**Task 4**

**Aim:** Implementing EDA and cover all the statistical specifications and note to write whole code in map-reduce distributed way.

**Algorithm:**

* Open Databricks account
* Create a Python Notebook
* Create a cluster and attach it with the Notebook.
* Import csv file.
* Create DataFrame and perform as many as statistical functions as I can.

**Implementation / Source Code**

1. Import SQLContext



### Load and creating DataFrame.

### 

### Print the schema using printSchema.

### 

### Describe the Schema.

### 

### Adding a New Column.

### 

### Calculating Percentage of marks of every student.

### 

### Plotting Bar Graph to represent number of Students Pass or Fail.

### 

### Sorting a column.

### 

### Dropping a column.

### 

### Print the schema.

### 