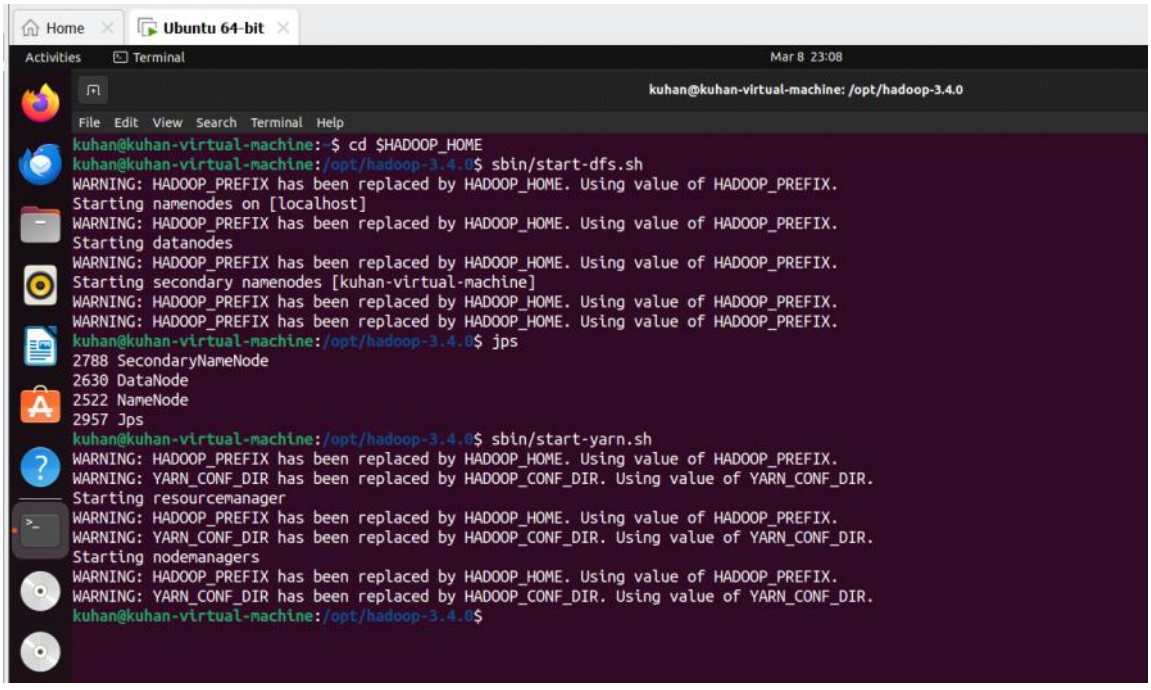
BigData Assignment 2

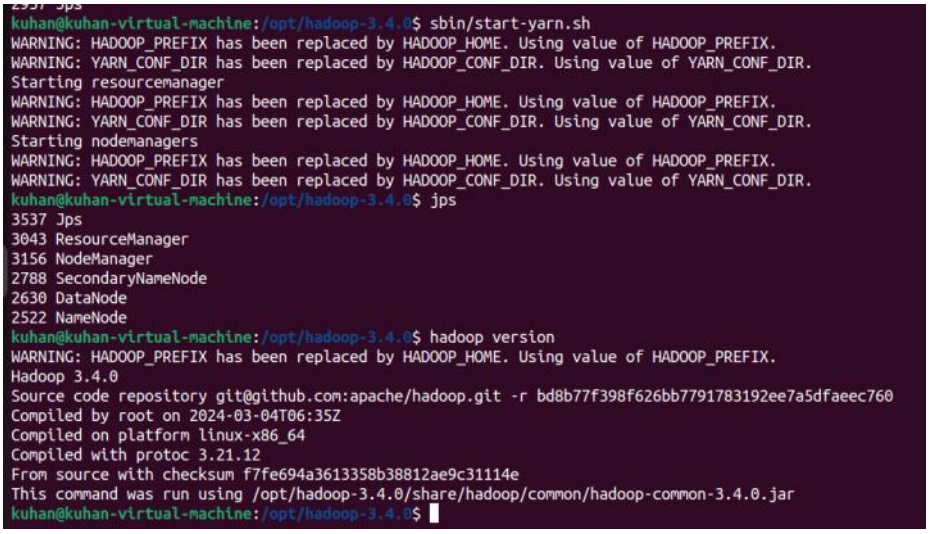
08 March 2025 23:05

Submitted by Kuhan
Roll No. 2024MCS110028

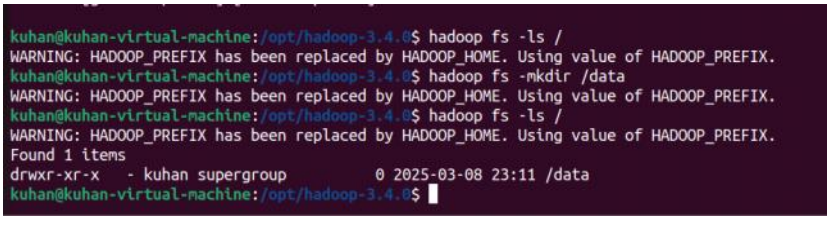
1. Start HDFS and YARN



2. Verify Running Services



List directory and files in hadoop and create a directory data



Create dummy file for uploading :

```
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ cat >file.txt <<EOF
> line 1: this a test file.
> line 2: Hadoop file system demo
> line 3: Storing files in HDFS
> EOF
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ echo "Hello, this is a sample file for Hadoop testing" >> file.txt
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ cat file.txt
line 1: this a test file.
line 2: Hadoop file system demo
line 3: Storing files in HDFS
Hello, this is a sample file for Hadoop testing
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$
```

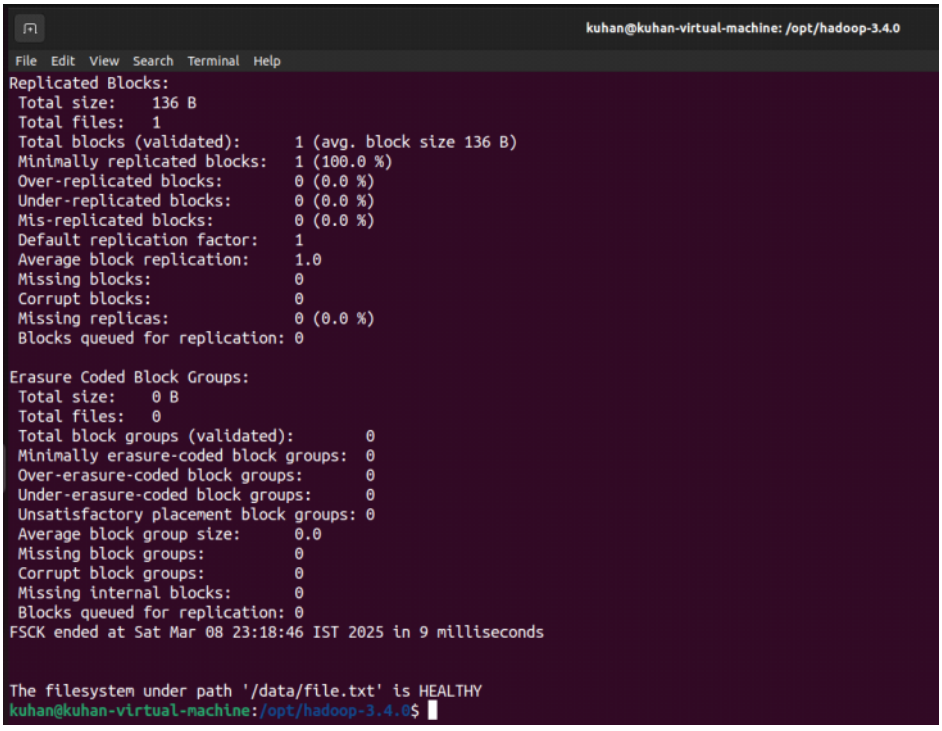
Put the file In hadoop and check the file is uploaded correctly or not

```
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ cat file.txt
line 1: this a test file.
line 2: Hadoop file system demo
line 3: Storing files in HDFS
Hello, this is a sample file for Hadoop testing
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hadoop fs -put file.txt /data/
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hadoop fs -ls /
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Found 1 items
drwxr-xr-x   - kuhan supergroup          0 2025-03-08 23:16 /data
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hadoop fs -ls /data
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Found 1 items
-rw-r--r--   1 kuhan supergroup        136 2025-03-08 23:16 /data/file.txt
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hadoop fs -cat /data/file.txt
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
line 1: this a test file.
line 2: Hadoop file system demo
line 3: Storing files in HDFS
Hello, this is a sample file for Hadoop testing
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$
```

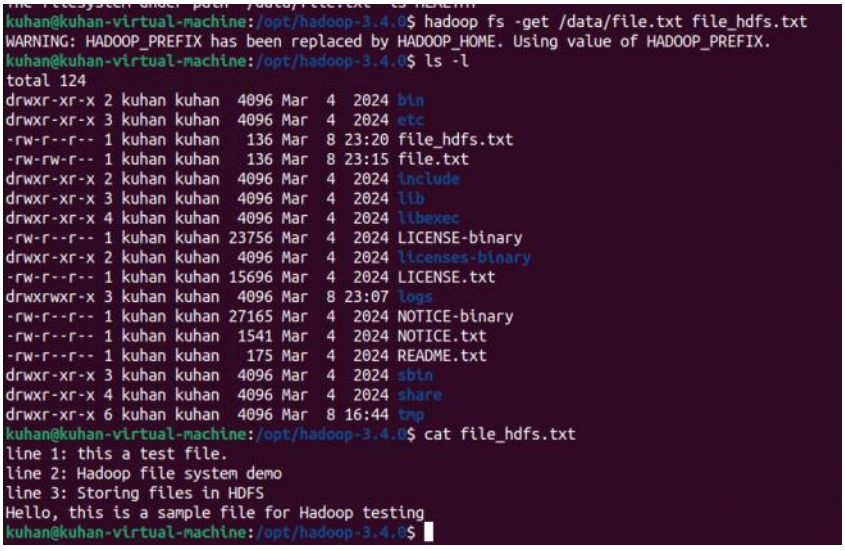
Verify file size and replication
And also check hdfs file system health

```
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hadoop fs -du -h /data/file.txt
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
136  136  /data/file.txt
```

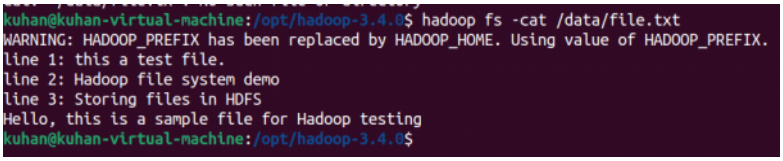
```
kuhan@kuhan-virtual-machine: /opt/hadoop-3.4.0
File  Edit  View  Search  Terminal  Help
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hdfs fsck /data/file.txt
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Connecting to namenode via http://localhost:9870/fsck?ugi=kuhan&path=%2Fdata%2Ffile.txt
FSCK started by kuhan (auth:SIMPLE) from /127.0.0.1 for path /data/file.txt at Sat Mar 08 23:18:46 IST 2025


Status: HEALTHY
 Number of data-nodes:  1
 Number of racks:               1
 Total dirs:                    0
 Total symlinks:                0

Replicated Blocks:
 Total size:    136 B
 Total files:   1
 Total blocks (validated):      1 (avg. block size 136 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:    1
 Average block replication:     1.0
 Missing blocks:                0
 Corrupt blocks:                0
 Missing replicas:              0 (0.0 %)
 Blocks queued for replication: 0

Erasure Coded Block Groups:
 Total size:    0 B
 Total files:   0
 Total block groups (validated):        0
 Minimally erasure-coded block groups:  0
 Over-erasure-coded block groups:       0
```



Download file and rename it



Display the file content



Create another directory "backup"
And move the file from data directory to backup directory

```
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hadoop fs -ls /
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Found 2 items
drwxr-xr-x   - kuhan supergroup          0 2025-03-08 23:22 /backup
drwxr-xr-x   - kuhan supergroup          0 2025-03-08 23:16 /data
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hadoop fs -mv /data/file.txt /backup/
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hadoop fs -ls /backup
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Found 1 items
-rw-r--r--   1 kuhan supergroup        136 2025-03-08 23:16 /backup/file.txt
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hadoop fs -ls /data
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hadoop fs -ls /
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Found 2 items
drwxr-xr-x   - kuhan supergroup          0 2025-03-08 23:23 /backup
drwxr-xr-x   - kuhan supergroup          0 2025-03-08 23:23 /data
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$
```

Copy files within hdfs

```
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hadoop fs -cp /backup/file.txt /data/file_copy.txt
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hadoop fs -ls /data
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Found 1 items
-rw-r--r--   1 kuhan supergroup        136 2025-03-08 23:25 /data/file_copy.txt
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hadoop fs -ls /backup
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Found 1 items
-rw-r--r--   1 kuhan supergroup        136 2025-03-08 23:16 /backup/file.txt
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$
```

Show last 2 lines of the file

```
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hadoop fs -cat /data/file_copy.txt | tail -n 2
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
line 3: Storing files in HDFS
Hello, this is a sample file for Hadoop testing
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$
```

4. Remove Files and Directories

```
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hadoop fs -rm /data/file_copy.txt
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Deleted /data/file_copy.txt
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hadoop fs -ls /data
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hadoop fs -ls /
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Found 2 items
drwxr-xr-x   - kuhan supergroup          0 2025-03-08 23:23 /backup
drwxr-xr-x   - kuhan supergroup          0 2025-03-08 23:28 /data
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hadoop fs -rm -r /backup/
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Deleted /backup
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hadoop fs -ls /
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Found 1 items
drwxr-xr-x   - kuhan supergroup          0 2025-03-08 23:28 /data
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$
```

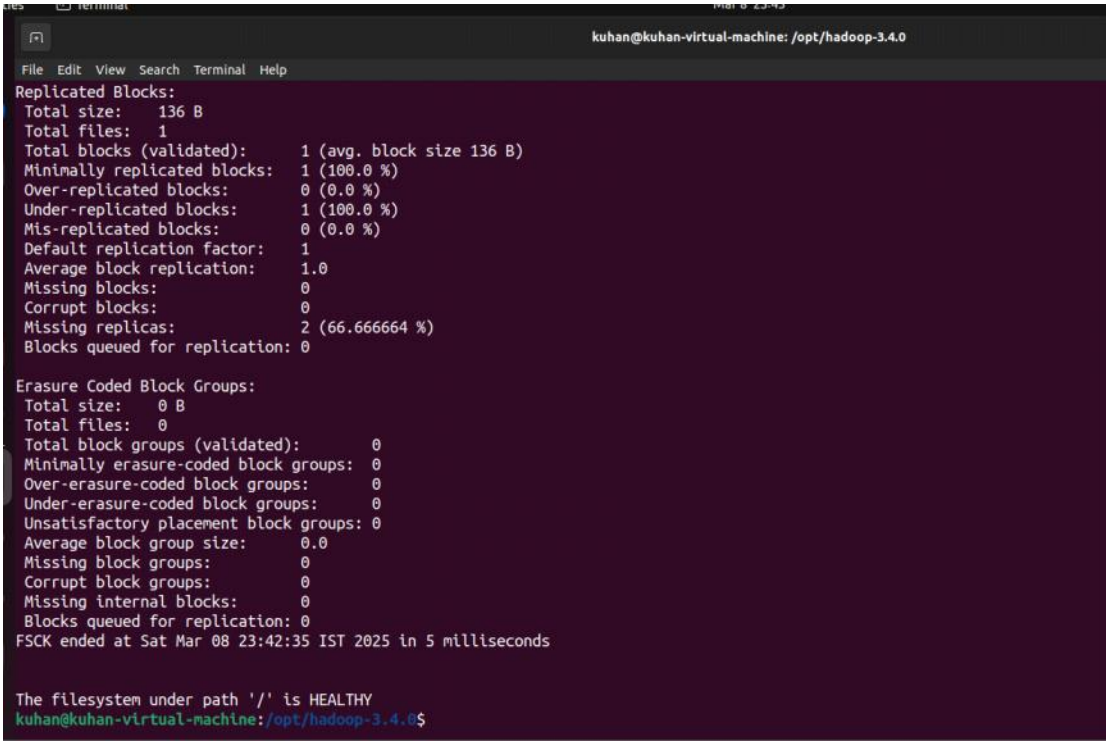
Set replication factor

```
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hadoop fs -setrep -w 3 /data/file.txt
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Replication 3 set: /data/file.txt
Waiting for /data/file.txt ...................................................^Ckuh
```

Check active data nodes

```
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hdfs dfsadmin -report
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Configured Capacity: 36139761664 (33.66 GB)
Present Capacity: 14784696320 (13.77 GB)
DFS Remaining: 14784647168 (13.77 GB)
DFS Used: 49152 (48 KB)
DFS Used%: 0.00%
Replicated Blocks:
        Under replicated blocks: 1
        Blocks with corrupt replicas: 0
        Missing blocks: 0
        Missing blocks (with replication factor 1): 0
        Low redundancy blocks with highest priority to recover: 1
        Pending deletion blocks: 0
Erasure Coded Block Groups:
        Low redundancy block groups: 0
        Block groups with corrupt internal blocks: 0
        Missing block groups: 0
        Low redundancy blocks with highest priority to recover: 0
        Pending deletion blocks: 0

-------------------------------------------------
Live datanodes (1):
```

```
kuhan@kuhan-virtual-machine: /opt/hadoop-3.4.0
File  Edit  View  Search  Terminal  Help
        Low redundancy block groups: 0
        Block groups with corrupt internal blocks: 0
        Missing block groups: 0
        Low redundancy blocks with highest priority to recover: 0
        Pending deletion blocks: 0

-------------------------------------------------
Live datanodes (1):

Name: 127.0.0.1:9866 (localhost)
Hostname: kuhan-virtual-machine
Decommission Status : Normal
Configured Capacity: 36139761664 (33.66 GB)
DFS Used: 49152 (48 KB)
Non DFS Used: 19486294016 (18.15 GB)
DFS Remaining: 14784647168 (13.77 GB)
DFS Used%: 0.00%
DFS Remaining%: 40.91%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 0
Last contact: Sat Mar 08 23:40:19 IST 2025
Last Block Report: Sat Mar 08 23:07:40 IST 2025
Num of Blocks: 1

kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hadoop fs -du -h /data
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
136  408  /data/file.txt
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hadoop fs -df -h
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Filesystem                Size  Used  Available  Use%
hdfs://localhost:9000  33.7 G  48 K     13.8 G    0%
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hdfs fsck /
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Connecting to namenode via http://localhost:9870/fsck?ugi=kuhan&path=%2F
FSCK started by kuhan (auth:SIMPLE) from /127.0.0.1 for path / at Sat Mar 08 23:42:35 IST 2025


/data/file.txt:  Under replicated BP-343335988-127.0.1.1-1741432315287:blk_1073741827_1003. Target Replicas is 3 but found 1 live replica(s), 0 decommissioned repl
ica(s), 0 decommissioning replica(s).

Status: HEALTHY
 Number of data-nodes:  1
 Number of racks:               1
 Total dirs:                    2
 Total symlinks:                0

Replicated Blocks:
```


kuhan@kuhan-virtual-machine: /opt/hadoop-3.4.0
File Edit View Search Terminal Help
Replicated Blocks:
Total size: 136 B
Total files: 1
Total blocks (validated): 1 (avg. block size 136 B)
Minimally replicated blocks: 1 (100.0 %)
Over-replicated blocks: 0 (0.0 %)
Under-replicated blocks: 1 (100.0 %)
Mis-replicated blocks: 0 (0.0 %)
Default replication factor: 1
Average block replication: 1.0
Missing blocks: 0
Corrupt blocks: 0
Missing replicas: 2 (66.666664 %)
Blocks queued for replication: 0
Erasure Coded Block Groups:
Total size: 0 B
Total files: 0
Total block groups (validated): 0
Minimally erasure-coded block groups: 0
Over-erasure-coded block groups: 0
Under-erasure-coded block groups: 0
Unsatisfactory placement block groups: 0
Average block group size: 0.0
Missing block groups: 0
Corrupt block groups: 0
Missing internal blocks: 0
Blocks queued for replication: 0
FSCK ended at Sat Mar 08 23:42:35 IST 2025 in 5 milliseconds
The filesystem under path '/' is HEALTHY
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$