

Big Data-Hadoop-L2

1) What is the command for checking disk usage in hadoop.

1. Hadoop fs -disk -space
2. Hadoop fs -diskusage
3. Hadoop fs -du
4. None of the above

▲ Hide Answer

Answer : c

2) How to set the replication factor of a file.

1. hadoop fs -setrep -w 3 -R path
2. hadoop fs -repset -w 3 -R path
3. hadoop fs -setrep -e 3 -R path
4. hadoop fs -repset -e 3 -R path

▲ Hide Answer

Answer : a

3) How to set an auto map side join in hive?

1. Set hive.exec.auto.map=true;
2. Set hive.auto.convert.join=true;
3. Set hive.mapred.auto.map.join=true;
4. Set hive.map.auto.convert=true;

▲ Hide Answer

Answer : b

4) If a database is having tables with data and you want to delete, then which one is the correct command.

1. Drop database database_name nonrestrict
2. Drop database database_name cascade
3. Drop schema database_name noncascade
4. Drop database database_name

▲ Hide Answer

Answer : b

5) What is the default serde used in hive?

1. Lazy serdy
2. Default serde
3. Binary serde
4. None of the above.

▲ Hide Answer

Answer : a

7) How can you add a catch file to a job?

1. DistributedCatch.addCatchFile()
2. DistributedCatch.addCatchArchive()
3. DistributedCatch.setCatchFiles()
4. All of the above.

▲ Hide Answer

Answer : d

6) Create table (ld int, dt string, ip int) //line1
partitioned by (dt string) //line2
stored as rcfile; //line 3

1. error in line 1;
2. error in line 2;
3. error in line 3;
4. no error;

▲ Hide Answer

Answer : a

8) Which one is not a master daemon?

1. Namenode
2. Jobtracker
3. Tasktracker
4. None of these.

▲ Hide Answer

Answer : b

9) How can you check your available space and total space in hadoop system?

1. HDFS dfsadmin -action
2. HDFS dfsadmin -property
3. HDFS dfsadmin -report
4. None of these

▲ Hide Answer

Answer : c

10) Job history is used to support job recovery after a jobtracker restart which parameter you need to set?

1. mapred.jobtracker.restart.recover
2. mapred.jobtracker.set.recover
3. mapred.jobtracker.restart.recover.history
4. None of the above

▲ Hide Answer

Answer : a

11) What is TTL in Hbase.

1. HBase will automatically delete rows once the expiration time is reached.
2. HBase will automatically disable rows once the expiration time is reached.
3. It's just a time taken for executing a job.
4. None.

▲ Hide Answer

Answer : a

12) Does HDFS allow appends to files.

1. True
2. False

▲ Hide Answer

Answer : a

13) In which file you can Set HBase environment variables.

1. hbase-env.sh
2. hbase-var.sh
3. hbase-update.sh
4. None.

▲ Hide Answer

Answer : a

14) Which file you need to edit to change rate at which HBase files are rolled and so as the level at which HBase logs messages.

1. log4j.properties
2. zookeeper.properties
3. hbase.properties
4. None

▲ Hide Answer

Answer : a

15) What is the default block size in apache HDFS.

1. 64MB
2. 128MB
3. 512MB
4. 1024MB

▲ Hide Answer

Answer : a

18) Data node decides where to store to data,

1. Yes
2. False

▲ Hide Answer

Answer : b

16) What is the default port for jobtracker web UI?

1. 50050
2. 50060
3. 50070
4. 50030

▲ Hide Answer

Answer : d

19) SSH is the communication channel between data node and name node

1. Yes
2. False

▲ Hide Answer

Answer : a

17) As HDFS works on the principle of

1. Write once , Read Many
2. Write Many, Read Many
3. Write Many, Read Once
4. None

▲ Hide Answer

Answer : a

20) Reading is parallel and writing is not parallel in HDFS

1. True
2. False

▲ Hide Answer

Answer : a

21) ___ command to check for various inconsistencies in HDFS

1. FSCK
2. FETCHDT
3. SAFEMODE
4. SAFEANDRECOVERY

▲ Hide Answer

Answer : a

22) Hive provides

1. SQL
2. HQL
3. PL/SQL
4. PL/HQL

▲ Hide Answer

Answer : B

23) HQL Stands for ?

1. Hibernate Query Language
2. Historical Query Language
3. Health Query Language
4. Hive Query Language

▲ Hide Answer

Answer : D

24) HIVE is _____

1. A data mart on hadoop
2. A dataware house on hadoop
3. a database on hadoop
4. None

▲ Hide Answer

Answer : B

25) HQL allows _____ programmers

1. C# programmers
2. Java programmers
3. Map-reduce programmers
4. python programmers

▲ Hide Answer

Answer : C

26) Hive data is organized into

1. Databases
2. Tables
3. Buckets/Clusters
4. All of the above

▲ Hide Answer

Answer : d

27) HQL has the statements

1. DDL,DCL
2. DML,TCL
3. DML,DDL
4. DCL,TCL

▲ Hide Answer

Answer : c

28) The Decimal datatype has _____precision in hive

1. 4
2. 8
3. 16
4. N/A

▲ Hide Answer

Answer : d

29) How many bytes takes TINYINT in hive

1. 1
2. 2
3. 4
4. 8

▲ Hide Answer

Answer : a

30) regexp_replace('sairam','ai|am') output is

1. sai|ram
2. sai
3. sr
4. ram

▲ Hide Answer

Answer : c

31) If explicit conversion fails then cast operator returns

1. zero
2. one
3. FALSE
4. Null

▲ Hide Answer

Answer : d

32) which clause can be used to filter rows from a table in HQL

1. group by
2. order by
3. where
4. having

▲ Hide Answer

Answer : c

33) Which one of the following we can use to list the columns and all properties of a table

1. DESCRIBE EXTENDED table_name;
2. DESCRIBE table_name;
3. DESCRIBE PROPERTIES table_name;
4. DESCRIBE EXTENDED PROPERTIES table_name;

▲ Hide Answer

Answer : a

34) Which clause can be used to restricts the query to a fraction of the buckets in the table rather than the whole table:

1. SAMPLE
2. TABLESAMPLE
3. RESTICTTABLE
4. NONE

▲ Hide Answer

Answer : b

35) TABLESAMPLE syntax is

1. TABLESAMPLE(BUCKET x OUT OF(Y))
2. TABLESAMPLE(BUCKET x OUT OF y)
3. TABLESAMPLE(BUCKET x IN y)
4. TABLESAMPLE(BUCKET x IN(y))

▲ Hide Answer

Answer : b

36) How many total default number of dynamic partitions could be created by one DML in hive.exec.max.dynamic.partitions parameter

1. 10
2. 100
3. 1000
4. N/A

▲ Hide Answer

Answer : c

37) When using a derby database for a Metastore, how many client instances can connect to Hive?

1. 1
2. 10
3. Any
4. Cannot Say

▲ Hide Answer

Answer : a

38) In Hadoop 2.0, Name Node High Availability feature is present

1. TRUE
2. FALSE

▲ Hide Answer

Answer : a

39) Name Node is horizontally scalable due to the facility in Namenode Federation

1. TRUE
2. FALSE

▲ Hide Answer

Answer : a

40) How will you identify when was the last checkpoint done in a cluster

1. Using the Name Node Web UI
2. Using the Secondary Name Node UI
3. Using the `hadoop dfsadmin -report` command
4. Using the `hadoop fsck` command

▲ Hide Answer

Answer : b

41) Hadoop fsck command is used to

1. Check the integrity of the HDFS
2. Check the status of data nodes in the cluster
3. check the status of the NameNode in the cluster
4. Check the status of the Secondary Name Node

▲ Hide Answer

Answer : a

42) How can you determine available HDFS space in your cluster

1. using `hadoop dfsadmin -report` command
2. using `hadoop fsck /` command
3. using secondary namenode web UI
4. Using Data Node Web UI

▲ Hide Answer

Answer : a

43) A existing Hadoop cluster has 20 slave nodes with quad-core CPUs and 24TB of hard drive space each. You plan to add 5 new slave nodes. How much disk space can your new nodes contain?

1. New nodes may have any amount of hard drive space
2. New nodes must have at least 24TB of hard drive space
3. New nodes must have exactly 24TB of hard drive space
4. New nodes must not have more than 24TB of hard drive space

▲ Hide Answer

Answer : b

44) Which is a recommended configuration of disk drives for a DataNode

1. 10 1TB disk drives in a RAID configuration
2. 10 2TB disk drives in a JBOD configuration
3. One 3TB disk drive
4. 48 2TB disk drives in a RAID configuration

▲ Hide Answer

Answer : b

45) How does the HDFS architecture provide data reliability?

1. Reliance on SAN devices as a DataNode interface.
2. Storing multiple replicas of data blocks on different DataNodes
3. DataNodes make copies of their data blocks, and put them on different local disks.
4. Reliance on RAID on each DataNode.

▲ Hide Answer

Answer : b

46) Hcatalog has APIs to connect to HBase

1. TRUE
2. FALSE

▲ Hide Answer

Answer : b

47) The path in which the HDFS data will be stored is specified in the following file

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

▲ Hide Answer

Answer : a

48) To access to a Web user interface for a specific daemon requires which details

1. The setting for dfs.http.address for the NameNode
2. The IP address or DNS/hostname of the NameNode in the cluster
3. The SSL password used to log in to the Hadoop Admin Console
4. The server IP address or DNS/hostname where the daemon is running and the TCP/IP port

▲ Hide Answer

Answer : d

49) What is the default partitioning machanisim?

1. Round Robin
2. User needs to configure
3. Hash Partitioning
4. None

▲ Hide Answer

Answer : c

50) Is it possible to change the HDFS block size

1. TRUE
2. FALSE

▲ Hide Answer

Answer : a

51) Name Node contain

1. Meta data, all data blocks
2. Metadata and recently used block
3. Meta data only
4. None of the above

▲ Hide Answer

Answer : c

52) What is variety means to Big Data

1. Related data from different source in different formats
2. Unrelated data from different source.

▲ Hide Answer

Answer : a

53) Where do you specify the HDFS file system and host location

1. hdsf-site.xml
2. core-site.xml
3. mapred-site.xml
4. hive-site.xml

▲ Hide Answer

Answer : b

54) Which file do you use to configure Job Tracker

1. core-site.xml
2. mapred-site.xml
3. hdfs-site.xml
4. job-tracker.xml

▲ Hide Answer

Answer : b

55) Which file is used to define worker nodes

1. core-site.xml
2. mapred-site.xml
3. master-slave.xml
4. None

▲ Hide Answer

Answer : d

56) Name Node can be formatted any time without data loss

1. TRUE
2. FALSE

▲ Hide Answer

Answer : b

57) How do you list the files in a HDFS directory

1. ls
2. hadoop ls
3. hadoop fs -ls
4. hadoop ls -fs

▲ Hide Answer

Answer : c

58) Formatting the Name Node first time will result in

1. Formats the Name Node disk
2. Cleans the HDFS data directory
3. Just creates the directory structure on the Data Node machine
4. None of the above

▲ Hide Answer

Answer : d

59) What create the empty directory structure on Name Node

1. Configure in hdfs-site.xml
2. start the Name Node demon
3. Format the Name Node
4. None of the above

▲ Hide Answer

Answer : c

60) Hadoop answer to Big Data challenge

1. Job Tracker and Name Node
2. Name Node and Data Node
3. Data blocks, keys and value
4. HDFS and MapReduce

▲ Hide Answer

Answer : d

61) HDFS Achieves High Availability and fault tolerance through

1. By splitting files into blocks
2. By keeping a copy of frequently accessing data block in Name Node
3. By replicating any blocks on multiple data node on the cluster
4. None of the above

▲ Hide Answer

Answer : c

62) Name node keeps metadata and data files

1. TRUE
2. FALSE

▲ Hide Answer

Answer : b

63) Big Data poses challenge to traditional system in terms of

1. Network bandwidth
2. Operating system
3. Storage and processing
4. None of the above

▲ Hide Answer

Answer : c

64) What is the function of Secondary Name Node

1. Backup to Name Node
2. Helps Name Node in merging fsimage and edit
3. When Name node is busy, it servers the request for the file system
4. None of the above

▲ Hide Answer

Answer : b

65) Hadoop data types are optimized for

1. Data processing
2. Encryption
3. Compression
4. Network transmissions

▲ Hide Answer

Answer : d

66) HCatalog uses hive metastore for schema operations

1. TRUE
2. FALSE

▲ Hide Answer

Answer : a

67) A HDFS file can be executed

1. TRUE
2. FALSE

▲ Hide Answer

Answer : b