**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.**

A screenshot of a computer

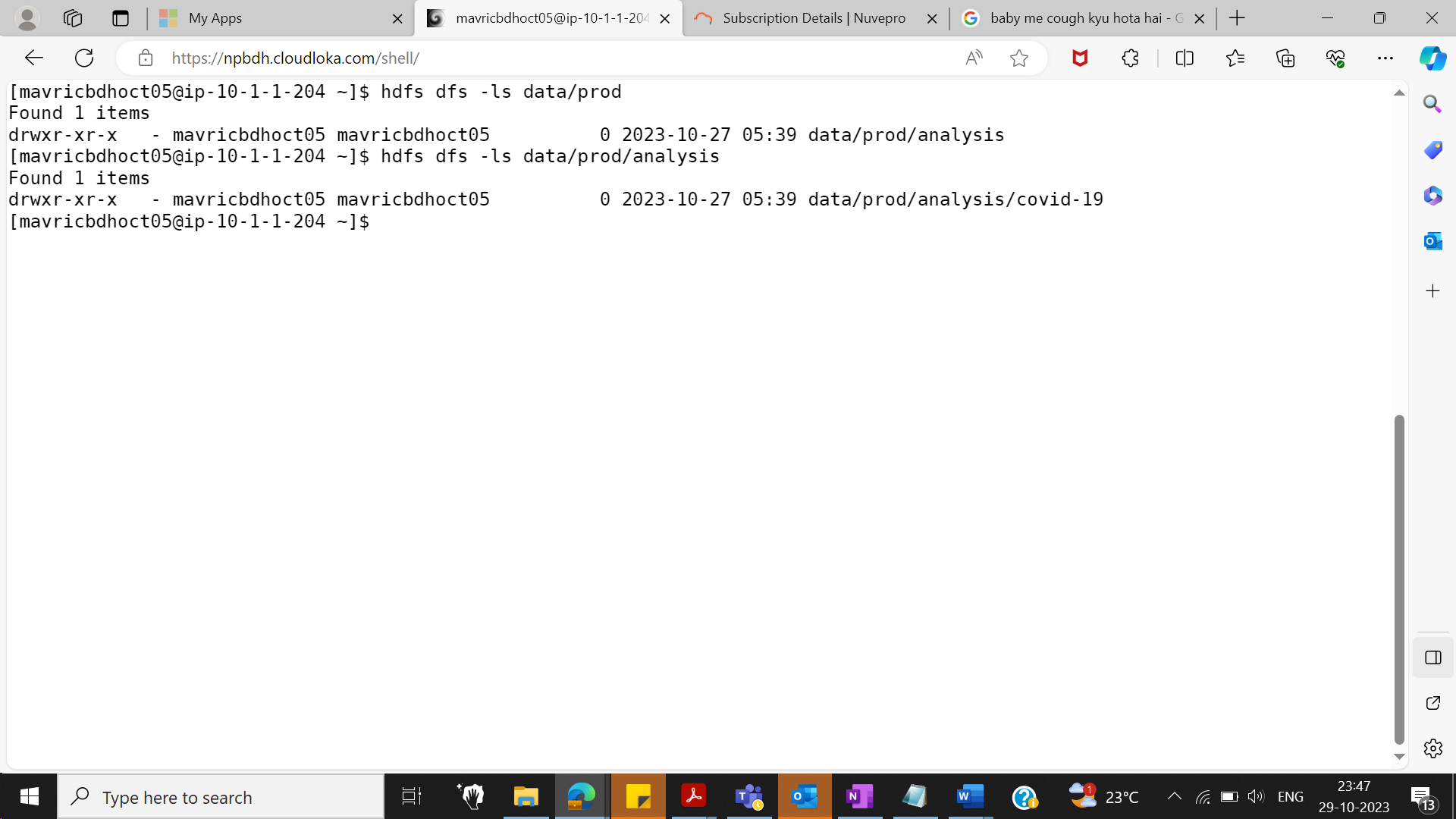
Description automatically generated

1. **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.**

A screenshot of a computer

Description automatically generated

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



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

A screenshot of a computer

Description automatically generated

A computer screen with text

Description automatically generated

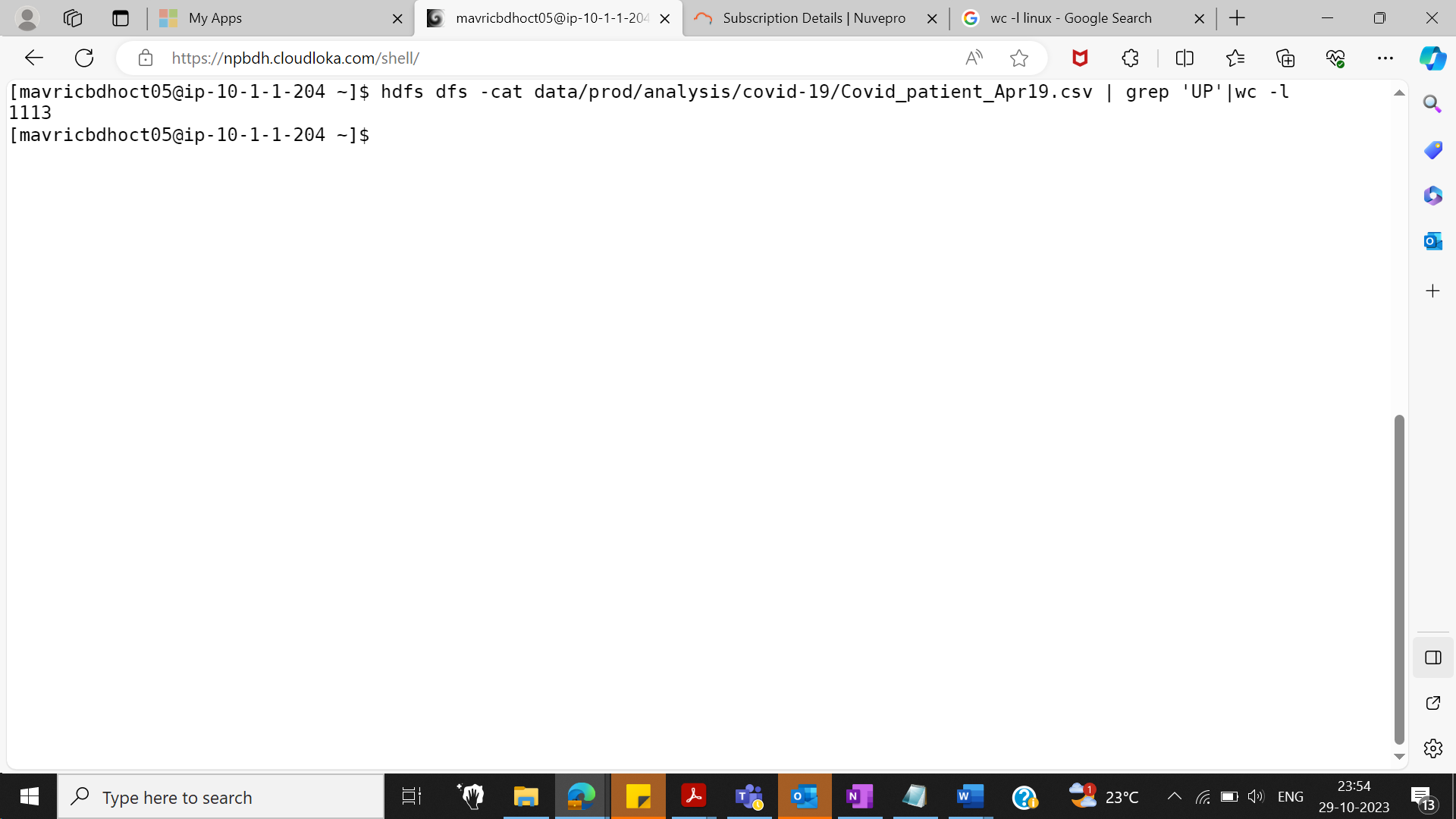
1. **Print the no of lines present in the HDFS file Covid\_patient\_Apr19.csv**

A screenshot of a computer

Description automatically generated

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

**state “UP”**

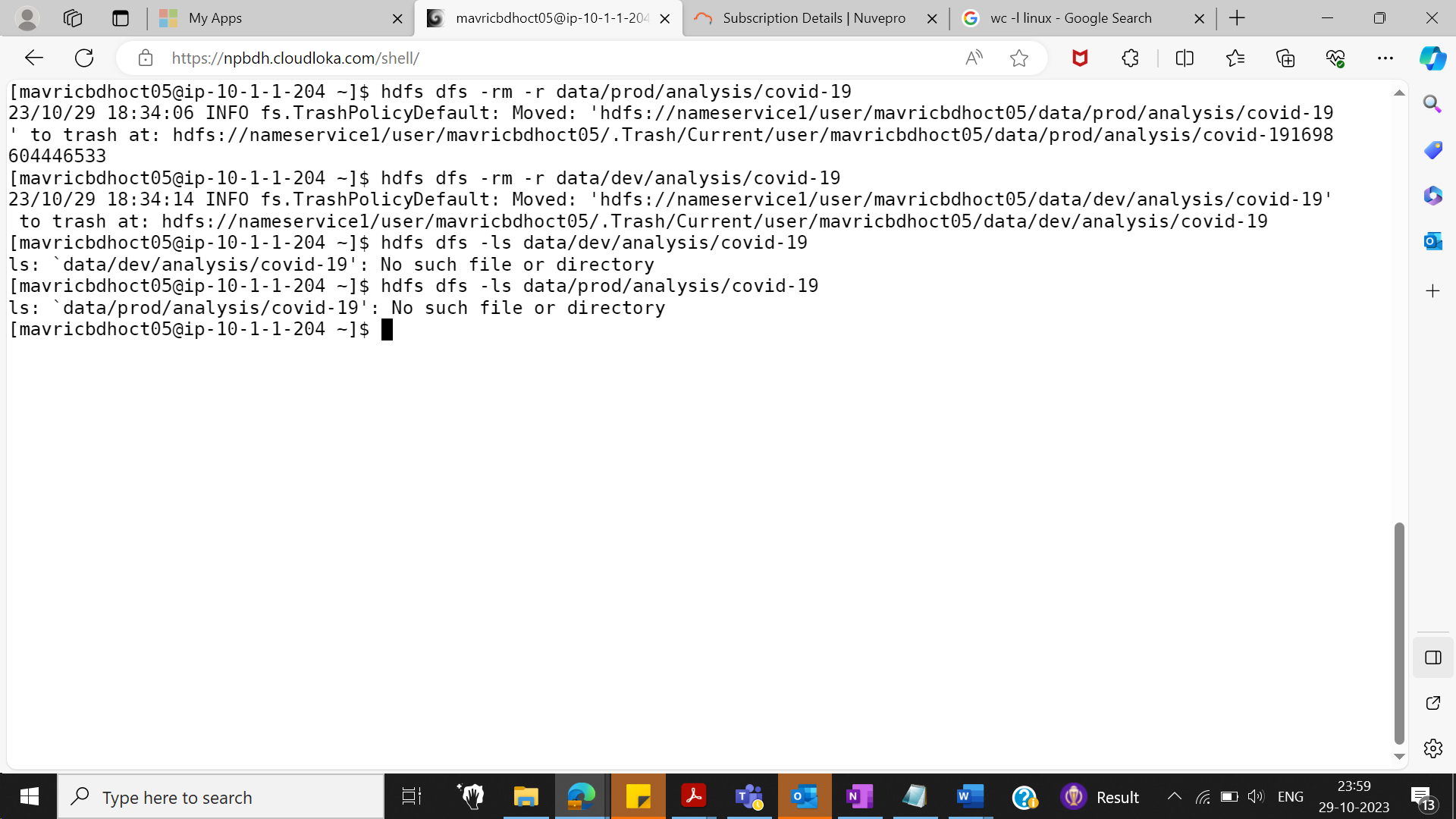


1. **Delete the file Covid\_patient\_Apr19.csv from the data/prod/analysis folder in HDFS**

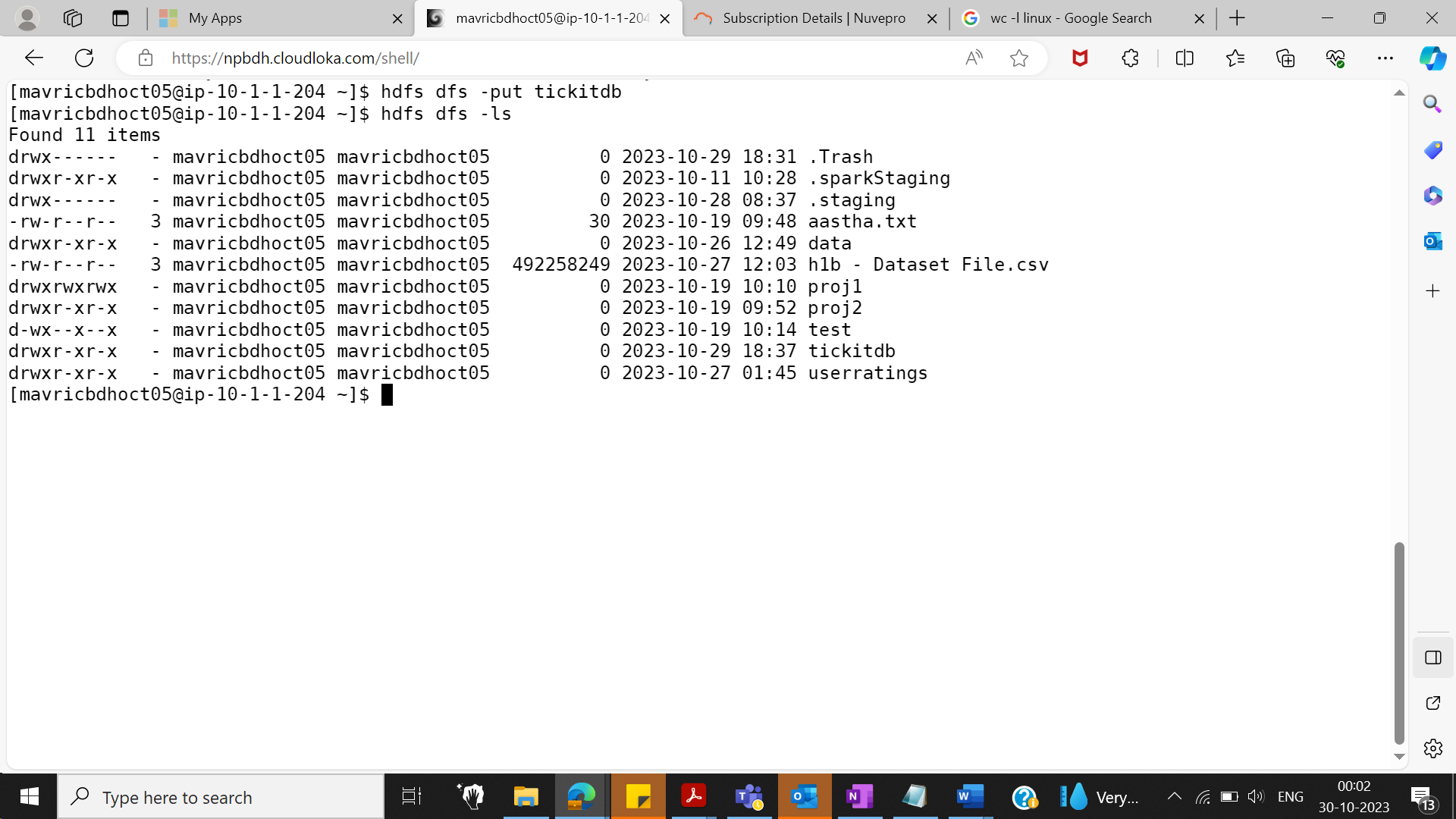
A screenshot of a computer

Description automatically generated

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



1. **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?**



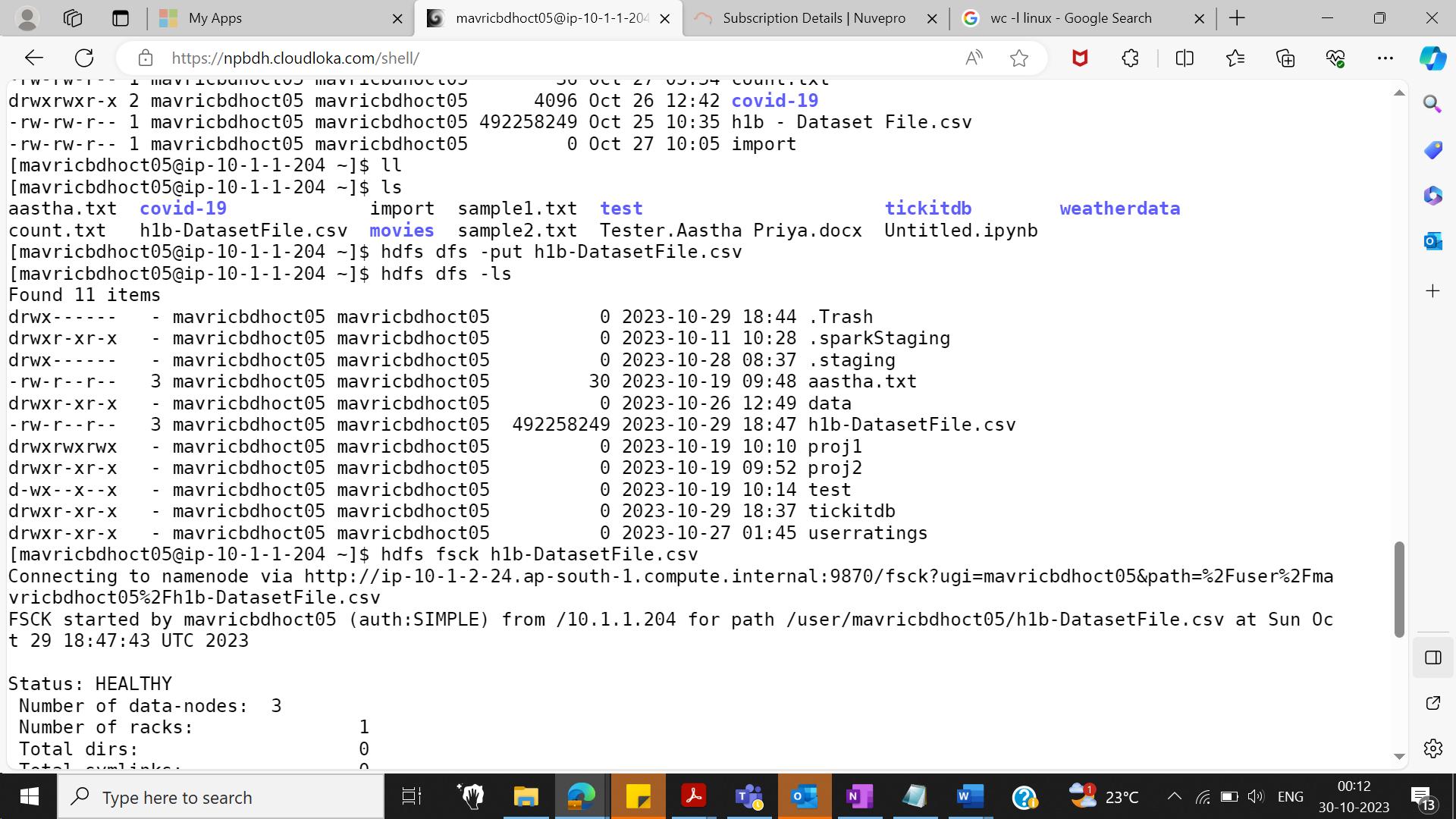
A screenshot of a computer

Description automatically generated

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



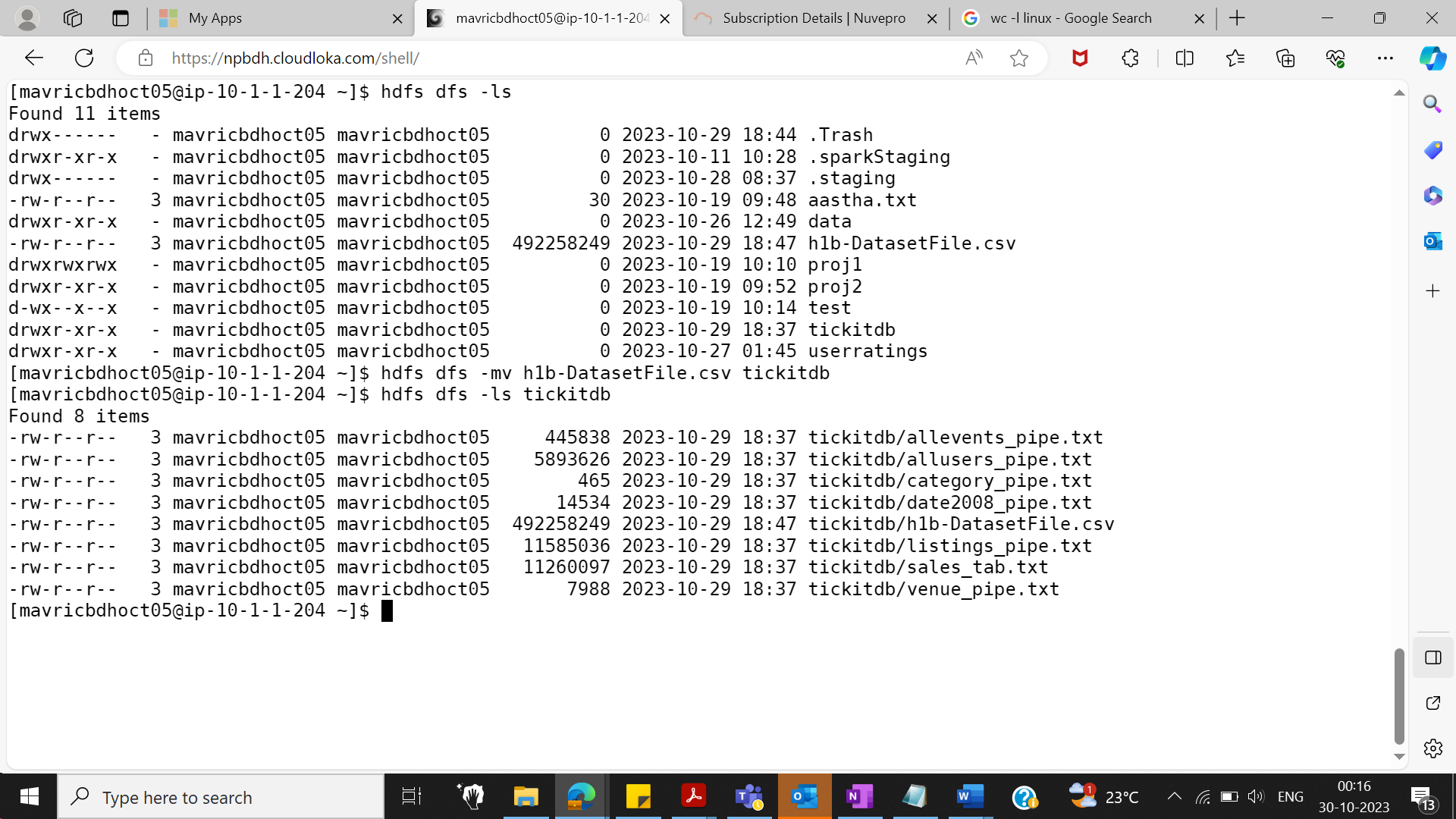
1. **Copy the h1b.csv file to hdfs home directory. How many blocks will be created for this file?**



A screenshot of a computer

Description automatically generated

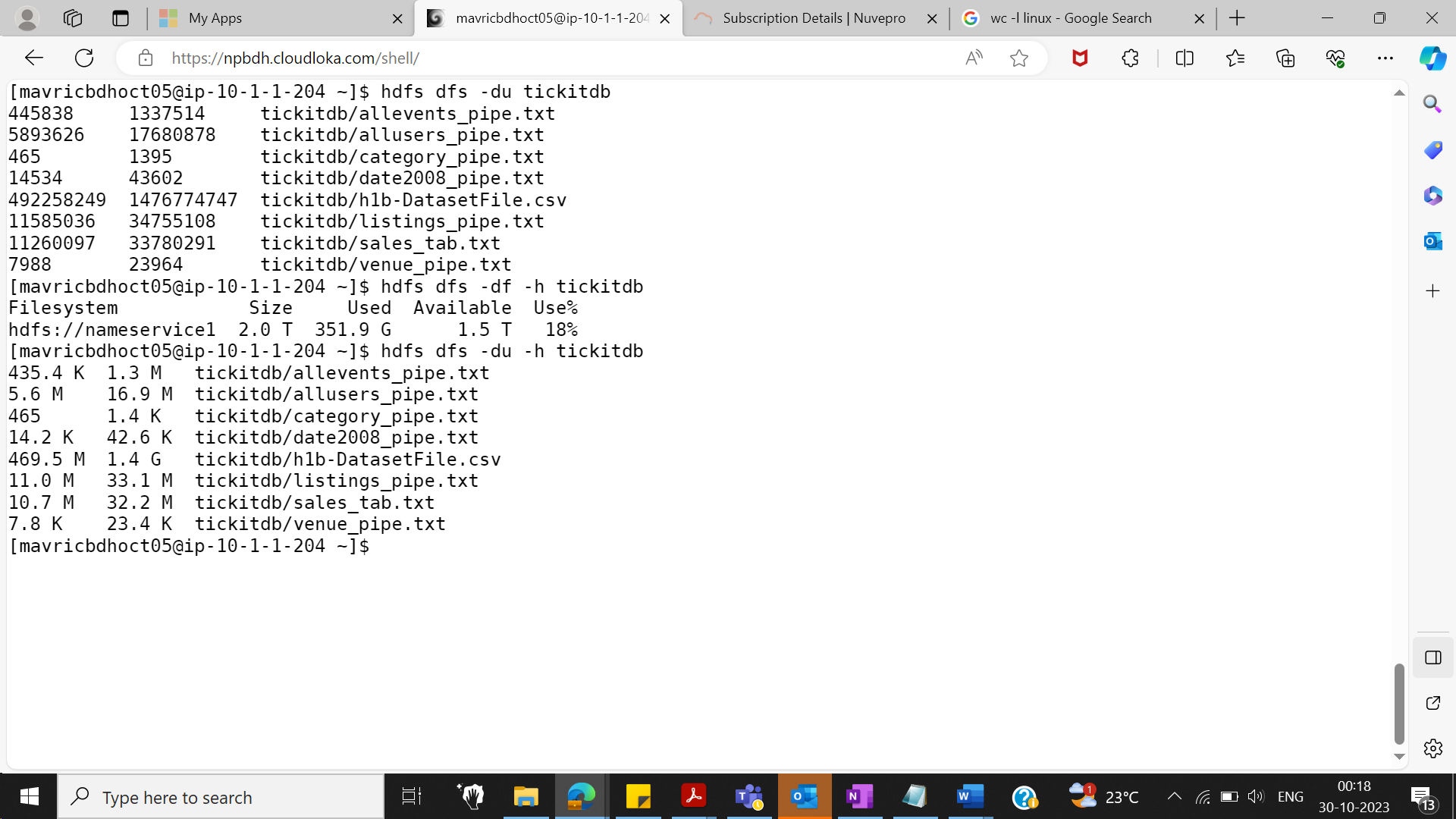
1. **Move the h1b.csv file (in the HDFS) to the tickitdb folder in hdfs. How many blocks will be there for the tickitdb folder?**



A screenshot of a computer

Description automatically generated

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



1. **Check the storage quota**A screenshot of a computer

   Description automatically generated **for the file directory.**
2. **Delete a specific snapshot 'snapshot1' from the file directory.**

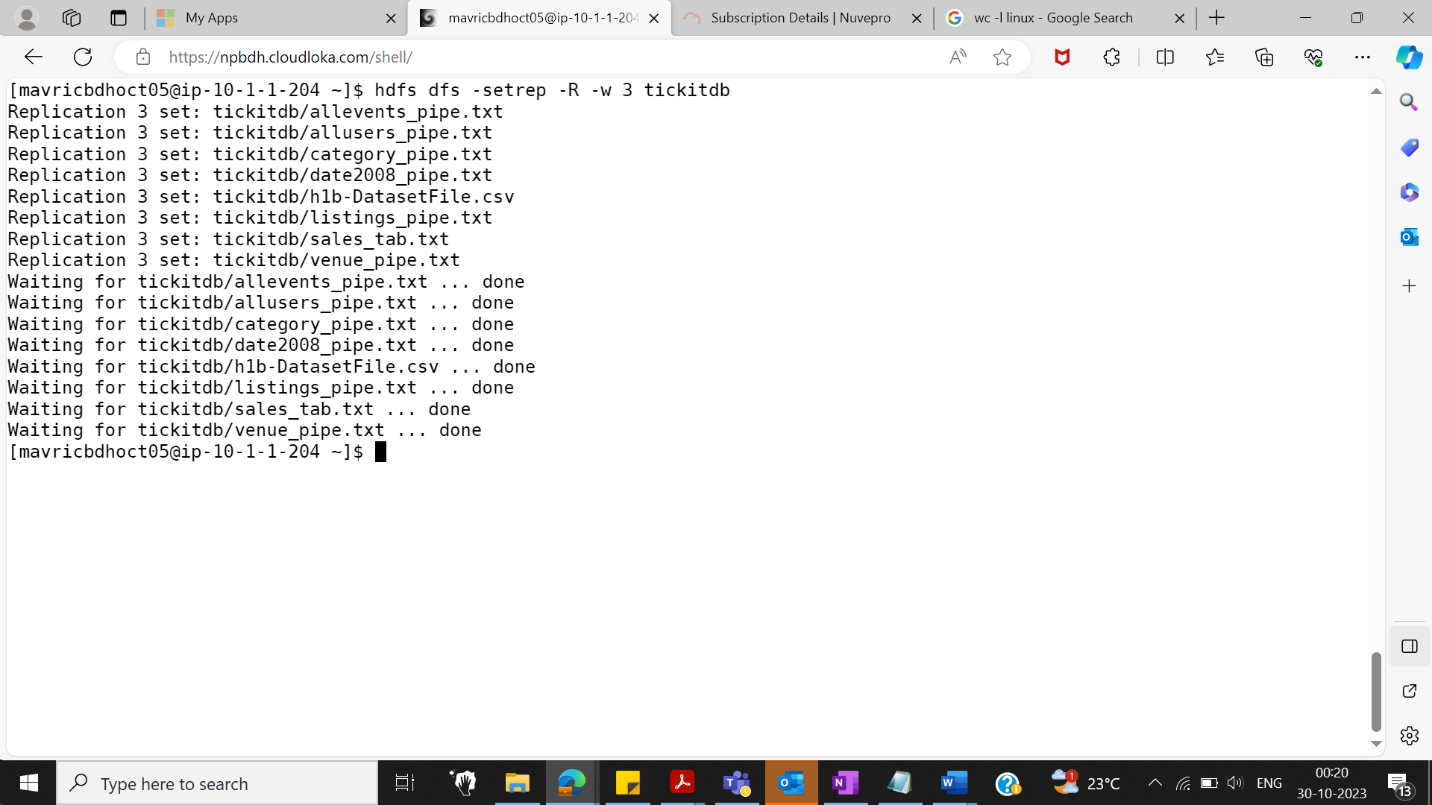
Not able to create snapshot because of access issue

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

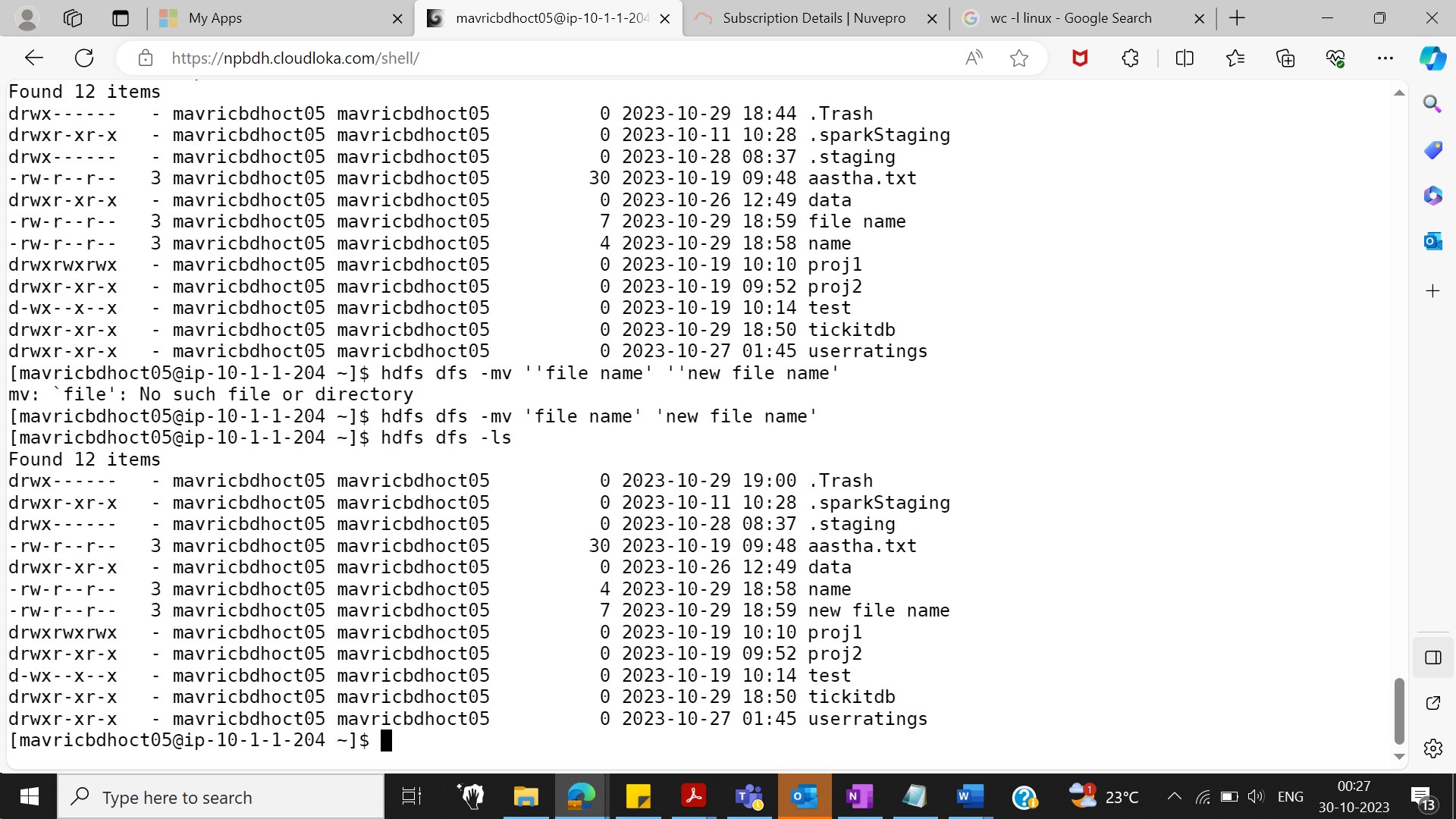
A screenshot of a computer

Description automatically generated

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

z

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



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

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

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

A screenshot of a computer

Description automatically generated

