1. **Which of the following is the daemon of Hadoop?**
2. Node manager
3. DataNode
4. NameNode
5. All of above
6. **Which one of the following is false about Hadoop?**
7. It is a distributed framework
8. The main algorithm used in Hadoop is Map Reduce
9. Hadoop can work with commodity hardware
10. All are true
11. **Which of the following is component of Hadoop?**
12. YARN
13. HDFS
14. MapReduce
15. All of the above
16. **Which of the following is true about Hadoop?**
17. Hadoop MapReduce processes the data at very low latency
18. Hadoop can't handle structured data
19. Hadoop is schema-less – we don't need to define a schema before writing the data
20. None of the above
21. **Hadoop Framework is written in**
22. Python
23. Java
24. C++
25. Scala
26. **The archive file created in Hadoop has the extension of**
27. .hrh
28. .har
29. .hrc
30. .hrar
31. **Apache Hadoop achieves reliability by replicating the data across multiple hosts, and hence does not require \_\_\_\_\_\_\_\_ storage on hosts.**
32. FileSystem
33. RAID
34. Local FS
35. Operating system
36. **Which of the following platforms does Apache Hadoop run on?**
37. Bare metal
38. Unix-like
39. Cross-platform
40. Debian
41. **Hadoop works in**
42. Centralized processing fashion
43. master – slave fashion
44. worker/slave fashion
45. All of the mentioned
46. **Which of the following property gets configured on mapred-site.xml?**
47. Replication factor
48. Java Environment variables
49. Directory names to store hdfs files
50. Which framework to use with MapReduce jobs
51. **Which type of data Hadoop can deal with is**
52. Structured
53. Semi – structured
54. Unstructured
55. All of the above
56. **Which of the below component deals with ingesting streaming data into Hadoop?**
57. Flume
58. Oozie
59. Hive
60. Kafka
61. **As compared to RDBMS, Apache Hadoop**
62. Has higher data Integrity
63. Does ACID transactions
64. Is suitable for read and write many times
65. Works better on unstructured and semi-structured data
66. **Hadoop command to list all the blocks corresponding to each file in the hdfs?**
67. hadoop fsck / -files -blocks
68. hadoop fsck / -blocks -files
69. hadoop ls / -blocks -files
70. hadoop fchk / -files -blocks
71. **In which all languages you can code in Hadoop?**
72. Java
73. Python
74. C++
75. All of the above
76. **Which Hadoop configuration file contains the setting for HDFS daemons?**
77. yarn-site.xml
78. hdfs-site.xml
79. mapred-site.xml
80. None of the above
81. **Which of the following is a distributed 3-dimensional database?**
82. HDFS
83. HBase
84. Both the above
85. None of the above
86. **Which of the file contains the configuration setting for NodeManager and ResourceManager?**
87. yarn-site.xml
88. hdfs-site.xml
89. mapred-site.xml
90. None of the above
91. **Which of the following is used for machine learning on Hadoop?**
92. Hive
93. Pig
94. HBase
95. Mahout

**20. Which of the following is used to ingest data into Hadoop clusters?**

1. Flume
2. Sqoop
3. Both the above
4. Nonw of the above

**21. The client reading the data from HDFS filesystem in Hadoop does which of the following?pla**

1. Gets only the block locations form the namenode
2. Gets the data from the namenode
3. Gets both the data and block location from the namenode
4. Gets the block location from the datanode

**22. Which of the following is the true about metadata?**

1. Metadata shows the structure of HDFS directories/files
2. FsImage & EditLogs are metadata files
3. Metadata contain information like number of blocks, their location, replicas
4. All of the above

**23. Which one of the following stores data?**

1. Name node
2. Data node
3. Master node
4. None of these

**24. For reading/writing data to/from HDFS, clients first connect to**

1. NameNode
2. DataNode
3. Secondary NameNode
4. None of the above

**25. Where is HDFS replication factor controlled?**

1. mapred-site.xml
2. yarn-site.xml
3. core-site.xml
4. hdfs-site.xml

**26. As compared to RDBMS, Hadoop**

1. Has higher data Integrity
2. Does ACID transactions
3. IS suitable for read and write many times
4. Works better on unstructured and semi-structured data

**27. Pig operates in mainly how many nodes?**

1. Two
2. Three
3. Four
4. Five

**28. Point out the correct statement.**

1. You can run Pig in either mode using the “pig” command
2. You can run Pig in batch mode using the Grunt shell
3. You can run Pig in interactive mode using the FS shell
4. None of the mentioned

**29. You can run Pig in batch mode using \_\_\_\_\_\_\_\_\_\_**

1. Pig shell command
2. Pig scripts
3. Pig options
4. All of the mentioned

**30. Pig Latin statements are generally organized in one of the following ways?**

1. A LOAD statement to read data from the file system
2. A series of “transformation” statements to process the data
3. A DUMP statement to view results or a STORE statement to save the results
4. All of the mentioned

**31. Point out the wrong statement.**

1. To run Pig in local mode, you need access to a single machine
2. The DISPLAY operator will display the results to your terminal screen
3. To run Pig in mapreduce mode, you need access to a Hadoop cluster and HDFS installation
4. All of the mentioned

**32. Which of the following function is used to read data in PIG?**

1. WRITE
2. READ
3. LOAD
4. None of the mentioned

**33. $ pig -x tez\_local … will enable \_\_\_\_\_\_\_\_ mode in Pig.**

1. Mapreduce
2. Tez
3. Local
4. None of the mentioned

**34. \_\_\_\_\_\_\_\_\_ operator is used to review the schema of a relation.**

1. DUMP
2. DESCRIBE
3. STORE
4. EXPLAIN

**35. Point out the wrong statement.**

1. ILLUSTRATE operator is used to review how data is transformed through a sequence of Pig Latin statements
2. ILLUSTRATE is based on an example generator
3. Several new private classes make it harder for external tools such as Oozie to integrate with Pig statistics
4. None of the mentioned

**36. The \_\_\_\_\_\_\_\_ class mimics the behavior of the Main class but gives users a statistics object back**.

1. PigRun
2. PigRunner
3. RunnerPig
4. None of the mentioned

**37. Which of the following will compile the Pigunit?**

1. $pig\_trunk ant pigunit-jar
2. $pig\_tr ant pigunit-jar
3. $pig\_ ant pigunit-jar
4. None of the mentioned

**38. The \_\_\_\_\_\_\_\_ class mimics the behavior of the Main class but gives users a statistics object back.**

1. PigRun
2. PigRunner
3. RunnerPig
4. None of the mentioned

**39. \_\_\_\_\_\_\_\_\_\_ abstract class has three main methods for loading data and for most use cases it would suffice to extend it.**

1. Load
2. LoadFunc
3. FuncLoad
4. None of the mentioned

**40. \_\_\_\_\_\_\_\_\_\_\_\_ method will be called by Pig both in the front end and back end to pass a unique signature to the Loader.**

1. relativeToAbsolutePath()
2. setUdfContextSignature()
3. getCacheFiles()
4. getShipFiles()