Q1. Define fsck command ?

Ans 1. Fsck stands for file system check. It tells the health of a system. It shows the number of files.

Ex- $hadoop fsck /

When executed on terminal, the result is as shown below.

.Status: HEALTHY

Total size: 1689130465 B

Total dirs: 137

Total files: 141 (Files currently being written: 1)

Total blocks (validated): 148 (avg. block size 11413043 B) (Total open file blocks (not validated): 1)

Minimally replicated blocks: 148 (100.0 %)

Over-replicated blocks: 0 (0.0 %)

Under-replicated blocks: 7 (4.7297297 %)

Mis-replicated blocks: 0 (0.0 %)

Default replication factor: 1

Average block replication: 1.0

Corrupt blocks: 0

Missing replicas: 21 (14.189189 %)

Number of data-nodes: 1

Number of racks: 1

FSCK ended at Tue Jun 21 10:25:49 PDT 2016 in 65 milliseconds

Q2. How to move a file in hdfs ?

Ans 2. bin/hadoop dfs -mkdir /hdfs

bin/hadoop dfs -copyFromLocal D:/eclipse/cygwin/home/313159/hadoop-0.19.1/dft/StartUnit.txt /hdfs/StartUnit.txt

Q3. Difference between local file system and hdfs ?

Ans 3. When you install hadoop in pseudo distributed mode, all the HDFS daemons namdenode, datanode and secondary name node run on the same machine. The temp dir which you configure is where a data node stores the data. So when you look at it from HDFS point of view, your data is still stored in block and read in blocks which are much bigger (and aggregation) on multiple file system level blocks. When you configure Hadoop it lays down a virtual FS on top of your local FS, which is the HDFS. HDFS stores data as blocks(similar to the local FS, but much much bigger as compared to it) in a replicated fashion. But the HDFS directory tree or the filesystem namespace is identical to that of local FS. When you start writing data into HDFS, it eventually gets written onto the local FS only, but you can't see it there directly.

Q4. Difference between root & local root ?

Ans 4. 'root' is traditionally the name given to the user account with superuser level rights. Root is generic where hdfs file can't be stored in other users. On the other hand, local root are nothing but the superuser which possess administration right & hence it is password oriented. In laymen language, there is one generic root & many hdfs missions or the local root or the superusers may exist.

Ex- training is a super user & local root with password training.

If we use forward slash '/' , then its the root & no hdfs file cant be saved in root.

Q5. What will happen when you will move a file from lfs to hdfs ?

Ans 5. n the background, the source file is split into HDFS blocks, the size of which is configurable (commonly 128 MB, 64 MB by default). For fault tolerance, each block is automatically replicated by HDFS. By default, three copies of each block are written to three different DataNodes. The replication factor is user-configurable (default is three). The DataNodes are servers which are physical machines or virtual machines/cloud instances. DataNodes form the Hadoop cluster into which you write your data and on which you run your MapReduce/Hive/Pig/Impala/Mahout/etc. programs.

The DataNodes are the workers of the Hadoop cluster, the NameNodes are the masters.

When a file is to be written into HDFS, the client writing the file obtains from the NameNode a list of DataNodes that can host replicas of the first block of the file.

The client arranges a pipeline through which all bytes of data from the first block of the source file will be transmitted to all participating DataNodes. The pipeline is formed from client to first DataNode to second DataNode to final (third in our case) DataNode. The data is split into packets for transmission, and each packet is tracked until all DataNodes return acks to indicate successful replication of the data. The packets are streamed to the first DataNode in the pipeline, which stores the packet and forwards it to the second DataNode, and so on. If one or more replications fail, the infrastructure automatically constructs a new pipeline and retries the copy.

When all three DataNodes confirm successful replication, the client will advance to the next block, again request a list of host DataNodes from the NameNode, and construct a new pipeline. This process is followed until all blocks have been copied into HDFS. The final block written may be smaller than the configured block size, but all blocks from the first to the penultimate block will be of the configured block size.

Ex - hadoop fs -put a.txt /user/hive/warehouse/sample\_db/sample\_table/

Q6. How to change replication factor ?

Ans 6. To set replication of an individual file to 4:

./bin/hadoop dfs -setrep -w 4 /path/to/file

You can also do this recursively. To change replication of entire HDFS to 1:

./bin/hadoop dfs -setrep -R -w 1 /

Q7. How to run jar file ?

Ans 7. Assume you have to run a jar file named worldcount.

Follow the steps in new terminal.

$cd NewData2 /

Choose Project

>Examples

>world Count

>Now create a class

>Edit the first line, add class say name.com

>Create a package  
>Now add the two jar files

>hadoop-cli

>hadoop.jar in hadop drop down list

> Check if the code is error free

$hadoop jar wc.jar name.com Word Count /Omlet.txt /op