*Spark*

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

Session 8: Build and Deployment

Assignment 2*Spark*

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

*Session 8: Build and Deployment*

*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. Read a whole CSV files into Spark application (You can take any csv file)

val movieFileData = sc.textFile("hdfs://manish.com:8020/tmp/movies.csv")

val header = movieFileData.first()

**6. Approximate Time to Complete Task**

30 min