

Hadoop 1.X

Instructor: Jayant Mohite

Name: _____

Period: Hadoop

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. Flume server resides inside HDFS
 - a. True
 - b. False
3. Flume and Sqoop do the same job with only difference that flume works with txt files and Sqoop works with databases
 - a. True
 - b. False
4. Hadoop cannot store Structured data
 - a. True
 - b. False
5. If there are 2 Shuffle and Sorts then there are 2 mappers and 1 reducer

- a. True
 - b. False
6. We cannot configure the Source and Sink of Flume
- a. True
 - b. False
7. Combiner is also called as Mini Mapper
- a. True
 - b. False
8. We can use Partitioner even if there is only one Reducer.
- a. True
 - b. False
9. In case of failure of Namenode, the Secondary Namenode can take its place
- a. True
 - b. False
10. Hadoop cannot work with commodity Hardware
- a. True
 - b. False
11. Hive and Pig are exactly the same

- a. True
 - b. False
12. Number of Mappers are 7 and number of Reducers is 0 then number of output files is 7.
- a. True
 - b. False
13. Flume sits between Webserver and Hadoop cluster.
- a. True
 - b. False
14. MapReduce can process data stored either in filesystem (unstructured) or in a database (structured).
- a. True
 - b. False
15. Hadoop is a replacement to RDBMS.
- a. True
 - b. False
16. All Map Reduce implementations implement exactly same algorithm.
- a. True
 - b. False
17. For using Hadoop to process your data, the data has to be moved/ingested into HDFS.

- a. True
- b. False

18. Use of commodity hardware is suitable for slave nodes.

- a. True
- b. False

19. Partioner reduces traffic from Reducer to Mapper.

- a. True
- b. False

20. Hadoop runs the jobs by dividing it into Map tasks and Reduce tasks.

- a. True
- b. False

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

- a. True
- b. False

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

- a. True
- b. False

23. 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

24. 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

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

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

26. 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

27. Flume is a _____.

- a. Import Export tool
- b. Messaging Service
- c. Scheduling and Workflow utility
- d. Data flow language

28. Which file is used for updating mapreduce settings?

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

29. 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

30. _____ 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

31. Which one of the following is an example of unstructured data?

- a. call detail records
- b. web server logs
- c. sensory data
- d. Voicemail

32. Which of the following is semi-structured format?

- a. email
- b. accord
- c. Huppa
- d. word document

33. If number of Reducers is 0 then Output of Mapper is stored in _____.

- a. Hadoop File system
- b. Linux File system
- c. Mapper Filesystem
- d. Partioner Filesystem

34. What is the default Block size?

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

35. Which file is used for updating mapreduce settings?

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

36. Partitioner is used to

- a. Reduce work of Reducer
- b. Reduce work of Combiner
- c. Chose one Reducer when there are more than 1 Mappers
- d. Chose one Reducer when there are more than 1 Reducers

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

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

38. Pentaho tools is installed on

- a. Windows
- b. Linux
- c. HDFS
- d. Hive

39. 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

40. 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

41. 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

42. If the number of Mappers is 3, then the number of Combiners is

- a. 1
- b. 2
- c. 3
- d. 4

43. If number of Reducers is 6, then number of Partitioner is

- a. 1
- b. 3
- c. 5
- d. 6

44. The port number of namenode is

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

45. The port number of Job Tracker is

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

46. Url of Namenode is

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

47. 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

48. Combiner can be used only when the operation is

- a. Commutative and Associative
- b. Commutative and Descriptive
- c. Associative and Descriptive
- d. Only descriptive

49. 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

50. 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