Big Data Model for Financial Assistance to Students- A Paradigm Shift

Harinder Pal Singh, Harpreet Singh, Jang Bahadur Singh

Abstract*: *Big Data Technologies for managing huge amounts of student data which is rising exponentially year after year specially in the countries like India, where there is a large no of young population is a need not luxury. Many students are searching for courses of their choice and for funding their education interests they look for scholarships. There are many scholarship schemes available for students of different status in the society. Educational institutions look for solutions to handle large amount of student data ranging from enrollments, scholarships, academic content, literature search, testing, results and finally placements. Such data require a lot of integrity and scalability. Memory saving schema for hardware and software platforms for the needy Universities and Colleges along with lot of mobility for their executives to access and analyze such data remain in focus. In this paper a new data base model has been suggested using modern day Big Data technologies rather than using conventional database technologies to handle complex data generated through student activity which can be analyzed to reach decisions. New model suggested after testing real time data sets of scholarships management of technical education system. Lot of memory and cost has been saved along with achieving data integrity, security and improvement in response time.

Keywords— Big Data Technologies, Big Data Model, Data analytics,, Educational Data, Student Scholarships, Student Data management, Technical Education data

I. INTRODUCTION

There are thousands of schools, ITI , Polytechnics and colleges affiliated to universities and millions of students are searching for courses of their choice and scholarships to fund their education. There are many scholarship options available for students of different status in the society. The educational institutions like universities and colleges look for solutions to handle large amount of student data ranging from enrollments, scholarships, academic content, literature search, testing, results and finally placements. The field of education is gaining insight from large volumes and variety of real time data known as Big Data. With the advent of online courses offered by many universities, the amount of data available to educational officials and students has exploded [2]. The Big Data paradigms are needed in today's world to support data mining approaches for increasing the efficacy of educational institutions [3]. A database model has been suggested which is more secure, cost effective and can handle ever increasing memory space requirements. The response time for data queries also increased considerably. Modern day open source tools like MongoDB coupled with cloud Technology used to test real time data samples on real time basis for analysis of students' scholarships data and comparisons thereof conventional technologies. The results are interesting as summarized in the later sections. Some characteristics of NoSQL databases are inherently schema-less and highly scalable [6] features listed below:

A Features:

Data from different universities and colleges stored in the form of JSON style documents and it uses simplified Java Script engine that supports GridFS for storing data. MongoDB is document database in which one collection (i.e. data store) can hold variety of documents. Number of fields, content and size of the document can be different from one document to another. In the above documentation conversion of application objects to structural format of database objects not needed on the other hand no complex joins, as in traditional database systems are required. The MongoDB is proffered as it supports dynamic queries on documents using a document-based query language. Due to program flexibility MongoDB is easy to scale. More importantly it uses internal memory for storing the (windowed) working set, enabling a faster response time. It also supports Replication, Sharing & High Availability. Data are stored in the form of JSON style documents and uses simplified Java Script engine. Advantages of MongoDB based data model are listed below: [5].

a) MongoDB supports GridFS for storing Data collection
b) This is document database in which one collection (i.e. data store) can hold variety of documents. Number of fields, content and size of the document can be different from one document to another.
c) Conversion of application objects to structural format of database objects not needed.
d) No complex joins, as in traditional database systems.
e) It supports dynamic queries on documents using a document-based query language.
f) It is easy to scale.
g) Uses internal memory for storing the (windowed) working set, enabling faster access of data.
h) Index on any attribute could be made and fast In-Place Updates on data.

Manuscript Received on May 20, 2020.

Harinder Pal Singh,Department of ECE, Desh Bhagat University Mandigobindgarh, India Email: harinderpal_s@yahoo.com
Harpreet Singh,Department of Electrical and Computer Engineering, Wayne State University Detroit, USA Email: hsingh1@wayne.edu
Jang Bahadur Singh,, Department of ECE Desh Bhagat University Mandi Gobind Garh, India Email: sonajbs@gmail.com

II. HANDLING BIG DATA OF STUDENT SCHOLARSHIPS

A. Existing Model

There are thousands of students applying for different scholarships every intake and are also applying for renewal after reaching next semester/year [1]. There are no of scholarships options available on the basis of social status of families, merit cum means or uplifting of minority communities etc. etc. The step by step procedure is as shown below which is currently in place in Fig1.

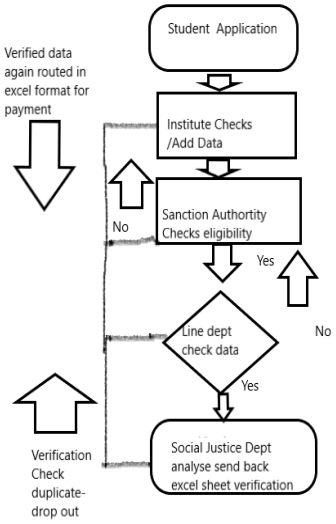

Fig.1. Flow Chart Existing Model

B. Implementation of new model

ford-fulerson algorithm for new framework is:
input: Applications form from students

output: send checks f to banks for awards

for each application (u, v) in *Database* do

implement clustering approach to distribute the applications

while there exists appropriate application to scholarship.

return f

It is based on the following example:
Here follows a longer example of mathematical-style pseudocode, for the ford-fulkerson algorithm :

Algorithm ford-fulkerson is
input: Graph G with flow capacity c,
source node s,
sink node t
output: Flow f such that f is maximal from s to t
(Note that $f_{(u,v)}$ is the flow from node u to node v, and $c_{(u,v)}$ is the flow capacity from node u to node v)

for each edge (u, v) in G_E do
$f_{(u, v)} \leftarrow 0$
$f_{(v, u)} \leftarrow 0$
while there exists a path p from s to t in the residual network G_f do
let c_f be the flow capacity of the residual network G_f
$c_f(p) \leftarrow \min\{c_f(u, v) \mid (u, v) \text{ in } p\}$
for each edge (u, v) in p do
$f_{(u, v)} \leftarrow f_{(u, v)} + c_f(p)$
$f_{(v, u)} \leftarrow -f_{(u, v)}$

The new model is cloud based plateform using virtualized cluster of servers over data centers over SLA [7]. Dynamic resource provisioning of the servers storage and the networks is cloud computing basically. The student fills in the application details from his/her mobile phone/ laptop. The UID server authenticates his/her identity from his UID (Unique Identification) number and opens up the application form. The student fills it up, attaches and uploads eligibility documents, and submits it online. College/University server checks his/her academic, enrollment and performance credentials and forwards his/her application online to District Sanctioning Authority. District Sanctioning authority ascertain the eligibility documents and sanctions the student claim which goes to line department which is Punjab Technical Education, office. The Line Department collects all claims checks authenticity of district sanctioning authority, University/College affiliation and recognition status and the upper limit of amount claimed, finally checks attendance performance from the linked University/Board server and send the claims for releasing payment to the Department of Social Justice. The interlink framework for new data model as shown in Fig-2. The department of social justice sends the money to UID linked bank account of the student through internet banking as the account no and IFS code all data in the student application. Many students are doing multiple times same activities year after year till they pass out. This kind of Big Data generated ranging from admission, claiming financial assistance, attendance, performance etc is stored in mongo cloud and available for analysis by the line department and Department of social justice for arranging funds estimations, budgeting and other decision making analytics. Mongo DB based data base architecture in the new data base model is more secure than MysQL based old data base model where lot of server memory and some manual processing of re-verifying the performance of student was required also manual deleting fake/duplicate claims and there were delays in releasing the scholarships and financial assistance to the students account [1].

C. Hardware Software Specification

The new data model require only a server and the application software giving access to mongo cloud platform, which can be hired for need based memory requirements. In the present case existing national informatics (NIC) server is sufficient for controlling the activity. The NIC server hosts the software application controls. There is no need for adding more hard disk memory or other memory which may become expensive year after year. The application software shall be connecting all the existing servers like mongo cloud, university /board server for student performance query, internet banking and (unique identification authority) UID server to student mobile phones, laptops or tabs.

For this software app the student can install on his mobile phone or use laptop to access the app from internet using normal browser. All the data can be added and processed simultaneously using the app interface. The hardware parallel processing diagram is given as simple illustration Fig-3.
In the following sections for developing a computer and mobile application the basic algorithms for importing existing data to Mongo DB are given.

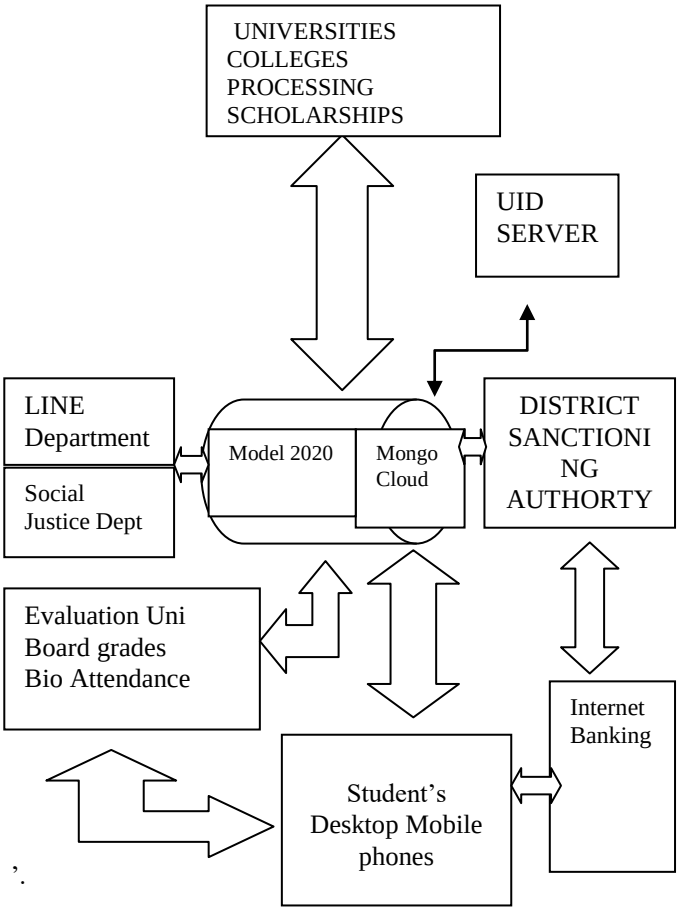

Fig.2. Big Data Scholarship Model

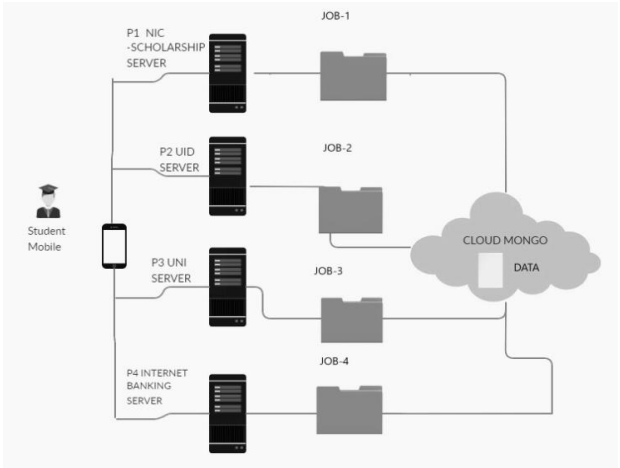

Fig3 Parallel processing layout

D. Importing CSV file into MongoDB

Create a folder on disk C – c:\importMongo then Download the file "ImportDataMongo.rar" from the Google drive then Extract file ImportDataMongo.exe from " ImportDataMongo.rar" in the folder: c:\importMongo then copy here the CSV file which you want to import to Mongo in the folder - c:\importMongo then launch command prompt and change folder to - c:\importMongo.Run the file c:\importMongo\ ImportDataMongo.exe
Note: the CSV file after import will be move to the folder c:\importMongo\Archive
Checking the imported data in MongoDB: Launch the application Compass from MongoDB. Click on Sample_StdRec Click on stdRecords Click on Table

E. The new model makes it a dynamic database architecture based on NoSQL data base. Moving from structured data management having rows and columns to mix of any kind of data like table the student fills application come as structured data, his attendance is captured from Biometric attendance system of college so it come as unstructured data .Progress of grades from university database server again as structured data-tables. Following Fig 4 shows mapping of RDBS (MySQL) to NoSQL (MongoDB) for migration of data base platform.

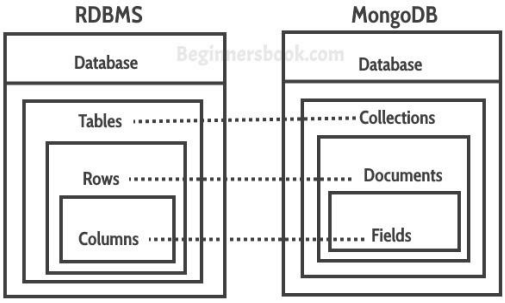

Fig.4 Migrating from MySqL to NoSQL

III. COMPARISON OF BIG DATA TECHNOLOGIES

Real time data sample of student scholarships were tested. The following table sums up the results:

TABLE 1
RESOURCE UTILIZATION MATRIX

Feature	MySql	PostgreSql	MongoDB
Data query time	30 ms	169 ms	3 ms
Memory space	23 MB	19MB	56.7 KB

Resource usage matrix of different data bases as shown in Table-1. Fig.5 shows the size of data on normal disk takes about 23MB of space using MySqL. Also same data sample is used for PostgreSqL as shown in Fig-6 and the disk space taken is 19MB. The disk space size of same data in MongoDB is 56.7k as shown in Fig-7. The response time to query data in MySqLis was 30.1 ms as shown in Fig-8. For same data sample the data query time for PostgreSqL was 169 ms as shown in Fig-9.

Data query time using same data sample with MongoDB as shown Fig-10 was only3 ms. Therefore a big percentage saving of space was achieved along with lowest data query times using MongoDb data base model.

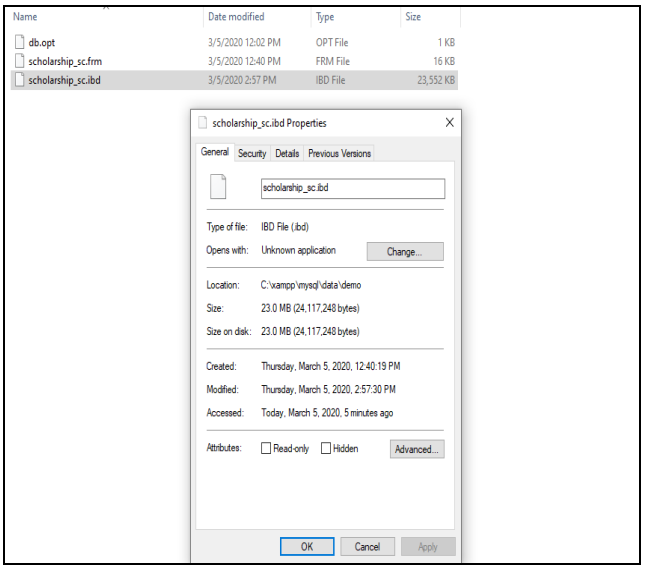

Fig .5 Using MySqL Memory size 23 MB

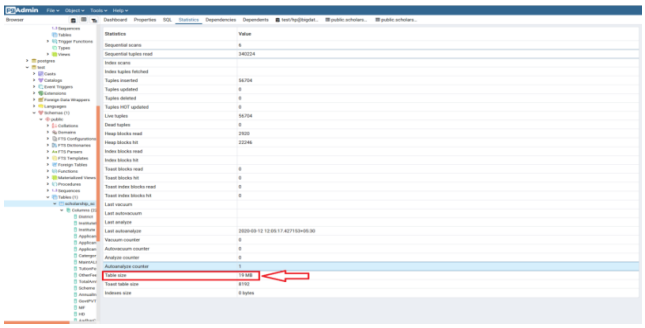

Fig. 6 Using postgreSqL memory size 19MB

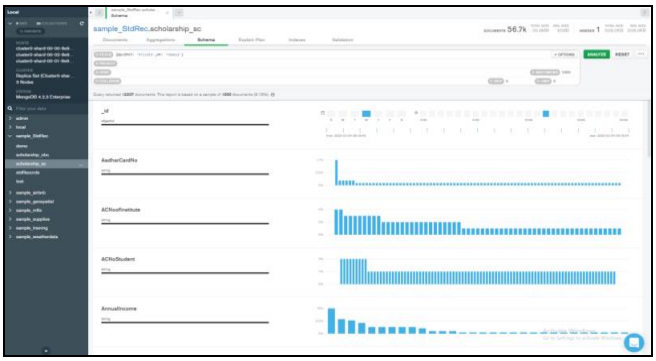

Fig.7 Using MongoDB Memory size 56.7 KB

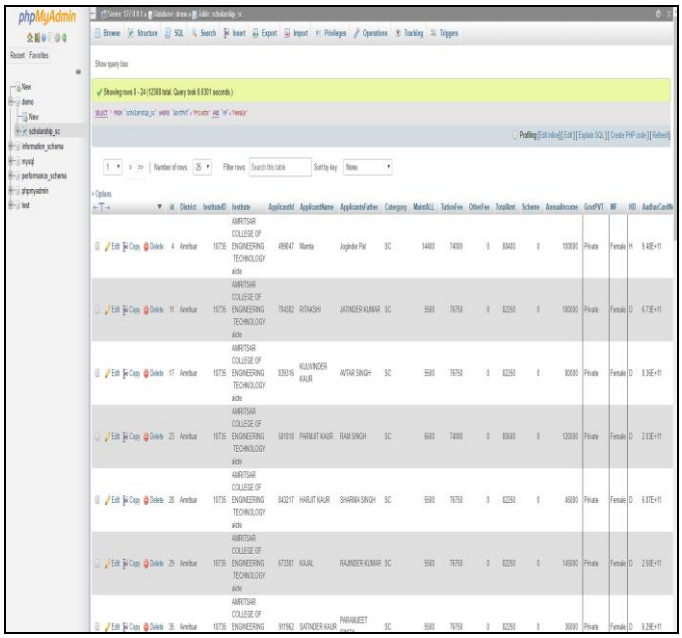

Fig.8 Query time using MySqL 30 ms

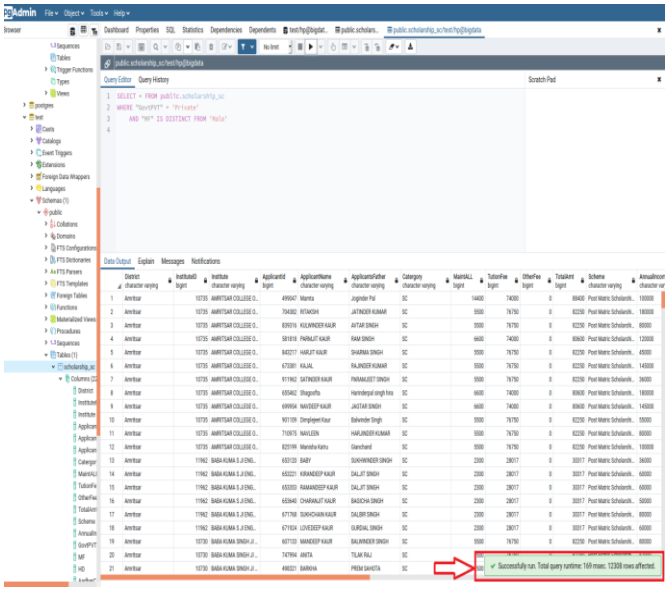

Fig.9 Query time using postgreSqL 169 ms

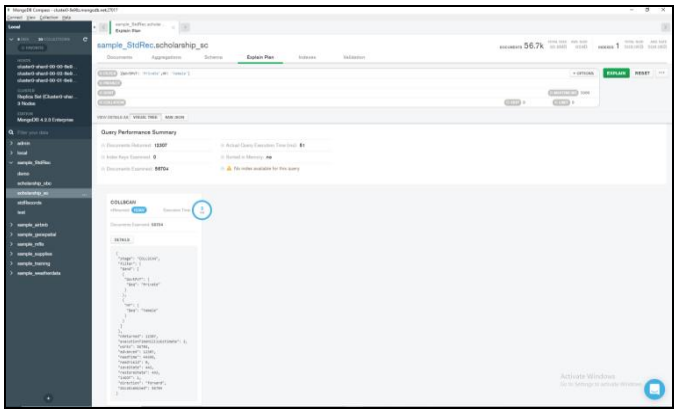

Fig. 10 MongoDB Query time 3ms

Above technology demonstrations it clearly shows the new model using latest BIG DATA tools like MongoDB are far better in space saving and reducing query time and reducing overall process time with better data security.

IV. RESULT AND DISCUSSION

Govt. is providing financial assistance to underrepresented students for Post Matric Scholarship (PMS) Scheme. This scheme enables free education for Scheduled Caste (SC) students and other backward class (OBC) students, whose parent' annual income is less than INR 250,000 and INR 100,000 respectively and their minimum education in each case is 10^{th} standard high school. As per the schedule of the Department of Social Justice and Empowerment of Minorities which is implementing Department the students can apply for Financial Assistance every year. The data for eligible students is processed by Punjab Technical Education Department also known as Directorate of Technical Education which is designated as one of the Line Department other line Departments are Department of Medical Education, Department of Higher Education and Department of School Education.

The following consolidated data table shows year wise Financial Assistance claimed by Underrepresented students belonging to scheduled caste (SC) category whose parents income is equal to or less than INR 250,000 and Other Backward Classes (OBC) category whose parents income is equal to or less than INR 100,000 . The data shown is for Punjab Technical Education only for under Represented Eligible/Duplicate Students and Claims in INR.

TABLE-2

YEAR WISE SC/OBC STUDENTS

Year	Number of S.C std	Claimed Amount	Number of Dropped / Duplicate S.C
2014-15	57440	2489978482/-	8283
2015-16	54276	2608049719/-	3538
2016-17	64029	3108739314/-	6342
2017-18	56825	2846101493/-	7698
2018-19	38885	2005609086/-	3207

Year	No. of OBC std	Claimed Amount	Dropped / Duplicate OBC
2014-15	9854	477288913	1373
2015-16	9153	447747836	451
2016-17	2448	121241411	289
2017-18	952	455664333	110
2018-19	435	21513161	37

Data Source DTE Punjab

Discussion on the above table shows lot of money is being disbursed to the students and there are considerable no of students applying for such financial assistance every year. Similarly students of other line Departments like Medical Education, Higher Education and School Education are also applying for the same as all the students seeking any kind of education are eligible to apply if the satisfy general eligible conditions as outline above. There is a risk of duplicate claims as same student may be applying with other line Departments as the data shows up in the last column of table. The new data base model is very important in the sense that duplicates are weeded out at the first instance using UID identification and dropouts can be weeded out using online university/college server being connected to mongo cloud. If any student drops out or his attendance is not continuous the server will not clear such students for transferring scholarship money for next semester until the student completes requisite attendance and his/her performance requirements. Also the data security risks are minimized as the UID numbers and bank accounts apart from personal information of the students were part of data. The new model would help eliminating duplicity and drop outs and shall provide inter se data analysis opportunities to the state with other all other scholarship schemes like minority scholarship scheme, fee waiver scheme of all the line departments including with other 29 states of the country offering similar financial assistances. This will lead to better decision making for uplifting the underrepresented students thereby contributing to overall development of the country.

V. CONCLUSION

It is successfully concluded that the new database model based on big data technologies for student scholarships will help in weeding out duplicity of claims with other scholarship schemes of the state and the country as the UID server authenticates the student id before application form opens up. It will also help in asserting the student performance like checking attendance, grades from university/server simultaneously without any delays whatsoever. It also helps in transparency in processing the student claims as no manual interface with students and authorities. It also helps with increased data security and authorized access of data due to capabilities of using MongoDB based tool including saving lots of hardware memory space as the cloud technology and need based cloud server can be hired on rent. So it will help in reducing costs related to adding hardware like memory discs and associated software. Considerable improvements in data query times can also be achieved. The new data base model is flexible enough to cater to any number of additions of new scholarship schemes state and national level like persons with disabilities based scholarship and minority scholarship can operated through same model. The model can be useful for other 29 Indian states, processing similar scholarships for millions of students every year. This may save them cost and time apart from avoiding possible frauds and scams. In future all the students may use smart phones for all kinds of educational activity, so this data base model will come handy for them.

Also in future the usage of combinations various parallel programming models based on Hadoop, MapReduce, PACT etc., for various data analytics techniques could be explored to accelerate the analysis of educational data for social justice objectives as these technologies even cater to the unstructured student data from social media networks, there by accelerating the economic growth. This will help in building scalable models in the field of education and may provide a better scope of improvement in the field of Educational analytics as the unstructured data from social media networks can also be taken into account to know the student's future interests as using platforms like MongoDB may also help in even building applications on social media networks thereby handling unstructured and structured students data with more flexibility. Data security has also been enhanced considerably as no one except IP address based pre defined devices are allowed to open the data base. In future the usage of various Big Data platforms like Hadoop, MongoDB, Cassandra etc. and parallel programming models like Hadoop, MapReduce, PACT etc., for various data analytics techniques could be explored to accelerate the analysis of educational data. This will help in building scalable models in the field of education and shall provide a better scope of improvement in the field of educational analytics.

REFERENCES

1. G Harinder Pal Singh and Harpreet Singh, "Big Data Technologies in Scholarship Management of Technical Education System", International Journal of Emerging Technologies and Innovative Research (www.jetir.org | UGC and issn Approved), ISSN:2349-5162, Vol.6, Issue 6, page no. pp492-496, June 2019, Available at : http://www.jetir.org/papers/JETIR1906H50.pdf
2. Yeon Hee Kim and Jin-Ho Ahn Hoseo "A Study on the Application of Big Data to the Korean College Education System", Conference paper (Information Technology and Quantitative Management (ITQM 2016) University of, Korea:
3. Dr. Ben Daniel, University of Otago , New Zealand "Big Data and analytics in higher education: Opportunities and challenges" https://doi.org/10.1111/bjet.12230 British Journal of Educational Technology: BJET, December 2014
4. A.S. Drigas, P.Leliopoulos, "The Use of Big Data in Education", *International Journal of Computer Science Issues (IJCSI)*, vol. 11, no. 5, September 2014
5. Darrell M , "Big Data for Education: Data Mining, Data Analytics, and Web Dashboards ",. West : September 2012
6. Alfredo Cuzzocrea, "Warehousing and Protecting Big Data: State-Of-The-Art-Analysis, Methodologies, Future Challenges", In ICC '16: Proceedings of the International Conference on Internet of things and Cloud Computing, Article number 14, 2016, DOI: http://dx.doi.org/10.1145/2896387. 2900335
7. Kai Hawang, Geoffrey C. Fox, Jack J. Dongarra ,"Distributed and Cloud Computing From Parallel processing to the Internet of Things " pp 35-48 publication Morgan Kaufmann, an imprint of Elsevier 225 Wyman Street, Waltham, MA 02451,USA

AUTHORS PROFILE

Harinder Pal Singh is M.E Electronics from Panjab University, pursuing PhD from Desh Bhagat University Punjab. He is working as Additional Director Technical Education Punjab, Chandigarh, India. He has 6 research publications in international conferences and journals**.** He is member of many professional bodies like ISTE, (INDIA), BERA (UK) and ASEE (USA)

Harpreet Singh is PhD from University of Roorkee, India and working as Professor Electrical and Computer Eng. At Wayne State University, Detroit Michigan USA, he has more than 250 research publications in International journals and conferences. He is a recipient of Khosla Award.

Jang Bahadur Singh is PhD from University of Jammu ,India and working in the Department of ECE at DeshBhagat University Mandi Gobind garh, Punjab, India