FILEMANAGEMENTIN HADOOP

OBJECTIVE:

Implementthefollowing filemanagementtasksinHadoop:

i.   Addingfilesanddirectories
ii. Retrievingfiles
iii. Deletingfiles

Hint:AtypicalHadoopworkflowcreatesdatafiles (suchas log files)elsewhereandcopiestheminto HDFS using one of the above command line utilities.

RESOURCES:

VMWarestack,4GBRAM, HardDisk80GB.

PROGRAMLOGIC:

AddingFilesandDirectoriesto HDFS

Before you can run Hadoop programs on data stored in HDFS, you,,ll need to put the data into HDFS first. Let,,screate a directoryand put afile in it. HDFS hasa default working directoryof

/user/$USER, where $USER is your login user name. This directory isn,,t automatically created for you, though, so let,,s create it with the mkdir command. For the purpose of illustration, we use chuck. You should substitute your user name in the example commands.

hadoopfs-mkdir/user/chuck

　　　　hadoop fs-put

hadoopfs-putexample.txt/user/chuck

RetrievingFilesfrom HDFS

TheHadoopcommandget copiesfilesfromHDFSbacktothelocalfilesystem.Toretrieveexample.txt, we can run the following command:
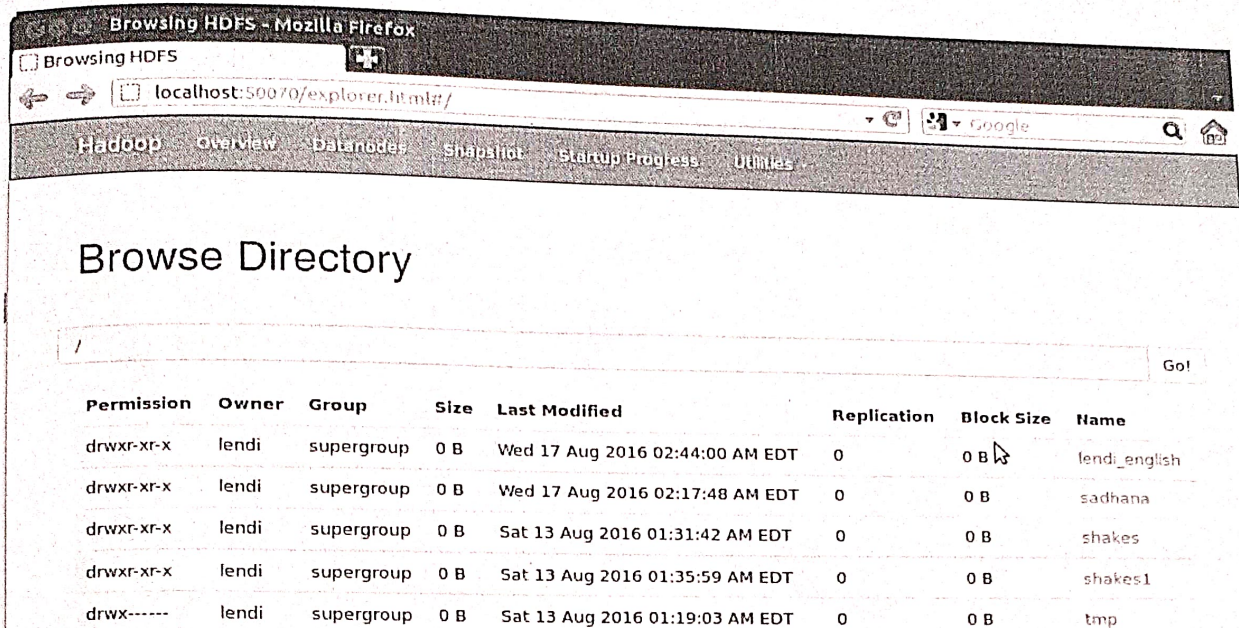
hadoopfs-catexample.txt

Deleting files from HDFS

hadoop fs -rmexample.txt

- Commandforcreatingadirectoryinhdfsis"hdfsdfs–mkdir/lendicse".
- Addingdirectoryisdonethroughthecommand"hdfsdfs–putlendi_english/".

INPUT/OUTPUT:

Browse Directory

/   Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	lendi	supergroup	0 B	Wed 17 Aug 2016 02:44:00 AM EDT	0	0 B	lendi_english
drwxr-xr-x	lendi	supergroup	0 B	Wed 17 Aug 2016 02:17:48 AM EDT	0	0 B	sadhana
drwxr-xr-x	lendi	supergroup	0 B	Sat 13 Aug 2016 01:31:42 AM EDT	0	0 B	shakes
drwxr-xr-x	lendi	supergroup	0 B	Sat 13 Aug 2016 01:35:59 AM EDT	0	0 B	shakes1
drwx------	lendi	supergroup	0 B	Sat 13 Aug 2016 01:19:03 AM EDT	0	0 B	tmp

PRELABVIVAQUESTIONS:
1) DefineHadoop?
2) ListoutthevarioususecasesofHadoop?

LABASSIGNMENT
1) WhatiscommandusedtolistoutdirectoriesofDataNodethroughwebtool

POSTLABVIVAQUESTIONS:

1. DistinguishtheHadoopEcosystem?
2. DemonstratedivideandconquerphilosophyinHadoopCluster?