

Predicting Academic Course Preference using Inspired MapReduce

Nandini Patil¹, Nageshwari²

¹Professor, CSE, Sharanbasva University, Kalaburagi, India

²Student, CSE, Sharanbasva University, Kalaburagi, India

Abstract - With the emergence of new technologies, new academic trends introduced into Educational system which results in large data which is unregulated and it is also challenge for students to prefer to those academic courses which are helpful in their industrial training and increases their career prospects. Another challenge is to convert the unregulated data into structured and meaningful information there is need of Data Mining Tools. Distributed File System is used to hold large amount of data. The Files are stored in a redundant fashion across multiple machines which ensure their endurance to failure and parallel applications. Knowledge extracted using Map Reduce will be helpful in decision making for students to determine courses chosen for industrial trainings. In this paper, we are deriving preferable courses for pursuing training for students based on course combinations. Here, using DFS, tasks run over Map Reduce and output is obtained after aggregation of results.

Keywords: Data Mining Tools, Distributed File System, Map Reduce

1. INTRODUCTION

Data mining is one of the most prominent areas in modern technologies for retrieving meaningful information from huge amount of unstructured and distributed data using parallel processing of data. There is huge advantage to Educational sector of following Data Mining Techniques to analyze data input from students, feedbacks, latest academic trends etc which helps in providing quality education and decision-making approach for students to increase their career prospects and right selection of courses for industrial trainings to fulfill the skill gap pertains between primary education and industry hiring students. Data Mining has great impact in academic systems where education is weighed as primary input for societal progress.

Big data is the emerging field of data mining. It is a term for datasets that are so large or complex that traditional data processing application software is incompetent to deal with them. Big data includes gathering of data for storage and analysis purpose which gain control over operations like searching, sharing, visualization of data, query processing, updation and maintain privacy of information. In Big data, there is extremely large dataset that is analyzed computationally to reveal patterns, trends and associations. It deals with unstructured data which may include MS Office files, PDF, Text etc whereas structured data may be the relational data. Big Data approach can help colleges, institutions, universities to get a comprehensive aspect about the students. It helps in answering questions related to the learning behaviors, better understanding and curriculum trends, and future course selection for students which helps to create captivating learning experiences for students. The problem of enormously large size of dataset can be solved using Map Reduce Techniques .Map Reduce jobs run and Clusters by splitting the big data into small chunks and process the data by running it parallel on distributed clusters.

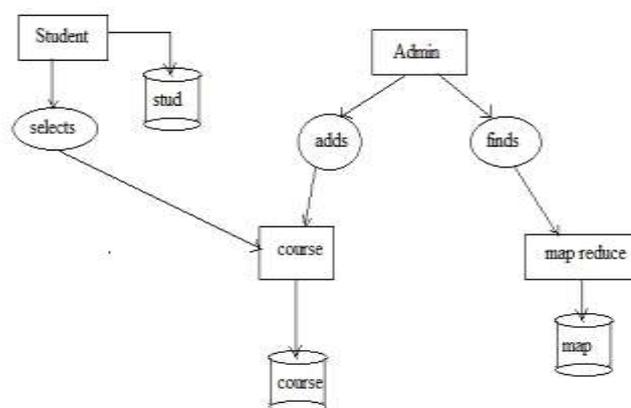


Fig. 1. Block Diagram of Industrial Training Course selection Using Map Reduce

1.1 RELATED WORK

- **Jongwook Woo, "Apriori-Map/Reduce Algorithm."** Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2012. Apriori algorithm is implemented and high performance is achieved using Map Reduce Technique of Hadoop framework to collect item sets frequently occurred in dataset.
- **Xin YueYang, Zhen Liu, Yan Fu, "MapReduce as a Programming Model for Association Rules Algorithm on Hadoop"**, Information Sciences and Interaction Sciences (ICIS), 2010 3rd International Conference on, pp. 99-102. IEEE, 2010. author has mainly addressed the challenges of using Map Reduce model for computing parallel application of Apriori.
- **Katrina Sin, Loganathan Muthu, "Application of Big Data in Education Data Mining and Learning Analytics – A Literature Review"**, ICTACT Journal on Soft Computing, ISSN: 2229-6956 (online), Vol5, Issue 4, July 2015.
- Big Data Techniques are the necessity in learning environments and present scenario with large amount of unstructured data and introduction of Massive open online courses in Education has stressed upon the need for data mining in Education.
- **B.Manjulatha, Ambica Venna, K.Soumya, "Implementation of Hadoop Operations for Big Data Processing in Educational Institutions"**, International Journal of Innovative Research in Computer and Communication Engineering, ISSN(Online) : 2320-9801, Vol. 4, Issue 4, April 2016. Tools of Data Mining like MangoDB, an open-source database and Apache Hadoop are discussed. Data Mining Techniques using these tools help students in choosing their course curriculum.

1.2 PROPOSED SYSTEM

Here, we have used, a single machine called the name node. The input dataset collected from students is shown in Table I is stored in the DFS for Map reduce.

Course Combination	Preferable Course
Java, J2EE	Android
HTML, Javascript	PHP
C, C++	Asp.Net
J2EE, Java	Android
C, C++	Java
C, C++, VB.Net	Java
C, C++	Java
HTML	PHP
HTML, Asp.net	PHP
C++	Java
-----	-----
n courses	n courses

Table I :Dataset for Course selection

The input data is then split into various clusters and provide it to the mapper that maps data to the output. The output from the mapper is represented as <key, value> pair. The output obtained from the mapper are then combined together in the combiner and then sent to the reducer.

ADVANTAGES OF PROPOSED SYSTEM

- Helps in providing quality education and decision-making approach for students to increase their career prospects
- For right selection of courses for industrial trainings to fulfil the skill gap pertains between primary education and industry hiring students
- Avoids ambiguity.
- Easy course finding for training centres.

MODULES

- **Admin**
- **Student Register**
- **Course Selection**
- **Map reducing**

2. RESULTS

- **Admin**

This module maintains course master and map reducing. Administrator Management allows the administrator to control who has access to leap within their company or organization. Administration Management allows the administrator to control the level of access each authorized person has to leap's administration and content modules. Trusted Users Groups. The System Administration Module (also called the Admin GUI) is an application that allows administrative users to manage CONNECT configuration, monitor gateway statistics and help with troubleshooting issues using CONNECT log files.

- **Student Register**

Registration Department is also interested in the documentation of processes to monitor grades and modify and add or delete courses during the regular student in the study. ... Student registration system is a structure that provides a simple set-up of programs for student enrolment.

With the advent of Information Technology in the last decade, the major focus has shifted from manual systems to computerised systems. Various systems viz. training, railway reservation, hospital management etc. involving manual work have been automated efficiently. Student course registration process in colleges involve filling registration forms manually, getting it signed by respective subject teachers, and then getting the documents acknowledged from the concerned Advisors, College Deans and Accounts Officers respectively. Finally the registration forms are submitted in the Administrative Branch. As is evident, this process is very laborious and time consuming. So an Online Student Course Registration System has been developed to simplify the current manual procedure. This module is used to register the student by entering fname, address, email, contact..etc.

- **Course Selection**

This is used to select the course by student. The student will select the course for industrial training purpose.

- **Map reducing**

This module loads the input data of the student. Then splits the data. Performs mapping, shuffling, reducing and provides the output data of the course selected by the student. Based on this result, the industrial course conducted admin can be able to find out how many number of student enroll for each course so that he can able to manage the staff for coaching.

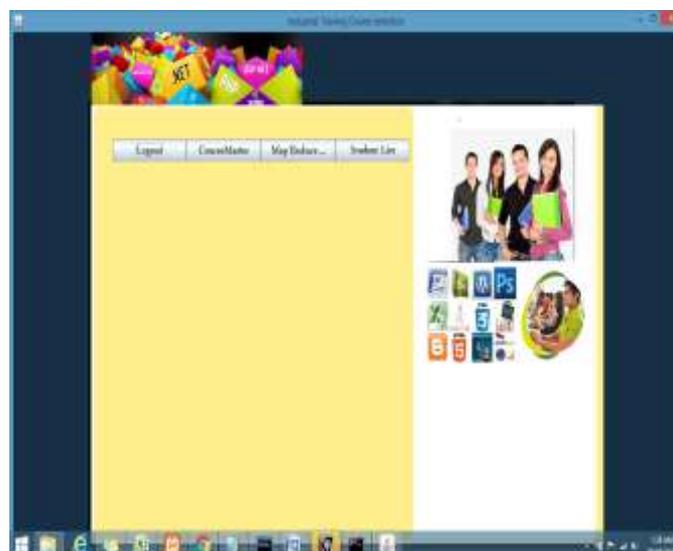
**Fig 2. Menu**



Fig 3. Split data

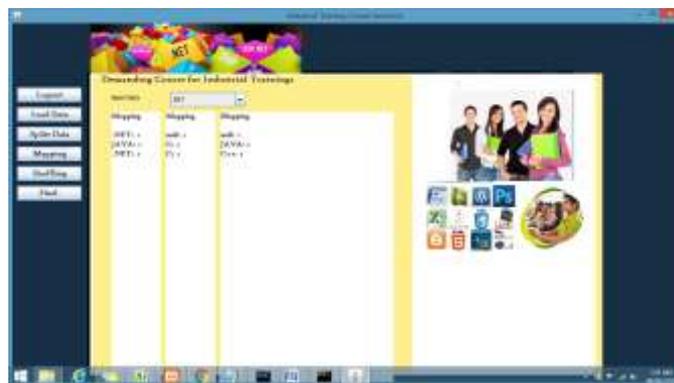


Fig 4. Mapping

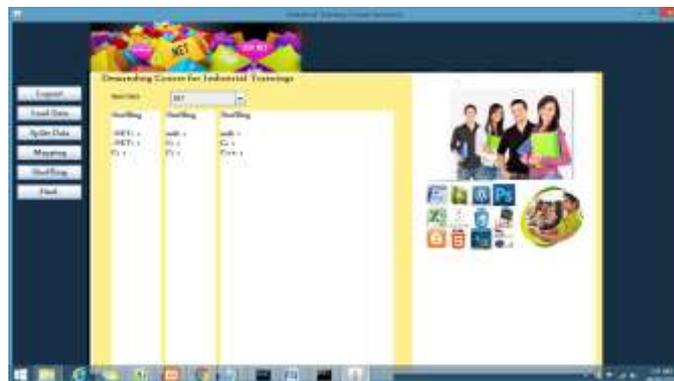


Fig.5 Shuffling



Fig.6 Final Result

3. CONCLUSION

The Map Reduce approach is used for running jobs over DFS. Using Map Reduce, the application can be scaled to run over multiple machines in a cluster and for that cluster is used. The Map Reduce Framework consists of Map and Reduce Functions with single Resource Manager which acts as a master and one Node manager which acts as slave per cluster node. The input dataset is fed into the mapper and after passing through shuffle phase, reducer displays the output after aggregating the tuples obtained from mapper and are in the form of <key, value> pair. The dataset shows that the students have opted for multiple course combinations for industrial trainings and the data becomes unstructured as well confusing for students to opt for course for trainings. The results in this paper shows that large volume of course combinations in the form of input dataset after passed through mapper function in Map Reduce Framework which runs the job in parallel on a single node cluster, it converts the data into individual tuples and the meaningful data obtained from Reducer function classifies the data of course combinations opted more by students and strengthens the decision-making of students as well institutions to prefer demanding course for Industrial trainings.

REFERENCES

- 1] Sonali Agarwal, G. N. Pandey, M. D. Tiwari, "Data Mining in Education: Data Classification and Decision Tree Approach", International Journal of e-Education, e-Business, e-Management and e-Learning, Vol. 2, No. 2, April 2012.
- [2] Jongwook Woo, "Apriori-Map/Reduce Algorithm." Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2012.
- [3] Xin YueYang, Zhen Liu, Yan Fu, "MapReduce as a Programming Model for Association Rules Algorithm on Hadoop", Information Sciences and Interaction Sciences (ICIS), 2010 3rd International Conference on, pp. 99-102. IEEE, 2010.
- [4] Katrina Sin, Loganathan Muthu, "Application of Big Data in Education Data Mining and Learning Analytics – A Literature Review", ICTACT Journal on Soft Computing, ISSN: 2229-6956 (online), Vol5, Issue 4, July 2015.
- [5] B.Manjulatha, Ambica Venna, K.Soumya, "Implementation of Hadoop Operations for Big Data Processing in Educational Institutions", International Journal of Innovative Research in Computer and Communication Engineering, ISSN(Online) : 2320-9801, Vol. 4, Issue 4, April 2016.