**HDFS ASSIGNMENT**

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

**Copy the covid-19 folder into it.**

A screenshot of a computer

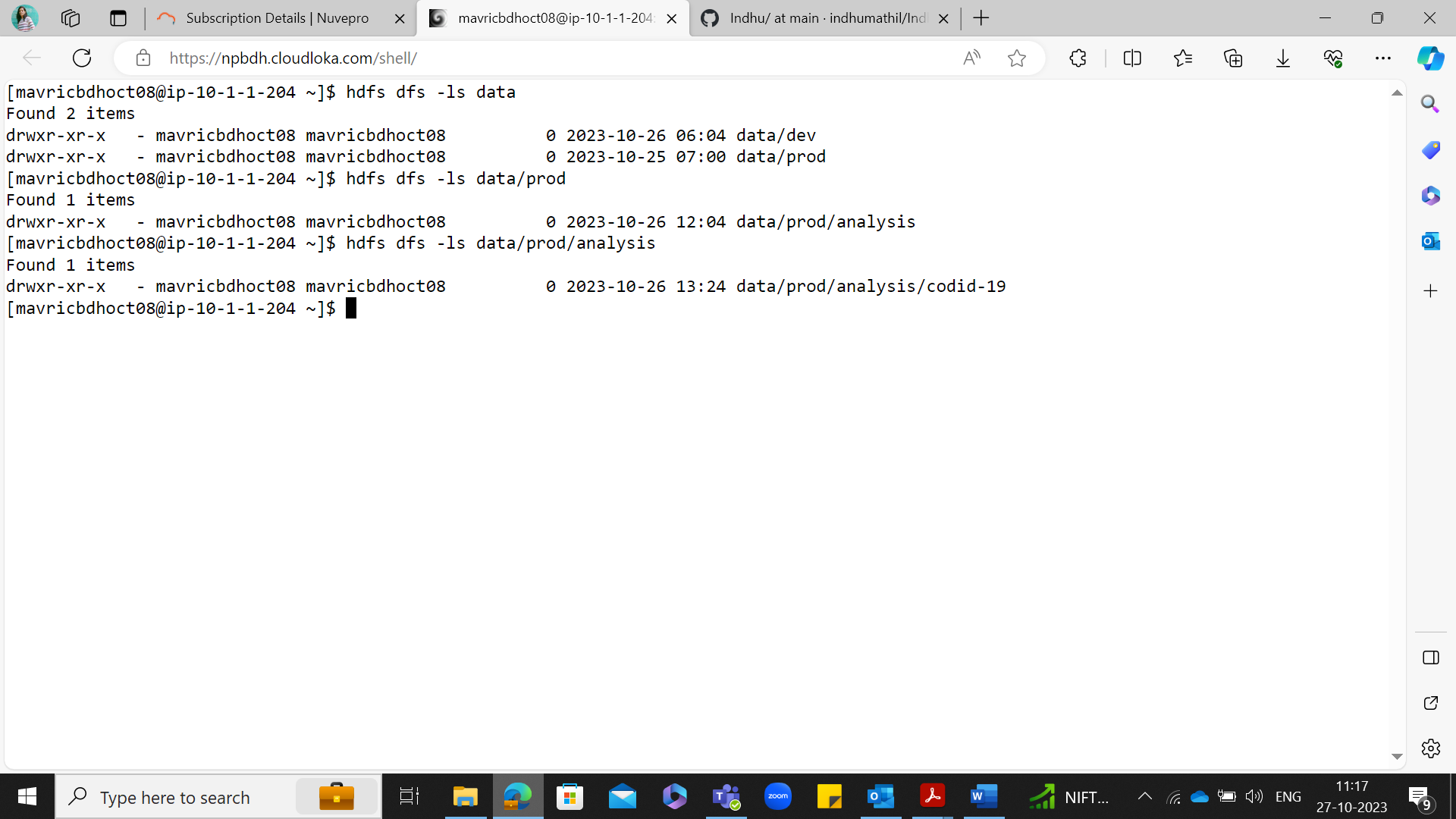
Description automatically generated

**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.**

A screenshot of a computer

Description automatically generated

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



**4.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 screenshot of a computer

Description automatically generated

**5.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”**

A screenshot of a computer

Description automatically generated

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

A screenshot of a computer

Description automatically generated

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

A screenshot of a computer

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

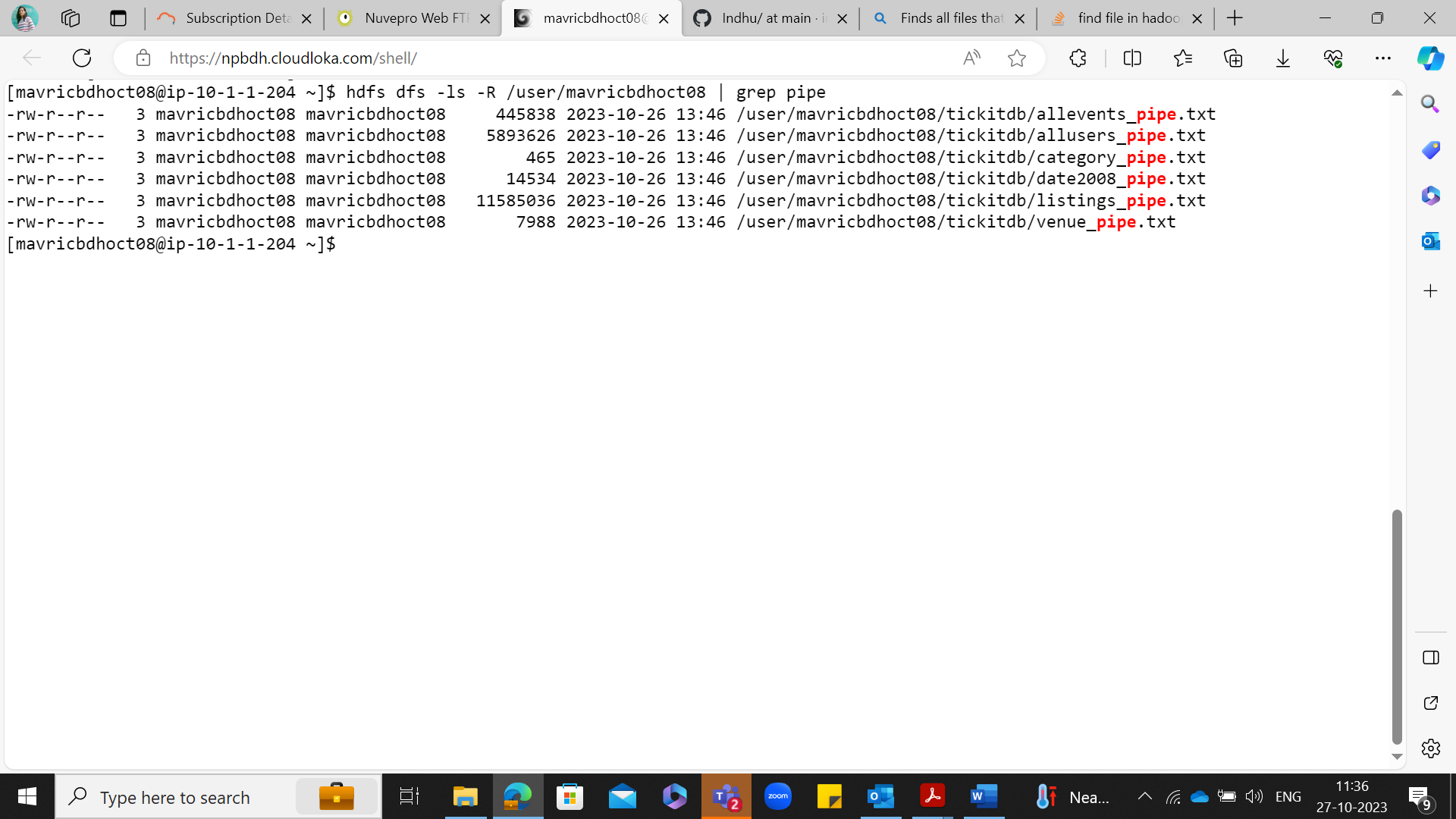
A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

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



**11.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

A screenshot of a computer

Description automatically generated

**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?**

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

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

A screenshot of a computer

Description automatically generated

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

**Note: for getting the result it needs admin access.**

A screenshot of a computer

Description automatically generated

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

**Note: to create Snapshot it needs DFS admin superuser**

A screenshot of a computer

Description automatically generated

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

**Note: to create Snapshot it needs DFS admin superuser**

A screenshot of a computer

Description automatically generated

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

A screenshot of a computer

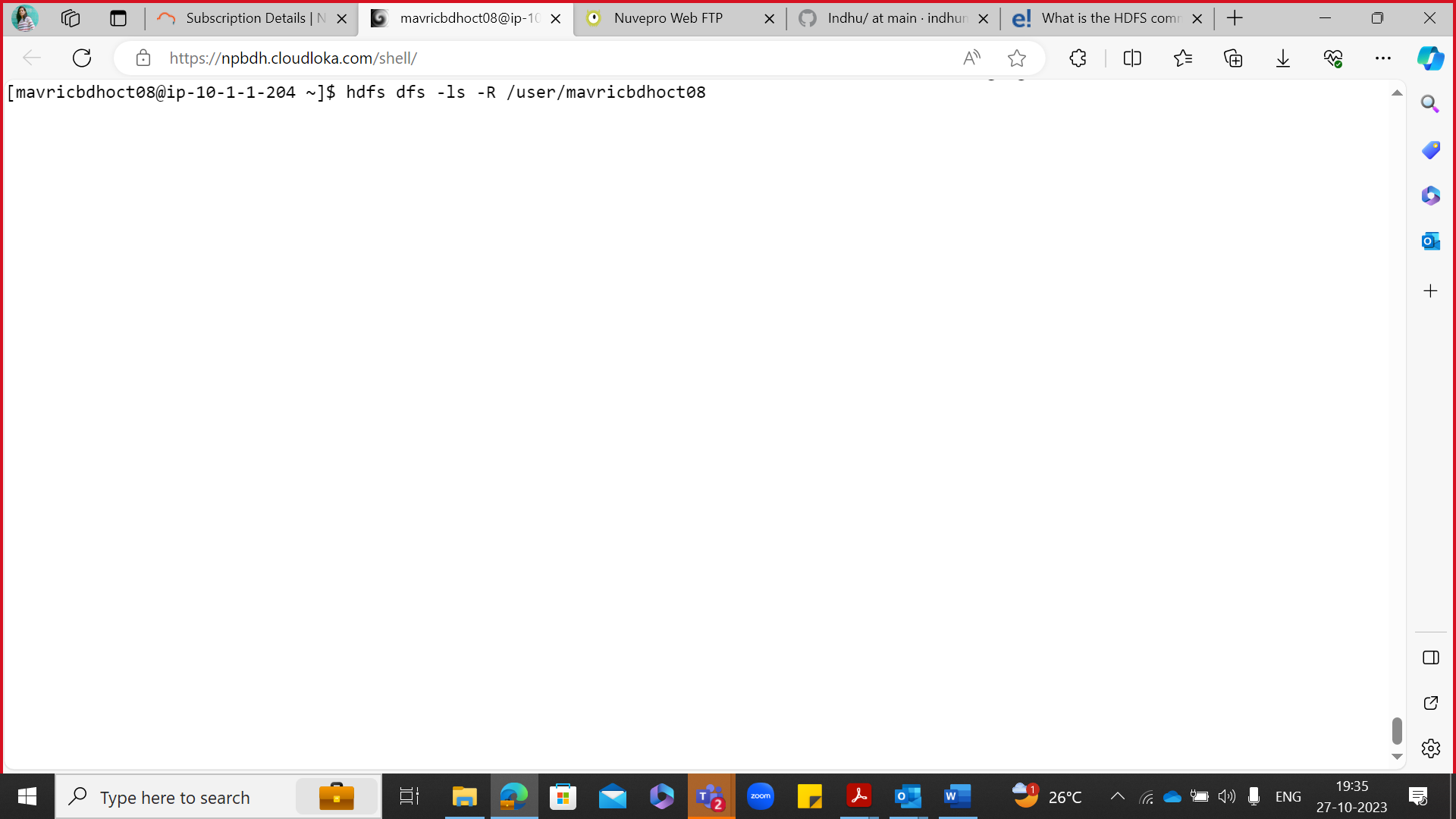
Description automatically generated

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

A screenshot of a computer

Description automatically generated

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



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 generatedA screenshot of a computer

Description automatically generated