International Journal of Management, IT & Engineering
Vol. 9 Issue 7, July 2019,
ISSN: 2249-0558 Impact Factor: 7.119
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

IMPLEMENTATION OF AM_EBIZ ALGORITHM FOR FINDING FREQUENT ITEMSETS

Dr.Shikha Dubey[*]

Abstract

In recent years, knowledge discovery has been focused on advancements with the vision of benefits to the corporate sector. This research work explicates the concepts of data mining, especially the importance of Frequent Pattern Mining (FPM) and Hadoop MapReduce programming model**.** The organization sectors such as social networks, web browsing, telecommunication, utilities, transportation, insurance, banking and many others generate vast amount of data every day. The proposed work for frequent item set mining of large volume of data is different from all the previous works, since it uses AM e_biz Algorithm with map reduce programming model. It provides compressed storage facility, reduced execution time. It is applicable in many of data mining applications like market-basket problem, decision support, web usage mining, and bioinformatics.

Keywords:

Association Rule Mining , Hadoop, Frequent itemset, Algorithm, Mapreduce.

[*] **HOD MCA, D. Y. P. I. M. R Pimpri, Pune-18**

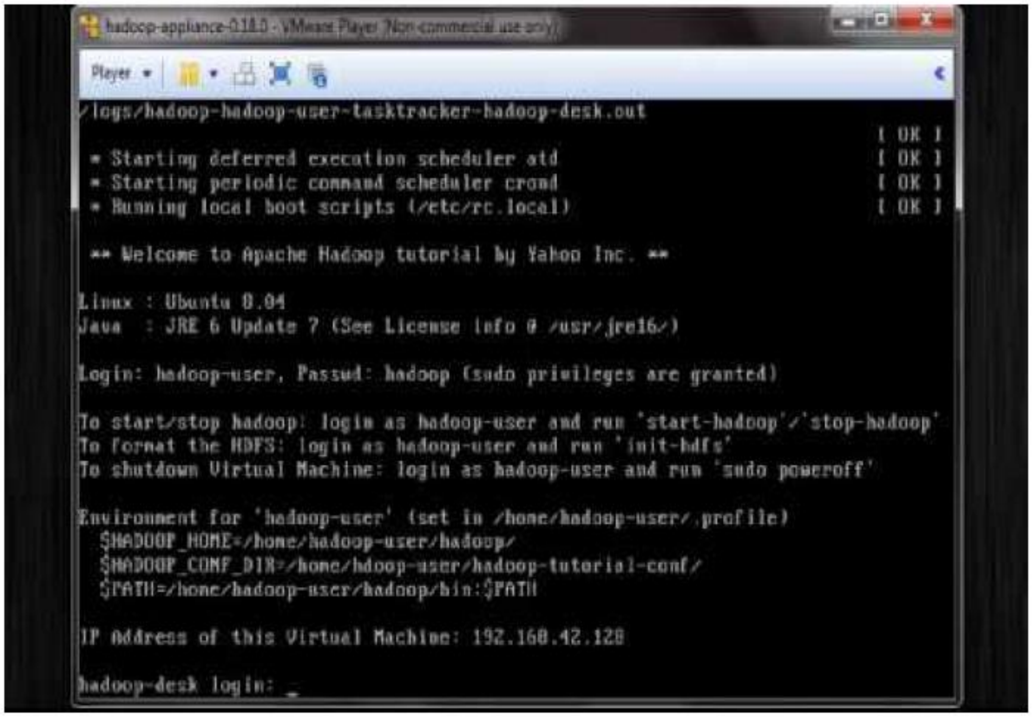
1. Introduction

To provide great user experience to users in their day to day activity the Big Data needs to be analyzed. But because of development of IT industries, services, technologies and data, the tremendous measure of complex data is produced from the different sources that can be in different frame. Such complex and massive data is hard to handle and process. So Size, Complexity, and variability of big data are the real difficulties to perceive association rules and frequent item sets.

Various existing data mining techniques are produced displayed to infer association rule and frequently happening item sets, yet with the quick entry of time of big data customary data mining algorithm have not been able to meet huge datasets analysis requirements. There is need to improve performance and accuracy of parallel processing with minimizing execution time complexity. Also assuring the output of a computation is insensitive to changes in any one personal record. So that it will restricting privacy leaks from results. Hence, there is need to provide better frequent item set mining approach using HDFS framework with privacy Preservation techniques.

Formally, frequent item set mining is to perform the following task: we are given a set B = i1, in of items, called the item base and a database T = (t1………tm) of transactions. An item may, for example, represent a product. In this case, the item base represents the set of all products offered by a supermarket. The term item set refers to any subset of the item base B. Each transaction is an item set and may represent, in the supermarket setting, a set of products that has been bought by a customer. As several customers may have bought the same set of products, the total of all transactions must be represented as a vector (as above) or as a multi set. Alternatively, each transaction may be enhanced by a transaction identifier (Tid). Note that the item base B is usually not given explicitly, but only implicitly as the union of all transactions, that is, B = k (-{1…..m} tk.

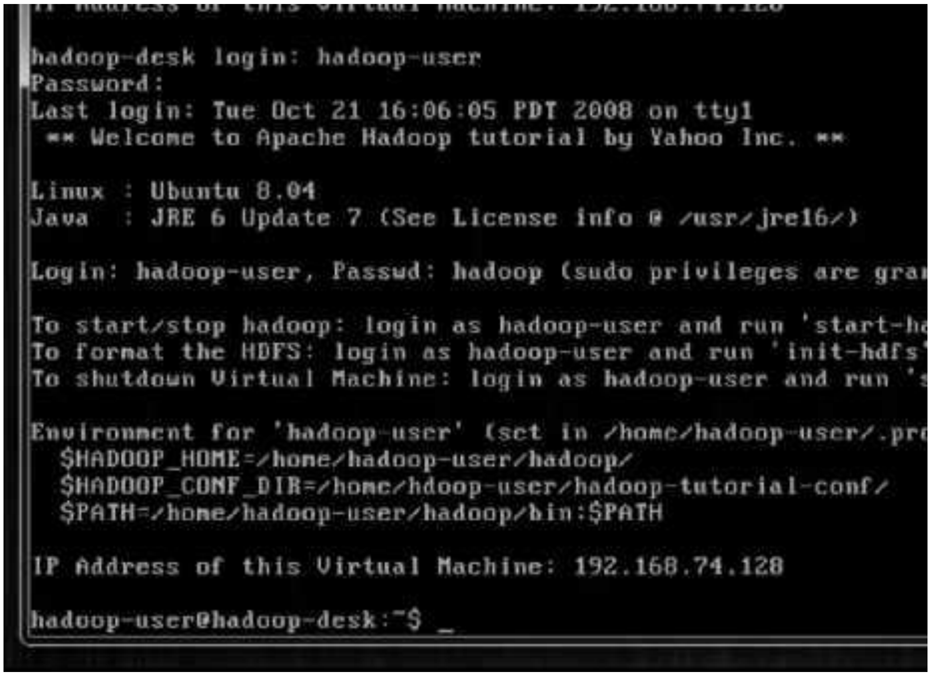
2 Research Methodology: The experiment is carried out in a machine with Intel core i5-M480 with 2.67GHz. And RAM of 4GB. VMware workstation is installed and two VMs are setup. One VM is for Hadoop 0.18.0 setup with configuration of 250GB hard disk and 512 MB main memory. The other VM is for Eclipse Juno, the virtual configuration is 20GB hard disk and 2 GB main memory. Ubuntu 14.04 is installed in it. Virtual machines are created using VMware Workstation 7.1.3. Necessary packages are installed for eucalyptus and virtual cloud setup is done in VMWare workstation. Eucalyptus platform is installed with Ubuntu packages in a single system with VMWare. Ubuntu image is used in this research work. Bridged connection will be set as 0 because of the same machine. If multiple machines are used IP address of the respective machines are to be used. Hadoop is installed in one virtual machine and eucalyptus is installed in another virtual machine with eclipse environment. Soon after the configuration is completed eucalyptus will be started. Figure 1.1 shows the screenshot of Hadoop login. It has been given to understand the clear picture of virtual machine setup of Hadoop.


hadoop-appliance-0.18.0 - VMware Player (Non-commercial use only)
Player

```
/logs/hadoop-hadoop-user-tasktracker-hadoop-desk.out
                                                                      [ OK ]
 * Starting deferred execution scheduler atd                          [ OK ]
 * Starting periodic command scheduler crond                          [ OK ]
 * Running local boot scripts (/etc/rc.local)                         [ OK ]

 ** Welcome to Apache Hadoop tutorial by Yahoo Inc. **

Linux : Ubuntu 8.04
Java  : JRE 6 Update 7 (See License info @ /usr/jre16/)

Login: hadoop-user, Passwd: hadoop (sudo privileges are granted)

To start/stop hadoop: login as hadoop-user and run 'start-hadoop'/'stop-hadoop'
To format the HDFS: login as hadoop-user and run 'init-hdfs'
To shutdown Virtual Machine: login as hadoop-user and run 'sudo poweroff'

Environment for 'hadoop-user' (set in /home/hadoop-user/.profile)
  $HADOOP_HOME=/home/hadoop-user/hadoop/
  $HADOOP_CONF_DIR=/home/hdoop-user/hadoop-tutorial-conf/
  $PATH=/home/hadoop-user/hadoop/bin:$PATH

IP Address of this Virtual Machine: 192.168.42.128

hadoop-desk login: _
```

Figure 1.1 Screenshot of Hadoop login

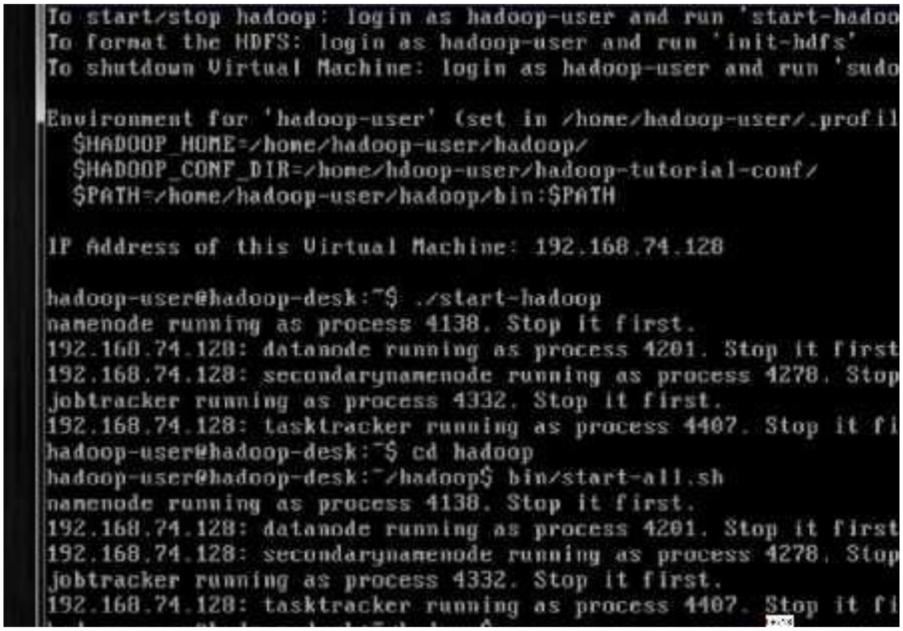
Figure 1.2 gives the screenshot of Hadoop environment that is appearing soon after the setup process of Hadoop login.



```
hadoop-desk login: hadoop-user
Password:
Last login: Tue Oct 21 16:06:05 PDT 2008 on tty1
 ** Welcome to Apache Hadoop tutorial by Yahoo Inc. **

Linux  : Ubuntu 8.04
Java   : JRE 6 Update 7 (See License info @ /usr/jre16/)

Login: hadoop-user, Passwd: hadoop (sudo privileges are gra

To start/stop hadoop: login as hadoop-user and run 'start-ha
To format the HDFS: login as hadoop-user and run 'init-hdfs'
To shutdown Virtual Machine: login as hadoop-user and run 's

Environment for 'hadoop-user' (set in /home/hadoop-user/.pro
  $HADOOP_HOME=/home/hadoop-user/hadoop/
  $HADOOP_CONF_DIR=/home/hdoop-user/hadoop-tutorial-conf/
  $PATH=/home/hadoop-user/hadoop/bin:$PATH

IP Address of this Virtual Machine: 192.168.74.128

hadoop-user@hadoop-desk:~$ _
```

Figure 1.2 Screenshot of Hadoop screen

Environment variables and IP address of the virtual machine are being set. The subsequent figure 1.3 presents with the Hadoop screen shot.



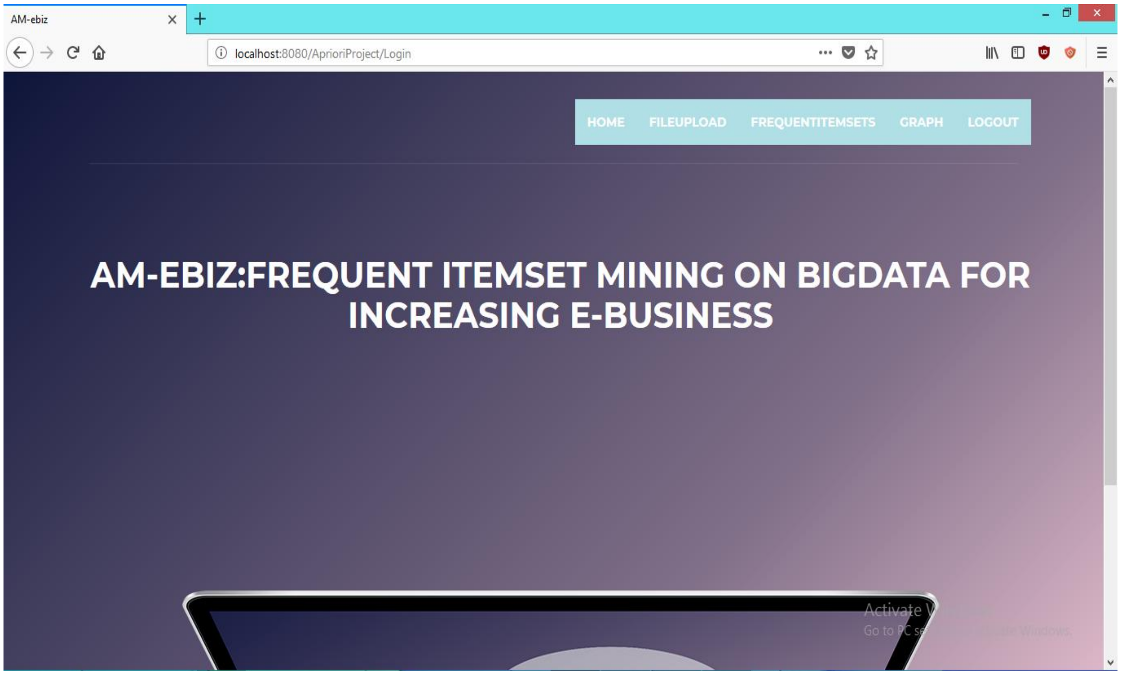
```
To start/stop hadoop: login as hadoop-user and run 'start-hadoo
To format the HDFS: login as hadoop-user and run 'init-hdfs'
To shutdown Virtual Machine: login as hadoop-user and run 'sudo

Environment for 'hadoop-user' (set in /home/hadoop-user/.profil
  $HADOOP_HOME=/home/hadoop-user/hadoop/
  $HADOOP_CONF_DIR=/home/hdoop-user/hadoop-tutorial-conf/
  $PATH=/home/hadoop-user/hadoop/bin:$PATH

IP Address of this Virtual Machine: 192.168.74.128

hadoop-user@hadoop-desk:~$ ./start-hadoop
namenode running as process 4138. Stop it first.
192.168.74.128: datanode running as process 4201. Stop it first
192.168.74.128: secondarynamenode running as process 4278. Stop
jobtracker running as process 4332. Stop it first.
192.168.74.128: tasktracker running as process 4407. Stop it fi
hadoop-user@hadoop-desk:~$ cd hadoop
hadoop-user@hadoop-desk:~/hadoop$ bin/start-all.sh
namenode running as process 4138. Stop it first.
192.168.74.128: datanode running as process 4201. Stop it first
192.168.74.128: secondarynamenode running as process 4278. Stop
jobtracker running as process 4332. Stop it first.
192.168.74.128: tasktracker running as process 4407. Stop it fi
```

Figure 1.3 Screenshot of Hadoop process running state

It is showing the Name Node, Secondary Name Node, Job Tracker and Task Trackers running. The algorithms are implemented which gives most frequent item set within very less time.

3. Results and Analysis

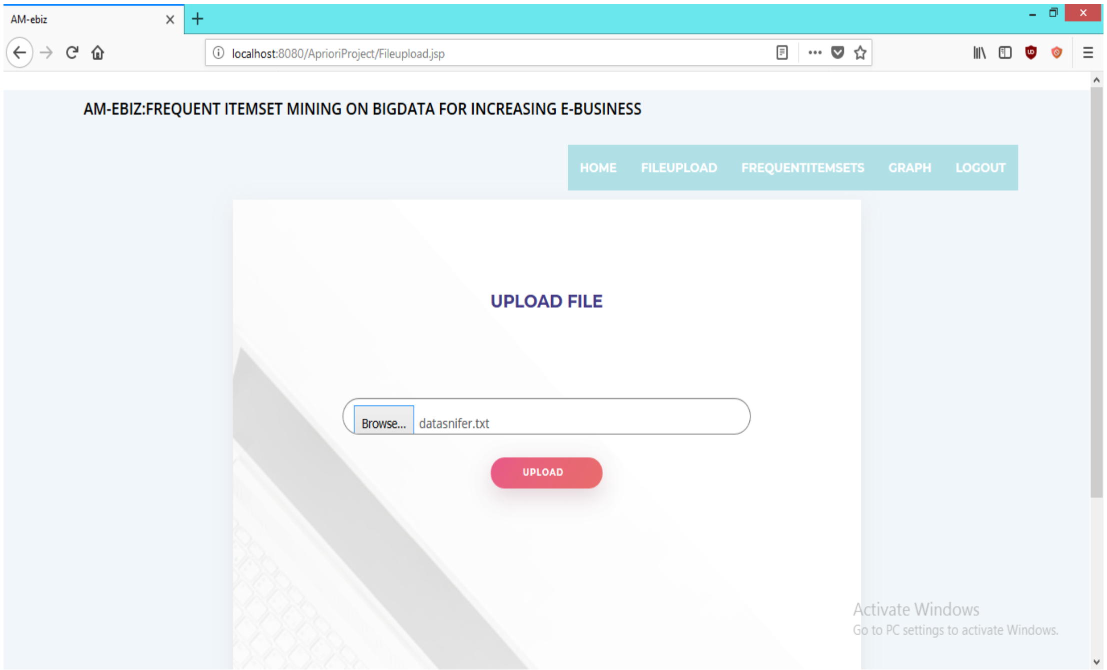
The following figure shows the main page of the system.



AM-EBIZ:FREQUENT ITEMSET MINING ON BIGDATA FOR INCREASING E-BUSINESS

Figure 1.4: Main page AM_ebiz

After this the system requests for uploading the data which can be browsed from the available files which is shown in Figure 1.5.Figure 1.6 shows the choice of dates to be selected by the user. Later the figures shows how 1 item frequent item set ,2 item frequent item set, three item frequent item set is retrieved.

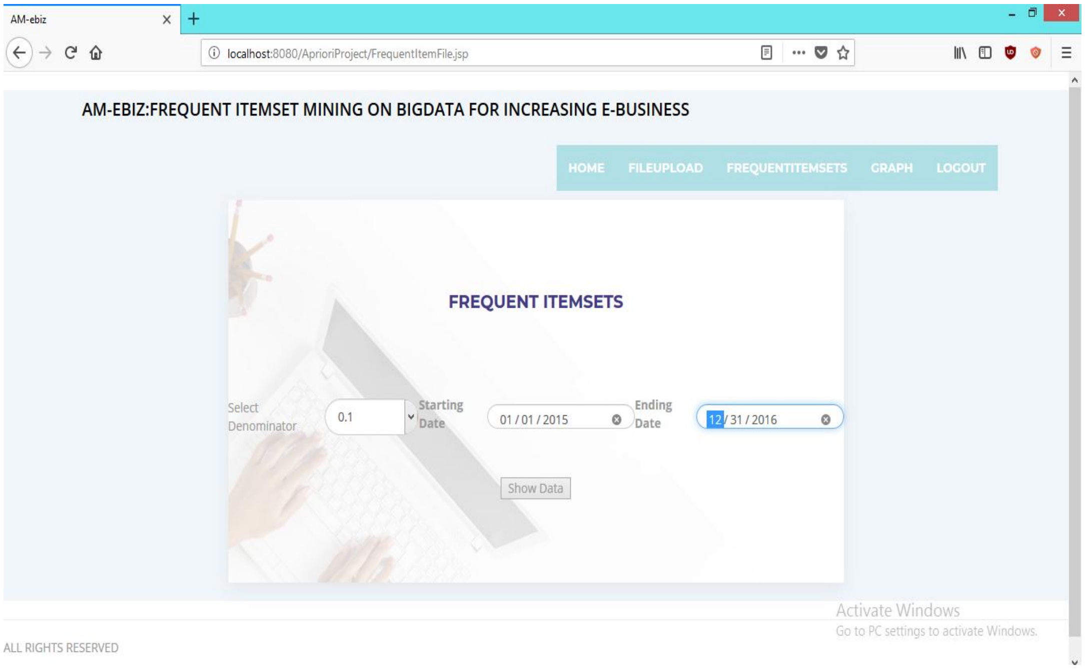

AM-ebiz
localhost:8080/AprioriProject/Fileupload.jsp
HOME
FILEUPLOAD
FREQUENTITEMSETS
GRAPH
LOGOUT
Browse...
UPLOAD
Activate Windows
Go to PC settings to activate Windows.

AM-EBIZ:FREQUENT ITEMSET MINING ON BIGDATA FOR INCREASING E-BUSINESS

UPLOAD FILE

datasnifer.txt

Figure 1.5: Dataset Uploading Phase

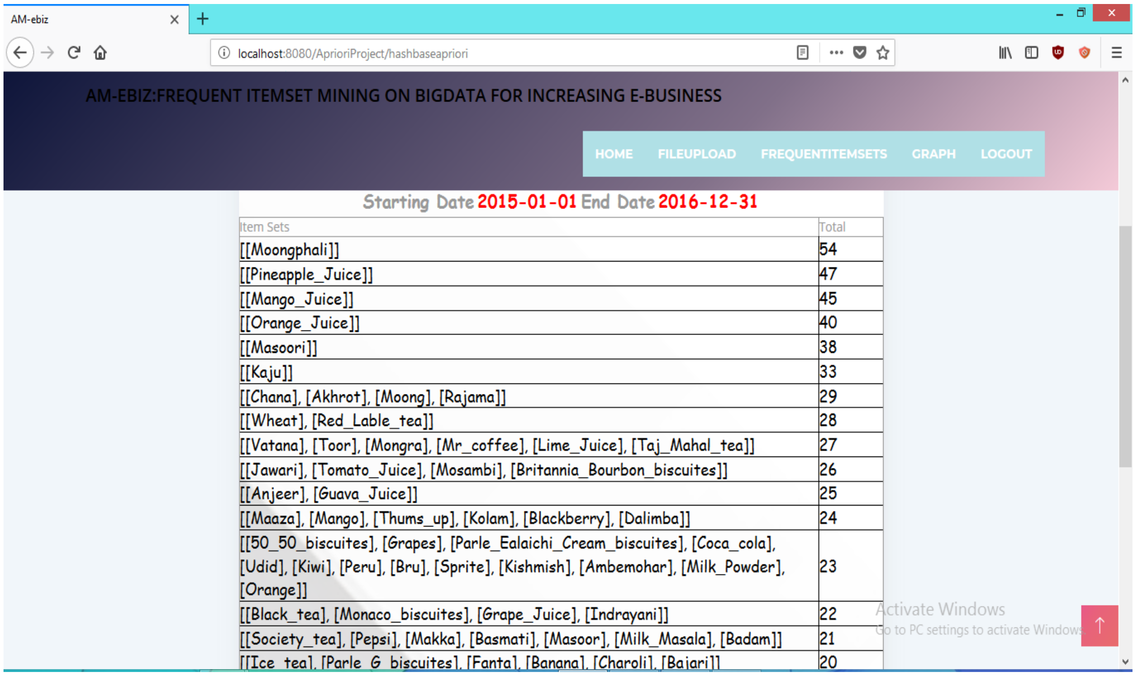


AM-EBIZ:FREQUENT ITEMSET MINING ON BIGDATA FOR INCREASING E-BUSINESS

FREQUENT ITEMSETS

Select Denominator 0.1 Starting Date 01 / 01 / 2015 Ending Date 12 / 31 / 2016

Figure 1.6: Find Frequent item set from specific Range

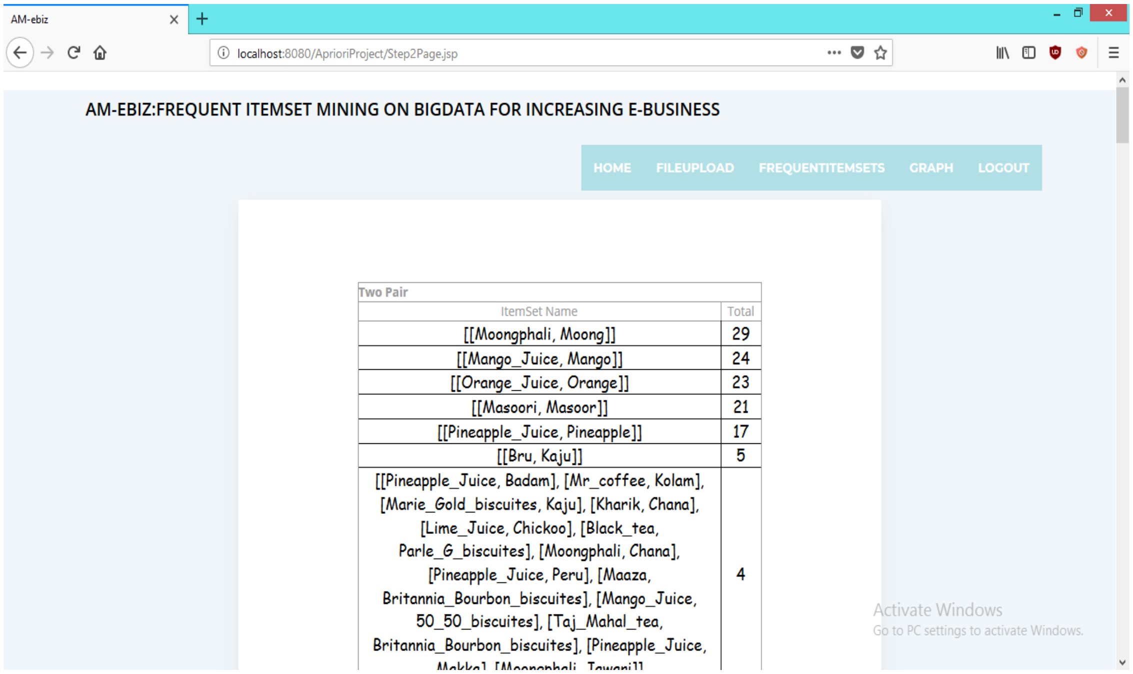

AM-ebiz
localhost:8080/AprioriProject/hashbaseapriori
HOME
FILEUPLOAD
FREQUENTITEMSETS
GRAPH
LOGOUT
Activate Windows
Go to PC settings to activate Windows

AM-EBIZ:FREQUENT ITEMSET MINING ON BIGDATA FOR INCREASING E-BUSINESS

Starting Date 2015-01-01 End Date 2016-12-31

Item Sets	Total
[[Moongphali]]	54
[[Pineapple_Juice]]	47
[[Mango_Juice]]	45
[[Orange_Juice]]	40
[[Masoori]]	38
[[Kaju]]	33
[[Chana], [Akhrot], [Moong], [Rajama]]	29
[[Wheat], [Red_Lable_tea]]	28
[[Vatana], [Toor], [Mongra], [Mr_coffee], [Lime_Juice], [Taj_Mahal_tea]]	27
[[Jawari], [Tomato_Juice], [Mosambi], [Britannia_Bourbon_biscuites]]	26
[[Anjeer], [Guava_Juice]]	25
[[Maaza], [Mango], [Thums_up], [Kolam], [Blackberry], [Dalimba]]	24
[[50_50_biscuites], [Grapes], [Parle_Ealaichi_Cream_biscuites], [Coca_cola], [Udid], [Kiwi], [Peru], [Bru], [Sprite], [Kishmish], [Ambemohar], [Milk_Powder], [Orange]]	23
[[Black_tea], [Monaco_biscuites], [Grape_Juice], [Indrayani]]	22
[[Society_tea], [Pepsi], [Makka], [Basmati], [Masoor], [Milk_Masala], [Badam]]	21
[[Ice_tea], [Parle_G_biscuites], [Fanta], [Banana], [Charoli], [Bajari]]	20

Figure 1.7: First Phase extracted item set group

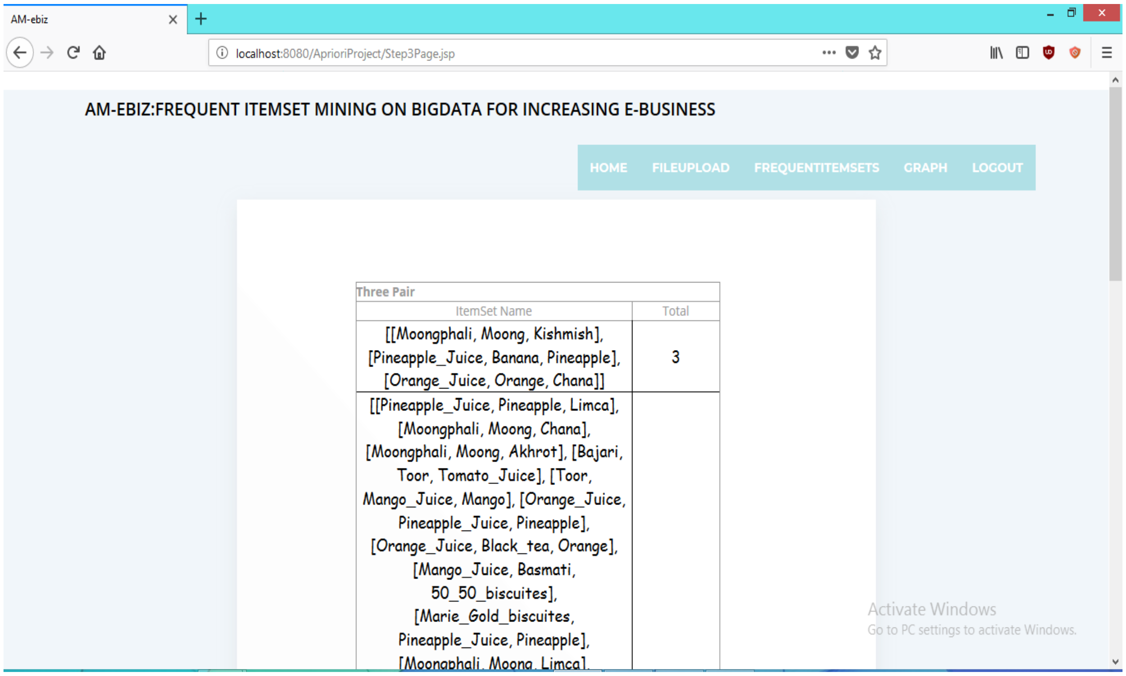

AM-ebiz
localhost:8080/AprioriProject/Step2Page.jsp
HOME
FILEUPLOAD
FREQUENTITEMSETS
GRAPH
LOGOUT
Activate Windows
Go to PC settings to activate Windows.

AM-EBIZ:FREQUENT ITEMSET MINING ON BIGDATA FOR INCREASING E-BUSINESS

Two Pair

ItemSet Name	Total
[[Moongphali, Moong]]	29
[[Mango_Juice, Mango]]	24
[[Orange_Juice, Orange]]	23
[[Masoori, Masoor]]	21
[[Pineapple_Juice, Pineapple]]	17
[[Bru, Kaju]]	5
[[Pineapple_Juice, Badam], [Mr_coffee, Kolam], [Marie_Gold_biscuites, Kaju], [Kharik, Chana], [Lime_Juice, Chickoo], [Black_tea, Parle_G_biscuites], [Moongphali, Chana], [Pineapple_Juice, Peru], [Maaza, Britannia_Bourbon_biscuites], [Mango_Juice, 50_50_biscuites], [Taj_Mahal_tea, Britannia_Bourbon_biscuites], [Pineapple_Juice, Makka], [Moongphali, Tawari]]	4

Figure 1.8: Two extracted pairs of item set

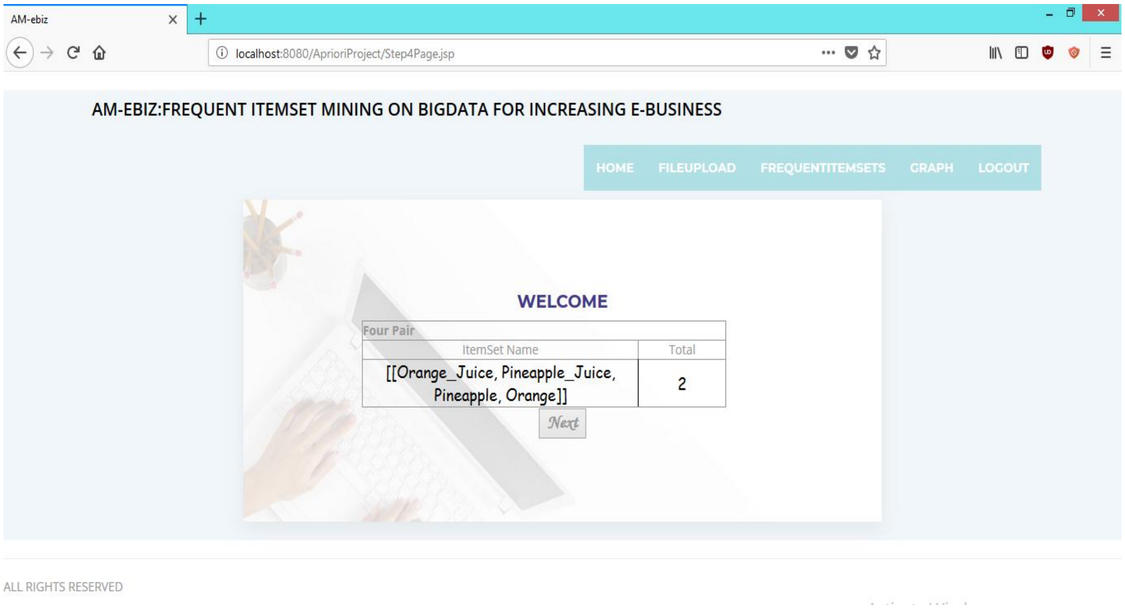

AM-ebiz
localhost:8080/AprioriProject/Step3Page.jsp
HOME
FILEUPLOAD
FREQUENTITEMSETS
GRAPH
LOGOUT
Activate Windows
Go to PC settings to activate Windows.

AM-EBIZ:FREQUENT ITEMSET MINING ON BIGDATA FOR INCREASING E-BUSINESS

Three Pair

ItemSet Name	Total
[[Moongphali, Moong, Kishmish], [Pineapple_Juice, Banana, Pineapple], [Orange_Juice, Orange, Chana]]	3
[[Pineapple_Juice, Pineapple, Limca], [Moongphali, Moong, Chana], [Moongphali, Moong, Akhrot], [Bajari, Toor, Tomato_Juice], [Toor, Mango_Juice, Mango], [Orange_Juice, Pineapple_Juice, Pineapple], [Orange_Juice, Black_tea, Orange], [Mango_Juice, Basmati, 50_50_biscuites], [Marie_Gold_biscuites, Pineapple_Juice, Pineapple], [Moongphali, Moong, Limca],	

Figure 1.9: Three extracted pairs of item set



AM-EBIZ:FREQUENT ITEMSET MINING ON BIGDATA FOR INCREASING E-BUSINESS

WELCOME

Four Pair

ItemSet Name	Total
[[Orange_Juice, Pineapple_Juice, Pineapple, Orange]]	2

Figure 1.10: Final extracted pairs of item set

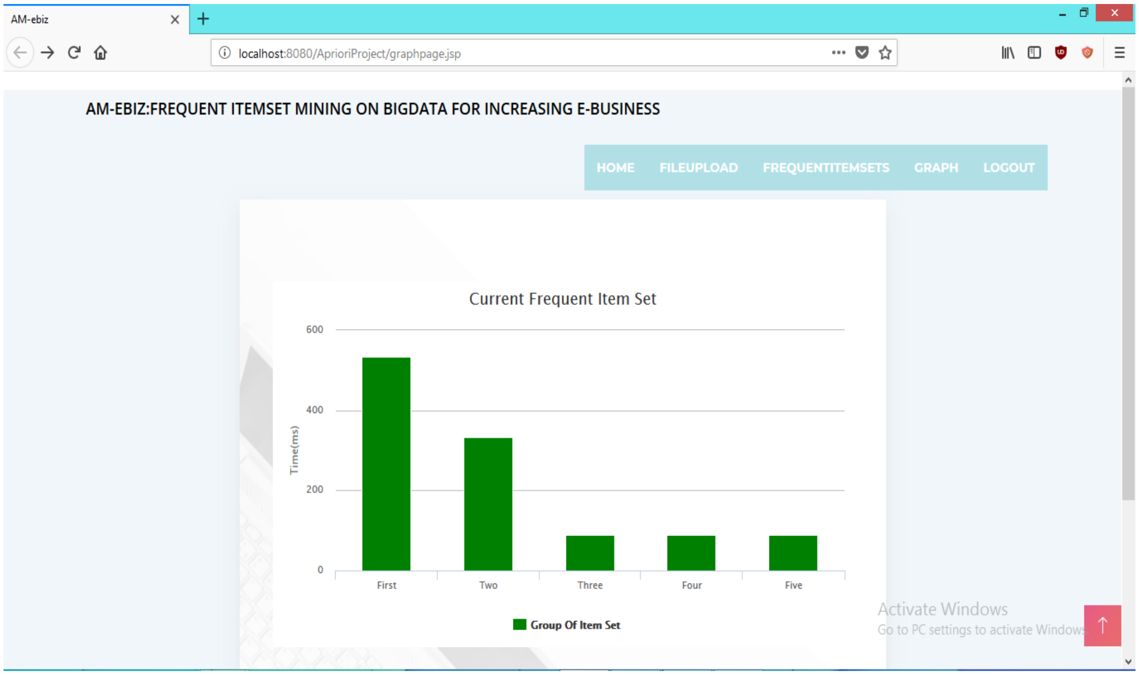

AM-ebiz
localhost:8080/AprioriProject/graphpage.jsp
AM-EBIZ:FREQUENT ITEMSET MINING ON BIGDATA FOR INCREASING E-BUSINESS
HOME
FILEUPLOAD
FREQUENTITEMSETS
GRAPH
LOGOUT
Current Frequent Item Set
Time(ms)
600
400
200
0
First
Two
Three
Four
Five
🟩 Group Of Item Set
Activate Windows
Go to PC settings to activate Windows

Figure 1.11: Time required for pair wise item set extraction

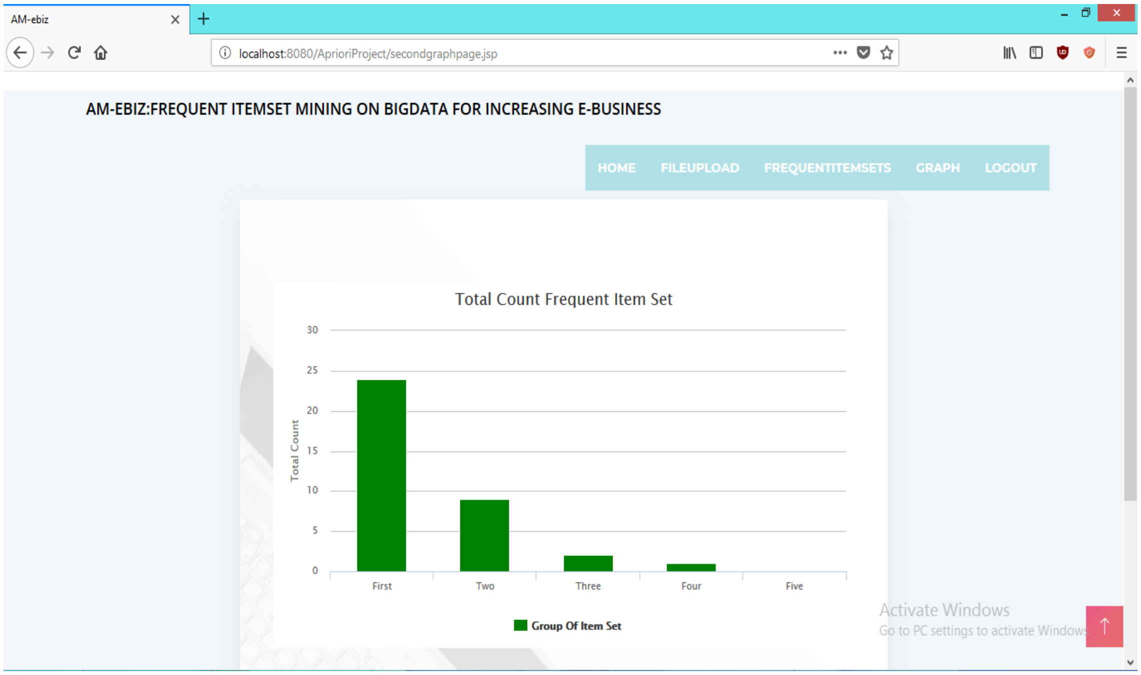

AM-ebiz
localhost:8080/AprioriProject/secondgraphpage.jsp
AM-EBIZ:FREQUENT ITEMSET MINING ON BIGDATA FOR INCREASING E-BUSINESS
HOME
FILEUPLOAD
FREQUENTITEMSETS
GRAPH
LOGOUT
Total Count Frequent Item Set
Total Count
30
25
20
15
10
5
0
First
Two
Three
Four
Five
🟩 Group Of Item Set
Activate Windows
Go to PC settings to activate Windows

Figure 1.12: Item set count of pair wise extraction

These figures show the time requirement for retrieving the frequent item set by AM_ebiz Algorithms.

Performance of AM_ebiz in terms of Time required in seconds with different support denominator with different dataset

Following table indicates the time taken in seconds by Proposed Algorithm AM_ebiz for the three different datasets of Grocery, Electronic & Sports for different support counts.

Table 1.1: Performance of Time required in seconds with different support denominator with different dataset

Support Count	**Time in Seconds**	**Time in Seconds**	**Time in Seconds**
Support 5	31	29	30
Support 8	26	24	25
Support 10	22	21	20
support 15	15	16	14

4. Conclusion

The evolution of information technology generates the back volume of data at an explosive rate in many disciplines that increases the challenge of storing and using those massive data in an efficient way. There is a constant improvement in the development of new techniques and algorithms to convert the data into knowledge. In this paper we have shown the implementation of a Novel association rule mining algorithm AM e_biz .The results shows the Performance of AM_ebiz in terms of Time required in seconds with different support denominator with different dataset for three different item set of Grocery, Electronic & sports for various support values of 5,8,10 and 15 respectively. It shows the time required to extract the frequent item set in seconds with different dataset. Utilization is very good in applying the algorithm AM_ebiz for frequent item set mining with retail item set for various support values.

References

[1] Wei Lu[1], YanyanShen[2], Su Chen[3], Beng Chin Ooi,"Efficient Processing of k Nearest Neighbor Joins using MapReduce", Proceedings of the VLDB Endowment, Vol. 5, No. 10, issued August 27th 31st 2012.

[2] Yaling Xun[1], Jifu Zhang[2], Xiao Qin[3], Senior Member, "FiDoop-DP: Data Partitioning in Frequent Itemset Mining on Hadoop Clusters", IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 28, NO. 1, JANUARY 2017

[3] RakeshAgrawal[1] Tomasz Imielinski[2]_ Arun Swami [3] : Mining Association Rules between Sets of Items in Large Databases IBM Almaden Research Center650 Harry Road, San Jose, CA 95120 2012 .

[4] Bay Vo [1],Bac Le [2:]fast algorithm for mining generalized association: International Journal of Database Theory and Application through IEEE proceedings. Vol. 3, No. 5, October 2012

[5]M.Jayakameswaraiah[1]: Design and Development of data mining system to estimate cars promotion using improved id3 algorithm,International Journal of Advanced Research in Computer and Communication Engineering, 2010

[6]Pappa**,** Gisele L,Automating The Design Of Data Mining Algorithms, 2010, XIII.

[7] Ramakrishna Shrikant [1] Rakesh Agarwal [2]: Comparison Of Various Association Rule Mining Algorithm On Frequent Itemsets in International Journal of Advanced Research in Computer Science and Software Engineering.