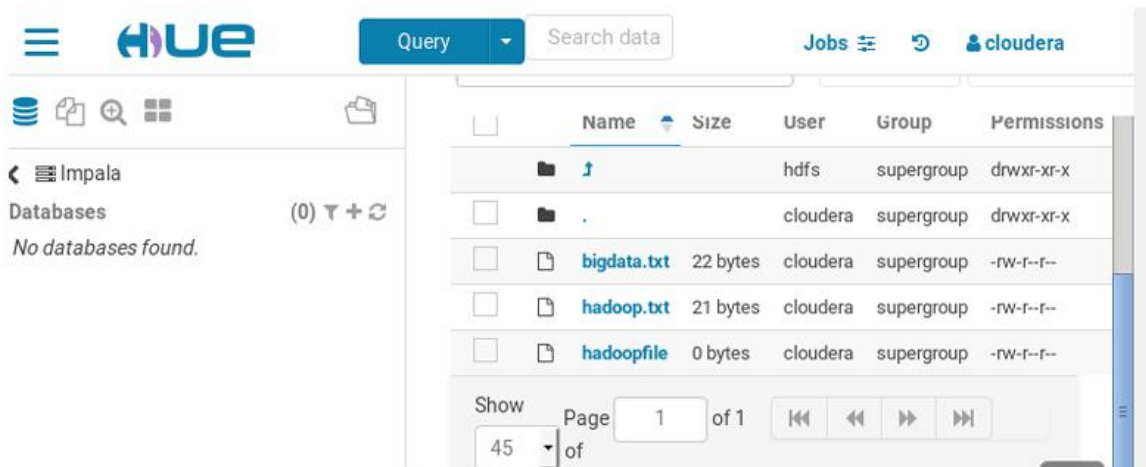


1. I've selected the two text files bigdata, hadoop files. These are the two files that you can view.

```
[cloudera@quickstart ~]$ ls
bigdata.txt      Downloads      kerberos      Public
cloudera-manager eclipse        lib           Templates
cm_api.py        enterprise-deployment.json Music          test1
Desktop          express-deployment.json parcels        Videos
Documents        hadoop.txt     Pictures      workspace
[cloudera@quickstart ~]$
```

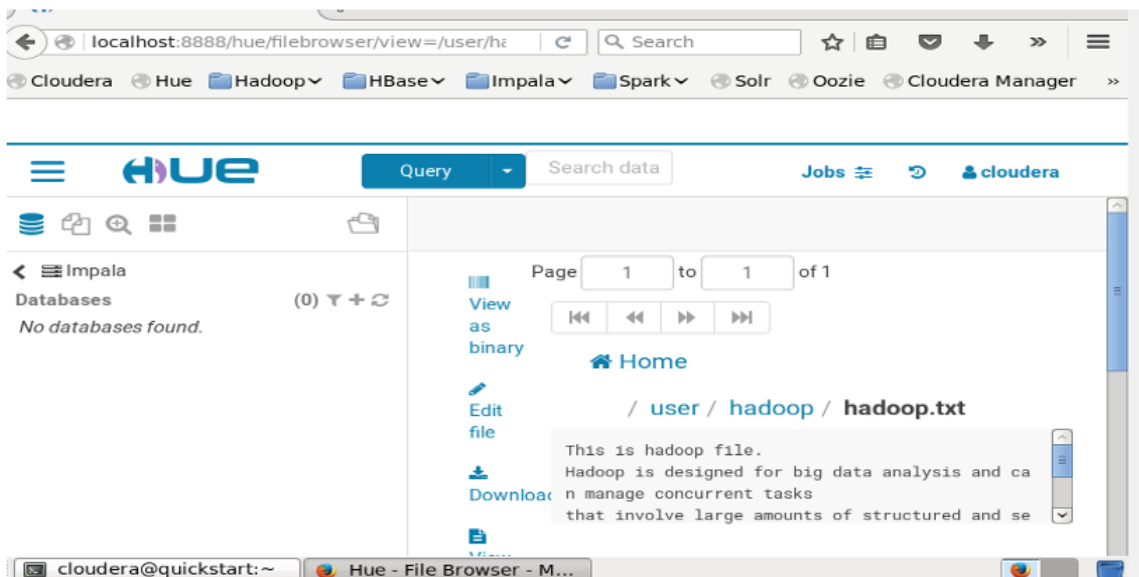
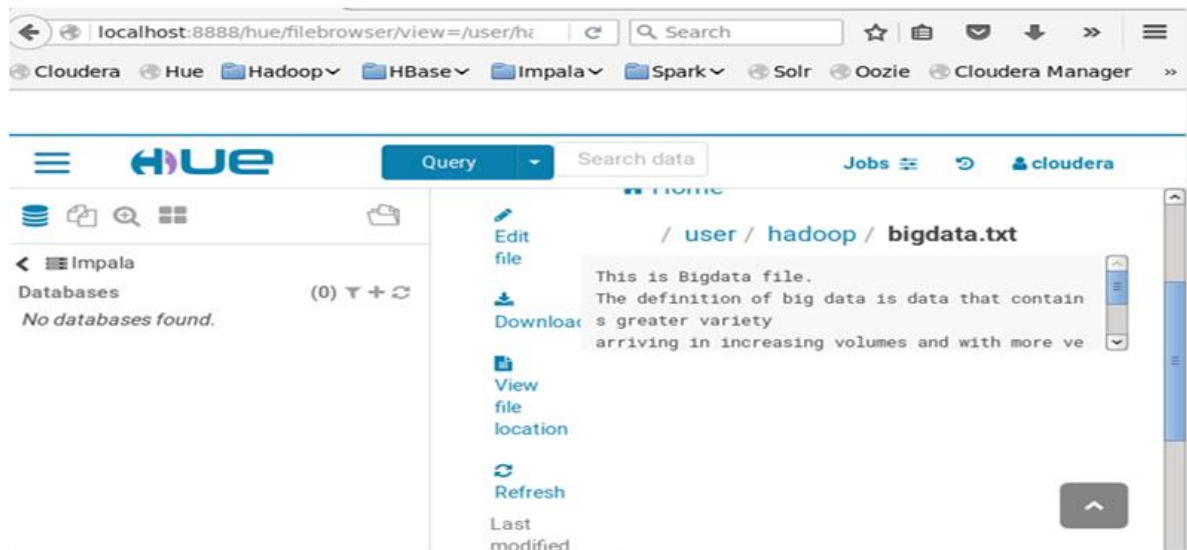
2. In this case, I uploaded the file to the Hadoop server using "hdfs dfs -put." I uploaded the text files Hadoop, bigdata files to the Hadoop server's /user/hadoop folder in my cloudera local.

```
[cloudera@quickstart ~]$ hdfs dfs -put bigdata.txt /user/hadoop/bigdata.txt
[cloudera@quickstart ~]$ hdfs dfs -put hadoop.txt /user/hadoop/hadoop.txt
[cloudera@quickstart ~]$
```



	Name	Size	User	Group	Permissions
	↑		hdfs	supergroup	drwxr-xr-x
	.		cloudera	supergroup	drwxr-xr-x
	bigdata.txt	22 bytes	cloudera	supergroup	-rw-r--r--
	hadoop.txt	21 bytes	cloudera	supergroup	-rw-r--r--
	hadoopfile	0 bytes	cloudera	supergroup	-rw-r--r--

3. I've loaded the files and their contents and am viewing them with Hue. The two .txt files that were uploaded to the Hadoop server are visible in the first screenshot. The contents of both files are shown in the ensuing screen grabs.



4. Here, I merged the two .txt files that were on the Hadoop server using the "hdfs dfs -getmerge" command. The combined file was saved locally and given the name merge.txt. I subsequently uploaded the merge.txt file to the Hadoop server.

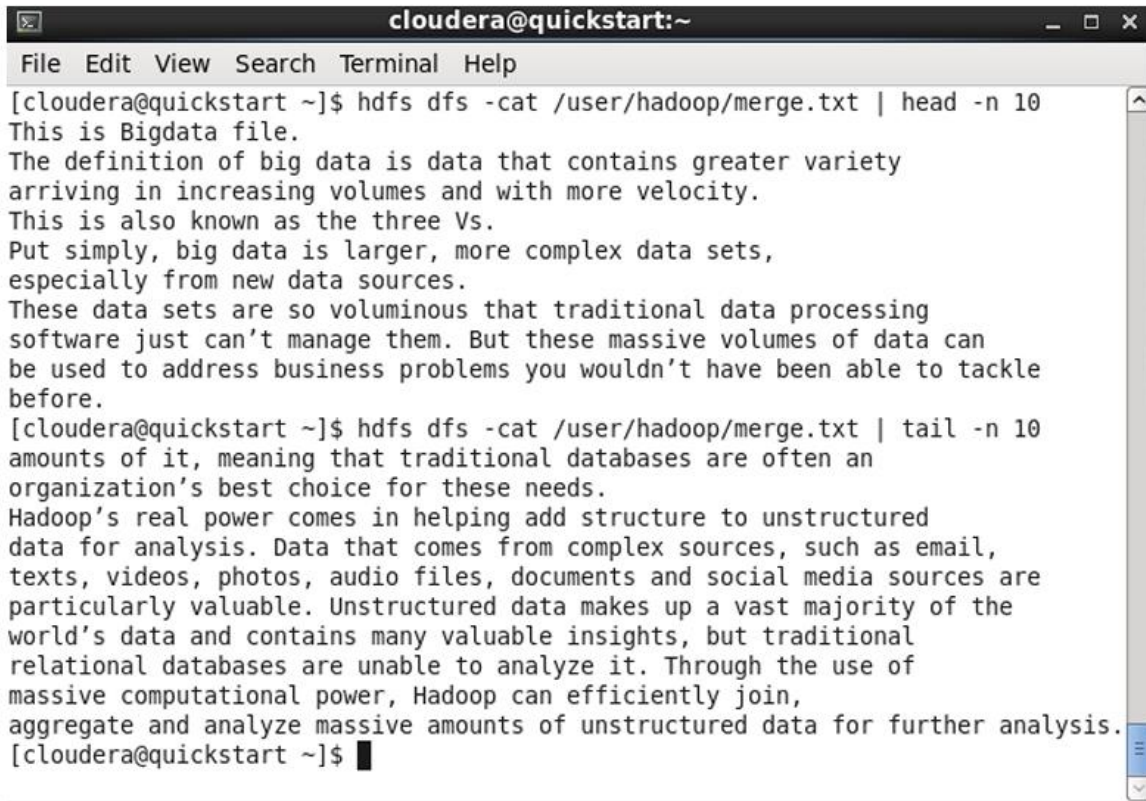
```
[cloudera@quickstart ~]$ hdfs dfs -getmerge /user/hadoop/bigdata.txt /user/hadoop/hadoop.txt merge.txt
[cloudera@quickstart ~]$ vim merge.txt
[cloudera@quickstart ~]$ hdfs dfs -put merge.txt /user/hadoop/merge.txt
[cloudera@quickstart ~]$
```

A screenshot of the Hue web interface showing the file browser view. The left sidebar shows the 'Impala' section with 'Databases' and a message 'No databases found.' The main area displays a file browser for the path '/user/hadoop'. It shows a table of files and directories with columns for Name, Size, User, Group, and Permissions. The files listed are 'bigdata.txt' (498 bytes), 'hadoop.txt' (978 bytes), and 'merge.txt' (1.4 KB), all owned by 'cloudera' and 'supergroup'.

5. I printed the first and last ten lines of the combined file using the commands listed below.

```
hdfs dfs -cat /user/Hadoop/merge.txt | head -n 10
```

```
hdfs dfs -cat /user/Hadoop/merge.txt | tail -n 10
```

A screenshot of a terminal window titled "cloudera@quickstart:~". The window has a menu bar with "File", "Edit", "View", "Search", "Terminal", and "Help". The terminal shows the following commands and output:  
[cloudera@quickstart ~]\$ hdfs dfs -cat /user/hadoop/merge.txt | head -n 10  
This is Bigdata file.  
The definition of big data is data that contains greater variety  
arriving in increasing volumes and with more velocity.  
This is also known as the three Vs.  
Put simply, big data is larger, more complex data sets,  
especially from new data sources.  
These data sets are so voluminous that traditional data processing  
software just can't manage them. But these massive volumes of data can  
be used to address business problems you wouldn't have been able to tackle  
before.  
[cloudera@quickstart ~]\$ hdfs dfs -cat /user/hadoop/merge.txt | tail -n 10  
amounts of it, meaning that traditional databases are often an  
organization's best choice for these needs.  
Hadoop's real power comes in helping add structure to unstructured  
data for analysis. Data that comes from complex sources, such as email,  
texts, videos, photos, audio files, documents and social media sources are  
particularly valuable. Unstructured data makes up a vast majority of the  
world's data and contains many valuable insights, but traditional  
relational databases are unable to analyze it. Through the use of  
massive computational power, Hadoop can efficiently join,  
aggregate and analyze massive amounts of unstructured data for further analysis.  
[cloudera@quickstart ~]\$