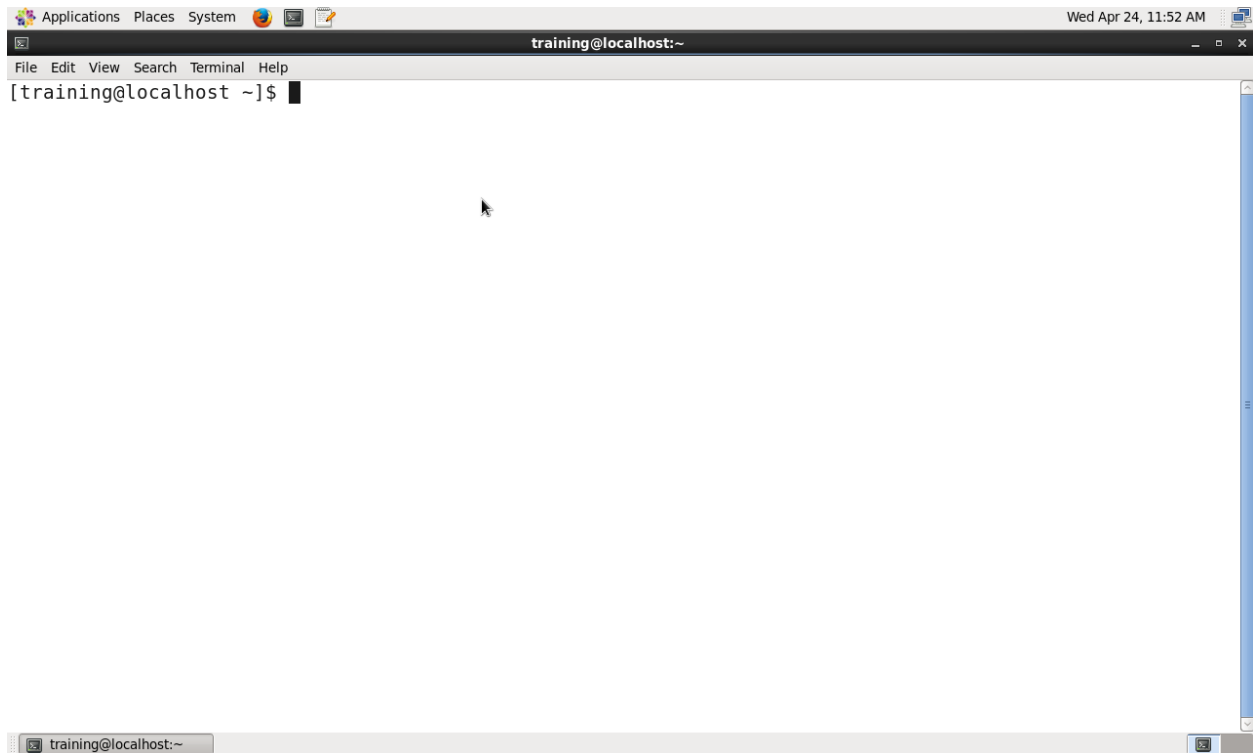
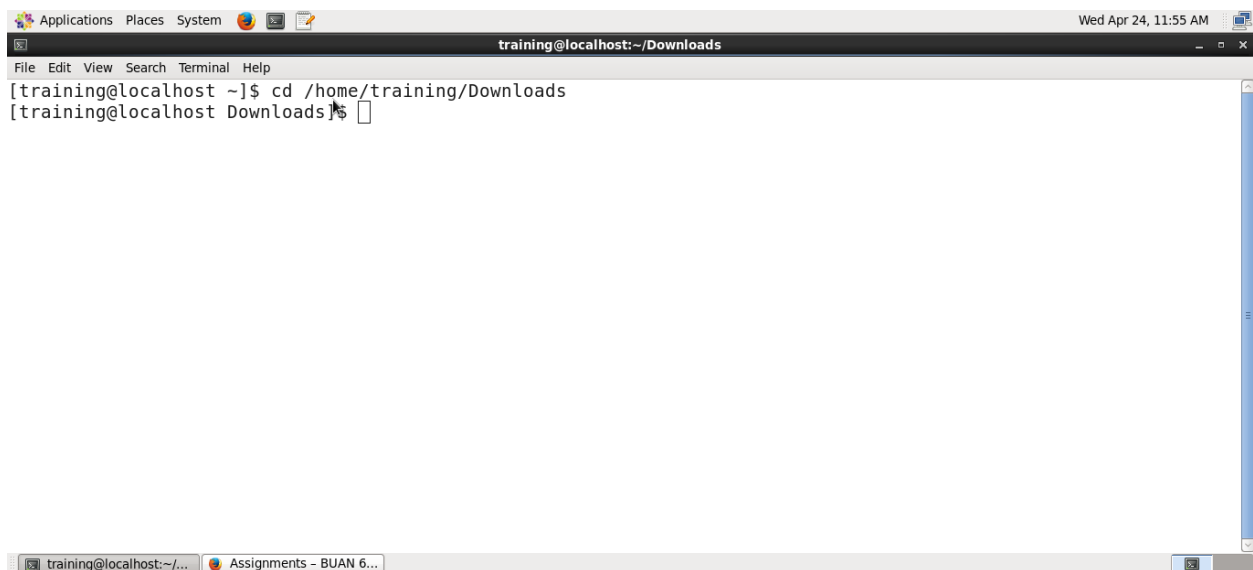


Sentiment Analysis Using Hadoop Tools

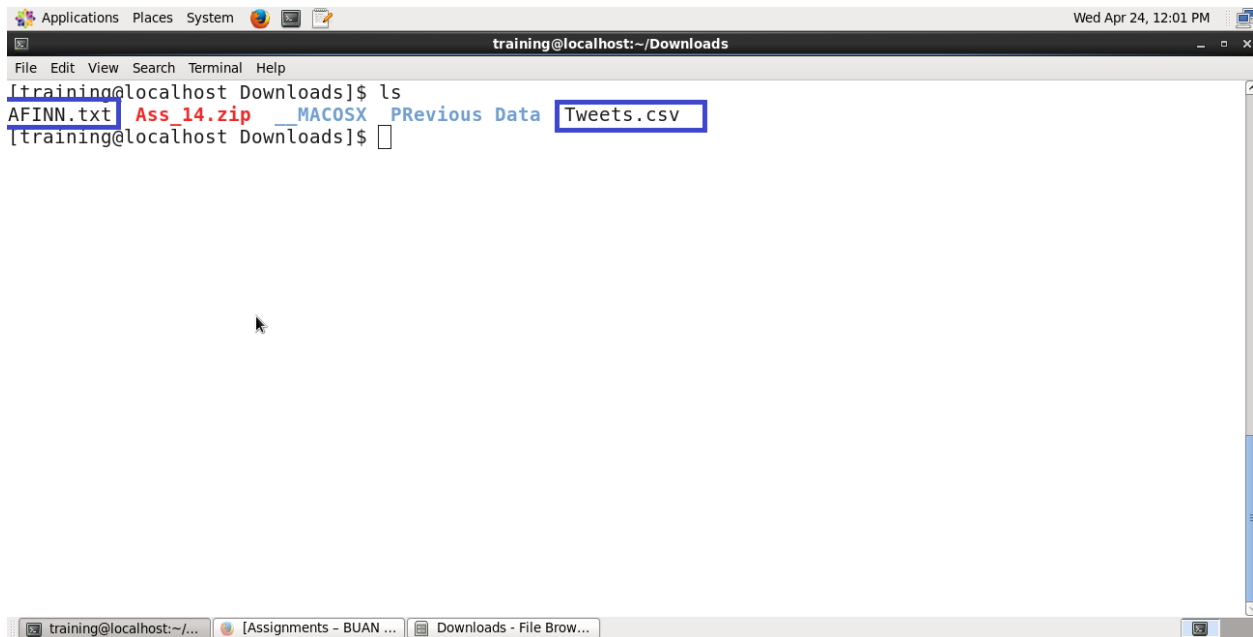
Step-1:



Step-2: Getting the Data from Download folder



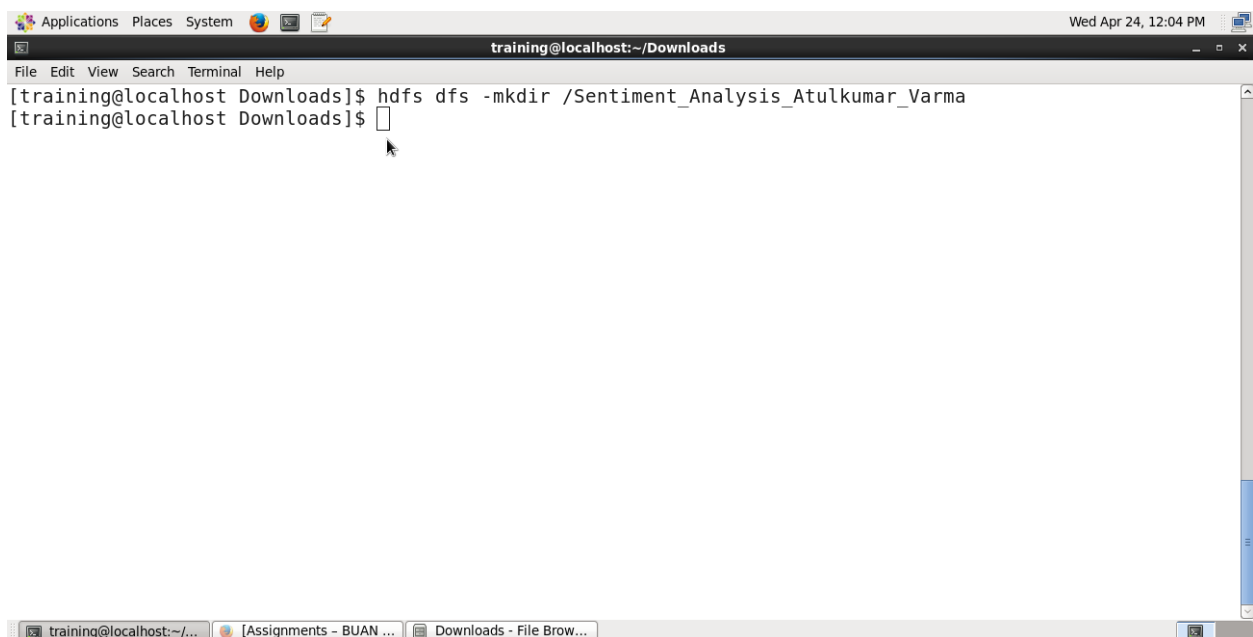
Step-3: Contents of the Downloads folder. Contains all the tweets file.



A terminal window titled "training@localhost: ~/Downloads" is shown. The window has a menu bar with "File", "Edit", "View", "Search", "Terminal", and "Help". The terminal output shows the command "ls" being executed, resulting in a list of files: "AFINN.txt", "Ass_14.zip", "_MACOSX", "Previous Data", and "Tweets.csv". The "Tweets.csv" file is highlighted with a blue box. The terminal prompt is "[training@localhost Downloads]\$".

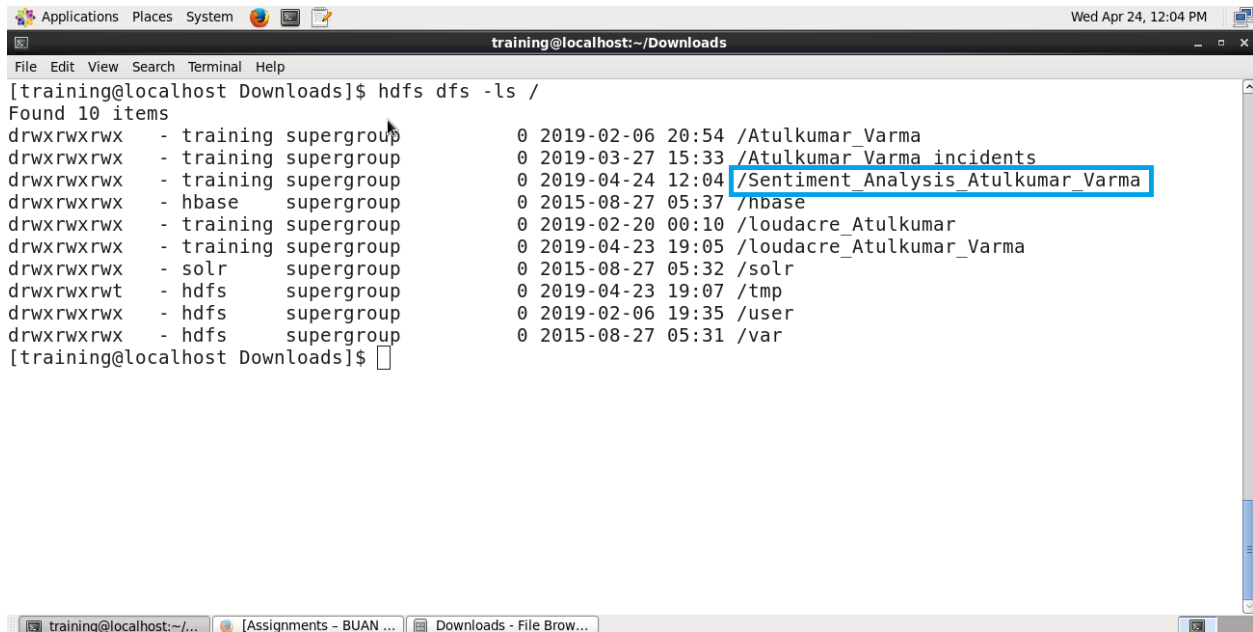
```
training@localhost: ~/Downloads
File Edit View Search Terminal Help
[training@localhost Downloads]$ ls
AFINN.txt Ass_14.zip _MACOSX Previous Data Tweets.csv
[training@localhost Downloads]$
```

Step-4: We are creating a directory at the root directory.



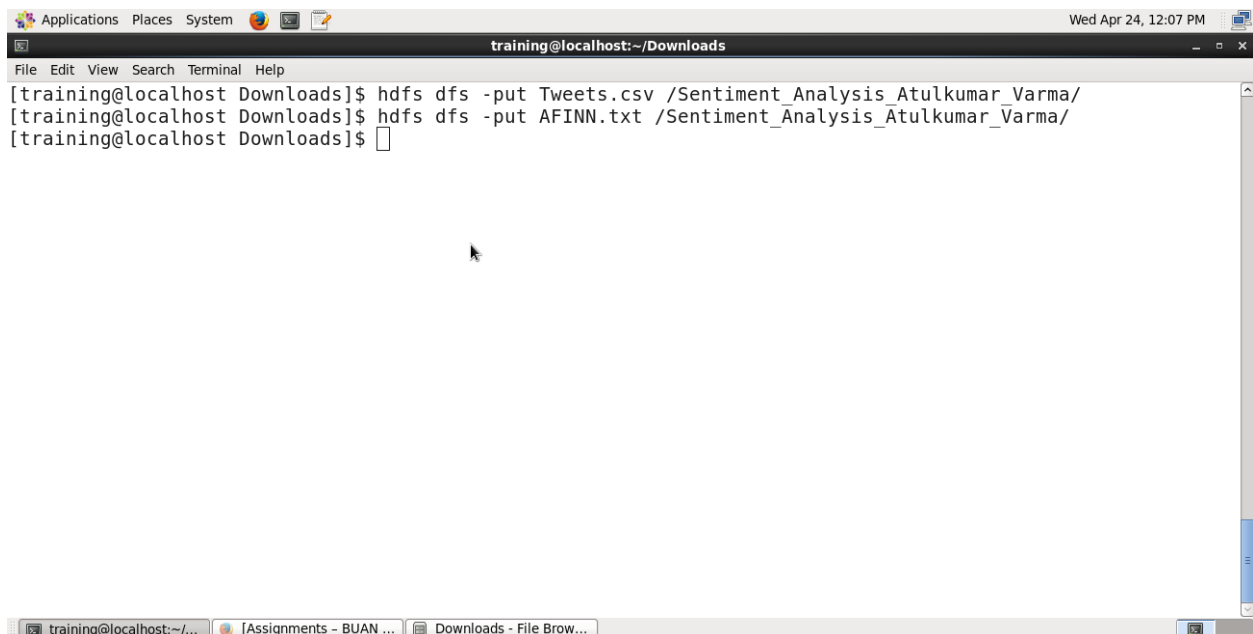
A terminal window titled "training@localhost: ~/Downloads" is shown. The window has a menu bar with "File", "Edit", "View", "Search", "Terminal", and "Help". The terminal output shows the command "hdfs dfs -mkdir /Sentiment_Analysis_Atulkumar_Varma" being executed. The terminal prompt is "[training@localhost Downloads]\$".

```
training@localhost: ~/Downloads
File Edit View Search Terminal Help
[training@localhost Downloads]$ hdfs dfs -mkdir /Sentiment_Analysis_Atulkumar_Varma
[training@localhost Downloads]$
```

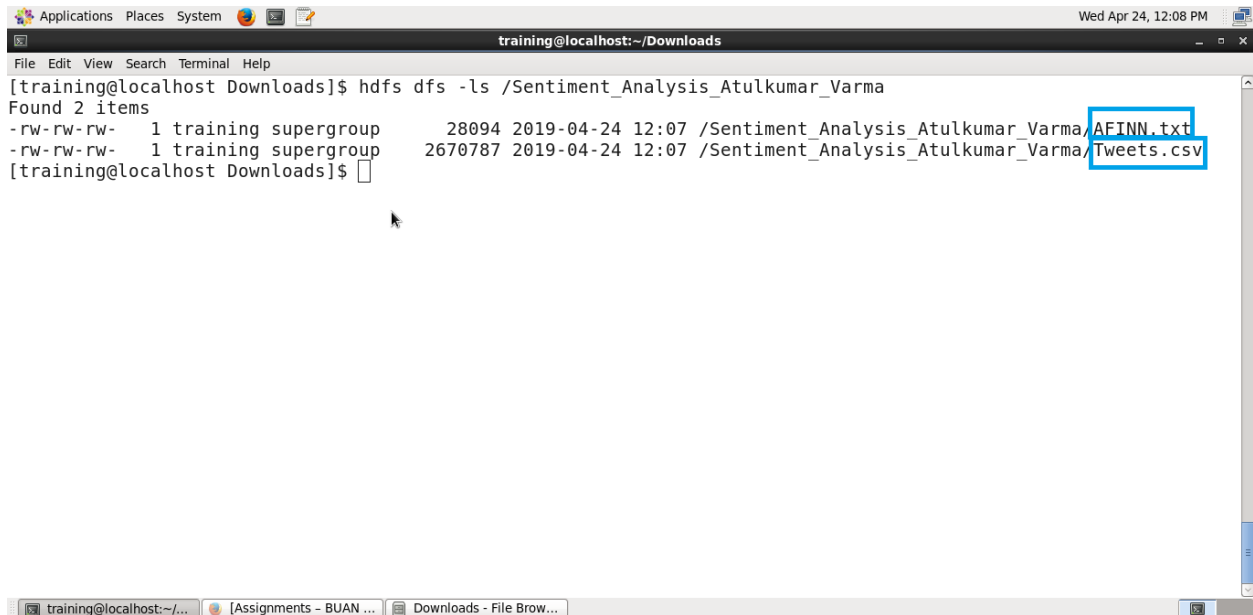
A terminal window titled 'training@localhost:~/Downloads' showing the output of the command 'hdfs dfs -ls /'. The output lists 10 items in the root directory, including files like 'Atulkumar_Varma', 'Atulkumar_Varma_incidents', and 'Sentiment_Analysis_Atulkumar_Varma' (which is highlighted with a blue box). The window also shows a taskbar at the bottom with icons for Applications, Places, System, and a file browser.

```
[training@localhost Downloads]$ hdfs dfs -ls /
Found 10 items
drwxrwxrwx - training supergroup 0 2019-02-06 20:54 /Atulkumar_Varma
drwxrwxrwx - training supergroup 0 2019-03-27 15:33 /Atulkumar_Varma_incidents
drwxrwxrwx - training supergroup 0 2019-04-24 12:04 /Sentiment_Analysis_Atulkumar_Varma
drwxrwxrwx - hbase supergroup 0 2015-08-27 05:37 /hbase
drwxrwxrwx - training supergroup 0 2019-02-20 00:10 /loudacre_Atulkumar
drwxrwxrwx - training supergroup 0 2019-04-23 19:05 /loudacre_Atulkumar_Varma
drwxrwxrwx - solr supergroup 0 2015-08-27 05:32 /solr
drwxrwxrwt - hdfs supergroup 0 2019-04-23 19:07 /tmp
drwxrwxrwx - hdfs supergroup 0 2019-02-06 19:35 /user
drwxrwxrwx - hdfs supergroup 0 2015-08-27 05:31 /var
[training@localhost Downloads]$
```

Step-5: Putting the data. We are copying the data files required into the directory that we created. AFINN dictionary is a dictionary consisting of 2500 words that are rated on a scale from -5 to +5 depending on the sentiment the word conveys.

A terminal window titled 'training@localhost:~/Downloads' showing two commands being executed to upload files to HDFS. The first command uploads 'Tweets.csv' and the second uploads 'AFINN.txt', both to the path '/Sentiment_Analysis_Atulkumar_Varma/'. The window also shows a taskbar at the bottom with icons for Applications, Places, System, and a file browser.

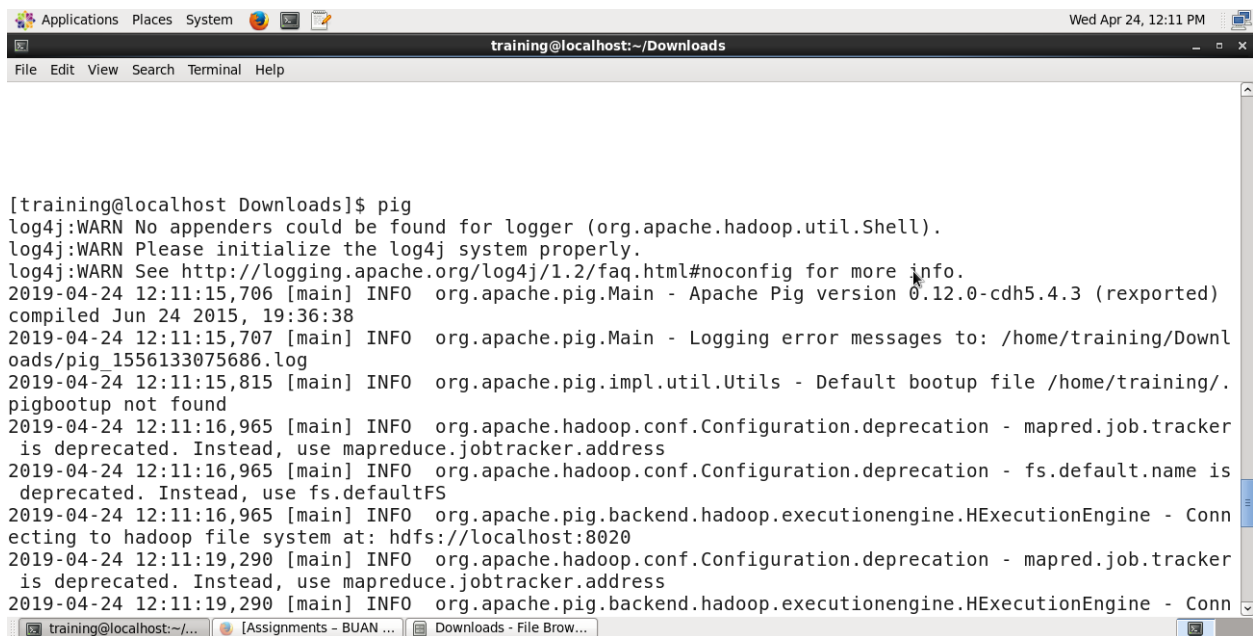
```
[training@localhost Downloads]$ hdfs dfs -put Tweets.csv /Sentiment_Analysis_Atulkumar_Varma/
[training@localhost Downloads]$ hdfs dfs -put AFINN.txt /Sentiment_Analysis_Atulkumar_Varma/
[training@localhost Downloads]$
```



A terminal window titled 'training@localhost:~/Downloads' showing the execution of the command 'hdfs dfs -ls /Sentiment_Analysis_Atulkumar_Varma'. The output lists two files: 'AFINN.txt' and 'Tweets.csv', both with permissions '-rw-rw-rw-', owned by 'training' and 'supergroup', and located in the directory '/Sentiment_Analysis_Atulkumar_Varma/'. The file 'Tweets.csv' is highlighted with a blue box.

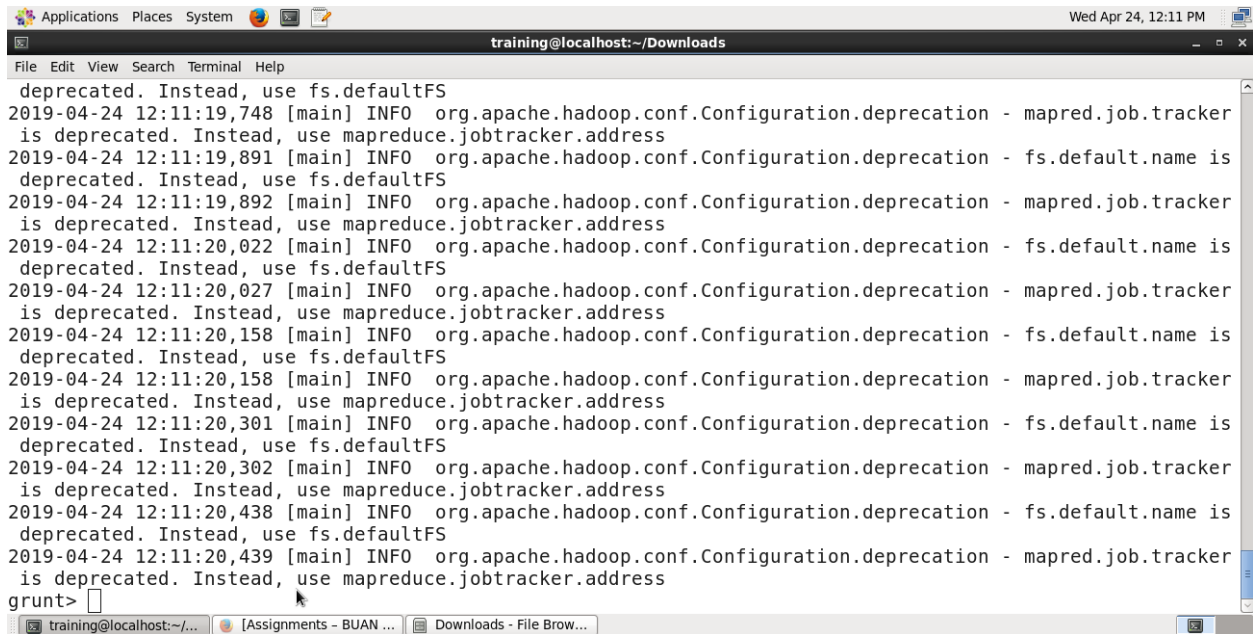
```
training@localhost:~/Downloads
[training@localhost Downloads]$ hdfs dfs -ls /Sentiment_Analysis_Atulkumar_Varma
Found 2 items
-rw-rw-rw- 1 training supergroup      28094 2019-04-24 12:07 /Sentiment_Analysis_Atulkumar_Varma/AFINN.txt
-rw-rw-rw- 1 training supergroup    2670787 2019-04-24 12:07 /Sentiment_Analysis_Atulkumar_Varma/Tweets.csv
[training@localhost Downloads]$
```

Step-6: Pig command invokes the grunt shell.



A terminal window titled 'training@localhost:~/Downloads' showing the execution of the command 'pig'. The output displays various log messages, including warnings about the log4j system and information about the Apache Pig version (0.12.0-cdh5.4.3) and its configuration. The output is truncated at the bottom.

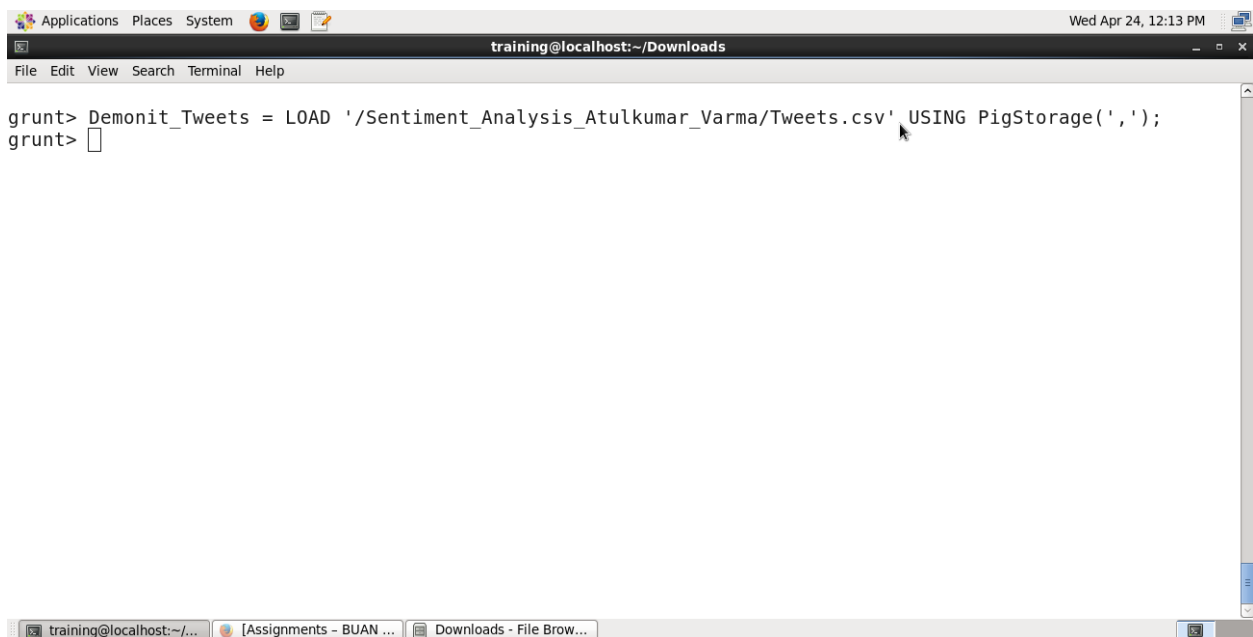
```
training@localhost:~/Downloads
[training@localhost Downloads]$ pig
log4j:WARN No appenders could be found for logger (org.apache.hadoop.util.Shell).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.
2019-04-24 12:11:15,706 [main] INFO org.apache.pig.Main - Apache Pig version 0.12.0-cdh5.4.3 (rexported)
compiled Jun 24 2015, 19:36:38
2019-04-24 12:11:15,707 [main] INFO org.apache.pig.Main - Logging error messages to: /home/training/Downl
oads/pig_1556133075686.log
2019-04-24 12:11:15,815 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/training/.
pigbootup not found
2019-04-24 12:11:16,965 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker
is deprecated. Instead, use mapreduce.jobtracker.address
2019-04-24 12:11:16,965 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is
deprecated. Instead, use fs.defaultFS
2019-04-24 12:11:16,965 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Conn
ecting to hadoop file system at: hdfs://localhost:8020
2019-04-24 12:11:19,290 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker
is deprecated. Instead, use mapreduce.jobtracker.address
2019-04-24 12:11:19,290 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Conn
```



A screenshot of a terminal window titled 'training@localhost:~/Downloads'. The window shows a series of log messages from the Hadoop configuration system. Each message follows the pattern: '2019-04-24 12:11:19,748 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address'. The messages are repeated for various configuration properties, including 'mapred.job.tracker', 'fs.default.name', and 'fs.defaultFS'. The terminal window has a menu bar with 'File', 'Edit', 'View', 'Search', 'Terminal', and 'Help'. The status bar at the bottom shows the current directory as '~/Downloads' and the user as 'training@localhost'.

```
deprecated. Instead, use fs.defaultFS
2019-04-24 12:11:19,748 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker
is deprecated. Instead, use mapreduce.jobtracker.address
2019-04-24 12:11:19,891 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is
deprecated. Instead, use fs.defaultFS
2019-04-24 12:11:19,892 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker
is deprecated. Instead, use mapreduce.jobtracker.address
2019-04-24 12:11:20,022 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is
deprecated. Instead, use fs.defaultFS
2019-04-24 12:11:20,027 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker
is deprecated. Instead, use mapreduce.jobtracker.address
2019-04-24 12:11:20,158 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is
deprecated. Instead, use fs.defaultFS
2019-04-24 12:11:20,158 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker
is deprecated. Instead, use mapreduce.jobtracker.address
2019-04-24 12:11:20,301 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is
deprecated. Instead, use fs.defaultFS
2019-04-24 12:11:20,302 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker
is deprecated. Instead, use mapreduce.jobtracker.address
2019-04-24 12:11:20,438 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is
deprecated. Instead, use fs.defaultFS
2019-04-24 12:11:20,439 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker
is deprecated. Instead, use mapreduce.jobtracker.address
grunt>
```

Step-7: The dump command displays the data for the variable `Demonit_Tweets` specified in the grunt shell.



A screenshot of a terminal window titled 'training@localhost:~/Downloads'. The window shows the execution of a grunt command. The command is: 'grunt> Demonit_Tweets = LOAD '/Sentiment_Analysis_Atulkumar_Varma/Tweets.csv' USING PigStorage(',');'. The terminal window has a menu bar with 'File', 'Edit', 'View', 'Search', 'Terminal', and 'Help'. The status bar at the bottom shows the current directory as '~/Downloads' and the user as 'training@localhost'.

```
grunt> Demonit_Tweets = LOAD '/Sentiment_Analysis_Atulkumar_Varma/Tweets.csv' USING PigStorage(',');
grunt>
```

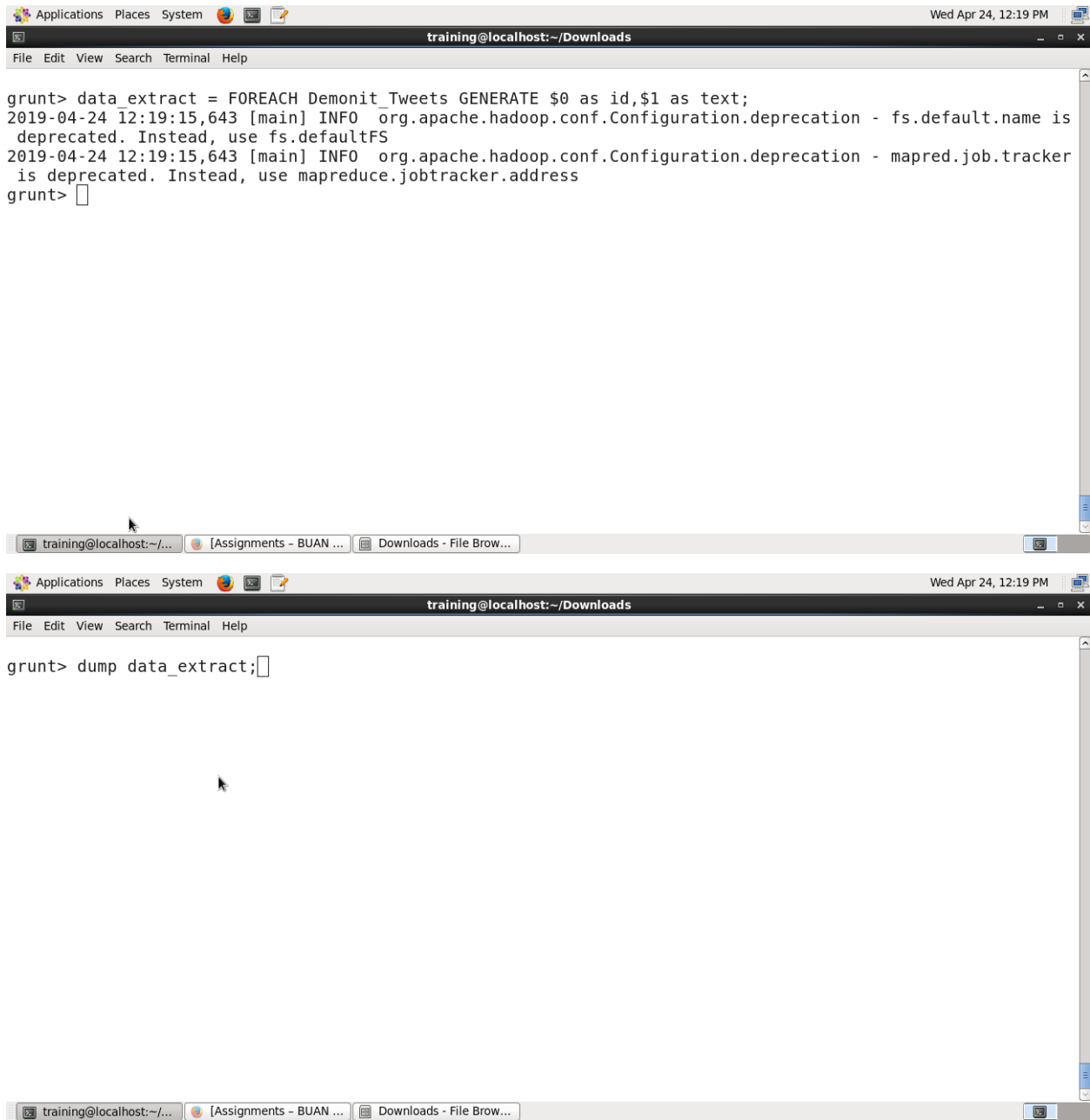
```
Applications Places System Wed Apr 24, 12:17 PM
training@localhost:~/Downloads
File Edit View Search Terminal Help

grunt> dump Demonit_Tweets;[]
```

```
Applications Places System Wed Apr 24, 12:15 PM
training@localhost:~/Downloads
File Edit View Search Terminal Help

("7995", "@baliramsingh2 So many restrictions. Not easy to avail the facility by anyone. Multiple U-turns b
y GOI on the issue. #DeMonetization #RBI", FALSE, 0, "baliramsingh2", "2016-11-22 10:58:03", FALSE, "80101126280
5012480", "801016889082163200", "2275972058", "<a href='\"http://twitter.com\"' rel='\"nofollow\"'>Twitter Web Cl
ient</a>", "RahulCipher", 0, FALSE, FALSE)
("7996", "FYI, I must tell and clear your doubts to enhance your GENERAL KNOWLEDGE that this wholesome expe
riment of #DEMONETIZATION is only", FALSE, 0, NA, "2016-11-22 10:57:59", FALSE, NA, "801016871856140289", NA, "<a h
ref='\"http://www.twitter.com\"' rel='\"nofollow\"'>Twitter for Windows Phone</a>", "Rishimathur18", 0, FALSE, FALSE)
("7997", "RT @sukanyaiyer2: #DeMonetization AAP protests by marching Against Govts move over DeMonetization
& he is also detained as he Tried 2 March", FALSE, 0, NA, "2016-11-22 10:57:58", FALSE, NA, "80101686747322
7776", NA, "<a href='\"http://twitter.com/download/android\"' rel='\"nofollow\"'>Twitter for Android</a>", "asita
wasthi_", 2, TRUE, FALSE)
("7998", "#demonetization will help combat terror because Pak won't be able to print new notes! And now, th
is. *slow claps* https://t.co/p0p93sXiMn", FALSE, 2, NA, "2016-11-22 10:57:44", FALSE, NA, "801016810006945792", NA, "<a href='\"http://twitter.com/download/iphone\"' rel='\"nofollow\"'>Twitter for iPhone</a>", "Modern_Gypsy",
1, FALSE, FALSE)
("7999", "#Demonetization Positive Effect: Cigarette Sales Slashed To Almost Half After Demonetization:Repo
rt @deepparihar @Tabassumzia @AshSharma1102", FALSE, 2, NA, "2016-11-22 10:57:43", FALSE, NA, "801016807825940481
", NA, "<a href='\"http://twitter.com\"' rel='\"nofollow\"'>Twitter Web Client</a>", "HP_Journo", 1, FALSE, FALSE)
("8000", "RT @UnSubtleDesi: Kejriwal posts pic of dead robber and claims it's #demonetization related death
? How shameless has this man become? https://t.co/...", FALSE, 0, NA, "2016-11-22 10:57:33", FALSE, NA, "801016763722797057
", NA, "<a href='\"http://twitter.com\"' rel='\"nofollow\"'>Twitter Web Client</a>", "prashik2107", 897, TRUE, FALSE)
)
grunt>
```

Step-8: We are defining and collecting the first two data elements in the csv file and dumping the data at the grunt shell.



The image shows a terminal window titled 'training@localhost:~/Downloads'. The window has a menu bar with 'File', 'Edit', 'View', 'Search', 'Terminal', and 'Help'. The terminal output shows the following commands and messages:

```
grunt> data_extract = FOREACH Demonit_Tweets GENERATE $0 as id,$1 as text;
2019-04-24 12:19:15,643 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2019-04-24 12:19:15,643 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
grunt> 
```

Below this, there is a task runner bar showing 'training@localhost:~/...' and two tabs: '[Assignments - BUAN ...]' and 'Downloads - File Brow...'. The terminal window is then shown again, displaying the command:

```
grunt> dump data_extract;
```

```
Applications Places System training@localhost:~/Downloads Wed Apr 24, 12:20 PM
File Edit View Search Terminal Help
early fishy and requires full disclosure &#0")
("7990","All weddings now need to be approved by RBI... Amazing times #demonetization isn't that what we a
re understanding")
("7991","RT @DrKumarVishwas: And the Oscar goes to ""Mr.<U+092D><U+093E><U+0935><U+0941><U+0915>"" <ed><U+
00A0><U+00BD><ed><U+00B8><U+00A9><ed><U+00A0><U+00BD><ed><U+00B8><U+00A5><ed><U+00A0><U+00BD><ed><U+00B8><
U+00A2><ed><U+00A0><U+00BD><ed><U+00B8><U+00AD>#demonetization https://t.co/ObQrhlnSL6")
("7992","RT @jan14anurag: Terrorists raided a bank to get new <U+20B9>2000 notes => their old finances
are indeed squeezed tight #Demonetization)
(https://t0",FALSE)
("7993","RT @jackerhack: Indore's collector would like you to shut up about #demonetization. At @internetf
reedom we think that is a problem. https://0")
("7994","@quizderek Post #Dmonetization the result will be totally different.The win is not because of #de
monetization an all knows about it")
("7995","@baliramsingh2 So many restrictions. Not easy to avail the facility by anyone. Multiple U-turns b
y GOI on the issue. #DeMonetization #RBI")
("7996","FYI)
("7997","RT @sukanyaiyer2: #DeMonetization AAP protests by marching Against Govts move over DeMonetization
&#0; he is also detained as he Tried 2 March0")
("7998","#demonetization will help combat terror because Pak won't be able to print new notes! And now)
("7999","#Demonetization Positive Effect: Cigarette Sales Slashed To Almost Half After Demonetization:Repo
rt @deeparihar @Tabassumzia @AshSharma1102")
("8000","RT @UnSubtleDesi: Kejriwal posts pic of dead robber and claims it's #demonetization related death
? How shameless has this man become? https0")
grunt>
```

Step-9: We are defining a third element by breaking down the text information in the variable 'text' into words. The describe operation shows us the variable names and types of the data in the variable specified.

```
Applications Places System training@localhost:~/Downloads Wed Apr 24, 12:22 PM
File Edit View Search Terminal Help
grunt> data tokens = FOREACH data extract generate id,text,FLATTEN(TOKENIZE(text)) as words;
2019-04-24 12:22:16,507 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARA
RRAY 1 time(s).
2019-04-24 12:22:16,507 [main] WARN org.apache.pig.PigServer - Encountered Warning USING_OVERLOADED_FUNCT
ION 1 time(s).
grunt>
```


Applications Places System Wed Apr 24, 12:22 PM
training@localhost:~/Downloads
File Edit View Search Terminal Help

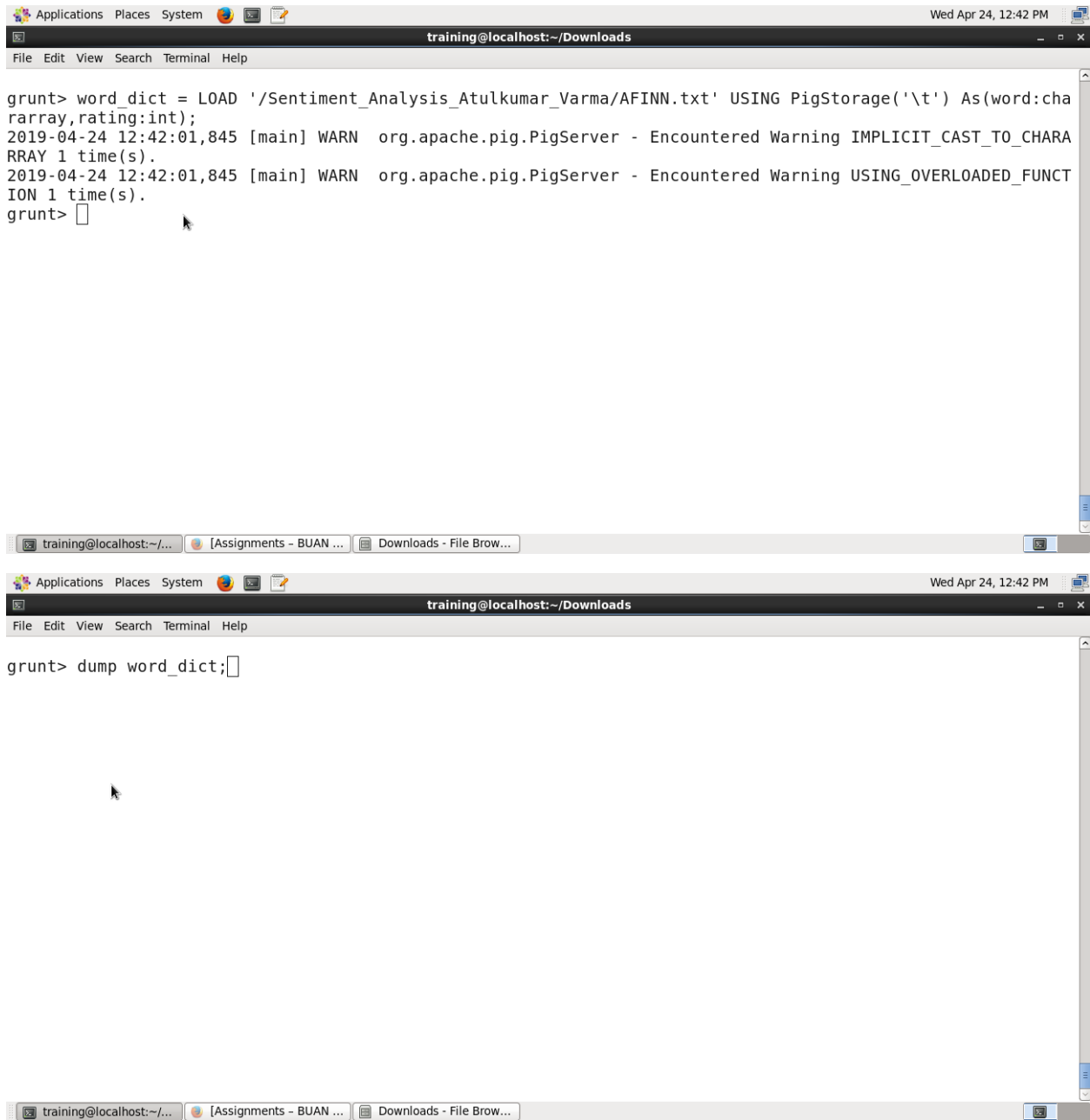
```
grunt> describe data_tokens;[]
```

training@localhost:~/... [Assignments - BUAN ... Downloads - File Brow...
Applications Places System Wed Apr 24, 12:23 PM
training@localhost:~/Downloads
File Edit View Search Terminal Help

```
grunt> describe data_tokens;  
2019-04-24 12:22:53,956 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARA  
RRAY 1 time(s).  
2019-04-24 12:22:53,956 [main] WARN org.apache.pig.PigServer - Encountered Warning USING_OVERLOADED_FUNC  
TION 1 time(s).  
data_tokens: {id: bytearray,text: bytearray,words: chararray}  
grunt> []
```

training@localhost:~/... [Assignments - BUAN ... Downloads - File Brow...

Step-10: We will be using AFINN dictionary to perform the sentiment analysis on our tweet data.



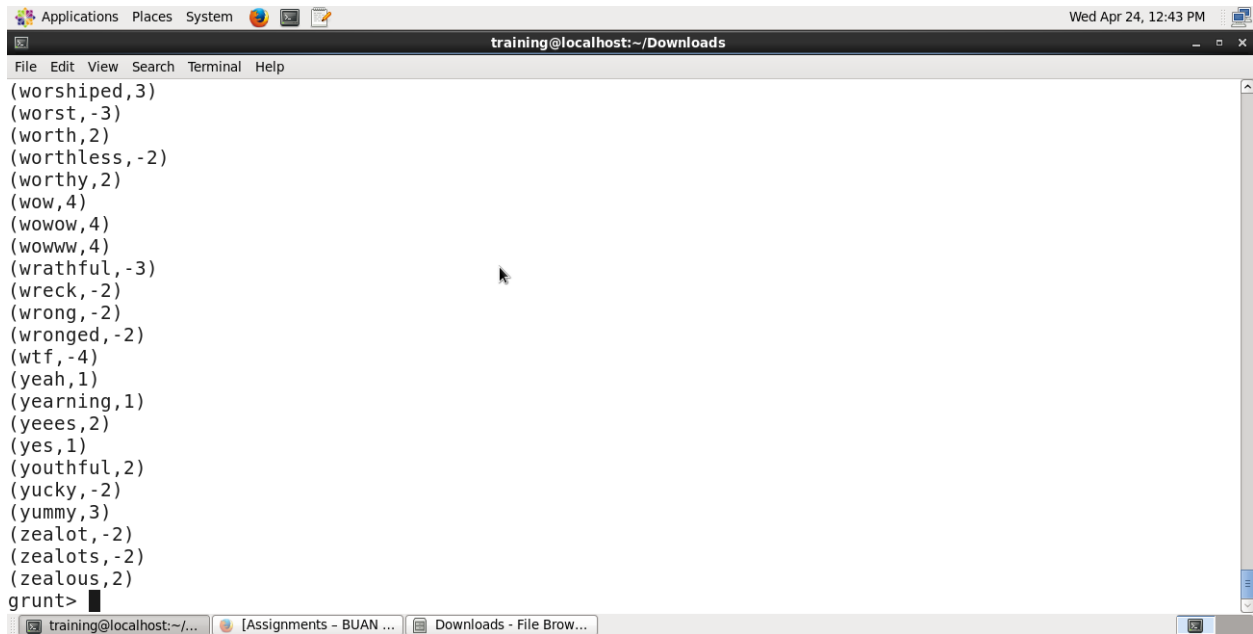
The image shows two screenshots of a terminal window. The top screenshot shows the execution of a Grunt command to load an AFINN dictionary. The command is: `grunt> word_dict = LOAD '/Sentiment_Analysis_Atulkumar_Varma/AFINN.txt' USING PigStorage('\t') As(word:chararray, rating:int);`. The output shows two warnings from the PigServer: `WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).` and `WARN org.apache.pig.PigServer - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).`. The bottom screenshot shows the command `grunt> dump word_dict;` being entered, but no output is visible yet.

```
training@localhost:~/Downloads
File Edit View Search Terminal Help

grunt> word_dict = LOAD '/Sentiment_Analysis_Atulkumar_Varma/AFINN.txt' USING PigStorage('\t') As(word:chararray, rating:int);
2019-04-24 12:42:01,845 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
2019-04-24 12:42:01,845 [main] WARN org.apache.pig.PigServer - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
grunt> 
```

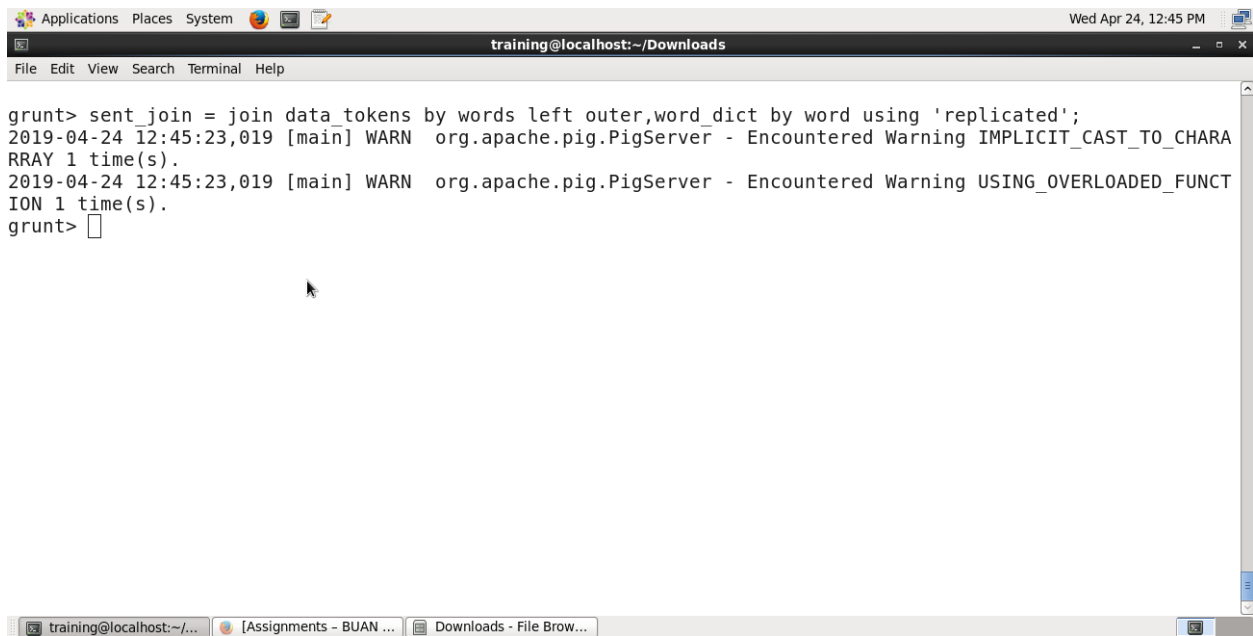
```
training@localhost:~/Downloads
File Edit View Search Terminal Help

grunt> dump word_dict;
```



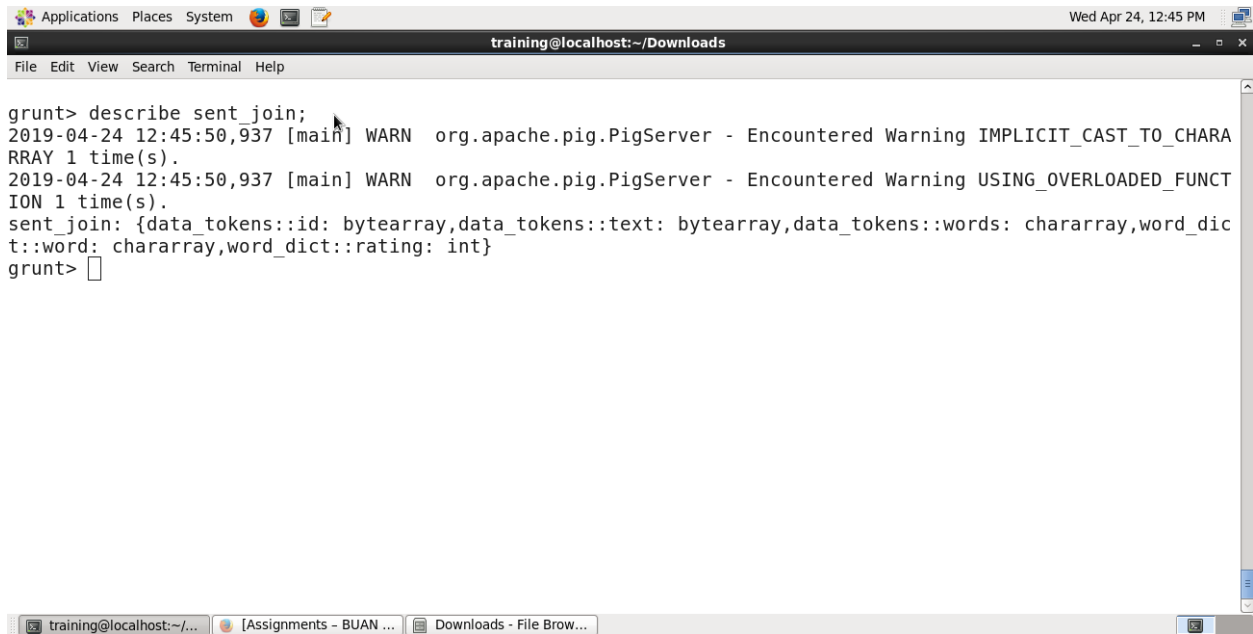
```
training@localhost:~/Downloads
File Edit View Search Terminal Help
(worshiped,3)
(worst,-3)
(worth,2)
(worthless,-2)
(worthy,2)
(wow,4)
(wowow,4)
(wowww,4)
(wrathful,-3)
(wreck,-2)
(wrong,-2)
(wronged,-2)
(wtf,-4)
(yeah,1)
(yearning,1)
(yees,2)
(yes,1)
(youthful,2)
(yucky,-2)
(yummy,3)
(zealot,-2)
(zealots,-2)
(zealous,2)
grunt>
```

Step-11: We are joining the sentiments values from the AFINN dictionary with the words in the text from each tweet.



```
training@localhost:~/Downloads
File Edit View Search Terminal Help

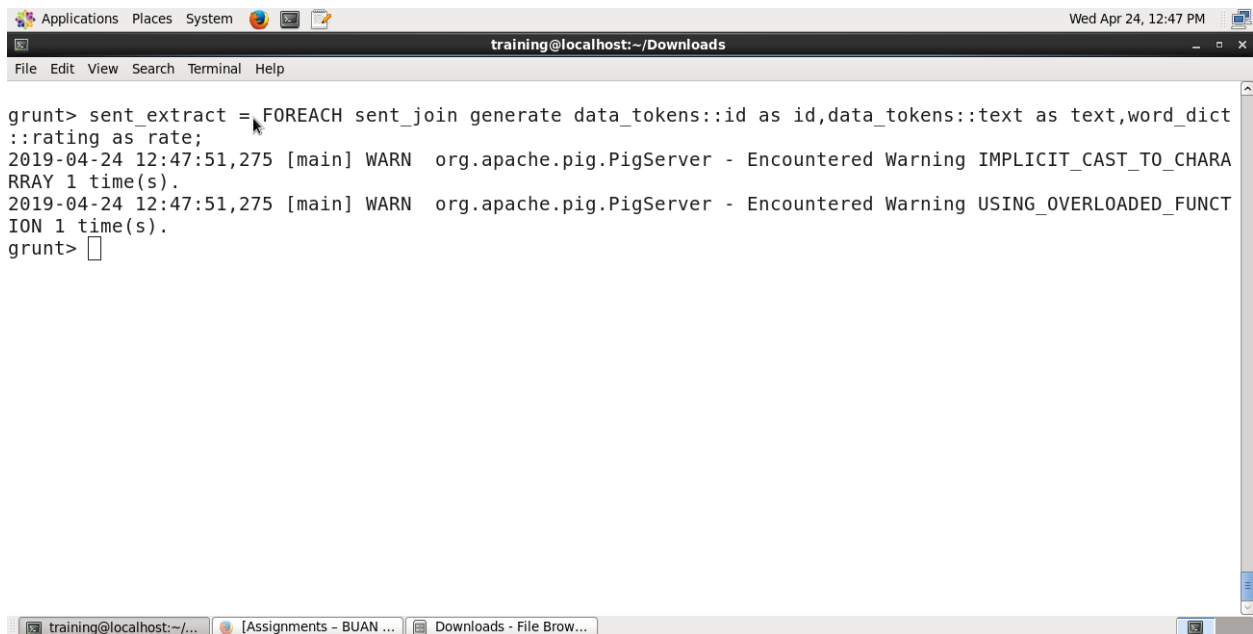
grunt> sent_join = join data_tokens by words left outer,word_dict by word using 'replicated';
2019-04-24 12:45:23,019 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARA
RRAY 1 time(s).
2019-04-24 12:45:23,019 [main] WARN org.apache.pig.PigServer - Encountered Warning USING_OVERLOADED_FUNCT
ION 1 time(s).
grunt>
```



```
training@localhost:~/Downloads
File Edit View Search Terminal Help

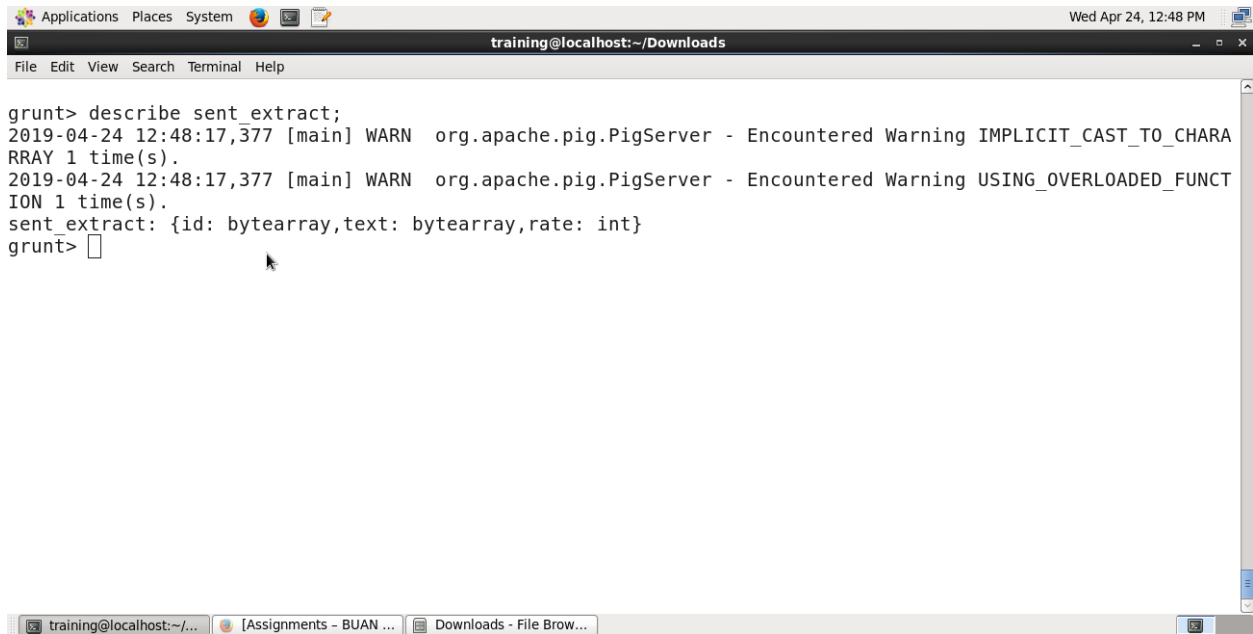
grunt> describe sent_join;
2019-04-24 12:45:50,937 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
2019-04-24 12:45:50,937 [main] WARN org.apache.pig.PigServer - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
sent_join: {data_tokens::id: bytearray,data_tokens::text: bytearray,data_tokens::words: chararray,word_dict::word: chararray,word_dict::rating: int}
grunt>
```

Step-12: : We are now extracting the data from the variable in step 11 that we need to complete our sentiment analysis.



```
training@localhost:~/Downloads
File Edit View Search Terminal Help

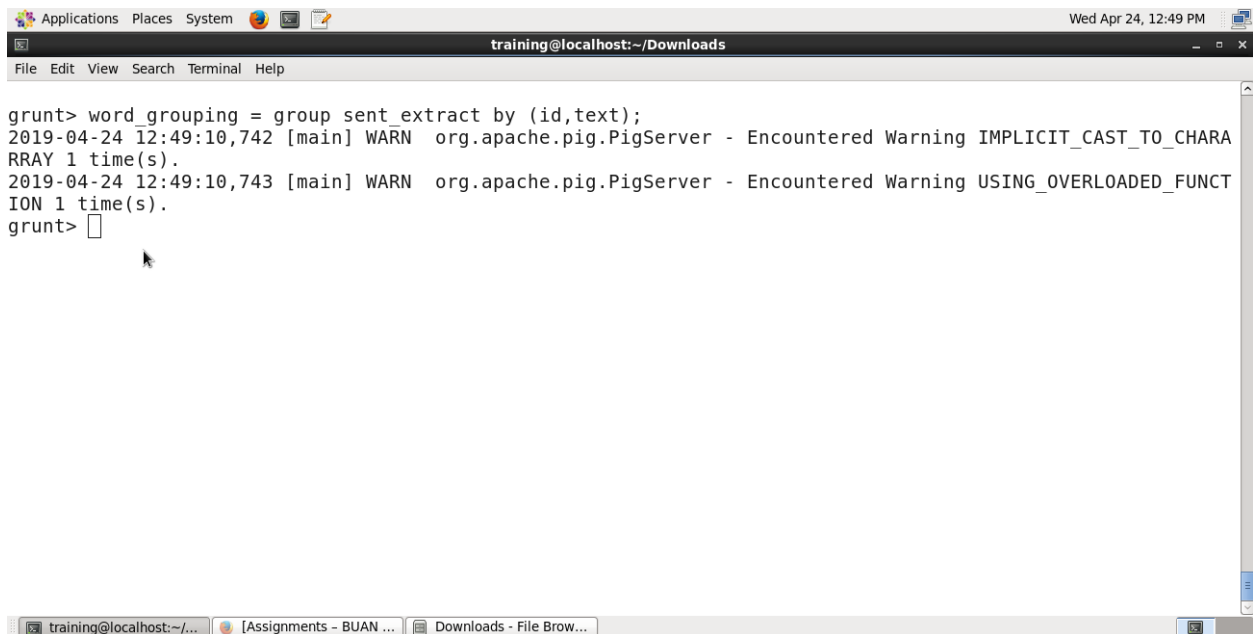
grunt> sent_extract = FOREACH sent_join generate data_tokens::id as id,data_tokens::text as text,word_dict::rating as rate;
2019-04-24 12:47:51,275 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
2019-04-24 12:47:51,275 [main] WARN org.apache.pig.PigServer - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
grunt>
```



```
training@localhost:~/Downloads
File Edit View Search Terminal Help

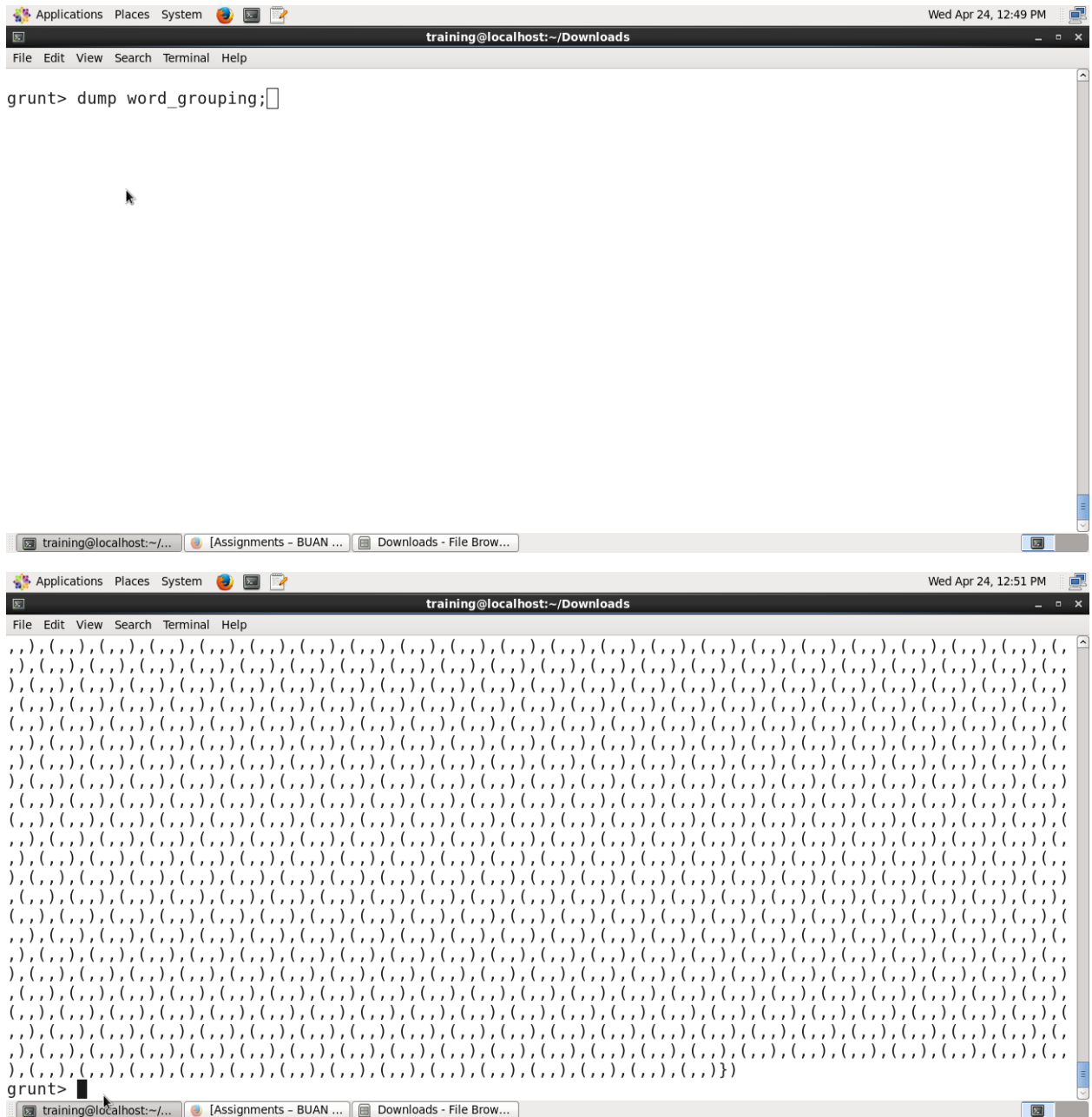
grunt> describe sent_extract;
2019-04-24 12:48:17,377 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
2019-04-24 12:48:17,377 [main] WARN org.apache.pig.PigServer - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
sent_extract: {id: bytearray,text: bytearray,rate: int}
grunt>
```

Step-13: We are grouping each record by its id and text.

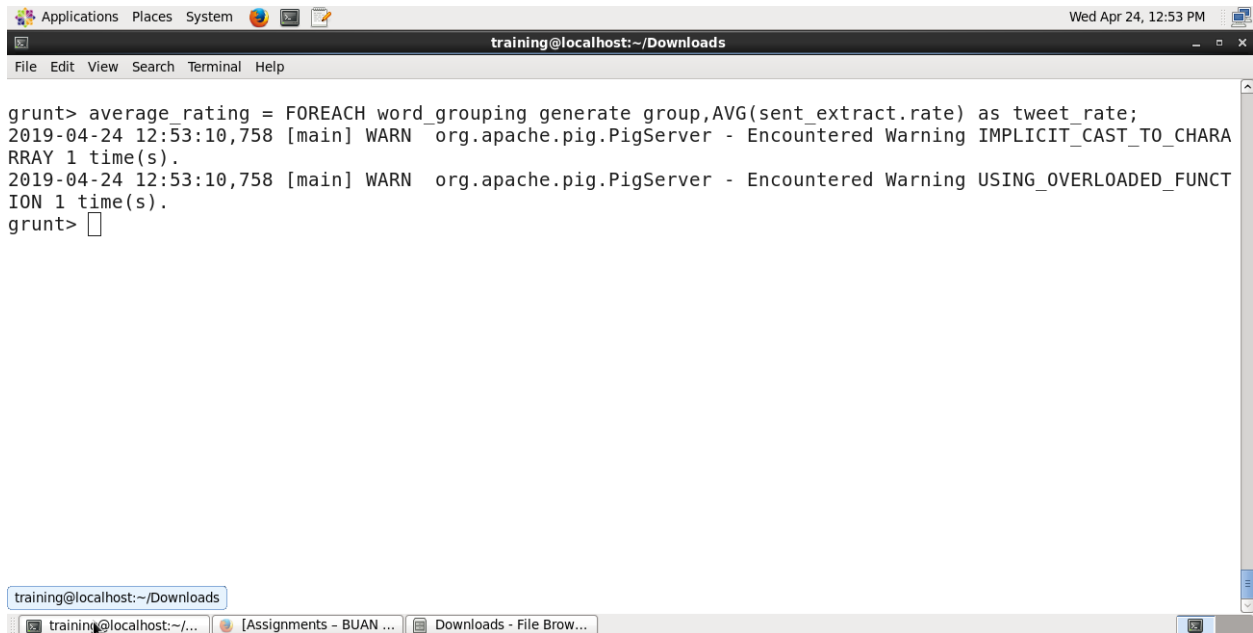


```
training@localhost:~/Downloads
File Edit View Search Terminal Help

grunt> word_grouping = group sent_extract by (id,text);
2019-04-24 12:49:10,742 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
2019-04-24 12:49:10,743 [main] WARN org.apache.pig.PigServer - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
grunt>
```



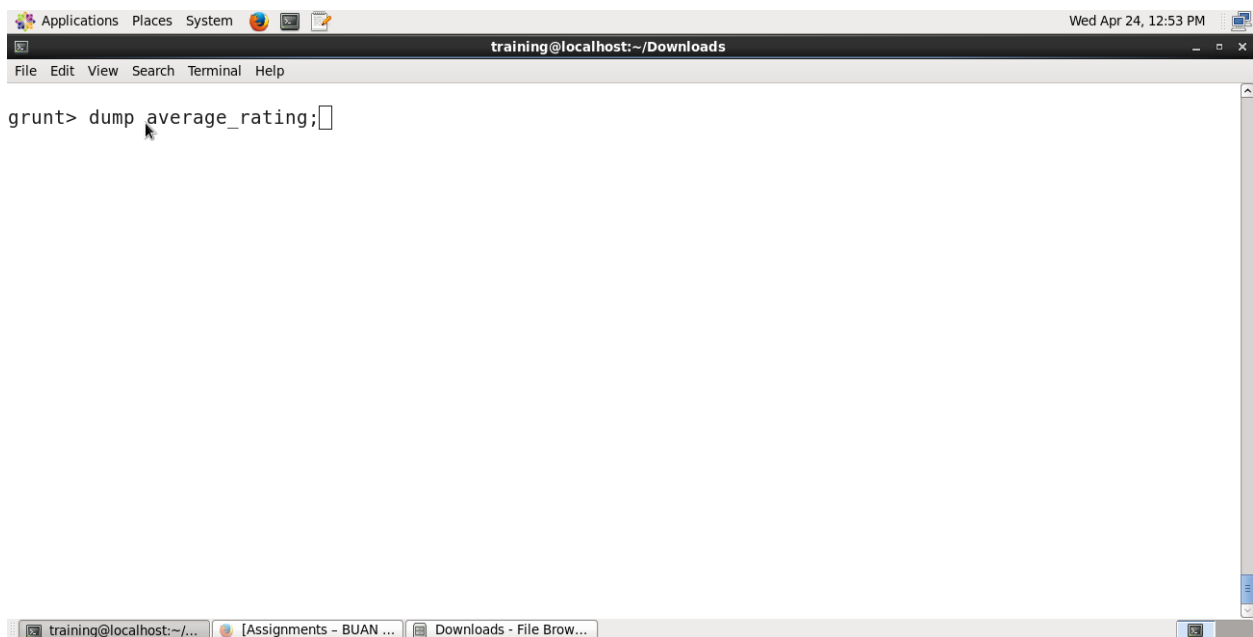
Step-14: We are averaging the rating for all the words in each tweet and storing the result in `average_rating`.



A screenshot of a terminal window titled "training@localhost:~/Downloads". The window shows the execution of a Grunt command to calculate the average rating for each word group. The command is: `grunt> average_rating = FOREACH word_grouping generate group,AVG(sent_extract.rate) as tweet_rate;`. The output shows two warnings from the `org.apache.pig.PigServer` class, both with a severity of "WARN". The first warning is about an "IMPLICIT_CAST_TO_CHARARRAY" and the second is about a "USING_OVERLOADED_FUNCTION". Both warnings occurred at 12:53:10,758 on 2019-04-24. The terminal window has a menu bar with "File", "Edit", "View", "Search", "Terminal", and "Help". The status bar at the bottom shows the current directory as `training@localhost:~/Downloads` and the file browser as `Downloads - File Brow...`.

```
training@localhost:~/Downloads
File Edit View Search Terminal Help

grunt> average_rating = FOREACH word_grouping generate group,AVG(sent_extract.rate) as tweet_rate;
2019-04-24 12:53:10,758 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHAR
ARRAY 1 time(s).
2019-04-24 12:53:10,758 [main] WARN org.apache.pig.PigServer - Encountered Warning USING_OVERLOADED_FUNCT
ION 1 time(s).
grunt>
```



A screenshot of a terminal window titled "training@localhost:~/Downloads". The window shows the execution of a Grunt command to dump the `average_rating` variable. The command is: `grunt> dump average_rating;`. The terminal window has a menu bar with "File", "Edit", "View", "Search", "Terminal", and "Help". The status bar at the bottom shows the current directory as `training@localhost:~/Downloads` and the file browser as `Downloads - File Brow...`.

```
training@localhost:~/Downloads
File Edit View Search Terminal Help

grunt> dump average_rating;
```

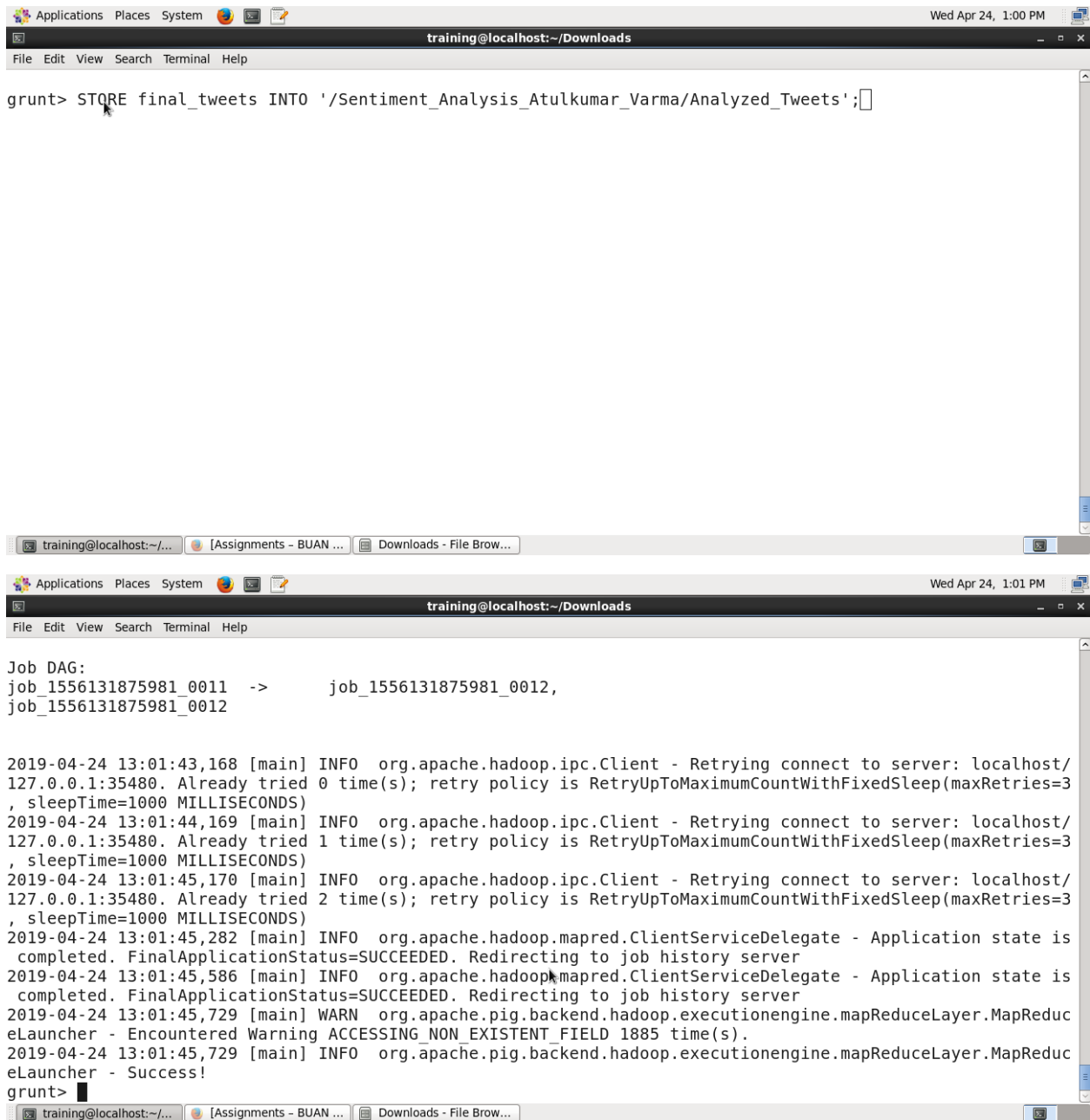


```
Applications Places System Wed Apr 24, 12:57 PM
training@localhost:~/Downloads
File Edit View Search Terminal Help

grunt> dump final_tweets;

by GOI on the issue. #DeMonetization #RBI"),1.0)
(("7997","RT @sukanyaiyer2: #DeMonetization AAP protests by marching Against Govts move over DeMonetization & he is also detained as he Tried 2 March"),-2.0)
(("7998","#demonetization will help combat terror because Pak won't be able to print new notes! And now",-0.6666666666666666)
(("8000","RT @UnSubtleDesi: Kejriwal posts pic of dead robber and claims it's #demonetization related death? How shameless has this man become? https"),-2.5)
(("How long, successful and sustainable will be this strategic game of #DeMonetization against Demons?"),3.0)
(("Only noise, chaos & disruptions by obstructionist #"),-2.0)
(("Only noise, chaos & disruptions by obstructi https://t.co/zVE7MYt04G"),-2.0)
(("5% bad idea, poor implementation"),-2.0)
(("25% good idea, poor implementation"),-2.0)
(("No there r many, we call them by many names like C#"),2.0)
(("Akhilesh=not good,black money is good),3.0)
(("If not for Aam Aadmi, listen to them no PM Modi?"),-1.0)
(("And respect their decision,but support oppositio"),2.0)
(("And respect their decision,but support opposition just b'coz of party"),2.0)
(("Aim of #demonetization laudable, but Govt has no road map2create... https://t.co/A4Geu9ch0v"),-1.0)
(("Enough jokes on #Demonetization, also no more posts on politics or social affairs..."),-1.0)
(("RT @kanimozhi: Everyone seems to hate the rich, even the rich hates richer and the richer hates the richest. #Demonetization"),-1.3333333333333333)
((" the avg indian wants corruptn free india.. So in d name of black money, everybody agrees),1.0)
grunt>
```

Step-16: We are viewing the output of the Analyzed_Tweets file in HUE



The screenshot shows a terminal window titled "training@localhost:~/Downloads" with a menu bar (File, Edit, View, Search, Terminal, Help) and a status bar (Wed Apr 24, 1:00 PM). The terminal displays the following command and output:

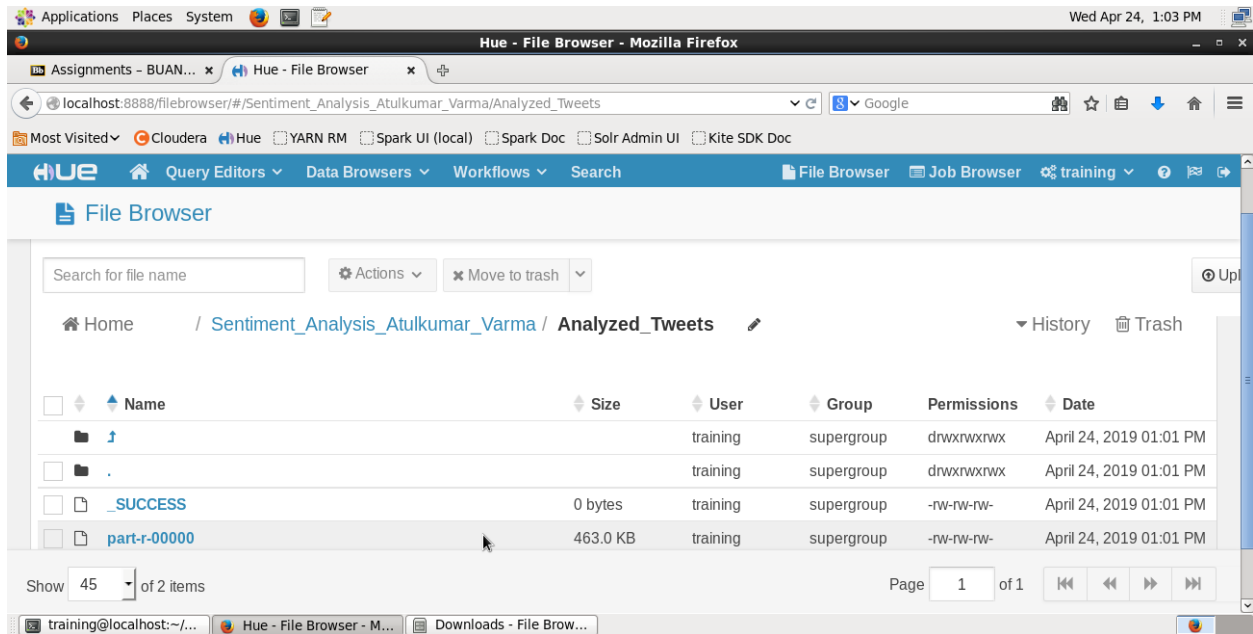
```
grunt> STORE final_tweets INTO '/Sentiment_Analysis_Atulkumar_Varma/Analyzed_Tweets';
```

The output shows the Job DAG and the execution of the STORE command:

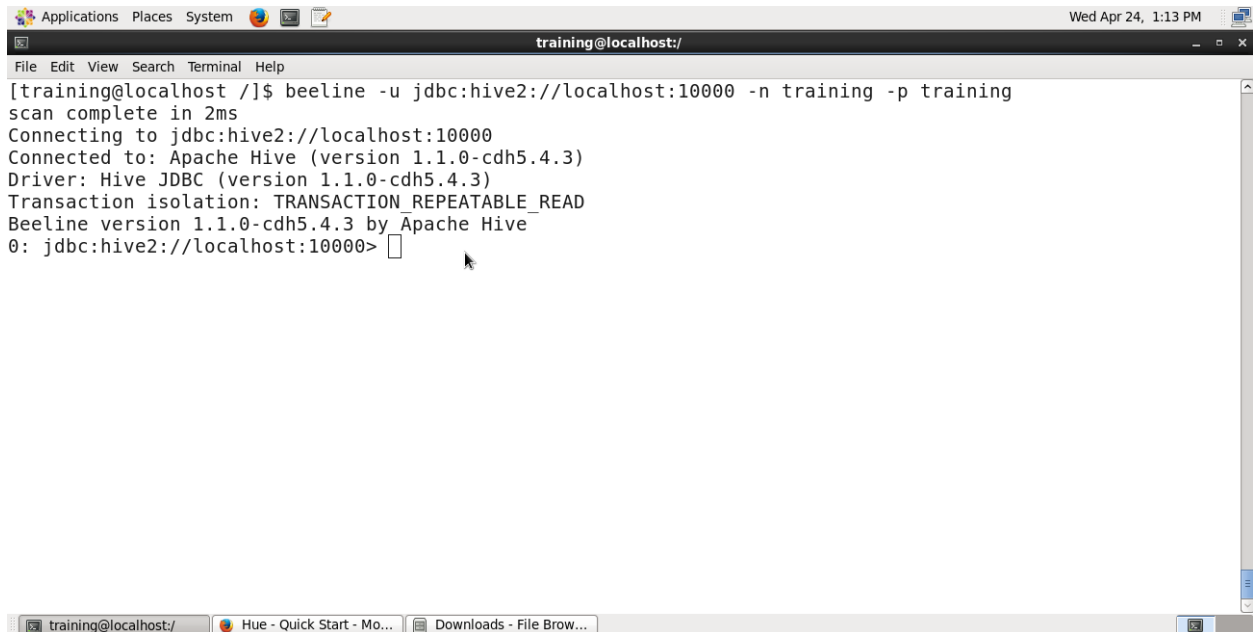
```
Job DAG:
job_1556131875981_0011  ->      job_1556131875981_0012,
job_1556131875981_0012

2019-04-24 13:01:43,168 [main] INFO  org.apache.hadoop.ipc.Client - Retrying connect to server: localhost/127.0.0.1:35480. Already tried 0 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)
2019-04-24 13:01:44,169 [main] INFO  org.apache.hadoop.ipc.Client - Retrying connect to server: localhost/127.0.0.1:35480. Already tried 1 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)
2019-04-24 13:01:45,170 [main] INFO  org.apache.hadoop.ipc.Client - Retrying connect to server: localhost/127.0.0.1:35480. Already tried 2 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)
2019-04-24 13:01:45,282 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2019-04-24 13:01:45,586 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2019-04-24 13:01:45,729 [main] WARN  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning ACCESSING_NON_EXISTENT_FIELD 1885 time(s).
2019-04-24 13:01:45,729 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt>
```

The terminal window also shows a taskbar with "training@localhost:~/...", "[Assignments - BUAN ...]", and "Downloads - File Brow...".



Step-17: We are starting up beeline, to execute the hive commands



Step-18: Creating a hive table to structure the unstructured data

The image shows a terminal window and a web browser window. The terminal window, titled 'training@localhost:/', displays the execution of a Hive query to create an external table named 'tweety'. The query is: `CREATE EXTERNAL TABLE tweety (text STRING, tweet_rating DOUBLE) ROW FORMAT DELIMITED FIELDS TERMINATED BY ' ' LOCATION '/Sentiment_Analysis_Atulkumar_Varma/Analyzed_Tweets';`. The output shows that the table was created successfully, with 0 rows affected in 4.57 seconds.

The web browser window, titled 'Hue - Hive Editor - Query - Mozilla Firefox', shows the 'Data sample for tweety' table. The table has two columns: 'text' and 'tweet_rating'. The data sample shows 7 rows of tweets and their ratings. The tweets are related to the demonetization move in India.

Data sample for tweety

text	tweet_rating
(So, if you really think that this #Demonetization move has struck at fake	
("1","RT @rssurjewala: Critical question: Was PayTM informed about #Demonetization edict by PM? It's clearly fishy and requires full disclosure &	
("10","National reform now destroyed even the essence of sagan. Such instances urge giving #demonetization a second though https://t.co/eyySIREUq"	
("12","RT @Joydas: Question in Narendra Modi App where PM is taking feedback if people support his #DeMonetization strategy https://t.co/pYgK8Rmg7r"	
("13","@Jaggesh2 Bharat band on 28??<ed><U+00A0><U+00BD><ed><U+00B8><U+0082>Those who are protesting #demonetization are all different pe	
("16","RT @Dipankar_cpiml: The Modi app on #DeMonetization proves once again that the govt is totally indifferent to the mounting misery and hards	
("19","RT @pGurus1: #Demonetization The co-operative banking sector in Kerala is as good as a tax haven. Is Kerala a Black Money HQ? https://t.co/	
("27","RT @kapil_kausik: #Doltiwal I mean #JaiChandKejriwal is ""hurt"" by #Demonetization as the same has rendered USELESS <ed><U+00A0><U+00BD>	

Step-19: Creating a tsv file from the table that we created in Step-18

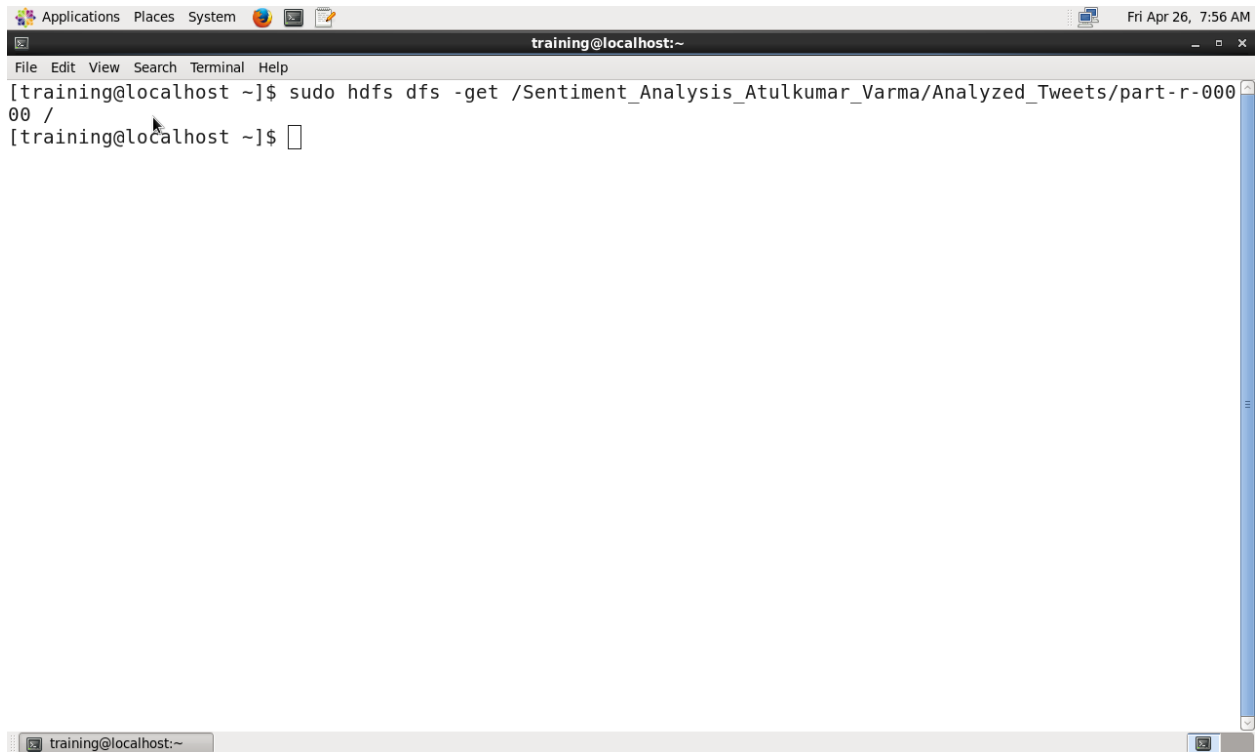
The image shows a terminal window and a web browser displaying the Hue interface. The terminal window, titled 'training@localhost:', shows the execution of Hive commands to create an external table and insert data. The web browser, titled 'Hue - File Browser - part-r-00000 - File Viewer - Mozilla Firefox', shows the file browser interface for the file 'part-r-00000' in the directory '/Sentiment_Analysis_Atulkumar_Varma/Analyzed_Tweets/'. The file content is displayed as a text file, showing a warning about binary data and the actual content of the file, which appears to be a tweet about demonetization.

```
0: jdbc:hive2://localhost:10000> CREATE EXTERNAL TABLE tweety (text STRING, tweet_rating DOUBLE) ROW FORM
T DELIMITED FIELDS TERMINATED BY ')' LOCATION '/Sentiment_Analysis_Atulkumar_Varma/Analyzed_Tweets';
No rows affected (4.57 seconds)
0: jdbc:hive2://localhost:10000> INSERT OVERWRITE LOCAL DIRECTORY '/Sentiment_Analysis_Atulkumar_Varma/Analyzed_
Tweets' ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' select * from tweety;
INFO : Number of reduce tasks is set to 0 since there's no reduce operator
WARN : Hadoop command-line option parsing not performed. Implement the Tool interface and execute your applicat
ion with ToolRunner to remedy this.
INFO : Starting Job = job_1556131875981_0013, Tracking URL = http://localhost:8088/proxy/application_1556131875
981_0013/
INFO : Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1556131875981_0013
INFO : Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
INFO : 2019-04-24 13:25:14,819 Stage-1 map = 0%, reduce = 0%
INFO : 2019-04-24 13:25:23,610 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.46 sec
INFO : MapReduce Total cumulative CPU time: 1 seconds 460 msec
INFO : Ended Job = job_1556131875981_0013
INFO : Copying data to local directory /Sentiment_Analysis_Atulkumar_Varma/Analyzed_Tweets from hdfs://localhos
t:8020/tmp/hive/training/227cd47d-fcc6-4839-8c8e-f4827721f1ed/hive_2019-04-24_13-25-01_144_733151001979179178-2/
-mr-10000
INFO : Copying data to local directory /Sentiment_Analysis_Atulkumar_Varma/Analyzed_Tweets from hdfs://localhos
t:8020/tmp/hive/training/227cd47d-fcc6-4839-8c8e-f4827721f1ed/hive_2019-04-24_13-25-01_144_733151001979179178-2/
-mr-10000
No rows affected (23.621 seconds)
0: jdbc:hive2://localhost:10000>
```

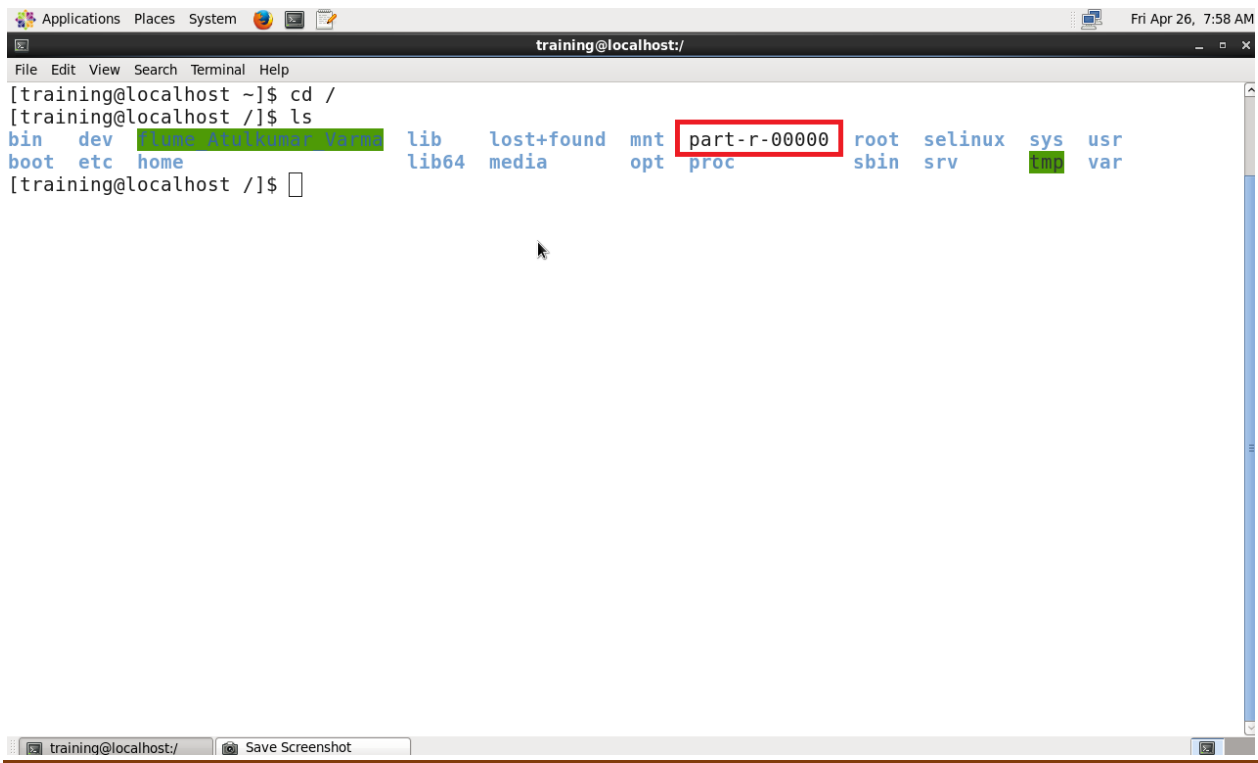
The web browser shows the Hue File Browser interface. The file 'part-r-00000' is selected, and its content is displayed. A warning message states: 'Warning: some binary data has been masked out with 'tfd;'. The content of the file is shown as a text file, displaying a tweet about demonetization.

```
(So, if you really think that this #Demonetization move has struck at fake) -1.0
("1","RT @rssurjewala: Critical question: Was PayTM informed about #Demonetization edict by PM? It's clearly f
ishy and requires full disclosure &#xftfd; 1.0
("10","National reform now destroyed even the essence of sagan. Such instances urge giving #demonetization a s
econd though https://t.co/eyySIREiUq") -3.0
```

Step-20: Copying the results of our sentiment analysis back on the local file system to visualize.

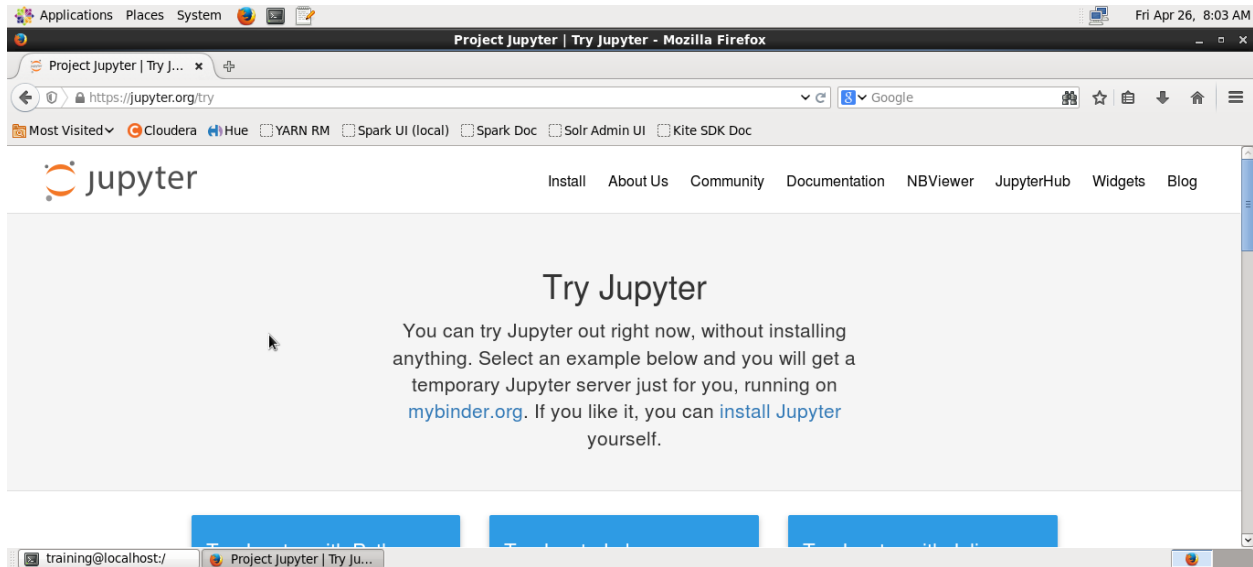


A terminal window titled "training@localhost:~" with a menu bar (File, Edit, View, Search, Terminal, Help) and a status bar (Fri Apr 26, 7:56 AM). The terminal shows the command `sudo hdfs dfs -get /Sentiment_Analysis_Atulkumar_Varma/Analyzed_Tweets/part-r-00000 /` being executed. The prompt is `[training@localhost ~]$`.

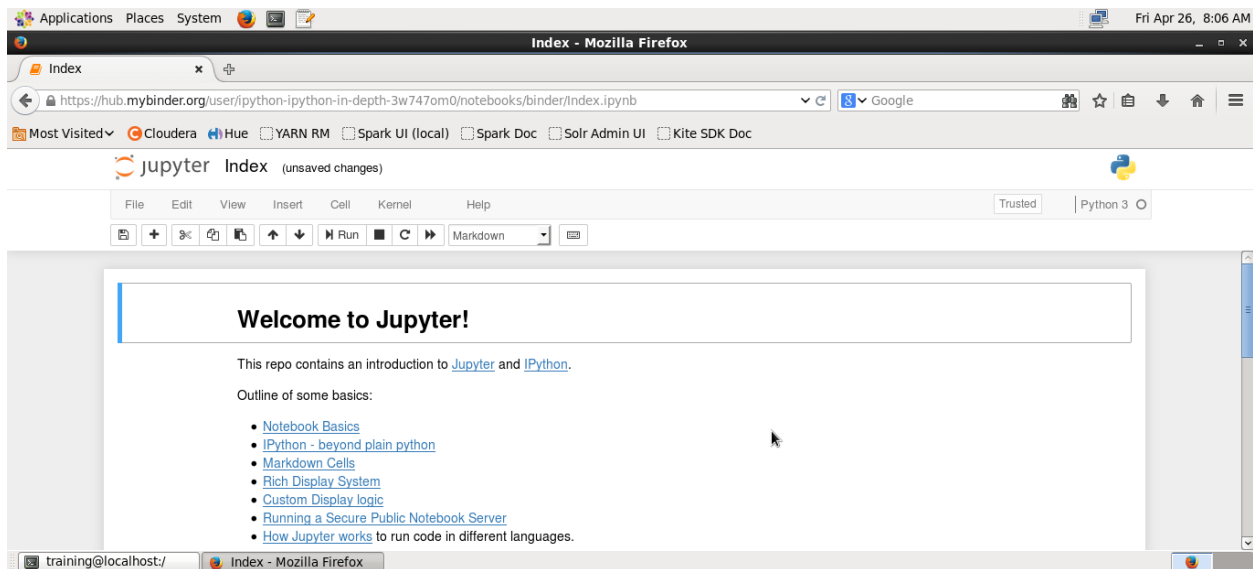


A terminal window titled "training@localhost:/" with a menu bar (File, Edit, View, Search, Terminal, Help) and a status bar (Fri Apr 26, 7:58 AM). The terminal shows the command `cd /` followed by `ls`. The output lists the root directory contents: `bin dev Sentiment_Analysis_Atulkumar_Varma lib lost+found mnt part-r-00000 root selinux sys usr boot etc home lib64 media opt proc sbin srv tmp var`. The file `part-r-00000` is highlighted with a red box. The prompt is `[training@localhost /]$`. A "Save Screenshot" button is visible in the bottom status bar.

Step-21:



Step-22:



Step-23:

The screenshot shows a Mozilla Firefox browser window with the address bar displaying `https://hub.mybinder.org/user/python-ipynb-in-depth-3w747om0/tree`. The page title is "Home - Mozilla Firefox". Below the browser window, the JupyterLab interface is visible. The "Files" tab is selected, showing a file browser view. The file list includes:

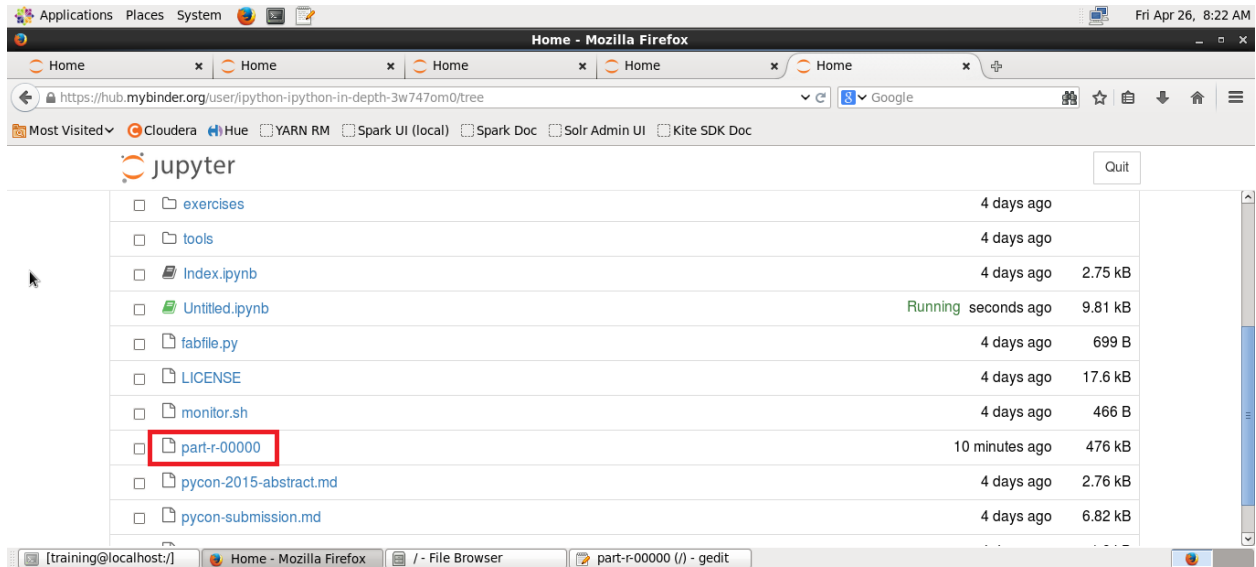
Name	Last Modified	File size
binder	seconds ago	
examples	4 days ago	
exercises	4 days ago	
tools	4 days ago	
Index.ipynb	4 days ago	2.75 kB
table.py	4 days ago	699 B
LICENSE	4 days ago	17.6 kB
monitor.sh	4 days ago	466 B
ovcon-2015-abstract.md	4 days ago	2.76 kB

The bottom status bar shows the terminal output: `training@localhost/`.

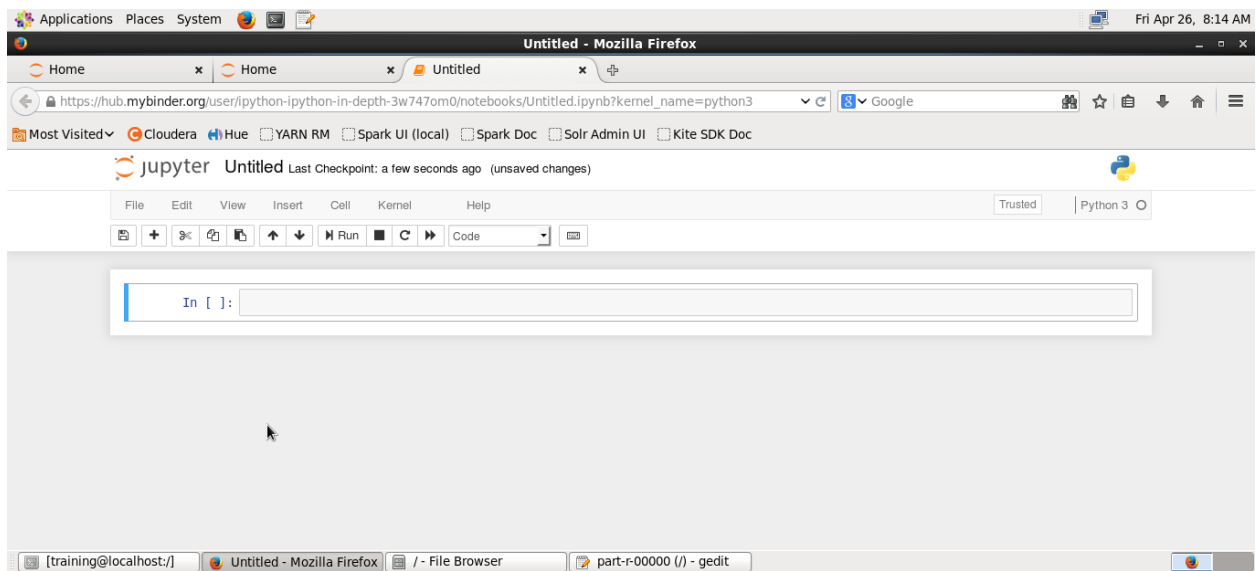
Step-24:

The screenshot shows a Mozilla Firefox browser window with the address bar displaying `https://hub.mybinder.org/user/python-ipynb-in-depth-3w747om0/edit/untitled.txt`. The page title is "untitled.txt - Mozilla Firefox". Below the browser window, the JupyterLab interface is visible. The "untitled.txt" file is open in the code editor, showing a single line of text: `1`. The status bar at the bottom shows the terminal output: `training@localhost/`.

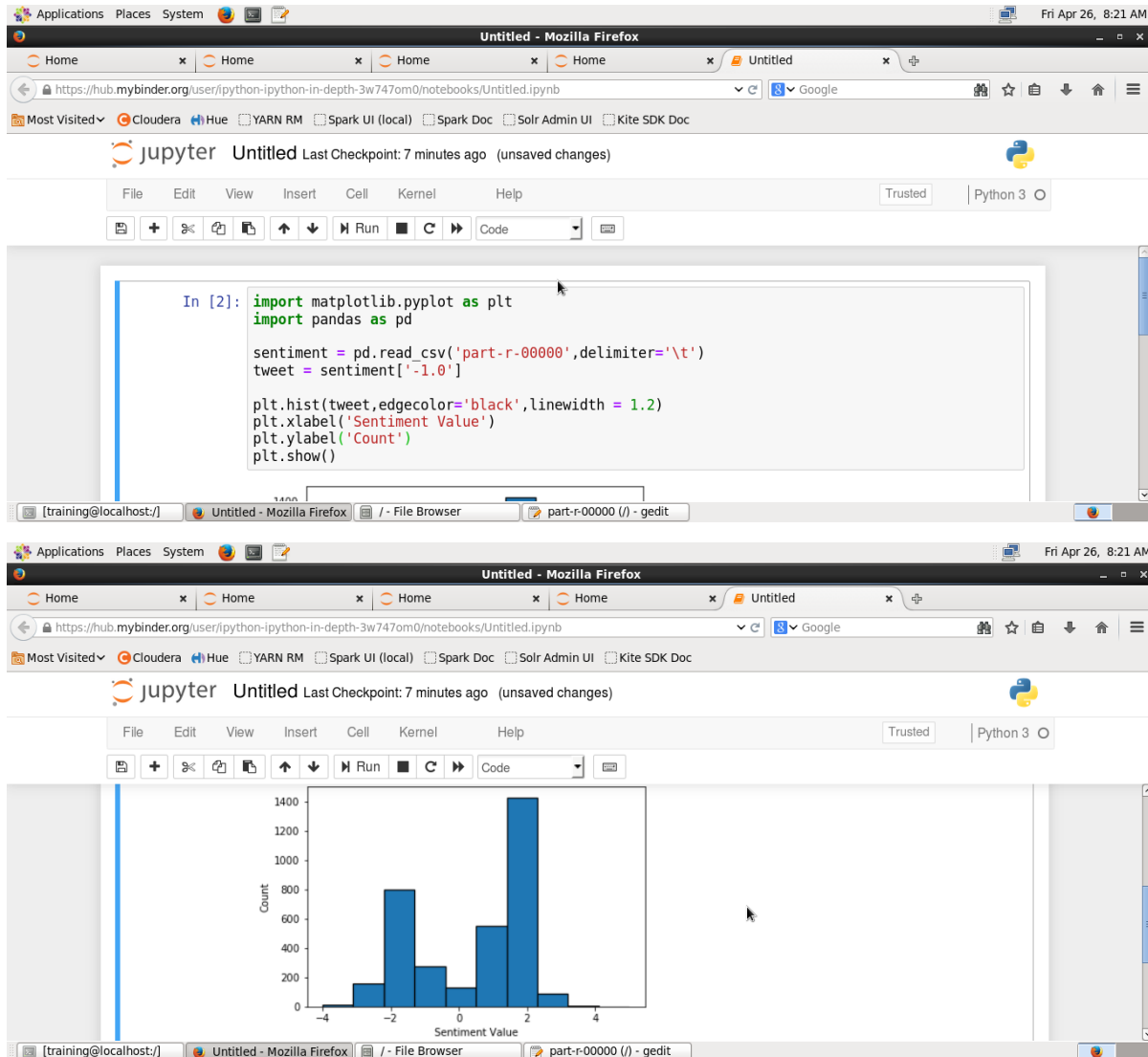
Step-25:



Step-26:



Step-27:



Y axis represents the count of sentiments

X axis whether the sentiment tweeted by people is negative or positive

Based on the AFINN Dictionary, few people(around 20-30) tweeted -4th level of negative sentiment

Around 800 tweeted -2nd level of negative sentiment

Around 100 tweeted neutral sentiment

Around 1400 tweeted 2nd level of positive sentiment

And finally, around 20-30 tweeted 4th level of positive sentiment