

**INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH  
TECHNOLOGY****STUDY OF PROS AND CONS IN THE EDUCATION SYSTEM USING BIG DATA****Arushi Gupta\*, Asmita Sharma, Astha Sahu, Anjali Mukati, Ashlesha Panse**

Computer Science and Engineering, Acropolis Institute of Technology and Research, Indore, INDIA

DOI: 10.5281/zenodo.48924

**ABSTRACT**

“Data is a precious thing and will last longer than the systems themselves.” – Tim Berners Lee, inventor of the World Wide Web. When Berners-Lee talked about data, he wouldn’t have imagined it growing to a gigantic size encompassing all spheres of life. Think of Big Data once and you’ll understand what we are talking about. From retail marketing to cloud computing, from improved patient outcomes to enhanced teaching learning processes, Big Data has a major role to play in every sector. Big Data in education could be leveraged to enhance creativity, potential, and problem solving ability among students. It can also open up a wide array of possibilities for students and teachers by freeing their minds, and setting them to do bigger things – and things that matter – rather than sticking to a standardized pattern of education. Imagine a world, where teachers can not only analyze students’ answer given in exam sheets to discover their learning abilities but also understand why do some of their students are not able to grasp the concepts clearly and why do they give wrong answers. Imagine a world in which the lecturer can analyze how their students use the information given in textbooks or through Internet and accordingly give the pointers and are able in bestowing tailor-made lectures to inspire and energize students. A pattern of wrong answers is no longer just a bad grade; teachers can get clues to why students picked the wrong answer. It has been promised for a long time that technology will change education for better – thus more affordable and accessible. The promise of educational technology is more important in Indian context because we have an enormous lack of access to high quality education due to a number of seemingly unbeatable challenges. Also, the cost of educating one of the world’s greatest populations has been steadily increasing. The educational technology promises to give the same platform for everyone. Even the rural children expect that technology would make education affordable and accessible for them. In education, digital learning technologies can help educators and researchers gain valuable insight into how to improve and personalize learning for students.

**KEYWORDS:** Big Data, Education, Challenges, Hadoop.**INTRODUCTION****A. BIG DATA**

Big data is a term for data sets that are so large or complex that traditional data processing applications are inadequate. Big data describes the large volume of data –structured, semi-structured and unstructured – that inundates a business on a day-to-day basis. But it’s not the amount of data that’s important. It’s what organizations do with the data that matters. Big data can be analyzed for insights that lead to better decisions and strategic business moves.

**Roger Magoulas** (director of market research at O’Reilly Media) coined the term Big Data.

**According to Gartner:** Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.

**From Wikipedia:** Big Data is a broad term for data sets so large or complex that they are difficult to process using traditional data processing applications. Challenges include analysis, capture, curation, search, sharing, storage, transfer, visualization, and information privacy.

**According to IBM:** Data coming from everywhere; posts on social media sites, transactional data during e-commerce, videos, texts etc.

### **BIG DATA CAN BE DESCRIBED BY THE FOLLOWING CHARACTERISTICS:**

- 1) *Volume:* The amount of data which is growing vastly every year. Social media sites are greatest contributor on increasing the volume of data. Now we have data in petabytes but in near future it will become zettabyte.
- 2) *Velocity:* The speed of increasing data or we can say the rate of data flow from various resources.
- 3) *Variety:* Data which is coming from various resources having different format such as structured, un-structured and semi- structured (log files, server logs, videos, emails etc) .
- 4) *Veracity:* Data that we have from various resources must be accurate. We cannot make decisions on the basis of inaccurate data.
- 5) *Complexity:* We have enormous amount of data in various formats such as structured, un-structured and semi-structured. So extracting information from social media sites is quite complex.

*“Big ideas don’t have to solve education’s problems. It is about objectives, using data and data mining to connect content and objectives.”*

The education has also transformed in tremendously. It has turned from offline/Indoors to online. The increasing potential of technologies to make learning available online to students globally is one of the most crucial educational concerns. With innovations in technology and increasing student mobility, educational delivery is changing constantly. Big data is a tool to drive innovation in education and teaching quality impacts student’s achievement. There are considerable opportunities for using Big Data in higher education. It is important that universities use Big Data to continue to deliver the very best learning environments for the good of society. There may also be possibilities around using Big Data to link research to education – both by making better use of latest research practices and outcomes to inform teaching and in enabling research activities to be undertaken as part of education. A key part of the higher education agenda is to harness cross-disciplinary intelligence to improve curriculum, content and delivery, enhance learners’ experiences and create an environment that qualifies them with the skills and knowledge necessary. Educators armed with data-driven insight can make a significant impact on school systems, students and curriculums. Digital technology makes informative content easier to find, to access, to manipulate and remix, and to disseminate. All of these steps are central to teaching, scholarship, and study. Together, they constitute a dynamic process of digital learning. By analyzing big data, they can identify at-risk students, make sure students are making adequate progress, and can implement a better system for evaluation and support of teachers and trainers. Online teaching can use BIG DATA to reform educational delivery and enhance learning in numerous ways, for example to: adapt and improve delivery through personalizing learners’ experience; create communities of practice; and standardize the presentation of knowledge.

### **B. HADOOP**

Hadoop is an open source software framework that allows the distributed processing of large data sets across clusters of commodity computers using a single programming mode. Hadoop was created by Doug Cutting and Mike Caffarella in 2005. Open source software-Hadoop is free to download, use, create and manage the program. Framework-Hadoop provides toolset, connections to develop and run software applications. Hadoop distributes the file in small chunks over thousands of nodes and process the data in parallel way. Hadoop stores the massive data and replicates the data. Big data is analyzed, handled, operated and utilized by Hadoop.

There are a number of pressing issues for education, including how to:

- increase educator effectiveness;
- harness insights from learning experiences;
- deliver education for all;
- prepare students with relevant skills for their future careers.

Below are some reasons why traditional educational approaches are being failed in India:

- **Education system provides rat race among the students.**  
The students have to learn and mug up the entire textbooks without understanding the topics. Sometimes, they don’t have any idea why are they learning these things or what are the applications of certain topics in real life.

- **No critical analysis or perspective.**  
Today's children are not able to analyze the things correctly. They take the line of establishment or the views of predominant majority.
- **Students are limited to local data rather than having access to global outlook.**
- **Some of the Teachers Themselves are Not Trained and Efficient.**  
The teachers themselves do not have sufficient knowledge and aren't properly trained. Sometimes, they don't even know what values they should impart to children. And more importantly, government doesn't pay the teachers enough salary. So, to improve our education system teachers should be better trained and more importantly better paid.
- **Missing innovation and creation.**

Big data is considered as the heart of digital revolution. Big Data can play a critical role in deciding the core focus of the education sector in the country by providing insights and outlining an illustrious path of growth.

### LITERATURE SURVEY

**Sagioglu, S.; Sinanc, D.** (20-24 May 2013), "Big Data: a Review" describe the big data content, its scope, methods, samples, advantages and challenges of Data.

**Amy Moynihan**, Content Manager at Hanover Research in her article titled "*The Roles of Big Data and Research in Improving Teacher Quality*" raises questions on how to connect Big Data, teacher quality and student achievement.

**Viktor Mayer-Schönberger, Kenneth Cukier** in *Big Data: A Revolution That Will Transform How We Live, Work, and Think* gives an introduction to what Big Data is and how it has changed the world around us. The book throws interesting examples that along with being interesting also showcases how big data have been utilized till date.

**S. Vikram Phaneendra & E. Madhusudhan Reddy et.al.** in "*Big Data- solutions for RDBMS problems-A Survey*" illustrated that in earlier days the data was less and therefore could be easily handled by RDBMS but eventually as the data increased, it became difficult to handle this amount of data through traditional RDBMS tools, which is referred as "big data". In this they told that big data differs from other data in 5 dimensions such as volume, velocity, variety, value and complexity. The authors focused on the challenges that need to be faced by enterprises when handling big data i.e. data privacy, search analysis, etc.

**Kiran kumara Reddi & Dnysl Indira** in "*Different Techniques to Transfer Big data: a Survey*" gives us an introduction that Big Data is a combination of structured, semi-structured, unstructured homogenous and heterogeneous data. The author also suggested to use a nice model to handle huge amount of data transfer over the network.

**Beheshti**, in his paper titled "Predictive performance of prevailing approaches to skills assessment techniques: Insights from real vs. synthetic data sets", uses synthetic as well as real data sets for assessing the skills of learners. He compares the differences between the details of the skills obtained by the different techniques and analyses his methodology. The results show that the real data provides more accurate results.

**Real Time Literature Review about the Big data according to 2016**, Facebook has over 1.591 billion monthly active users, 1.038 billion daily active users and 1.442 billion mobile monthly active users. Almost five new profiles are created every second. Another example is Instagram with 400 million monthly active users and around 75 million people use Instagram daily. Around 990 million people use another common messenger Whatsapp. 30 billion messages are sent via Whatsapp daily.

### CHALLENGES

Currently, a key concern is that Big Data projects are highly dependent on skilled developers and computer programmers and thus there is a necessity of user-friendly developer tools. Appropriate education and skills training from an early age to encourage users to play with Big Data datasets in a fun and informative way is required. The

internet, cloud computing, live stream, and other comparable technological developments bring new forms for delivering and increasing access to learning. In the process the widespread ability to collect, integrate and analyze Big Data data from these activities is generating important opportunities for improving education, but it also poses new challenges. Some of the challenges are:

- Big Data involves learning complex technology, which restricts the rate at which Big Data resources are available for education users.
- Data integration is crucial in managing Big Data, but governance and data quality need to be key areas of focus for Big Data projects.
- Poor data quality can significantly impact the effectiveness of Big Data projects. As analysis on Big Data grows, so too will the need for validation, standardization, amendment, enrichment and resolution of data.
- There is a lack of a common language across platforms that should be addressed to ensure clear and strong communications.
- Storage: While the common capacity of hard disks nowadays is in the range of terabytes, the amount of data generated through internet everyday is in the order of exabytes. Though the data generated in education is not as large as all the data generated through internet, it is large enough, and would get much larger in future. The traditional RDBMS tools will be unable to store or process such Big Data.
- Much of the data available is of no interest, so it needs to be filtered. The challenge is to define these filters in such a way that they do not eliminate useful information.
- Another challenge is to automatically generate the right metadata to describe what data is recorded and how it is recorded and measured.
- Privacy is the most sensitive issue, with conceptual, legal, and technological implications. It is a fundamental human right that has both intrinsic and instrumental values.
- Heterogeneity, scale, timeliness, complexity and privacy are certain challenges of big data.

## METHODOLOGY

Big data is one of the largest streams to be discussed under one heading, so let us discuss some of the key topics under BD data. Here in this research paper we are discussing about big data in education field. Education is something that can't be defined under the range of pointer and percentile that is scored on basis of cramming things one night before the examination. There is a great relation between education and knowledge. The main element which acts as a link between them is nothing but a book. But in today's fast growing metro life it is inconvenient to handle a book and all the issues that come as a combo pack for a book holder.

- Book requires space as well as maintenance.
- If you are having a library, you will have to keep track of record for the book issued and deposited.
- Library is not a one-time investment because technology keeps on changing on daily basis.
- Books can't be updated as per requirement.

In this do or die condition, big data acts as a lead role by proving a great solution for all the above mentioned issues in one time investment plan. It is a profit investment for any institute to have such a high-tech and advance library which includes recommended as well as renowned authors of each and every field. Often when the terminology "library" strikes our mind the image of a very large data collection on a hard paper is created in our minds. Big data and world's largest network i.e. internet are interlinked with each other so it becomes easy in transportation of data all over the globe. In this hi-tech and advance level, library data collection can be of three types:

- E-books
- Audio type
- Video type i.e. video lectures

This type of library is helpful not only for people who are able but also for differently able person. Let us take an example of an institute having a library area of approximately 854.33 sq m and overall collection of books as 21,757. If we assume an average cost of 150 per book and total number of books be 21757 then the total cost will be 32, 63,550/-. Besides these expenses it will also require annual maintenance as well as per year increment of approximately 500-600 books to keep it up to date i.e. revised editions. Overhead expenses will comprise of 70 cupboards to hold such a large amount of books.

Apart from these, it will also require human resources having a team of at least 4 members for performing tasks such as issuing/returning/due submissions, scholars' library, arranging the books in their shelves properly and replacing the damaged books. But it will be a time consuming and tedious work.

#### What We Can Do?

Minimum cost of a tablet is 2000/-. If we buy around 200 tablets, then the total cost would be 40,000/-.

If we are having a setup in which we provide facility of reading any desired course book or reference, it will encourage students to acquire knowledge in a comparatively easy way. By using this simple and user-friendly technology of big data we can make some best of changes which will help in the evolution of our society. Thus this technology will contribute towards the growth of nation in the field of education.

It will act as a boon because we can control all the data using one main server computer. It will be eco-friendly and it does not require large money to be invested for updating. There will be more variety of books available on the tablets as compared to the books in the library shelves. It will provide a larger scope for the students as the books will be readily available. Here we are talking about online availability of books with the help of one application similar to Whatsapp. We can easily keep a track on books which can be enrolled only by the registered numbers entered in the computer server. If there is any transaction of book on number which is not registered in the server, the student will have to pay some extra charges.

#### CONCLUSION

Big Data is a data whose scale, diversity, and complexity require new architecture, techniques, algorithms, and analytics to manage it and extract value and hidden knowledge from it. We have entered an era of Big Data. The paper describes the concept of Big Data along with 3 Vs: Volume, Velocity and variety of Big Data along with Veracity and complexity. In this paper, we discussed the traditional educational system and its flaws with some solutions. We also gave a methodology for converting the library into E-Library using the concept of Big Data. Big Data techniques can be used in a variety of ways in learning as listed below:

- *Performance Prediction*: Student's performance can be predicted by analyzing student's interaction in a learning environment with other students and teachers.
- *Data Visualization*: Reports on educational data become more and more complex as educational data grow in size. Data can be visualized using data visualization techniques to easily identify the trends and relations in the data just by looking on the visual reports.
- *Intelligent feedback*: Learning systems can provide intelligent and immediate feedback to students in response to their inputs which will improve student interaction and performance.
- *Course Recommendation*: New courses can be recommended to students based on the interests of the students identified by analyzing their activities. That will ensure that students are not misguided in choosing fields in which they may not have interest.
- *Attrition Risk Detection*: By analyzing the student's behavior, risk of students dropping out from courses can be detected and measures can be implemented in the beginning of the course to retain students.
- *Student skill estimation*: It refers to the estimation of the skills of the students so that the learning environment can be adjusted to suit the student's skills. Skills were calculated based on the interaction of the student with the system or in the message boards or discussion forums. Paulo Blikstein in his paper "Using learning analytics to assess students' behavior in open-ended programming tasks" notes the use of a tool named "NetLogo". He logs the mouse inputs of students into the lab machines through the software and with help of the data logged finds the error rates and progress rates of the students.
- *Behavior Detection*: Detection of student behaviors in community based activities or games which help in developing a student model. Joseph Grafsgaard, in his paper "Predicting Learning and Affect from Multimodal Data Streams in Task-Oriented Tutorial Dialogue", presented a system for recognizing the facial expressions of the students to predict the engagement, frustration and learning outcomes of students after the learning session. He also used gesture detection and posture tracking algorithms to capture non-verbal behaviors of students and associates them with the learning patterns.
- *Grouping & collaboration of students.*
- *Social network analysis.*

- *Constructing courseware.*
- *Planning and scheduling.*

Big Data have immense influence to predict the future of education. Nowadays there is a growing need of analysis techniques and technology in the entire domain such as in government, business etc. In education domain big data will help to improve learners and learning skills and to achieve immense productivity and efficiency of the organization. It helps in making decisions. Big Data in education and analytics helps us to achieve successful future of education for the learners.

## REFERENCES

- [1] Utsow Pradhan: “*Big Data in education sector can do wonders- Let’s see how!*” - NextERP Blog, 4<sup>th</sup> August 2015.
- [2] See, [https://en.wikipedia.org/wiki/Big\\_data](https://en.wikipedia.org/wiki/Big_data)
- [3] Oliver Marsh; Lajos Maurovich-Horvat; Dr Olivia Stevenson: *UCL Policy Briefing* – September 2014.
- [4] Sagioglu, S.; Sinanc, D. (20-24 May 2013): “*Big Data: A Review*”.
- [5] Amy Moynihan: “*The Roles of Big Data and Research in Improving Teacher Quality*”.
- [6] Viktor Mayer-Schönberger; Kenneth Cukier : “*Big Data: A Revolution That Will Transform How We Live, Work, and Think*”.
- [7] Magoulas, Roger: “*Introduction to Big Data*” (O’Reilly)
- [8] Katrina Sin; Loganathan Muthu: “*Application of Big Data in Education Data Mining and Learning Analytics- A Literature Review*” Volume: 05, Issue: 04, July 2015.
- [9] Karan Kyanam: “*6 problems with the Indian Higher Education System*” <https://www.linkedin.com/pulse/201407091711442669075126problemswiththeindianeducationsystem>
- [10] Sachin Sharma; Diksha Sharma; Pankaj Vaidya: “*Analytics in Education Using Big Data*” Volume: 04, Issue: 11, November 2014
- [11] Harshawardhan S. Bhosale, Prof. Devendra P. Gadekar: “*A Review Paper on Big Data and Hadoop*” in International Journal of Scientific and Research Publications, Volume: 04, Issue: 10, October 2014.
- [12] S.Vikram Phaneendra & E.Madhusudhan Reddy: “*Big Data- solutions for RDBMS problems-A Survey*” in 12th IEEE/IFIP Network Operations & Management Symposium (NOMS 2010) (Osaka, Japan, Apr 19{23 2013).
- [13] Kiran Kumar Reddi; DnvsIs Indira: “*Different Techniques to Transfer Big data: a Survey*” in International Journal of Engineering Research and Applications, Volume: 03, Issue:06, Nov-Dec 2013.
- [14] Paulo Blikstein, “Using learning analytics to assess students' behavior in open-ended programming tasks”, *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, pp. 110-116, 2011.
- [15] Joseph Grafsgaard; Joseph Wiggins; Kristy Elizabeth Boyer; Eric Wiebe and James Lester: “*Predicting learning and affect from multimodal data streams in task-oriented tutorial dialogue*”, Proceedings of the 7th International Conference on Educational Data Mining, 2014.