#Author: Harish G

#Purpose: Making a documentation of my learning in Hadoop environment using linux terminal, exploring various cmds

Before starting this document, I want to thank my instructor Mr Ajay Singhal under whom, I am taking a course **Big Data Hadoop and Spark Developer** who motivated me to document my Hadoop learning in git.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Using a virtual lab environment Web Console- The Linux terminal of the system on which Hadoop is installed.

#changing the ip values here manually wherever present to a random value just as a precaution because it can help prevent any accidental sharing of any imp details or trespassing of my access to labs by 3rd parties

#to search my sub-folder name among the list of user sub-folders under user folder, in this hdfs dfs listing , will use (pipeline) and grep .i.e. | grep

[harishga\_buffalo@ip-10-0-1-99 ~]$ **hdfs dfs -ls /user | grep harishga\_**

drwxr-xr-x - harishga\_buffalo harishga\_buffalo 0 2020-02-02 13:26 /user/harishga\_buffalo

[harishga\_buffalo@ip-10-0-1-99 ~]$ hdfs dfs -ls /user/harishga\_buffalo

[harishga\_buffalo@ip-10-0-1-99 ~]$ hdfs dfs -ls /user/harishga\_buffalo/

# to create a directory : using mkdir command below

[harishga\_buffalo@ip-10-0-1-99 ~]$ **hdfs dfs -mkdir /user/harishga\_buffalo/1stfebdata2**

[harishga\_buffalo@ip-10-0-1-99 ~]$ hdfs dfs -ls /user/harishga\_buffalo/

Found 1 items

drwxr-xr-x - harishga\_buffalo harishga\_buffalo 0 2020-02-08 17:52 /user/harishga\_buffalo/1stfebdata2

[harishga\_buffalo@ip-10-0-1-99 ~]$ hdfs dfs -mkdir 1stfebdata

[harishga\_buffalo@ip-10-0-1-99 ~]$ hdfs dfs -ls /user/harishga\_buffalo/

Found 2 items

drwxr-xr-x - harishga\_buffalo harishga\_buffalo 0 2020-02-08 17:53 /user/harishga\_buffalo/1stfebdata

drwxr-xr-x - harishga\_buffalo harishga\_buffalo 0 2020-02-08 17:52 /user/harishga\_buffalo/1stfebdata2

[harishga\_buffalo@ip-10-0-1-99 ~]$ hdfs dfs -ls

Found 2 items

drwxr-xr-x - harishga\_buffalo harishga\_buffalo 0 2020-02-08 17:53 1stfebdata

drwxr-xr-x - harishga\_buffalo harishga\_buffalo 0 2020-02-08 17:52 1stfebdata2

[harishga\_buffalo@ip-10-0-1-99 ~]$

#command line utility for printing the current working directory

[harishga\_buffalo@ip-10-0-1-99 ~]$ **pwd**

/home/harishga\_buffalo

[harishga\_buffalo@ip-10-0-1-99 ~]$ ls

harishga\_my\_first\_hdfs\_cmds\_webconsole\_02\_08\_2020.txt

#using the following copyFromLocal we can copy from the local path which is the working directory path (got by pwd command), to another path

[harishga\_buffalo@ip-10-0-1-99 ~]$ **hdfs dfs -copyFromLocal harishga\_my\_first\_hdfs\_cmds\_webconsole\_02\_08\_2020.txt /user/harishga\_buffalo/1stfebdata**

[harishga\_buffalo@ip-10-0-1-99 ~]$ ls

harishga\_my\_first\_hdfs\_cmds\_webconsole\_02\_08\_2020.txt

[harishga\_buffalo@ip-10-0-1-99 ~]$ hdfs dfs -ls /user/harishga\_buffalo/

Found 2 items

drwxr-xr-x - harishga\_buffalo harishga\_buffalo 0 2020-02-08 18:03 /user/harishga\_buffalo/1stfebdata

drwxr-xr-x - harishga\_buffalo harishga\_buffalo 0 2020-02-08 17:52 /user/harishga\_buffalo/1stfebdata2

[harishga\_buffalo@ip-10-0-1-99 ~]$ hdfs dfs -ls /user/harishga\_buffalo/1stfebdata/

Found 1 items

-rw-r--r-- 2 harishga\_buffalo harishga\_buffalo 1852 2020-02-08 18:03 /user/harishga\_buffalo/1stfebdata/harishga\_my\_first\_hdfs\_cmds\_webconso

\_02\_08\_2020.txt

[harishga\_buffalo@ip-10-0-1-99 ~]$ hdfs dfs -ls /user | grep ajaykuma

drwxr-xr-x - ajaykuma24\_gmail ajaykuma24\_gmail 0

020-02-08 17:24 /user/ajaykuma24\_gmail

# the task was to copy a file from my directory onto the common4all folder of the instructor Ajay Singhal of the course Big Data and Spark Developer

#so using copyFromLocal cmd to copy from pwd above

[harishga\_buffalo@ip-10-0-1-99 ~]$ hdfs dfs -copyFromLocal harishga\_my\_first\_hdfs\_cmds\_webconsole\_02\_08\_2020.txt /user/ajaykuma24\_gmail/common4all

[harishga\_buffalo@ip-10-0-1-99 ~]$ hdfs dfs -ls /user/ajaykuma24\_gmail/

Found 72 items

drwx------ - ajaykuma24\_gmail ajaykuma24\_gmail 0 2020-02-08 18:00 /user/ajaykuma24\_gmail/.Trash

drwxr-xr-x - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-10-19 07:13 /user/ajaykuma24\_gmail/.sparkStaging

drwx------ - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-08-29 17:27 /user/ajaykuma24\_gmail/.staging

drwxr-xr-x - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-08-21 17:20 /user/ajaykuma24\_gmail/19thaug19

drwxr-xr-x - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-08-26 15:35 /user/ajaykuma24\_gmail/19thaug19bck

drwxr-xr-x - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-08-26 17:05 /user/ajaykuma24\_gmail/19thaug19i1

drwxr-xr-x - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-08-26 17:07 /user/ajaykuma24\_gmail/19thaug19i2

drwxr-xr-x - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-08-26 17:08 /user/ajaykuma24\_gmail/19thaug19i3

drwxr-xr-x - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-08-26 17:10 /user/ajaykuma24\_gmail/19thaug19i4

drwxr-xr-x - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-08-26 17:12 /user/ajaykuma24\_gmail/19thaug19i5

drwxr-xr-x - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-08-26 17:13 /user/ajaykuma24\_gmail/19thaug19i6

drwxr-xr-x - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-08-26 17:14 /user/ajaykuma24\_gmail/19thaug19i7

drwxr-xr-x - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-08-26 17:22 /user/ajaykuma24\_gmail/19thaug19i8

drwxr-xr-x - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-08-20 17:24 /user/ajaykuma24\_gmail/19thaugB

drwxr-xr-x - ajaykuma24\_gmail ajaykuma24\_gmail 0 2020-02-08 16:43 /user/ajaykuma24\_gmail/1stfebdata

-rw-r--r-- 2 ajaykuma24\_gmail ajaykuma24\_gmail 378811 2019-08-26 15:36 /user/ajaykuma24\_gmail/FederalFundsHighRange.xml

drwxr-xr-x - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-07-07 13:06 /user/ajaykuma24\_gmail/URLop-1

drwxr-xr-x - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-07-07 13:09 /user/ajaykuma24\_gmail/URLop-2

drwxr-xr-x - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-07-07 13:11 /user/ajaykuma24\_gmail/URLop-3

drwxr-xr-x - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-07-07 13:16 /user/ajaykuma24\_gmail/URLop-4

drwxr-xr-x - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-07-07 13:42 /user/ajaykuma24\_gmail/URLoput-1

drwxr-xr-x - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-07-07 14:16 /user/ajaykuma24\_gmail/URLoput-2

-rw-r--r-- 2 ajaykuma24\_gmail ajaykuma24\_gmail 4670 2020-02-02 17:17 /user/ajaykuma24\_gmail/a10.csv

-rw-r--r-- 2 ajaykuma24\_gmail ajaykuma24\_gmail 3912 2020-02-02 17:17 /user/ajaykuma24\_gmail/a10\_missings.csv

-rw-r--r-- 2 ajaykuma24\_gmail ajaykuma24\_gmail 66 2019-08-01 12:04 /user/ajaykuma24\_gmail/abc1.txt

drwxrwxrwx - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-08-22 12:06 /user/ajaykuma24\_gmail/common

drwxrwxrwx - ajaykuma24\_gmail ajaykuma24\_gmail 0 2020-02-08 18:06 /user/ajaykuma24\_gmail/common4all

-rw-r--r-- 2 ajaykuma24\_gmail ajaykuma24\_gmail 6712 2019-06-30 18:09 /user/ajaykuma24\_gmail/custglobal.txt

-rwxrwxrwx 2 ajaykuma24\_gmail ajaykuma24\_gmail 1068 2019-06-30 18:09 /user/ajaykuma24\_gmail/custpref.txt

drwxr-xr-x - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-09-09 16:34 /user/ajaykuma24\_gmail/flmDataAug

drwxrwxrwx - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-06-30 18:09 /user/ajaykuma24\_gmail/flume

drwxr-xr-x - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-07-20 16:53 /user/ajaykuma24\_gmail/hbtbl1

drwxr-xr-x - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-07-13 17:07 /user/ajaykuma24\_gmail/help\_keyword

drwxr-xr-x - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-07-13 17:08 /user/ajaykuma24\_gmail/helpky1

drwxr-xr-x - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-07-13 17:09 /user/ajaykuma24\_gmail/helpky2

drwxrwxrwx - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-06-30 18:09 /user/ajaykuma24\_gmail/hivedata

drwxrwxrwx - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-06-30 18:09 /user/ajaykuma24\_gmail/hiveq

-rw-r--r-- 2 ajaykuma24\_gmail ajaykuma24\_gmail 232 2019-07-27 15:47 /user/ajaykuma24\_gmail/lending\_types.txt

drwxr-xr-x - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-07-07 13:40 /user/ajaykuma24\_gmail/mydata

drwxr-xr-x - ajaykuma24\_gmail ajaykuma24\_gmail 0 2019-06-30 17:19 /user/ajaykuma24\_gmail/mydata2

[harishga\_buffalo@ip-10-0-1-99 ~]$ hdfs dfs -ls /user/ajaykuma24\_gmail/common4all

Found 4 items

-rw-r--r-- 2 rajesh\_mereddy\_gmail ajaykuma24\_gmail 3706094 2020-02-08 17:28 /user/ajaykuma24\_gmail/common4all/Bank\_full\_rajeshmereddy.csv

-rw-r--r-- 2 harishga\_buffalo ajaykuma24\_gmail 1852 2020-02-08 18:06 /user/ajaykuma24\_gmail/common4all/harishga\_my\_first\_hdfs\_cmds\_webc

sole\_02\_08\_2020.txt

-rw-r--r-- 2 mathisang7\_gmail ajaykuma24\_gmail 0 2020-02-08 17:38 /user/ajaykuma24\_gmail/common4all/test\_mathi.txt

drwxr-xr-x - vinod\_duri\_gmail ajaykuma24\_gmail 0 2020-02-08 17:55 /user/ajaykuma24\_gmail/common4all/test\_vinodduri

[harishga\_buffalo@ip-10-0-1-99 ~]$

#to check the contents of a file using cat unix/linux cmd

[harishga\_buffalo@ip-10-0-1-99 ~]$ **cat harishga\_my\_first\_hdfs\_cmds\_webconsole\_02\_08\_2020.txt**

# to recursively list the files/folders in my hdfs path, using R option .i.e. it checks folders within folders until it reaches final file

[harishga\_buffalo@ip-10-0-1-99 ~]$ **hdfs dfs -ls -R**

drwxr-xr-x - harishga\_buffalo harishga\_buffalo 0 2020-02-08 18:03 1stfebdata

-rw-r--r-- 2 harishga\_buffalo harishga\_buffalo 1852 2020-02-08 18:03 1stfebdata/harishga\_my\_first\_hdfs\_cmds\_webconsole\_02\_08\_2020.txt

drwxr-xr-x - harishga\_buffalo harishga\_buffalo 0 2020-02-08 17:52 1stfebdata2

#now checking how many name nodes are there and how many blocks and what is the replication factor etc, using fsck file system command check utility.

[harishga\_buffalo@ip-10-0-1-99 ~]$ **hdfs fsck /user/harishga\_buffalo/1stfebdata -files -blocks**

Connecting to namenode via http://10-0-1-98.ec2.internal:50070/fsck?ugi=harishga\_buffalo&files=1&blocks=1&path=%2Fuser%2Fharishga\_buffalo%2F1stfeb

data

FSCK started by harishga\_buffalo (auth:SIMPLE) from /10.0.1.10 for path /user/harishga\_buffalo/1stfebdata at Mon Feb 10 03:22:06 UTC 2020

/user/harishga\_buffalo/1stfebdata <dir>

/user/harishga\_buffalo/1stfebdata/harishga\_my\_first\_hdfs\_cmds\_webconsole\_02\_08\_2020.txt 1852 bytes, 1 block(s): OK

0. BP-1418814966-10.0.1.20-1522062197056:blk\_1079863811\_6127028 len=1852 Live\_repl=2

Status: HEALTHY

Total size: 1852 B

Total dirs: 1

Total files: 1

Total symlinks: 0

Total blocks (validated): 1 (avg. block size 1852 B)

Minimally replicated blocks: 1 (100.0 %)

Over-replicated blocks: 0 (0.0 %)

Under-replicated blocks: 0 (0.0 %)

Mis-replicated blocks: 0 (0.0 %)

Default replication factor: 2

Average block replication: 2.0

Corrupt blocks: 0

Missing replicas: 0 (0.0 %)

Number of data-nodes: 5

Number of racks: 1

FSCK ended at Mon Feb 10 03:22:06 UTC 2020 in 0 milliseconds

The filesystem under path '/user/harishga\_buffalo/1stfebdata' is HEALTHY

# /**opt**. This directory is reserved for all the software and add-on packages that are not part of the default installation

#checking the softwares present in opt path

[harishga\_buffalo@ip-10-0-1-99 ~]$ ls /opt

**aws** **cloudera** **SPARK2\_ON\_YARN-2.2.0.cloudera2.jar**

[harishga\_buffalo@ip-10-0-1-99 ~]$ ls /opt/cloudera/parcels

**CDH** **CLABS\_PHOENIX-4.7.0-1.clabs\_phoenix1.3.0.p0.000** **KAFKA-3.0.0-1.3.0.0.p0.40**

**CDH-5.14.0-1.cdh5.14.0.p0.24** **CLABS\_PHOENIX-4.7.0-1.clabs\_phoenix1.3.0.p0.000-el5.parcel** **SPARK2**

**CLABS\_PHOENIX** **KAFKA**

#now checking the services provided by cloudera parcels

[harishga\_buffalo@ip-10-0-1-99 ~]$ ls /opt/cloudera/parcels/CDH-5.14.0-1.cdh5.14.0.p0.24/lib

**avro** **debug** **hadoop-hdfs** **hadoop-yarn** **hive-hcatalog** **kite** **oozie** **sentry** **sqoop2**

**bigtop-tomcat** **flume-ng** **hadoop-httpfs** **hbase** **hue** **kudu** **parquet** **solr** **whirr**

**bigtop-utils** **hadoop** **hadoop-kms** **hbase-solr** **impala** **llama** **pig** **spark** **zookeeper**

**crunch** **hadoop-0.20-mapreduce** **hadoop-mapreduce** **hive** **impala-shell** **mahout** **search** **sqoop** **zookeeper-native**

#Checking the path of Hadoop config files:

[harishga\_buffalo@ip-10-0-1-99 ~]$ ls /opt/cloudera/parcels/CDH-5.14.0-1.cdh5.14.0.p0.24/lib/hadoop/etc/hadoop

\_\_cloudera\_generation\_\_ container-executor.cfg hadoop-env.sh log4j.properties ssl-client.xml **topology.py**

\_\_cloudera\_metadata\_\_ core-site.xml hdfs-site.xml mapred-site.xml topology.map yarn-site.xml

#or we can use the following and see the config files going into conf sub folder below

[harishga\_buffalo@ip-10-0-1-99 ~]$ ls /etc/hadoop

**conf** **conf.cloudera.hdfs** **conf.cloudera.yarn**

#checking Hadoop classpath

[harishga\_buffalo@ip-10-0-1-99 ~]$ **hadoop classpath**

/etc/hadoop/conf:/opt/cloudera/parcels/CDH-5.14.0-1.cdh5.14.0.p0.24/lib/hadoop/libexec/../../hadoop/lib/\*:/opt/cloudera/parcels/CDH-5.14.0-1.cdh5.14.

0.p0.24/lib/hadoop/libexec/../../hadoop/.//\*:/opt/cloudera/parcels/CDH-5.14.0-1.cdh5.14.0.p0.24/lib/hadoop/libexec/../../hadoop-hdfs/./:/opt/cloudera

/parcels/CDH-5.14.0-1.cdh5.14.0.p0.24/lib/hadoop/libexec/../../hadoop-hdfs/lib/\*:/opt/cloudera/parcels/CDH-5.14.0-1.cdh5.14.0.p0.24/lib/hadoop/libexe

c/../../hadoop-hdfs/.//\*:/opt/cloudera/parcels/CDH-5.14.0-1.cdh5.14.0.p0.24/lib/hadoop/libexec/../../hadoop-yarn/lib/\*:/opt/cloudera/parcels/CDH-5.14

.0-1.cdh5.14.0.p0.24/lib/hadoop/libexec/../../hadoop-yarn/.//\*:/opt/cloudera/parcels/CDH/lib/hadoop-mapreduce/lib/\*:/opt/cloudera/parcels/CDH/lib/had

oop-mapreduce/.//\*

#checking the contents of config file we can get the hdfs path; Config files contain Property, Name, Value

[harishga\_buffalo@ip-10-0-1-99 ~]$ **cat /etc/hadoop/conf/core-site.xml**

<?xml version="1.0" encoding="UTF-8"?>

<!--Autogenerated by Cloudera Manager-->

<configuration>

<property>

<name>fs.defaultFS</name>

<value>hdfs://10-0-1-98.ec2.internal:8020</value>

</property>