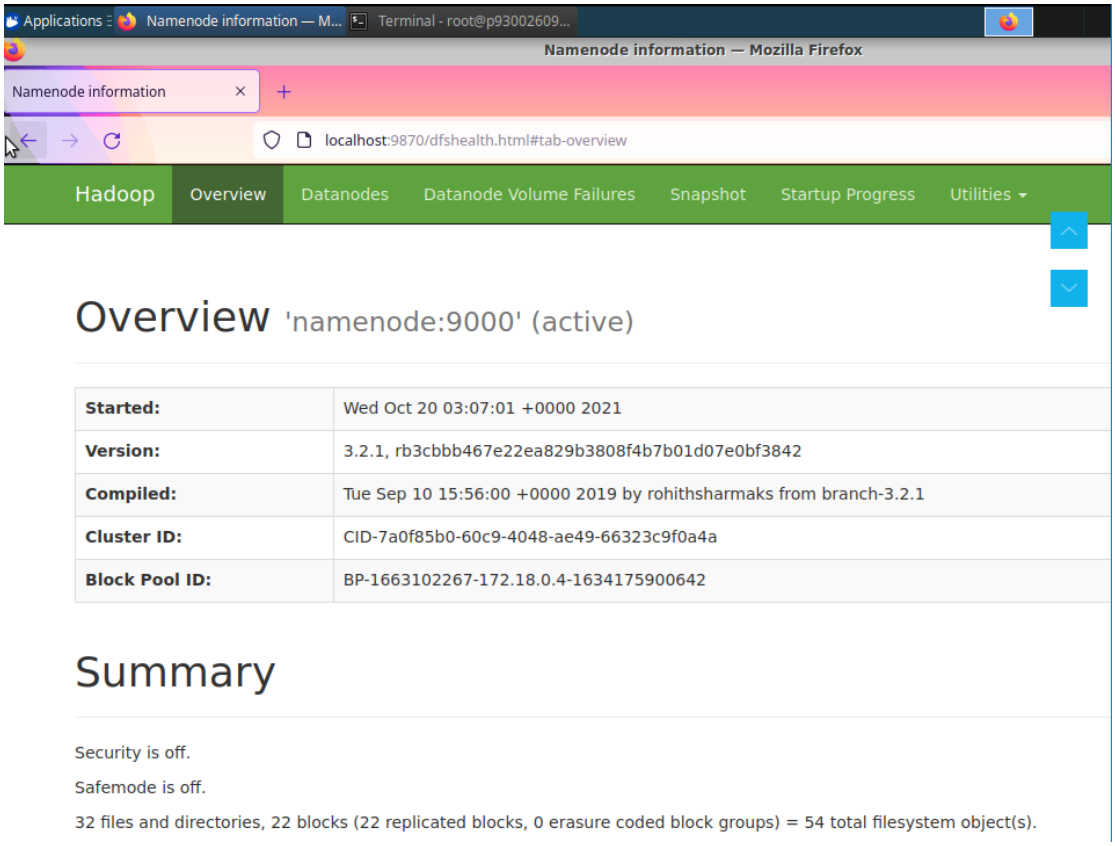
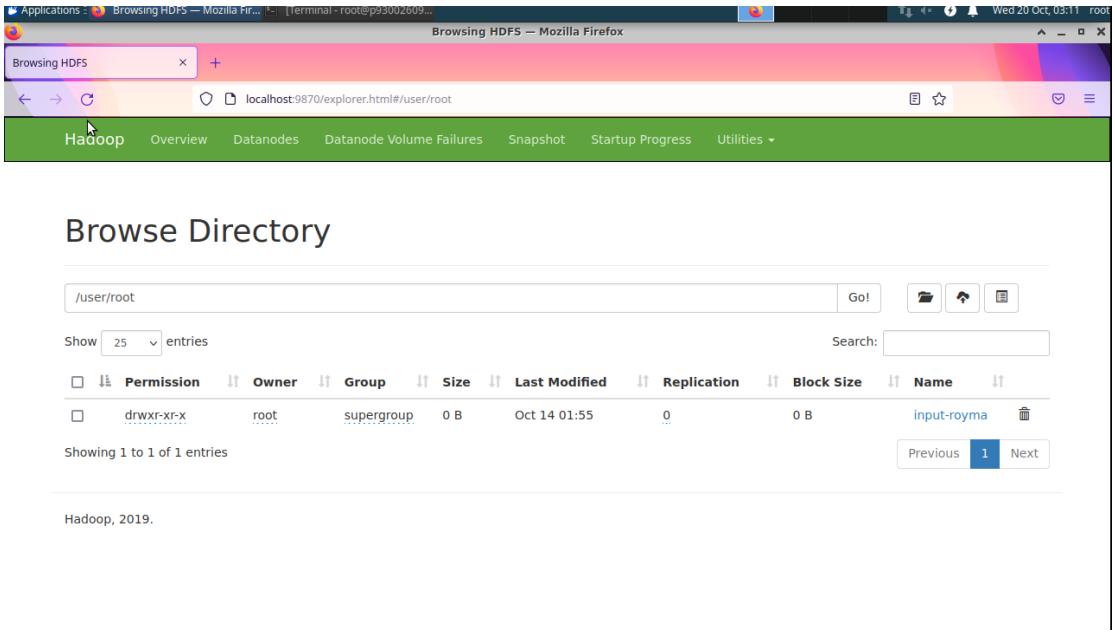
Basic

```
root@p930026096:~# cd docker-hadoop-master/
root@p930026096:~/docker-hadoop-master# docker-compose up -d
nodemanager is up-to-date
namenode is up-to-date
resourcemanager is up-to-date
historyserver is up-to-date
datanode is up-to-date
root@p930026096:~/docker-hadoop-master#
```

Start the docker. Here docker already started.



Make sure the namenode is running by checking the port 9870.



The directory input-royma exists. For storing two .txt file (f1.txt & f2.txt).

```
root@p930026096:~/Downloads# ls
WordCount.java
```

Download WordCount.java under "~/Downloads/"

```
root@p930026096:~/Downloads# docker container ls
CONTAINER ID   IMAGE                                                    COMMAND                  CREATED      STATUS                    PORTS
                                                NAMES
2351e46b2519   bde2020/hadoop-historyserver:2.0.0-hadoop3.2.1-java8     "/entrypoint.sh /run…"   6 days ago   Up 13 minutes (healthy)   8188/tcp
                                                historyserver
bbe3b7f06d38   bde2020/hadoop-namenode:2.0.0-hadoop3.2.1-java8          "/entrypoint.sh /run…"   6 days ago   Up 13 minutes (healthy)   0.0.0.0:9000->9000/tcp, :::9000->9000/tcp, 0
.0.0.0:9870->9870/tcp, :::9870->9870/tcp   namenode
ea3ae59bf34c   bde2020/hadoop-datanode:2.0.0-hadoop3.2.1-java8          "/entrypoint.sh /run…"   6 days ago   Up 13 minutes (healthy)   9864/tcp
                                                datanode
a45d74f53a72   bde2020/hadoop-resourcemanager:2.0.0-hadoop3.2.1-java8   "/entrypoint.sh /run…"   6 days ago   Up 12 minutes (healthy)   8088/tcp
                                                resourcemanager
a3fa26117290   bde2020/hadoop-nodemanager:2.0.0-hadoop3.2.1-java8       "/entrypoint.sh /run…"   6 days ago   Up 13 minutes (healthy)   8042/tcp
                                                nodemanager
root@p930026096:~/Downloads#
```

Find out the ID of my own namenode container.

```
root@p930026096:~/Downloads# docker cp ~/Downloads/WordCount.java bbe3b7f06d38:WordCount.java
```

Copy the WordCount.java file from ~/Downloads to my docker Hadoop cluster.

```
root@p930026096:~/Downloads# docker exec -it namenode bash
```

Start and enter namenode.

```
root@bbe3b7f06d38:/# export HADOOP_CLASSPATH=${JAVA_HOME}/lib/tools.jar
```

Create a jar

```
root@bbe3b7f06d38:/# hadoop com.sun.tools.javac.Main WordCount.java
/opt/hadoop-3.2.1/libexec/hadoop-functions.sh: line 2401: HADOOP_COM.SUN.TOOLS.JAVAC.MAIN_USER: bad substitution
/opt/hadoop-3.2.1/libexec/hadoop-functions.sh: line 2366: HADOOP_COM.SUN.TOOLS.JAVAC.MAIN_USER: bad substitution
/opt/hadoop-3.2.1/libexec/hadoop-functions.sh: line 2461: HADOOP_COM.SUN.TOOLS.JAVAC.MAIN_OPTS: bad substitution
```

```
root@bbe3b7f06d38:/# jar cf wc.jar WordCount*.class
```

"wc.jar" file is under current working directory and ready to run the Java program.

```
root@bbe3b7f06d38:/# ls
KEYS                        WordCount$TokenizerMapper.class  WordCount.java  boot  entrypoint.sh  hadoop       home   lib    media  opt   root  run.sh  srv  tmp  var
WordCount$IntSumReducer.class  WordCount.class                bin             dev   etc            hadoop-data  input  lib64  mnt    proc  run   sbin    sys  usr  wc.jar
```

```
root@bbe3b7f06d38:/# hadoop jar wc.jar org.apache.hadoop.examples.WordCount input-royma output
```

Run the WordCount.java program and count the word in two files.

```
2021-10-20 03:57:32,353 INFO client.RMProxy: Connecting to ResourceManager at resourcemanager/172.18.0.6:8032
2021-10-20 03:57:32,559 INFO client.AHSProxy: Connecting to Application History server at historyserver/172.18.0.2:10200
2021-10-20 03:57:32,773 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/root/.staging/job_1634701928272_0001
2021-10-20 03:57:32,908 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-10-20 03:57:33,081 INFO input.FileInputFormat: Total input files to process : 2
2021-10-20 03:57:33,126 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-10-20 03:57:33,152 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-10-20 03:57:33,158 INFO mapreduce.JobSubmitter: number of splits:2
2021-10-20 03:57:33,283 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-10-20 03:57:33,314 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1634701928272_0001
2021-10-20 03:57:33,314 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-10-20 03:57:33,501 INFO conf.Configuration: resource-types.xml not found
2021-10-20 03:57:33,502 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-10-20 03:57:33,842 INFO impl.YarnClientImpl: Submitted application application_1634701928272_0001
2021-10-20 03:57:33,878 INFO mapreduce.Job: The url to track the job: http://resourcemanager:8088/proxy/application_1634701928272_0001/
2021-10-20 03:57:33,878 INFO mapreduce.Job: Running job: job_1634701928272_0001
2021-10-20 03:57:41,028 INFO mapreduce.Job: Job job_1634701928272_0001 running in uber mode : false
2021-10-20 03:57:41,036 INFO mapreduce.Job:  map 0% reduce 0%
2021-10-20 03:57:59,560 INFO mapreduce.Job:  map 100% reduce 0%
2021-10-20 03:58:52,994 INFO mapreduce.Job:  map 100% reduce 100%
2021-10-20 03:59:16,094 INFO mapreduce.Job: Job job_1634701928272_0001 completed successfully
2021-10-20 03:59:16,236 INFO mapreduce.Job: Counters: 54
        File System Counters
                FILE: Number of bytes read=62
                FILE: Number of bytes written=688017
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=257
                HDFS: Number of bytes written=31
                HDFS: Number of read operations=11
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
                HDFS: Number of bytes read erasure-coded=0
        Job Counters
                Launched map tasks=2
                Launched reduce tasks=1
                Rack-local map tasks=2
```

```
                Rack-local map tasks=2
                Total time spent by all maps in occupied slots (ms)=119924
                Total time spent by all reduces in occupied slots (ms)=218880
                Total time spent by all map tasks (ms)=29981
                Total time spent by all reduce tasks (ms)=27260
                Total vcore-milliseconds taken by all map tasks=29981
                Total vcore-milliseconds taken by all reduce tasks=27260
                Total megabyte-milliseconds taken by all map tasks=122802176
                Total megabyte-milliseconds taken by all reduce tasks=223313920
        Map-Reduce Framework
                Map input records=2
                Map output records=5
                Map output bytes=49
                Map output materialized bytes=80
                Input split bytes=228
                Combine input records=5
                Combine output records=5
                Reduce input groups=4
                Reduce shuffle bytes=80
                Reduce input records=5
                Reduce output records=4
                Spilled Records=10
                Shuffled Maps =2
                Failed Shuffles=0
                Merged Map outputs=2
                GC time elapsed (ms)=490
                CPU time spent (ms)=1680
                Physical memory (bytes) snapshot=518209536
                Virtual memory (bytes) snapshot=18604883968
                Total committed heap usage (bytes)=398458880
                Peak Map Physical memory (bytes)=205385728
                Peak Map Virtual memory (bytes)=5086400512
                Peak Reduce Physical memory (bytes)=113324032
                Peak Reduce Virtual memory (bytes)=8432193536
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=29
        File Output Format Counters
                Bytes Written=31
root@bbe3b7f06d38:/#
```

It takes few minutes to finish map-reduce process. It will show the message, “map 100% reduce 100%”, if it runs well.

```
root@bbe3b7f06d38:/# hdfs dfs -cat output/part-r-00000
2021-10-20 04:11:38,740 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
Data    1
Hello   2
Science 1
UIC     1
root@bbe3b7f06d38:/#
```

To print out the result of WordCount.java program.

```
root@bbe3b7f06d38:/# exit
exit
```

Leave the container.

```
root@p930026096:~# cd docker-hadoop-master/
root@p930026096:~/docker-hadoop-master# docker-compose down
Stopping historyserver   ... done
Stopping namenode        ... done
Stopping datanode        ... done
Stopping resourcemanager ... done
Stopping nodemanager     ... done
Removing historyserver   ... done
Removing namenode        ... done
Removing datanode        ... done
Removing resourcemanager ... done
Removing nodemanager     ... done
Removing network docker-hadoop-master_default
root@p930026096:~/docker-hadoop-master#
```

Safely shut down the cluster.