*Spark*

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

Session 6: In-Memory Management

Assignment 1*Spark*

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

*Session 6: In-Memory Management*

*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 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. What are the movies that were released in 2012 and have >= 4.0 rating

val input = sc.textFile(“hdfs://manish.com:8020/data/movies.txt”)

val mapIn = input.map(x => x.split(“,”))

val filterMap = mapIn.filter(x=> x(2).contains(“2012”))

val filteredMap = filterMap.filter(x=>x(3) >= “4.0”)

filteredMap.collect()

2. List the movies per year

val in = sc.textFile(“hdfs://manish.com:8020/data/movies.txt”)

val mapInput = in.map(x => x.split(“,”))

val mapByYear = mapInput.map(x => (x,x(2)))

mapByYear.collect()

**6. Approximate Time to Complete Task**

30 min