



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 5, May 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

 9940 572 462

 6381 907 438

 ijirset@gmail.com

 www.ijirset.com

Big Data Technology in Education System

Ashwini Narayan Shirke, Prof.Harshada Salavi

Department of MCA, Finolex Academy of Management and Technology, Maharashtra, India

ABSTRACT: Big Data technology is a powerful tool that can be used for many different real-life problems that involve handling large amounts of information. When paired with cloud-based systems, it becomes a cost-effective way to deal with these big data challenges. One area that can really benefit from this technology is education and research. In education, there's a lot of data being generated, especially with online learning becoming more common. This has led to what we call "Big Data in education." Now, with Big Data analytics, we can look at all this information and use it to improve how we teach and run schools.

For example, we can use Big Data to predict how well students might do in certain subjects, or even recommend courses based on how they've performed in other classes. Think of it like how Netflix suggests movies based on what you've watched before, but instead of movies, it's suggesting classes or subjects based on your grades.

This kind of recommendation system can be really helpful for students, teachers, and schools. It's like having a personal advisor who knows your strengths and weaknesses and can suggest what might be best for you to learn next.

KEYWORDS: Education Big Data, Data Analysis, System Architecture, Hadoop

I. INTRODUCTION

In modern education, digital tools like online systems and games collect a ton of information as students go through lessons and activities. This data gives us a better understanding of how learning happens compared to just looking at grades and test scores. Big Data technology helps us handle and analyze all this diverse information, which will be really important for education in the future.

A lot of this educational data comes from things like online courses, learning management systems, educational apps, and even students using social media. Traditional systems can't handle all this data, so we need Big Data technology to manage and make sense of it. Big Data refers to really large sets of data that regular software can't handle efficiently.

Using techniques like data mining, we have a field called Educational Data Mining (EDM) which focuses on analyzing educational data to understand students and their learning environment better. EDM uses various algorithms to improve education quality by finding patterns in how teachers and students use educational systems. It helps us understand things like how students learn and use different learning tools. Techniques like clustering, classification, and prediction are used to analyze this data and improve education.

II. BIG DATA

Big data refers to the gathering of massive amounts of data that can't be handled efficiently using traditional database technologies. It's not just one thing but involves many different areas of business and technology. This data can be structured (like in databases) or unstructured (like text or images). The actual size of the data doesn't matter as much as how we organize and use it.

When we manage big data effectively, it can help us make better decisions and strategic business moves. The concept of big data isn't entirely new; we've been collecting and storing large amounts of data analysis for quite some time.

However, the term "big data" gained popularity in the early 2000s, especially when industry analyst Doug Laney described it using four key aspects, often referred to as the four V's: Volume, Velocity, Variety, and Veracity.

Volume: Organizations collect data from various sources like business transactions, social media, and sensors on machines. Storing all this data used to be a problem, but new technologies have made it easier.

Velocity: Data comes in really fast and needs to be managed quickly. Things like RFID tags, sensors, and smart devices are constantly generating large amounts of data that need to be handled almost instantly.

Variety: Data comes in many different forms, from structured data in traditional databases to unstructured data like emails, videos, and social media posts. It's a mix of numbers, text, images, and more.

Variability: On top of the fast pace and different types of data, there can be fluctuations in data streams. For example, there might be sudden spikes in data when something goes viral on social media, or during holidays or events. Managing these fluctuations, especially with unstructured data, can be challenging.

Today, data comes from many different sources, making it challenging to connect, organize, clean, and transform it across various systems. However, it's crucial to establish connections and understand relationships between data points, or else things can quickly become chaotic.

The amount of data being generated and stored globally is enormous and keeps growing. But despite this vast amount of data, only a small percentage is actually analyzed. What does this mean for businesses? How can they make better use of the data flowing into their organizations every day?

The importance of big data isn't about how much data you have, but rather what you do with it. By leveraging big data analytics, businesses can analyze structured, semi-structured, and unstructured data to gain valuable insights. This can lead to: 1)Cost reduction 2)Time savings 3)Innovation in product development and optimization 4)Making informed decisions quickly, such as identifying the main causes of failures or detecting fraudulent behavior before it impacts the organization.

For example, businesses can generate coupons based on customer buying habits, detect fraud in real-time, or quickly recalculate risk portfolios. These are just some of the ways big data can drive value for businesses.

It's important to understand that the real value of big data comes from analyzing it and extracting insights, rather than just collecting raw data. As big data technologies advance, organizations need to rethink how they use data to make decisions and develop products and services.

Now, let's see how various industries can benefit from big data:

Banking: Banks deal with large amounts of data from various sources. They need to use big data analytics to understand customers better, manage risks, prevent fraud, and comply with regulations.

Education: Teachers can use data-driven insights to improve educational systems, identify at-risk students, monitor progress, and support teachers and principals more effectively.

Government: Governments can use big data analytics to improve services, manage utilities, handle traffic congestion, and predict crime. However, they must also address issues of transparency and security.

Healthcare: Big data helps healthcare providers improve patient care by analyzing medical records, treatment plans, and prescription data quickly and accurately.

Manufacturing: Manufacturers can use big data to improve quality, increase yield, and reduce waste, which is crucial in today's competitive market.

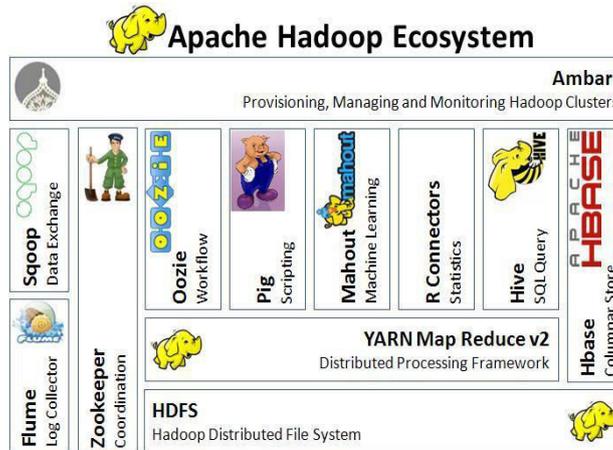
Retail: Retailers rely on big data to enhance customer relationships, optimize marketing strategies, manage transactions, and win back customers. Big data is at the core of these efforts.

A. Hadoop Architecture

Hadoop is an open source framework developed by Doug Cutting and his team. Hadoop runs applications utilizing the MapReduce algorithms, where the information is handled in parallel with others. In short, Hadoop is utilized to create applications that could perform final factual examination on tremendous measures of information.

Hadoop has two critical layers to be particular:

- 1.Processing/calculation layer (MapReduce)
- 2.Storage layer (Hadoop Distributed File System)



Hadoop Framework: Picture this as the team leader. It's a set of tools created by Doug Cutting and his team. Their job is to handle massive amounts of data efficiently. They do this by breaking down tasks and spreading them out across many computers, working together.

MapReduce: This is like the task manager. It divides the work among the team members, called nodes, and then gathers all the results to put them together. Think of it as sorting and organizing all those Lego blocks into neat piles based on color or shape.

Hadoop Distributed File System (HDFS): Now, imagine you need a giant storage space to keep all those organized Lego piles. HDFS is like a massive storage room with many shelves. It's built to handle lots of data across multiple computers, making sure nothing gets lost. Plus, it's tough, so even if one computer breaks down, your data stays safe.

Hadoop Common: Think of this as the toolbox for Hadoop. It provides all the necessary tools and utilities, written in Java, that other parts of Hadoop need to work properly. It's like having a set of essential tools that every handyman needs before they can start building or fixing things.

Hadoop YARN (Yet Another Resource Negotiator): This is like the traffic controller for a busy intersection. Its job is to manage and schedule jobs (tasks) across all the computers (nodes) in a Hadoop cluster. Instead of relying on one super powerful server, YARN makes use of many regular computers connected together. It's like having multiple lanes on a highway rather than one big road. This way, tasks can be processed faster and more efficiently.

III. RECOMMENDER SYSTEMS

In this section we will discuss about the recommender system techniques.

Fig. 1 shows the steps involved in recommendation system for educational data. First step is to obtain data from the educational resources. This involves integrating multiple databases, data cubes, files etc and removing the inconsistencies of data. Data and attribute redundancy is taken care at this stage. Next step is to select the required attributes of data using various feature selection and extraction techniques. This step reduces the volume of data such that analytical results are not affected. As data can originate from different sources, it is susceptible to noise, missing values and inconsistency. In order to get the correct result it is required to preprocess it. Following tasks is performed at data preprocessing stage

- **Data Cleaning-** Data is cleaned at this stage. Missing values are handled by filling in those values. Noisy data those data which are not conveying meaningful information are handled using smoothing techniques. Smoothing techniques creates a function to extract meaningful information from data and ignore the noise.

- **Transformation of Data-** Data discretization or hierarchy generation can be used for data transformation. Data discretization is to convert expression data into finite value. Hierarchy generation reduces the data by converting low level concept for example age into higher level concept for example youth. Next step is to use appropriate data mining technique to obtain the desired result. At this stage identified data mining technique is used to generate the recommendation for users. The common tasks in recommender systems are Top-N item recommendation and rating prediction. The recommender suggests a ranked list of items to a user. Rating prediction predicts a preference score for

a given user-item combination As explained in Fig-2, the recommendation techniques can be broadly categorized into collaborative filtering, content-based filtering knowledge-based system, hybrid systems etc.

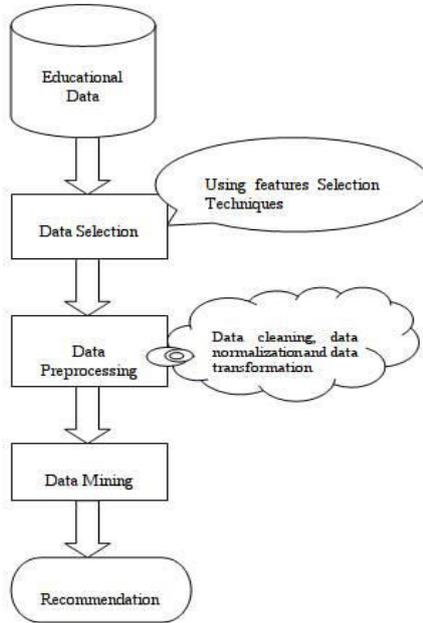


Fig 1

IV. METHODOLOGY

The "Analyzing Big Education Data" aims to develop a tool for analyzing large volumes of education-related data. Here's a summary of the chosen technologies and their rationale:

Storage Solution: Google BigQuery is selected for its ability to handle vast amounts of data in the cloud. It's cost-effective for data sizes under 10 GB per month, making it practically free for users. Alternatives like MongoDB and Cassandra are complex and suitable only for highly intricate designs.

Processing Language: R programming language is chosen for its suitability in interfacing with both the storage solution and the web interface. It's open-source, has numerous packages for various functionalities, and simplifies project complexity, reducing development time and effort. Alternatives like Hadoop, Hadoop with R, and Spark are expected to increase project complexity and cost, leading to unnecessary overheads.

Integration: R has a dedicated package for interfacing with Google BigQuery, enabling data retrieval and query processing. It also offers graphics packages for visual data analysis. For developing the web interface, R provides the Shiny package.



Technology	<i>Reasons for Choosing the Technology</i>
Google BigQuery	1) Free data storage of 10 GB per month is allowed on BigQuery, making it a cost-effective solution. 2) Cloud-based, distributed storage allow easy storage of massive datasets, which is one of the fundamental requirements of the project as far as storage is concerned. 3) Basic data access, retrieval and query abilities of BigQuery provide for the fundamental processing requirements of the project.
R Programming Language	1) Open-source solution, contributing to the costeffectiveness of the developed tool. 2) Availability of packages and inbuilt functionalities for establishing interface with BigQuery, creation of graphics and development of web views simplifies solution’s development

V. ADVENT OF BIG DATA IN EDUCATION

In the current landscape of learning environments, students are increasingly turning to online platforms, such as discussion forums, chats, and Learning Management Systems like Moodle, to engage with their studies. The widespread use of smartphones further facilitates access to educational content, allowing students to participate in learning activities from anywhere with internet connectivity. This shift towards online learning has been accelerated by the global COVID-19 pandemic, which has compelled educational institutions to adapt to remote teaching and learning methods. As a result, the demand for online courses and eLearning programs has surged, leading to a significant increase in web traffic and the adoption of video streaming software for conducting virtual classes.

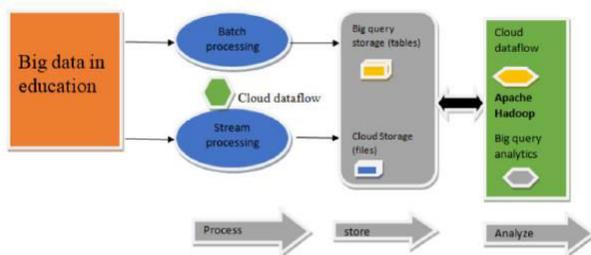
With this transition to digital learning comes a deluge of data generated by student activities and institutional operations. Traditional processing techniques are inadequate for handling the vast amounts of data produced in these scenarios. Consequently, educational institutions are turning to "Big Data" technologies to analyze and derive insights from this wealth of information. Data scientists play a crucial role in examining user behavior, identifying patterns, and understanding preferences and challenges faced by students. By analyzing data on software usage, student engagement, test performance, and attendance patterns, these experts can provide valuable insights to universities and corporations.

These insights, gleaned from big data analysis, enable educational institutions and companies to make data-driven decisions aimed at enhancing the overall learning experience. By understanding user behavior and preferences, institutions can tailor educational content and resources to better meet the needs of students. Additionally, insights derived from big data analysis can inform the development of new educational tools and strategies, ultimately leading to improvements in teaching effectiveness and student outcomes. Thus, big data analytics holds the potential to make a

significant impact on education by empowering institutions to adapt and innovate in response to evolving learning trends and challenges.

VI. ADVENT OF BIG DATA IN EDUCATION

The objective of this approach is that use of big data technology in education implemented using Apache Hadoop within the system process. Batch processing and stream processing are utilized in processing the educational data in this proposed approach. Batch processing is that the approach of evaluating the blocks of information over a particular 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 information and performs the function by using Hadoop platform. 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. These two types of processing are used to improve the performance of the educational data. With the assistance of cloud data flow, this process is going to be accessed efficiently.



VII. ADVANTAGES OF BIG DATA TECHNOLOGY IN EDUCATION

Infrastructure for Sustainable Change: Big Data technology provides the infrastructure and capacity for sustainable change in education. By institutionalizing data acquisition, exchange, and creation, it enables educators and administrators to make data-driven decisions that positively impact learner experience and educator effectiveness.

Data-Driven Decision Making: Big Data technology allows decision-makers to extract meaningful insights from educational data. This includes identifying learning achievement rates, patterns, and challenges faced by students, as well as monitoring their academic progress over time. By leveraging this data, institutions can implement targeted interventions to support student success and improve overall educational outcomes.

Adaptive Learning Systems: Developed models based on data processing algorithms enable the construction of adaptive learning systems. These systems leverage predictive analytics to personalize the learning experience for individual students. By using data-driven insights to adapt course content and recommendations, educators can better meet the diverse needs of learners and enhance their learning progress.

Predictive Teaching and Assessment: Big Data technology empowers teachers with real-time feedback on the effectiveness of their teaching methods and assessment strategies. By monitoring the learning process and detecting student weaknesses early on, educators can tailor their instructional approach to better support student learning. Additionally, predictive analytics help identify program and content flaws, enabling continuous improvement of educational offerings.

Enhanced Collaboration and Engagement: Big Data technology fosters collaboration and engagement among learners and educators. Through personalized instruction, responsive assessments, and interactive teaching methods, educators can create a more engaging learning environment. Furthermore, collaborative learning platforms facilitate knowledge sharing and peer-to-peer interaction, enriching the overall educational experience.

VIII. ADVANTAGES OF BIG DATA TECHNOLOGY IN EDUCATION

Storage: With the amount of data generated by the web and educational institutions reaching exabytes, traditional databases struggle to handle it all. To address this, new databases that don't rely on traditional SQL queries are used. Compression technology helps store and manage this vast amount of data both on disk and in memory.

Security: As the digital universe grows to 40 zettabytes by the end of the decade, ensuring data security becomes paramount. Traditional security measures like strong passwords and firewalls are insufficient for protecting Big Data, IoT, and Cloud Computing technologies. Training institutions need robust policies to govern data access, storage, and processing, as well as protect intellectual property.

Privacy: Big Data in education raises concerns about student privacy. Massive online courses gather extensive data on students' performance and behavior, which can be exploited if not properly secured. Predictive analytics tools may also infringe on student privacy by predicting future outcomes without consent. The decentralized nature of Big Data storage can further exacerbate privacy breaches, with numerous educational breaches already recorded.

Analysis: Analyzing the vast and diverse data generated by online learning platforms requires significant time and resources. To address this, scalable architectures distribute data processing across a network of computers, breaking data into smaller chunks for efficient processing.

Reporting: Traditional reports can become difficult to interpret when dealing with large datasets. Big Data technologies offer various techniques to represent data in easily understandable formats, overcoming the challenges of interpreting massive amounts of data effectively.

IX. CONCLUSION

Through this study we can see that Big Data can truly improve the learning. Can bear to shape a modern and dynamic schooling framework, which each individual student can have the most extreme benefit from that. Moreover, educators have important tools, where they don't have previously, which can settle on their choices more explicit and can pick a major assortment of new learning strategies. Thus, the Big Data are really included to change the method of enterprises including the education. In the new time of Data, the conventional challenges will not exist anymore, keeping the great strategies. The education framework will be enhanced with new learning ways, making more productive and designated. However, the method of this new time, have quite recently started and there are numerous troubles, for example, the absence of experienced faculty on the study of Big Data and Data investigation. Moreover, the instructors and academics should really prepare and involved on them and the students should acknowledge and utilize these new tools.

REFERENCES

- [1] Educational Intelligence:Applying Cloud-based Big Data Analytics to the Indian Education Sector by Samiya Khan, Kashish A. Shakil, Mansaf Alam.
- [2] PABED – A Tool for Big Education Data Analysis by Samiya Khan, Kashish Ara Shakil, Mansaf Alam.
- [3] Extending R Boxplot analysis to Big Data in Education by Dr. Rajiv Pandey, Nidhi Srivastava, Dr Shahnaz Fatima.
- [4] Recommender System for Big Data in Education by Surabhi Dwivedi1, Dr Kumari Roshni.
- [5] Systematic Approach on Big Data Analytics in Education Systems by 1R.Swathi, 2N.Pavan Kumar, 3 L.KiranKranth.
- [6] Homogenizing Social Networking with Smart Education by means of Machine Learning and Hadoop: A Case Study by Ashish Jagtap, Balaji Bodkhe.
- [7] Research on Architecture of Education Big Data Analysis System by Jinhua Chen, Jing Tang, Qin Jiang, Yuxin Wang, Chunmei Tao, Xu Zhang, Jingya Liao.
- [8] Improving the quality of education system using Data Science Technologies : Survey by Pattinige Ravindra R Fernando.
- [9] The Application of Big Data Technology and Method in Moral Education in Colleges and Universities by Fang xi, Wang Juanjuan.
- [10] Big Data Technology in Education By 1devika Subhash, 2tina Babu C.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details