Hadoop HDFS tutorial

- First you can display the usage for a given command or all commands using the help

hdfs dfs -help

- Also, you can display the directory information (either all directories or a specific directory)

  hdfs dfs –ls /

  hdfs dfs -ls /test

- You can create new directory

  hdfs dfs –mkdir /hadoop

  hdfs dfs –mkdir / data

  hdfs dfs –ls /

- You can upload a local file to a specified directory in HDFS
  example

  hdfs dfs -put file:///home/bigdata/Desktop/word.txt /hadoopor

  hdfs dfs -copyFromLocal file:///home/bigdata/Desktop/word.txt /hadoop

****

On your vm desktop create trial.txt

By right clicking=>create document>>empty file

Open your document and write down the following text

What is Big Data Analytics?

Big data analytics helps businesses to get insights from today's huge data resources. People, organizations, and machines now produce massive amounts of data. Social media, cloud applications, and machine sensor data are just some examples. Big data can be examined to see big data trends, opportunities, and risks, using big data analytics tools.

Big Data Basics

Until recently, data was mostly produced by people working in organizations. The data usually had a specific structure. It was the basis of records for money paid, deliveries made, employees hired, and so on. This data is still vital to businesses. Now, big data concepts mean that data processing must manage:

- High volume (lots of data)
- High velocity (data arriving at high speed)
- High variety (many different data sources and formats)

Big data can be structured, but with high volume, like historical payment transaction data. It can be semi-structured as in XML and other user-defined content. It can also be totally unstructured. Free form text used in social networks is an example.
Save your document

Then upload it to the hdfs directory named hadoop (create hadoop directory if not exist)

hdfs dfs -copyFromLocal /home/bigdata/Desktop/trial /hadoop

hdfs dfs -ls /hadoop

- now, you can download a specified file from HDFS to a local host

hdfs dfs -get /hadoop/trial /home/bigdata/Documents

hdfs dfs -copyToLocal /hadoop/trial /home/bigdata/Documents

- you can also download through the HDFS GUI

Browse the GUI through the following URL:

http://localhost:50070/

Then from Utilities menu choose browse the File system

You can choose any file in a directory and by clicking on the file name you'll have the option to download it on your machine

- If you 'd like to display the file contents , then use the following command

hdfs dfs -cat /hadoop /trial.txt

you can append your file if needed:

create a new document appendtxt.txt and insert the following text then save

Using Big Data Analytics
The more data you have, the more chance you have of getting useful insights from it. However, the size of big data usually makes it impossible to use manual or even conventional computing methods (learn more here: big data and Hadoop). Instead, big data analytics is based on:

- Data mining to sift through data to find patterns and relationships
- Statistical algorithms to build models and predict outcomes
- Machine learning to handle changing and new data, to adapt and enrich models
- Text analytics and natural language processing to analyze free form text and speech

Big data analytics tools can also be grouped like this:
Descriptive analytics to show what happened
Diagnostic analytics to explain why something happened
Predictive analytics to suggest what will likely happen next
Prescriptive analytics to tell users what to do, to obtain a given result

If you want to display the content of your append file use the command :cat appendtxt.txt

Then you can append your old file using the created appendtxt

hdfs dfs -appendToFile /home/bigdata/Desktop/appendtxt.txt /hadoop/trial

Display your document to check your process was done successfully

hdfs dfs -cat /hadoop/trial

The result will be as follows:

What is Big Data Analytics?
Big data analytics helps businesses to get insights from today's huge data resources. People, organizations, and machines now produce massive amounts of data. Social media, cloud applications, and machine sensor data are just some examples. Big data can be examined to see big data trends, opportunities, and risks, using big data analytics tools.
Big Data Basics
Until recently, data was mostly produced by people working in organizations. The data usually had a specific structure. It was the basis of records for money paid, deliveries made, employees hired, and so on. This data is still vital to businesses. Now, big data concepts mean that data processing must manage:

- High volume (lots of data)
- High velocity (data arriving at high speed)
- High variety (many different data sources and formats)

Big data can be structured, but with high volume, like historical payment transaction data. It can be semi-structured as in XML and other user-defined content. It can also be totally unstructured. Free form text used in social networks is an example.
Using Big Data Analytics
The more data you have, the more chance you have of getting useful insights from it. However, the size of big data usually makes it impossible to use manual or even conventional computing methods (learn more here: big data and Hadoop). Instead, big data analytics is based on:

- Data mining to sift through data to find patterns and relationships
- Statistical algorithms to build models and predict outcomes
- Machine learning to handle changing and new data, to adapt and enrich models
- Text analytics and natural language processing to analyze free form text and speech

Big data analytics tools can also be grouped like this:
Descriptive analytics to show what happened
Diagnostic analytics to explain why something happened
Predictive analytics to suggest what will likely happen next
Prescriptive analytics to tell users what to do, to obtain a given result
Well Done! The file was appended.

****

- Every directory and every document has its own permissions either read or write or execute. You can modify such permissions

  To check permissions, list the hdfs contents by using the list command

  hdfs dfs -ls /

you can modify the trial document permissions using the following command

hdfs dfs -chmod 777 /hadoop/trial.txt

Permissions will look like the following representation:

drwx------ 2 Ahmed Ahmed 2048 Jun 12 2012 private

This is a directory named "private", owned by user Ahmed and associated with Unix group Ahmed. The directory has read, write, and execute permissions for the owner, and no permissions for any other user.

-rw------- 2 Ahmed Ahmed 1327 Apr 9 2012 try.f90

This is a normal file named "try.f90", owned by user Ahmed and associated with group Ahmed. It is readable and writable by the owner, but is not accessible to any other user.

-rwx------ 2 Ahmed Ahmed 12040 Apr 9 2012 a.out

This is a normal file named "a.out", owned by user Ahmed and associated with group Ahmed. It is executable, as well as readable and writable, for the owner only.

drwxr-x--- 2 Rana bigdata 2048 Oct 17 2011 share

This is a directory named "share", owned by user Rana and associated with group bigdata. The owner can read and write the directory; all members of the file group bigdata can list the contents of the directory. Presumably, this directory would contain files that also have "group read" permissions.

drwxr-xr-x 3 Aya bigsci 2048 Nov 13 2011 public

This is a directory named "public", owned by user Aya and associated with group bigsci. The owner can read and write the directory; all other users can only read the contents of the directory. A directory such as this would most likely contain files that have "world read" permissions.

- The chmod ("change mode") command is used to change the permission flags on existing files. It can be applied recursively using the -R option. It can be invoked with either octal values representing the permission flags, or with symbolic representations of the flags. The octal values have the following meaning:

Octal Digit	Binary Representation (rwx)	Permission
0	000	none
1	001	execute only
2	010	write only
3	011	write and execute
4	100	read only
5	101	read and execute
6	110	read and write
7	111	read, write, and execute (full permissions)

To reassure the changes are activated list the directory content

hdfs dfs -ls /hadoop

- To check the available space of the file system

```
hdfs dfs -df
```

- To delete file or directory

```
hdfs dfs –rm /hadoop/trial.txt
hdfs dfs –rm -r  /hadoop
hdfs dfs  -getmerge  -nl  /datasource/trial.txt  /datasink/animals.txt  files.merge
```

**********GOOD JOB!*******

Zookeeper_Lab

#open the path
cd zookeeper
#start zookeeper process
./bin/zkServer.sh start
#stop zookeeper process
./bin/zkServer.sh stop
connect to the client
./bin/zkCli.sh -server 127.0.0.1:2181
[zk: 127.0.0.1:2181(CONNECTED) 0] help
#create a new znode by running
create /zk_test my_data
create /zk_test my_data
#verify that the data was associated with the znode by running the get command
get /zk_test
#We can change the data associated with zk_test by issuing the set command
set /zk_test junk
#Create Sub znode
create /zk_test/Child1 “firstchild”
#List Children
ls /zk_test
#Check Status
stat /zk_test

```
]zk: 127.0.0.1:2181(CONNECTED) 1]  create /zk_test my_data
Created /zk_test
]zk: 127.0.0.1:2181(CONNECTED) 2] get /zk_test
my_data
]zk: 127.0.0.1:2181(CONNECTED) 3] set /zk_test junk
]zk: 127.0.0.1:2181(CONNECTED) 4] create /zk_test/Child1 “firstchild”
Created /zk_test/Child1
]zk: 127.0.0.1:2181(CONNECTED) 5] ls /zk_test
]Child1[
]zk: 127.0.0.1:2181(CONNECTED) 6] stat /zk_test
cZxid = 0x17
ctime = Mon Mar 06 11:57:22 EET 2023
mZxid = 0x18
mtime = Mon Mar 06 11:58:00 EET 2023
pZxid = 0x19
cversion = 1
dataVersion = 1
aclVersion = 0
ephemeralOwner = 0x0
dataLength = 4
numChildren = 1
```

delete the node
delete /node
delete /zk_test
Delete a node and all subnodes : deleteall / zk_test
#Finally to exit from command line interface
Quit

```
[zk: 127.0.0.1:2181(CONNECTED) 8] delete /zk_test
Node not empty: /zk_test
[zk: 127.0.0.1:2181(CONNECTED) 10] deleteall /zk_test
```