Q1.Explain what is High availability of Namenode?

A1

1.There is a pair of Name Nodesin an active-standby configuration.

2.In the event of failure an active NameNode, the standby takes over its duties without a significant interruption.

**Architectural Changes:**

a)The NameNodes must use highly-available shared storage to share the edit log.Edit logs are read by the StandbyNameNode

when it takes the responsibility of ActiveNameNode.

b)DataNodes must send block reports to both the NameNodes because of the block mappings.

c)The secondary NameNode’s role is subsumed by the StandBy NameNode.StandbyNameNode takes periodic checkpoints of the

active namenode.

d)Checkpointing is done by the StandBy NameNode .



Each DataNodereports block stored by it to both the NameNodes.

Q2.Explain what is check pointing and how it is useful?

A2

The Checkpoint node periodically creates checkpoints of the namespace. It downloads fsimage and edits from the active NameNode, merges them locally, and uploads the new image back to the active NameNode. The Checkpoint node usually runs on a different machine than the NameNode since its memory requirements are on the same order as the NameNode. The Checkpoint node is started by bin/hdfs namenode -checkpoint on the node specified in the configuration file.

The Checkpoint node stores the latest checkpoint in a directory that is structured the same as the NameNode's directory. This allows the checkpointed image to be always available for reading by the NameNode if necessary. See Import checkpoint.



Q3.Explain what is HDFS federation?

A3

1. Let HDFS contain two sub-directories in the root directory.

/usr

/share

2) In the HDFS Federation, there are multiple Name Nodes, each storing the metadata and block mapping of files and directories contained in particular sub-directories.

3) The list of sub-directories managed by a Name Node is called a namespace volume.

4) Blocks for files belonging to a namespace is called a block pool.

For example, here we can have two name nodes, one for storing the metadata and block mapping for namespace volume /usr and one for /share.

5) This way, if one Name Node fails, a namespace volume managed by another name node is still accessible. So the entire cluster doesn’t become inaccessible.

6) Let,

a) Sub-directories and files inside /usr constitute NameSpaceVolume1

b) Sub-directories and files inside /share constitute Name Space Volume 2

7) One Data Node can contain blocks of different namespace volumes

8) So, namespace volumes are divided among Name Nodes, but not among Data Nodes.



Q4What are the configuration files that are to be edited for sure while installing a hadoop cluster?

A4



**Settings that need to be done in Core-site.xml**

Some of the important properties are:

* Configuring the name node address
* Configuring the rack awareness factor
* Selecting the type of security

### ****Settings To Be Done In HDFS-site.xml****

The properties inside this xml file deals with storage procedure inside HDFS of Hadoop. Some of the important properties are:

* Configure port access
* Manages ssl client authentication
* Controls Network interface
* Changes file permission

**Settings In yarn-site.xml**

* WebAppProxy Configuration
* MapReduce Configuration
* NodeManager Configuration
* ResourceManager Configuration
* IPC Configuration

**mapred-site.xml**

* Node health script variables
* Proxy Configuration
* Job Notification Configuration