→ su : switch user
→ mkdir : creating a directory
→ nano, vi editors : for editing any file
→ cd / : It will move to the root folder
→ pwd : present working directory
→ whoami : to find the current user
→ ls : lists out folders
→ ll : long list with permissions
→ d; belongs to directory
→ l : linking

Installation

→ nano ~/.bashrc

① sudo apt-get remove ssh
sudo apt-get remove pdsh

Configure Password-less SSH

sudo apt install openssh-server openssh-client -y
ssh

5/3/24

1) In hadoop ecosystem files are always in read only mode so, you can't write in hadoop ecosystem. If you want to edit a file then you need to import that file into LFS (Local File System) or any storage other than hadoop

2) If you are uploading a file to hadoop you can use either copy from Local or put commands.

3) If you want to download a file or export a file from hadoop you can use either copy to local or get commands.

Note:

→ If you are using copyFromLocal or put command copy of a file exists in both ecosystems.
1) LFS 2) HDFS.

→ If you don't want to copy in LFS then you can use move from local

4)
→ In hadoop ecosystem copy are used to copy between LFS to LFS, HDFS to HDFS only

5) DU: It will give the size of each file in directory.
SU: Switching user

6) DUS: This command will give total size of directory or a file.

7) STAT: It will give the last modified time of a directory or path.

8) SETREP: This command is used to change the replication factor of a file or a directly in HDFS only.

Example: hadoop fs -setrep -R -w any_number
space filename or directory

Note:
→ -w means wait till replication is completed.
→ -R means recursively

Commands
1) localhost:50070
2) su

→ for copying:
cp old file-name new-filename

→ rename
mv old-name new_name

→ remove:
file: rm filename
directory: rm -r directoryname

→ compress:
file: gzip filename

→ uncompress:
gunzip filename.gz

→ date

→ cal

→ touch filename

→ gedit filename.txt

→ du : shows the amt of space in bytes

du filename.log

→ df : shows the capacity used and free space

df samplenew.log

→ hostname

quickstart.cloudera

→ To know the configuration:

ifconfig

HDFS commands:

1) hadoop

2) creating a new directory in HDFS:

```
hadoop fs -mkdir /bda-6717
```

3) To list the directories in HDFS under /:

```
hadoop fs -ls /
```

4) If we create the directory by the following way it will create the dir under default path /user/root/

```
hadoop fs -mkdir bda-6717
hadoop fs -ls /user/root
```

5) To load the data from local file system to HDFS

```
hadoop fs -put file.txt /bda-6717
```

6) To see the files & directories which is located under a particular directory in HDFS

```
hadoop fs -ls /bda-6717
```

7) To see the content of a file which is located under a particular directory in HDFS

```
hadoop fs -cat /bda-6717/file.txt
```

8) moveFromLocal

to move the data from LFS to HDFS

hadoop fs -moveFromLocal samplenew.log /bda_6717

9) copyFromLocal

To copy the data from LFS to HDFS

hadoop fs- copyFromLocal dummy.txt / bda_6717

hadoop fs -ls / bda-6717

10) hadoop fs -get /bda_6717/samplenew.log /home/ bda_6717_LFS

11) To remove a file under the HDFS directory

hadoop fs -rm /bda-6717/samplenew.log

12) hadoop fs -rmr /CiscoOut1

13) To copy the file btw 2 HDFS directories

fs -mkdir / bda-6717-2

fs -cp /bda_6717 / file.txt / bda_6717_2

fs-ls /bda-6717-2

14) To move a file from one dir to another in HDFS

hadoop fs -mv / bda_6717_HDFS/dummy.txt /bda_6717 -HDFS2

hadoop fs -ls /bda_6717_HDFS2

15) To know the safemode status

hadoop dfsadmin -safemode get

16) To leave the safemode

hadoop dfsadmin -safemode leave

17) To enter into the safemode

hadoop dfsadmin -safemode enter

18) To know the cluster health we use the following command

hadoop dfsadmin -report

19) We can create a dummy file in the HDFS by using

hadoop fs -touchz /bda_6717_hdfs/zerobytefile

20) To access the namenode in GUI

localhost : 50070

21) To see the hadoop configuration files

PATH : /usr/lib/hadoop/conf

22) To access job tracker by using GUI

localhost : 50030

23) By typing JPS we can get all 5 deamons

JPS

Map Reduce Programming:

→ This requires 3 things:

1) Driver class
2) Mapper class
3) Reducer class

Driver class:

→ This class specifies job configuration details

Mapper class

→ This class overrides the map function based on the problem statement

Reducer class:

→ This class overrides the reduce function based on the problem statement

Driver program:

Note:

Open Eclipse

File → new → java project → Project name
BDA WordCount Sample

Package → src → BDA → source → new → class (create 3 classes)

add external jars → usr → lib

① /usr/lib/hadoop-0.20-mapreduce/hadoop-core-2.6.0-mr1-cdh5.13.0.jar

② /usr/lib/hadoop-common-2.6.0-cdh5.13.0.jar