*Spark*

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

Session 4: Spark Shell

Assignment 1*Spark*

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

*Session 4: Spark Shell*

*Assignment 1*

**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**

None

**4. Associated Data Files**

N/A

**5. Problem Statement**

1. Find the average of numbers.

Initialize the base RDD with a list of 1,2,3,4,5,6,7,8,9,10

scala> val num=sc.parallelize(1 to 10,2)

num: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[0] at parallelize at <console>:27

scala> val y = num.reduce(\_ + \_)

y: Int = 55

**Another way :**

scala> val x = sc.parallelize(1 to 10, 2)

x: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[1] at parallelize at <console>:27

scala> val y = x.reduce((accum,n) => (accum + n))

y: Int = 55

2. Find the average of numbers.

Read the numbers from a file, each number in separate line. 1,2,3,4,5,6,7,8,9,10

package com.spark.sparkExample

import org.apache.spark.SparkConf;

import org.apache.spark.SparkContext;

object Average {

def main(args: Array[String]) = {

val conf = new SparkConf()

.setAppName("Average ")

.setMaster("local")

val sc = new SparkContext(conf)

//Read some example file to a test RDD

val test = sc.textFile("/home/work/file.txt")

val result= test.reduce(\_ + \_)

println(result)

}

}

**6. Approximate Time to Complete Task**

30 min