*Spark*

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

Session 2: Introduction to Spark

Assignment 1*Spark*

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

*Session 2: Introduction to Spark*

*Assignment 1*

**Table of Contents**

1. Introduction ........................................................................................................................................ 2

2. Objective ............................................................................................................................................. 2

3. Prerequisites ....................................................................................................................................... 2

4. Associated Data Files ............................................................................................................................ 2

5. Problem Statement ............................................................................................................................... 3

6. Approximate Time to Complete Task ................................................................................................... 3

**1. Introduction**

In this assignment you need to select the correct answer for the given questions.

**2. Objective**

This assignment will help you to understand Big Data basics.

**3. Prerequisites**

None

**4. Associated Data Files**

N/A *Spark*

**A C A D G I L D Page 3**

**5. Problem Statement**

1. What is in-memory processing in Spark?

a) Processing data in each node

b) Storing data in RDBMS during processing

c) Maximum effective utilization of RAM during the processing

d) Using more no of CPU threads

ANS c) Maximum effective utilization of RAM during the processing

2. What are the features of Apache Spark?

a) In-memory processing

b) Ease of use APIs

c) Unified high level tools

d) Runs Everywhere (Hadoop, Mesos, standalone, or in the cloud. It can access diverse data sources including HDFS, Cassandra, HBase, S3.)

e) All the above

f) None of the above

Ans e) All the above

3. Apache Hadoop is more active project than Apache Spark in open source community in the last year.

a) true

b) false

Ans a)true

4. Driver program will be launched in every node of the worker.

a) true

b) false

Ans a)true

5. Spark only supports Stream processing.

a. Yes

b. No

Ans ) b.No

**6. Approximate Time to Complete Task**

10 min

*Spark*

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