|  |  |
| --- | --- |
| **Ex. No: 5** | **Spark and PySpark** |
|  |

**Aim: To Install and understand the usage of the Pyspark**

**Program:**

Word\_count.py

from pyspark import SparkContext

# Initialize SparkContext

sc = SparkContext("local", "Word Count")

# Read the file

text\_file = sc.textFile("u.txt")

# Perform word count

counts = text\_file.flatMap(lambda line: line.split("\t")) \

                  .map(lambda word: (word, 1)) \

                  .reduceByKey(lambda a, b: a + b)

# Collect the results

output = counts.collect()

# Print the output

for (word, count) in output:

    print(f"{word}: {count}")

# Stop the SparkContext

sc.stop()

movie\_rating.py

from pyspark import SparkContext

# Initialize SparkContext

sc = SparkContext("local", "Movie Ratings Distribution")

# Read the file

text\_file = sc.textFile("u.txt")

# Parse the lines to extract movie ID and rating

movie\_ratings = text\_file.map(lambda line: line.split("\t")) \

                         .map(lambda fields: (fields[1], int(fields[2])))

# Count the number of each rating for each movie

ratings\_distribution = movie\_ratings.map(lambda x: ((x[0], x[1]), 1)) \

                                    .reduceByKey(lambda a, b: a + b) \

                                    .map(lambda x: (x[0][0], (x[0][1], x[1]))) \

                                    .groupByKey() \

                                    .mapValues(list)

# Collect and print the results

output = ratings\_distribution.collect()

for (movie\_id, ratings) in output:

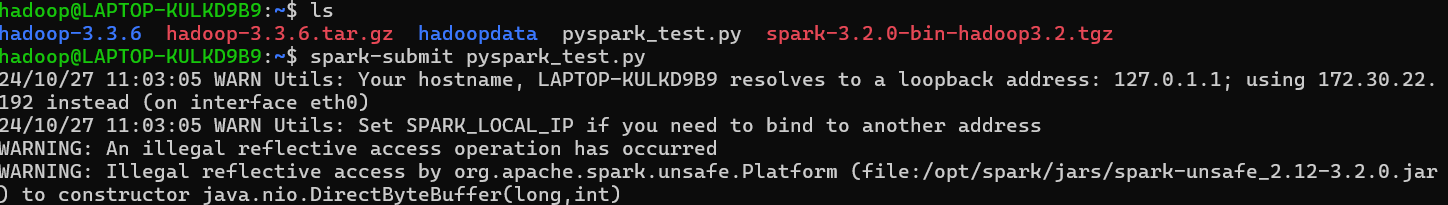
    print(f"Movie ID: {movie\_id}, Ratings Distribution: {ratings}")

# Stop the SparkContext

sc.stop()

**Output:**

A black screen with white text

Description automatically generatedA screenshot of a computer

Description automatically generated

A screen shot of a computer program

Description automatically generatedA computer screen with white text

Description automatically generatedA screenshot of a computer

Description automatically generatedA computer screen shot of a computer program

Description automatically generatedA screenshot of a computer

Description automatically generated

**Result: Required results achieved by performing following codes**