1. Which command lists the blocks that make up each file in the file system?

a) Hdfs fsck / -files -blocks

b) Hdfs fsck / -blocks -files

c) Hdfs fchk / -blocks -files

d) Hdfs fchk / -files -blocks

Answer: a) Hdfs fsck / -files -blocks

2. In Hadoop, Snappy and LZO are examples of which of the following?

a) Mechanisms of file transport between data nodes

b) Mechanisms of Data synchronization

c) Mechanisms of data compression

d) Mechanisms of data Replication

Answer: c) Mechanisms of data compression

3. What is the default input format?

a) There is no default input format. The input format always should be specified.

b) The default input format is xml. Developer can specify other input formats as appropriate if xml is not the correct input.

c) The default input format is a sequence file format. The data needs to be preprocessed before using the default input format.

d) The default input format is TextInputFormat with byte offset as a key and entire line as a value.

Answer: d) The default input format is TextInputFormat with byte offset as a key and entire line as a value.

4. Which of the following are among the duties of the Data Nodes in HDFS?

a) Maintain the file system tree and metadata for all files and directories.

b) Store and retrieve blocks when told to by clients or the NameNode.

c) Manage the file system namespace.

d) Control the execution of an individual map task or a reduce task.

e) None of these options is correct.

Answer: b) Store and retrieve blocks when told to by clients or the NameNode.

5. The namenode loses its only copy of fsimage file. We can recover this from which of the following?

a) Secondary namenode

b) It can't be recovered.

c) Checkpoint node

d) Datanodes

Answer: c) Checkpoint node

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

a) Gets the block location from the datanode

b) Gets the data from the namenode

c) Gets both the data and block location from the namenode

d) Gets only the block locations form the namenode

Answer: d) Gets only the block locations form the namenode

7. The HDFS command to create the copy of a file from a local system is which of the following?

a) copyFromLocal

b) CopyLocal

c) CopyFromLocal

d) copyfromlocal

Answer: a) copyFromLocal

8. In the local disk of the namenode the files which are stored persistently are which of the following?

a) Name space image and edit log

b) Block locations and namespace image

c) Edit log and block locations

d) Namespace image, edit log and block locations

Answer: a) Name space image and edit log

9. What is distributed cache?

a) The distributed cache is a component that caches java objects.

b) The distributed cache is special component on name node that will cache frequently used data for faster client response. It is used during reduce step.

c) The distributed cache is special component on data node that will cache frequently used data for faster client response. It is used during map step.

d) The distributed cache is a component that allows developers to deploy jars for Map-Reduce processing.

Answer: c) The distributed cache is special component on data node that will cache frequently used data for faster client response. It is used during map step.

10. The datanode and namenode are, respectiviley, which of the following?

a) Worker and Master nodes

b) Both are worker nodes

c) Master and worker nodes

d) None of these

Answer: a) Worker and Master nodes

11. Which of these provides a Stream processing system used in Hadoop ecosystem?

a) Tez

b) Hive

c) Solr

d) Spark

Answer: d) Spark

12. If we increase the size of files stored in HDFS without increasing the number of files, then the memory required by namenode does which of the following?

a) Remains unchanged

b) Decreases

c) Increases

d) May or may not increase

Answer: b) Decreases

13. When a backup node is used in a cluster there is no need of which of the following?

a) Secondary name node

b) Secondary data node

c) Rack awareness

d) Check point node

Answer: d) Check point node

14. Which of the following components retrieves the input splits directly from HDFS to determine the number of map tasks?

a) The NameNode

b) The TaskTrackers

c) The JobTracker

d) None of these

e) The JobClient

Answer: c) The JobTracker

15. The default replication factor for HDFS file system in Hadoop is which of the following?

a) 1

b) 2

c) 3

d) 4

Answer: c) 3

16. "Under replication" in HDFS means which of the following?

a) The frequency of replication in data nodes is very low.

b) Replication process is very slow in the data nodes.

c) No replication is happening in the data nodes.

d) The number of replicated copies is less than as specified by the replication factor.

Answer: d) The number of replicated copies is less than as specified by the replication factor.

17. When a client contacts the namenode for accessing a file, the namenode responds with which of the following?

a) Block ID and hostname of any one of the data nodes containing that block

b) Block ID of the file requested

c) Size of the file requested

d) Block ID and hostname of all the data nodes containing that block

Answer: d) Block ID and hostname of all the data nodes containing that block

18. All the files in a directory in HDFS can be merged together using which of the following?

a) merge all

b) remerge

c) put merge

d) get merge

Answer: d) get merge

19. The source of HDFS architecture in Hadoop originated as which of the following?

a) Yahoo distributed filesystem

b) Facebook distributed filesystem

c) Azure distributed filesystem

d) Google distributed filesystem

Answer: d) Google distributed filesystem

20. The concept using multiple machines to process data stored in distributed system is not new.

The High-performance computing (HPC) uses many computing machines to process large volume of data stored in a storage area network (SAN). As compared to HPC, Hadoop

A - Can process a larger volume of data.

B - Can run on a larger number of machines than HPC cluster.

C - Can process data faster under the same network bandwidth as compared to HPC.

D - Cannot run compute intensive jobs.

Answer: c) Can process data faster under the same network bandwidth as compared to HPC.

21. Hadoop differs from volunteer computing in

A - Volunteers donating CPU time and not network bandwidth.

B - Volunteers donating network bandwidth and not CPU time.

C - Hadoop cannot search for large prime numbers.

D - Only Hadoop can use mapreduce.

Answer: a) Volunteers donating CPU time and not network bandwidth.

22. As compared to RDBMS, Hadoop

A - Has higher data Integrity.

B - Does ACID transactions

C - IS suitable for read and write many times

D - Works better on unstructured and semi-structured data.

Answer: d) Works better on unstructured and semi-structured data.

23. What is the main problem faced while reading and writing data in parallel from multiple disks?

A - Processing high volume of data faster.

B - Combining data from multiple disks.

C - The software required to do this task is extremely costly.

D - The hardware required to do this task is extremely costly.

Answer: b) Combining data from multiple disks.

24. Which of the following is true for disk drives over a period of time?

A - Data Seek time is improving faster than data transfer rate.

B - Data Seek time is improving more slowly than data transfer rate.

C - Data Seek time and data transfer rate are both increasing proportionately.

D - Only the storage capacity is increasing without increase in data transfer rate.

Answer: b) Data Seek time is improving more slowly than data transfer rate.

25. Data locality feature in Hadoop means

A - store the same data across multiple nodes.

B - relocate the data from one node to another.

C - co-locate the data with the computing nodes.

D - Distribute the data across multiple nodes.

Answer: c) co-locate the data with the computing nodes.

26. Which of these provides a Stream processing system used in Hadoop ecosystem?

A - Solr

B - Tez

C - Spark

D – Hive

Answer: c) Spark

27. HDFS files are designed for

A - Multiple writers and modifications at arbitrary offsets.

B - Only append at the end of file

C - Writing into a file only once.

D - Low latency data access.

Answer: b) Only append at the end of file

28. A file in HDFS that is smaller than a single block size

A - Cannot be stored in HDFS.

B - Occupies the full block's size.

C - Occupies only the size it needs and not the full block.

D - Can span over multiple blocks.

Answer: c) Occupies only the size it needs and not the full block.

29. HDFS block size is larger as compared to the size of the disk blocks so that

A - Only HDFS files can be stored in the disk used.

B - The seek time is maximum

C - Transfer of a large files made of multiple disk blocks is not possible.

D - A single file larger than the disk size can be stored across many disks in the cluster.

Answer: d) A single file larger than the disk size can be stored across many disks in the cluster.

30. In a Hadoop cluster, what is true for a HDFS block that is no longer available due to disk corruption or machine failure?

A - It is lost for ever

B - It can be replicated form its alternative locations to other live machines.

C - The namenode allows new client request to keep trying to read it.

D - The Mapreduce job process runs ignoring the block and the data stored in it.

Answer: b) It can be replicated form its alternative locations to other live machines.

31. Which utility is used for checking the health of a HDFS file system?

A - fchk

B - fsck

C - fsch

D – fcks

Answer: b) fsck

32. Which command lists the blocks that make up each file in the filesystem.

A - hdfs fsck / -files -blocks

B - hdfs fsck / -blocks -files

C - hdfs fchk / -blocks -files

D - hdfs fchk / -files -blocks

Answer: a) hdfs fsck / -files -blocks

33. The datanode and namenode are respectively

A - Master and worker nodes

B - Worker and Master nodes

C - Both are worker nodes

D – None

Answer: b) Worker and Master nodes

34. In the local disk of the namenode the files which are stored persistently are

A - namespace image and edit log

B - block locations and namespace image

C - edit log and block locations

D - Namespace image, edit log and block locations.

Answer: a) namespace image and edit log

35. When a client communicates with the HDFS file system, it needs to communicate with

A - only the namenode

B - only the data node

C - both the namenode and datanode

D - None of these

Answer: c) both the namenode and datanode

36. What mechanisms Hadoop uses to make namenode resilient to failure.

A - Take backup of filesystem metadata to a local disk and a remote NFS mount.

B - Store the filesystem metadata in cloud.

C - Use a machine with at least 12 CPUs

D - Using expensive and reliable hardware.

Answer: a) Take backup of filesystem metadata to a local disk and a remote NFS mount.

37. The main role of the secondary namenode is to

A - Copy the filesystem metadata from primary namenode.

B - Copy the filesystem metadata from NFS stored by primary namenode

C - Monitor if the primary namenode is up and running.

D - Periodically merge the namespace image with the edit log.

Answer: d) Periodically merge the namespace image with the edit log.

38. For the frequently accessed HDFS files the blocks are cached in

A - the memory of the datanode

B - in the memory of the namenode

C - Both A&B

D - In the memory of the client application which requested the access to these files.

Answer: a) the memory of the datanode

39. User applications can instruct the namenode to cache the files by

A - adding cache file names to cache pool

B - adding cache config to cache pool

C - adding cache directive to cache pool

D - passing the file names as parameters to the cache pool

Answer: c) adding cache directive to cache pool

40. In Hadoop 2.x release HDFS federation means

A - Allowing namenodes to communicate with each other.

B - Allow a cluster to scale by adding more datanodes under one namenode.

C - Allow a cluster to scale by adding more namenodes.

D - Adding more physical memory to both namenode and datanode.

Answer: c) Allow a cluster to scale by adding more namenodes.

41. Under HDFS federation

A - Each namenode manages metadata of the entire filesystem.

B - Each namenode manages metadata of a portion of the filesystem.

C - Failure of one namenode causes loss of some metadata availability from the entire filesystem.

D - Each datanode registers with each namenode.

Answer: b) Each namenode manages metadata of a portion of the filesystem.

42. The main goal of HDFS High availability is

A - Faster creation of the replicas of primary namenode.

B - To reduce the cycle time required to bring back a new primary namenode after existing primary fails.

C - Prevent data loss due to failure of primary namenode.

D - Prevent the primary namenode form becoming single point of failure.

Answer: b) To reduce the cycle time required to bring back a new primary namenode after existing primary fails.

43. As part of the HDFS high availability a pair of primary namenodes are configured. What is true for them?

A - When a client request comes, one of them chosen at random serves the request.

B - One of them is active while the other one remains powered off.

C - Datanodes send block reports to only one of the namenodes.

D - The standby node takes periodic checkpoints of active namenode's namespace.

Answer: d) The standby node takes periodic checkpoints of active namenode's namespace.

44. Zookeeper ensures that

A - All the namenodes are actively serving the client requests

B - Only one namenode is actively serving the client requests

C - A failover is triggered when any of the datanode fails.

D - A failover can not be started by hadoop administrator.

Answer: b) Only one namenode is actively serving the client requests

45. Under Hadoop High Availability, Fencing means

A - Preventing a previously active namenode from start running again.

B - Preventing the start of a failover in the event of network failure with the active namenode.

C - Preventing the power down to the previously active namenode.

D - Preventing a previously active namenode from writing to the edit log.

Answer: d) Preventing a previously active namenode from writing to the edit log.

46. Which of the following is not a fencing mechanism for a previously active namenode?

A - Disabling its network port via a remote management command.

B - Revoking its access to shared storage directory.

C - Formatting its disk drive.

D – STONITH

Answer: c) Formatting its disk drive.

47. The property used to set the default filesystem for Hadoop in core-site.xml is

A - filesystem.default

B - fs.default

C - fs.defaultFS

D - hdfs.default

Answer: b) fs.default

48. The default replication factor for HDFS file system in hadoop is

A - 1

B - 2

C - 3

D – 4

Answer: c) 3

49. When running on a pseudo distributed mode the replication factor is set to

A - 2

B - 1

C - 0

D – 3

Answer: b) 1

50. For a HDFS directory the replication factor(RF) is

A - same as the RF of the files in that directory

B - Zero

C - 3

D - Does not apply.

Answer: d) Does not apply.

51. The following is not permitted on HDFS files

A - Deleting

B - Renaming

C - Moving

D - Executing.

Answer: d) Executing.

52. Hadoop differs from volunteer computing in

A - Volunteers donating CPU time and not network bandwidth.

B - Volunteers donating network bandwidth and not CPU time.

C - Hadoop cannot search for large prime numbers.

D - Only Hadoop can use mapreduce.

Answer: a) Volunteers donating CPU time and not network bandwidth.

53. Zookeeper ensures that

A - All the namenodes are actively serving the client requests

B - Only one namenode is actively serving the client requests

C - A failover is triggered when any of the datanode fails.

D - A failover can not be started by hadoop administrator.

Answer: b) Only one namenode is actively serving the client requests

54. - The client reading the data from HDFS filesystem in Hadoop

A - gets the data from the namenode

B - gets the block location from the datanode

C - gets only the block locations form the namenode

D - gets both the data and block location from the namenode

Answer: c) gets only the block locations form the namenode

55. The hadoop frame work is written in

A - C++

B - Python

C - Java

D – GO

Answer: c) Java

56. HDFS stands for

A - Highly distributed file system.

B - Hadoop directed file system

C - Highly distributed file shell

D - Hadoop distributed file system.

Answer: d) Hadoop distributed file system.

57. The inter process communication between different nodes in Hadoop uses

A - REST API

B - RPC

C - RMI

D - IP Exchange

Answer: b) RPC

58. The data from a remote hadoop cluster can

A - not be read by another hadoop cluster

B - be read using http

C - be read using hhtp

D - be read using hftp

Answer: d) be read using hftp

59. In a HDFS system with block size 64MB we store a file which is less than 64MB. Which of the following is true?

A - The file will consume 64MB

B - The file will consume more than 64MB

C - The file will consume less than 64MB.

D - Can not be predicted.

Answer: c) The file will consume less than 64MB.

60. What is writable?

A - Writable is a java interface that needs to be implemented for streaming data to remote servers.

B - Writable is a java interface that needs to be implemented for HDFS writes.

C - Writable is a java interface that needs to be implemented for MapReduce processing.

D - None of these answers are correct.

Answer: c) Writable is a java interface that needs to be implemented for MapReduce processing.

61. Which of the following are among the duties of the Data Nodes in HDFS?

A - Maintain the file system tree and metadata for all files and directories.

B - None of the options is correct.

C - Control the execution of an individual map task or a reduce task.

D - Store and retrieve blocks when told to by clients or the NameNode.

Answer: d) Store and retrieve blocks when told to by clients or the NameNode.

62. Which of the following deal with small files issue

a) Hadoop archives

b) Sequence files

c) HBase

d) All of the above

Answer: d) All of the above

63. Which of the following feature overcomes this single point of failure

a) None of the above

b) HDFS federation

c) High availability

d) Erasure coding

Answer: b) HDFS federation

64. Which statement is true about NameNode High Availability

a) Solve Single point of failure

b) For high scalability

c) Reduce storage overhead to 50%

d) None of the above

Answer: a) Solve Single point of failure

65. In NameNode HA, when active node fails, which node takes the responsibility of active node

a) Secondary NameNode

b) Backup node

c) Standby node

d) Checkpoint node

Answer: c) Standby node

66. What are the advantages of 3x replication schema in Hadoop

a) Fault tolerance

b) High availability

c) Reliability

d) All of the above

Answer: d) All of the above

67. What are the advantages of HDFS federation in Hadoop?

a) Isolation

b) Namespace scalability

c) Improves throughput

d) All of the above

Answer: d) All of the above

68. Rack Awareness improves

a) Data high availability and reliability

b) Performance of the cluster

c) Network bandwidth

d) All of the above

Answer: d) All of the above

69. Which property is used to enable/disable speculative execution

a) mapred.map.tasks.speculative.execution

b) mapred.reduce.tasks.speculative.execution

c) Both the above

d) None of the above

Answer: c) Both the above

70. In which process duplicate task is created to improve the overall execution time

a) Erasure coding

b) Speculative execution

c) HDFS federation

d) None of the above

Answer: b) Speculative execution

71. In which mode each daemon runs on a single node but there is separate java process for each daemon

a) Local (Standalone) mode

b) Fully distributed mode

c) Pseudo-distributed mode

d) None of the above

Answer: c) Pseudo-distributed mode

72. In which mode each daemon runs on a single node as a single java process

a) Local (Standalone) mode

b) Pseudo-distributed mode

c) Fully distributed mode

d) None of the above

Answer: a) Local (Standalone) mode

73. In which mode all daemons execute in separate nodes

a) Local (Standalone) mode

b) Pseudo-distributed mode

c) Fully distributed mode

d) None of the above

Answer: c) Fully distributed mode

74. Which configuration file is used to control the HDFS replication factor?

a) mapred-site.xml

b) hdfs-site.xml

c) core-site.xml

d) yarn-site.xml

Answer: b) hdfs-site.xml

75. How to adjust the size of a distributed cache

a) local.cache.size.

b) mapred.cache.size.

c) hdfs.cache.size.

d) distributedcache.size.

Answer: a) local.cache.size.

76. What is the default size of distributed cache

a) 8 GB

b) 10 GB

c) 12 GB

d) 16 GB

Answer: b) 10 GB

77. Distributed cache can cache files

a) Jar Files

b) Read-only text files

c) Archives

d) All of the above

Answer: d) All of the above

78. The total number of partitioner is equal to

a) The number of reducer

b) The number of mapper

c) The number of combiner

b) None of the above

Answer: a) The number of reducer

79. Which of the following Hadoop config files is used to define the heap size?

a) hdfs-site.xml

b) core-site.xml

c) hadoop-env.sh

d) mapred-site.xml

Answer: c) hadoop-env.sh

80. Which of the following feature you will use submit jars, static files for MapReduce job during runtime

a) Distributed cache

b) Speculative execution

c) Data locality

d) Erasure coding

Answer: a) Distributed cache

81. Which of the following method used to set the output directory

a) FileOutputFormat.setOutputgetpath()

b) OutputFormat.setOutputpath()

c) FileOutputFormat.setOutputpath()

d) OutputFormat.setOutputgetpath()

Answer: c) FileOutputFormat.setOutputpath()

82. Which tool is used to distributes data evenly across datanode

a) Balancer

b) Disk balancer

Answer: a) Balancer

83. Which tool is used to distributes data evenly on all disks of a datanode

a) Balancer

b) Disk Balancer

Answer: b) Disk Balancer

84. Which of the following must be set true enable diskbalnecr in hdfs-site.xml

a) dfs.balancer.enabled

b) dfs.disk.balancer.enabled

c) dfs.diskbalancer.enabled

d) dfs.disk.balancer.enabled

Answer: d) dfs.disk.balancer.enabled

85. In disk balancer datanode uses which volume choosing the policy to choose the disk for the block.

a) Round-robin

b) Available space

c) All of the above

d) None of the above

Answer: d) None of the above

86. Which among the following is configuration files in Hadoop

a) core-site.xml

b) hdfs-site.xml

c) yarn-site.xml

d) All of the above

Answer: d) All of the above

87. Which of the following is used for large inter/intra-cluster copying

a) fsck

b) distch

c) DistCp

d) dtutil

Answer: c) DistCp

88. Hadoop uses hadoop

a) Troubleshooting

b) Performance reporting purpose

c) Monitoring

d) All of the above

Answer: c) Monitoring

89. Is it possible to provide multiple inputs to Hadoop?

a) Yes

b) No

Answer: a) Yes

90. Is it possible to have Hadoop job output in multiple directories?

a) Yes

b) No

Answer: a) Yes

91. Which of the following is used to provide multiple outputs to Hadoop?

a) MultipleOutputFormat

b) MultipleOutputs class

c) FileOutputFormat

d) DBInputFormat

Answer: b) MultipleOutputs class

92. Which of the following command is used to check for various inconsistencies

a) zkfc

b) fs

c) fsck

d) fetchdt

Answer: c) fsck

93. What does commodity Hardware in Hadoop world mean?

a) Very cheap hardware

b) Industry standard hardware

c) Discarded hardware

d) Low specifications Industry grade hardware

Answer: d) Low specifications Industry grade hardware

94. Distributed cache files can't be accessed in Reducer.

a) True

b) False

Answer: b) False

95. Pig is a:

a) Programming Language

b) Data Flow Language

c) Query Language

d) Database

Answer: b) Data Flow Language

96. Distributed Cache can be used in

a) Mapper phase only

b) Reducer phase only

c) In either phase, but not on both sides simultaneously

d) In either phase

Answer: d) In either phase

97. Which of the following is not a valid Hadoop config file?

a) core-default.xml

b) hdfs-default.xml

c) hadoop-default.xml

d) mapred-default.xml

Answer: c) hadoop-default.xml

98. It is necessary to default all the properties in Hadoop config files.

a) True

b) False

Answer: b) False

99. Which of the following is a column-oriented database that runs on top of HDFS

a) Hive

b) Sqoop

c) HBase

d) Flume

Answer: c) HBase

100. Which command is used to show all the Hadoop daemons that are running on the machine

a) distcp

b) jps

c) fsck

Answer: b) jps

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

a) MapReduce, Hive and HBase

b) MapReduce, MySQL and Google Apps

c) MapReduce, Hummer and Iguana

d) MapReduce, Heron and Trumpet

Answer: a) MapReduce, Hive and HBase

102. Which of the following is the daemon of Hadoop?

a) NameNode

b) Node manager

c) Data Node

d) All of the above

Answer: d) All of the above

103. Which one of the following is false about Hadoop?

a) It is a distributed framework

b) The main algorithm used in it is Map Reduce

c) It runs with commodity hard ware

d) all are true

Answer: d) all are true

104. Hadoop Framework is written in

a) Python

b) Java

c) C++

d) Scala

Answer: b) Java

105. Which of the following is component of Hadoop?

a) YARN

b) HDFS

c) MapReduce

d) All of the above

Answer: d) All of the above

106. The archive file created in Hadoop has the extension of

.a) hrh

.b) har

c) hrc

d) hrar

Answer: .b) har

107. Which command is used to check the status of all daemons running in the HDFS.

a) jps

b) fsck

c) distcp

d) None of the above

Answer: a) jps

108. What license is Apache Hadoop distributed under?

a) Apache License 2.0

b) Shareware

c) Mozilla Public License

d) Commercial

Answer: a) Apache License 2.0

109. Which of the following platforms does Apache Hadoop run on ?

a) Bare metal

b) Unix-like

c) Cross-platform

d) Debian

Answer: c) Cross-platform

110. Apache Hadoop achieves reliability by replicating the data across multiple hosts, and hence does not require _____ storage on hosts.

a) Standard RAID levels

b) RAID

c) ZFS

d) Operating system

Answer: b) RAID

111. Which of the following is the correct statement

a) Data locality means moving computation to data instead of data to computation

b) Data locality means moving data to computation instead of computation to data

c) Both the above

d) None of the above

Answer: a) Data locality means moving computation to data instead of data to computation

112. Hadoop works in

a) master-worker fashion

b) master – slave fashion

c) worker/slave fashion

d) All of the mentioned

Answer: b) master – slave fashion

113. Which type of data Hadoop can deal with is

a) Structured

b) Semi - structured

c) Unstructured

d) All of the above

Answer: d) All of the above

114. Which statement is false about Hadoop

a) It runs with commodity hard ware

b) It is a part of the Apache project sponsored by the ASF

c) It is best for live streaming of data

d) None of the above

Answer: c) It is best for live streaming of data

115. Which of the following property gets configured on mapred-site.xml ?

a) Replication factor

b) Java Environment variables.

c) Directory names to store hdfs files.

d) Host and port where MapReduce job runs.

Answer: d) Host and port where MapReduce job runs.

116. Which of the below apache system deals with ingesting streaming data to hadoop

a) Flume

b) Oozie

c) Hive

d) Kafka

Answer: a) Flume

117. As compared to RDBMS, Apache Hadoop

a) Has higher data Integrity

b) Does ACID transactions

c) Is suitable for read and write many times

d) Works better on unstructured and semi-structured data.

Answer: d) Works better on unstructured and semi-structured data.

118. Which command lists the blocks that make up each file in the filesystem

a) hdfs fsck / -files -blocks

b) hdfs fsck / -blocks -files

c) hdfs fchk / -blocks -files

d) hdfs fchk / -files -blocks

Answer: a) hdfs fsck / -files -blocks

119. In which all languages you can code in Hadoop

a) Java

b) Python

c) C++

d) All of the above

Answer: d) All of the above

120. Whch of the file contains the configuration setting for HDFS daemons

a) yarn-site.xml

b) hdfs-site.xml

c) mapred-site.xml

d) None of the above

Answer: b) hdfs-site.xml

121. All of the following accurately describe Hadoop, EXCEPT

a) Open source

b) Real-time

c) Java-based

d) Distributed computing approach

Answer: b) Real-time

122. Whch of the file contains the configuration setting for NodeManager and ResourceManager

a) yarn-site.xml

b) hdfs-site.xml

c) mapred-site.xml

d) None of the above

Answer: a) yarn-site.xml

123. Hadoop can be used to create distributed clusters, based on commodity servers, that provide low-cost processing and storage for unstructured data

a) True

b) False

Answer: a) True

124. Which of the following is a distributed multi-level database?

a) HDFS

b) HBase

c) Both the above

d) None of the above

Answer: b) HBase

125. Which of the following is used for machine learning on Hadoop?

a) Hive

b) Pig

c) HBase

d) Mahoot

Answer: d) Mahoot

126. Which of the following is used to ingest streaming data into Hadoop clusters

a) Flume

b) Sqoop

c) Both the above

d) None of the above

Answer: a) Flume

127. Hadoop distributed file system behaves similarly to which of the following:

a) RAID-1 Filesystem

b) RAID-0 Filesystem

c) Both the above

d) All of the above

Answer: a) RAID-1 Filesystem

128. Zookeeper ensures that

a) All the namenodes are actively serving the client requests

b) A failover is triggered when any of the datanode fails.

c) Only one namenode is actively serving the client requests

d) A failover can not be started by hadoop administrator.

Answer: c) Only one namenode is actively serving the client requests

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

a) Flume

b) Sqoop

c) Both the above

d) None of the above

Answer: c) Both the above

130. Which of the following is a data processing engine for clustered computing?

a) Drill

b) Oozie

c) Spark

d) All of the above

Answer: c) Spark

131. Which tool could be used to move data from RDBMS data to HDFS?

a) Sqoop

b) Flume

c) Both the above

d) None of the above

Answer: d) None of the above

132. All the files in a directory in HDFS can be merged together using which of the following?

a) Put merge

b) Get merge

c) Remerge

d) Merge all

Answer: b) Get merge

133. Which of these provides a Stream processing system used in Hadoop ecosystem?

a) Hive

b) Solr

c) Tez

d) Spark

Answer: d) Spark

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

a) Gets only the block locations form the namenode

b) Gets the data from the namenode

c) Gets both the data and block location from the namenode

d) Gets the block location from the datanode

Answer: a) Gets only the block locations form the namenode

135. Which of the following jobs are optimized for scalability but not latency

a) Mapreduce

b) Drill

c) Oozie

d) Hive

Answer: d) Hive

136. Can multiple clients write into an HDFS file concurrently?

a) Yes

b) No

Answer: b) No

137. Which of the following command is used to enter Safemode

a) hadoop dfsadmin –safemode get

b) bin dfsadmin –safemode get

c) hadoop dfsadmin –safemode enter

d) None of the above

Answer: c) hadoop dfsadmin –safemode enter

138. Which of the following command is used to come out of Safemode

a) hadoop dfsadmin -safemode leave

b) hadoop dfsadmin -safemode exit

c) hadoop dfsadmin -safemode out

d) None of the above

Answer: a) hadoop dfsadmin -safemode leave

139. During Safemode Hadoop cluster is in

a) Read-only

b) Write-only

c) Read-Write

d) None of the above

Answer: a) Read-only

140. HDFS allows a client to read a file which is already opened for writing?

a) False

b) True

Answer: b) True

141. What happens when a file is deleted from the command line?

a) It is permanently deleted if trash is enabled.

b) It is permanently deleted and the file attributes are recorded in a log file.

c) It is placed into a trash directory common to all users for that cluster.

d) None of the above

Answer: b) It is permanently deleted and the file attributes are recorded in a log file.

142. Which one of the following statements is false about Distributed Cache?

a) Hive

b) Drill

c) Oozie

d) Mapreduce

Answer: a) Hive

143. Which among the following are the features of Hadoop

a) Open source

b) Fault-tolerant

c) High Availability

d) All of the above

Answer: d) All of the above

144. Checkpoint node download the FsImage and EditLogs from the NameNode & then merge them & store the modified FsImage

a) Into persistent storage

b) Back to the active NameNode

Answer: a) Into persistent storage

145. Which statement is true about Safemode

a) It is a maintenance state of NameNode

b) In Safemode, HDFS cluster is in read only

c) In Safemode, NameNode doesn't allow any modifications to the file system

d) All of the above

Answer: d) All of the above

146. Which command is used to know the current status of the safe mode

a) hadoop dfsadmin –safemode get

b) hadoop dfsadmin –safemode getStatus

c) hadoop dfsadmin –safemode status

d)None of the above

Answer: a) hadoop dfsadmin –safemode get