

ASSIGNMENT – V

TITLE :

Develop and Implement Applications Using Apache Spark.

NAME : Shinde Shubham Dnyandev. **ROLL NO :** 23107121. **CLASS :** TY-B

PROGRAM :**INPUT :**

I am Student of VPKBIET. I am Studying in AI&DS Department.

CODE :

```
from pyspark import SparkContext
```

```
sc = SparkContext("spark://localhost:7077", "WordCount")
```

```
text_file = sc.textFile("/mnt/spark/README.md")
```

```
counts = text_file.flatMap(lambda line: line.split()) \  
    .map(lambda word: (word, 1)) \  
    .reduceByKey(lambda a, b: a + b)
```

```
# Save output
```

```
counts.saveAsTextFile("/home/admin1/wordcount_output")
```

```
sc.stop()
```

OUTPUT :

The screenshot shows the Apache Spark 3.4.3 Web UI interface. At the top, it displays the URL 'spark://localhost:7077' and the title 'Spark Master at spark://localhost:7077'. Below this, there's a summary section with metrics like 'Alive Workers: 1', 'Cores In Use: 2 Total: 0 Used', 'Memory in use: 4.0 GiB Total: 0.0 B Used', 'Resources in use: 0 Running, 3 Completed', 'Applications: 0 Running, 3 Completed', 'Drivers: 0 Running, 0 Completed', and 'Status: ALIVE'. There are three main sections: 'Workers (1)', 'Running Applications (0)', and 'Completed Applications (3)'. The 'Workers (1)' section lists one worker with ID 'worker-20250813160945-172.16.9.35-42207', Address '172.16.9.35:42207', State 'ALIVE', Cores '2 (0 Used)', and Memory '4.0 GiB (0.0 B Used)'. The 'Running Applications (0)' section is empty. The 'Completed Applications (3)' section lists three completed applications: 'WordCount' (ID app-20250813161849-0002), 'JavaWordCount' (ID app-20250813161430-0001), and 'Spark shell' (ID app-20250813161201-0000). Each application entry includes its name, cores, memory per executor, resources per executor, submitted time, user, state, and duration.

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration
app-20250813161849-0002	WordCount	2	2.0 GiB		2025/08/13 16:18:49	admin1	FINISHED	0.2 s
app-20250813161430-0001	JavaWordCount	2	2.0 GiB		2025/08/13 16:14:30	admin1	FINISHED	0.2 s
app-20250813161201-0000	Spark shell	2	2.0 GiB		2025/08/13 16:12:01	admin1	FINISHED	0.2 s