Session 1: How to Solve Big Data

Assignment 2

Big Data Developer

Assignment 2 – Try the given quiz questions and forward the answers in a word document.

Table of Contents

Big Data Developer

1. Introduction

In this assignment you need to select one right choice for the questions given on the topics discussed in the first session.

2. Objective

This assignment will help you to consolidate the concepts learnt in the session 1.

3. Prerequisites:
None

4. Associated Data Files
None

5. Problem Statement:

1. What constitutes Big Data?

a. volume

b. velocity

c. variety

d. All the above　　*Answer: d*

2. Solution Identified to handle Big Data is?

a. Hadoop　　*Answer: a*

b. storm

c. spark

d. All of these

3. Doug cutting added DFS & MapReduce support to:

a) Nutch

b) spark

c) GFS   *Answer: c*

d) facebook

4. Which file system is followed by Hadoop?

a. DFS

b. GFS

c. HDFS   *Answer: c*

d. NTFS

5. How much data can be considered as a BigData with regards to volume?

a. Mega Bytes

b. Terabytes

c. Petabytes

d. b & c   *Answer: d*

6. Can Hadoop process data in real time?

a. Yes   *Answer: a*

b. No

c. None of these

7. Hadoop is developed using?

a. C

b. C++

c. Java   *Answer: c*

d. Ruby

8. Hadoop reduces cost of operation by?

a. Using commodity Hardware

b. Using less number of nodes

c. Limited cluster size

d. None of these   *Answer: d*

9. Search engine adopted in Hadoop is?

a. Google

b. Yahoo   *Answer: b*

c. Solr

c. AOL

10. Stream processing can be achieved in Hadoop using?

a. Storm   *Answer: a*

b. Spark

c. Both of these

d. None of these

11. What is the highest data unit we have at present?

a. Petabyte   *Answer: a*

b. Exabyte

c. Terabyte

d. Brontobyte

12. Hadoop is well suited for clusters where:

a. Number of nodes keep changing   **Answer: a**

b. Number of nodes are constant

c. Most number of free nodes

d. None of these

13. What is batch processing?

a. processing of previously collected jobs in a single batch.   **Answer: a**

b. processing of previously collected jobs in different batches.

c. processing of jobs at that moment they are collected

d. None of these

14. What is stream processing?

a. Processing data as streams   **Answer: a**

b. Processing data in batches

c. Both of these

d. None of these

15. Hadoop uses?

a. Batch processing

b. Near Real time processing

c. Real time processing

d. All the above   **Answer: d**

6. Expected Output

None.

7. Approximate Time to Complete Task

10 mins