1. Create a new HDFS file structure data/dev/analysis in the HDFS home directory. Copy the covid-19 folder in to it. [mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -mkdir -p /data/dev/analysis

[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -ls

Found 12

items

drwx------ - mavricbdhoct07 mavricbdhoct07 0 2023-10-21 07:00 .Trash

drwx------ - mavricbdhoct07 mavricbdhoct07 0 2023-10-26 09:35 .staging

drwxr-xr-x - mavricbdhoct07 mavricbdhoct07 0 2023-10-29 11:24 data

drwxr-xr-x - mavricbdhoct07 mavricbdhoct07 0 2023-10-19 10:34 file1.txt

drwxr-xr-x - mavricbdhoct07 mavricbdhoct07 0 2023-10-19 10:34 file2.txt

drwxr-xr-x - mavricbdhoct07 mavricbdhoct07 0 2023-10-19 10:34 file3.txt

drwxr-xr-x - mavricbdhoct07 mavricbdhoct07 0 2023-10-20 07:21 project1

drwxr-xr-x - mavricbdhoct07 mavricbdhoct07 0 2023-10-20 07:23 project2

-rw-r--r-- 3 mavricbdhoct07 mavricbdhoct07 33 2023-10-20 07:28 project2hdfs

drwxr-xr-x - mavricbdhoct07 mavricbdhoct07 0 2023-10-19 10:32 project3

drwxr-xr-x - mavricbdhoct07 mavricbdhoct07 0 2023-10-26 07:52 savita\_userratings

drwxrwxrwx - mavricbdhoct07 mavricbdhoct07 0 2023-10-20 07:04 test

[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -ls data

Found 1 items

drwxr-xr-x - mavricbdhoct07 mavricbdhoct07 0 2023-10-29 11:24 data/dev

[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -ls data/dev

Found 1 items

drwxr-xr-x - mavricbdhoct07 mavricbdhoct07 0 2023-10-29 11:24 data/dev/analysis

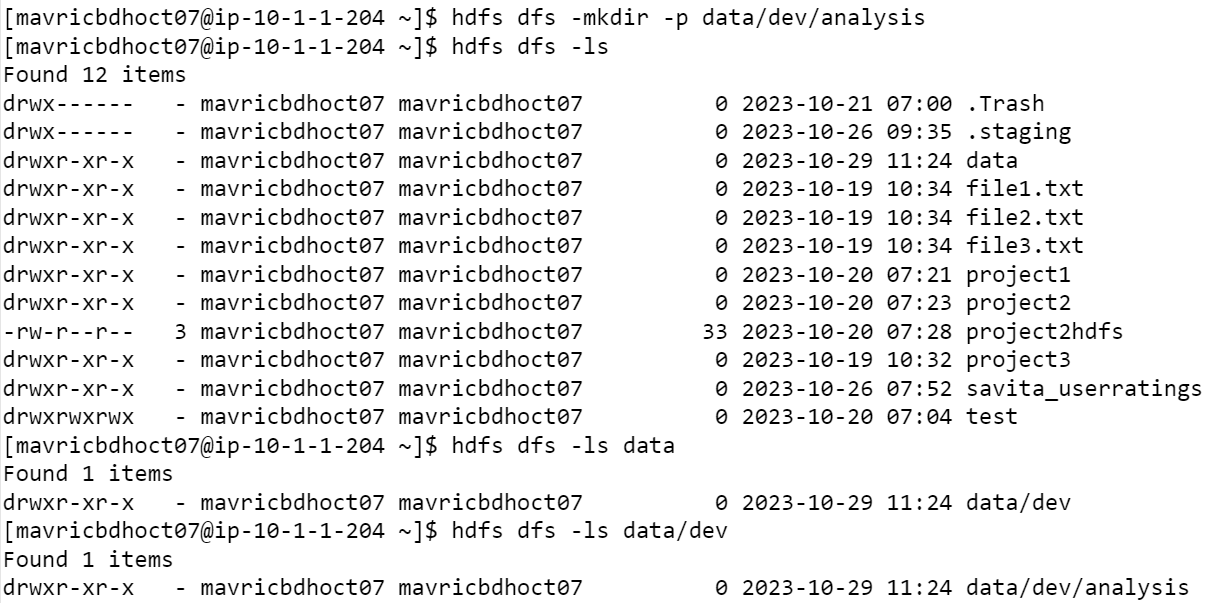
**Copy the covid-19 folder in to it**

[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -put codid-19 /user/mavricbdhoct07/data/dev/analysis

[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -ls data/dev/analysis

Found 1 items

drwxr-xr-x - mavricbdhoct07 mavricbdhoct07 0 2023-10-29 11:54 data/dev/analysis/codid-19



1. Change the Permission of Covid\_patient\_Apr26.csv file to have only read/write/execute for the owner/group and read and write permission for other

[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -chmod -R 777 /user/mavricbdhoct07/data/dev/analysis/ codid-19/codid-19/

[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -ls /user/mavricbdhoct07/data/dev/analysis/codid-19/codid-19/

Found 3 items

-rw-r--r-- 3 mavricbdhoct07 mavricbdhoct07 2878318 2023-10-29 12:03 /user/mavricbdhoct07/data/dev/analysis/codid-19/codid-19/Covid\_pat

ient\_Apr19.csv

-rw-r--r-- 3 mavricbdhoct07 mavricbdhoct07 1576844 2023-10-29 12:03 /user/mavricbdhoct07/data/dev/analysis/codid-19/codid-19/Covid\_pat

ient\_Apr26.csv

-rw-r--r-- 3 mavricbdhoct07 mavricbdhoct07 3642061 2023-10-29 12:03 /user/mavricbdhoct07/data/dev/analysis/codid-19/codid-19/covid-cou

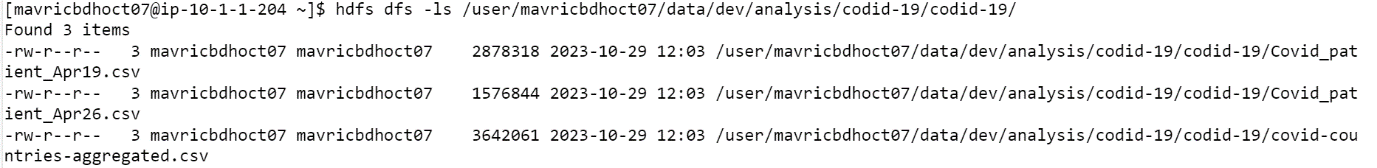
ntries-aggregated.csv

[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -ls

/user/mavricbdhoct07/data/dev/analysis/codid-19/codid-19/Covid\_patient\_Apr26.csv

-rw-r--r-- 3 mavricbdhoct07 mavricbdhoct07 1576844 2023-10-29 12:03 /user/mavricbdhoct07/data/dev/analysis/codid-19/codid-19/Covid\_pat

ient\_Apr26.csv



1. Copy Covid-19 folder into another hdfs directory data/proud/analysis

Created another hdfs directory data/proud/analysis

[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -ls data

Found 2 items

drwxr-xr-x - mavricbdhoct07 mavricbdhoct07 0 2023-10-29 11:24 data/dev

drwxr-xr-x - mavricbdhoct07 mavricbdhoct07 0 2023-10-29 12:23 data/prod

Copy Covid-19 folder into another hdfs directory data/proud/analysis

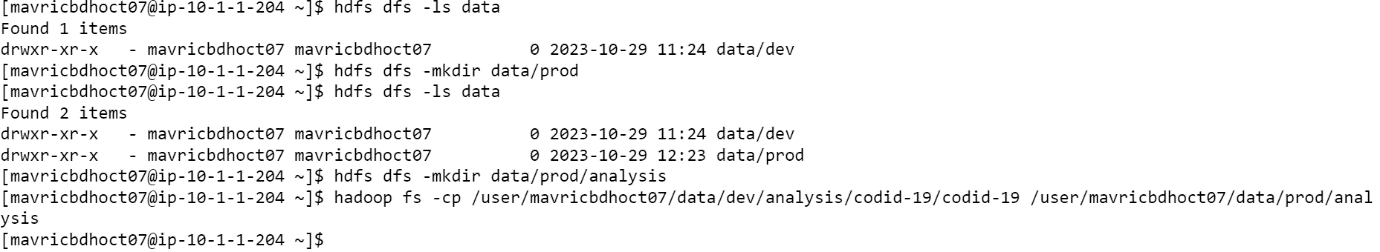
[mavricbdhoct07@ip-10-1-1-204 ~]$ hadoop fs -cp /user/mavricbdhoct07/data/dev/analysis/codid-19/codid-19 /user/mavricbdhoct07/data/prod/anal

ysis

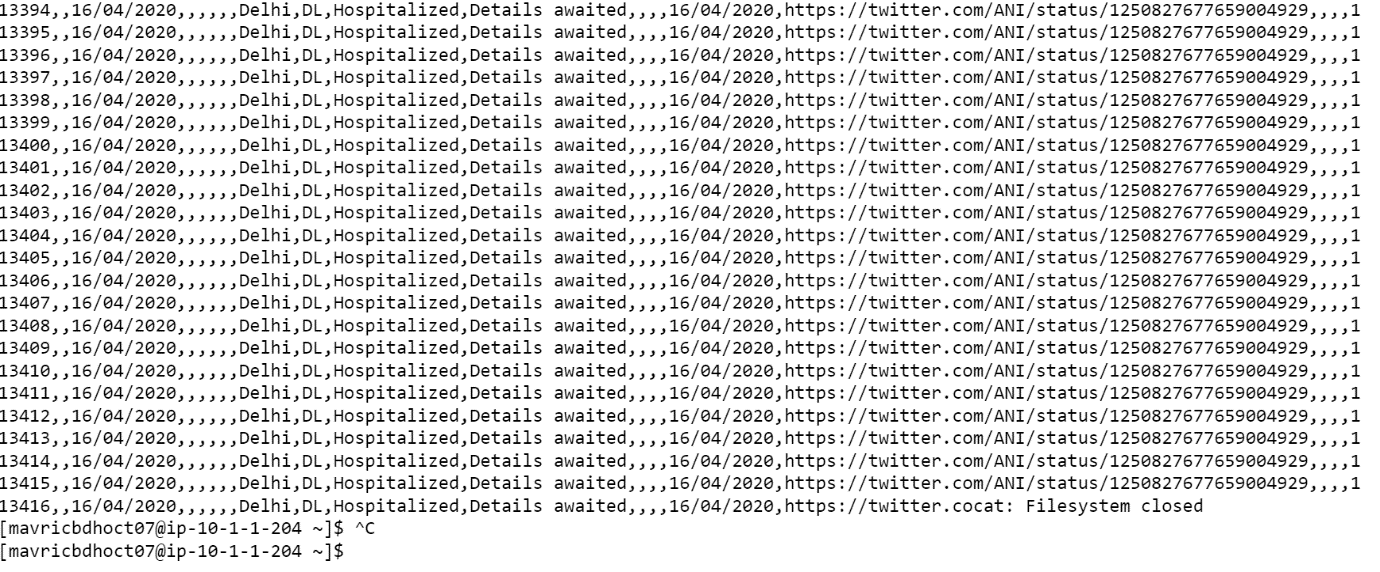
[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -ls data/prod/analysis

Found 1 items

drwxr-xr-x - mavricbdhoct07 mavricbdhoct07 0 2023-10-29 12:28 data/prod/analysis/codid-19



1. Read the contents of the file Covid\_patient\_Apr19.csv and print the content on screen

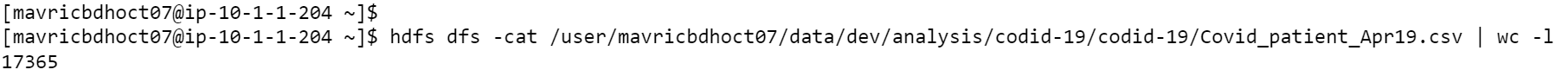


5)Print the no of lines present in the HDFS file Covid\_patient\_Apr19.csv

[mavricbdhoct01233@ip-10-1-1-204 ~]$ hdfs dfs -cat /user/mavricbdhoct07/data/dev/analysis/codid-19/codid-19/Covid\_patient\_Apr19.csv | head -n 10

[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -cat /user/mavricbdhoct07/data/dev/analysis/codid-19/codid-19/Covid\_patient\_Apr19.csv | wc -l

17365

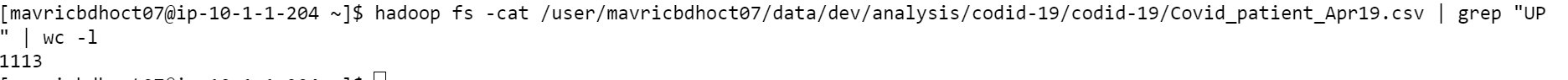


1. Count the no of records present in the HDFS file Covid\_patient\_Apr19.csv for the state “Up”

[mavricbdhoct07@ip-10-1-1-204 ~]$ hadoop fs -cat /user/mavricbdhoct07/data/dev/analysis/codid-19/codid-19/Covid\_patient\_Apr19.csv | grep "UP

" | wc -l

1113



1. Delete the file Covid\_patient\_Apr19.csv from data/proud/analysis folder in HDFS

Deleted Covid\_patient\_Apr19.csv file

[mavricbdhoct07@ip-10-1-1-204 ~]$ hadoop fs -rm /user/mavricbdhoct07/data/dev/analysis/codid-19/codid-19/Covid\_patient\_Apr19.csv

Remaining files in codid-19 are

[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -ls /user/mavricbdhoct07/data/dev/analysis/codid-19/codid-19

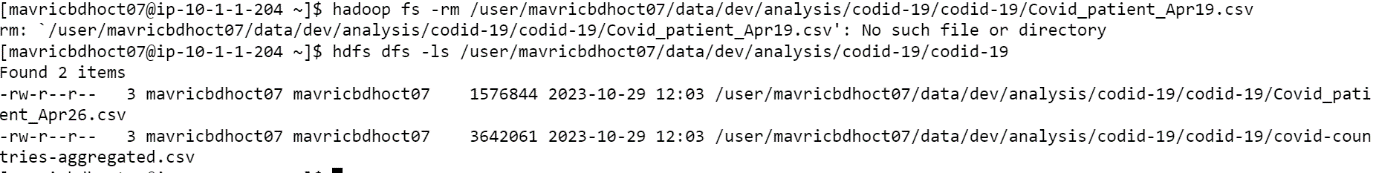
Found 2 items

-rw-r--r-- 3 mavricbdhoct07 mavricbdhoct07 1576844 2023-10-29 12:03 /user/mavricbdhoct07/data/dev/analysis/codid-19/codid-19/Covid\_pati

ent\_Apr26.csv

-rw-r--r-- 3 mavricbdhoct07 mavricbdhoct07 3642061 2023-10-29 12:03 /user/mavricbdhoct07/data/dev/analysis/codid-19/codid-19/covid-coun

tries-aggregated.csv



1. Delete the file covid-19 folder from both the dev and prod folders created above

Removed dev folder

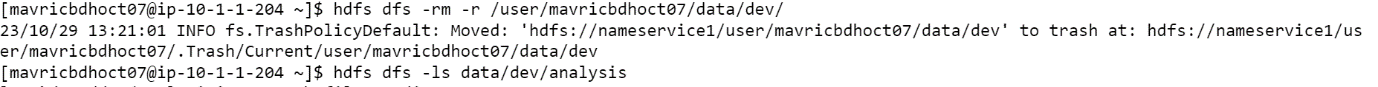
[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -rm -r /user/mavricbdhoct07/data/dev/

23/10/29 13:21:01 INFO fs.TrashPolicyDefault: Moved: 'hdfs://nameservice1/user/mavricbdhoct07/data/dev' to trash at: hdfs://nameservice1/us

er/mavricbdhoct07/.Trash/Current/user/mavricbdhoct07/data/dev

[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -ls data/dev/analysis

ls: `data/dev/analysis': No such file or directory



Removed prod folder

[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -rm -r /user/mavricbdhoct07/data/prod

23/10/29 13:23:06 INFO fs.TrashPolicyDefault: Moved: 'hdfs://nameservice1/user/mavricbdhoct07/data/prod' to trash at: hdfs://nameservice1/u

ser/mavricbdhoct07/.Trash/Current/user/mavricbdhoct07/data/prod

[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -ls data/prod/analysis

ls: `data/prod/analysis': No such file or directory



1. Copy the tickitdb folder (in the local file system)in to HDFS home directory. How many blockes will be created for the files in the directory

Created folder Savita\_T in hdfs file

[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -mkdir /user/mavricbdhoct07/savita\_T

Put the local tickitdb file to hdfs Savita\_T folder

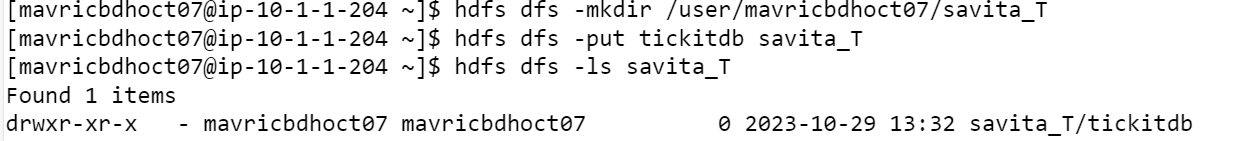
[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -put tickitdb savita\_T

tickitdb is present in hdfs home directory

[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -ls savita\_T

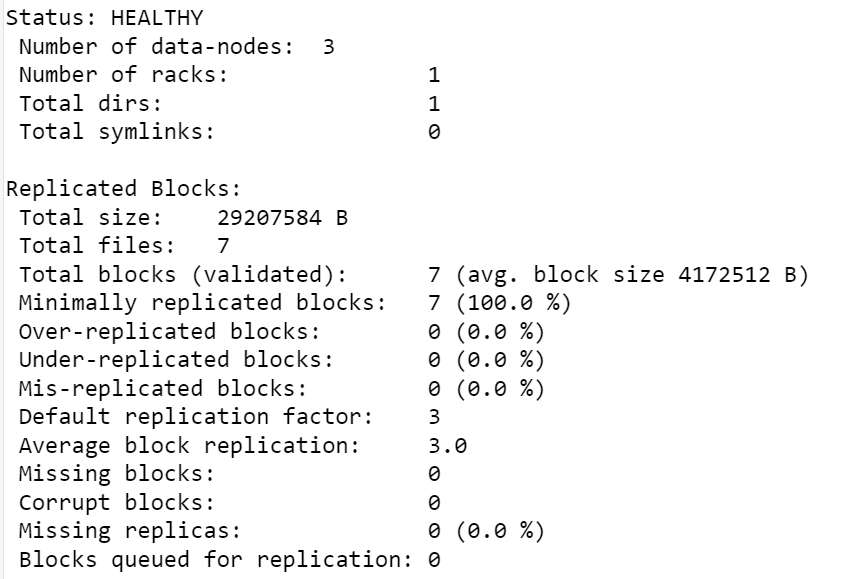
Found 1 items

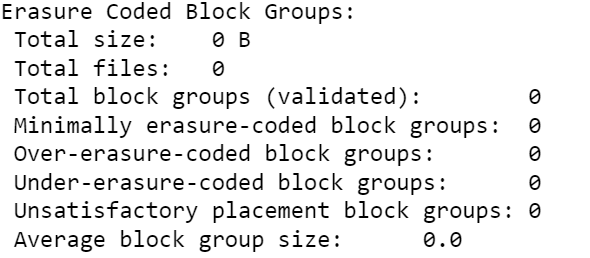
drwxr-xr-x - mavricbdhoct07 mavricbdhoct07 0 2023-10-29 13:32 savita\_T/tickitdb



**How many blockes will be created for the files in the directory**

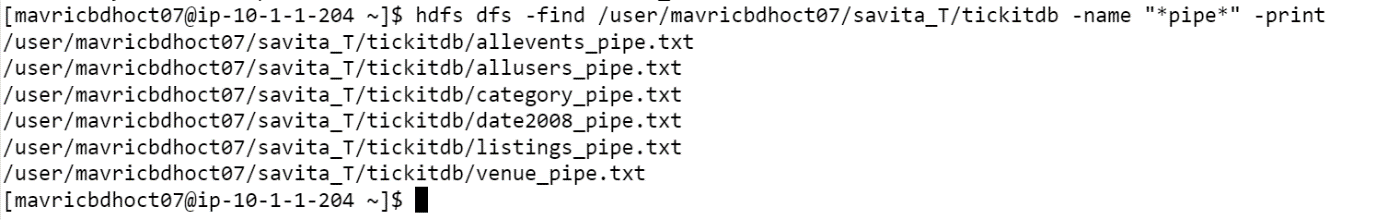


****

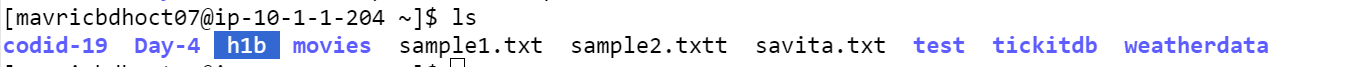
****

1. Find all the file name which has the word “pipe” in their name

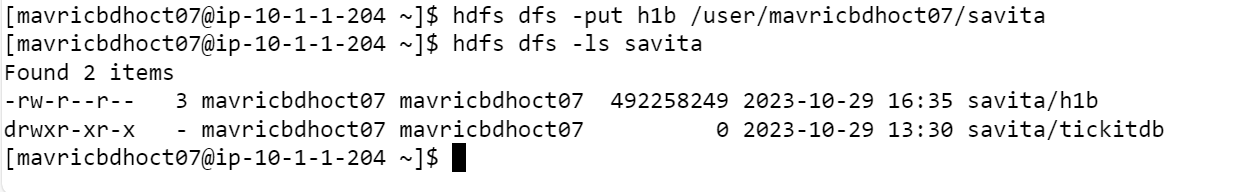
[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -find /user/mavricbdhoct07/savita\_T/tickitdb -name "\*pipe\*" -print



11) Copy the H1b.csv file to hdfs home directory. How many blocks will be created for this file

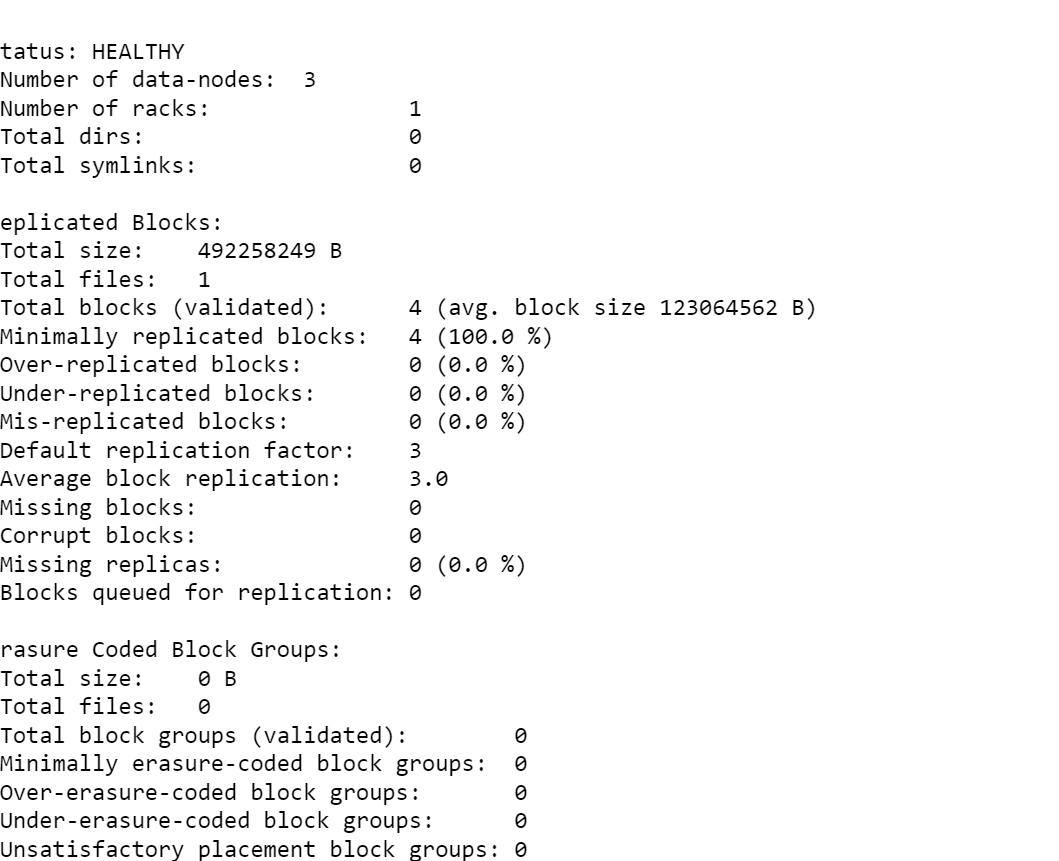


**Put h1b to hdfs Savita folder**



[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs fsck /user/mavricbdhoct07/savita/h1b -files -blocks



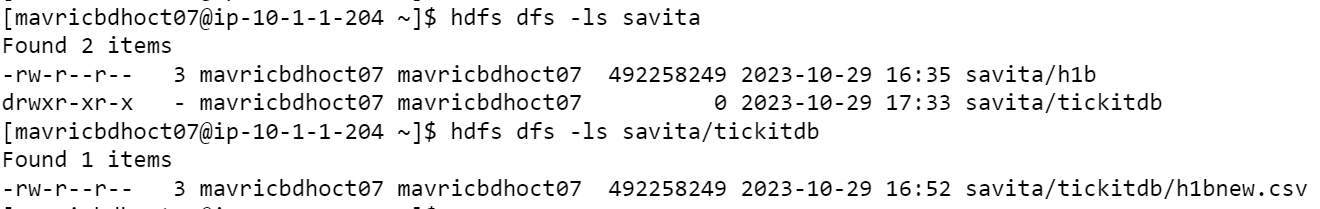


**12. Move the h1b.csv file (in the HDFS) to the tickitdb folder in hdfs. How many blocks will be there for the tickitdb folder?**

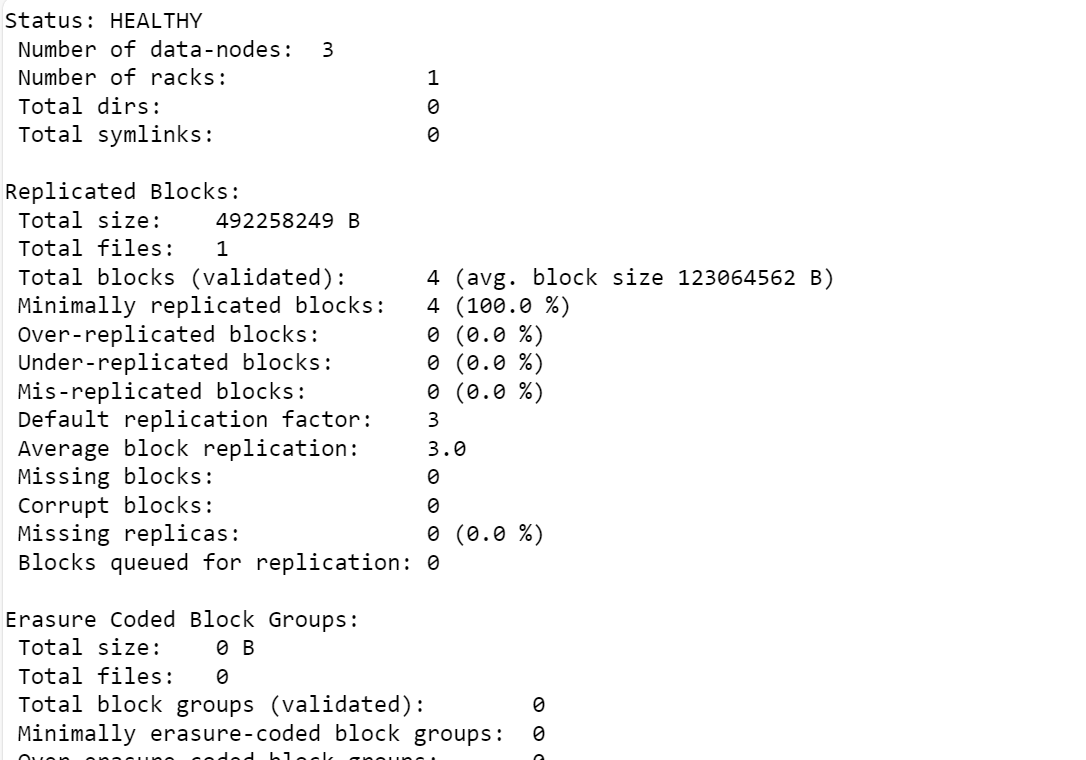
[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -put h1b.csv /user/mavricbdhoct07/Savita

[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -mv savita/h1b.csv savita/tickitdb/h1bnew.csv



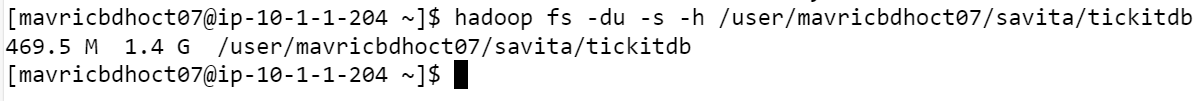


[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs fsck /user/mavricbdhoct07/savita/tickitdb/h1bnew.csv -files -blocks



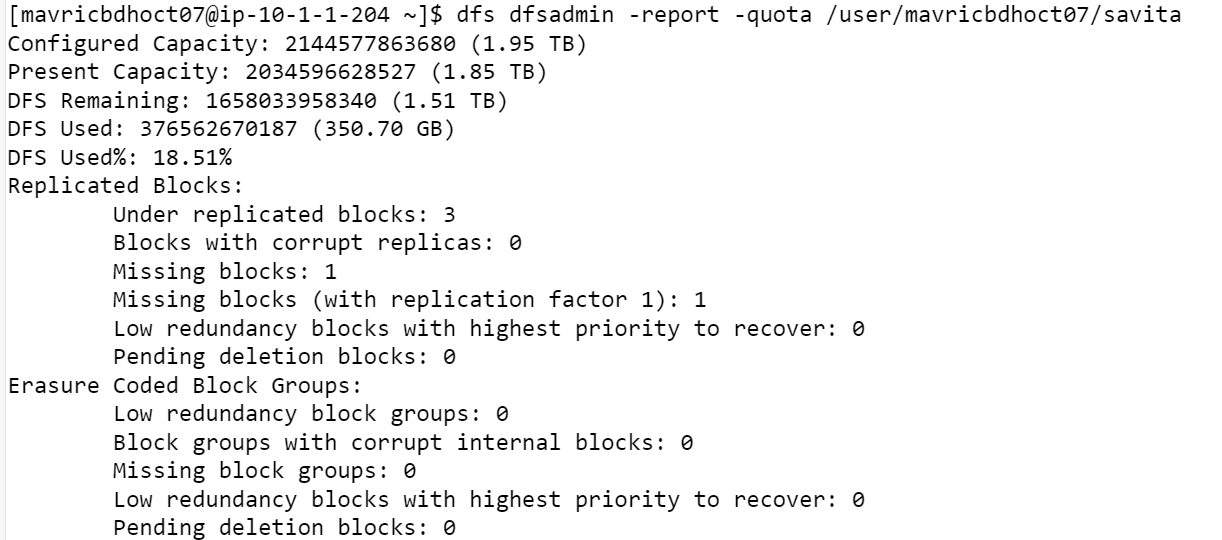
**13. Calculate the total storage space occupied by the files directory and its contents.**

[mavricbdhoct07@ip-10-1-1-204 ~]$ hadoop fs -du -s -h /user/mavricbdhoct07/savita/tickitdb



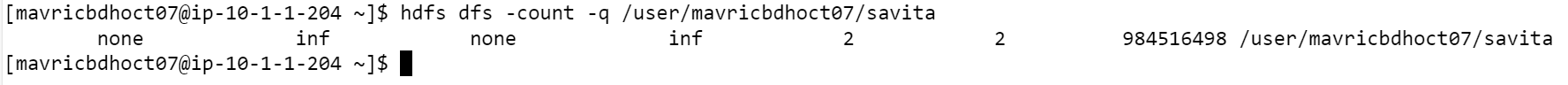
**14. Check the storage quota for the file directory.**

[mavricbdhoct07@ip-10-1-1-204 ~]$ dfs dfsadmin -report -quota /user/mavricbdhoct07/savita



Or

[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -count -q /user/mavricbdhoct07/Savita



**15. Delete a specific snapshot 'snapshot1' from the file directory.**

[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -deleteSnapshot /user/mavricbdhoct07/savita snapshot1

deleteSnapshot: `/user/mavricbdhoct07/savita': No such file or directory

**16. Create a snapshot of the file directory in HDFS.**

[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -createSnapshot /user/mavricbdhoct07/savita/ snapshot\_20231101

createSnapshot: Directory is not a snapshottable directory: /user/mavricbdhoct07/savita

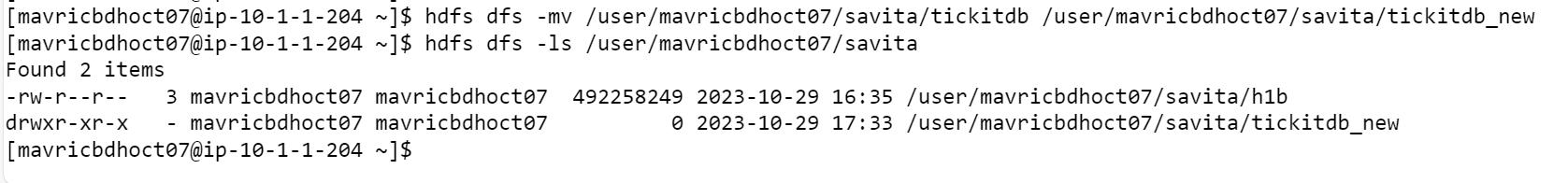
**17. Set a replication factor of 3 for all files within the file directory.**

[mavricbdhoct07@ip-10-1-1-204 ~]$ hadoop fs -setrep -w 3 /user/mavricbdhoct07/Savita



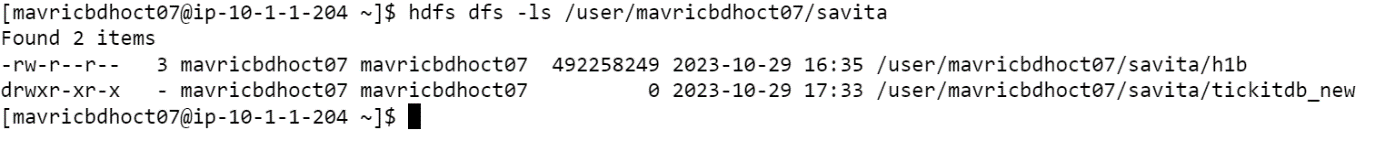
**18. Rename the 'file name' file to 'new file name' within the file directory in HDFS.**

[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -mv /user/mavricbdhoct07/savita/tickitdb /user/mavricbdhoct07/savita/tickitdb\_new



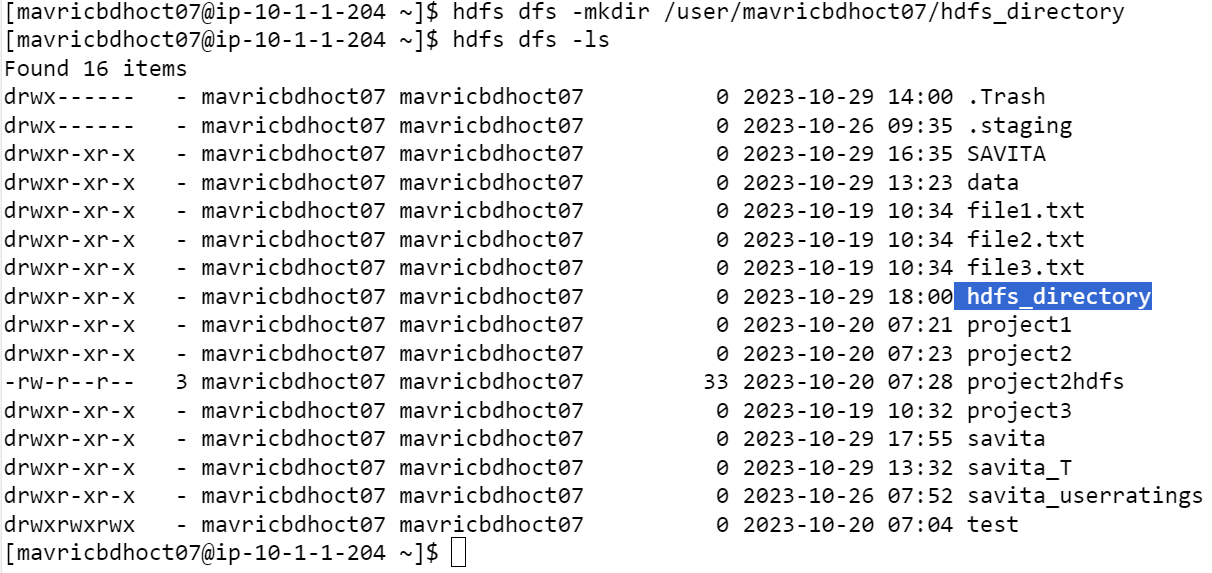
**19. List all the files and directories in the files directory in HDFS.**

[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -ls /user/mavricbdhoct07/Savita



**20. Create an HDFS directory file names in your home directory and upload a local file 'test.txt' to it.**

[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -mkdir /user/mavricbdhoct07/hdfs\_directory



[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -copyFromLocal /home/mavricbdhoct07/test.txt /user/mavricbdhoct07/hdfs\_directory

