*Spark*

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

Session 6: In-Memory Management

Assignment 3*Spark*

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

*Session 6: In-Memory Management*

*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 questions

**2. Objective**

This assignment will help you to understand Big Data basics.

**3. Prerequisites**

None

**4. Associated Data Files**

https://drive.google.com/open?id=0ByJLBTmJojjzYWxickJ5U3JHbGM

The above dataset consists of the details of movies. The dataset description is as follows

Metadata header: id, name, year, rating, duration

**5. Problem Statement**

1. Find the top 5 longest movies according to their duration

val textFile = sc.textFile("/tmp/movie.txt")

val InputSplit = textFile.map(x => x.split(“,”))

val mapInput = InputSplit.map(x => (x,x.last.toDouble))

val sort = mapInput.sortBy(\_.\_2,false)

val output = sort.take(5)

output.collect()

2. Find the top 5 rated movies in the year 1995

val textFile = sc.textFile("/tmp/movie.txt")

val InputSplit = textFile.map(x => x.split(“,”))

val filterInput = InputSplit.filter(x => x.contains(“1995”))

val mapInput = filterInput.map(x => (x,x(3))

val sortInput = mapInput.sortBy(\_.\_2,false)

val output = sortInput.take(5)

output.collect

**6. Approximate Time to Complete Task**

30 min