**BDSP Screenshots**

1. Import sentiment\_analysis.csv file into ‘Test’ database and ‘sentiment’ collections for MongoDB

A screenshot of a computer

Description automatically generated

1. Checking the data loaded correctly in the ‘test’ database within MongoDB shell

A screen shot of a computer screen

Description automatically generated

1. Moving sentiment\_analysis.csv file into Lab04 directory for apache spark processing

A computer screen shot of a computer program

Description automatically generated

1. Starting up dfs.sh and yarn.sh for cluster mode

A computer screen with white text

Description automatically generated

1. Ensuring appropriate ipynb file and csv file are in HDFS for processing activity

A computer screen shot of a program

Description automatically generated

1. Launching pyspark

A screenshot of a computer program

Description automatically generated

1. Confirming yarn mode is running from within pyspark

A screenshot of a computer

Description automatically generated

A screenshot of a computer

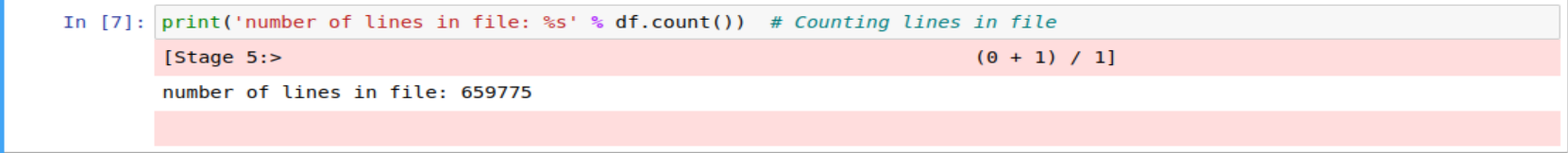
Description automatically generated

1. Reading in and displaying csv file, shortened file to stay within VM capacity

A screenshot of a computer

Description automatically generated

1. Counting number of lines in file



1. Processing text data and then displaying word counts in descending order

A screen shot of a computer code

Description automatically generated

Initial dataframe

A screenshot of a computer error

Description automatically generated

Tokenized words

A screen shot of a computer

Description automatically generated

Most frequent words and their respective counts

A screenshot of a computer

Description automatically generated

1. Export data out to dfs

A screen shot of a computer

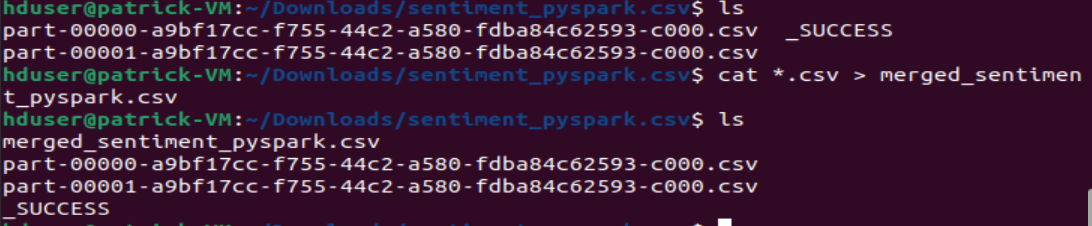
Description automatically generated

1. Copying sentiment\_pyspark.csv to local file

A computer screen with white text

Description automatically generated

1. Merged the split dataframe into one file ‘merged\_sentiment\_pyspark.csv’



1. Imported merged\_sentiment\_pyspark.csv file to MongoDB

A screenshot of a computer screen

Description automatically generated

1. Successfully imported, initial query to confirm

A screen shot of a computer

Description automatically generated