

Enhancing Performance of Educational Data Using Big Data and Hadoop

¹Abdullah M. Alghamdi, ²Fahad A. Alghamdi

¹ Assistant Professor, Department of Management Information Systems,
Collage of applied studies and community Service, Imam Abdulrahman Bin Faisal University, Saudi Arabia.

² Assistant Professor, Department of Management Information Systems,
Collage of applied studies and community Service, Imam Abdulrahman Bin Faisal University, Saudi Arabia.

Abstract

Nowadays big data allows the education institutions to conduct organizational analytics and perform new business intelligence using learning management system. This data visualization helps to evaluate the performance indicators in teaching, administration and research. The real time analytics provide potential to track individuals and offer interventions to improve the learning by reshaping and personalizing learning experience. The education sector has faced several challenges in teaching effectiveness, student acquisition and student retention, ineffectiveness in storing or processing or analyzing the data. In order to enhance the performance, the study uses Hadoop platform and big data in the education sector. Further, the study uses exploratory research design to determine the challenges and needs of big data in the education sector. The study results explain that big data with Hadoop provides effective results in the education industry.

Index Terms: Big data, education, performance, Hadoop

I. INTRODUCTION

Business is highly dependent on the analytics which tend to determine hidden aspects and trends. Over a few decades, the emergence of big data analytics helps entrepreneurs to explore the data manually to carry out useful patterns in the market. Big data analytics has derived various opportunities for the institutions, policy makers, educationalists, administrators and learners. The opportunities are enhanced knowledge flow and learning success over the organization, cross collaboration over the institutions become comfortable and learning effectiveness would be enhanced, cost reduction over organizing financial performance become possible and academic risk would be lowered. Through traditional application software, big data won't be processed. Hence, it requires cloud based technologies like Hadoop and Spark to mine huge amount of data. This big data approach offer organizations with effective way to stay strong and active in the business. In addition to this, Hadoop platform has received attentions as it renders various advantages to the institutions and learners. This study aims at the influence of big data in the education and how the education system will be enhanced by using big data analytics.

II. PROBLEM STATEMENT

Nowadays big data analytics has been used in the education. Besides various opportunities the educationalist experience some challenges to deploy big data analytics. The challenges are enunciating data flow, training practitioners and decision making and actions. Retaining data for the analysis is significant challenge for the deployment of educational analytics . It is difficult to access required data from the incorporated database system and hard to create data warehouse for all institutions. Unstructured data and lack of quality can leads to essential issues. In order to create a understanding of the system among the educators, the trainees need to involve in learning the system and takes more time. It would be difficult for educators and learners to offer information in an informative way. However, the big data influence the education sector in an effective. To sort out these challenges, this study will be proposed.

III. RELATED WORK

Big data analytics provide effective assistance to the organization in order to use the data to determine new field in the business. This mining will develop new opportunities and enhance smart business. This data analytics has been resulting in profits, effective operations and customer relationship. Enterprise can able to obtain cost advantages as this cloud based analytics focus on particular issues. It is significant to notice that usage of Hadoop in organizations work faster and make effective decisions as this platform has capability to determine source of data [1]. As per the customer needs, the new products and services can be produced by using analytics. Hence companies are focusing on the needs by enhancing services to fulfill the customer needs.

A. *Shaping the education sector*

Educational institutes like universities, colleges, schools has carried huge amount of data. It can be determined to focus on which enhance the operational effectiveness of the educational institutions [3]. Student exam results and development of educational needs is highly relied on changing educational requirements will be computed by using statistical analysis.

Big data provides a way for the revolutionary system where the students can learn in innovative ways.

Education sector has adopted big data technique where the educational institutions, students and parents acquires effective advantages. It is used to evaluate the academic performance of the student over exams. Each student produce a unique data which can be evaluated for determining the student behaviour to develop learning environment [5]. Big data analytics administer the student activity like classroom performance, curricular activity interests, favourite subjects and time to complete the exam. Because of processing data driven system the educators can receive advantages from big data analytics. This data driven approach helps the institutions develop a learning experience based on student ability, learning capability and preference. Multiple programs will be predicted which will motivate person to determine what they need to learn. Many reports will be produced about their future and predict what they want to do in future. After receiving feedback from the learning experience for students the educators can enhance teaching skills [6].

B. Career prediction

Big analytics helps to determine the student performance report will enhance the authority to know about the student strength and weakness. Such report will implicates some solutions to student about the areas to be focused in future. If the student focuses on learning specific subject, the student must be motivated and choice should be recommended to prefer what they want to follow [6]. Big data is found in all different fields and render valuable information to human beings. However, the significant decision will be taken to enhance the present scenario and determine predictive big data analytics.

Educational data mining defined as evolving concept focus on the creating methods for determining type of data which obtains from the educational settings. This method is used to understand students in which they learn. Educational data mining deals with applying and researching the computerized methods to evaluate the patterns in the educational data. Due to high volume of data the patterns would be difficult to analyze. The process involves in educational data mining projects is data mining, data acquisition, pre processing and validation. These methods are obtained from data mining, psychometrics and machine learning, information visualization and modelling [3]. Educational data mining and learning analytics are similar in the aspects of processes, data sources, data types and objectives. Educational data mining and learning analytics are quietly differ in the approaches in the process of applying data analysis. It also emphasize the application of data mining approaches which includes association rule, classification, clustering. These approaches support students and teachers in learning and analyzing the process [7].

C. Learning analytics

Learning analytics has received enhanced attention as it provides various advantages to the institutions of higher education and enhance student retention, student success and provide accountability. Learning analytics focus on managing the capacity of the analytics such as acting on predictions and forecast behaviour. The objective of learning analytics is to enhance the prediction over time. It allows schools and teachers to utilize the educational opportunities to the level of needs. It enhances the efficiency of approaches in interpretation, mining to enhance the understanding of learning and teaching [8]. It tends to enhance the education to each student effectively. It aims at managing large amount of data generated by the students in academic activities. This learning analytics focus on student success in their education. It is the phenomenon of collecting, analysing and reporting the data about learners in terms of understanding the learning.

An increased development of big data analytics becomes significant for the organization to manage the measurement and management processes [9]. The educations sector has become significant to determine the data for the enhancement of academic and learning activities. Learning and academic analytics are the process of analytics in the education sector. This learning analytics is the process of collecting data about the learners and tends to learn more effective. It deals with enhancing the learner success.

Academic analytics aims at development of the processes, resources and workflows of the institution over use of academic, learner and institutional data. It also focuses on enhancing the organizational effectiveness [7]. Due to enhance accreditation, competition, regulations the adoption of the academic analytics is increased. In the education sector, a large amount of data has been collected but they are not analyzed properly. This higher education leadership is highly focused on significant and critical decision as per insufficient information when that could be attained by analyzing the data. The business analytics, predictive analytics and action analytics are taken into consideration of the big data analytics [11]. In the academic domain, learning analytics focused on learning processes, learners and learning behaviours.

D. Slow progress of big data in education

In the education sector, big data is considered as the game changer in the academic performance. The learning company like Knewton is helping this analytics in an efficient way. The company has increased \$157 million dollars to develop an adaptive learning system which suggest the learning paths for the students over grade and subject levels [10]. There are some barriers preventing the adoption of these solutions. Teachers and parents have increased student privacy and fearing that data could be misused. The other challenges like difficult approach of deployment and working in tight deadlines are enhanced.

The system which implicates products to buy and suggests what to learn is difficult. If the student course will be monitored, the offline activities like homework and classroom discussions are complex to gather. Big data has capability to enhance academic performance and the machine learning need to enhance the personal learning path for each student.

IV. METHODOLOGY

Qualitative research has been applied to this study. The qualitative research is used to acquire the knowledge of people reasons, opinions and motivations. It offer extensive view of the problem and provide ideas for making quantitative research. It also determines the thoughts and opinions about the problem[2]. However, this qualitative research will add perceptions to the existing problem statement. Such research is called exploratory research.

Data collection:

In order to evaluate the results, the researcher may use primary method and secondary method.

Primary method:

Primary method of the research is used to gather data by performing interviews, surveys and observations. By doing these methods, the researcher will experience confidentiality issues. It also aims at market research, experimenting and observations. However, these methods are not applied to this study.

Secondary method:

Secondary is the approach of gathering data from the sources like articles, magazines, newspaper, and journals. This method prefers existing sources of distinct researchers. This method has been applied to this research for data collection. The data has collected from the literary sources of researchers and introduce a effective solutions for education using big data analytics.

RESEARCH DESIGN:

In this research, the exploratory research design has been applied. Exploratory research can able to determine the research questions and do not offer final solutions to the existing issues [2]. In order to determine the problem, the research designs offer a effective understanding of problem. From this research, the researcher can able to decide their decisions as per the findings of new insights and new data. This research explores problems and offer new solutions when there is no existing research has made.”

V. ANALYSIS AND FINDINGS

The study findings revealed that NoSQL databases are highly scalable and schema less. It provides great effective support to the framework such as Dryad and MapReduce for accessing huge amount of data. It manages the data mapped on the

distributed file system and intermediate data stored on local disks. Nowadays some platform like Apache Hadoop, Big table, HBASE, Mongo DB has increased to manage huge amount of data. Mongo DB is used to store the educational data. Because of scalability Mongo DB is widely used in the educational platform [15]. Similarly, Hadoop platforms are used in the education sector to manage huge amount of data in an efficient way. In order to enhance the process, the education sector use batch and stream processing for storing the data. Hadoop is the platform which allows the distributed processing of data using single programming model. The characteristics of Hadoop platform are reliable, scalable, flexible and economical [16].

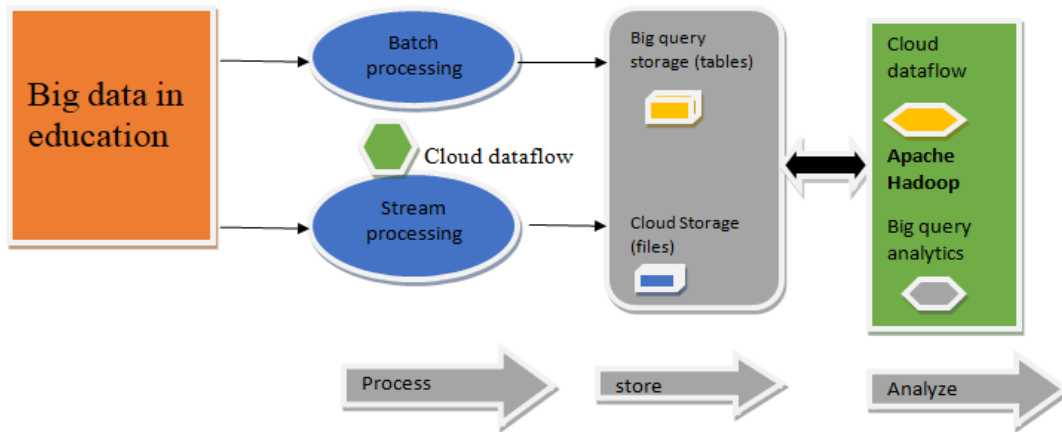


Source: Author

Hadoop is the modern big data processing framework used for extensive storage and effective scalability[11].

VI. PROPOSED APPROACHES OF BIG DATA TECHNIQUE IN EDUCATION

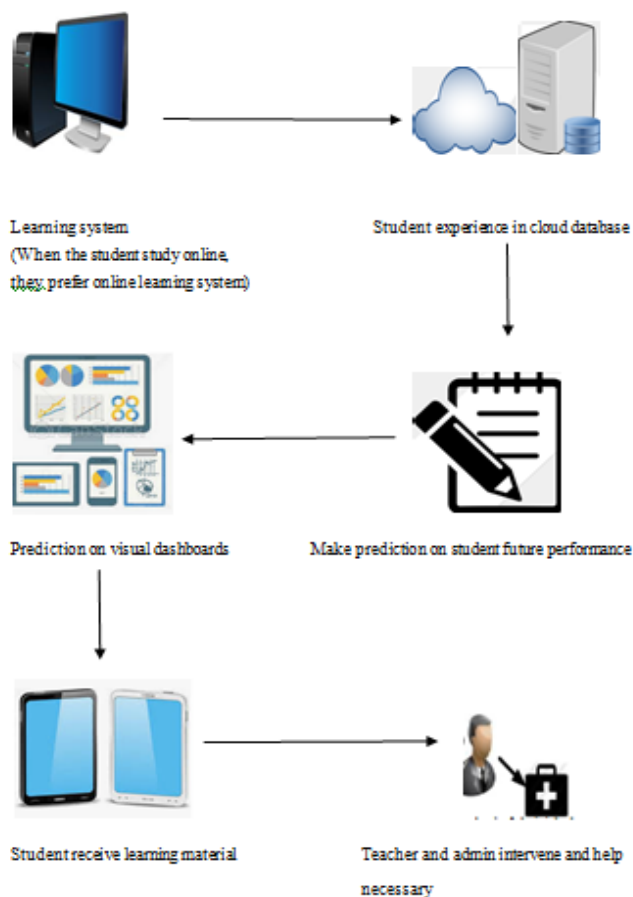
The proposed approach of big data technique in education is implemented Apache Hadoop in the system process. This proposed approach use batch processing and stream processing in processing the educational data. Batch processing is the approach of evaluating the blocks of data over a certain period of time. Hadoop MapReduce is the effective platform for processing the data in batches. In the educational sectors, the stream of data has divided into blocks of data and performs the function by using Hadoop platform[13]. The stream processing tends to process the data in real time as they determine the condition in small amount of time. It enables to give data into analytics tools and receive analytics results[14]. These two types of processing are used to enhance the performance of the educational data. With the help of cloud data flow, this process will be accessed efficiently.



Source: Author

A. Storing

In order to store the data, our study uses cloud storage to store the files. The educational data can be stored in the cloud in forms of files and tables. The tables are stored in big query storage and files are stored in cloud storage. The below diagram will represent the storage process of extracting knowledge of big data from cloud.



Source: Author

The students prefer the online learning system and the cloud will help to store the data in an efficient way. The above process will be accessed in the education system to store the data.

B. Analysing

The big data Hadoop used in education industry for accountable, economical and reliable purpose. There are some challenges in the education sectors such as enhancing evaluation and monitoring, creating industry ready education system, getting trained teachers for enhancing quality education, making the system more accountable and understanding the industry demand. By having these challenges in the education sector, Hadoop has the potential to overcome these challenges with the help of big data.

Student acquisition optimization:

We can find unemployment in many countries. If we ask the human resource department, they will answer that they are not finding the right candidates. Through campus drive, they can find hundreds of candidates and fill the vacancies. Many companies state that employers are experiencing issues in getting the right resources. With the help of the Hadoop system, the industry may conduct sentiment analysis on the students who are involved in the drive. Many students are focusing on different domains as they are studying. They are paying more attention towards the engineering stream. As a result of this intention, it may lead to depressions and unemployment. By having a Hadoop solution, they must do analysis before they can select any branch for study. It can be determined by posts and shares on social media data. They must take feedback from the teachers and need to know their interest.

Teaching effectiveness:

Many instructors are not able to give their best in teaching. It is significant to determine that instructors and enhance their performance and reward them. Inadequate teaching effectiveness is the expressed sentiment about the instructor will allow the educational institutes to take necessary steps.

The institution may reward the instructors who have successful students. With the help of big data, the educational institutes determine the trends in the industry

Student retention:

Many educational institutions could not able to retain the students for longer. The students leave the educational institution early and the problem exists within the colleges, schools and institutes. The educational institutes should determine the posts and analyze the trends and feedback to serve the student better. The significant challenges in the education are to evaluate how the data is stored, processes, used for the future outcome. An essential issue experienced by the education institutions is the amount of student transfer or dropouts. The challenge in high education sector has been restricted to business intelligence and students can make decisions. This become ineffective in enhancing the student retention and affects the student performance.

VII. RECOMMENDATIONS

The usage of big data in the education is increased with privacy and security concerns. As the big data focused on digitalizing data, there is no roadblocks on how to process, store and access the student learning data when preventing such data from being misused or abused. The student learning data are stored and collected in online learning system, mobile devices and school district offices. The disconnection among these aspects leads to data security as the influence of data security breach in one database will affects the whole data system. It may also cause blocking the linkage of various databases. In order to make a significant balance between sharing and securing data, the data security protocols makes the linkage and reduce insignificant data facts. The other significant challenges in implementation of big data in education are the private information and student learning data could be used against the students even they move in educational system or workplace. It allows the students to learn by offering learning materials which are enhanced by the adaptive learning algorithms and the academic performance of the students are digitally tracked. In order to resolve these problems, our proposed system will be used. The education sector may use this Hadoop platform to enhance the reliability and scalability process.

VIII. CONCLUSIONS

This study clearly highlights how big data influence the education sector. There are some challenges in the education sectors such as data privacy, data security, ineffective decisions, unable to capture or access or store the data. These issues can be solved using Hadoop framework and big data. It also uses batch and stream processing for processing the data. cloud storage is used for storing the data and Hadoop platform for analyzing the educational data in an efficient manner. By using this framework, the education sector can enhance the student retention, improve teaching effectiveness, transform into effective decision thinking and actions and student acquisition optimization. For future research, the practitioner

can integrate Spark and Hadoop for enhancing the performance of education sector.

REFERENCES

- [1] A. Picciano, "The Evolution of Big Data and Learning Analytics in American Higher Education", *Online Learning*, vol. 16, no. 3, 2012.
- [2] B. D'Ippolito, "An Exploratory Review of the Design Literature: Gaps and Avenues for Future Research", *SSRN Electronic Journal*, 2012.
- [3] D. Dommmnguez Figaredo, "Big Data, analiitica del aprendizaje y educacion basada en datos (Big Data, Learning Analytics & Data-driven Education)", *SSRN Electronic Journal*, 2018.
- [4] K. Madhavan and M. Richey, "Problems in Big Data Analytics in Learning", *Journal of Engineering Education*, vol. 105, no. 1, pp. 6-14, 2015.
- [5] B. Daniel, "Big Data and analytics in higher education: Opportunities and challenges", *British Journal of Educational Technology*, vol. 46, no. 5, pp. 904-920, 2014.
- [6] T. Brock, "Performance Analytics: The Missing Big Data Link Between Learning Analytics and Business Analytics", *Performance Improvement*, vol. 56, no. 7, pp. 6-16, 2017.
- [7] S. Bajpai and S. Mani, "Big Data in Education and Learning Analytics", *TechnoLearn: An International Journal of Educational Technology*, vol. 7, no. 12, p. 45, 2017.
- [8] R. Ellaway, M. Pusic, R. Galbraith and T. Cameron, "Developing the role of big data and analytics in health professional education", *Medical Teacher*, vol. 36, no. 3, pp. 216-222, 2014.
- [9] Y. Wang, "Big Opportunities and Big Concerns of Big Data in Education", *TechTrends*, vol. 60, no. 4, pp. 381-384, 2016.
- [10] C. Jurkiewicz, "Big Data, Big Concerns: Ethics in the Digital Age", *Public Integrity*, vol. 20, no. 1, pp. S46-S59, 2018.
- [11] R. Desai, "Real Time Analysis using Hadoop", *International Journal Of Engineering And Computer Science*, 2016.
- [12] R. De Rosa and B. Aragona, "Unpacking Big Data in Education. A Research Framework", *Statistics, Politics and Policy*, vol. 8, no. 2, pp. 123-137, 2017.
- [13] S. Shahrivari, "Beyond Batch Processing: Towards Real-Time and Streaming Big Data", *Computers*, vol. 3, no. 4, pp. 117-129, 2014.
- [14] D. Sun, G. Fu, X. Liu and D. Ren, "Understanding data flow graph for improving big data stream computing environments", *International Journal of Computing Science and Mathematics*, vol. 5, no. 4, p. 394, 2014.

- [15] M. Goudarzi, "Heterogeneous Architectures for Big Data Batch Processing in MapReduce Paradigm", *IEEE Transactions on Big Data*, pp. 1-1, 2017.
- [16] J. Reyes, "The skinny on big data in education: Learning analytics simplified", *TechTrends*, vol. 59, no. 2, pp. 75-80, 2015.