**Terna Engineering College**

**Computer Engineering Department**

Program: Sem VII

[**Course: Big Data Analytics & Computational Lab -I (BDA&CL-I)**](https://github.com/Amey-Thakur/BIG-DATA-ANALYTICS-AND-COMPUTATIONAL-LAB-I)

**Experiment No. 07**

**PART B**

**(PART B: TO BE COMPLETED BY STUDENTS)**

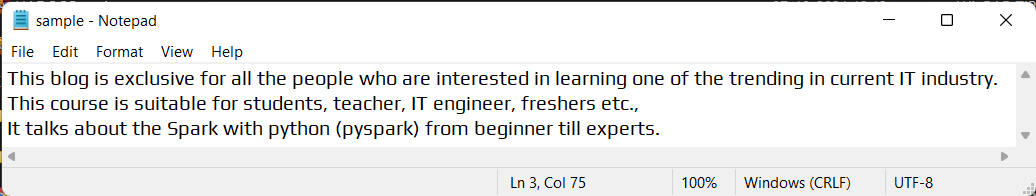
***(Students must submit the soft copy as per the following segments within two hours of the practical. The soft copy must be uploaded on the Blackboard or emailed to the concerned lab in charge faculties at the end of the practical in case there is no Blackboard access available)***

| Roll No. 50 | Name: AMEY THAKUR |
| --- | --- |
| Class: BE-COMPS-50 | Batch: B3 |
| Date of Experiment: 05-10-2021 | Date of Submission: 05-10-2021 |
| Grade : |  |

**Aim:** To implement a word count program using Map Reduce.

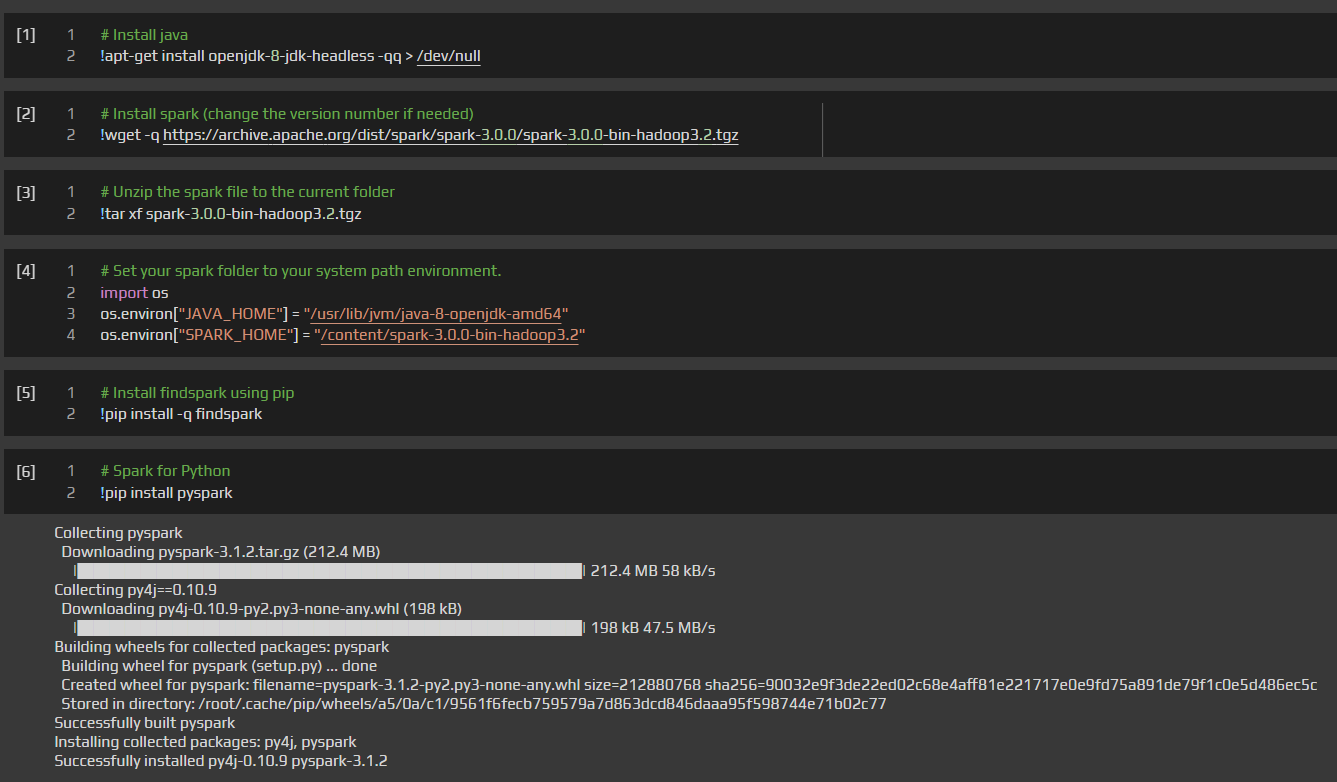
**B.1 Software Code written by a student:**

* **Input file:** sample.txt

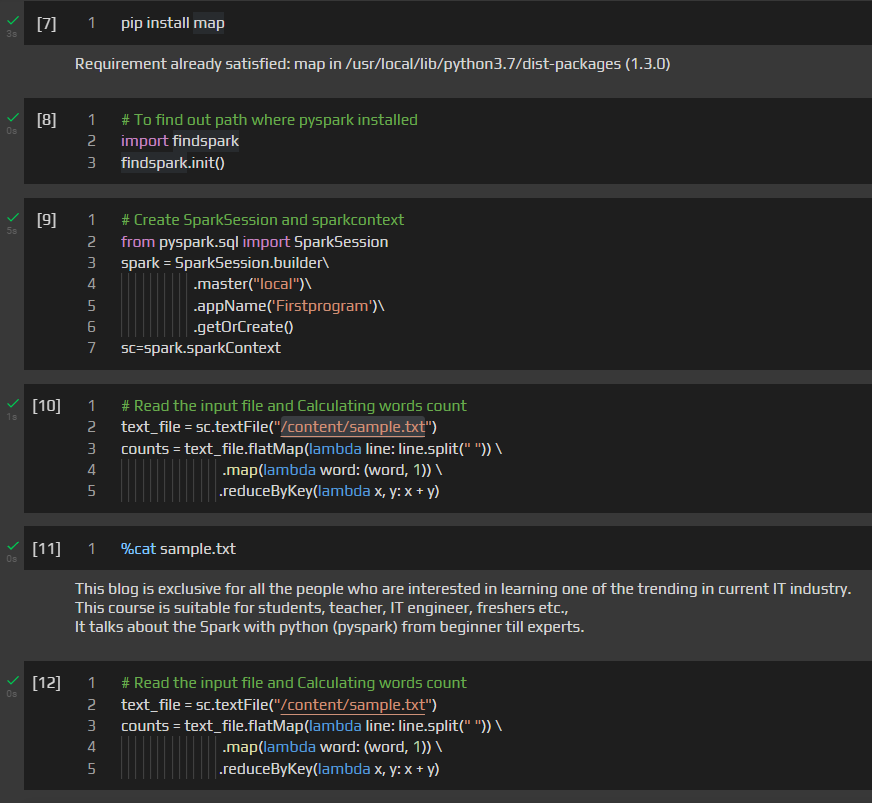


**B.2 Input and Output:**

* **Step 1:**  Installing Hadoop and Pyspark.



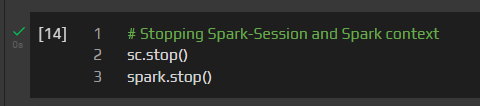
* **Step 2:** Installing Map and reading the sample text file and calculating words counts.



* **Step 3:** Printing each word with its respective count.



* **Step 4:** Stopping the Spark-session and Spark context.



**B.3 Observations and learning:**

We are able to acquire fundamental enabling techniques and scalable algorithms like Hadoop, Map Reduce and NO SQL in big data analytics.

**B.4 Conclusion:**

We have learned To implement a word count program using Map Reduce.

**B.5 Question of Curiosity:**

1. What is Flatten?

**Ans:**

The FLATTEN operator looks like a UDF syntactically, but it is actually an operator that changes the structure of tuples and bags in a way that a UDF cannot. Flatten un-nests tuples as well as bags. The idea is the same, but the operation and result are different for each type of structure.

1. How does Pig differ from MapReduce?

**Ans:**

| **Sr. No.** | **Pig** | **MapReduce** |
| --- | --- | --- |
| **1** | It is a Data Flow Language. | It is a Data Processing Language. |
| **2** | It converts the query into map-reduce functions. | It converts the job into map-reduce functions. |
| **3** | It is a High-level Language. | It is a Low-level Language. |
| **4** | Makes it easy for the user to perform Join operations. | It is difficult for the user to perform join operations. |
| **5** | There is less compilation time as the Pig operator converts it into MapReduce jobs. | It has several jobs therefore execution time is more. |