**1.**Because, this information is reconstructed from DataNodes when the system starts. NN builds it dynamically with the help of **block reports** sent by DNs.

## 2.Because, without the NameNode, the filesystem can’t be used. In fact, if the machine running the NameNode were obliterated(destroyed), all the files on the filesystem would be lost since there would be no way of knowing how to reconstruct the files from the blocks on the datanodes.

**3.**An fsimage file contains the complete state of the file system at a point in time. Every file system modification is assigned a unique, monotonically increasing transaction ID. An fsimage file represents the file system state after all modifications up to a specific transaction ID.

**4**.It merges the fsimage and the editlog files periodically and keeps editlog size within a limit. This copy of the merged namespace image can be used in the event of the NameNode failing.

5.No, the **NameNode** leaves **Safemode** automatically after DataNodes report that most file system blocks are available. Depending on how the Safemode parameters are configured, the NameNode stays in Safemode until a specific percentage of system blocks are minimally replicated.

6. difference between Hadoop 1 vs Hadoop 2 are

a. Hadoop 1 has one namenode but Hadoop 2 has multiple namenode.

b. haddop 1 has single point failure but Hadoop 2 has active or passive (if active namenode is failed , the passive namenode generates automitically

c. Hadoop 1 use only java lanuages but Hadoop 2 beyond java language

d. Hadoop 1.x supports 4000 nodes per cluster but Hadoop 2.x supports more than 10,000 nodes per cluster.

7. map reduce 1 has two types of daemons that control the job execution process that are JobTracker(coordinates all the jobs runs on the system by scheduling tasks to run on TaskTrackers) and one or more TaskTrackers. But map reduce 2 has types of daemons that control the resource manager and oner or more node manager.

8.HDFS federation is partitions the filesystem namespace over multiple separated namenodes each of which manages a portion of the filesystem namespace. The advantage is that allows a cluster to scale by adding Nodename and also removes tight coupling of block storage and namespace.

9.The high availability (HA) feature in Hadoop 2 addresses the node name SPOF(single point of failure) problem by providing the option of running two redundant nameNodes in the same cluster in an Active/Passive configuration with a hot standby. It achieves when two separate machines are configured as namenode. At any point in time ,exactly one of the namenodes is on an active state and the other is in standby state.

10.The role of application master is to coordinates the execution of all tasks within its application and asks for appropriate resource containers to run tasks.

**Mussie Ghebregziabher**

**Id: 108445**