

HDFS

1. What are all the factors we need to go for Big data? Explain with example.
2. Why we need hadoop even we have plenty of tools in market for processing Big Data?
3. What is data locality in hadoop?
4. What is the version of Hadoop you are using?
5. Replication causes data redundancy then why is it still preferred in HDFS?
6. What is fsimage and editlogs in HDFS?
7. What is the purpose of Secondary Name Node?
8. What is checkpoint name node and backup node in Hadoop?
9. What do you mean by the High Availability of a Name node? How is it achieved?
10. What is the difference between HDFS1 and HDFS2?
11. What is the difference between an "HDFS Block" and an "Input Split?"
12. What is edge node in hadoop?
13. What is inode in hadoop?
14. What is Safe mode? How to come out of Safe mode in Hadoop?
15. What do you understand by Active and Passive Name node?
16. What is Data Integrity in HDFS? How does HDFS ensure data integrity of data blocks stored in HDFS?
17. What are all the conditions HDFS cluster will follow while replicating data?
18. Is it a good practice to use HDFS for multiple small files?
19. What happens if two clients try writing into the same HDFS file?
20. What does the HDFS error "File could only be replicated to 0 nodes, instead of 1" mean?
21. What is a rack awareness and on what basis is data stored in a rack?
22. What is the difference between HDFS and GFS?