

LAB 4:

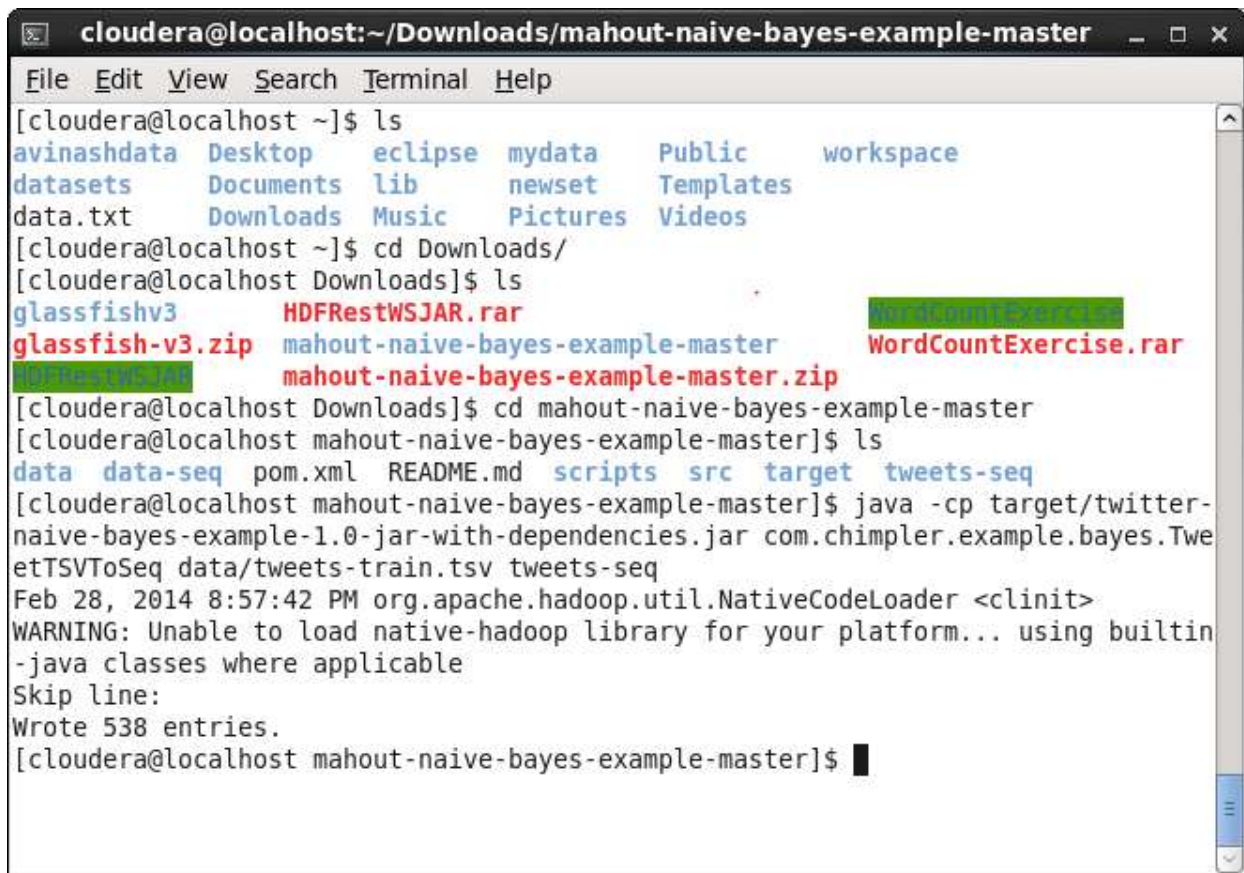
Submitted by

Mahesh Vemula

Task 1: Create a Mahout application

Data is taken from twitter tweets and stored in target folder under main directory

Step 1: Storing training data using HDFS



```
cloudera@localhost:~/Downloads/mahout-naive-bayes-example-master
File Edit View Search Terminal Help
[cloudera@localhost ~]$ ls
avinashdata Desktop eclipse mydata Public workspace
datasets Documents lib newset Templates
data.txt Downloads Music Pictures Videos
[cloudera@localhost ~]$ cd Downloads/
[cloudera@localhost Downloads]$ ls
glassfishv3 HDFRestWSJAR.rar WordCountExercise.rar
glassfish-v3.zip mahout-naive-bayes-example-master WordCountExercise.rar
HDFRestWSJAR mahout-naive-bayes-example-master.zip
[cloudera@localhost Downloads]$ cd mahout-naive-bayes-example-master
[cloudera@localhost mahout-naive-bayes-example-master]$ ls
data data-seq pom.xml README.md scripts src target tweets-seq
[cloudera@localhost mahout-naive-bayes-example-master]$ java -cp target/twitter-
naive-bayes-example-1.0-jar-with-dependencies.jar com.chimpler.example.bayes.Twe
etTSVToSeq data/tweets-train.tsv tweets-seq
Feb 28, 2014 8:57:42 PM org.apache.hadoop.util.NativeCodeLoader <clinit>
WARNING: Unable to load native-hadoop library for your platform... using builtin
-java classes where applicable
Skip line:
Wrote 538 entries.
[cloudera@localhost mahout-naive-bayes-example-master]$
```

Step 2: Copy file into Hadoop directory and run mahout naïve bayes algorithms on training data

```
cloudera@localhost:~/Downloads/mahout-naive-bayes-example-master
File Edit View Search Terminal Help
Running on hadoop, using /usr/lib/hadoop/bin/hadoop and HADOOP_CONF_DIR=/etc/hadoop/conf
MAHOUT-JOB: /usr/lib/mahout/mahout-examples-0.7-cdh4.4.0-job.jar
14/02/28 21:03:38 WARN driver.MahoutDriver: No split.props found on classpath, will use command-line arguments only
14/02/28 21:03:38 INFO common.AbstractJob: Command line arguments: [--endPhase=[2147483647], --input=[tweets-vectors/tfidf-vectors], --method=[sequential], --overwrite=null, --randomSelectionPct=[40], --sequenceFiles=null, --startPhase=[0], --tempDir=[temp], --testOutput=[test-vectors], --trainingOutput=[train-vectors]]
14/02/28 21:03:40 INFO common.HadoopUtil: Deleting train-vectors
14/02/28 21:03:40 INFO common.HadoopUtil: Deleting test-vectors
14/02/28 21:03:45 INFO utils.SplitInput: part-r-000000 has 471 lines
14/02/28 21:03:45 INFO utils.SplitInput: part-r-000000 test split size is 188 based on random selection percentage 40
14/02/28 21:03:45 INFO zlib.ZlibFactory: Successfully loaded & initialized native-zlib library
14/02/28 21:03:45 INFO compress.CodecPool: Got brand-new compressor [.deflate]
14/02/28 21:03:45 INFO compress.CodecPool: Got brand-new compressor [.deflate]
14/02/28 21:03:45 INFO utils.SplitInput: file: part-r-000000, input: 471 train: 324, test: 188 starting at 0
14/02/28 21:03:45 INFO driver.MahoutDriver: Program took 7593 ms (Minutes: 0.12655)
[cloudera@localhost mahout-naive-bayes-example-master]$
```

Step 3: Test algorithm on training data and then on test data to make a classifier

```
cloudera@localhost:~/Downloads/mahout-naive-bayes-example-master
File Edit View Search Terminal Help

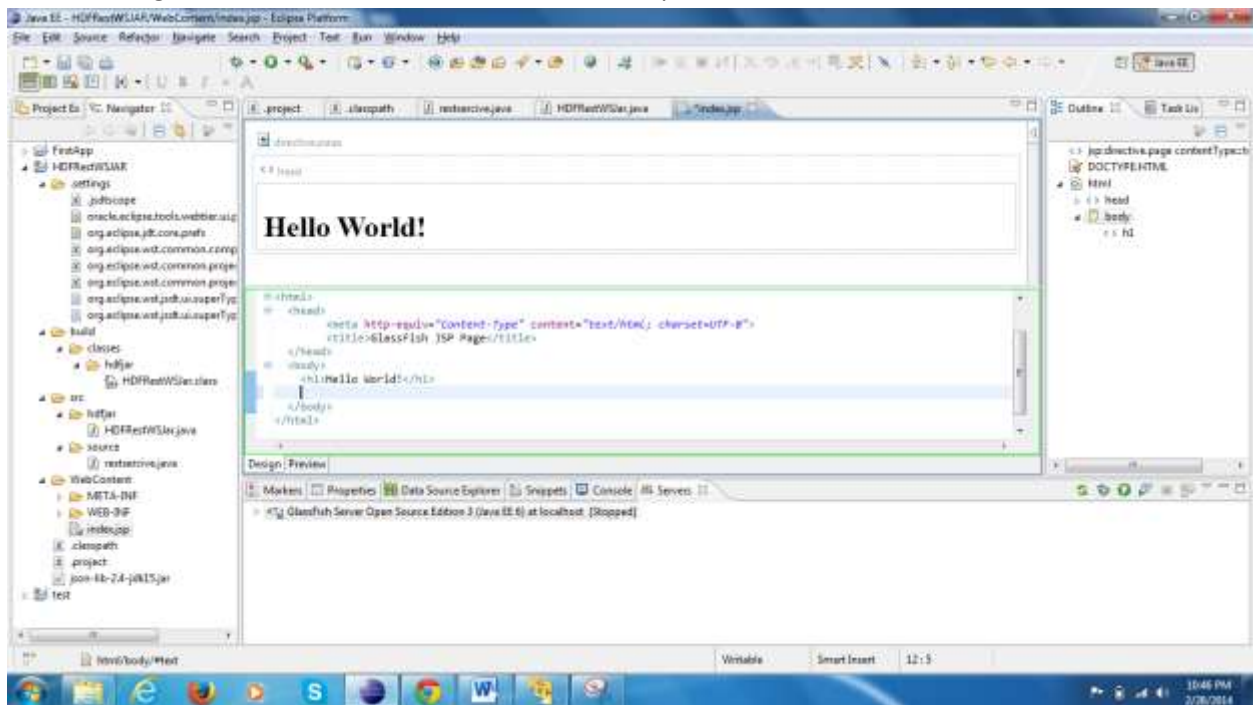
=====
Confusion Matrix
-----
a      b      c      d      e      f      g      <--Classified as
61      0      0      0      0      1      0      | 62      a      =
apparel
0      42      0      0      1      0      0      | 43      b      =
art
0      0      34      1      1      0      0      | 36      c      =
camera
0      0      0      37      0      0      0      | 37      d      =
event
0      0      0      0      29      0      0      | 29      e      =
health
1      2      0      0      1      31      0      | 35      f      =
home
0      0      1      0      0      0      81      | 82      g      =
tech

14/02/28 21:07:48 INFO driver.MahoutDriver: Program took 30822 ms (Minutes: 0.51
37)
[cloudera@localhost mahout-naive-bayes-example-master]$
```

#### Step 4: Finally classification of tweets

```
cloudera@localhost:~/Downloads/mahout-naive-bayes-example-master
File Edit View Search Terminal Help
Tweet: 309143843301912576      SubscriptionSave: £10 #discount code - £10 off a
subscription to Air Gunner magazine from SubscriptionSave using... http://t.co/mf
rV2FamMD
  apparel: -303.62569459307616  art: -259.23151687856205  camera: -328.263911830
57245  event: -287.0577685053108  health: -277.3669870598712  home: -300.4151380
970932  tech: -310.35689248314566 => art
Tweet: 309143735508291584      Weekly Deal 4 – Lorell 44553 Floor Fan on Sale.
special offer up to 50 off http://t.co/q5e4rKL87f #discount #promo
  apparel: -226.5291307187531  art: -264.09446486003844  camera: -288.0421843223
376  event: -238.4576987430118  health: -227.83759766803902  home: -216.00171779
383788  tech: -233.94028598393194 => home
Tweet: 309143443412770817      RT @Katheleen_Souza: House, M.D.: Seasons 3-4: H
OUSE:SEASON THREE & SEASON FOUR - DVD Movie http://t.co/mXHlcabMDm #discount
#deal
  apparel: -326.1933399124089  art: -269.73661955440804  camera: -319.9860890414
002  event: -316.62464605591845  health: -309.7337560386189  home: -302.66826307
152144  tech: -322.8530096492342 => art
Tweet: 309143199207804928      DS Miller Inc. Equivalent of ACER 5600 SERIES La
ptop AC Adapter: 19-Volt 90-Watt Laptop AC Ada... http://t.co/N88a3niE3j #discou
nt #deal
  apparel: -580.3848905202378  art: -568.2297450700363  camera: -580.97834766434
78  event: -561.7880448612377  health: -541.4020083342314  home: -558.9224055423
573  tech: -463.80509989155166 => tech
[cloudera@localhost mahout-naive-bayes-example-master]$
```

Now creating a rest server: Install Glassfish and add require libraries and start the server

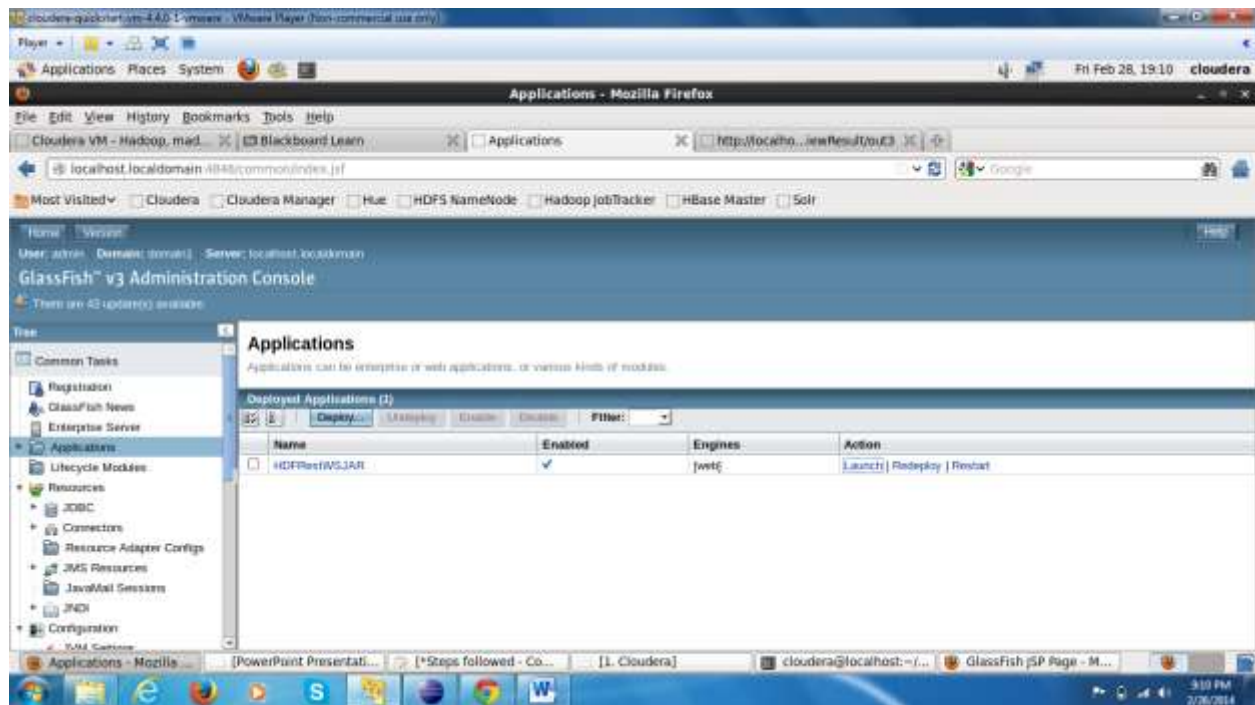


Download glassfish server in cloudera and start it

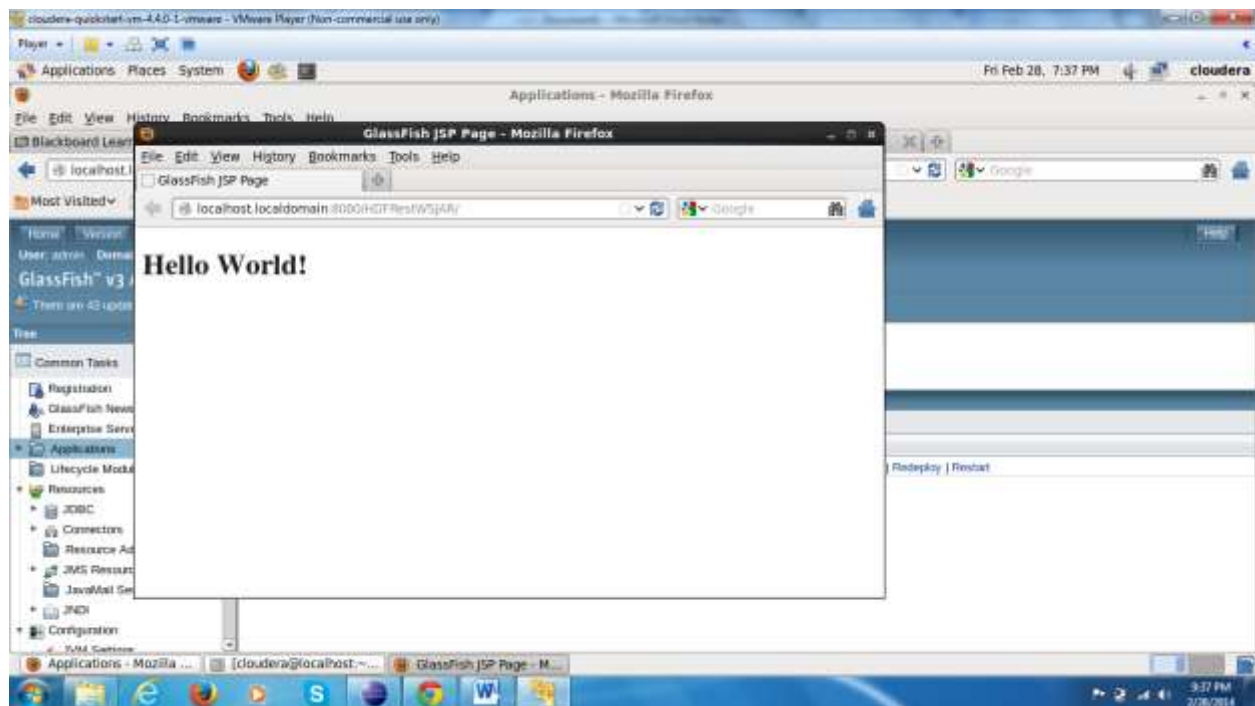




Load the rest service created into the server



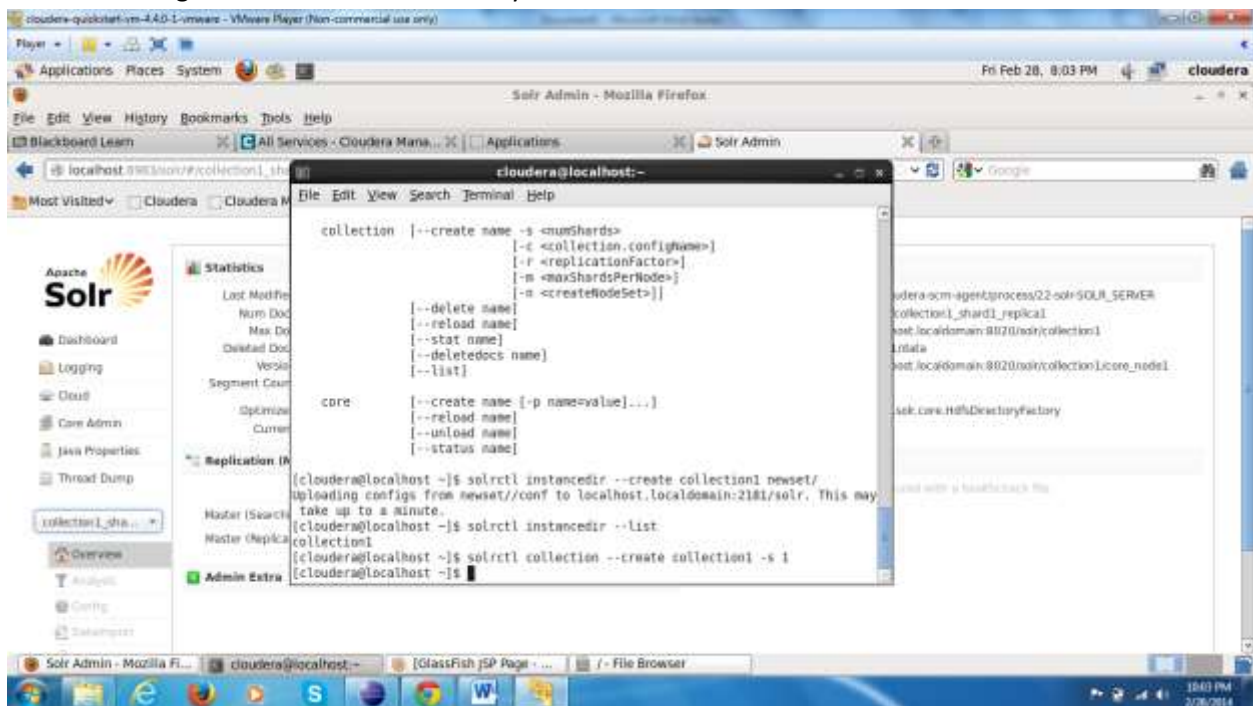
Launch the server



## Viewing data from Hadoop file system using server



## Now initializing solar and create a directory name collection1



## Details of collection1 directory

The screenshot shows the Solr Admin interface in a Mozilla Firefox browser. The address bar displays `localhost:8983/solr/#collection1_shard1_replica1`. The left sidebar contains navigation links: Dashboard, Logging, Cloud, Core Admin, Java Properties, and Thread Dump. The main content area is divided into two sections: Statistics and Instance.

**Statistics**

- Last Modified: 0
- Num Docs: 0
- Max Doc: 0
- Deleted Docs: 0
- Version: 1
- Segment Count: 0
- Optimized: ☒
- Current: ☒

**Replication (Master)**

Version	Size	Size
Master (Searching): 0	1	45 bytes
Master Replicable: -	-	-

**Instance**

- CWD: `/var/run/cloudera-scm-agent/process/22-solr-SOLR_SERVER`
- Instance: `/var/lib/solr/collection1_shard1_replica1`
- Data: `hdfs://localhost.localdomain:8020/solr/collection1`
- Index: `hdfs://localhost.localdomain:8020/solr/collection1/core_node1`
- Index: `/data/index`
- Impl: `org.apache.solr.core.HdfsDirectoryFactory`

**Admin Extra**

The bottom status bar shows the system time as 10:03 PM on Fri Feb 28, 2014.

## Insert the json data into solr

The screenshot shows the Solr Admin interface in a Mozilla Firefox browser. The address bar displays `localhost:8983/solr/#collection1_shard1_replica1/documents`. The left sidebar contains navigation links: Dashboard, Logging, Cloud, Core Admin, Java Properties, and Thread Dump. The main content area is divided into two sections: Request Handler and Response.

**Request Handler**

- Request Handler: `update`
- Document Type: `JSON`
- Document(s): `{ "id": "Mahesh", "role": "Student" }`
- Commit: ☐
- Overwrite: ☐
- Base: ☐

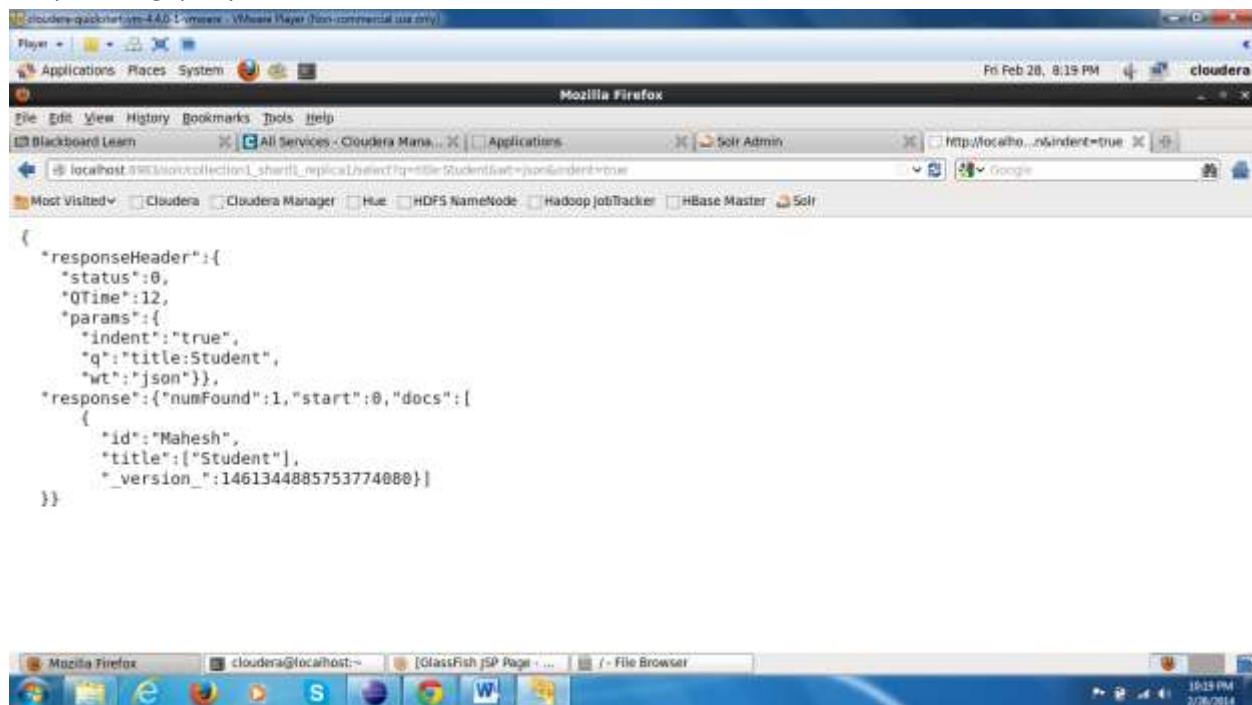
**Response**

```
{
  "responseHeader": {
    "status": 0,
    "offset": 100
  }
}
```

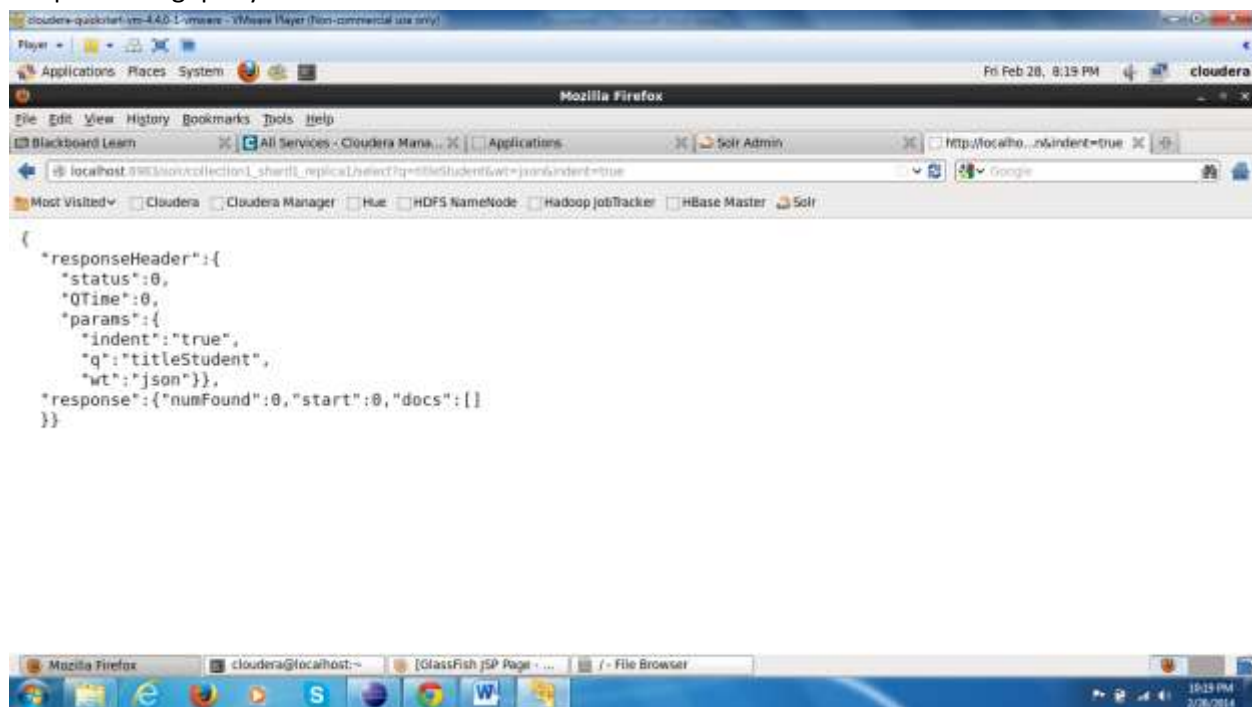
The bottom status bar shows the system time as 10:12 PM on Fri Feb 28, 2014.



## Output using query url



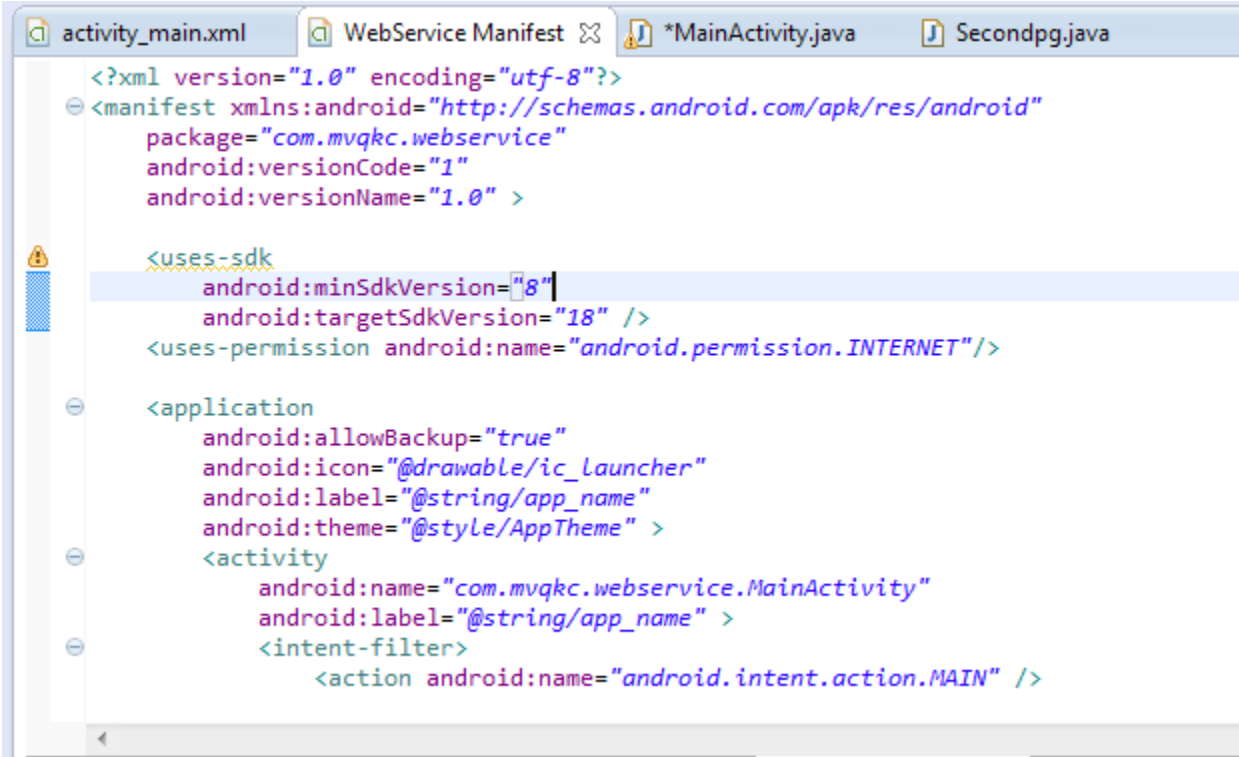
## Output 2 using query url from solr



## Task 2: Creating android service application

Step 1: Create a new project and edit java file, layout file and manifest file

Adding internet permission in manifest file:



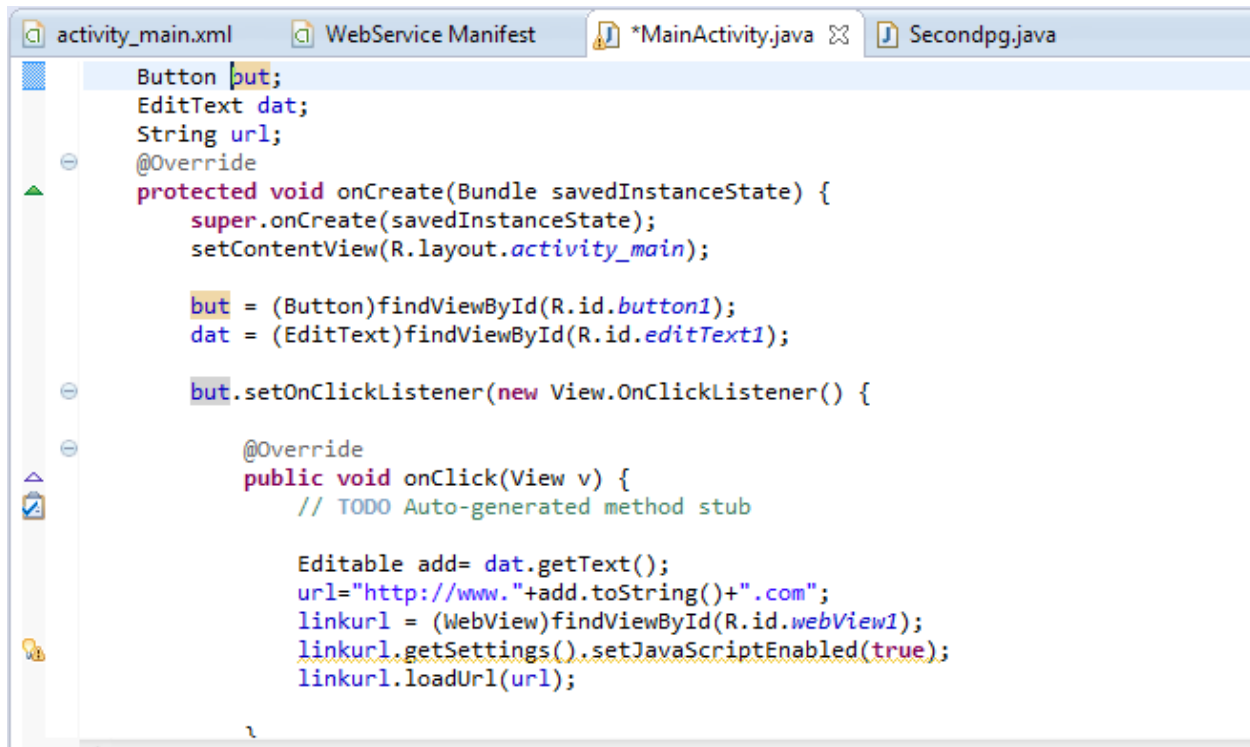
```
<?xml version="1.0" encoding="utf-8"?>
<manifest xmlns:android="http://schemas.android.com/apk/res/android"
    package="com.mvqkc.webservice"
    android:versionCode="1"
    android:versionName="1.0" >

    <uses-sdk
        android:minSdkVersion="8"
        android:targetSdkVersion="18" />

    <uses-permission android:name="android.permission.INTERNET"/>

    <application
        android:allowBackup="true"
        android:icon="@drawable/ic_launcher"
        android:label="@string/app_name"
        android:theme="@style/AppTheme" >
        <activity
            android:name="com.mvqkc.webservice.MainActivity"
            android:label="@string/app_name" >
            <intent-filter>
                <action android:name="android.intent.action.MAIN" />
            </intent-filter>
        </activity>
    </application>
</manifest>
```

Defining java file for dynamic behavior:



```
activity_main.xml | WebService Manifest | *MainActivity.java | Secondpg.java
Button but;
EditText dat;
String url;
@Override
protected void onCreate(Bundle savedInstanceState) {
    super.onCreate(savedInstanceState);
    setContentView(R.layout.activity_main);

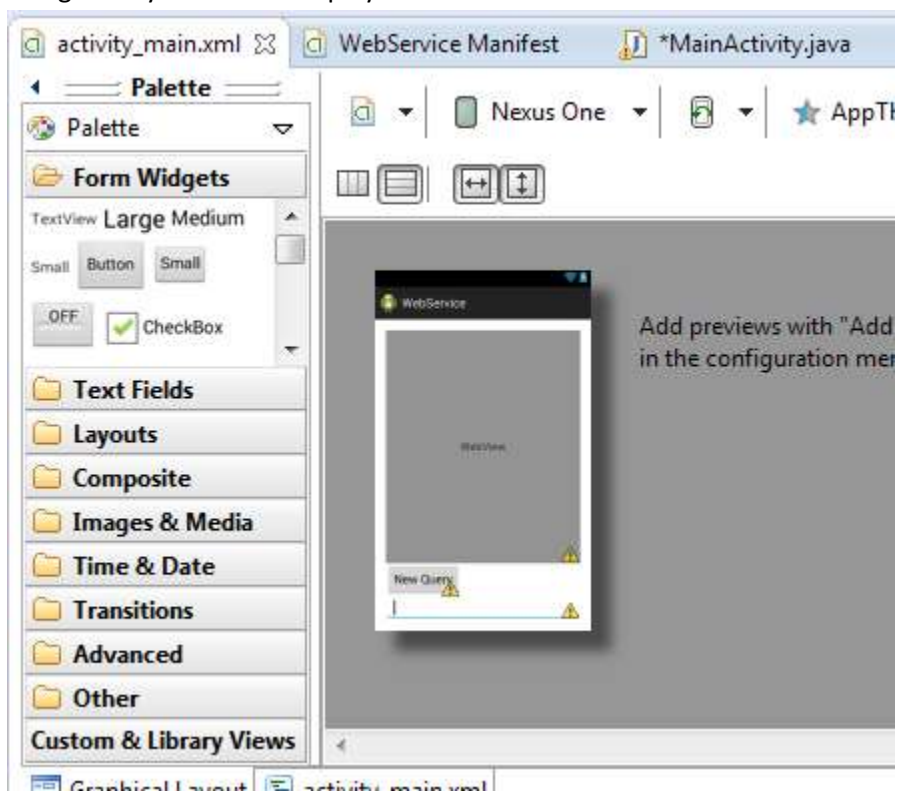
    but = (Button)findViewById(R.id.button1);
    dat = (EditText)findViewById(R.id.editText1);

    but.setOnClickListener(new View.OnClickListener() {

        @Override
        public void onClick(View v) {
            // TODO Auto-generated method stub

            Editable add= dat.getText();
            url="http://www."+add.toString()+".com";
            linkurl = (WebView)findViewById(R.id.webView1);
            linkurl.getSettings().setJavaScriptEnabled(true);
            linkurl.loadUrl(url);
        }
    });
}
```

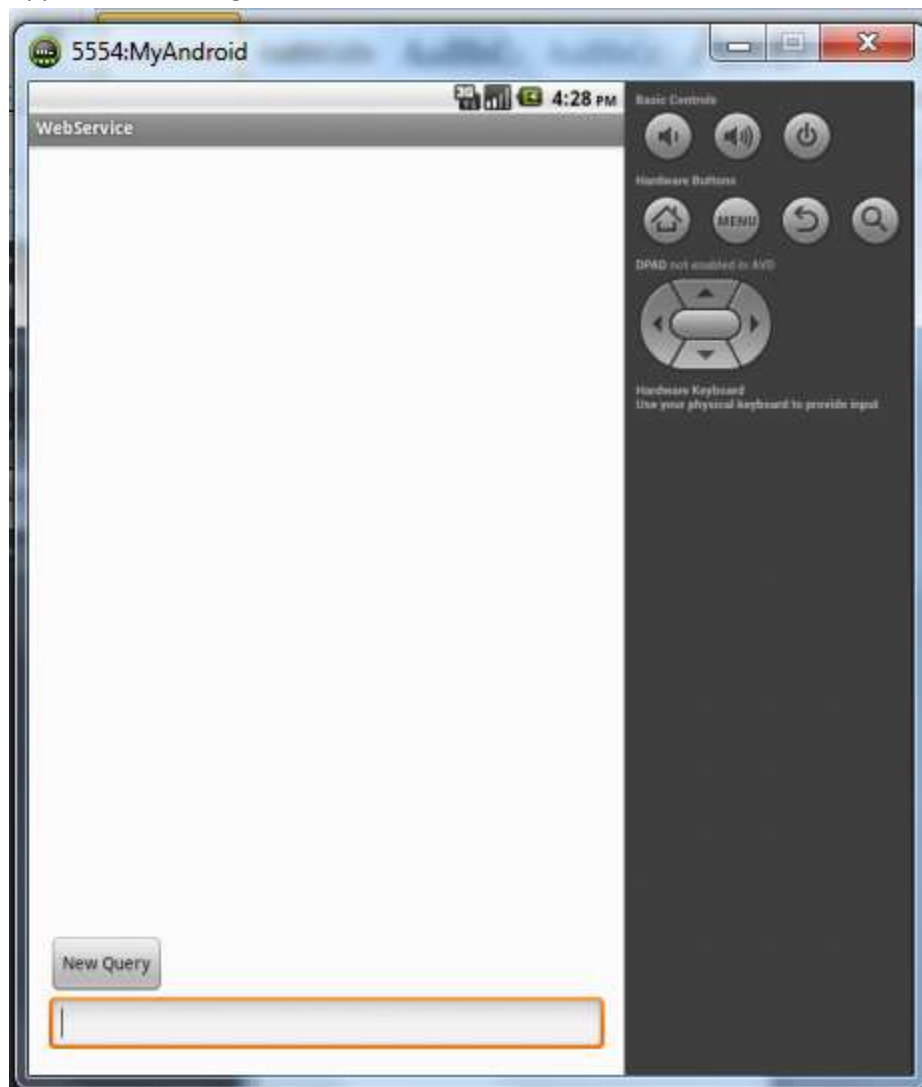
Design of layout for the display:



Running application on android emulator



Application running first view:





After entering Student in and click new query in localhost of cloudera where we stored data the response:

