenshot of a computer

AI-generated content may be incorrect.

1. Create and query a table

A screenshot of a computer

AI-generated content may be incorrect.

1. Delete Azure Databricks resources

A screenshot of a computer

AI-generated content may be incorrect.

Summary:

Azure Databricks is a unified, open analytics platform for building, deploying, sharing, and maintaining enterprise-grade data, analytics, and AI solutions at scale. We begin the lab by provisioning resources. We then proceed to create an Apache Spark Cluster. We create a notebook and attach it to the Spark Cluster we created in the previous step. The first step takes time due to Spark initiation, but subsequent steps execute fast. We import the products.csv file and store it in a dataframe. Once we have data within the dataframe, we visualize it as a bar graph. We also explore the functionality of creating and executing SQL queries on our dataset. Lastly we delete the Azure databricks resources and end the lab.