Big Data Analysis

Practical 3: Single node Hadoop cluster

Objective:

Setup single node Hadoop cluster and apply HDFS commands on single node Hadoop Cluster.

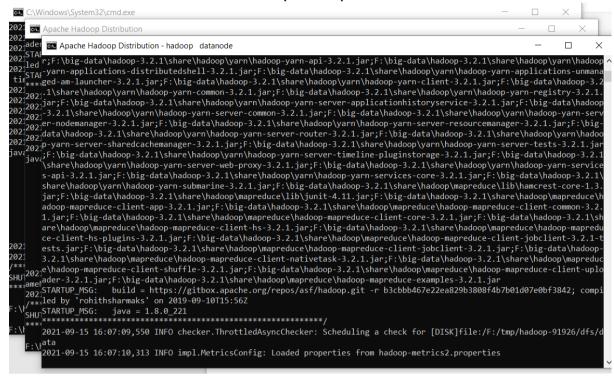
Roll No. & Name: 18bce 183 & Prince Prajapati

Submitted to: Prof. Jaiprakash Verma

1. Initialise HDFS

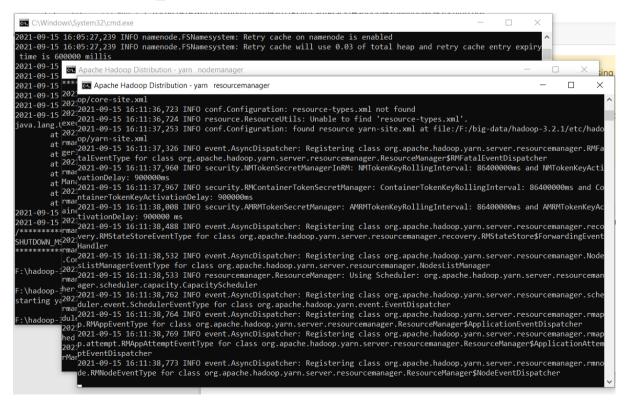
2. Start HDFS daemons

% HADOOP_HOME% \sbin\start-dfs.cmd



3. Start YARN daemons

%HADOOP_HOME%\sbin\start-yarn.cmd



4. Print the Hadoop version

hadoop version

```
C:\Users\91926\Documents\sem 7\BDA\Lab\Prac3>hadoop version

Hadoop 3.2.1

Source code repository https://gitbox.apache.org/repos/asf/hadoop.git -r b3cbbb467e22ea829b3808f4b7b01d07e0bf3842

Compiled by rohithsharmaks on 2019-09-10T15:56Z

Compiled with protoc 2.5.0

From source with checksum 776eaf9eee9c0ffc370bcbc1888737

This command was run using /F:/big-data/hadoop-3.2.1/share/hadoop/common/hadoop-common-3.2.1.jar
```

5. jps: Java Virtual Machine Process Status tool.

```
F:\hadoop-3.2.1\hadoop-3.2.1\sbin>jps
12736 Jps
2560 ResourceManager
```

- 6. make directry
- 7. List the contents of the root directory in HDFS
 - hadoop fs -ls /
 - Hadoop fs -mkdir /directory

```
F:\hadoop-3.2.1\hadoop-3.2.1>hadoop fs -mkdir /prince
F:\hadoop-3.2.1\hadoop-3.2.1>hadoop fs -ls /
Found 10 items
```

. . .

- 8. Report the amount of space used and available on currently mounted files system
 - hadoop fs -du hdfs:/

```
F:\hadoop-3.2.1\hadoop-3.2.1>hadoop fs -du
1032755
            1032755
                        bin
106682
           106682
            55827
                        include
55827
            27854844
27854844
                        lib
152850
            152850
                        libexec
150569
           150569
                        LICENSE.txt
22125
           22125
                       NOTICE.txt
           1361
1361
                        README.txt
88570
           88570
                        shin
852761395
           852761395
                        share
```

- 9. Count the number of directories, files and bytes under the paths that match the specified file pattern
 - hadoop fs -count hdfs:/

```
C:\Users\91926\Documents\sem 7\BDA\Lab\Prac3\Hadoop>hadoop fs -count hdfs:/
count: Incomplete HDFS URI, no host: hdfs:/
```

- 10. Run a DFS file system checking utility
 - hadoop fsck /

```
C:\Users\91926\Documents\sem 7\BDA\Lab\Prac3\Hadoop>hadoop fsck - /
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.
EileSystem is file:///
DFSck exiting.
```

- 11. Run a cluster balancing utility
 - hadoop balancer

```
C:\Users\91926\Documents\sem 7\BDA\Lab\Prac3\Hadoop>hadoop balancer

DEPRECATED: Use of this script to execute hdfs command is deprecated.

Instead use the hdfs command for it.

2021-09-12 18:44:02,058 INFO balancer.Balancer: namenodes = []

2021-09-12 18:44:02,064 INFO balancer.Balancer: parameters = Balancer.BalancerParameters [Balancing 0, #blockpools = 0, run during upgrade = false]

2021-09-12 18:44:02,064 INFO balancer.Balancer: included nodes = []

2021-09-12 18:44:02,064 INFO balancer.Balancer: excluded nodes = []

2021-09-12 18:44:02,064 INFO balancer.Balancer: source nodes = []

Time Stamp Iteration# Bytes Already Moved Bytes Left To Move Bytes Being Moved

12 Sep, 2021 6:44:02 PM Balancing took 381.0 milliseconds
```

12. hadoop fs -put F:/sample.txt /prince : add local directory file to hdfs directory

```
F:\hadoop-3.2.1\hadoop-3.2.1>hadoop fs -put F:/sample.txt /prince

F:\hadoop-3.2.1\hadoop-3.2.1>hadoop fs -ls /prince

Found 1 items

2021-09-15 16:33:21,285 WARN nativeio.NativeIO: NativeIO.getStat error (3): The system cannot find the path specified.

-- file path: prince/sample.txt

-rw-rw-r-- 1 91926 91926 10 2021-09-15 16:32 /prince/sample.txt
```

13. hadoop fs -put C:/Users/Hp/retail/hadoop: add local directory folder to hdfs directory

14. hadoop fs -du -s -h /prince : See how much space directory occupies in HDFS.

```
F:\hadoop-3.2.1\hadoop-3.2.1>hadoop fs -du -s -h /prince
10  10 /prince
```

15. hadoop fs -rm /prince/sample.txt : deleting a file from directory in hdfs

```
F:\hadoop-3.2.1\hadoop-3.2.1>hadoop fs -rm /prince/sample.txt
2021-09-15 16:42:07,380 INFO Configuration.deprecation: io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
Deleted /prince/sample.txt
```

16. hadoop fs -rm /prince/prac3/* : deleting all the files from directory in hdfs

F:\hadoop-3.2.1\hadoop-3.2.1>hadoop fs -rm /prince/prac3/*
2021-09-15 16:45:30,097 INFO Configuration.deprecation: io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
Deleted /prince/prac3/sample.txt

17. hadoop fs –expunge: to empty the trash

```
F:\hadoop-3.2.1\hadoop-3.2.1>hadoop fs -expunge
F:\hadoop-3.2.1\hadoop-3.2.1>
```

18. hadoop fs -rmr /prince/prac3 : remove entire directory and all its content in hdfs

```
F:\hadoop-3.2.1\hadoop-3.2.1>hadoop fs -rmr /prince/prac3
rmr: DEPRECATED: Please use '-rm -r' instead.
2021-09-15 16:48:34,931 INFO Configuration.deprecation: io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
Deleted /prince/prac3
```

19. hadoop fs -copyFromLocal C:/Users/Hp/purchases.txt /hadoop : add local directory file to hadoop directory in hdfs

```
F:\hadoop-3.2.1\hadoop-3.2.1>hadoop fs -copyFromLocal F:/java/hello.java/PP2.java /prince
F:\hadoop-3.2.1\hadoop-3.2.1>_
```

20. hadoop fs -cat /prince/PP2.java : view the content of file which is present in hdfs directory

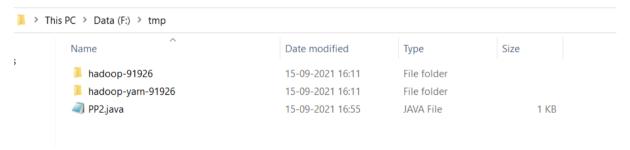
```
F:\hadoop-3.2.1\hadoop-3.2.1>hadoop fs -cat /prince/PP2.java
abstract class A
{

void callmetoo()
{
   System.out.println("This is a concrete method.");
} }

class B extends A
{
   void callme()
{
   System.out.println("B's implementation of callme.");
} }

class PP2
{
   public static void main(String args[])
{
   B b = new B();
   b.callme();
}
}
```

21. hadoop fs -copyToLocal /prince/PP2.java F:\tmp: add file which is present in hdfs directory to local directory



Conclusion: After implementing this practical now I have good practice in working with Hadoop and its commands.