. how fsimage and edit logs works in Hadoop

<http://www.bidwbooks.com/understanding-hdfs-architecture-part-2-2/>

Why a block size in HDFS is large?

We all know that transferring a large file made of multiple blocks operates at the disk transfer rate i.e. let consider 100mbps. Also we know that average seek time of a Hard Disk is 10ms.

Hence, Transfer time = Size of a block / Transfer rate.

Let consider in HDFS a block size is 128 MB then transfer time = 128 / 100mbps i.e. 1.28 sec.

Therefore total time taken to transfer 128 MB block = seek time + transfer time i.e. (10/1000) sec + 1.28 sec i.e. about 1.29 sec.

Suppose if we make 10 blocks instead of 1 block then seek time for all 10 blocks = 10 \* (10 / 1000) sec i.e. 0.1 sec. hence time taken to dump all blocks could be 1.39 secs. Apart from this we come across lot of other performance latency issues like…

1. If block size is less, its difficult to maintain seek time as 1% of total data transfer time. Hence block size should be large which means it takes more time to transfer all data then its easy to maintain seek time as 1% of total transfer time for continuous data stream access pattern.
2. If blocks are more with small size it restricts the NN namespace in memory space (i.e. each metadata line entry for each block file in namespace is about 150 bytes)
3. Also its difficult to maintain seek time consistency due to track seek, entire disc seek or cylinder rotation time for many small size block chunks.

* HDFS commands to get storage capacity of Data Node and how many blocks can be stored in disk
* HDFS commands to get file, blocks and their location
* How it’s possible to get over-replicated blocks for a HDFS file
* To get the information about the blocks being replicated or waiting to be replicated using hdfs dfsadmin -metasave
* To get misreplicated blocks scenarios using command line hdfs fsck…we need to change the replication factor as well as have file size of more than 200 MB in single node cluster
* Can I say Corrupted blocks as Missing replicas or what is exact diff between these two…
* Get the Datanode block scanner using DN Web interface i.e. <http://localhost:50075/blockScannerReport> along with listblocks parameter i.e. <http://localhost:50075/blockScannerReport?listblocks>

Date : 5th October 2015

Q) How to check whether two files in HDFS have same contents.

🡪 Use “Hadoop fs –checksum” command or distcp command

Q) Usage of “Hadoop dfsadmin –saveNamespace” command

🡪 This helps admin to run manually when NN is in safemode for roll the in progress edits file so that all the edit transaction can tracked in new file.

Q) How do load namenode metadata from Secondary NameNode to Primary NN.

🡪 By using the -importCheckpoint option when starting the namenode daemon.

The -importCheckpoint option will load the namenode metadata from the latest

checkpoint in the directory defined by the dfs.namenode.checkpoint.dir property, but

only if there is no metadata in the dfs.namenode.name.dir directory, to ensure that there

is no risk of overwriting precious metadata.

Q) When does NN creates its persistent storage directory?

🡪 When we format the NN first time. But this is not required by DataNode.

Q) About Audit logging and how to enable it?

Date: 6th October 2015