*Spark*

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

Session 8: Build and Deployment

Assignment 1*Spark*

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

*Session 8: Build and Deployment*

*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**

APIs in Spark

**4. Associated Data Files**

N/A

**5. Problem Statement**

1. Parse JSON records using Spark (You can take any JSON record)

val sqlcontext = new org.apache.spark.sql.SQLContext(sc) // creates sql context from spark context

val jsonFile = sqlcontext.read.json("hdfs://manish.com:8020/data/testFile.json") //read json file from hdfs

**6. Approximate Time to Complete Task**

30 min