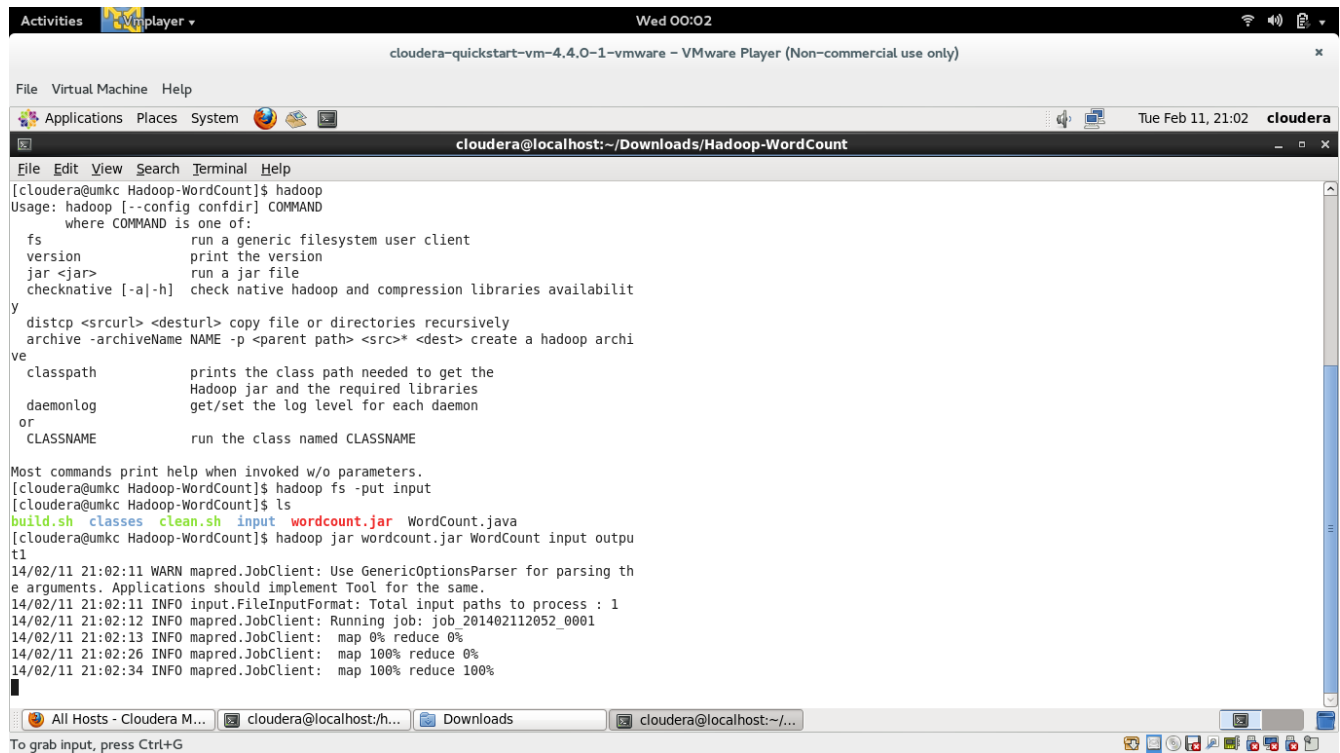


Lab 3

Cloudera/MapReduce:

Executing existing WordCount mapreduce problem



```
Activities  VMware Player  Wed 00:02
cloudera-quickstart-vm-4.4.0-1-vmware - VMware Player (Non-commercial use only)

File Virtual Machine Help

Applications Places System  Tue Feb 11, 21:02 cloudera

cloudera@localhost:~/Downloads/Hadoop-WordCount

File Edit View Search Terminal Help

[cloudera@umkc Hadoop-WordCount]$ hadoop
Usage: hadoop [--config confdir] COMMAND
    where COMMAND is one of:
        fs                run a generic filesystem user client
        version           print the version
        jar <jar>         run a jar file
        checknative [-a|-h] check native hadoop and compression libraries availability
        distcp <srcurl> <desturl> copy file or directories recursively
        archive -archiveName NAME -p <parent path> <src>* <dest> create a hadoop archive
        classpath         prints the class path needed to get the Hadoop jar and the required libraries
        daemonlog         get/set the log level for each daemon
        or
        CLASSNAME         run the class named CLASSNAME

Most commands print help when invoked w/o parameters.
[cloudera@umkc Hadoop-WordCount]$ hadoop fs -put input
[cloudera@umkc Hadoop-WordCount]$ ls
build.sh  classes  clean.sh  input  wordcount.jar  WordCount.java
[cloudera@umkc Hadoop-WordCount]$ hadoop jar wordcount.jar WordCount input output
14/02/11 21:02:11 WARN mapred.JobClient: Use GenericOptionsParser for parsing the arguments. Applications should implement Tool for the same.
14/02/11 21:02:11 INFO input.FileInputFormat: Total input paths to process : 1
14/02/11 21:02:12 INFO mapred.JobClient: Running job: job_201402112052_0001
14/02/11 21:02:13 INFO mapred.JobClient: map 0% reduce 0%
14/02/11 21:02:26 INFO mapred.JobClient: map 100% reduce 0%
14/02/11 21:02:34 INFO mapred.JobClient: map 100% reduce 100%

All Hosts - Cloudera M... cloudera@localhost:/h... Downloads cloudera@localhost:~/...

To grab input, press Ctrl+G
```

Activities Wed 00:04

cloudera-quickstart-vm-4.4.0-1-vmware - VMware Player (Non-commercial use only)

File Virtual Machine Help

Applications Places System Tue Feb 11, 21:04 cloudera

cloudera@localhost:~/Downloads/Hadoop-WordCount

File Edit View Search Terminal Help

```
you, 21
you," 12
you---" 1
you---" 1
you--I 1
you--the 1
you--wait 1
you. 14
you." 15
you; 1
you? 2
you?" 23
young 42
young, 7
young. 2
young." 1
younger 1
youngest 3
your 74
yours," 1
yours?" 1
yourself 5
yourself, 1
yourself. 1
yourself." 1
yourself; 1
youth 7
youth, 1
youth. 1
youthful 1
[cloudera@umkc Hadoop-WordCount]$
```

All Hosts - Cloudera M... cloudera@localhost:h... Downloads cloudera@localhost:~/...

To grab input, press Ctrl+G

Installing mahout and maven

Activities Wed 00:26

cloudera-quickstart-vm-4.4.0-1-vmware - VMware Player (Non-commercial use only)

File Virtual Machine Help

Applications Places System Tue Feb 11, 21:26 cloudera

Index of /maven2/org/apache/mahout/mahout-distrib

File Edit View History Bookmarks Tools Help

All Hosts - Cloudera Manager Blackboard Learn Index of /maven2/org

repo2.maven.org/maven2/org/apache/mahout/mahout-distribution/0.6/

Most Visited ☐ Cloudera ☐ Cloudera Manager ☐ Hue ☐ HDFS NameNode ☐ Hadoop Jo

mahout-distribution-0.6-src.zip.md5	01-Feb-2012 13:46	32
mahout-distribution-0.6-src.zip.sha1	01-Feb-2012 13:46	40
mahout-distribution-0.6.pom	01-Feb-2012 13:38	3948
mahout-distribution-0.6.pom.asc	01-Feb-2012 13:46	881
mahout-distribution-0.6.pom.asc.md5	01-Feb-2012 13:46	32
mahout-distribution-0.6.pom.asc.sha1	01-Feb-2012 13:46	40
mahout-distribution-0.6.pom.md5	01-Feb-2012 13:38	32
mahout-distribution-0.6.pom.sha1	01-Feb-2012 13:38	40
mahout-distribution-0.6.tar.bz2	01-Feb-2012 13:41	58676099
mahout-distribution-0.6.tar.bz2.asc	01-Feb-2012 13:46	881
mahout-distribution-0.6.tar.bz2.asc.md5	01-Feb-2012 13:46	32
mahout-distribution-0.6.tar.bz2.asc.sha1	01-Feb-2012 13:46	40
mahout-distribution-0.6.tar.bz2.md5	01-Feb-2012 13:43	32
mahout-distribution-0.6.tar.bz2.sha1	01-Feb-2012 13:43	40
mahout-distribution-0.6.tar.gz	01-Feb-2012 13:38	60072600
mahout-distribution-0.6.tar.gz.asc	01-Feb-2012 13:46	881
mahout-distribution-0.6.tar.gz.asc.md5	01-Feb-2012 13:46	32
mahout-distribution-0.6.tar.gz.asc.sha1	01-Feb-2012 13:46	40
mahout-distribution-0.6.tar.gz.md5	01-Feb-2012 13:41	32
mahout-distribution-0.6.tar.gz.sha1	01-Feb-2012 13:41	40
mahout-distribution-0.6.zip	01-Feb-2012 13:43	65070887
mahout-distribution-0.6.zip.asc	01-Feb-2012 13:46	881
mahout-distribution-0.6.zip.asc.md5	01-Feb-2012 13:46	32
mahout-distribution-0.6.zip.asc.sha1	01-Feb-2012 13:46	40
mahout-distribution-0.6.zip.md5	01-Feb-2012 13:45	32
mahout-distribution-0.6.zip.sha1	01-Feb-2012 13:45	40

Find: < Previous > Next Highlight all ☐ Match case

Index of /maven2/... cloudera@localho... Downloads [cloudera@localho... Downloads PowerPoint Presen... PowerPoint Presen...

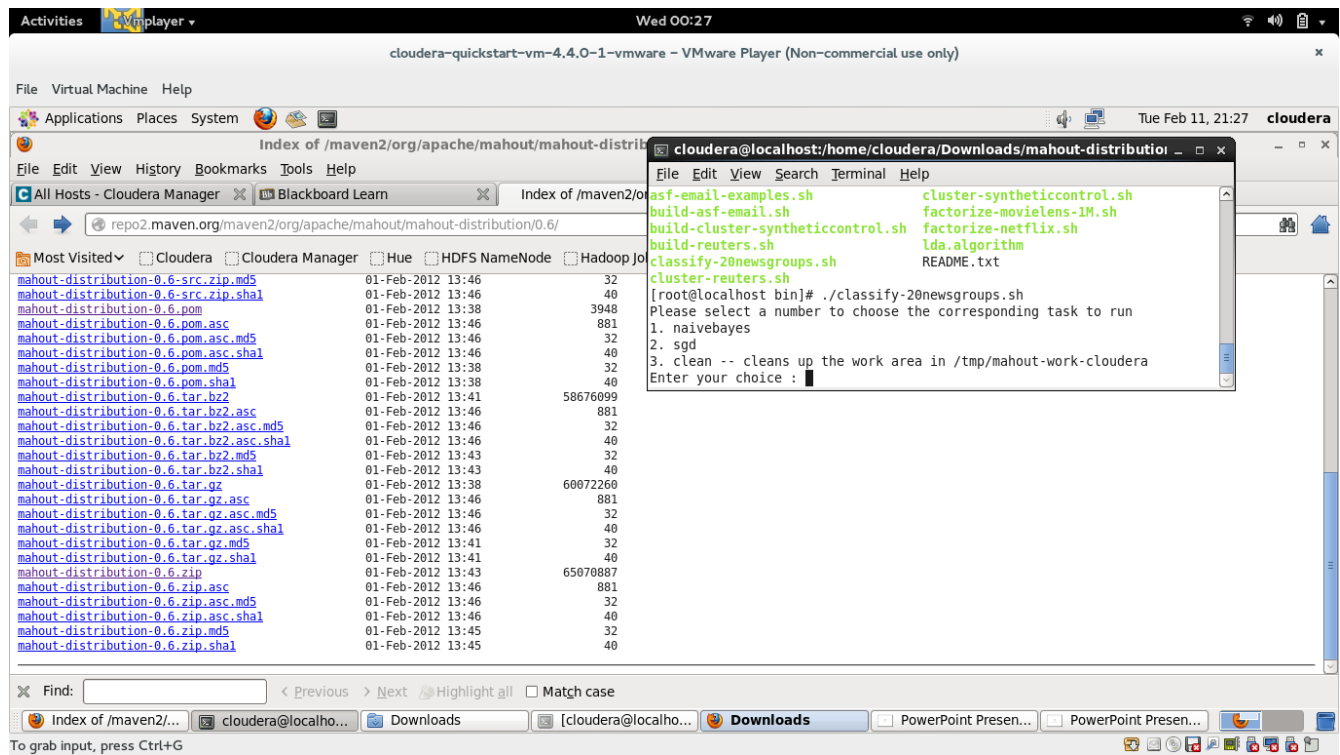
To grab input, press Ctrl+G

cloudera@localhost:~/Downloads/mahout-distribution-0.6

File Edit View Search Terminal Help

```
[INFO] -----
[INFO] BUILD SUCCESS
[INFO] -----
[INFO] Total time: 18.510s
[INFO] Finished at: Tue Feb 11 21:20:35 PST 2014
[INFO] Final Memory: 6M/57M
[INFO] -----
[root@localhost mahout-distribution-0.6]#
```

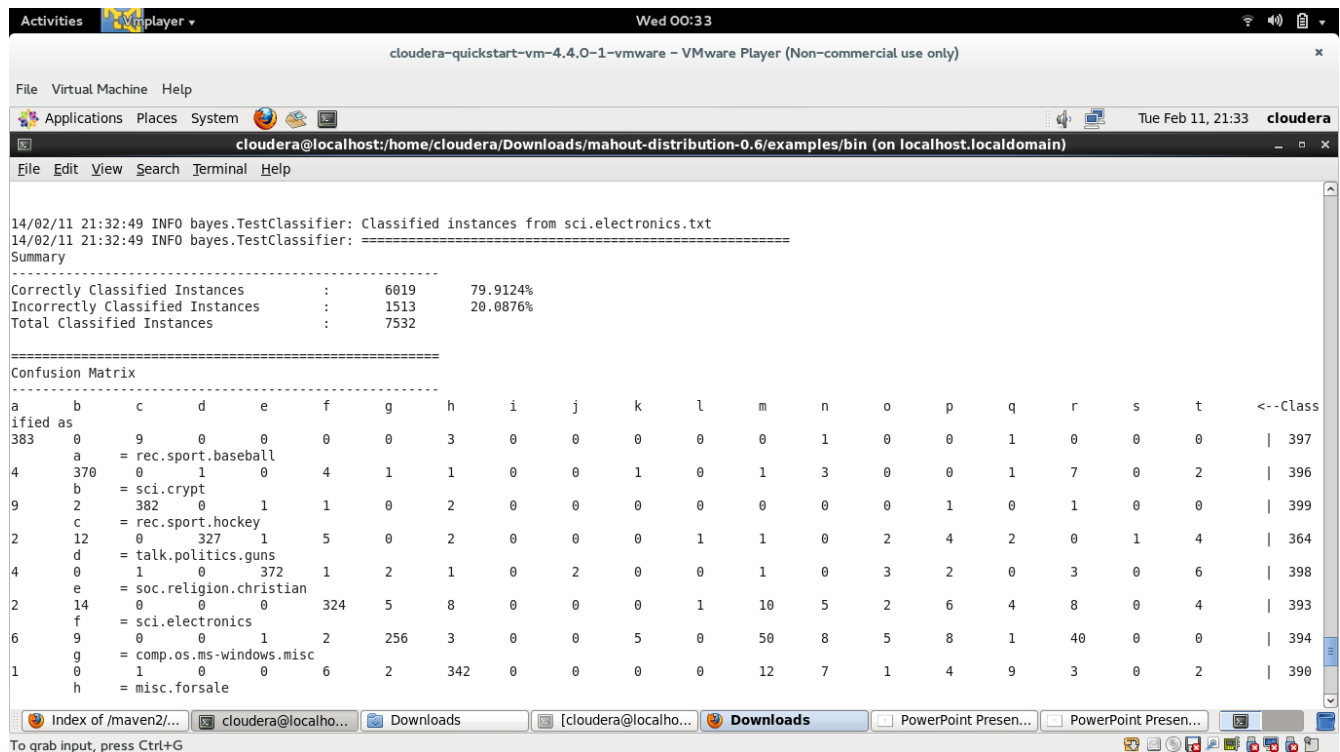
Running Mahout Example



The screenshot shows a VMware Player window titled "cloudera-quickstart-vm-4.4.0-1-vmware - VMware Player (Non-commercial use only)". The main window displays the "Index of /maven2/org/apache/mahout/mahout-distribution-0.6/" page. A terminal window is open in the foreground, showing a script menu with the following options:

```
asf-email-examples.sh      cluster-syntheticcontrol.sh
build-asf-email.sh          factorize-movielens-1M.sh
build-cluster-syntheticcontrol.sh  factorize-netflix.sh
build-reuters.sh            lda.algorithm
classify-20newsgroups.sh    README.txt
cluster-reuters.sh

[root@localhost bin]# ./classify-20newsgroups.sh
Please select a number to choose the corresponding task to run
1. naivebayes
2. sgd
3. clean -- cleans up the work area in /tmp/mahout-work-cloudera
Enter your choice :
```



The screenshot shows the same VMware Player window, but the terminal window now displays the output of the Mahout script. It shows the classification results for the "sci.electronics.txt" dataset, including a summary of correctly and incorrectly classified instances, and a confusion matrix.

14/02/11 21:32:49 INFO bayes.TestClassifier: Classified instances from sci.electronics.txt
14/02/11 21:32:49 INFO bayes.TestClassifier: =====
Summary

Correctly Classified Instances : 6019 79.9124%
Incorrectly Classified Instances : 1513 20.0876%
Total Classified Instances : 7532

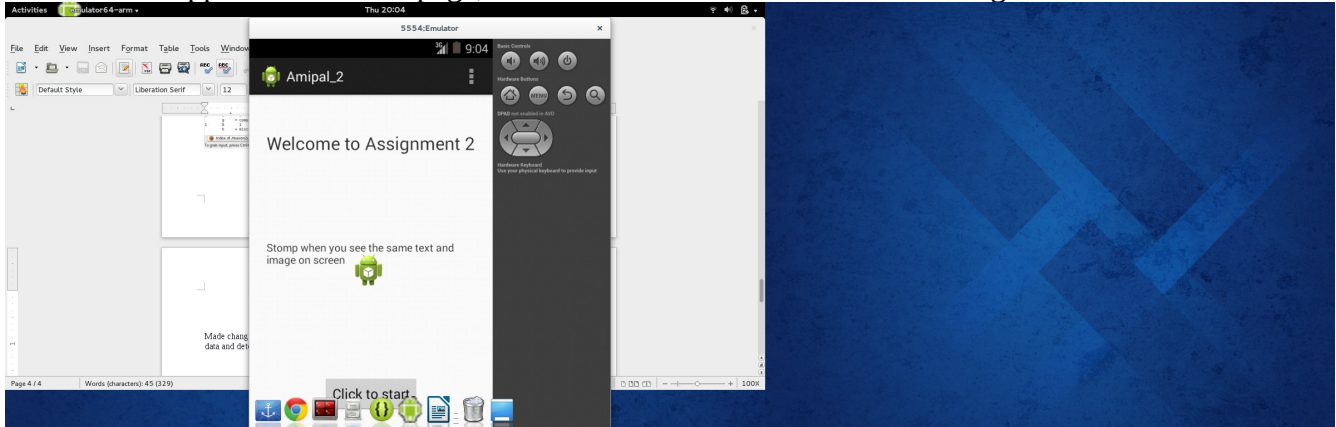
Confusion Matrix

a b c d e f g h i j k l m n o p q r s t <--Class
ified as
383 0 9 0 0 0 0 3 0 0 0 0 1 0 0 1 0 0 0
a = rec.sport.baseball | 397
4 370 0 1 0 4 1 1 0 0 1 0 1 3 0 0 1 7 0 2
b = sci.crypt | 396
9 2 382 0 1 1 0 2 0 0 0 0 0 0 1 0 1 0 0
c = rec.sport.hockey | 399
2 12 0 327 1 5 0 2 0 0 0 1 1 0 2 4 2 0 1 4
d = talk.politics.guns | 364
4 0 1 0 372 1 2 1 0 2 0 0 1 0 3 2 0 3 0 6
e = soc.religion.christian | 398
2 14 0 0 0 324 5 8 0 0 0 1 10 5 2 6 4 8 0 4
f = sci.electronics | 393
6 9 0 0 1 2 256 3 0 0 5 0 50 8 5 8 1 40 0 0
g = comp.os.ms-windows.misc | 394
1 0 1 0 0 6 2 342 0 0 0 0 12 7 1 4 9 3 0 2
h = misc.forsale | 390

Mashup Application

Made changes to existing android application which could collect real time data from the accelerometer data and detect the peak value of the acceleration for a particular motion.

Below is the application launch page, user selects the start button to start the game.



The image on screen “Apple” is a friendly image. If the user stomps his leg when this image is displayed he will loose points and negative score will be updated. A hurt sound will be played as well.

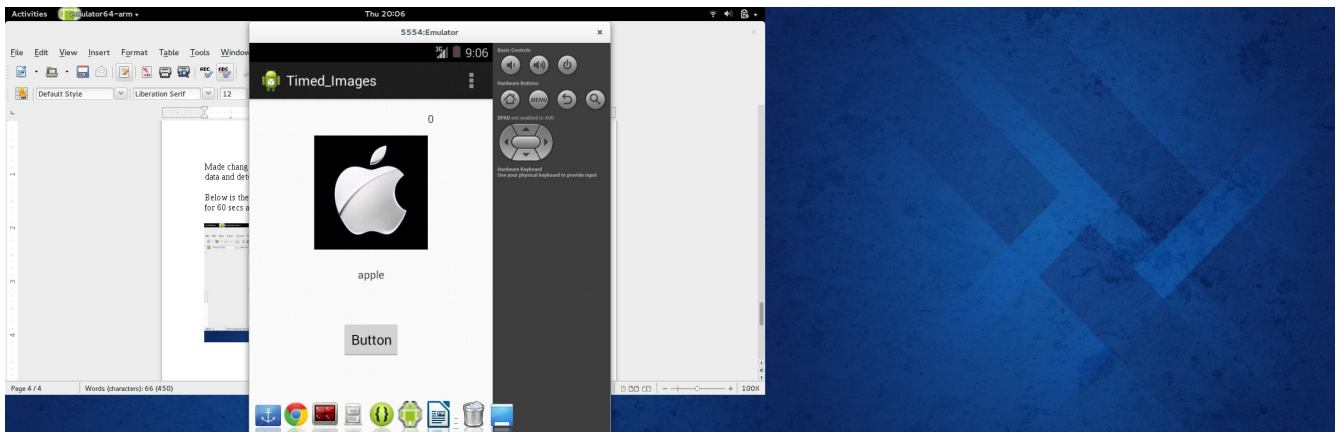


Image waits for 2 secs for user inputs , if non found score will be deducted from the user and next image is displayed.

Score value is the amount of acceleration the user applied on the device during the stomp.

Final score of the user is displayed after 60 secs of the game start.

