

Hadoop 1.X

Instructor: Jayant Mohite

Results: _____

Name: _____

Class: Wisdom Sprouts

Period: Hadoop

Date: 6th August 2016

1. Due to Hadoop's ability to manage unstructured and semi-structured data and because of its scale-out support for handling ever-growing quantities of data, many experts view it as a replacement for the enterprise data warehouse.
 - a. True
 - b. False

2. Hadoop cannot store Structured data
 - a. True
 - b. False

3. If there are 2 Shuffle and Sorts then there are 2 mappers and 1 reducer
 - a. True
 - b. False

4. In case of failure of Namenode, the Secondary Namenode can take its place
 - a. True
 - b. False

5. Number of Mappers are 7 and number of Reducers is 0 then number of output files is 7.
- a. True
 - b. False
6. MapReduce can process data stored either in filesystem (unstructured) or in a database (structured).
- a. True
 - b. False
7. Hadoop is a replacement to RDBMS.
- a. True
 - b. False
8. All Map Reduce implementations implement exactly same algorithm.
- a. True
 - b. False
9. For using Hadoop to process your data, the data has to be moved/ingested into HDFS.
- a. True
 - b. False
10. Hadoop runs the jobs by dividing it into Map tasks and Reduce tasks.
- a. True
 - b. False

11. Hadoop can handle petabytes of data which are store across hundreds or thousands of nodes and physical storage servers.

- a. True
- b. False

12. This MapReduce the take the input data and break it for further processing across various nodes within the Hadoop instance.

- a. True
- b. False

13. HDFS instances are classified into two components: the Name Node, which maintains metadata to keep the track on placement of physical data across the Hadoop instance and Data Nodes, which store the data.

- a. True
- b. False

14. Hadoop is a framework that works with a variety of related tools. Common stack includes:

- a. MapReduce, Hive and HBase
- b. MapReduce, MySQL and Google Apps
- c. MapReduce, Hummer and Iguana
- d. MapReduce, Heron and Trumpet

15. The MapReduce programming model widely used in analytics was developed at_____.

- a. Apache Foundation
- b. Google
- c. Microsoft Research
- d. Apple

16. What was Hadoop named after?

- a. Creator Doug Cutting's favorite circus act
- b. Cutting's high school rock band
- c. The toy elephant of Cutting's son
- d. A sound Cutting's laptop made during Hadoop's development

17. Which file is used for updating mapreduce settings?

- a. core-site
- b. mapred-site
- c. hdfs-site
- d. hadoop-env.sh

18. Which programming model is also used as the associated implementation for processing and generating large data sets with a parallel, distributed algorithm on a cluster in Hadoop?

- a. MapReduce
- b. Pig
- c. Hive
- d. SQL

19. _____ is an optimized file system for distributed processing of very large datasets on commodity hardware.

- a. RDBMS
- b. Flat File System
- c. HDFS File System
- d. MapReduce

20. What is the default Block size?

- a. 128Mb
- b. 64Mb
- c. 256Mb
- d. 512Mb

21. Which file is used for updating mapreduce settings?

- a. core-site
- b. mapred-site
- c. hdfs-site
- d. hadoop-env.sh

22. All of the following accurately describe Hadoop, EXCEPT:

- a. Open Source
- b. Real Time
- c. Java Based
- d. Distributed Processing

23. Hadoop is a framework that works with a variety of related tools. Common cohorts include:

- a. Pig, Hive, Oozie
- b. Pig, Sheep, Goat
- c. Linux, Windows, Hadoop
- d. Pig, Hive, JDK

24. Consider case scenario: In M/R system, - HDFS block size is 64 MB. We have 3 files of size 64K, 65Mb and 127Mb. How many input splits will be made by Hadoop framework?

- a. 1
- b. 2
- c. 4
- d. 5

25. Compulsory Phases of Hadoop are

- a. Map, Reduce
- b. Map, Shuffle and Sort, Reduce
- c. Map, Combiner, Partitioner, Shuffle and Sort, Reduce
- d. Map, Combiner, Reduce

26. The port number of namenode is

- a. 8020
- b. 8021
- c. 8022
- d. 8024

27. The port number of Job Tracker is

- a. 8020
- b. 8021
- c. 8022
- d. 8023

28. Url of Namenode is

- a. localhost:55070
- b. localhost:55030
- c. localhost:50090
- d. localhost:60090

29. Sqoop can be used to

- a. Only Import data from SQL Databases
- b. Import Export data from to SQL Databases
- c. Only Export data to SQL Databases
- d. Delete data from SQL Databases

30. Which of the following statements is correct? In Hadoop there exists

- a. one JobTracker per Hadoop job
- b. one JobTracker per Mapper
- c. one JobTracker per node
- d. one JobTracker per cluster

31. Which of the following is true for Namenode?

- a. The master node that manages the file system namespace.
- b. It maintains the file system tree and metadata for all the files/directories within.
- c. Keeps track of the location of all the Data Nodes for a given file.
- d. All of the above