

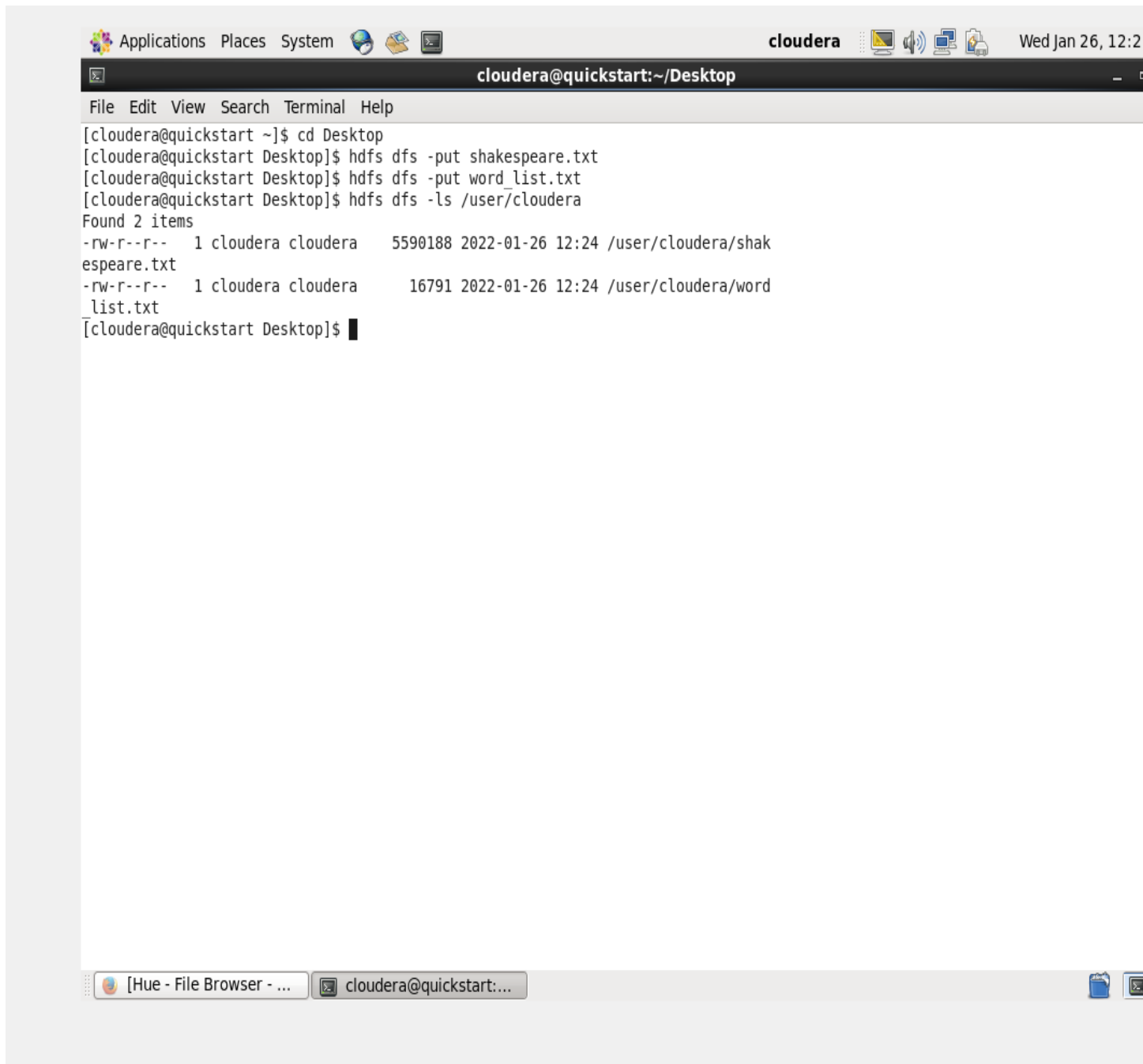
ICE-1

Name: Krishna Reddy Chaduvu

UNT ID:11551974

Step1:Loading the data into hadoop hdfs.

Explanation:Here,we use the given datasets and load them into hadoop.



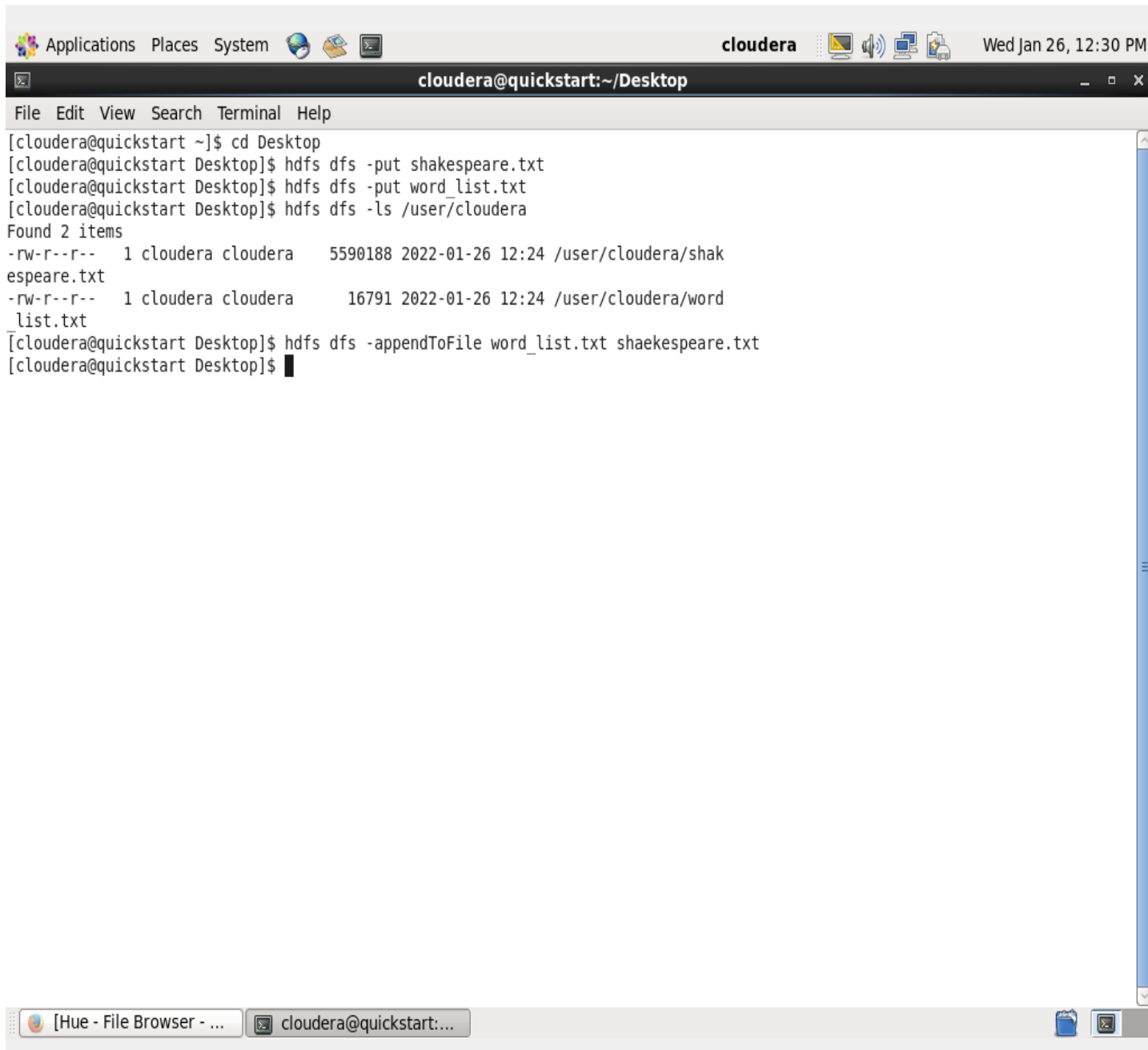
The screenshot shows a terminal window titled "cloudera@quickstart:~/Desktop". The terminal output shows the following commands and results:

```
[cloudera@quickstart ~]$ cd Desktop
[cloudera@quickstart Desktop]$ hdfs dfs -put shakespeare.txt
[cloudera@quickstart Desktop]$ hdfs dfs -put word_list.txt
[cloudera@quickstart Desktop]$ hdfs dfs -ls /user/cloudera
Found 2 items
-rw-r--r--  1 cloudera cloudera    5590188 2022-01-26 12:24 /user/cloudera/shak
espeare.txt
-rw-r--r--  1 cloudera cloudera    16791 2022-01-26 12:24 /user/cloudera/word
list.txt
[cloudera@quickstart Desktop]$
```

The terminal window is part of a desktop environment with a top bar showing "Applications", "Places", "System", and "cloudera". The bottom taskbar shows a "Hue - File Browser - ..." window and a "cloudera@quickstart:..." terminal window.

Step2: Using the second file and appending it to the first file

Explanation: Here, we append the second file(word_list.txt) to first file(shakespeare.txt).



The screenshot shows a terminal window titled "cloudera@quickstart:~/Desktop". The terminal displays the following commands and output:

```
[cloudera@quickstart ~]$ cd Desktop
[cloudera@quickstart Desktop]$ hdfs dfs -put shakespeare.txt
[cloudera@quickstart Desktop]$ hdfs dfs -put word_list.txt
[cloudera@quickstart Desktop]$ hdfs dfs -ls /user/cloudera
Found 2 items
-rw-r--r--  1 cloudera cloudera  5590188 2022-01-26 12:24 /user/cloudera/shakespeare.txt
-rw-r--r--  1 cloudera cloudera   16791 2022-01-26 12:24 /user/cloudera/word_list.txt
[cloudera@quickstart Desktop]$ hdfs dfs -appendToFile word_list.txt shakespeare.txt
[cloudera@quickstart Desktop]$
```

The terminal window is part of a desktop environment with a top bar showing "Applications", "Places", "System", and system status "cloudera" and "Wed Jan 26, 12:30 PM". The bottom taskbar shows a "Hue - File Browser" window and the terminal window.

Step3:Visualizing the file with Hue

Explanation:Using the Hue,we visualize the files .

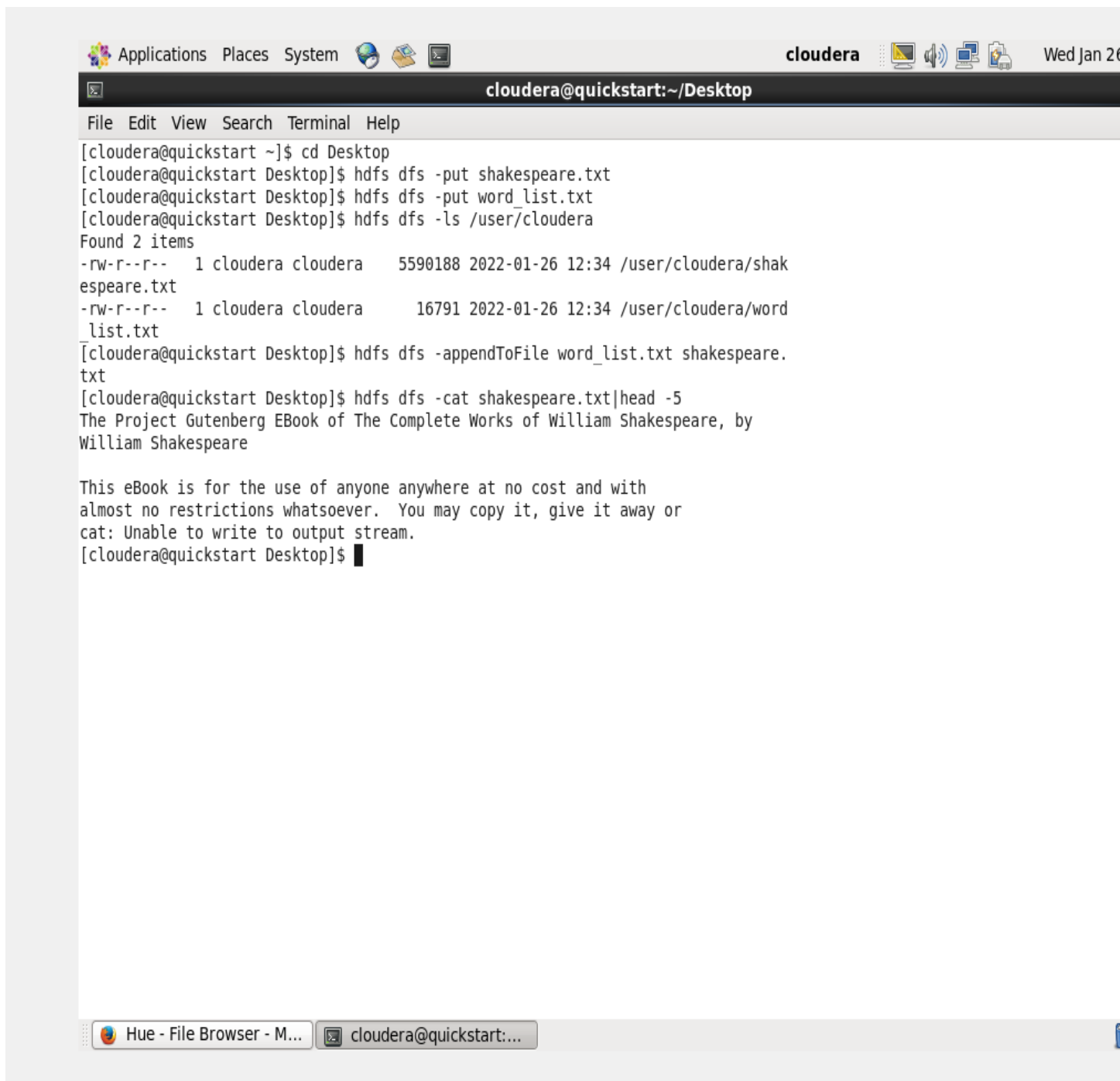
The screenshot displays the Hue File Browser interface within a Mozilla Firefox browser window. The browser's address bar shows the URL `quickstart.cloudera:8888/hue/filebrowser/view%3Duser/cloudera/#/`. The Hue interface includes a top navigation bar with the Hue logo, a search bar, and a 'Query' dropdown. Below this, the 'File Browser' section is active, showing a breadcrumb path of `Home / user / cloudera`. A table lists the files and directories in the `user/cloudera` directory:

	Name	Size	User	Group	Permissions	Date
<input type="checkbox"/>	f		hdfs	supergroup	drwxr-xr-x	January 22, 2022 08:13 PM
<input type="checkbox"/>	.		cloudera	cloudera	drwxr-xr-x	January 26, 2022 12:34 PM
<input type="checkbox"/>	shakespeare.txt	5.3 MB	cloudera	cloudera	-rw-r--r--	January 26, 2022 12:35 PM
<input type="checkbox"/>	word_list.txt	16.4 KB	cloudera	cloudera	-rw-r--r--	January 26, 2022 12:34 PM

At the bottom of the table, it indicates 'Show 45 of 2 items' and 'Page 1 of 1'. The left sidebar shows the 'Tables' section with the message 'The database has no tables'. The bottom of the browser window shows the taskbar with the active window 'Hue - File Browser - M...' and the user 'cloudera@quickstart:...'.

Step4:Viewing the first lines(approximately 5) of merged dataset

Explanation:As,we merged the file to shakespeare.txt file,retrieving the first 5 lines from the it.



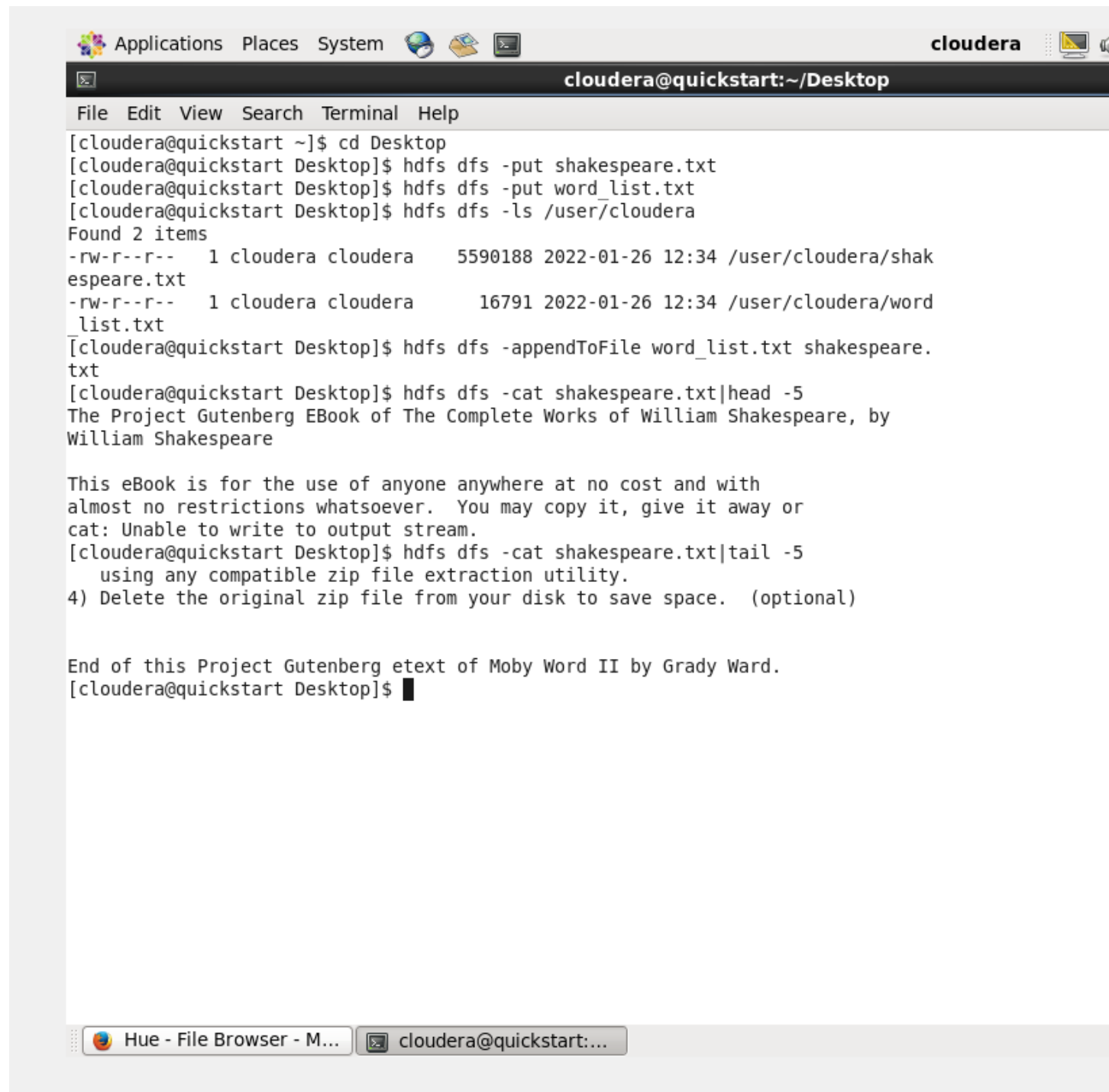
The screenshot shows a desktop environment with a terminal window titled "cloudera@quickstart:~/Desktop". The terminal displays the following commands and output:

```
[cloudera@quickstart ~]$ cd Desktop
[cloudera@quickstart Desktop]$ hdfs dfs -put shakespeare.txt
[cloudera@quickstart Desktop]$ hdfs dfs -put word_list.txt
[cloudera@quickstart Desktop]$ hdfs dfs -ls /user/cloudera
Found 2 items
-rw-r--r--  1 cloudera cloudera  5590188 2022-01-26 12:34 /user/cloudera/shakespeare.txt
-rw-r--r--  1 cloudera cloudera   16791 2022-01-26 12:34 /user/cloudera/word_list.txt
[cloudera@quickstart Desktop]$ hdfs dfs -appendToFile word_list.txt shakespeare.txt
[cloudera@quickstart Desktop]$ hdfs dfs -cat shakespeare.txt|head -5
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
cat: Unable to write to output stream.
[cloudera@quickstart Desktop]$
```

The desktop environment includes a top bar with "Applications", "Places", "System", and "cloudera" status. The bottom taskbar shows "Hue - File Browser - M..." and "cloudera@quickstart:..." windows. The date "Wed Jan 26" is visible in the top right corner.

Step5:Viewing the last lines(approximately 5) of merged data



The screenshot shows a desktop environment with a terminal window titled "cloudera@quickstart:~/Desktop". The terminal displays the following commands and output:

```
[cloudera@quickstart ~]$ cd Desktop
[cloudera@quickstart Desktop]$ hdfs dfs -put shakespeare.txt
[cloudera@quickstart Desktop]$ hdfs dfs -put word_list.txt
[cloudera@quickstart Desktop]$ hdfs dfs -ls /user/cloudera
Found 2 items
-rw-r--r--  1 cloudera cloudera    5590188 2022-01-26 12:34 /user/cloudera/shakespeare.txt
-rw-r--r--  1 cloudera cloudera    16791 2022-01-26 12:34 /user/cloudera/word_list.txt
[cloudera@quickstart Desktop]$ hdfs dfs -appendToFile word_list.txt shakespeare.txt
[cloudera@quickstart Desktop]$ hdfs dfs -cat shakespeare.txt|head -5
The Project Gutenberg EBook of The Complete Works of William Shakespeare, by
William Shakespeare

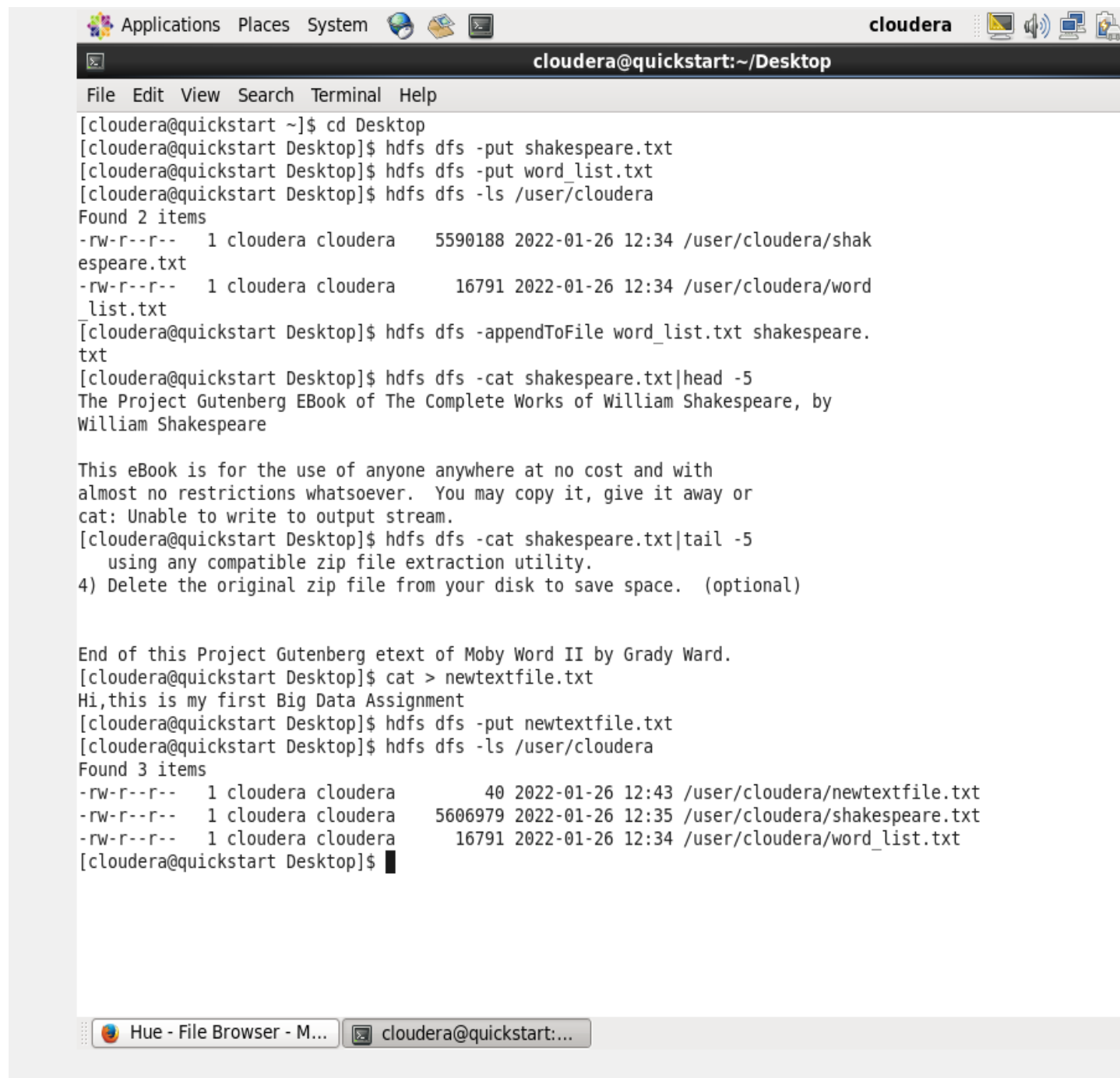
This eBook is for the use of anyone anywhere at no cost and with
almost no restrictions whatsoever.  You may copy it, give it away or
cat: Unable to write to output stream.
[cloudera@quickstart Desktop]$ hdfs dfs -cat shakespeare.txt|tail -5
    using any compatible zip file extraction utility.
4) Delete 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 Desktop]$
```

The desktop taskbar at the bottom shows two open windows: "Hue - File Browser - M..." and "cloudera@quickstart:..."

Step6: Creation of new text file and loading it into hdfs

Explanation: Here, we create a new file called newtextfile.txt and after adding some data, we load the file into hadoop hdfs



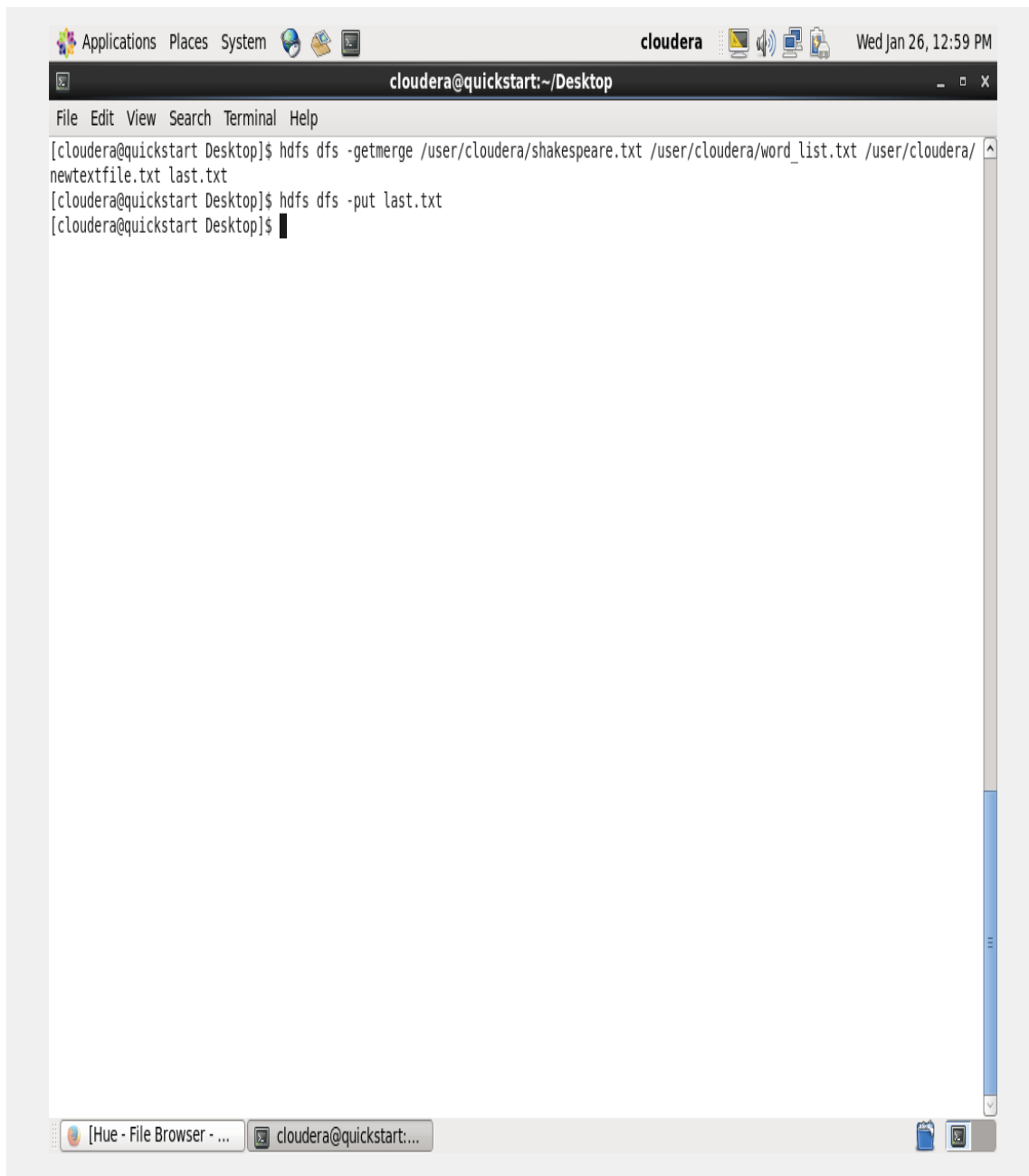
```
Applications  Places  System  cloudera@quickstart:~/Desktop
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ cd Desktop
[cloudera@quickstart Desktop]$ hdfs dfs -put shakespeare.txt
[cloudera@quickstart Desktop]$ hdfs dfs -put word_list.txt
[cloudera@quickstart Desktop]$ hdfs dfs -ls /user/cloudera
Found 2 items
-rw-r--r--  1 cloudera cloudera    5590188 2022-01-26 12:34 /user/cloudera/shakespeare.txt
-rw-r--r--  1 cloudera cloudera    16791 2022-01-26 12:34 /user/cloudera/word_list.txt
[cloudera@quickstart Desktop]$ hdfs dfs -appendToFile word_list.txt shakespeare.txt
[cloudera@quickstart Desktop]$ hdfs dfs -cat shakespeare.txt|head -5
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
cat: Unable to write to output stream.
[cloudera@quickstart Desktop]$ hdfs dfs -cat shakespeare.txt|tail -5
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 Desktop]$ cat > newtextfile.txt
Hi, this is my first Big Data Assignment
[cloudera@quickstart Desktop]$ hdfs dfs -put newtextfile.txt
[cloudera@quickstart Desktop]$ hdfs dfs -ls /user/cloudera
Found 3 items
-rw-r--r--  1 cloudera cloudera      40 2022-01-26 12:43 /user/cloudera/newtextfile.txt
-rw-r--r--  1 cloudera cloudera 5606979 2022-01-26 12:35 /user/cloudera/shakespeare.txt
-rw-r--r--  1 cloudera cloudera   16791 2022-01-26 12:34 /user/cloudera/word_list.txt
[cloudera@quickstart Desktop]$
```

Step7:Appending all the three datasets.

Explanation:Here,we append the datasets(1.shakespeare.txt 2.word_list.txt 3.newtextfile.txt) to a new file called last.txt and then we load the last.txt file into hadoop hdfs



The screenshot shows a terminal window titled "cloudera@quickstart:~/Desktop". The window contains the following commands and output:

```
[cloudera@quickstart Desktop]$ hdfs dfs -getmerge /user/cloudera/shakespeare.txt /user/cloudera/word_list.txt /user/cloudera/newtextfile.txt last.txt
[cloudera@quickstart Desktop]$ hdfs dfs -put last.txt
[cloudera@quickstart Desktop]$
```

The terminal window is part of a desktop environment. The top bar shows "Applications Places System" and "cloudera" with system icons. The bottom bar shows a taskbar with "Hue - File Browser - ..." and "cloudera@quickstart:..." open. The system clock in the top right corner indicates "Wed Jan 26, 12:59 PM".

Step8:Visualizing the all datasets with Hue

The screenshot displays the Hue File Browser interface within a Mozilla Firefox browser window. The browser's address bar shows the URL: `quickstart.cloudera:8888/hue/filebrowser/view%3D/user/cloudera/`. The interface features a sidebar on the left with a 'Tables' section indicating 'The database has no tables'. The main content area is titled 'File Browser' and shows the path `Home / user / cloudera`. A table lists the files and directories in the current view:

	Name	Size	User	Group	Permissions	Date
<input type="checkbox"/>	f		hdfs	supergroup	drwxr-xr-x	January 22, 2022 0
<input type="checkbox"/>	.		cloudera	cloudera	drwxr-xr-x	January 26, 2022 12
<input type="checkbox"/>	last.txt	5.4 MB	cloudera	cloudera	-rw-r--r--	January 26, 2022 12
<input type="checkbox"/>	newtextfile.txt	40 bytes	cloudera	cloudera	-rw-r--r--	January 26, 2022 12
<input type="checkbox"/>	shakespeare.txt	5.3 MB	cloudera	cloudera	-rw-r--r--	January 26, 2022 12
<input type="checkbox"/>	word_list.txt	16.4 KB	cloudera	cloudera	-rw-r--r--	January 26, 2022 12

At the bottom of the file list, there is a pagination control showing 'Show 45 of 4 items' and 'Page 1 of 1'.

