*Spark*

**A C A D G I L D Page 1**

Session 7: Using Spark’s API in Scala, Java and Python

Assignment 2*Spark*

**A C A D G I L D Page 2**

*Session 7: Spark’s API in Scala, Java and Python*

*Assignment 2*

**Table of Contents**

1. Introduction .......................................................................................................................................... 3

2. Objective ............................................................................................................................................... 3

3. Prerequisites ......................................................................................................................................... 3

4. Associated Data Files ............................................................................................................................ 3

5. Problem Statement ............................................................................................................................... 3

6. Approximate Time to Complete Task ................................................................................................... 3 *Spark*

**A C A D G I L D Page 3**

**1. Introduction**

In this assignment you need to write the code for the given question.

**2. Objective**

This assignment will help you to understand Big Data basics.

**3. Prerequisites**

APIs in Spark

**4. Associated Data Files**

N/A

**5. Problem Statement**

1. Execute python Word count program.

|  |
| --- |
| input= sc.textFile("hdfs://manish.com:8020/data/word.txt") |
|  |

|  |
| --- |
| count = input.flatMap(lambda line: line.split(" ")) |
|  |

|  |
| --- |
| .map(lambda word: (word, 1)) |
|  |

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

count.saveAsTextFile("hdfs://manish.com:8020/data/output.txt")

**6. Approximate Time to Complete Task**

30 min