

# Analytical Study of Big Data Using Data Analytics in Highly Developed Education Applications

Dogbe Komlan Florient<sup>1</sup>, Ruchika Chandel<sup>2</sup>, Dr. Raman Chadha

<sup>1</sup>Research Scholar, M.Tech CSE, APG Shimla University, Shimla

<sup>2</sup>Assistant Professor, CSE department, APG Shimla University, Shimla

<sup>3</sup>dean research development, CGC Jhanjeri, Mohali

florientdogbe@gmail.com<sup>1</sup>, chandel.ruchika25@gmail.com, deanresearchdevelopment@cgc.ac.in<sup>3</sup>

**Abstract**— New improvements in the field of information advancement have changed the way or the route toward instructing. Encouraging technique has change fundamentally. With the enormous measure of information accessible on the web, works in the study hall become virtual. Under the foundation of huge information, we should choose the best showing mode in advanced education, improve the capacity of large information investigation and improve the quality instructing mode. Comprehensively, the scene of advanced education part is feeling the squeeze to change its operational and overseeing structure; to oblige new monetary, social and social plans; applicable to local, national and worldwide requests. Thus, colleges are continually scanning for significant bits of knowledge from information, to produce methodologies they can use to fulfill these new needs. Enormous Data and investigation can possibly empower foundations to altogether analyze their present difficulties, distinguish approaches to address them just as anticipate conceivable future results. Notwithstanding, in light of the fact that Big Data is another wonder in advanced education, its applied pertinence, just as the chances and constraints it would bring, is as yet obscure. This part portrays the theoretical supporting of Big Data research and shows potential open doors just as constraints related with opening the estimation of Big Data in advanced education. With the fast development of online assets, data is become over-burden during the time spent web based learning. A model for choosing training mode in advanced education dependent on large information. To improve understudy, instructor and any individual who enthusiasm of the information, the new advancements has been received in training. Right now, look at and present the examination part of huge information that can be applied in training and how we can investigate large information investigation in the fields of instruction to make expectation.

**Keywords**— Education, student, teaching, big data

## I. INTRODUCTION

Big data refers to the generation, storage and processing of large amounts of data or information. With the development and application of information technology at the high speed, big data, are influencing Education. The impact of information technology on education is becoming increasingly significant. The information-based teaching, the integration of information technology and curriculums has become the hot

focus of education. In order to meet the needs of the times better, the teaching of higher vocational courses is bound to face significant changes. [1]

This paper studies the concept of Big data in higher education and examine the analytics part apply in education. It establishes a new carrier and channel of communication between teachers and students. The system consists of online resource teaching and big data recommendation. The big data recommendation are more appropriate learning resources to users by collecting and analyzing [2]. Students are able to learn, desire to learn, predict to learn, and improve the teaching efficiency, so as to be helpful for the development of higher education. Functional implementation and performance test result show that the system implements a student centered teaching mode.

According to Edd Dumbill: “Big data is data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn’t fit the strictures of your database architectures. To gain value from this data, you must choose an alternative way to process it.”

As a catch-all term, “big data” can be pretty nebulous, in the same way that the term “cloud” covers diverse technologies. Input data to big data systems could be chatter from social networks, web server logs, traffic flow sensors, satellite imagery, broadcast audio streams, banking transactions, MP3s of rock music, the content of web pages, scans of government documents, GPS trails, telemetry from automobiles, financial market data, the list goes on. To clarify matters, there are “four V’s” that characterize this data: Volume, Velocity, Variety, and Veracity.

## II. APPLICATION OF BIG DATA IN EDUCATION

### 2.1 The Necessity of reform in Higher education

In the data age, modern information technology marked by the Internet, multimedia, cloud-computing, big data and so on is growing quickly in the world facing the fierce competition of national strength increasingly, most of countries all over the world are concerned about the important role of education reform and education.

The African continent will need to process, update and apply a new reform.

□ The impact of modern teaching from like “Online courses”

Information Technology as the main carrier, MOOC are become a new area of education. At present, teaching methods have changed significantly. The use of information technology like MOOC breaks the limitations of time and space, and learners can learn anytime and anywhere. However, traditional classroom learning is conducive to the emotional communication between teachers and students. Traditional classroom learning and network-learning have their own advantages and disadvantages.

## 2.2 MapReduce Computational Framework

MapReduce is a programming model for expressing distributed computation on massive amount of data and an execution framework for large-scale data processing on clusters of commodity servers. It was originally developed by Google and built on well known principles in parallel and distributed processing[3]. This model abstracts computation problems through two functions: map and reduce. All problems formulated in this way can be parallelized automatically.

Key-value pair frames the essential information structure in MapReduce. Keys and value qualities might be natives, for example, integers, floating point values, strings, and raw bytes or they may be arbitrary complex structures (lists, tuples, associative array, etc.). Developers ordinarily need to characterize their custom information types. The map function takes the input record and generates intermediate key and value pairs. The reduce function takes an intermediate key and a set of values to form a smaller set of values.[4]

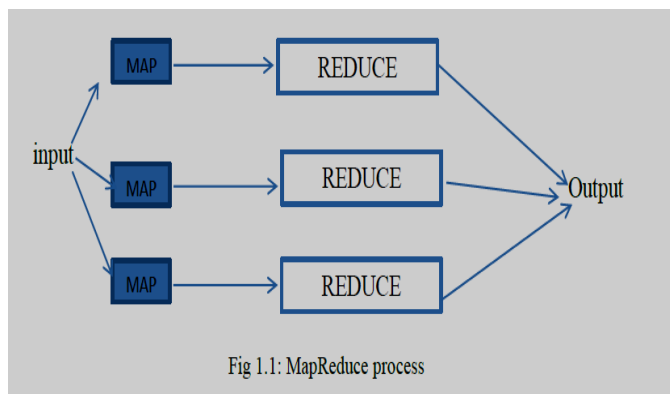


Fig 1.1: MapReduce process

## III. LITERATURE ANALYSIS

In literature, several papers presented different approaches to improve higher educational performance using big data analysis. In this chapter, we will discuss some of them:

**Deng et al.[1]** Presented guide of the background of education informationization 2.0, explores the informatized-teaching reform of business English course in higher vocational colleges to promote students' performance, so as to be helpful for the development of higher vocational education.

**Zhang et al. [2]** Integrates the functions of data acquisition, data storage, data analysis and data virtualization, and provides a friendly human-computer interaction interface to shield the details of large data analysis algorithm, to teaching

staff use large data analysis technology simply and efficiently for high concurrency and rapid query. Classification and design the basic operation model based on MapReduce for users to configure complex data processing logic freely. The system can add reusable and extensible computing and analysis module to the platform according to the field experience, to expand the data analysis and processing capabilities of the system, which facilitates users to explore iterative and incremental data.

**S.Sathya, Victor Jose. [3]** Virtual Database Technology is one of the effective solutions for integration of data from heterogeneous sources. MapReduce is a new framework specifically designed for processing huge datasets on distributed sources. Apache's Hadoop is an implementation of MapReduce. Currently Hadoop has been applied successfully for file based datasets. It is proposes to utilize the parallel and distributed processing capability of Hadoop MapReduce for handling heterogeneous query execution on large datasets. So Virtual Database Engine built on top of this will result in effective high performance distributed data integration.

**Zhao et al. [5]** Optimizes the research on the precise service technology and methods of distance learning services, realizes the individualized teaching under the support of computer and network information technology, enhances the level of distance learning service and knowledge dissemination efficiency, and promote internet plus education data service industry development.

**Vincent Koon Ong. [6]** Present the impact appear even in Higher Education sector. The strategic use and applications of big data in higher education would to better student and staff experience. Using the output from UK JISC'S BI projects and apply relevant theory for further research implications of big data in the higher education sector. Business Intelligence Programme aims to help further and higher education institutions identify their Business Intelligence (BI) maturity and build appropriate BI systems or solutions which will provide senior managers with better and more timely access to accurate data resulting in improved predictions, forecasts, benchmarking and other report in order to meet business goals.

**Zeng et al.[7]** Big data analysis model of hybrid teaching in the optimal higher vocational colleges, the information fusion of the mixed teaching mode in the optimal higher vocational colleges is carried out by using the structured big data information recombination method. Big data fusion scheduling and optimization selection of the mixed teaching mode based on the piecewise information fusion is adopted. Characteristic clustering results, the self-regression analysis of the evaluation ability of hybrid teaching in the optimal higher vocational colleges is carried out, and the test statistic model is constructed to optimize the selection of hybrid teaching model in higher vocational colleges.

**Chen et al.[8]** Discusses the curriculum design process for the "Data science & Big data technology" bachelor's degree program, and some detailed approaches to improving teaching experiments. The authors main ideas include that, a suitable graduating curriculum in big data education should be

determined as the big data talent needs are hierarchical, the redesigned curriculum in big data education should provide students more practical capabilities and knowledge, the teaching of the existing mainstream big data technologies and tools should be significant components in the syllabi of big data education.

#### IV. PROBLEM FORMULATION

Today, one of the most promising areas of application of big data technology is education. This is a small proportion of those who need big data, integrated into the educational system.

The first one is individualization. Data analysis allows you to work with individual student programs, that is, to personalize learning. Analyzing it is possible to determine which type of student interacts with which part of the content, what frequency of the given interaction, with certain studies. It is also possible to determine the weaknesses in the data provided and the advantages in them.

The second point is that the fact of changing the perception of the format of educational programs with the help of educational analytics based on data. It is important to understand that all the data used in the educational process can be not only digitized, but also dated, that is, translated into numerical data. Users move through the material with great freedom, then they analyse how the users interacted with the material, which turned out to be effective and not.

The result of such analytics should be a change in content. Therefore, the educational program is transformed from a format of approved text into a content format that dynamically changes through the analysis of data appearing as a result of interaction with students' online content. It can be assumed that the programs of training courses will also undergo changes.

The third point is a change in the approaches to monitoring and evaluating both the educational process itself and the educational result. Monitoring becomes permanent. The interest of students in continuous monitoring is due to the fact that data analysis makes the curriculum individual, the interest of teachers is related to the possibility of obtaining information about productive groups, feedback from students to the content created, when they can express an opinion whether they are interested or not, or easy, useful for a future specialty or not.

Finally, when using big data, new methods will be an integral part of educational analytics, such as forecasting, when a combination of know data will allow to predict the unknown unknown, the use of the structure detection and clustering methods, and network analysis.

#### V. PROPOSED METHODOLOGY

In this section, we will give a complete overview of our approach. A web-based service that provides the development of course contents with the use of machine learning and big data in high education. The development of personal

educational systems, the analysis of non-relational data and relational data such as access logs to students' all materials such as age, education, gender, location, course to determine new course contents, social media analysis and so on with Apache Hadoop and machine learning has been proposed. The proposed system is shown in Fig 5.1

The design of the function, we are supposed as follows:

- Map reduce function design
- Algorithm Description
- Theoretical Analysis
- Parallel implementation

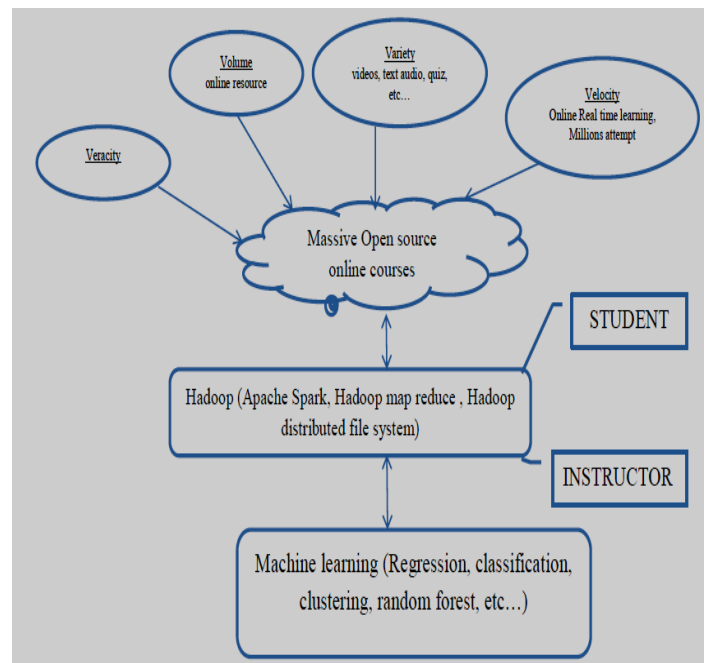


Fig 5.1 Block diagram for proposal solution

#### VI. CONCLUSION

Big Data Analytics and its relevance in Educational systems with a view of helping educational institutions adopt Big Data Analytics. Big data and the correct mechanisms for working are able to bring science and education to the new level. As a supplement and extension of traditional ideological and political education, network platform plays an increasingly important role in guiding students to establish correct values and outlook of life, by improving his performance. Moocs, one of today's most popular educational systems, comprise course materials and related learning activities such as video/text, quizzes and projects. The analysis platform can help university educators to complete scientific management of user information, online setting videos courses and interactive teaching.

## VII. REFERENCES

1. Deng, J., The Teaching Reform of Higher Vocational Business English Course Informationization on the Background of Education Informationization 2.0, in 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS). 2019. p. 324-326.
2. Zhang, T., Construction and Application of Big Data Analysis Platform for Ideological and Political Education in Colleges, in 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS). 2019. p. 249-252.
3. S.Sathya and P.M.V. Jose, Application of Hadoop MapReduce technique to Virtual Database system design. IEEE Access, 2011.
4. Djenouri, Y., et al., Frequent Itemset Mining in Big Data With Effective Single Scan Algorithms. IEEE Access, 2018. 6: p. 68013-68026.
5. Zhao, J.-c. and J.-x. Guo, Online Distance Learning Precision Service Technology Based on Big Data Analysis. International Conference on Cloud Computing and Big Data Analytics, 2019 IEEE.
6. Zan, K.K., Prospects For Using Big Data To Improve The Effectiveness Of An Education Organization. IEEE Access, 2019.
7. Ong, V.K., Big Data and Its Research Implications for Higher Education: Cases from UK Higher Education Institutions, in 2015 IIAI 4th International Congress on Advanced Applied Informatics. 2015. p. 487-491.
8. Liu, Q. and M. Zeng, Selection Model of Optimal Mixed Teaching Mode in Higher Vocational Colleges Based on Big Data, in 2018 International Conference on Virtual Reality and Intelligent Systems (ICVRIS). 2018. p. 139-142.
9. Li, X., et al., Curriculum Reform in Big Data Education at Applied Technical Colleges and Universities in China. IEEE Access, 2019. 7: p. 125511-125521.
10. Geng, W., Research on Academic Evaluation of College Students Based on Big Data. IEEE Access 2019.
11. Yunus Santur, M.K., Erhan Akin, Improving of Personal Educational Content Using MOOC. IEEE Access, 2016.
12. Xiong, Z., L. Zhi, and J. Jiang, <Research on Art Education Digital Platform Based on Big Data.pdf>. IEEE Access 2019.
13. Segooa, M.A. and B.M. Kalema, The Big potential of Big Data towards universities outcome based funding. IEEE Access, 2019.
14. Muthukrishnan, S.M., N.B.M. Yasin, and M. Govindasamy, Big Data Framework for Students' Academic REVIEW. IEEE Access 2018.
15. Rezania, D., Ali Al Hadwer, and D. Gillis, Big Data Analytics for Higher Education in The Cloud Era. IEEE Access, 2019.