International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.7 March - April – 2023
DOI:10.46647/ijetms.2023.v07i02.096 ISSN: 2581-4621

BIG DATA FRAMEWORK FOR EDUCATIONAL ANALYSIS

Najna Nazir M K[a], Ambili Antharjanam[b]

[a]*PG Scholar, Dept. of Computer Science and Engineering, M.Dasan Institute of technology Ulliyeri, Kozhikode, 673620 ,*

[b]*Assistant Professor Dept. of Computer Science and Engineering M.Dasan Institute of technology Ulliyeri Kozhikode 673620*

Abstract

Huge amounts of educational data are being produced, and a common challenge that many educational organizations confront,is finding an effective method to harness and analyze this data for continuously delivering enhanced education. Nowadays, the educational data is evolving and has become large in volume, wide in variety and high in velocity. This produced data needs to behandled in an efficient manner to extract value and make informed decisions. For that, the proposed system confronts such dataas a big data challenge and presents a comprehensive platform tailored to perform educational big data analytical applications. It presents an effective environment for non-data scientists and people in the educational sector to apply their demanding educational big data applications.

Keywords: Big Data Framework,, Educational Big Data Analytics, Data Visualization.

1. INTRODUCTION

Massive complex educational related data is being producedand with proper management, immense knowledge can be extracted. The desire to enhance the quality of education is continuous. The application of big data analytics in education helps the educators to track the performance of students. Early indications regarding students' progress help aca- demics to optimise their learning strategies and can focus on diverse educational practices to make the learning experience successfully. It can analyse the expected weaknesses in learning processes and as a result the faculties can proactively engage such students in better learning experience. Analysis can also be performed to identify the significant and impact ofstudent background.

There are numerous challenges in handling educational data efficiently. In general, these challenges can be categorized into technical and organizational challenges. The technical challenges can be summed into four main challenges: The capability of the infrastructure at the universities premises to process the produced data, The ability to monitor the effectof the made decisions, The absence of a comprehensive platform tailored for educational organizations for gathering and analyzing the produced data effectively, Deploying and using technologies that handle educational data require skilledand talented practitioners.

In order to tackle those challenges, the characteristics of educational data must be considered to find appropriate tools that may assist in the gathering and analyzing of the produced data. The educational data is: Produced at large volumes, The educational data produced varies in type, (the produced data could be in a structured or unstructured form), The rate of educational data produced is high in velocity, The produced data needs to be handle in an appropriate manner to extract value.

From those aforementioned characteristics, educational datacan be considered as a big data challenge. The characteristics of big data is having data that incorporates the 4V's, i.e., volume, variety, velocity and value, which matches the nature of educational data. While handling big data requires costly infrastructure and expertise, there has been much progressusing big data analytics in business and industry sectors for enhanced working, effectiveness and informed decision making. The constructing a big data platform for a specific area can be a complex task due to the lack of rigidity in the produced data. This makes it more difficult to precisely definewhat the constructed big data platform will achieve. This applies to communication interfaces, communication with other applications and computation on various types of data. Accordingly, constructing and big data

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.7 March - April – 2023
DOI:10.46647/ijetms.2023.v07i02.096 ISSN: 2581-4621

platform for a specific sector is data and application dependant.
In education, numerous research has been conducted to im- prove the quality of education in many aspects. In predicting student academic success, predicting the students' final grades, and course recommendation. As the volume, variety and ve- locity of educational data is increasing, there has been growinginterest in the education community to utilize this produced bigdata for improving the educational outcomes in many aspects including enhancing the learning performance of students, en- hancing the working effectiveness of instructors and reducing administrative workload. Big data is being considered a revo- lutionary significance to education, and educational big data isbecoming of research interest. Many works highlight the chal-lenges and opportunities of big data analytics for education.

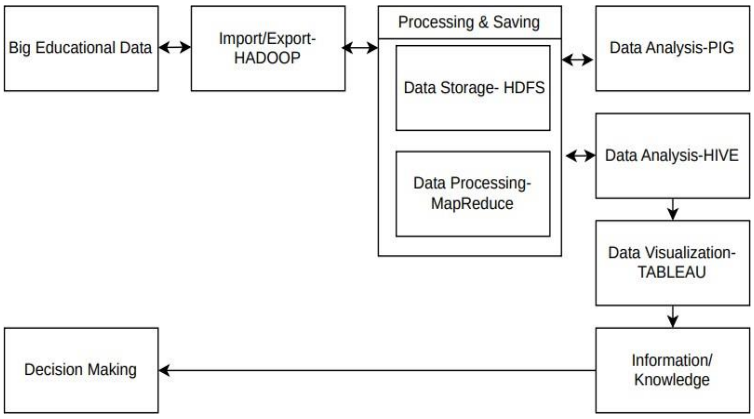

Figure 1: Proposed Architecture

2. EDUCATIONAL BIG DATA FRAMEWORK

Data is collection of facts. Data analysis is the collec- tion,transformation,and organization of data in order to draw conclusions,make predictions,and drive informed decision making. There are 6 phases in data analysis: Ask, Prepare, Process, Analyze, Share and Act. In the first phase objective and problem is defined.Then prepare data by collecting and storing the information. There are different ways to collect data.They are-Surveys,Interviews,Observations,Forms, Ques- tionnaires and Cookies.After prepare data,process data by cleaning and checking the data integrity.Then analyze datato find variety of problems. Six common types are: Making Predictions, Categorizing things, Spotting something unusual, Identifying themes, Discovering connections and Finding patterns. After performing analysis share data made by the analysis for communicating and interpreting result. Here data visualization play a major role. Finally act on data, putting insights to work and to solve the problem.
The implementation of educational big data platform is per- formed using Hadoop framework. Hadoop is a framework used to store, process and analyse big data. It has the abil-ity to distribute data storage, parallel processing, and process data at higher volume, higher velocity, variety, value, and ve- racity. The three major components Hadoop framework are HDFS, MapReduce, and YARN . HDFS uses name nodes anddata nodes to store extensive data. MapReduce manages these nodes for processing, and YARN acts as an Operating system for Hadoop in managing cluster resources. The implementationof the educational big data platform goes through the followingstages:

2.1. DATA PRODUCTION

At this stage, the data sources and produced data are speci- fied. This includes data such as, present and past student demo-graphic data, projects and research data, building's information, laboratory needs, courses data, multimedia data, events and so- cial media related data. This data includes structured, semi- structured and unstructured data that is growing at a vast rate. Thus, the types of data and data sources are specified. This data

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.7 March - April – 2023
DOI:10.46647/ijetms.2023.v07i02.096 ISSN: 2581-4621

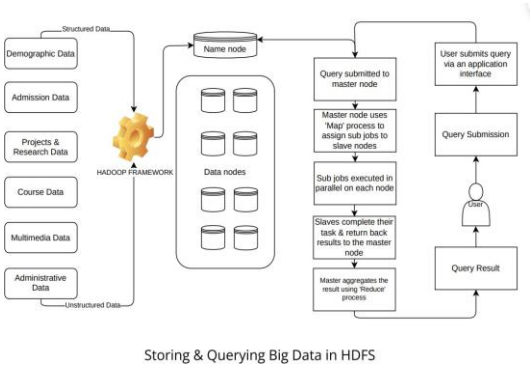

Figure 2: Basic Hadoop Principles

produced from a variety of data sources needs to be stored and organized in the Hadoop framework. This allows the related data to be sent from anywhere to the framework via a network connection.

2.2. . PROCESSING OF EDUCATIONAL BIG DATA

Data storage in Hadoop framework is performed using HDFS. It store Big Data in a distributed manner. It has the ability to write or read terabytes of data per second and offers comprehensive support for huge files. It also provides very high fault-tolerance and high availability to the storage layer.

HDFS consists of master and worker nodes for processing data.They are: Name node, Secondary Name node, Data node,Node Manager and Resource Manager.

Name node: There is only one name node.It also known as Master node.This name node maintains metadata of all data node.

Secondary Name node: It is also a master node.It is responsi-ble to create checkpoint of metadata maintained by name node.**Data node:** There are multiple data nodes. It also known as slave node or worker node.The data is stored like blocks of data.

Node Manager: Node Manager runs on the slave daemons and are responsible for the execution of a task on every single data node.

Resource Manager: Resource Manager runs on a master daemon and manages the resource allocation in the cluster.

Once the data is stored and organized in the Hadoop frame- work, analytical tools to process and analyze data are possible.Hadoop, which is a software framework that stores and pro- cesses huge amounts of data. It relays on distributed clusters of commodity servers for storing and processing data. This allows to processes huge amounts of data and to perform batchdata analytical processing. The Hadoop framework is setup forstoring and processing the educational big data.

MapReduce is the primary processing engine of Hadoop Twophases of MapReduce are Mapping and Reducing. Mapping classifies data into nodes, and The Reducer generates the final on. HiveQL allows multiple users to query data simultaneously.

User writing queries in the HiveQL language, is converted into MapReduce tasks. Then, the data is processed and ana- lyzed. HiveQL works on structured data, such as numbers, ad-dresses, dates, names, and so on. HiveQL allows multiple usersto query data simultaneously.

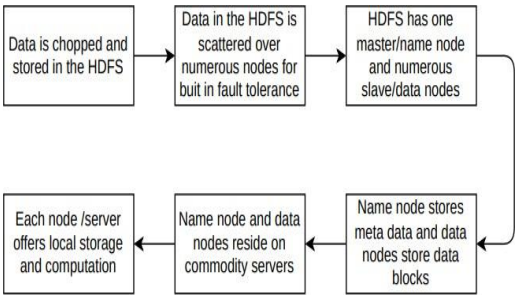

Figure 3: Storing and Querying Big Data in HDFS

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.7 March - April – 2023
DOI:10.46647/ijetms.2023.v07i02.096 ISSN: 2581-4621

product by aggregating and reducing the output. It can process and compute significantly large volumes of data.

2.3. *EDUCATIONAL BIG DATA QUERYING*

At this stage, the data has been stored in the Hadoop frame-work are ready to receive workloads to perform on data. Query-ing the stored data can be accomplished by utilizing querying components. Two querying components namely, Pig and Hivethat enable batch processing of educational data are utilized. Each component differs in the way it queries data. Hive and Pig are the two integral parts of the Hadoop ecosystem, both ofwhich enable the processing and analyzing of large data sets. Pig is utilized for batch processing workloads, whereas, Hive is utilized for low-latency processing workloads. Once the data isprocessed, it can be stored back into the HDFS.

PIG: Pig is the leading scripting platform to process and analyze Big Datasets. It can use structured and unstructured data to get actionable insights and then stores the result in HDFS. Pig has two essential components; first, a Pig Latin script language along with a runtime engine to process and analyze MapReduce programs. Pig operates in three stages first by loading data and writing script, then Pig operations, andthen execution of the plan. Pig is extensible, self-optimizing, and easily programmed. It provides a high-level scripting language- Pig Latin. Programmers can use Pig to write data transformations without knowing Java.

HIVE: Hive is a data warehouse system used to queryand analyze large datasets stored in HDFS. Acting as a Data warehouse software, Hive uses SQL like language, HiveQL, for querying through distributed databases. There are mainly two Hive data types; first, as Primitive data types with numeric, string, date/time, and miscellaneous data types, and secondary Complex data types include arrays, maps, structs, and units. Hive architecture first performs a compiler for checking and an- alyzing, then optimizes with MapReduce and HDFS tasks and executors to accomplish the query. User writing queries in the HiveQL language, is converted into MapReduce tasks. Then, the data is processed and analyzed. HiveQL works on struc- tured data, such as numbers, addresses, dates, names, and so.

2.4. *EDUCATIONAL BIG DATA ANALYTICS*

The data analytics stage is where useful information and insights are extracted from the educational data that was ingested previously into HDFS. This includes batch and real- time applications that promote the operation of the educational organization. The components that allow batch and real-time applications are setup and can be run on top of the querying layer or directly on data stored in the database. The component utilized is Tableau Desktop. It can cover applications including, querying, cleaning, data mining, statistics, and visual analytics.

TABLEAU: Tableau is an excellent data visualization tool used for reporting and analyzing vast volumes of data. It helpsusers create different charts, graphs, maps, dashboards, and sto-ries for visualizing and analyzing data, to help in making de- cisions. It is very secure and highly available. It can run on both the physical machines and virtual machines. It is a multi- process, multi-user, and multi-threaded system. It simplifies raw data in a very easily understandable format. It also allows non-technical users to create customized dashboards. Tableau provides Data Blending , Real time analysis and Collaboration of data.

3. RESULTS AND DISCUSSION

The data provided through learning analytics can improve productivity and efficiency and allow to respond effectivelyto any educational challenges. This ensures optimal use of resources and funding to improve outcomes. Currently, student success is a major challenge for institutions. Though institutions are looking for new ways to offer the right supportto the right student at the right moment. Student Analytics offers new ways to improve student success. Through better insight into students' needs and behavior, Student Analytics helps institutions develop and implement personalized and proactive services.

Any kinds of analysis based on the context can be performedon the dataset, which leads to generate insights. Modelling stu-dent success is possible by relating available information to fu-ture student

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.7 March - April – 2023
DOI:10.46647/ijetms.2023.v07i02.096 ISSN: 2581-4621

success measures such as retention and grade rates.It provide better insights into students based on Background information, Motivation and engagement,Tests and Academic results.
The graphs in figure 6.1 shows that internet usage in school GP is higher than the school MS. It leads to the increase in fail-ure rate in the school GP.
The graphs in figure 6.2 shows that the failure rate of students

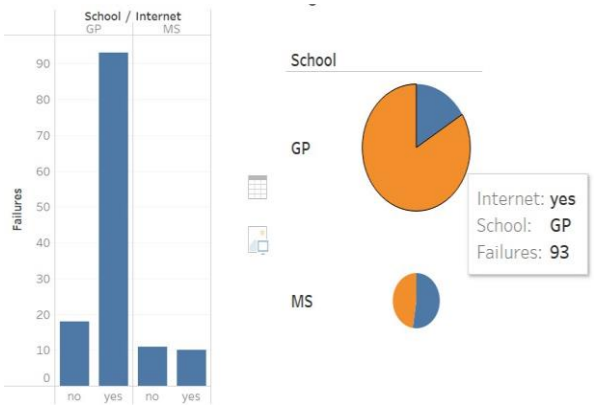

Figure 4: Internet usage vs Failures in school

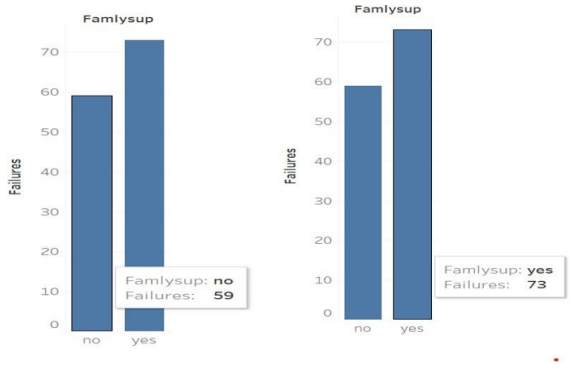

Figure 5: Family support vs Failures

having family support is higher than students having no familysupport.
Similarly the School support for the students is directly propor-tional to their success.

4. CONCLUSION

The proposed system presents a comprehensive platform for educational big data analytics.The proposed system presents a comprehensive platform for educational big data analytics. Theobjective of the platform was to handle complex educational big data while considering four main contributions to field: Aninfrastructure that is able to handle the educational big data from data production to analytics. Presenting a comprehensiveplatform to perform various educational data analytical applications. Including a feedback loop to monitor the short-term and longterm effects of the decision-making process. Introduc-ing an effective environment for non-data scientists and peoplein the educational organizations to apply their demanding edu-cational applications. The decision makers at the organization can take corrective actions to enhance the educational outputs.The presented educational big data platform features a feed- back loop to monitor the effects of the decisions taken.

DOI:10.46647/ijetms.2023.v07i02.096 ISSN: 2581-4621

References

[1]E. Kurilovas, , "Advanced machine learning approaches to personalise learning: Learning analytics and decision making," *Behaviour Inf. Technol.,vol. 38, no. 4, pp. 410–421, Apr. 2019.*

[2]C. Romero and S. Ventura , "Educational data mining and learning ana- lytics: An updated survey," *WIREs Data Mining Knowl. Discovery, vol. 10,no. 3, May 2020, Art. no. e1355*

[3]]U. K. Saba, S. U. Islam, H. Ijaz, J. J. P. C. Rodrigues, A. Gani, and K. Munir, "Planning fog networks for time-critical IoT requests", *Comput. Commun., vol. 172, pp. 75–83, Apr. 2021.*

[4]G.-E. Zaharia, T.-A.-I. Sosea, R.-I. Ciobanu, and C. Dobre,"Machine learning-based traffic offloading in fog networks," *Simul. Model. Pract. The- ory, vol. 101, May 2020, Art. no. 102045.*

[5]A. Shakarami, M. Ghobaei-Arani, and A. Shahidinejad, "A survey on the computation offloading approaches in mobile edge computing: A machine learning-based perspective",*Comput. Netw., vol. 182, Dec. 2020, Art. no. 107496.*

[6]C. Dede, A. Ho, and P. Atan "Big data analysis in higher education: Promises and pitfalls",*Scientific Research and Essays Vol. 6(7), pp. 1573- 1582, 4 April, 2021.*

[7]K. L.-M. Ang, F. L. Ge, and K. P. Seng, "Big educational data an- alytics: Survey, architecture and challenges",*IEEE Access, vol. 8, pp. 116392–116414, 2020.*

[8] **Ching-Chieh Kiu,** "Data Mining Analysis on Student's Academic Perfor- mance through Exploration of Student's Background and Social Activities ",*IEEE Access, 978-1-5386-7167-2/18.*

[9] **Rodrigo da Fonseca Silveira, Marcio de Carvalho Victorino** "Educa-
tional Data Mining: Analysis of Drop out of Engineering Majors at the UnB
- Brazil ",*2019 18th IEEE International Conference on Machine Learning and Applications.*

[10] **Hina Gull , Madeeha Saqib , Sardar Zafar Iqbal , Saqib Saeed,** "Im- proving Learning Experience of Students by Early Prediction of Student Performance using Machine Learning",*2020 IEEE International Conference for Innovation in Technology.*

[11] **Boddeti Sravani, Myneni Madhu Bala** "Prediction of Student Perfor- mance Using Linear Regression",*2020 International Conference for Emerg-ing Technology (INCET) Belgaum, India. Jun 5-7, 2020*

[12] **Kennneth Li-Minn Ang, Feng Lu Ge2, Kah Phooi Seng** "Big Educa- tional Data Analytics: Survey, Architecture and Challenges",*IEEE AC- CESS.2020.2994561.*

[13] **Ahmed Mjhool, Salam Hassan Mhesan Al augby, Ahmed Hazim Al- hilali** "A Proposed Architecture of Big Educational Data using Hadoop at the university of Kufa", *International Journal of Electrical and Computer Engineering · December 2019.*

[14] **Abdullah M. Alghamdi, Fahad A. Alghamdi** "Enhancing Performance of Educational Data Using Big Data and Hadoop", *International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 19 (2019) pp. 3814-3819.*

[15] **Fezile Matsebula, Ernest Mnkandla** "A Big Data Architecture for Learn- ing Aanalytics in Higher Education", *IEEE Africon 2019 Proceedings*

[16]1Abdullah M. Alghamdi, Fahad A. Alghamdi "Enhancing Performance of Educational Data Using Big Data and Hadoop" *International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 19 (2019) pp. 3814-3819 .*