**HOME ASSIGNMENT 4**

**Roll No: 208W1A1299**

**Name: MOHAMMAD RIZWANULLAH**

**Aim**: To do analysis using Apache spark in data bricks.

**Execution**:

1)Open databricks and create a new cluster with latest configuration version.

2) Now create a Notebook to work with spark.

3) Import csv file or any file for using it and do analysis using pyspark.

4) Now write code to import dataset and create some RDD to read input in spark.

5) Run the cell.

**Below are the screenshots of execution:**

Twitter followers count and protected data analysis.





