1)Hadoop provides

Hadoop is an open-source software framework for storing data and running applications on clusters of commodity hardware. It provides massive storage for any kind of data, enormous processing power and the ability to handle virtually limitless concurrent tasks or jobs.

2)What is Hadoop?

Apache Hadoop is a collection of open-source software utilities that facilitates using a network of many computers to solve problems involving massive amounts of data and computation. It provides a software framework for distributed storage and processing of big data using the MapReduce programming model

3)Hadoop block size should be multiple of which unit?

A 'block' is the minimum amount of data that can be read or written. In HDFS, the default block size is 64 MB as contrast to the block size of 8192 bytes in Unix/Linux. Files in HDFS are broken down into block-sized chunks, which are stored as independent units.

4)Which component of the Hadoop cluster manages data on slave nodes?

Ans) Name Node is the master node in the Apache Hadoop HDFS Architecture that maintains and manages the blocks present on the Data Nodes (slave nodes). Name Node is a very highly available server that manages the File System Namespace and controls access to files by clients

5)Hadoop block size should be multiple of which unit?

Ans) The default size of the HDFS data block is 128 MB.

This results in the transfer of multiple blocks at the disk transfer rate. If blocks are small, there will be too many blocks in Hadoop HDFS and thus too much metadata to store

6)The functions performed by Data Nodes in Hadoop Cluster is/are?

Ans) Data Nodes sends information to the Name Node about the files and blocks stored in that node and responds to the Name Node for all filesystem operations

7). If a file size of size 100 MB is stored on HDFS, what would be the split size?

Ans) If you have not defined any input split size in Map/Reduce program then default HDFS block split will be considered as input split.

8)Which component of Hadoop Cluster manages the file system namespace and regulates access

Ans) The HDFS cluster's Name Node is the primary server that manages the file system namespace and controls client access to files. As the central component of the Hadoop Distributed File System, the Name Node maintains and manages the file system namespace and provides clients with the right access permissions.

9) Which component in Hadoop Cluster is responsible for serving read and write requests from

Ans) The Data Nodes are responsible for serving read and write requests from the file system's clients. The Data Nodes also perform block creation, deletion, and replication upon instruction from the

8) Which component of Hadoop Cluster manages the file system namespace and regulates access

Ans) The HDFS cluster's Name Node is the primary server that manages the file system namespace and controls client access to files. As the central component of the Hadoop Distributed File System, the Name Node maintains and manages the file system namespace and provides clients with the right access permissions.

9) Which component in Hadoop Cluster is responsible for serving read and write requests from

Ans) The Data Nodes are responsible for serving read and write requests from the file system's clients. The Data Nodes also perform block creation, deletion, and replication upon instruction from the Name Node.

10)Which Eco system component of Hadoop is good for non sql programmers?

Ans) Spark can easily coexist with MapReduce and with other ecosystem components that perform other tasks. Spark is also popular because it supports SQL, which helps overcome a shortcoming in core Hadoop technology. The Spark programming environment works interactively with Scala, Python, and R shells.

11) Which command of HDFS helps copy files from HDFS to Local file system?

Ans) You can use the put command in the HDFS. This command is used to copy files from the HDFS file system to the local file system, just the opposite to put command.

12). State true or false: MR2 support various MPP modes for data processing?

Ans) Big Data goes hand-in-hand with Hadoop + MapReduce. But MPP (Massively Parallel Processing) and data warehouse appliances are Big Data technologies too. The MapReduce and MPP worlds have been pretty separate, but are now starting to collide.

13). Which component of Hadoop Cluster manages the file system namespace and regulates access

Ans) The HDFS cluster's Name Node is the primary server that manages the file system namespace and controls client access to files. As the central component of the Hadoop Distributed File System, the Name Node maintains and manages the file system namespace and provides clients with the right access permissions.

14)Which component in Hadoop Cluster is responsible for serving read and write requests from the file system's clients?

Ans) The Data Nodes are responsible for serving read and write requests from the file system's clients. The Data Nodes also perform block creation, deletion, and replication upon instruction from the Name Node.

15)What is the default replication factor of Hadoop cluster?

Ans) default the Replication Factor for Hadoop is set to 3 which can be configured means you can change it Manually as per your requirement like in above

10) Which Eco system component of Hadoop is good for non sql programmers?

Ans) Spark can easily coexist with MapReduce and with other ecosystem components that perform other tasks. Spark is also popular because it supports SQL, which helps overcome a shortcoming in core Hadoop technology. The Spark programming environment works interactively with Scala, Python, and R shells.

11) Which command of HDFS helps copy files from HDFS to Local file system?

Ans) You can use the put command in the HDFS. This command is used to copy files from the HDFS file system to the local file system, just the opposite to put command.

12). State true or false: MR2 support various MPP modes for data processing?

Ans) Big Data goes hand-in-hand with Hadoop + MapReduce. But MPP (Massively Parallel Processing) and data warehouse appliances are Big Data technologies too. The MapReduce and MPP worlds have been pretty separate, but are now starting to collide.

13). Which component of Hadoop Cluster manages the file system namespace and regulates access

Ans) The HDFS cluster's Name Node is the primary server that manages the file system namespace and controls client access to files. As the central component of the Hadoop Distributed File System, the Name Node maintains and manages the file system namespace and provides clients with the right access permissions.

14) Which component in Hadoop Cluster is responsible for serving read and write requests from the file system's clients?

Ans) The Data Nodes are responsible for serving read and write requests from the file system's clients. The Data Nodes also perform block creation, deletion, and replication upon instruction from the Name Node.

15) What is the default replication factor of Hadoop cluster?

Ans) default the Replication Factor for Hadoop is set to 3 which can be configured means you can change it Manually as per your requirement like in above example, we have made 4 file blocks which means that 3 Replica or copy of each file block is made means total of 4×3 = 12 blocks are made for the backup purpose.