**Big Data Programming**

**ICP -1**

**Introduction:**

* Cloudera, provides a software platform for [data engineering](https://en.wikipedia.org/wiki/Data_engineering), [data warehousing](https://en.wikipedia.org/wiki/Data_warehouse), [machine learning](https://en.wikipedia.org/wiki/Machine_learning) and analytics that runs in the cloud or on premises.

**Installations:**

The following are the screenshots of Cloudera Setup File and Oracle VirtualBox Manager where Cloudera Environment was setup.

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

**In Class Exercise:**

**A screenshot of a social media post

Description automatically generated**

* So, here in order to load the dataset Shakespeare.txt into hdfs we have used the command Hadoop fs -put /home/cloudera/Dowloads/Shakespeare.txt and then in order to append the second dataset word\_list.txt to the first dataset we have used the command Hadoop fs -appendToFile /home/cloudera/Downloads/word\_list.txt Shakespeare.txt.
* In order to view the first and the last 5 lines of the merged dataset and store them in a file we have used the commands hadoop fs -cat shakespeare.txt | head -n 5 | hadoop fs -put - first5lines.txt and hadoop fs -cat shakespeare.txt | tail -n 5 | hadoop fs -put – last5lines.txt
* And all the three datasets are of the form .txt, the following command appends all .txt files and then the appended data is stored in a HDFS file outputfinalfile.txt

**A screenshot of a cell phone

Description automatically generated**

**Output:**

**Hadoop User Interface:**

This is the screenshot of the dataset Shakespeare.txt to which word\_list.txt has been appended.

**A screenshot of a cell phone

Description automatically generated**

This is the screenshot of the file which contains the first 5 lines of the dataset in which both the datasets Shakespeare.txt and word\_list.txt are merged.

**A screenshot of a cell phone

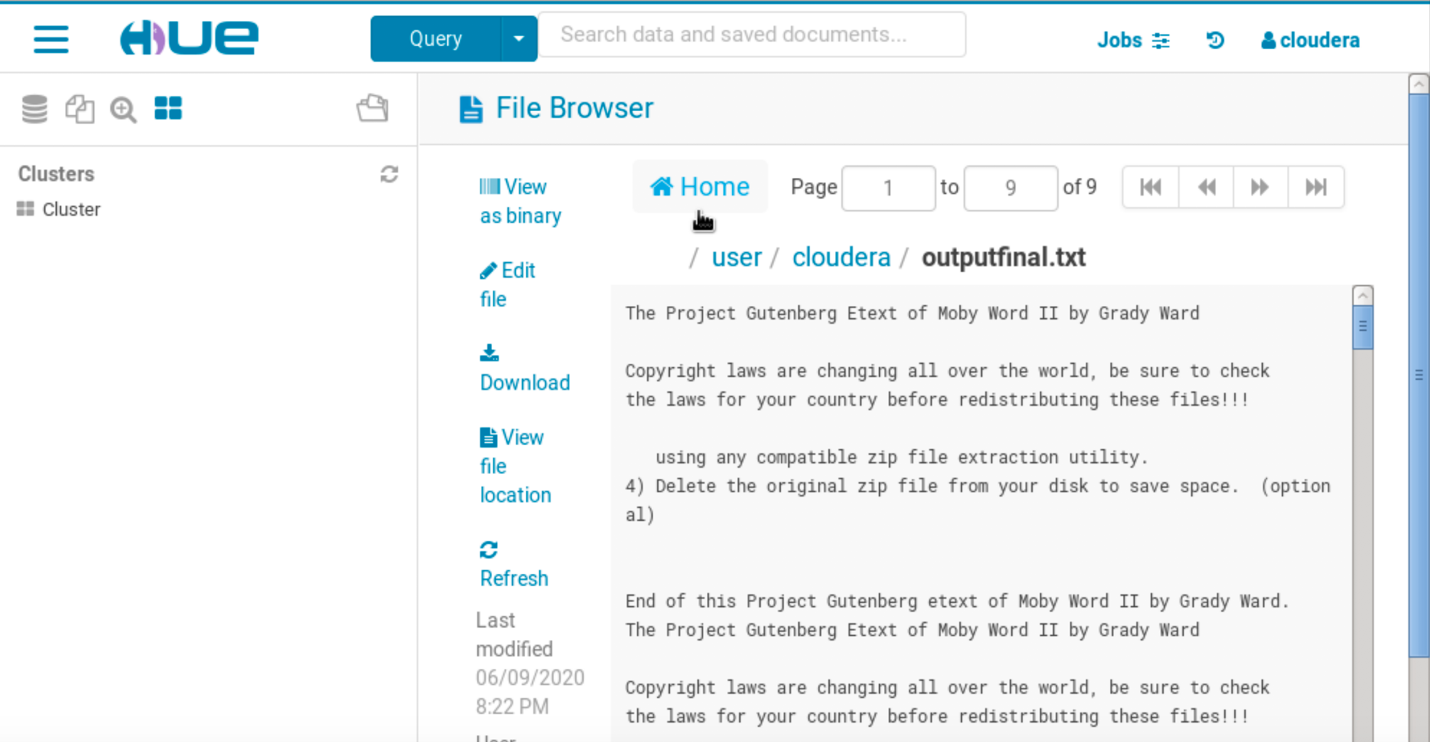
Description automatically generated**

This is the screenshot of the file which contains the last 5 lines of the dataset in which both the datasets Shakespeare.txt and word\_list.txt are merged.

**A screenshot of a cell phone

Description automatically generated**

This is the screenshot of the file in which all the three datasets are merged.

****