Review: BDAT - Tutorial 4 - HDFS

Respondent

99 Dr. V. Sivakumar

41:40
Time to complete

0/24
Points

Needs review

Score / 2 pts
Needs review

1. What is a Namenode? *

Name node is a master node which stores meta data.

Needs review

Score / 2 pts
Needs review

2. Can Namenode be a commodity hardware? Why? *

No. Name Node can never be a commodity hardware, because the entire HDFS relies on it. Hence, Name node has to be a high available machine.

Needs review

Score / 2 pts
Needs review

3. What is a heartbeat in HDFS? *

A heartbeat is a signal sent by data nodes to name node (needs to be sent every 3 seconds by default) thereby indicating that data nodes are alive in the cluster. If name node does not receive any heartbeat from any of the data nodes, then it will assume/understand that there is an issue with that data node in terms of connection to the cluster and thereby not able to perform any task.

Needs review

Score / 2 pts
Needs review

4. If NameNode doesn't have any data, what would happen? *

If name node doesn't have any data it means that it cannot be a part of cluster anymore.

Needs review

Score / 2 pts
Needs review

5. What do you mean by hadoop daemon? *

It is a process or service that runs in the background (doesn't have physical appearance)

Needs review

Score / 2 pts
Needs review

6. What is a rack in HDFS? *

It is the storage location where all the data nodes are put together. Thus, it is a physical location of data nodes stored in single location.

Needs review

Score / 2 pts
Needs review

7. When a data node gets crashed or failure, how does name node would manage it? *

If there is a crash with data node, then the jobs running in that data node would be automatically redirected by the name node to one of its replica (which has the similar data) thereby retrieving all the operations being done.

Needs review

Score / 2 pts
Needs review

8. What is block placement strategy/Replica Placement Policy? *

Block placement/Replica placement strategy is storing the replica of data nodes in racks. The strategy of this placement is storing one block in one rack and the rest two blocks in another rack.

Needs review

Score / 2 pts
Needs review

9. What is Rack Awareness in Hadoop? *

Rack awareness is the storage information of blocks being put up in the respective racks. This rack awareness is maintained by the name node.

Needs review

Score / 2 pts
Needs review

10. What is the default block size in Hadoop 1 and in Hadoop 2? *

Hadoop 1 - 64 MB
Hadoop 2 - 128 MB

Needs review

Score / 2 pts
Needs review

11. Is it advisable to use HDFS for storing multiple small files? *

Not advisable. Name node is an expensive high available machine and hence using name node for storing the meta data of all small files is not advisable. Otherwise needless metadata would be generated for each of those small files. Rather if we have huge file with lots of data, then using of HDFS is a good idea.

Needs review

Score / 2 pts
Needs review

12. What are the basic differences between relational database and HDFS? *

Relational database - Structured data, Scheme on write validation and Reading is fast.
HDFS - Both structured and unstructured , No schema on write validation and hence writes are fast.