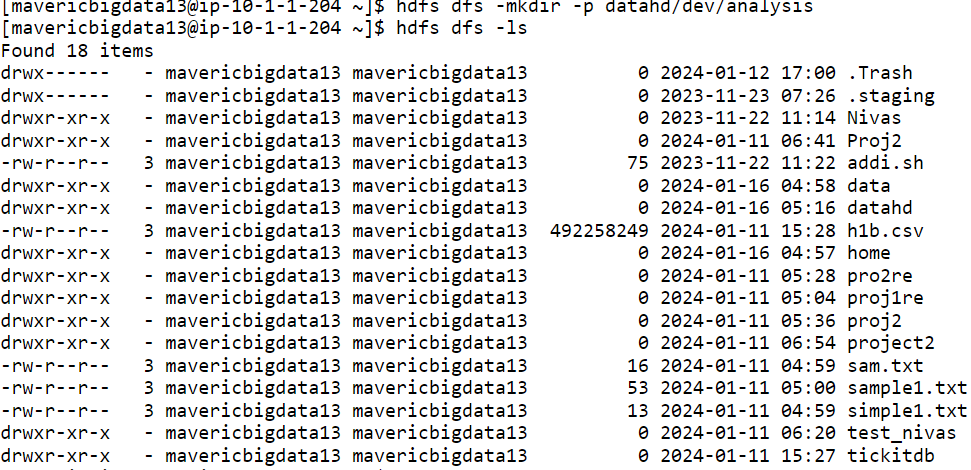
HDFS Assignment

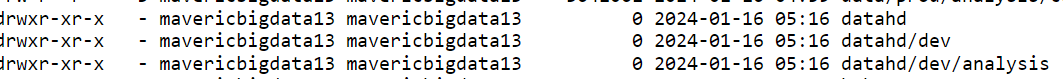
1. Create a new HDFS file structure data/dev/analysis in the HDFS home directory.

Copy the covid-19 folder in to it.

Hdfs dfs -mkdir-p data/dev/analysis



hdfs dfs -ls -R



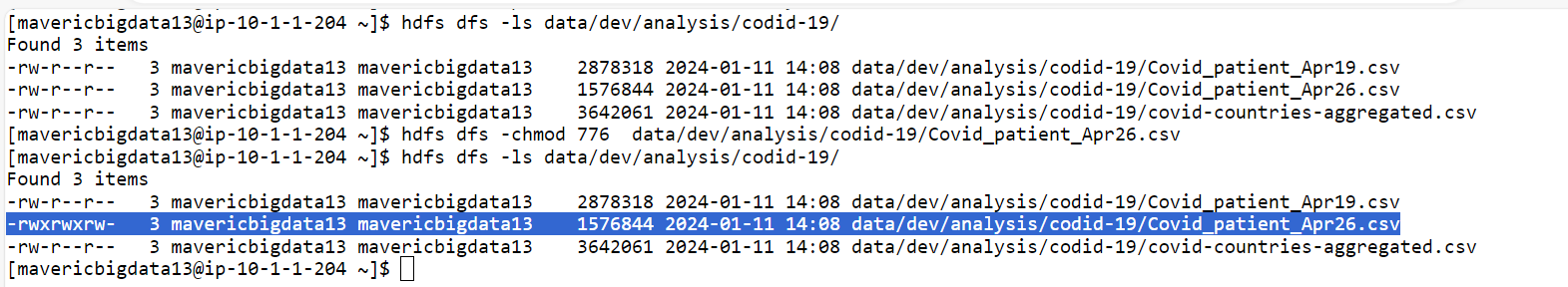
Hdfs dfs -put covid19   data/dev/analysis

A screenshot of a computer code

Description automatically generated

2. Change the permission of Covid\_patient\_Apr26.csv file to have only

read/write/execute for the owner/ group and read/write permission for others.



3. Copy covid-19 folder in to another hdfs directory data/prod/analysis

A close up of a number

Description automatically generated with medium confidence

4. Read the contents of the file Covid\_patient\_Apr19.csv and print the contents on the

Screen\

Hdfs dfs -cat data/dev/analysis/codid-19/ covid\_patient\_Apr19.csv

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

Hdfs dfs -cat data/dev/analysis/codid-19/Covid\_patient\_Apr19.csv | wc -l

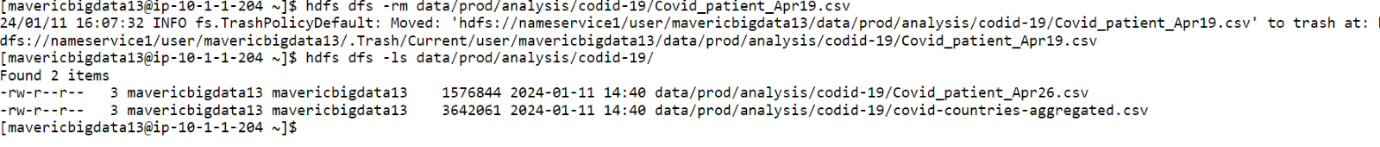
6. Count the no of records present in the HDFS file Covid\_patient\_Apr19.csv for the

state “UP”

Hdfs dfs- count covid\_patient\_Apr19.csv | grep ‘,UP,’

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

Hdfs dfs -rm  data/prod/analysis/codid-19/Covid\_patient\_Apr19.csv



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

Hdfs dfs -rm -R data

Hdfs dfs -rm -R  data/prod/analysis/codid-19/Covid\_patient\_Apr19.csv

Hdfs dfs -rm -R  data/dev/analysis/codid-19/Covid\_patient\_Apr19.csv

A close-up of a number

Description automatically generated

9. Copy the tickitdb folder(in the local file system) in to HDFS home directory. How

many blocks will be created for the files in this directory?

A screenshot of a computer

Description automatically generated  
Blocks creation check using fsck command

A white screen with black text

Description automatically generated

10. Find all the file names which has the word “pipe” in their names.

A screenshot of a computer code

Description automatically generated

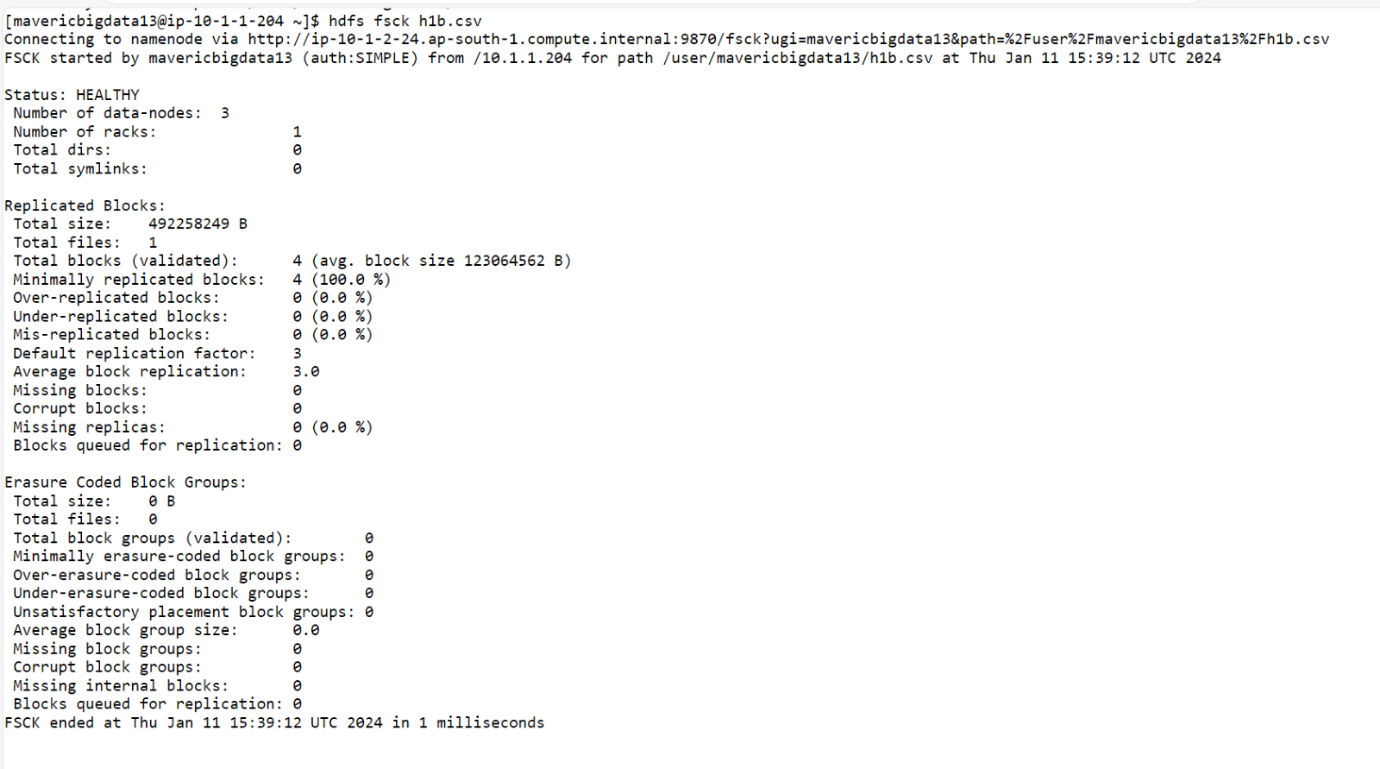
11. Copy the h1b.csv file to hdfs home directory. How many blocks will be created for

this file?

Hdfs dfs -put h1b.csv

A screen shot of a computer code

Description automatically generated  
Block creation checking using fsck cmd



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

A screenshot of a computer code

Description automatically generated  
Block Creation check using fsck cmd

A close-up of a document

Description automatically generated

A screenshot of a computer

Description automatically generated