TAB-1

[cloudera@quickstart ~]$ cd Downloads/big-data-2

[cloudera@quickstart big-data-2]$ cd image

[cloudera@quickstart image]$ eog Australia.jpg&

[1] 10008

[cloudera@quickstart image]$ ./dimensions.py Australia.jpg

size = 5250 columns x 4320 rows

mode = RGB 3x8-bit pixels, true colour

[1]+ Done eog Australia.jpg

[cloudera@quickstart image]$ ./pixel.py Australia.jpg

[cloudera@quickstart image]$ cd ..

[cloudera@quickstart json]$ ./print\_json.py

Enter filename: twitter.json

Which Tweet Number are you interested in ? 99

Enter path (ex: user/id) : text

RT @IGN: #Beyond: Sony's game plan for the rest of 2015 http://t.co/AIzUexSjrM http://t.co/o7jVHjnVuw

[cloudera@quickstart json]$ ./

json\_schema.py LiveTweets.py PlotTweets.py print\_json.py

[cloudera@quickstart json]$ ./

json\_schema.py LiveTweets.py PlotTweets.py print\_json.py

[cloudera@quickstart json]$ ./

json\_schema.py LiveTweets.py PlotTweets.py print\_json.py

[cloudera@quickstart json]$ ./print\_json.py

Enter filename: twitter.json

Which Tweet Number are you interested in ? 34

Enter path (ex: user/id) : himanshu

Check your path : himanshu

Problem at : himanshu

[cloudera@quickstart json]$ ./print\_json.py

Enter filename: twitter.json

Which Tweet Number are you interested in ? 99

Enter path (ex: user/id) : entities/hastags

Check your path : entities/hastags

Problem at : hastags

[cloudera@quickstart json]$ cd..

bash: cd..: command not found

[cloudera@quickstart json]$ cd..

bash: cd..: command not found

[cloudera@quickstart json]$

[cloudera@quickstart json]$ cd ..

[cloudera@quickstart big-data-2]$ ls

csv graph image json sensor setup.sh vector

[cloudera@quickstart big-data-2]$ cd sensor/

[cloudera@quickstart sensor]$ ls -l

total 5992

-rwx------ 1 cloudera cloudera 888 Apr 19 2016 plot-data.py

-rwxr-xr-x 1 cloudera cloudera 246 Apr 19 2016 stream-data.py

-rwx------ 1 cloudera cloudera 1457 Jan 17 2017 stream-plot-data.py

-rw------- 1 cloudera cloudera 6119195 Apr 19 2016 wx-data.txt

-rw------- 1 cloudera cloudera 1018 Mar 14 2016 wxt-520-format.txt

[cloudera@quickstart sensor]$ ./plot-data.py wx

Traceback (most recent call last):

File "./plot-data.py", line 13, in <module>

file = open(sys.argv[1], 'r')

IOError: [Errno 2] No such file or directory: 'wx'

[cloudera@quickstart sensor]$ ./plot-data.py wx-data.txt Th

/usr/lib64/python2.6/site-packages/matplotlib/backends/backend\_gtk.py:621: DeprecationWarning: Use the new widget gtk.Tooltip

self.tooltips = gtk.Tooltips()

[cloudera@quickstart sensor]$ cd ..

[cloudera@quickstart big-data-2]$ cd data

bash: cd: data: No such file or directory

[cloudera@quickstart big-data-2]$ ls

csv graph image json sensor setup.sh vector

[cloudera@quickstart big-data-2]$ cd data/

bash: cd: data/: No such file or directory

[cloudera@quickstart big-data-2]$ ls

csv graph image json sensor setup.sh vector

[cloudera@quickstart big-data-2]$ ls -l

total 28

drwx------ 2 cloudera cloudera 4096 Jan 27 22:23 csv

drwx------ 2 cloudera cloudera 4096 Apr 27 2016 graph

drwx------ 2 cloudera cloudera 4096 Apr 15 2016 image

drwx------ 2 cloudera cloudera 4096 May 7 2016 json

drwxr-xr-x 2 cloudera cloudera 4096 Jan 17 2017 sensor

-rwx------ 1 cloudera cloudera 281 Apr 19 2016 setup.sh

drwx------ 3 cloudera cloudera 4096 Apr 25 2016 vector

[cloudera@quickstart big-data-2]$ cd vector

[cloudera@quickstart vector]$ ls

data LuceneQuery.class LuceneTFIDF.class runLuceneQuery.sh runLuceneTFIDF.sh

[cloudera@quickstart vector]$ cd data

[cloudera@quickstart data]$ more news1.csv

[cloudera@quickstart data]$ ls -l

total 84

-rw------- 1 cloudera cloudera 20595 Apr 25 2016 news1.csv

-rw------- 1 cloudera cloudera 25600 Apr 25 2016 news2.csv

-rw------- 1 cloudera cloudera 31529 Apr 25 2016 news3.csv

[cloudera@quickstart data]$ cd ..

[cloudera@quickstart vector]$ ls -l

total 28

drwx------ 2 cloudera cloudera 4096 Apr 25 2016 data

-rw------- 1 cloudera cloudera 7031 Apr 25 2016 LuceneQuery.class

-rw------- 1 cloudera cloudera 6863 Apr 25 2016 LuceneTFIDF.class

-rwx------ 1 cloudera cloudera 166 Apr 25 2016 runLuceneQuery.sh

-rwx------ 1 cloudera cloudera 165 Apr 25 2016 runLuceneTFIDF.sh

[cloudera@quickstart vector]$ ./runLuceneQuery.sh

Index Location:/index

ERROR: could not create directory: /index

[cloudera@quickstart vector]$ ./runLuceneQuery.sh data

Index Location:data/index

Skipping (not csv/htm/html/xml/txt) : write.lock

Indexed : data/news1.csv

Indexed : data/news3.csv

Indexed : data/news2.csv

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

3 new documents added.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Enter query for Lucene (q=quit):

delegates

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Displaying 2 results.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

1) data/news2.csv score :0.041339863

2) data/news1.csv score :0.01953125

Enter query for Lucene (q=quit):

voters

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Displaying 3 results.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

1) data/news1.csv score :0.043995064

2) data/news2.csv score :0.024887364

3) data/news3.csv score :0.011129968

Enter query for Lucene (q=quit):

voters delegates

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Displaying 3 results.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

1) data/news2.csv score :0.04811

2) data/news1.csv score :0.041432917

3) data/news3.csv score :0.0032286723

Enter query for Lucene (q=quit):

voters^5 delegates

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Displaying 3 results.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

1) data/news1.csv score :0.047636837

2) data/news2.csv score :0.035135828

3) data/news3.csv score :0.005357802

Enter query for Lucene (q=quit):

q

[cloudera@quickstart vector]$ ls -l

total 28

drwx------ 3 cloudera cloudera 4096 Jan 27 23:52 data

-rw------- 1 cloudera cloudera 7031 Apr 25 2016 LuceneQuery.class

-rw------- 1 cloudera cloudera 6863 Apr 25 2016 LuceneTFIDF.class

-rwx------ 1 cloudera cloudera 166 Apr 25 2016 runLuceneQuery.sh

-rwx------ 1 cloudera cloudera 165 Apr 25 2016 runLuceneTFIDF.sh

[cloudera@quickstart vector]$ ./runLuceneTFIDF.sh data

Index Location:data/index

Skipping (not csv,htm,html,xml,txt : write.lock

Indexed : data/news1.csv

Indexed : data/news3.csv

Indexed : data/news2.csv

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

3 new documents added.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Enter a term to calculate TF-IDF (q=quit):

voter delegates

Error looking for the term: voter delegates

Enter a term to calculate TF-IDF (q=quit):

voters

Doc # 0: data/news1.csv TF-IDF = 2.252547264099121

Doc # 1: data/news3.csv TF-IDF = 0.712317943572998

Doc # 2: data/news2.csv TF-IDF = 1.5927913188934326

Enter a term to calculate TF-IDF (q=quit):

TAB-2

[cloudera@quickstart ~]$ eog Australia.pg

[cloudera@quickstart ~]$ Eog Australia.jpg

bash: Eog: command not found

[cloudera@quickstart ~]$ eog Australia.jpg

[cloudera@quickstart ~]$ cd image

bash: cd: image: No such file or directory

[cloudera@quickstart ~]$ cd Downloads/big-data-2

[cloudera@quickstart big-data-2]$ cd image

[cloudera@quickstart image]$ eog Australia.jpg

cd image

Message from syslogd@quickstart at Jan 29 00:05:56 ...

kernel:BUG: soft lockup - CPU#0 stuck for 97s! [SelectionManage:9343]

[cloudera@quickstart image]$

[cloudera@quickstart image]$

[cloudera@quickstart image]$ cd image

bash: cd: image: No such file or directory

[cloudera@quickstart image]$

TAB-3

[cloudera@quickstart ~]$ cd Downloads/

[cloudera@quickstart Downloads]$ ls -l

total 3544

-rw-rw-r-- 1 cloudera cloudera 3625605 Jan 27 22:00 big-data-2.zip

[cloudera@quickstart Downloads]$ unzip -o big-data-2.zip

Archive: big-data-2.zip

creating: big-data-2/

inflating: big-data-2/setup.sh

creating: big-data-2/graph/

inflating: big-data-2/graph/diseaseGraph.csv

creating: big-data-2/csv/

inflating: big-data-2/csv/census.csv

creating: big-data-2/vector/

inflating: big-data-2/vector/runLuceneTFIDF.sh

inflating: big-data-2/vector/LuceneQuery.class

creating: big-data-2/vector/data/

inflating: big-data-2/vector/data/news3.csv

inflating: big-data-2/vector/data/news1.csv

inflating: big-data-2/vector/data/news2.csv

inflating: big-data-2/vector/runLuceneQuery.sh

inflating: big-data-2/vector/LuceneTFIDF.class

creating: big-data-2/image/

inflating: big-data-2/image/Australia.jpg

inflating: big-data-2/image/dimensions.py

inflating: big-data-2/image/pixel.py

creating: big-data-2/sensor/

inflating: big-data-2/sensor/plot-data.py

inflating: big-data-2/sensor/wxt-520-format.txt

inflating: big-data-2/sensor/stream-plot-data.py

inflating: big-data-2/sensor/wx-data.txt

inflating: big-data-2/sensor/stream-data.py

creating: big-data-2/json/

inflating: big-data-2/json/LiveTweets.py

inflating: big-data-2/json/auth

inflating: big-data-2/json/json\_schema.py

inflating: big-data-2/json/print\_json.py

inflating: big-data-2/json/twitter.json

inflating: big-data-2/json/PlotTweets.py

[cloudera@quickstart Downloads]$ cd big-data-2

[cloudera@quickstart big-data-2]$ pwd

/home/cloudera/Downloads/big-data-2

[cloudera@quickstart big-data-2]$ ls

csv graph image json sensor setup.sh vector

[cloudera@quickstart big-data-2]$ ls -l

total 28

drwx------ 2 cloudera cloudera 4096 Apr 11 2016 csv

drwx------ 2 cloudera cloudera 4096 Apr 27 2016 graph

drwx------ 2 cloudera cloudera 4096 Apr 15 2016 image

drwx------ 2 cloudera cloudera 4096 May 7 2016 json

drwxr-xr-x 2 cloudera cloudera 4096 Jan 17 2017 sensor

-rwx------ 1 cloudera cloudera 281 Apr 19 2016 setup.sh

drwx------ 3 cloudera cloudera 4096 Apr 25 2016 vector

[cloudera@quickstart big-data-2]$ cd csv

[cloudera@quickstart csv]$ ls

census.csv

[cloudera@quickstart csv]$ ls –l

[cloudera@quickstart csv]$ oocalc

bash: oocalc: command not found

[cloudera@quickstart csv]$

Message from syslogd@quickstart at Jan 29 00:05:56 ...

kernel:BUG: soft lockup - CPU#0 stuck for 97s! [SelectionManage:9343]

TAB-4(WORD MEDIAN)

[cloudera@quickstart ~]$ cd

[cloudera@quickstart ~]$ pwd

/home/cloudera

[cloudera@quickstart ~]$ cd^C

[cloudera@quickstart ~]$ cd /home/cloudera/desktop

bash: cd: /home/cloudera/desktop: No such file or directory

[cloudera@quickstart ~]$ pwd

/home/cloudera

[cloudera@quickstart ~]$ cd Desktop

[cloudera@quickstart Desktop]$ hadoop fs -put work.txt

[cloudera@quickstart Desktop]$ pwd

/home/cloudera/Desktop

[cloudera@quickstart Desktop]$ ls

Eclipse.desktop Kerberos.desktop untitled folder 2

Enterprise.desktop Parcels.desktop work.txt

Express.desktop untitled folder work.txt~

[cloudera@quickstart Desktop]$ hadoop fs -ls

Found 2 items

drwxr-xr-x - cloudera cloudera 0 2019-01-23 03:00 Desktop

-rw-r--r-- 1 cloudera cloudera 5458200 2019-01-27 21:19 work.txt

[cloudera@quickstart Desktop]$ hadoop jar /usr/lib/hadoop-0.20-mapreduce/hadoop-examples.jar wordmedian work.txt himanshu.txt

19/01/27 21:24:01 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032

19/01/27 21:24:06 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.

19/01/27 21:24:08 INFO input.FileInputFormat: Total input paths to process : 1

19/01/27 21:24:09 WARN hdfs.DFSClient: Caught exception

java.lang.InterruptedException

at java.lang.Object.wait(Native Method)

at java.lang.Thread.join(Thread.java:1281)

at java.lang.Thread.join(Thread.java:1355)

at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)

at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)

at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)

19/01/27 21:24:09 INFO mapreduce.JobSubmitter: number of splits:1

19/01/27 21:24:10 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1548651821670\_0001

19/01/27 21:24:15 INFO impl.YarnClientImpl: Submitted application application\_1548651821670\_0001

19/01/27 21:24:15 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application\_1548651821670\_0001/

19/01/27 21:24:15 INFO mapreduce.Job: Running job: job\_1548651821670\_0001

19/01/27 21:25:05 INFO mapreduce.Job: Job job\_1548651821670\_0001 running in uber mode : false

19/01/27 21:25:05 INFO mapreduce.Job: map 0% reduce 0%

19/01/27 21:25:41 INFO mapreduce.Job: map 100% reduce 0%

19/01/27 21:26:22 INFO mapreduce.Job: map 100% reduce 100%

19/01/27 21:26:24 INFO mapreduce.Job: Job job\_1548651821670\_0001 completed successfully

19/01/27 21:26:25 INFO mapreduce.Job: Counters: 49

File System Counters

FILE: Number of bytes read=296

FILE: Number of bytes written=287541

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=5458319

HDFS: Number of bytes written=197

HDFS: Number of read operations=6

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

Job Counters

Launched map tasks=1

Launched reduce tasks=1

Data-local map tasks=1

Total time spent by all maps in occupied slots (ms)=33645

Total time spent by all reduces in occupied slots (ms)=37400

Total time spent by all map tasks (ms)=33645

Total time spent by all reduce tasks (ms)=37400

Total vcore-milliseconds taken by all map tasks=33645

Total vcore-milliseconds taken by all reduce tasks=37400

Total megabyte-milliseconds taken by all map tasks=34452480

Total megabyte-milliseconds taken by all reduce tasks=38297600

Map-Reduce Framework

Map input records=124457

Map output records=901325

Map output bytes=7210600

Map output materialized bytes=296

Input split bytes=119

Combine input records=901325

Combine output records=29

Reduce input groups=29

Reduce shuffle bytes=296

Reduce input records=29

Reduce output records=29

Spilled Records=58

Shuffled Maps =1

Failed Shuffles=0

Merged Map outputs=1

GC time elapsed (ms)=448

CPU time spent (ms)=6460

Physical memory (bytes) snapshot=385482752

Virtual memory (bytes) snapshot=3015073792

Total committed heap usage (bytes)=226627584

Shuffle Errors

BAD\_ID=0

CONNECTION=0

IO\_ERROR=0

WRONG\_LENGTH=0

WRONG\_MAP=0

WRONG\_REDUCE=0

File Input Format Counters

Bytes Read=5458200

File Output Format Counters

Bytes Written=197

The median is: 4

[cloudera@quickstart Desktop]$

Message from syslogd@quickstart at Jan 29 00:05:56 ...

kernel:BUG: soft lockup - CPU#0 stuck for 97s! [SelectionManage:9343]

TAB-5

[cloudera@quickstart ~]$ cd Downloads/big-data-2/vector

[cloudera@quickstart vector]$ cd data

[cloudera@quickstart data]$ cat news3.csv | grep delegate

[cloudera@quickstart data]$ cat news2.csv | grep delegates

TAB-6

CREATE TABLE error\_log

> ROW FORMAT DELIMITED FIELDS TERMINATED by '\t'

> STORED AS TEXTFILE LOCATION '/data/error\_log'

> AS

> SELECT datetime,source,event\_id,details

> FROM system\_log

> WHERE level='Error';

CREATE TABLE staged\_log

> > (level STRING, datetime STRING,

> > source STRING,

> > event\_id INT,

> > category STRING,

> > details STRING)

> > Row FORMAT DELIMITED FIELDS TERMINATED BY '\t'

> > STORED AS TEXTFILE LOCATION '/data/s