*Spark*

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

Session 3: Spark Core

Assignment 3*Spark*

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

*Session 3: Spark Core*

*Assignment 3*

**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. Write the code to Turn a collection into a RDD and perform map operation on it to cube every number and filter the number which are divided by two and three.

val a=sc.parallelize(List(1,2,3,4,5,6,7,8))

a: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[12] at parallelize at <console>:27

scala> val c=a.map(x=>(x,(x\*x\*x)))

c: org.apache.spark.rdd.RDD[(Int, Int)] = MapPartitionsRDD[13] at map at <console>:29

scala> c.collect

res2: Array[(Int, Int)] = Array((1,1), (2,8), (3,27), (4,64), (5,125), (6,216), (7,343), (8,512))

scala> val div1 = a.filter(\_ % 2==0)

div1: org.apache.spark.rdd.RDD[Int] = MapPartitionsRDD[14] at filter at <console>:29

scala> div1.collect

res3: Array[Int] = Array(2, 4, 6, 8)

scala> val div2 = a.filter(\_ % 3==0)

div2: org.apache.spark.rdd.RDD[Int] = MapPartitionsRDD[15] at filter at <console>:29

scala> div2.collect

res4: Array[Int] = Array(3, 6)

**6. Approximate Time to Complete Task**

20 min