

Hussein Mohamed

Day02

Task01 HDFS

1: exec the docker

```

h@ja:~/Documents/ITI/BIG DATA/hadoop-hive-spark-docker-master$ docker ps
CONTAINER ID   IMAGE                                COMMAND                  CREATED        STATUS        PORTS
8ae7f572f5e6   hadoop-hive-spark-worker           "entrypoint.sh run.sh"  11 hours ago  Up 16 minutes  0.0.0.0:8043->8042/tcp, [::]:8043->8042/tcp, 0.0.0.0:8082->8081/tcp, [::]:8082->8081/tcp, 0.0.0.0:9865->9864/tcp, [::]:9865->9864/tcp
037cce20e592   hadoop-hive-spark-worker           "entrypoint.sh run.sh"  11 hours ago  Up 16 minutes  0.0.0.0:8042->8042/tcp, [::]:8042->8042/tcp, 0.0.0.0:8081->8081/tcp, [::]:8081->8081/tcp, 0.0.0.0:9864->9864/tcp, [::]:9864->9864/tcp
9817ed9d456e   hadoop-hive-spark-worker           "entrypoint.sh run.sh"  11 hours ago  Up 16 minutes  0.0.0.0:8044->8042/tcp, [::]:8044->8042/tcp, 0.0.0.0:8083->8081/tcp, [::]:8083->8081/tcp, 0.0.0.0:9866->9864/tcp, [::]:9866->9864/tcp
20e7ac3e37a7   hadoop-hive-spark-master           "entrypoint.sh run.sh"  11 hours ago  Up 16 minutes  0.0.0.0:4040->4040/tcp, [::]:4040->4040/tcp, 0.0.0.0:7077->7077/tcp, [::]:7077->7077/tcp, 0.0.0.0:8020->8020/tcp, [::]:8020->8020/tcp, 0.0.0.0:8080->8080/tcp, [::]:8080->8080/tcp, 0.0.0.0:8088->8088/tcp, [::]:8088->8088/tcp, 0.0.0.0:9083->9083/tcp, [::]:9083->9083/tcp, 0.0.0.0:9870->9870/tcp, [::]:9870->9870/tcp, 0.0.0.0:10000->10000/tcp, [::]:10000->10000/tcp
08546d4e68e9   postgres:11                        "docker-entrypoint.s..."  11 hours ago  Up 16 minutes  0.0.0.0:15432->5432/tcp, [::]:15432->5432/tcp

hadoop-hive-spark-docker-master-metastore-1
h@ja:~/Documents/ITI/BIG DATA/hadoop-hive-spark-docker-master$ docker exec -it hadoop-hive-spark-docker-master-master-1 bash

```

2: chmod

```

jupyter@master:/$ ls
bin  dev  file2.txt  lib  media  myfile.txt  proc  run  srv  tmp  var
boot  etc  home      lib64  mnt    opt        root  sbin  sys  usr
jupyter@master:/$ touch f
jupyter@master:/$ sudo -i
root@master:~# chmod 777 /
root@master:~# |

```

3: hdfs dfs -ls / and cat

```

jupyter@master:/$ touch file3.txt
jupyter@master:/$ echo hi hi > file3.txt
jupyter@master:/$ hdfs dfs -ls /
Found 6 items
-rw-r--r--  3 jupyter supergroup      8 2025-09-27 22:53 /file2.txt
-rw-r--r--  3 root   supergroup     10 2025-09-27 12:23 /myfile.txt
-rw-r--r--  3 jupyter supergroup     10 2025-09-27 12:26 /myfile2.txt
drwxr-xr-x  - jupyter supergroup      0 2025-09-27 12:12 /spark
drwxr-xr-x  - jupyter supergroup      0 2025-09-27 12:12 /tmp
drwxr-xr-x  - jupyter supergroup      0 2025-09-27 12:12 /var
jupyter@master:/$ hdfs dfs -cat /file2.txt
hi hdfs

```

4: get and put

```
jupyter@master:/$ ls
bin  dev  home  lib64  mnt  proc  run  srv  tmp  var
boot  etc  lib   media  opt  root  sbin  sys  usr
jupyter@master:/$ touch filetoput.txt
jupyter@master:/$ echo hi hi hadoop > filetoput.txt
jupyter@master:/$ hdfs dfs -put /filetoput.txt /
jupyter@master:/$ hdfs dfs -cat /filetoput.txt
hi hi hadoop
jupyter@master:/$ hdfs dfs -ls /
Found 8 items
-rw-r--r--  3 jupyter supergroup      8 2025-09-27 22:53 /file2.txt
-rw-r--r--  3 jupyter supergroup      6 2025-09-27 22:58 /file3.txt
-rw-r--r--  3 jupyter supergroup     13 2025-09-28 11:27 /filetoput.txt
-rw-r--r--  3 root supergroup        10 2025-09-27 12:23 /myfile.txt
-rw-r--r--  3 jupyter supergroup     10 2025-09-27 12:26 /myfile2.txt
drwxr-xr-x  - jupyter supergroup      0 2025-09-27 12:12 /spark
drwxr-xr-x  - jupyter supergroup      0 2025-09-27 12:12 /tmp
drwxr-xr-x  - jupyter supergroup      0 2025-09-27 12:12 /var
jupyter@master:/$ hdfs dfs -get /file2.txt /
jupyter@master:/$ cat file2.txt
hi hdfs
jupyter@master:/$ |
```

5: stop a data node and try cat

```
h@ja: ~/Documents/ITI/BIG DATA/hadoop-hive-spark-docker-master
jupyter@master:/$ exit
exit
(.venv) h@ja:~/Documents/ITI/BIG DATA/hadoop-hive-spark-docker-master$ docker stop hadoop-hive-spark-docker-master-worker1-1
hadoop-hive-spark-docker-master-worker1-1
(.venv) h@ja:~/Documents/ITI/BIG DATA/hadoop-hive-spark-docker-master$ docker exec -it hadoop-hive-spark-docker-master-master-1 bash
jupyter@master:/$ hdfs dfs -ls /
Found 7 items
-rw-r--r--  3 jupyter supergroup      8 2025-09-27 22:53 /file2.txt
-rw-r--r--  3 jupyter supergroup      6 2025-09-27 22:58 /file3.txt
-rw-r--r--  3 root supergroup        10 2025-09-27 12:23 /myfile.txt
-rw-r--r--  3 jupyter supergroup     10 2025-09-27 12:26 /myfile2.txt
drwxr-xr-x  - jupyter supergroup      0 2025-09-27 12:12 /spark
drwxr-xr-x  - jupyter supergroup      0 2025-09-27 12:12 /tmp
drwxr-xr-x  - jupyter supergroup      0 2025-09-27 12:12 /var
jupyter@master:/$ hdfs dfs -cat /file2.txt
```

6: a result with error and long time loading

```
h@ja: ~/Documents/ITI/BIG DATA/hadoop-hive-spark-docker-master
at org.apache.hadoop.net.NetUtils.connect(NetUtils.java:603)
at org.apache.hadoop.hdfs.DFSClient.newConnectedPeer(DFSClient.java:3033)
at org.apache.hadoop.hdfs.client.impl.BlockReaderFactory.nextTcpPeer(BlockReaderFactory.java:829)
at org.apache.hadoop.hdfs.client.impl.BlockReaderFactory.getRemoteBlockReaderFromTcp(BlockReaderFactory.java:75
4)
at org.apache.hadoop.hdfs.client.impl.BlockReaderFactory.build(BlockReaderFactory.java:381)
at org.apache.hadoop.hdfs.DFSInputStream.getBlockReader(DFSInputStream.java:715)
at org.apache.hadoop.hdfs.DFSInputStream.blockSeekTo(DFSInputStream.java:645)
at org.apache.hadoop.hdfs.DFSInputStream.readWithStrategy(DFSInputStream.java:845)
at org.apache.hadoop.hdfs.DFSInputStream.read(DFSInputStream.java:918)
at java.io.DataInputStream.read(DataInputStream.java:100)
at org.apache.hadoop.io.IOUtils.copyBytes(IOUtils.java:96)
at org.apache.hadoop.io.IOUtils.copyBytes(IOUtils.java:69)
at org.apache.hadoop.io.IOUtils.copyBytes(IOUtils.java:133)
at org.apache.hadoop.fs.shell.Display$Cat.printToStdout(Display.java:101)
at org.apache.hadoop.fs.shell.Display$Cat.processPath(Display.java:96)
at org.apache.hadoop.fs.shell.Command.processPathInternal(Command.java:382)
at org.apache.hadoop.fs.shell.Command.processPaths(Command.java:345)
at org.apache.hadoop.fs.shell.Command.processPathArgument(Command.java:318)
at org.apache.hadoop.fs.shell.Command.processArgument(Command.java:300)
at org.apache.hadoop.fs.shell.Command.processArguments(Command.java:284)
at org.apache.hadoop.fs.shell.FsCommand.processRawArguments(FsCommand.java:121)
at org.apache.hadoop.fs.shell.Command.run(Command.java:191)
at org.apache.hadoop.fs.FsShell.run(FsShell.java:327)
at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:82)
at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:97)
at org.apache.hadoop.fs.FsShell.main(FsShell.java:390)
2025-09-27 23:01:47,884 INFO hdfs.DFSClient: Successfully connected to /172.28.1.5:9866 for BP-1553536198-172.28.1.2-17
58975076798:blk_1073742077_1253
hi hdfs
jupyter@master:/$ |
```

7: stop worker2, another error but still gives result

```
h@ja:~/Documents/ITI/BIG DATA/hadoop-hive-spark-docker-master$ docker stop hadoop-hive-spark-docker-
master-worker2-1
hadoop-hive-spark-docker-master-worker2-1
h@ja:~/Documents/ITI/BIG DATA/hadoop-hive-spark-docker-master$ docker exec -it hadoop-hive-spark-doc
ker-master-master-1 bash
jupyter@master:/$ hdfs dfs -cat /
.dockerenv dev/ lib/ mnt/ root/ srv/ usr/
bin/ etc/ lib64/ opt/ run/ sys/ var/
boot/ home/ media/ proc/ sbin/ tmp/
jupyter@master:/$ hdfs dfs -cat /file2.txt
2025-09-28 11:18:02,943 WARN impl.BlockReaderFactory: I/O error constructing remote block reader.
java.net.NoRouteToHostException: No route to host
at sun.nio.ch.SocketChannelImpl.checkConnect(Native Method)
at sun.nio.ch.SocketChannelImpl.finishConnect(SocketChannelImpl.java:716)
at org.apache.hadoop.net.SocketIOWithTimeout.connect(SocketIOWithTimeout.java:205)
at org.apache.hadoop.net.NetUtils.connect(NetUtils.java:600)
at org.apache.hadoop.hdfs.DFSClient.newConnectedPeer(DFSClient.java:3033)
at org.apache.hadoop.hdfs.client.impl.BlockReaderFactory.nextTcpPeer(BlockReaderFactory.java
:829)
at org.apache.hadoop.hdfs.client.impl.BlockReaderFactory.getRemoteBlockReaderFromTcp(BlockRe
aderFactory.java:754)
at org.apache.hadoop.hdfs.client.impl.BlockReaderFactory.build(BlockReaderFactory.java:381)
at org.apache.hadoop.hdfs.DFSInputStream.getBlockReader(DFSInputStream.java:715)
at org.apache.hadoop.hdfs.DFSInputStream.blockSeekTo(DFSInputStream.java:645)
at org.apache.hadoop.hdfs.DFSInputStream.readWithStrategy(DFSInputStream.java:845)
at org.apache.hadoop.hdfs.DFSInputStream.read(DFSInputStream.java:918)
```

8: the result

```
at org.apache.hadoop.hdfs.client.impl.BlockReaderFactory.build(BlockReaderFactory.java:381)
at org.apache.hadoop.hdfs.DFSInputStream.getBlockReader(DFSInputStream.java:715)
at org.apache.hadoop.hdfs.DFSInputStream.blockSeekTo(DFSInputStream.java:645)
at org.apache.hadoop.hdfs.DFSInputStream.readWithStrategy(DFSInputStream.java:845)
at org.apache.hadoop.hdfs.DFSInputStream.read(DFSInputStream.java:918)
at java.io.DataInputStream.read(DataInputStream.java:100)
at org.apache.hadoop.io.IOUtils.copyBytes(IOUtils.java:96)
at org.apache.hadoop.io.IOUtils.copyBytes(IOUtils.java:69)
at org.apache.hadoop.io.IOUtils.copyBytes(IOUtils.java:133)
at org.apache.hadoop.fs.shell.Display$Cat.printToStdout(Display.java:101)
at org.apache.hadoop.fs.shell.Display$Cat.processPath(Display.java:96)
at org.apache.hadoop.fs.shell.Command.processPathInternal(Command.java:382)
at org.apache.hadoop.fs.shell.Command.processPaths(Command.java:345)
at org.apache.hadoop.fs.shell.Command.processPathArgument(Command.java:318)
at org.apache.hadoop.fs.shell.Command.processArgument(Command.java:300)
at org.apache.hadoop.fs.shell.Command.processArguments(Command.java:284)
at org.apache.hadoop.fs.shell.FsCommand.processRawArguments(FsCommand.java:121)
at org.apache.hadoop.fs.shell.Command.run(Command.java:191)
at org.apache.hadoop.fs.FsShell.run(FsShell.java:327)
at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:82)
at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:97)
at org.apache.hadoop.fs.FsShell.main(FsShell.java:390)
2025-09-28 11:18:17,397 INFO hdfs.DFSClient: Successfully connected to /172.28.1.5:9866 for BP-1553536198-172.28.1.2-1758975076798:blk_1073742077_1253
hi hdfs
jupyter@master:/$ |
```

9: stop all workers, no result

```
a:~/Documents/ITI/BIG DATA/hadoop-hive-spark-docker-master$ docker stop hadoop-hive-spark-docker-
ter-worker3-1
oap-hive-spark-docker-master-worker3-1
a:~/Documents/ITI/BIG DATA/hadoop-hive-spark-docker-master$ docker exec -it hadoop-hive-spark-doc
-master-master-1 bash
/ter@master:/$ hdfs dfs -cat /file2.txt
5-09-28 11:20:44,415 WARN impl.BlockReaderFactory: I/O error constructing remote block reader.
a.net.NoRouteToHostException: No route to host
at sun.nio.ch.SocketChannelImpl.checkConnect(Native Method)
at sun.nio.ch.SocketChannelImpl.finishConnect(SocketChannelImpl.java:716)
at org.apache.hadoop.net.SocketIOWithTimeout.connect(SocketIOWithTimeout.java:205)
at org.apache.hadoop.net.NetUtils.connect(NetUtils.java:600)
at org.apache.hadoop.hdfs.DFSClient.newConnectedPeer(DFSClient.java:3033)
at org.apache.hadoop.hdfs.client.impl.BlockReaderFactory.nextTcpPeer(BlockReaderFactory.java
9)
at org.apache.hadoop.hdfs.client.impl.BlockReaderFactory.getRemoteBlockReaderFromTcp(BlockRe
rFactory.java:754)
at org.apache.hadoop.hdfs.client.impl.BlockReaderFactory.build(BlockReaderFactory.java:381)
at org.apache.hadoop.hdfs.DFSInputStream.getBlockReader(DFSInputStream.java:715)
at org.apache.hadoop.hdfs.DFSInputStream.blockSeekTo(DFSInputStream.java:645)
at org.apache.hadoop.hdfs.DFSInputStream.readWithStrategy(DFSInputStream.java:845)
at org.apache.hadoop.hdfs.DFSInputStream.read(DFSInputStream.java:918)
at java.io.DataInputStream.read(DataInputStream.java:100)
at org.apache.hadoop.io.IOUtils.copyBytes(IOUtils.java:96)
at org.apache.hadoop.io.IOUtils.copyBytes(IOUtils.java:69)
at org.apache.hadoop.io.IOUtils.copyBytes(IOUtils.java:133)
```


10: the output

```
at org.apache.hadoop.hdfs.DFSInputStream.read(DFSInputStream.java:918)
at java.io.DataInputStream.read(DataInputStream.java:100)
at org.apache.hadoop.io.IOUtils.copyBytes(IOUtils.java:96)
at org.apache.hadoop.io.IOUtils.copyBytes(IOUtils.java:69)
at org.apache.hadoop.io.IOUtils.copyBytes(IOUtils.java:133)
at org.apache.hadoop.fs.shell.Display$Cat.printToStdout(Display.java:101)
at org.apache.hadoop.fs.shell.Display$Cat.processPath(Display.java:96)
at org.apache.hadoop.fs.shell.Command.processPathInternal(Command.java:382)
at org.apache.hadoop.fs.shell.Command.processPaths(Command.java:345)
at org.apache.hadoop.fs.shell.Command.processPathArgument(Command.java:318)
at org.apache.hadoop.fs.shell.Command.processArgument(Command.java:300)
at org.apache.hadoop.fs.shell.Command.processArguments(Command.java:284)
at org.apache.hadoop.fs.shell.FsCommand.processRawArguments(FsCommand.java:121)
at org.apache.hadoop.fs.shell.Command.run(Command.java:191)
at org.apache.hadoop.fs.FsShell.run(FsShell.java:327)
at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:82)
at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:97)
at org.apache.hadoop.fs.FsShell.main(FsShell.java:390)
cat: Could not obtain block: BP-1553536198-172.28.1.2-1758975076798:blk_1073742077_1253 file=/file2.
txt No live nodes contain current block Block locations: DatanodeInfoWithStorage[172.28.1.5:9866,DS-
0043516a-f3cc-43e9-97bc-41159dd52ed2,DISK] DatanodeInfoWithStorage[172.28.1.4:9866,DS-604723ab-f6d8-
462b-83a7-2cfd83663caf,DISK] DatanodeInfoWithStorage[172.28.1.3:9866,DS-708e7b5a-1e02-4423-a510-673a
2961d0ea,DISK] Dead nodes: DatanodeInfoWithStorage[172.28.1.5:9866,DS-0043516a-f3cc-43e9-97bc-41159
dd52ed2,DISK] DatanodeInfoWithStorage[172.28.1.3:9866,DS-708e7b5a-1e02-4423-a510-673a2961d0ea,DISK]
DatanodeInfoWithStorage[172.28.1.4:9866,DS-604723ab-f6d8-462b-83a7-2cfd83663caf,DISK]
jupyter@master:/$ |
```

10: stop the master

```
Sep 28 02:02
h@ja:~$ docker ps
CONTAINER ID        IMAGE               COMMAND             CREATED             STATUS              PORTS
8ae7f572f5e6        hadoop-hive-spark-worker  "entrypoint.sh run.sh"  11 hours ago       Up 22 minutes      0.0.0.0:8043->8042/tcp, [::]:8043->8042/tcp, 0.0.0.0:8082->8081/tcp, [::]:8082->8081/tcp, 0.0.0.0:9865->9864/tcp, [::]:9865->9864/tcp
park-docker-master-worker2-1
9017ed9d456e        hadoop-hive-spark-worker  "entrypoint.sh run.sh"  11 hours ago       Up 22 minutes      0.0.0.0:8044->8042/tcp, [::]:8044->8042/tcp, 0.0.0.0:8083->8081/tcp, [::]:8083->8081/tcp, 0.0.0.0:9866->9864/tcp, [::]:9866->9864/tcp
park-docker-master-worker3-1
20e7ac3e37a7        hadoop-hive-spark-master  "entrypoint.sh run.sh"  11 hours ago       Up 22 minutes      0.0.0.0:4040->4040/tcp, [::]:4040->4040/tcp, 0.0.0.0:7077->7077/tcp, [::]:7077->7077/tcp, 0.0.0.0:8020->8020/tcp, [::]:8020->8020/tcp, 0.0.0.0:8080->8080/tcp, [::]:8080->8080/tcp, 0.0.0.0:8088->8088/tcp, [::]:8088->8088/tcp, 0.0.0.0:9083->9083/tcp, [::]:9083->9083/tcp, 0.0.0.0:9870->9870/tcp, [::]:9870->9870/tcp, 0.0.0.0:10000->10000/tcp, [::]:10000->10000/tcp
park-docker-master-master-1
08546d4e68e9        postgres:11          "docker-entrypoint.s..."  11 hours ago       Up 22 minutes      0.0.0.0:15432->5432/tcp
park-docker-master-metastore-1
h@ja:~$ docker stop hadoop-hive-spark-docker-master-m
hadoop-hive-spark-docker-master-master-1      hadoop-hive-spark-docker-master-metastore-1
h@ja:~$ docker stop hadoop-hive-spark-docker-master-m
hadoop-hive-spark-docker-master-master-1      hadoop-hive-spark-docker-master-metastore-1
h@ja:~$ docker stop hadoop-hive-spark-docker-master-master-1
```

11: try any command, no response

```
Sep 28 02:02
jupyter@worker2:/

heja:~$ docker ps
CONTAINER ID   IMAGE                                COMMAND                  CREATED        STATUS        PORTS
NAMES
8ae7f572f5e6   hadoop-hive-spark-worker            "entrypoint.sh run.sh"  11 hours ago  Up 22 minutes  0.0.0.0:8043->8042/tcp, [::]:8043->8042/tcp, 0.0
.0.0:8082->8081/tcp, [::]:8082->8081/tcp, 0.0.0.0:9865->9864/tcp, [::]:9865->9864/tcp
hadoop-hive-spark-docker-master-worker2-1
9017ed9d456e   hadoop-hive-spark-worker            "entrypoint.sh run.sh"  11 hours ago  Up 22 minutes  0.0.0.0:8044->8042/tcp, [::]:8044->8042/tcp, 0.0
.0.0:8083->8081/tcp, [::]:8083->8081/tcp, 0.0.0.0:9866->9864/tcp, [::]:9866->9864/tcp
hadoop-hive-spark-docker-master-worker3-1
20e7ac3e37a7   hadoop-hive-spark-master            "entrypoint.sh run.sh"  11 hours ago  Up 22 minutes  0.0.0.0:4040->4040/tcp, [::]:4040->4040/tcp, 0.0
.0.0:7077->7077/tcp, [::]:7077->7077/tcp, 0.0.0.0:8020->8020/tcp, [::]:8020->8020/tcp, 0.0.0.0:8080->8080/tcp, [::]:8080->8080/tcp, 0.0.0.0:8088->
8088/tcp, [::]:8088->8088/tcp, 0.0.0.0:9083->9083/tcp, [::]:9083->9083/tcp, 0.0.0.0:9870->9870/tcp, [::]:9870->9870/tcp, 0.0.0.0:10000->10000/tcp,
[::]:10000->10000/tcp
hadoop-hive-spark-docker-master-master-1
08546d4e68e9   postgres:11                        "docker-entrypoint.s..."  11 hours ago  Up 22 minutes  0.0.0.0:15432->5432/tcp, [::]:15432->5432/tcp
hadoop-hive-spark-docker-master-metastore-1
heja:~$ docker stop hadoop-hive-spark-docker-master-m
hadoop-hive-spark-docker-master-master-1 hadoop-hive-spark-docker-master-metastore-1
heja:~$ docker stop hadoop-hive-spark-docker-master-m
hadoop-hive-spark-docker-master-master-1 hadoop-hive-spark-docker-master-metastore-1
heja:~$ docker stop hadoop-hive-spark-docker-master-master-1
hadoop-hive-spark-docker-master-master-1
heja:~$ docker exec -it hadoop-hive-spark-docker-master-worker-1 bash
hadoop-hive-spark-docker-master-worker2-1 hadoop-hive-spark-docker-master-worker3-1
heja:~$ docker exec -it hadoop-hive-spark-docker-master-worker2-1 bash
jupyter@worker2:/# ls
bin  dev  home  lib64  mnt  proc  run  srv  tmp  var
boot  etc  lib  media  opt  root /sbin  sys  usr
jupyter@worker2:/# hdfs dfs -ls /
ls: No Route to Host from worker2/172.28.1.4 to master:8020 failed on socket timeout exception: java.net.NoRouteToHostException: No route to host
; For more details see: http://wiki.apache.org/hadoop/NoRouteToHost
jupyter@worker2:/#
```