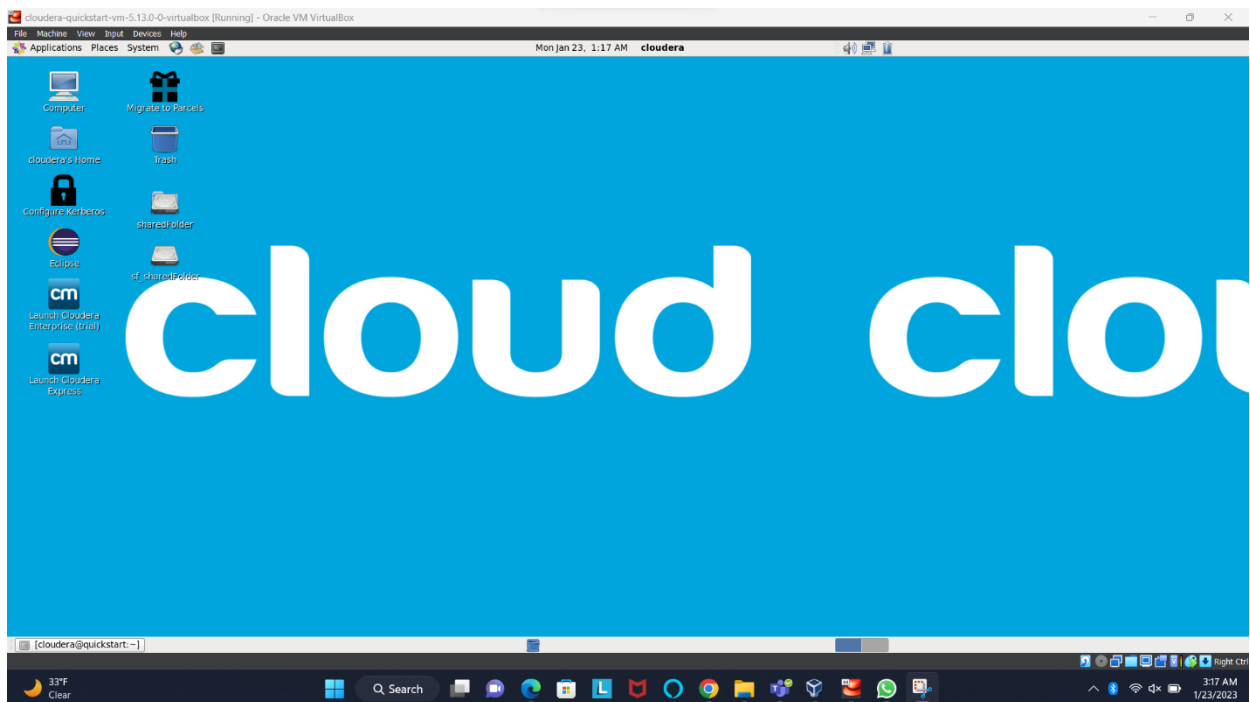


INTRODUCTION TO BIG DATA AND DATA SCIENCE

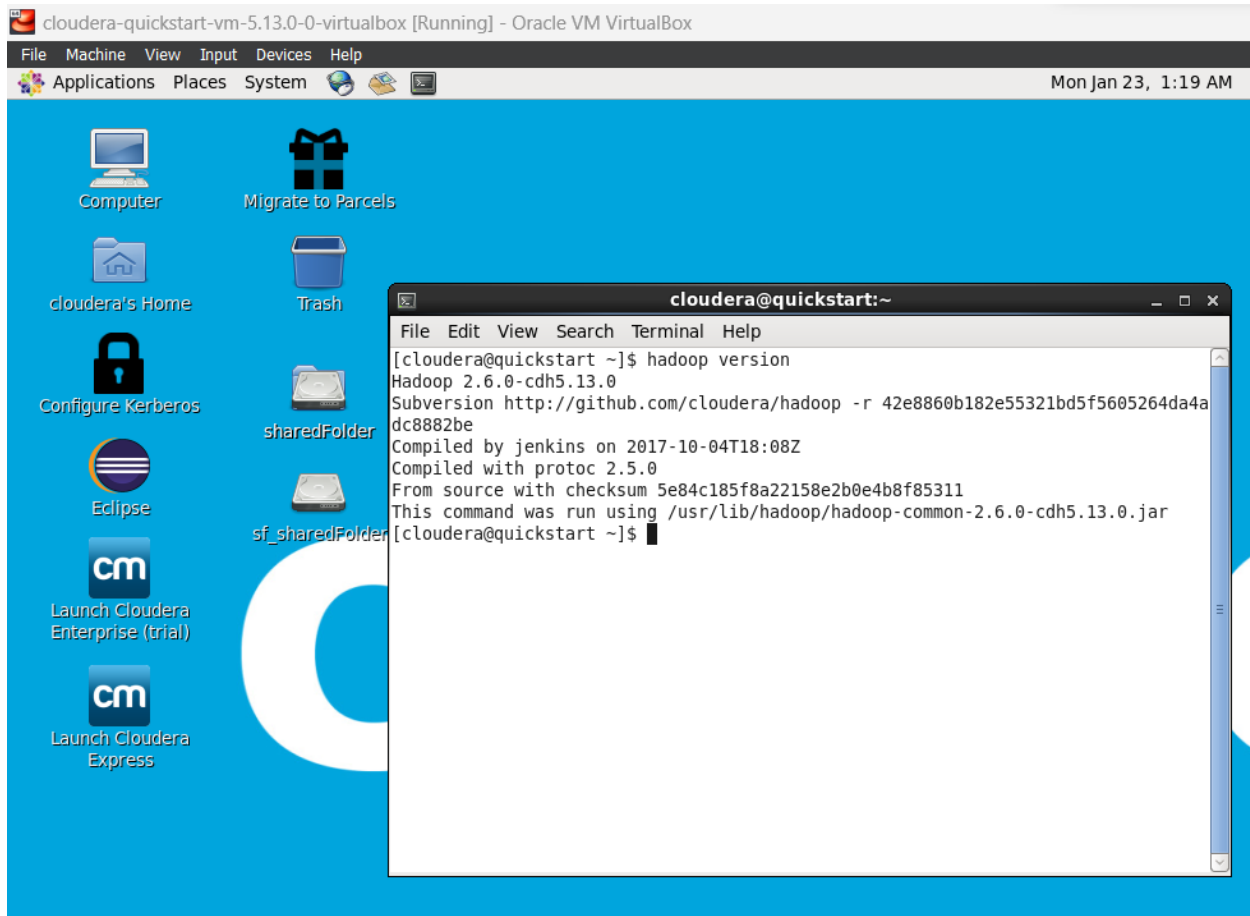
ASSIGNMENT 1

Downloaded Oracle Virtual Machine 7.0 from Oracle official website and downloaded Cloudera quick start VM from canvas.

Installed Virtual Machine and imported Cloudera into it. Below image shows the successful installation of Virtual Machine and Cloudera.

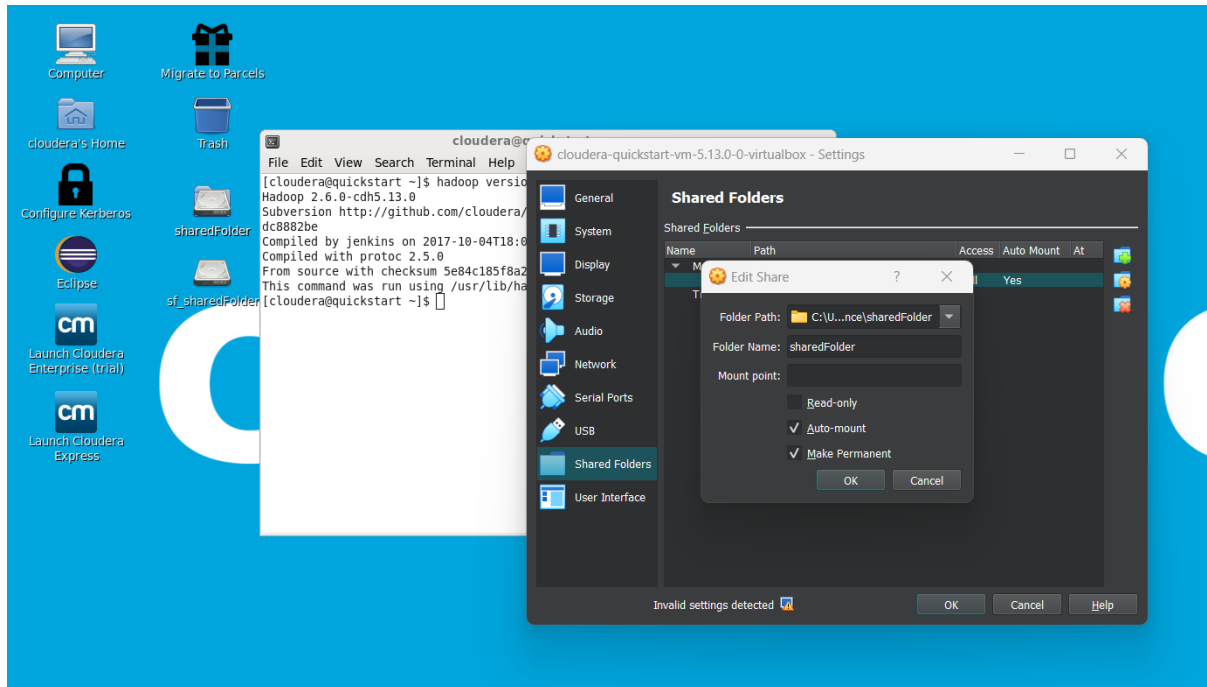


Checking Hadoop version in terminal by using the command '**hadoop version**' to check if Hadoop is working or not. Below screenshot shows that the current version of Hadoop is 2.6.0.



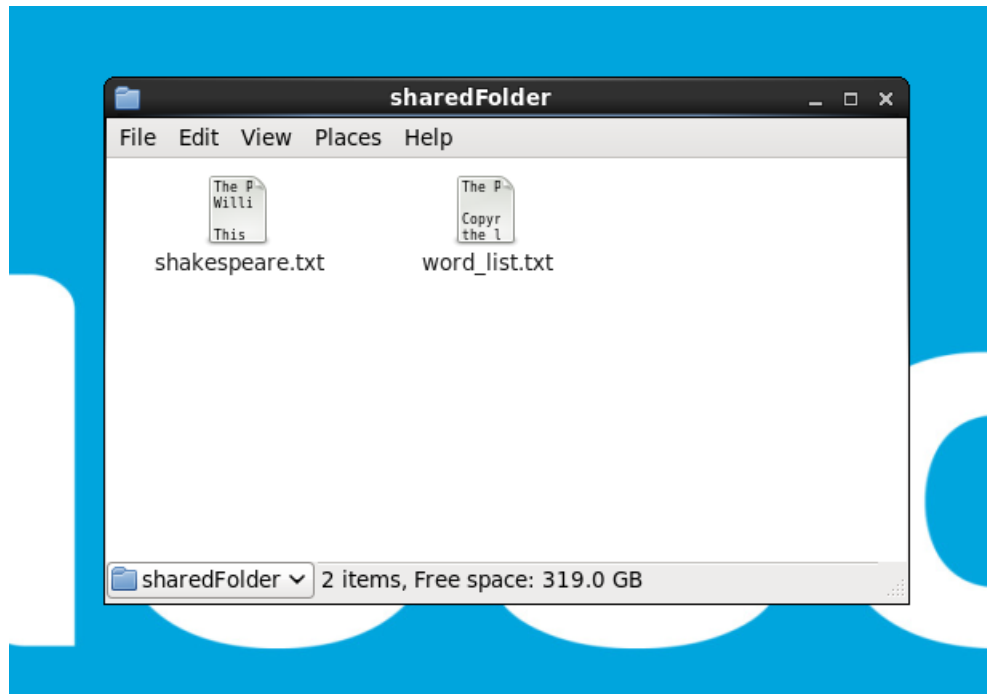
Created a shared folder:

Downloaded the files word_list.txt and shakespeare.txt from canvas to a directory named sharedFolder on windows. Made the sharedFolder available to Cloudera by adding the path of sharedFolder in shared folder settings. Now the sharedFolder in windows environment is available in Cloudera also. Below image is all about, how to create a shared folder.



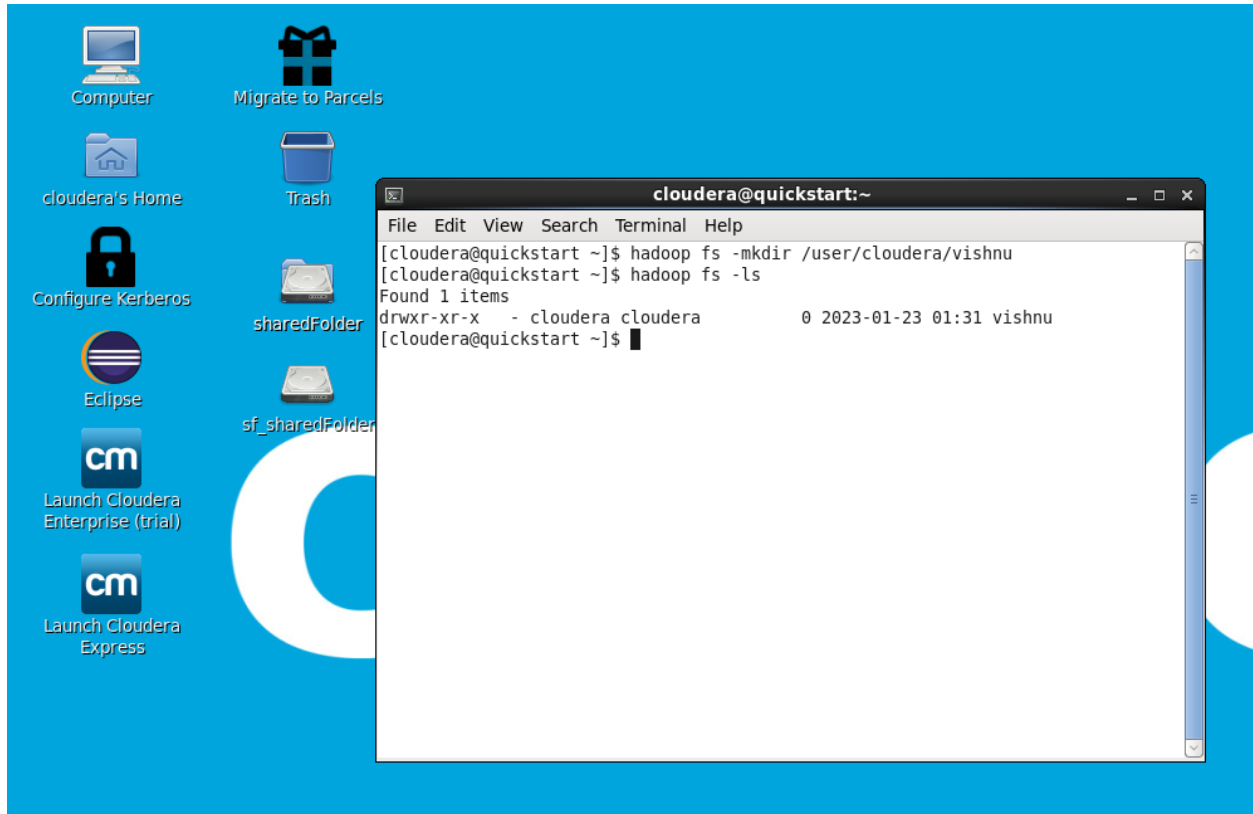
- Under Oracle VM settings added the path of the sharedFolder directory in windows environment

Got the files in sharedFolder in windows environment in to the sharedFolder of Cloudera.



Created a new Directory using Hadoop commands:

Created a directory with name in '**vishnu**' in Hadoop at path '/user/cloudera/vishnu' with the command – 'hadoop fs -mkdir vishnu /user/cloudera/vishnu'. Verified the existence of directory with the command 'hadoop fs -ls /user/cloudera/' or 'hadoop fs -ls'



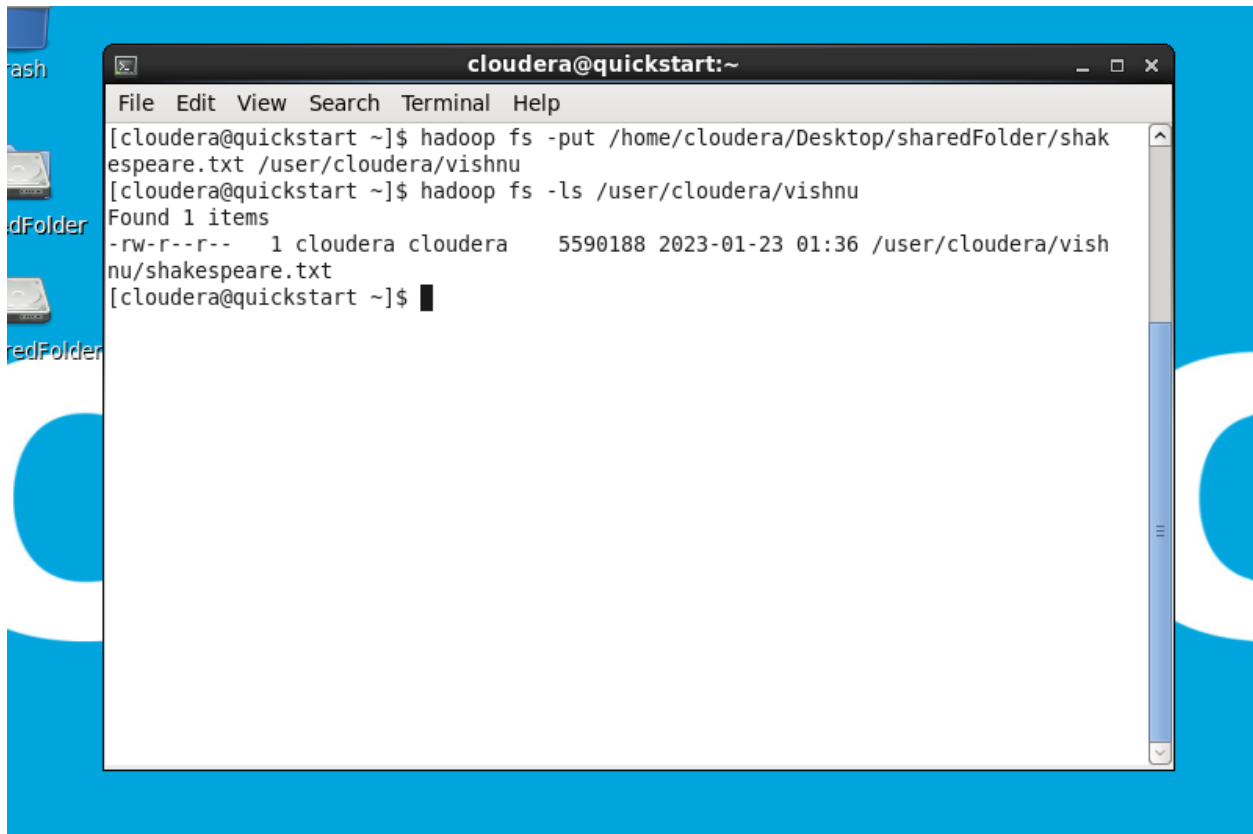
Moved files from cloudera environment to Hadoop:

Moved shakespeare.txt file from Cloudera to Hadoop using the below command.

'hadoop fs -put /home/cloudera/Desktop/sharedFolder/shakespeare.txt /user/cloudera/vishnu'.

Verified the existence of shakespeare.txt file in the Directory 'vishnu' in Hadoop using the below command.

'hadoop fs -ls /user/cloudera/vishnu'

A screenshot of a terminal window titled 'cloudera@quickstart:~'. The terminal shows the following commands and output:

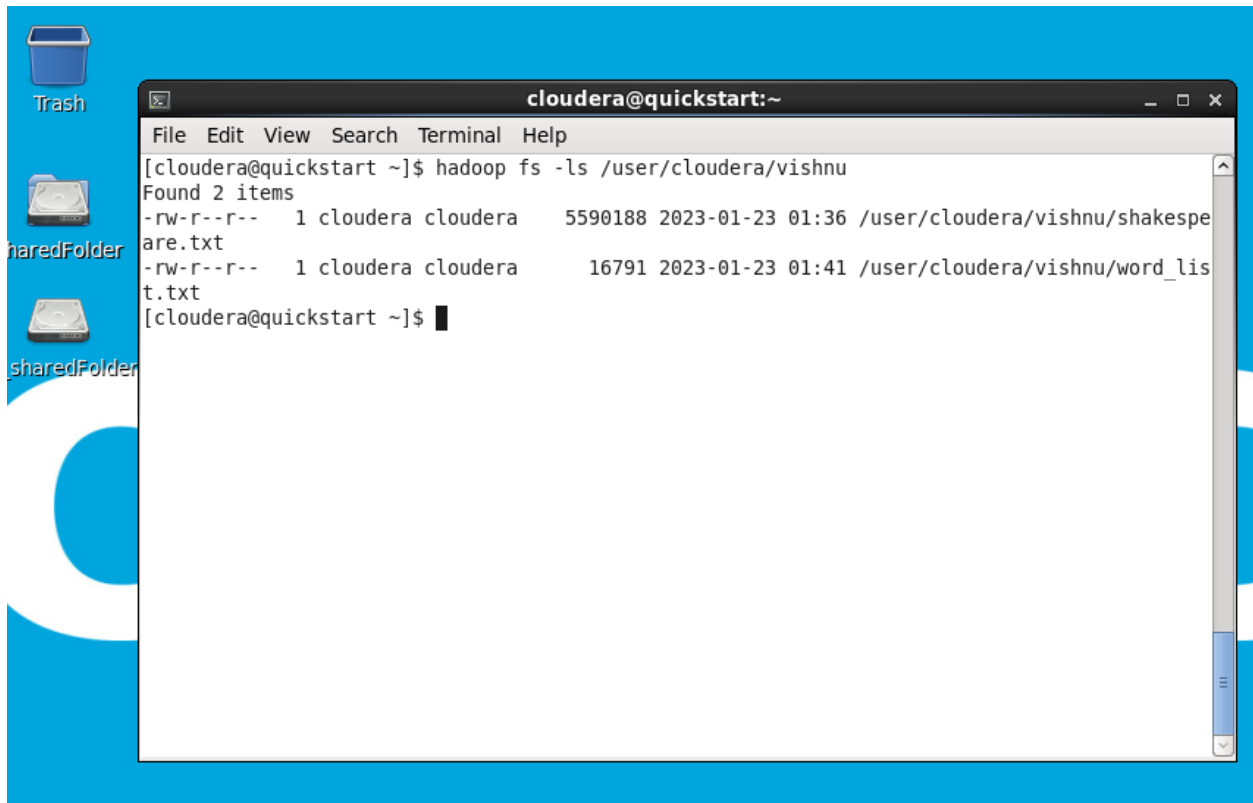
```
[cloudera@quickstart ~]$ hadoop fs -put /home/cloudera/Desktop/sharedFolder/shakespeare.txt /user/cloudera/vishnu
[cloudera@quickstart ~]$ hadoop fs -ls /user/cloudera/vishnu
Found 1 items
-rw-r--r--  1 cloudera cloudera    5590188 2023-01-23 01:36 /user/cloudera/vishnu/shakespeare.txt
[cloudera@quickstart ~]$
```

The terminal window has a menu bar with 'File', 'Edit', 'View', 'Search', 'Terminal', and 'Help'. The background of the desktop is blue with a large white 'C' logo.

Similarly moved word_list.txt file from Cloudera to Hadoop using the below command.
'hadoop fs -put /home/cloudera/Desktop/sharedFolder/word_list.txt /user/cloudera/vishnu'.

Verified the existence of shakespere.txt file in the Directory 'vishnu' in Hadoop using the below command.

'hadoop fs -ls /user/cloudera/vishnu'



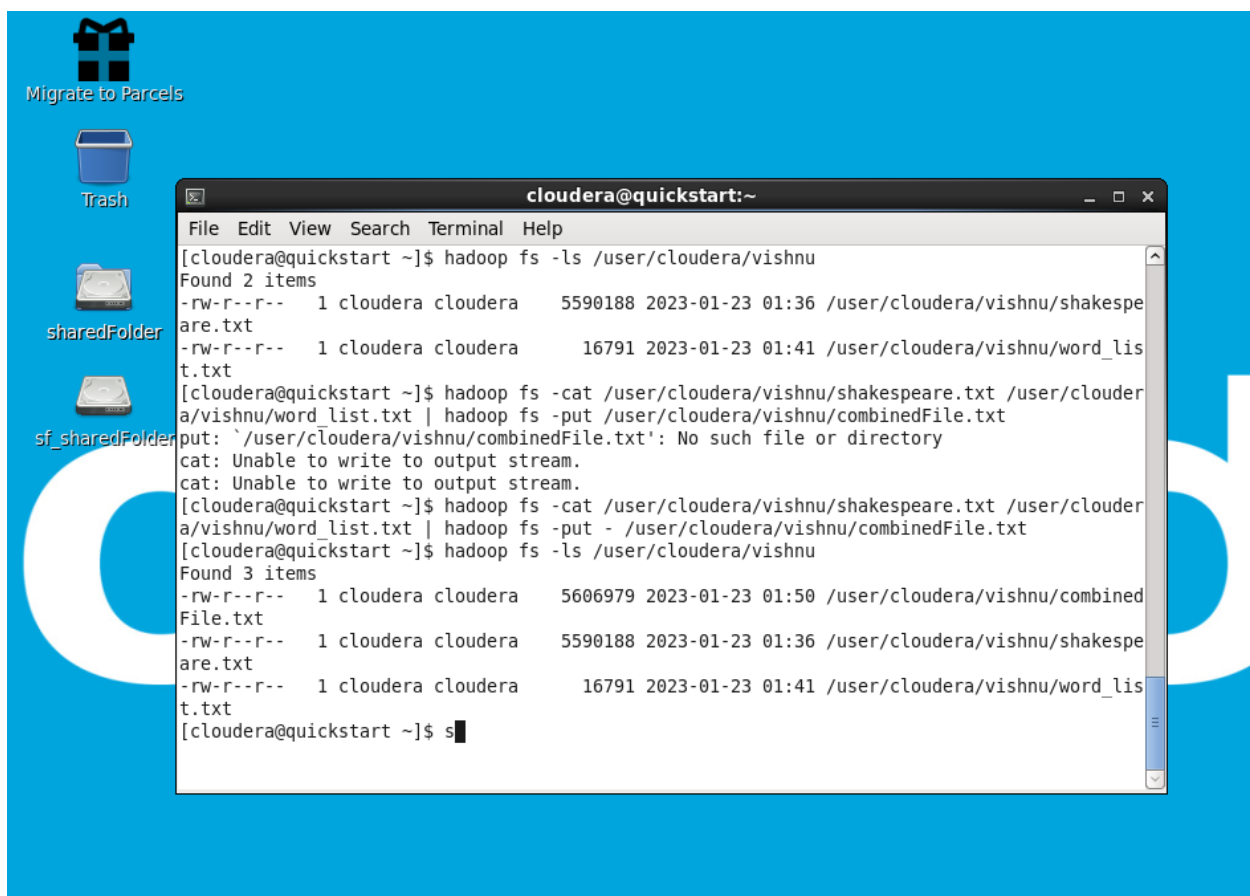
```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hadoop fs -ls /user/cloudera/vishnu  
Found 2 items  
-rw-r--r--  1 cloudera cloudera    5590188 2023-01-23 01:36 /user/cloudera/vishnu/shakespe  
are.txt  
-rw-r--r--  1 cloudera cloudera    16791 2023-01-23 01:41 /user/cloudera/vishnu/word_lis  
t.txt  
[cloudera@quickstart ~]$
```

Merged Files in Hadoop:

Concatenated the files `shakesperae.txt` and `word_list.txt` and generated a new file '`combinedFile.txt`'. This file contains the whole content of the files `shakesperae.txt` and `word_list.txt`. The command used to generate this file is '`hadoop fs -cat /user/cloudera/vishnu/shakespeare.txt /user/cloudera/vishnu/word_list.txt | hadoop fs -put - /user/cloudera/vishnu/combinedFile.txt`'.

Verified the availability of new file generated by using the listing command.

'`hadoop fs -ls /user/cloudera/vishnu`'



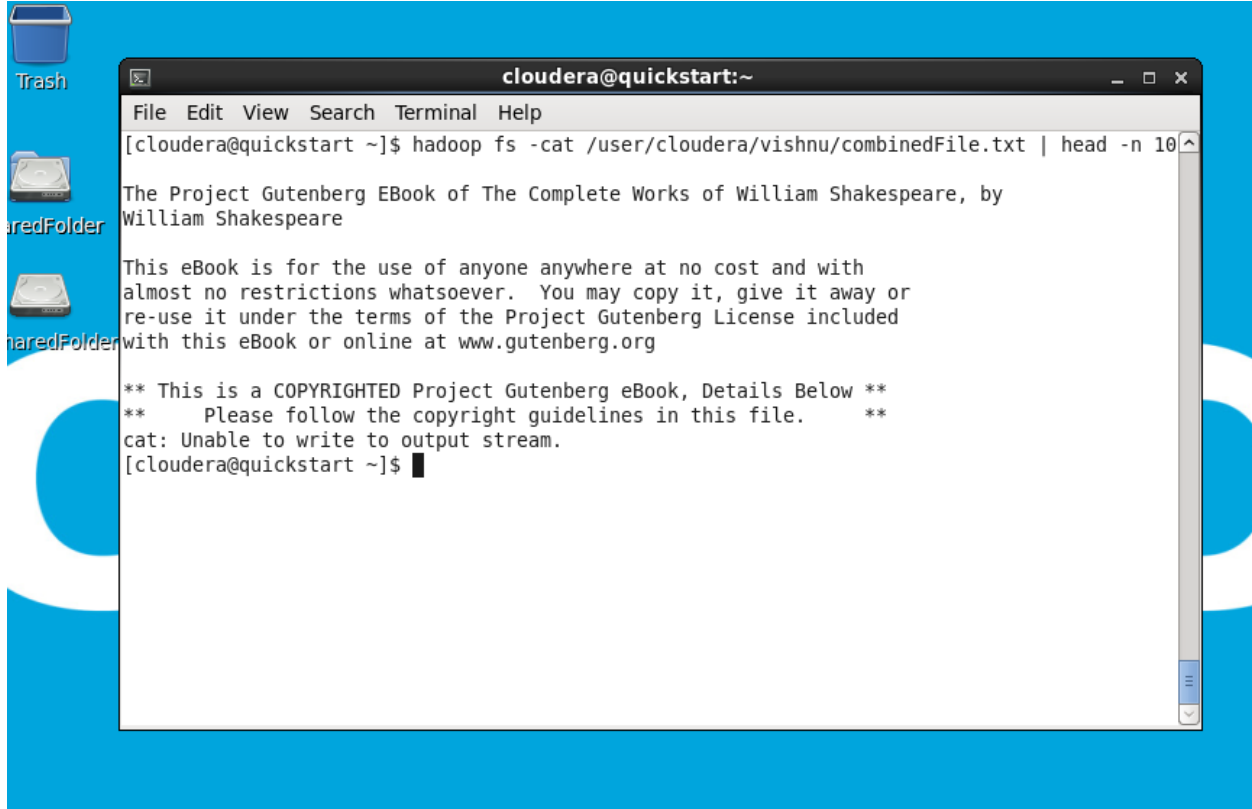
The screenshot shows a Linux desktop with a blue background. On the left, there are icons for 'Migrate to Parcels', 'Trash', 'sharedFolder', and 'sf_sharedFolder'. A terminal window titled 'cloudera@quickstart:~' is open, displaying the following commands and output:

```
cloudera@quickstart:~$ hadoop fs -ls /user/cloudera/vishnu
Found 2 items
-rw-r--r-- 1 cloudera cloudera 5590188 2023-01-23 01:36 /user/cloudera/vishnu/shakespeare.txt
-rw-r--r-- 1 cloudera cloudera 16791 2023-01-23 01:41 /user/cloudera/vishnu/word_list.txt
cloudera@quickstart ~]$ hadoop fs -cat /user/cloudera/vishnu/shakespeare.txt /user/cloudera/vishnu/word_list.txt | hadoop fs -put /user/cloudera/vishnu/combinedFile.txt
put: `/user/cloudera/vishnu/combinedFile.txt': No such file or directory
cat: Unable to write to output stream.
cat: Unable to write to output stream.
cloudera@quickstart ~]$ hadoop fs -cat /user/cloudera/vishnu/shakespeare.txt /user/cloudera/vishnu/word_list.txt | hadoop fs -put - /user/cloudera/vishnu/combinedFile.txt
cloudera@quickstart ~]$ hadoop fs -ls /user/cloudera/vishnu
Found 3 items
-rw-r--r-- 1 cloudera cloudera 5606979 2023-01-23 01:50 /user/cloudera/vishnu/combinedFile.txt
-rw-r--r-- 1 cloudera cloudera 5590188 2023-01-23 01:36 /user/cloudera/vishnu/shakespeare.txt
-rw-r--r-- 1 cloudera cloudera 16791 2023-01-23 01:41 /user/cloudera/vishnu/word_list.txt
cloudera@quickstart ~]$ s
```


Displaying starting 10 lines of a file:

Displayed first 10 lines of the newly created combinedFile.txt using the below command.

`'hadoop fs -cat /user/cloudera/vishnu/combinedFile.txt | head -n 10`



The screenshot shows a terminal window titled "cloudera@quickstart:~" with a menu bar (File, Edit, View, Search, Terminal, Help). The command `[cloudera@quickstart ~]$ hadoop fs -cat /user/cloudera/vishnu/combinedFile.txt | head -n 10` has been executed. The output displays the first 10 lines of the file "combinedFile.txt", which contains the Project Gutenberg EBook of "The Complete Works of William Shakespeare". The output is as follows:

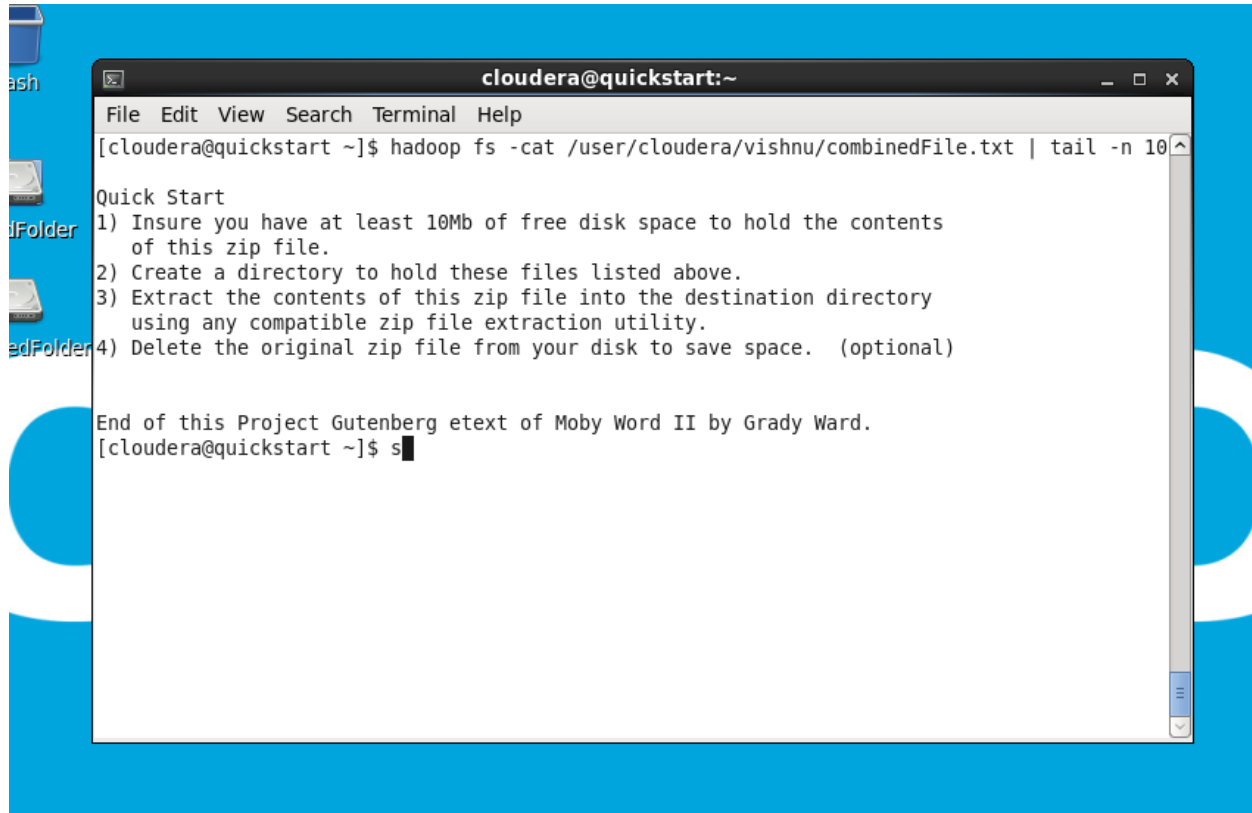
```
The Project Gutenberg EBook of The Complete Works of William Shakespeare, by
William Shakespeare

This eBook is for the use of anyone anywhere at no cost and with
almost no restrictions whatsoever. You may copy it, give it away or
re-use it under the terms of the Project Gutenberg License included
with this eBook or online at www.gutenberg.org

** This is a COPYRIGHTED Project Gutenberg eBook, Details Below **
** Please follow the copyright guidelines in this file. **
cat: Unable to write to output stream.
[cloudera@quickstart ~]$
```

Displaying last 10 lines of a file:

Displayed the last 10 lines of the newly created combinedFile.txt using the below command.
'`hadoop fs -cat /user/cloudera/vishnu/combinedFile.txt | tail -n 10`'



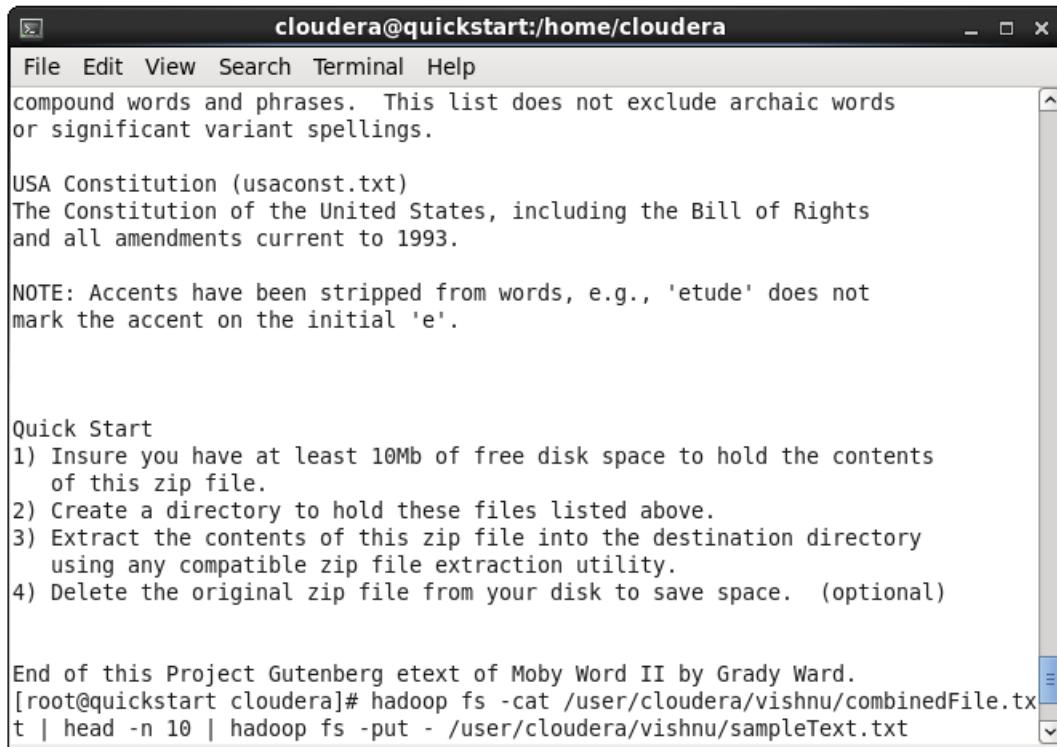
The screenshot shows a terminal window titled 'cloudera@quickstart:~'. The command executed is `hadoop fs -cat /user/cloudera/vishnu/combinedFile.txt | tail -n 10`. The output consists of a 'Quick Start' section with four numbered instructions, followed by the end of a Project Gutenberg text. The terminal window has a menu bar with 'File', 'Edit', 'View', 'Search', 'Terminal', and 'Help'. The background of the desktop is blue with some icons visible on the left.

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hadoop fs -cat /user/cloudera/vishnu/combinedFile.txt | tail -n 10  
Quick Start  
1) Insure you have at least 10Mb of free disk space to hold the contents  
   of this zip file.  
2) Create a directory to hold these files listed above.  
3) Extract the contents of this zip file into the destination directory  
   using any compatible zip file extraction utility.  
4) Delete the original zip file from your disk to save space. (optional)  
  
End of this Project Gutenberg etext of Moby Word II by Grady Ward.  
[cloudera@quickstart ~]$ s
```

Created a new file with 10 lines in it:

Created a sample text file with file name sampleText.txt with 10 lines of content in it using below hadoop command.

```
'hadoop fs -cat /user/cloudera/vishnu/combinedFile.txt | head -n 10 | hadoop fs -put - /user/cloudera/vishnu/sampleText.txt
```



The screenshot shows a terminal window titled "cloudera@quickstart:/home/cloudera". The window contains the following text:

```
File Edit View Search Terminal Help
compound words and phrases. This list does not exclude archaic words
or significant variant spellings.

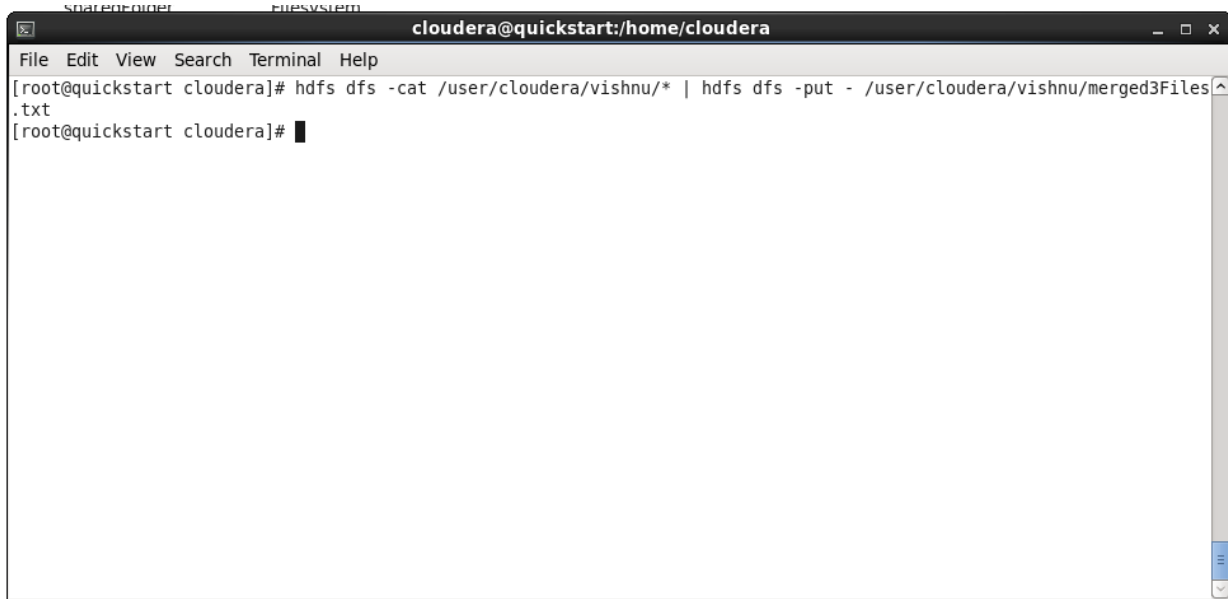
USA Constitution (usaconst.txt)
The Constitution of the United States, including the Bill of Rights
and all amendments current to 1993.

NOTE: Accents have been stripped from words, e.g., 'etude' does not
mark the accent on the initial 'e'.

Quick Start
1) Insure you have at least 10Mb of free disk space to hold the contents
of this zip file.
2) Create a directory to hold these files listed above.
3) Extract the contents of this zip file into the destination directory
using any compatible zip file extraction utility.
4) Delete the original zip file from your disk to save space. (optional)

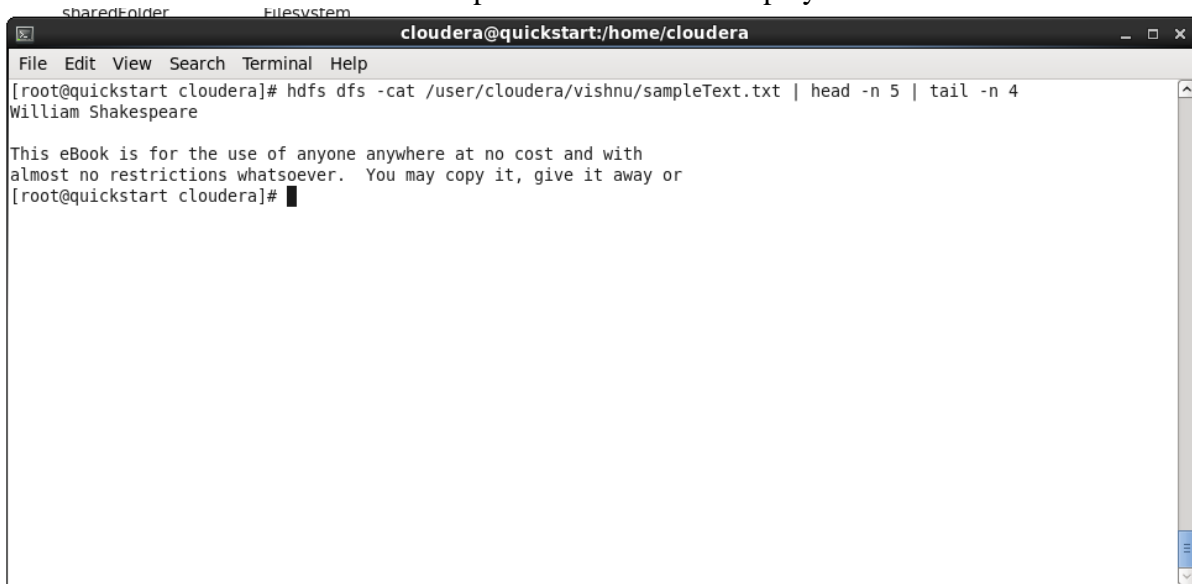
End of this Project Gutenberg etext of Moby Word II by Grady Ward.
[root@quickstart cloudera]# hadoop fs -cat /user/cloudera/vishnu/combinedFile.tx
t | head -n 10 | hadoop fs -put - /user/cloudera/vishnu/sampleText.txt
```

Merged all the 3 files (Shakespeare.txt, word_file.txt and sampleText.txt) to a new file named merged3Files.txt



```
cloudera@quickstart:/home/cloudera
File Edit View Search Terminal Help
[root@quickstart cloudera]# hdfs dfs -cat /user/cloudera/vishnu/* | hdfs dfs -put - /user/cloudera/vishnu/merged3Files.txt
[root@quickstart cloudera]#
```

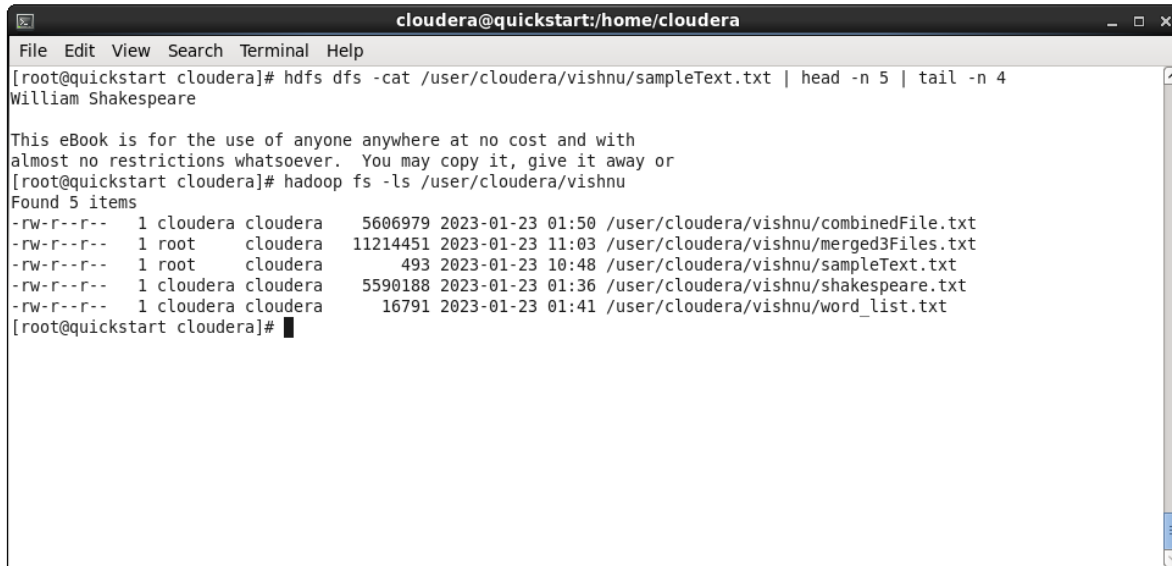
Extracted 2-5 lines from the new sampleText.txt file and displayed it in the terminal.



```
cloudera@quickstart:/home/cloudera
File Edit View Search Terminal Help
[root@quickstart cloudera]# hdfs dfs -cat /user/cloudera/vishnu/sampleText.txt | head -n 5 | tail -n 4
William Shakespeare

This eBook is for the use of anyone anywhere at no cost and with
almost no restrictions whatsoever. You may copy it, give it away or
[root@quickstart cloudera]#
```

Listed all the files in folder named 'vishnu' using 'ls' command.



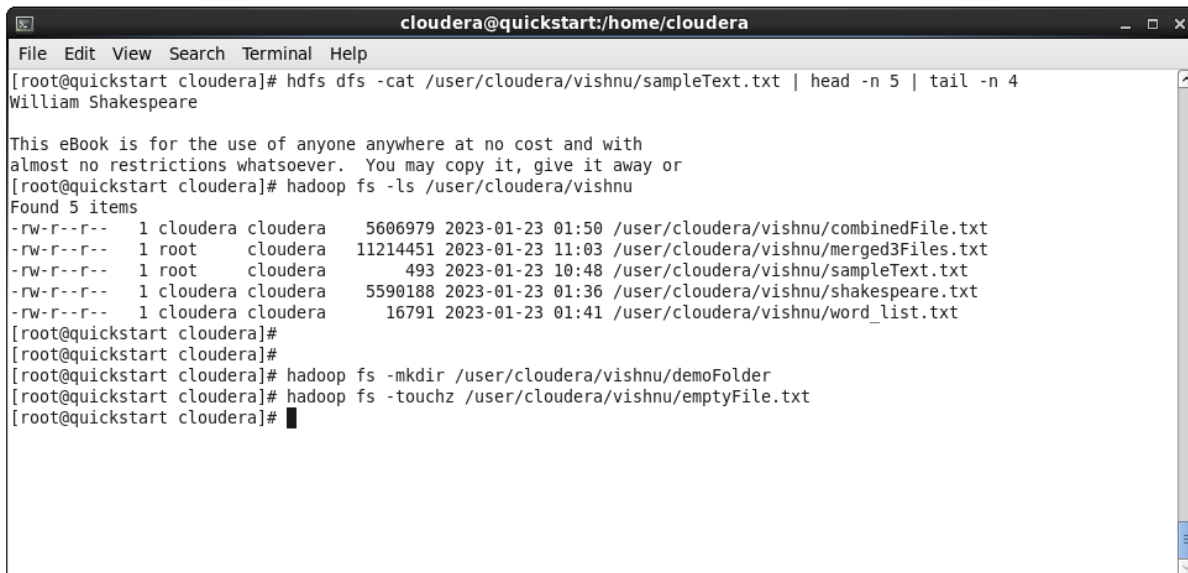
```
cloudera@quickstart:/home/cloudera
File Edit View Search Terminal Help
[root@quickstart cloudera]# hdfs dfs -cat /user/cloudera/vishnu/sampleText.txt | head -n 5 | tail -n 4
William Shakespeare

This eBook is for the use of anyone anywhere at no cost and with
almost no restrictions whatsoever. You may copy it, give it away or
[root@quickstart cloudera]# hadoop fs -ls /user/cloudera/vishnu
Found 5 items
-rw-r--r-- 1 cloudera cloudera 5606979 2023-01-23 01:50 /user/cloudera/vishnu/combinedFile.txt
-rw-r--r-- 1 root cloudera 11214451 2023-01-23 11:03 /user/cloudera/vishnu/merged3Files.txt
-rw-r--r-- 1 root cloudera 493 2023-01-23 10:48 /user/cloudera/vishnu/sampleText.txt
-rw-r--r-- 1 cloudera cloudera 5590188 2023-01-23 01:36 /user/cloudera/vishnu/shakespeare.txt
-rw-r--r-- 1 cloudera cloudera 16791 2023-01-23 01:41 /user/cloudera/vishnu/word_list.txt
[root@quickstart cloudera]#
```

Created a new directory named 'demoFolder' and a text file named 'emptyFile.txt'

'hadoop fs -mkdir' is used to create a directory.

'hadoop fs -touchz' is used to create an empty file.



```
cloudera@quickstart:/home/cloudera
File Edit View Search Terminal Help
[root@quickstart cloudera]# hdfs dfs -cat /user/cloudera/vishnu/sampleText.txt | head -n 5 | tail -n 4
William Shakespeare

This eBook is for the use of anyone anywhere at no cost and with
almost no restrictions whatsoever. You may copy it, give it away or
[root@quickstart cloudera]# hadoop fs -ls /user/cloudera/vishnu
Found 5 items
-rw-r--r-- 1 cloudera cloudera 5606979 2023-01-23 01:50 /user/cloudera/vishnu/combinedFile.txt
-rw-r--r-- 1 root cloudera 11214451 2023-01-23 11:03 /user/cloudera/vishnu/merged3Files.txt
-rw-r--r-- 1 root cloudera 493 2023-01-23 10:48 /user/cloudera/vishnu/sampleText.txt
-rw-r--r-- 1 cloudera cloudera 5590188 2023-01-23 01:36 /user/cloudera/vishnu/shakespeare.txt
-rw-r--r-- 1 cloudera cloudera 16791 2023-01-23 01:41 /user/cloudera/vishnu/word_list.txt
[root@quickstart cloudera]#
[root@quickstart cloudera]# hadoop fs -mkdir /user/cloudera/vishnu/demoFolder
[root@quickstart cloudera]# hadoop fs -touchz /user/cloudera/vishnu/emptyFile.txt
[root@quickstart cloudera]#
```

Below is the view of the files and directories created in hadoop in hue.

Query

Search data and saved documents...

Jobs

cloudera

Sources

- Impala
- Hive

File Browser

Search for file name

Actions

Move to trash

Upload

New

Home

/ user / cloudera / vishnu

Trash

<input type="checkbox"/>	Name	Size	User	Group	Permissions	Date
<input type="checkbox"/>	.		cloudera	cloudera	drwxr-xr-x	January 23, 2023 09:15 AM
<input type="checkbox"/>	.		cloudera	cloudera	drwxr-xr-x	January 23, 2023 11:33 AM
<input type="checkbox"/>	combinedFile.txt	5.3 MB	cloudera	cloudera	-rw-r--r--	January 23, 2023 01:50 AM
<input type="checkbox"/>	demoFolder		root	cloudera	drwxr-xr-x	January 23, 2023 11:33 AM
<input type="checkbox"/>	emptyFile.txt	0 bytes	root	cloudera	-rw-r--r--	January 23, 2023 11:33 AM
<input type="checkbox"/>	mergedFiles.txt	10.7 MB	root	cloudera	-rw-r--r--	January 23, 2023 11:03 AM
<input type="checkbox"/>	sampleText.txt	493 bytes	root	cloudera	-rw-r--r--	January 23, 2023 10:48 AM
<input type="checkbox"/>	shakespeare.txt	5.3 MB	cloudera	cloudera	-rw-r--r--	January 23, 2023 01:36 AM
<input type="checkbox"/>	word_list.txt	16.4 KB	cloudera	cloudera	-rw-r--r--	January 23, 2023 01:41 AM

Show 45 of 7 items

Page 1 of 1

END