PART-B

SPARK PROGRAMS USING PYTHON

1. Write a **Spark program using Python**, to analyse the given **Weather Report Data** and to generate a report with cities having maximum and minimum temperature for a particular year.

**Code:**  weather.py

import sys

if(len(sys.argv)!=4):

    print("Provide Input File and Output Directory")

    sys.exit(0)

from pyspark import SparkContext

sc =SparkContext()

f = sc.textFile(sys.argv[1])

temp=f.map(lambda x: (int(x[15:19]),int(x[87:92])))

mini=temp.reduceByKey(lambda a,b:a if a<b else b)

mini.saveAsTextFile(sys.argv[2])

maxi=temp.reduceByKey(lambda a,b:a if a>b else b)

maxi.saveAsTextFile(sys.argv[3])

**Execution:**

spark-submit weather.py input.txt minimum maximum

