**UStudy**

**A complete educational portal**

**Arapally Ebenezer Anand**

**Bandaru Sai Kishore**

**Kommineni Sivakrishna**

**Ponnam Balakrishna**

**Summary:**

For the second iteration the main focus is made on the data analysis and web services part. We used certain tools for analyzing the data and finding out the right algorithm for data classification and analysis. Our back end work majorly consists of analyzing the data and cluster the list of colleges based on certain parameters. Some amount of work has also been spent on the developing of UI part and investigations were made for storing the user profiles in a NoSQL store. In order to not to maintain monotony in the look of web pages we are planning include some graphics in the user profiles which may be forwarded and developed during the third phase.

**Framework Specifications:**

In this increment we made progress in the deployment of Solr on a web server. The web server here we used is glassfish and now we were able to upload the files from the Hadoop database to the solr. We have also applied the Naïve Bayes algorithm for classification of data on a sample of data that we have and were able to build the sequence files and vectors. But due some errors we couldn’t generate the matrix required for classification. Currently we are continuing on resolving the issues. The mobile client is entirely developed in Android and it is a native application for android users. The version supports from Froyo to JellyBeans.

**Application Specification:**

* **Software Specification**
  + Tools: Mahout, Solr, Android Development Kit, WEKA
  + Operating System: Android
  + Development Operating System: Windows 8
  + Programming Language: Java 7.0
  + Databases: Oracle, Hadoop

**Implementation:**

The implementation part typically consisted of three parts. They are:

* Identifying the algorithms for the data classification and applying them on to the data to generate the output.
* Deploying Solr on a web server and uploading the output files generated from Hadoop system to the Solr.
* Designing of new web pages for the user home pages and search pages.

For the data classification we basically considered the Naïve Bayes approach for classification of data. To test this algorithm we have taken a sample of data as training data and applied the algorithm on the data for generating the sequence files and the vectors required for the data classification on the training data. The final goal of applying these algorithm on the data is to classify the data on the basis of region and financial aid. The number of students enrolled will be the key parameter for this classification. However in our implementation process we are facing certain exceptions like OutOfMemory exceptions, and we are in the process of resolving them. The following are the screenshots taken during the implementation of Naïve Bayes algorithm on sample data:

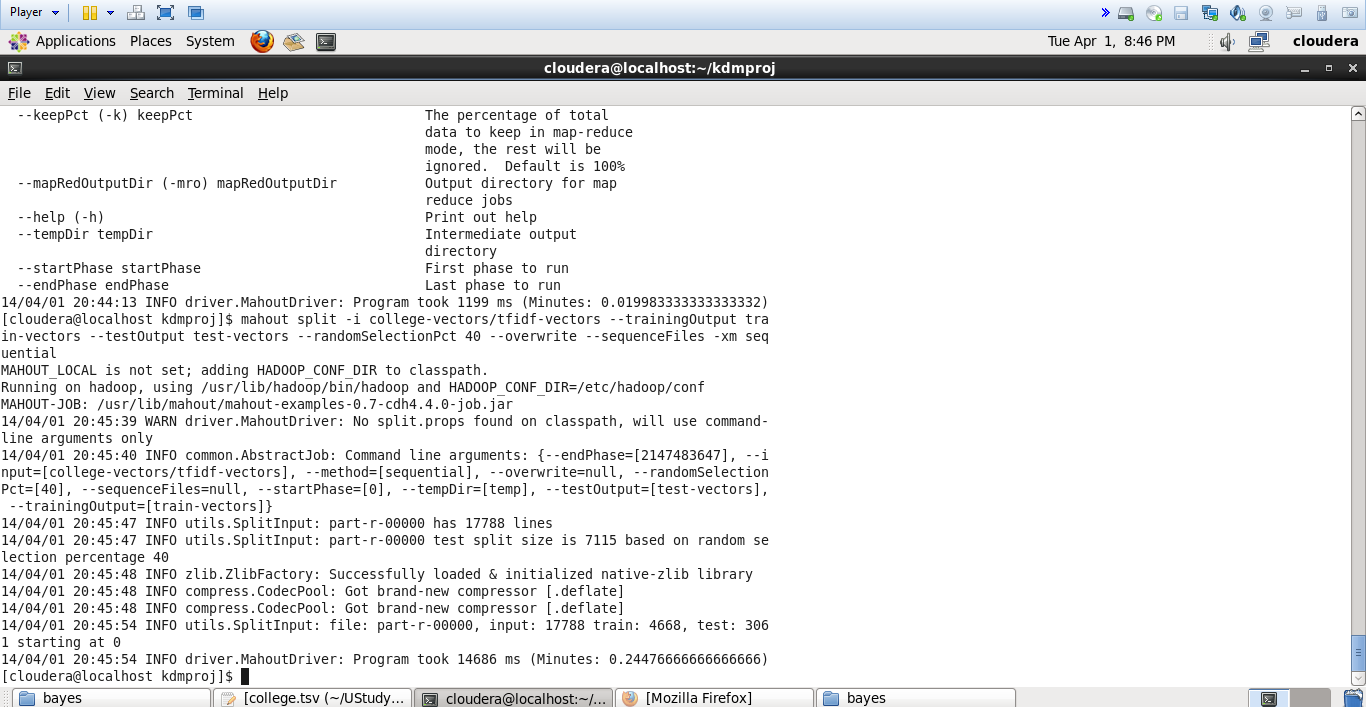


Figure 1 naive bayes application

The below figure depicts the generation of sequence file using mahout:

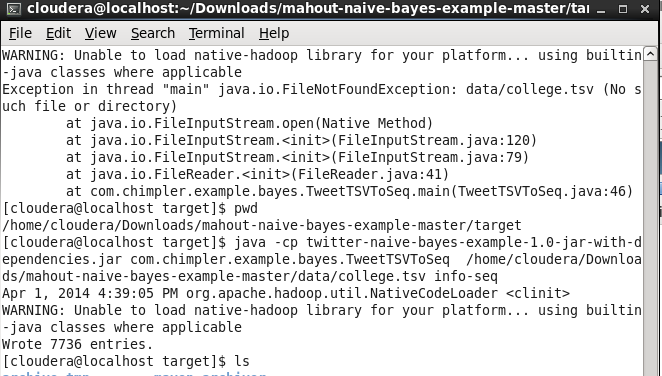


Figure 2sequence file generation

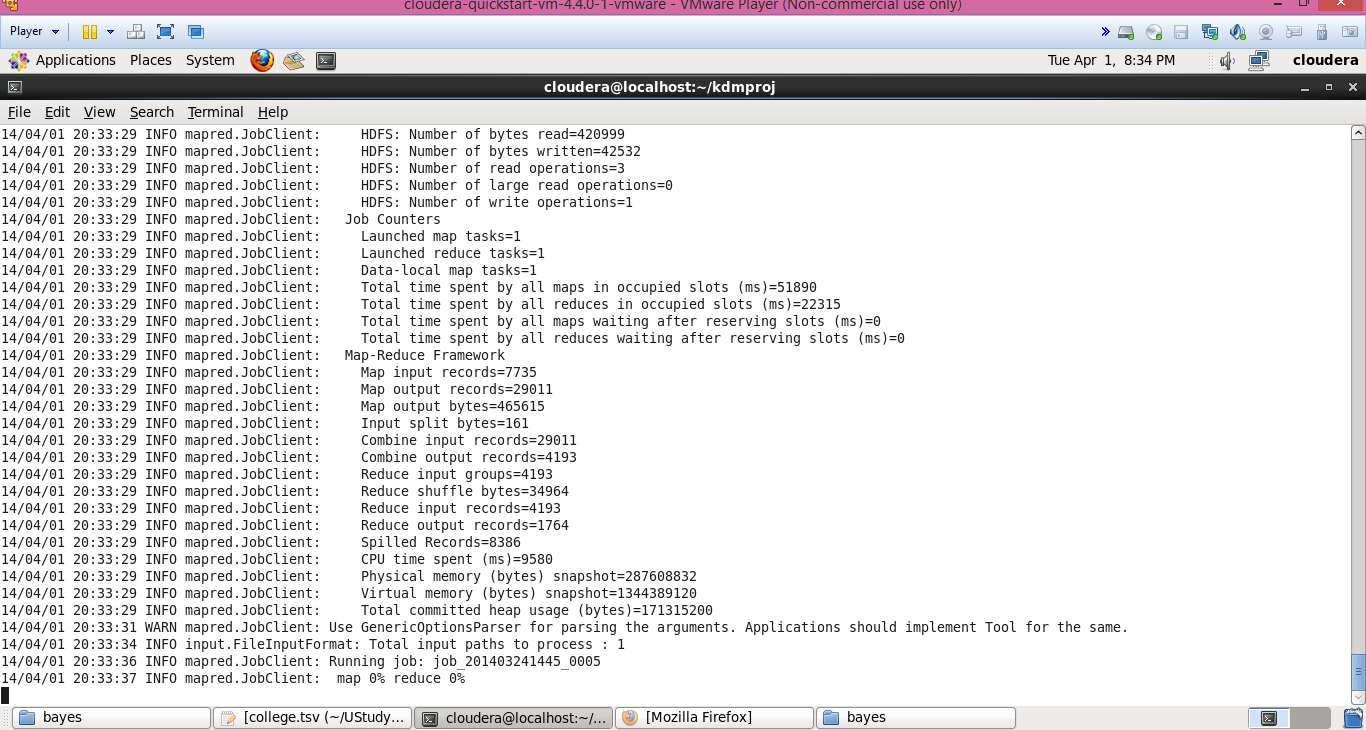


Figure 3Mapreduce execution

But while running the command for getting the matrix we were facing out of memory exceptions which we are trying to resolve, so that we can apply this classification for entire data.

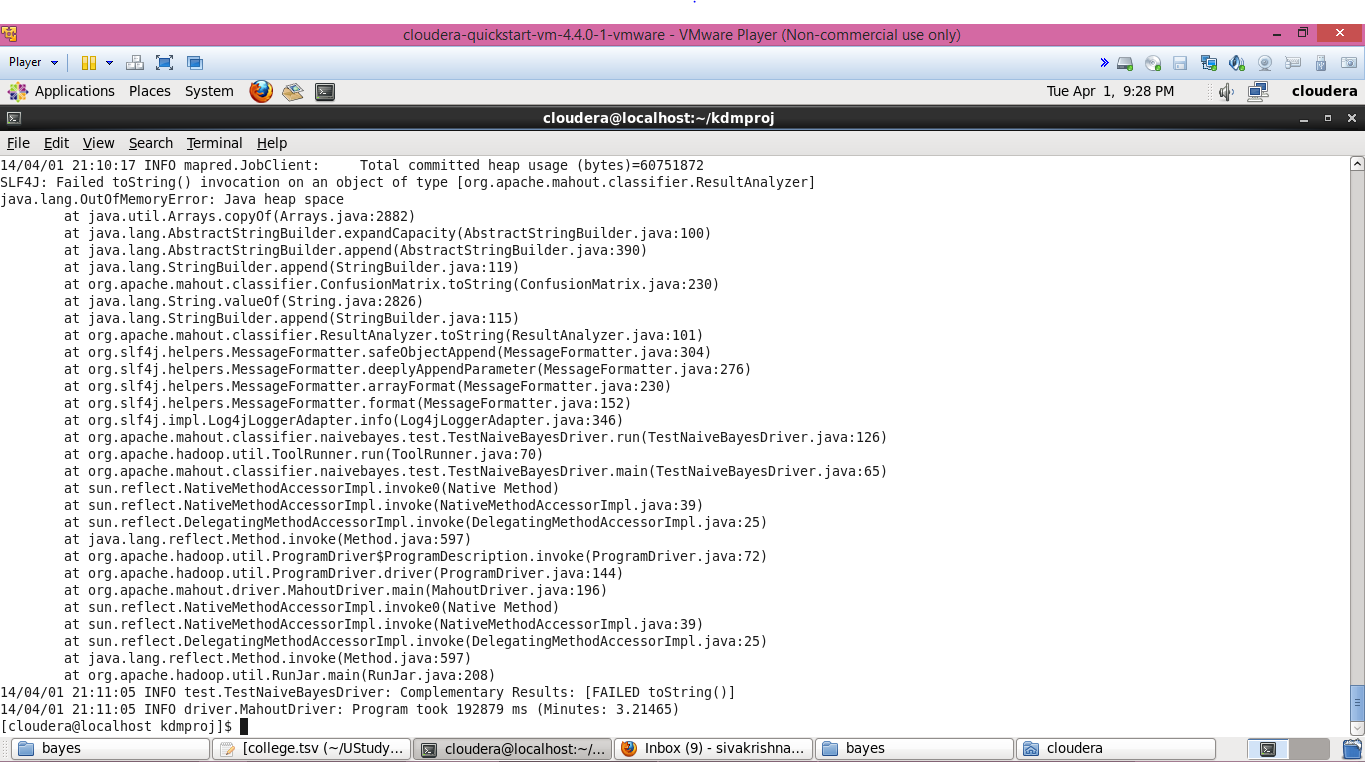


Figure 4 out of memory exceptions

In addition to these we are also trying to use the tool weka for finding out the right algorithm for our data classification. Considerable amount of hours has been spent on exploring the tool for finding the appropriate algorithms.

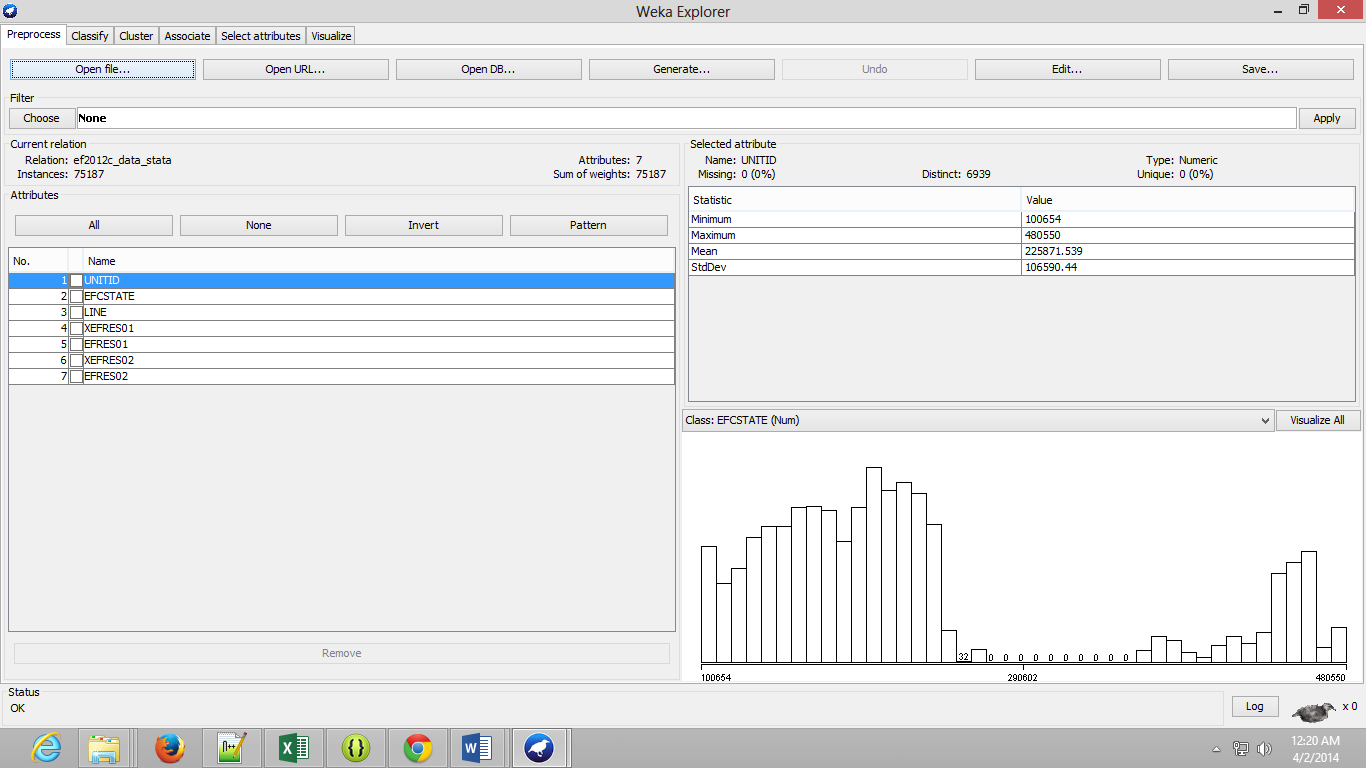


Figure 5 weka analysis for the sample data

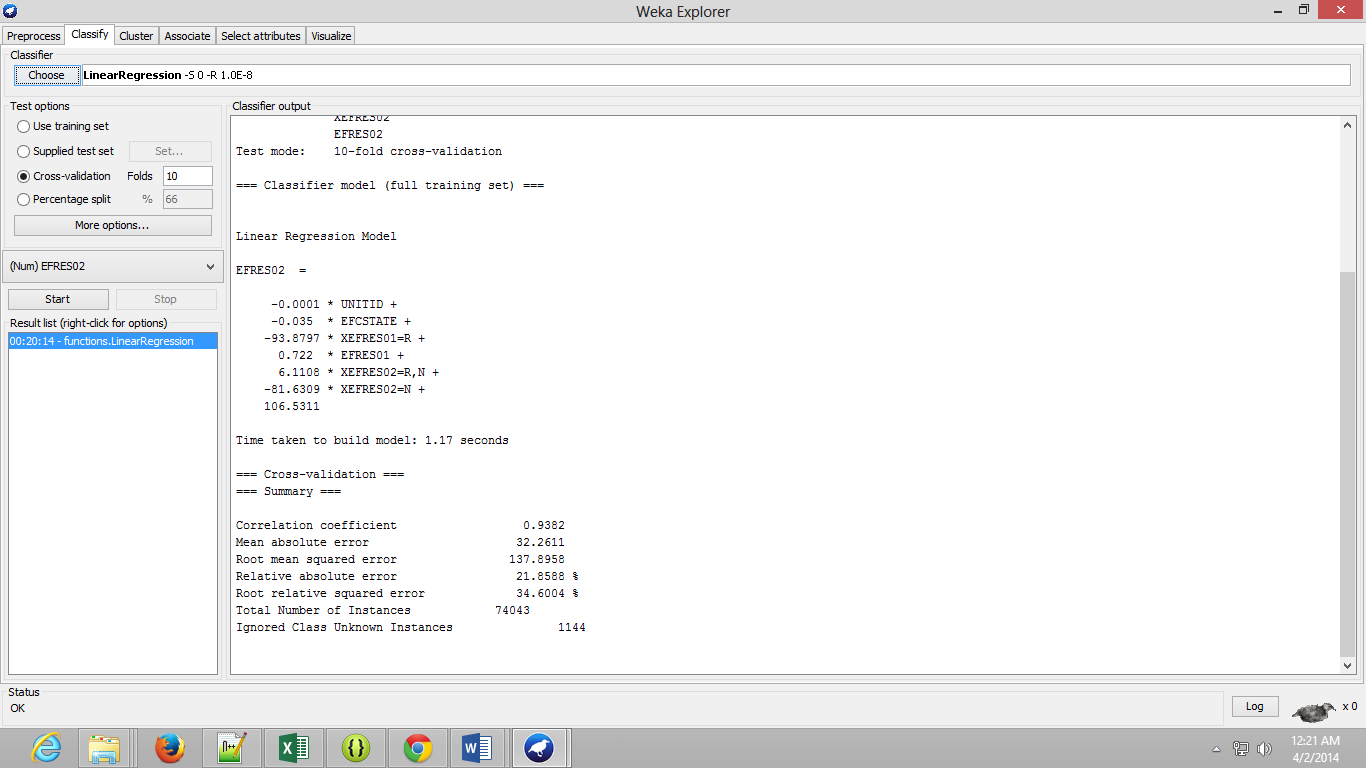


Figure 6 Results obtained for using linear regression in weka for classification

The second task we completed was deploying of solr on the glassfish server and upload the output files from Hadoop database to solr directly. We have completed these task successfully for JSON format files and currently checking compatibility for txt or csv files. We found certain online tools which can convert the txt or any other format files to json format. The following are the snapshots for the deploying of the web server:

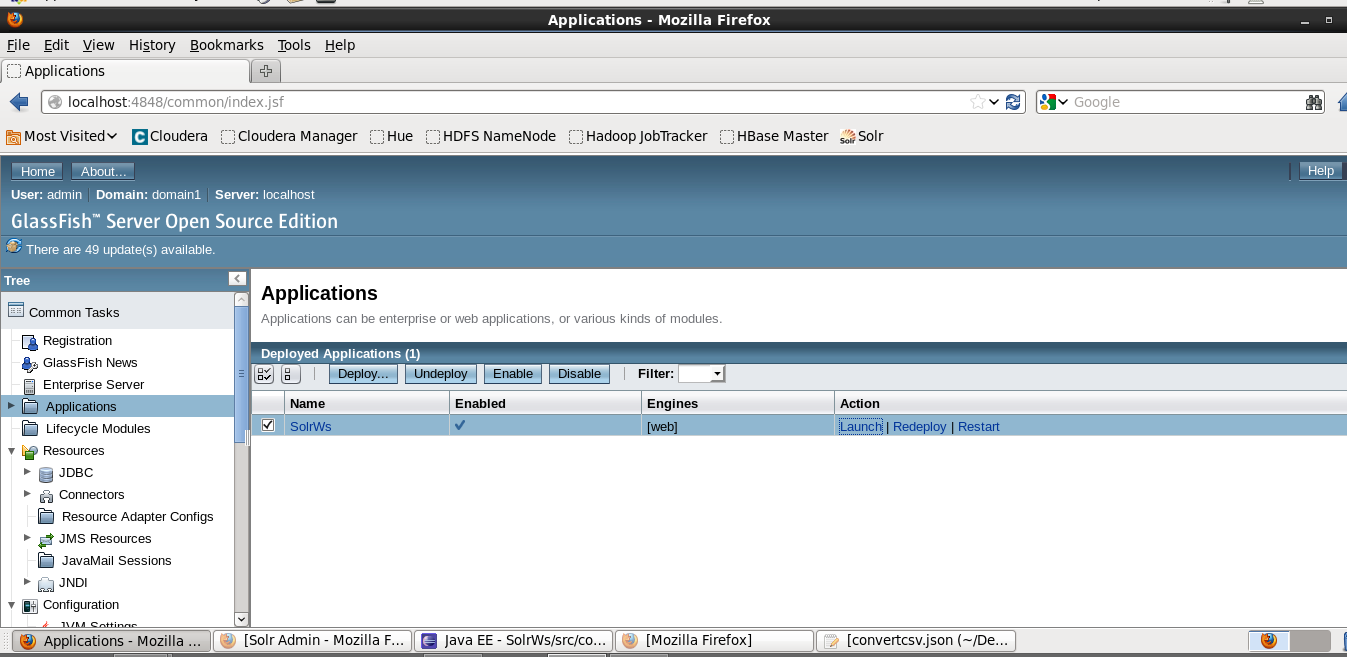


Figure 7deployment of solr onto the glassfish web server

We have also created certain android pages for the user interface which includes the user’s home page and guest home page. Due to more focus on data classification and analysis we couldn’t approach much on the GUI part of our project.

**Project Management:**

All the tasks and their day today increments are being updated in the scrumdo.

* <https://www.scrumdo.com/organization/university-of-missouri-kansas-city5/dashboard>

**Third Increment:**

The tasks that will be included for third increment are:

1. Completion of the entire client web pages – Ponnam Balakrishna
2. Registration of user profiles onto NoSQL store and analyzing those data – Ebenezer Anand Arapally
3. Data Classification and Data recommendation based on the data available and retrieving the output from classification algorithms – Sai Kishore Bandaru, Ebenezer Anand Arapally
4. Retrieving the output from Solr based on the inputs specified by user and also pushing the data onto the Solr from the Hadoop database system – Kommineni Sivakrishna

The above mentioned tasks will be uploaded the scrumdo tool with specified timelines.