

An Efficient Way of Applying Big Data Analytics in Higher Education Sector for Performance Evaluation

Minimol Anil Job, PhD
Faculty of Computer Studies
Arab Open University
Kingdom of Bahrain

ABSTRACT

This research paper evaluate how big data analytics can be used an efficient way for performance evaluation in education sector. The researcher is presenting a suggesting a framework for the big data analytics and highlights the comparison of the institutional data characteristics with big data characteristics. Due to the advancement of technology modern conventional approaches are adopted by educational institutions in teaching and learning. The number of students enrolling for advanced studies are registering for various courses is increasing day by day globally. The higher education sectors are increasingly getting technology centric and institutions under this sectors need to look for tools and technology for data acquisition and storage for further analysis as well as decision making. The huge amount of students' data in these institutions can be considered as big data. Big Data refers to the large volume of the data as well as the technology and tools used to processes and analyze data into usable information. Academic institutions should make use of advanced technologies to yield the benefits from this huge amount of data. Also there is a need to understand how this big data analytics tools and technologies can be utilized in a way to take decisions and drive the institution towards benefitting from the big data.

Keywords

Data Analysis, Big Data, Data Analytics, performance evaluation, Higher Education

1. INTRODUCTION

Higher education institutions need to record all academic related data from various activities such as students' data, tutor data, courses data, registration data, assessment data etc. Due to the advancement of information technology social interactions have increasingly moved to online. These online interactions can be traced and collected and will form a huge amount of data and can be considered as big data. Thus data collection and analysis have become a challenging task in higher education sectors. Most of the higher education sectors lack proper IT infrastructure, tools as well as human expertise required for effective data collection, cleansing, analysis and visualization also while collecting and analyzing student data, higher education institutions face some challenges like privacy, safety, and security issues [12].

The types of big data available in higher education sectors will provide a variety of opportunities which helps to improve student learning such as better learning as a result of faster and more in-depth diagnosis of learning needs or course trouble areas, including assessment of skills such as systems thinking, collaboration, and problem solving in the context of deep, authentic subject-area knowledge assessments [15]. By integrating the collected big data and other sources of data can be used to measure learning, complex problem-solving and collaborative skills. There is a need to find methods and tools

to analyze big data from various sources. Data analysis tools use many types of analysis techniques to store, manipulate and find meaningful information for decision making. These tools will also help in deriving accurate results with minimum efforts.

Student recruitment is continuing as a driving force for higher education institutes. Most of the higher education institutions are facing challenges in addressing problems of progression and retention rates of students. We can say that retaining students is not a new problem in institutions but the ways for student retention is to be considered with highest priority and more researches can be done in this area. Big data analytics tools can be used in performance evaluation and subsequently this help in retention rate analysis. The two main objectives in this research are to recognize available higher education sectors' data as big data by comparing with the four V's of big data characteristics and to suggest a framework for big data analysis in higher education sector.

2. LITERATURE REVIEW

There is no single standard definition for big data. "*Big Data*" is data whose scale, diversity, and complexity require new architecture, techniques, algorithms, and analytics to manage it and extract value and hidden knowledge from it.[17][18] Big Data is used to describe data which is too large in size, which makes it difficult to analyze in traditional ways. [7][4]. Gartner refers to data sets that have "high variety, volume and velocity as Big Data" [8]. Data-driven decision making, popularized in the 1980s and 1990s, is evolving into a vastly more sophisticated concept known as big data that relies on software approaches generally referred to as analytics [2].In education, the emergence of "big data" through new widespread educational mass media, shared with progresses in computation [3] embraces potential for improving learning processes in formal education, and beyond as well. Increasingly, very large data sets are available from students' communications with educational software and online learning - among other sources - with public data warehouses supporting researchers in obtaining this data [11].Technology is at the center of much of the turbulence in our times.While big data and analytics are not answers for addressing all of the issues and decisions being faced by higher education management, they can become part of the solutions and integrated into administrative and instructional functions. [11] Modern Internet and scientific research project produce a huge amount of data with complex interrelationship. These big data need to be supported by a new type of Infrastructure tailored for big data, which must have the performance to provide fast data access and process to satisfy users' just in time needs.[7]

In information technology, big data has emerged as a widely recognized trend, attracting attentions from government, industry and academia. Big data are high volume, high

velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization. [14] Over time, administrative decision making evolved as well, as more data were made available from integrated information systems that could dabble in “what if” questions using database query languages and decision-support systems [13]. The responsibilities of institutional research offices changed from conducting static yearly studies to rejecting information from the institution’s database management systems on an ongoing basis. Regional accrediting bodies began to require colleges to determine a command of the information in their institutions and necessitated proof of data-informed rational planning and decision processes. Most colleges have been able to meet these requirements and have integrated technology into these processes. [4] To take advantage of big data and learning analytics, it is almost a requirement that transaction processing be electronic rather than manual. Traditional face-to-face instruction can support traditional data-driven decision-making processes, however, to move into the more extensive and especially time-sensitive learning analytics applications, it is important that instructional transactions are collected as they occur.[1].

In a white paper published by IBM entitled Analytics for Achievement, eight categories of instructional applications were described. While developed for education in general, they are nevertheless appropriate for the discussion here. The eight categories are as follows [10].

1. Monitoring individual student performance
2. Disaggregating student performance by selected characteristics such as major, year of study, ethnicity, etc.
3. Identifying outliers for early intervention
4. Predicting potential so that all students achieve optimally
5. Preventing attrition from a course or program
6. Identifying and developing effective instructional techniques
7. Analyzing standard assessment techniques and instruments (i.e. departmental and licensing exams)
8. Testing and evaluation of curricula. [Of the above, monitoring individual student performance and participation in a course is among the most popular type of learning analytics application.

3. THE FOUR V’S OF BIG DATA

Big Data is a term for a variety of strategies and tactics that involve massive data sets and technologies that make sense out of these mindboggling reams of data. The Big Data trend has impacted all industries, including the education sectors as new technologies are being developed to automate and simplify the process of data analysis and prediction of results [18][19]. The specific attributes that define big data are called the four V’s: volume, variety, velocity, and veracity. Figure 1 shows the four Vs and the data types

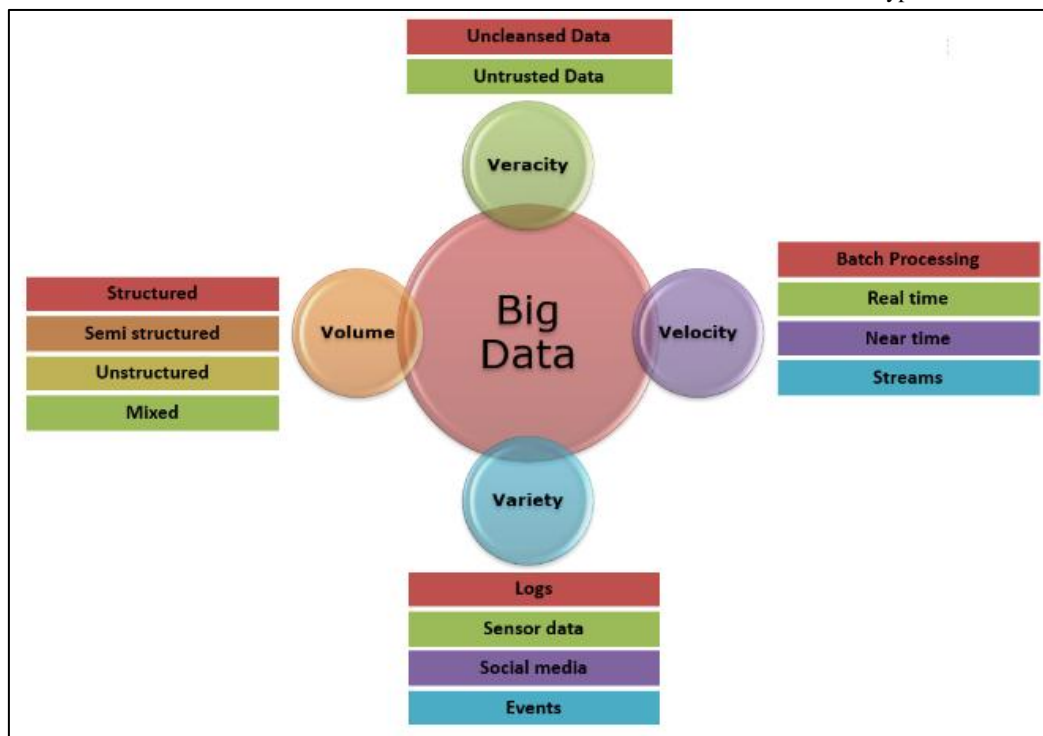


Fig 1: The four Vs of Big Data

- **Volume:** In the context of big data volume is considered as one of the characteristics because of the fact that large volumes of structured and unstructured data are collected and analyzed in organizations from various sources . Data is generated by human interaction through social networks, data from M2M technology and enterprise systems processes and reports.
- **Variety:** Data is available from various sources and various types such as relational data such as tables, transaction, legacy data available from Data Base Management Systems, Text Data from Web sources, XML data availability from online sources and database environment can be considered as semi-structured data, data from social networks, streaming data etc. The

analytic environment in organizations is extracting data from enterprise systems to include unstructured big data from various sources.

- **Velocity:** Data acquisition is an ongoing process in organizations and it is required that data is being generated fast and there is a need to be processed fast for immediate analysis and decision making. Organizations use online data analytics for immediate data processing by considering the fact that late decisions will lead to missing opportunities in their business. Organizations realize the speed of access to report difference between effective and ineffective analytics.
- **Veracity:** The collected data in organizations must meet the quality standards and produce accurate results which will lead to proper action especially in critical decision making. By implementing data validation managerial levels as well as the operational level employees can trust the organizational data and can be used for successful analytics.

4. METHODOLOGY

The research methodology used in this paper is studies by literature review by reading case studies and white papers. Sample references to the latest researches have been taken into consideration. The researcher has identified the available data in Higher Education sector as Big Data by comparing Big

Data Attributes with the attributes of the available data and the factors for predicative analysis using big data in higher education sector. A Suggested model of Big Data Architecture and a framework for Big Data analysis in higher education sector are presented. The future of data mining lies in predictive analytics. The terms *data mining* and *data extraction* are often confused with each other; but there is a significant difference [20]. Big Data is an umbrella term for a variety of strategies and tactics that involve massive data sets and technologies that make sense out of these mindboggling reams of data. The Big Data trend has impacted all industries, including the education sectors as new technologies are being developed to automate and simplify the process of data analysis and prediction of results. The higher education sectors can think of Big Data as the Four Vs, including volume of data; velocity of data, meaning it needs to be analyzed quickly (especially student results); in a variety of structured and increasingly unstructured data formats[14]; which all have potential value in terms of high quality education and skilled graduates.

5. BIG DATA ATTRIBUTES VS HIGHER EDUCATION SECTORS' DATA ATTRIBUTES

Following four tables show the comparison of big data's four V attributes with higher education data and thus identifying as big data.

Table 1: Big Data Attributes comparison

	Big Data	Higher Education Data
Volume	<p>Large volumes of data acquisition in organizations, these data is in the form of</p> <ul style="list-style-type: none"> • structured data • unstructured data • Data is generated by machines for example M2M technology • networks • human interaction on systems like social media the volume of data to be analyzed is massive. 	<p>Large volume of student data from different sources such as:</p> <ul style="list-style-type: none"> • Learning Management System • Registration Data • Assessment/ Performance Data • Student Information System Data • Quality Assurance Survey Data • Graduate Data, Employees Data • Complaint/Appeal Data • Social Media Communication Data etc.
Variety	<p>The data receiving from enterprise systems is huge and from unstructured sources.</p> <ul style="list-style-type: none"> • From DBMS, Relational Data (Tables, Transactional data etc. • , Text Data from Web • XML data - Semi-structured Data • Graphical Data, • Data from Social Network, • Streaming Data. etc. 	<p>Everyday availability of structured and increasingly unstructured data formats</p> <p>Examples:</p> <ul style="list-style-type: none"> • Variety of teaching methods • Virtual Classes • Live Streaming lectures • Tables, SQL • Online Quizzes (web data) • Social media data etc.
Velocity	<p>Speed of access to reports will have the difference between effective and ineffective analytics.</p> <ul style="list-style-type: none"> • Data is begin generated fast and need to be processed fast 	<p>The collected data needs to be analyzed quickly in order to present the reports</p> <p>Examples:</p> <ul style="list-style-type: none"> • Student Admission results

	<ul style="list-style-type: none"> • Through Online Data Analytics methods • Missing opportunities from late decisions 	<ul style="list-style-type: none"> • Assessment Results • Appeal /Complaint Results • List of student with warnings • Expected graduate list
Veracity	<p>By implementing data validation employees in an organization can trust the organizational data and can be used for successful analytics.</p> <ul style="list-style-type: none"> • The data must have quality • The Data should produce credible results • right action at right time • critical decision making. 	<p>The data will have potential value in terms of high quality education and skilled graduates</p> <p>Examples:</p> <ul style="list-style-type: none"> • Preparing Admission results • Preparing end-of-semester results • Evaluating Assessment Quality • Identifying Learning Objectives • Identifying market needs and contextualizing the curriculum • Preparing Graduate list with criteria etc.

The Table-1 presented above depict the comparison of attributes of available Higher Education sectors' data with the four Big Data attributes volume, velocity, variety and veracity. From the comparison results it can be concluded that the higher education data possess the characteristics of big data and can be recognized as big data.

6. PREDICTIVE ANALYSIS OF BIG DATA IN HIGHER EDUCATION SECTOR

According to T. Erl, W. Khattak, P. Buhler (2016), data analytics includes the management of complete data life cycle, from collecting data, cleansing, organizing, storing, analyzing, to data governing. In the context of Big Data, data analytics is focused on managing large volumes of data using highly scalable distributed technologies. Data analytics can be divided into descriptive, diagnostic, predictive, and prescriptive analytics [20].

6.1 Big Data Analysis types practiced by organizations:

- **Descriptive:** A set of techniques for reviewing and examining the data set(s) to understand the data and analyze business performance.
- **Diagnostic:** A set of techniques for determine what has happened and why.
- **Predictive:** A set of techniques that analyze current and historical data to determine what is most likely to (not) happen.
- **Prescriptive:** A set of techniques for computationally developing and analyzing alternatives that can become courses of action – either tactical or strategic – that may discover the unexpected.
- **Decisive:** A set of techniques for visualizing information and recommending courses of action to facilitate human decision-making when presented with a set of alternatives.

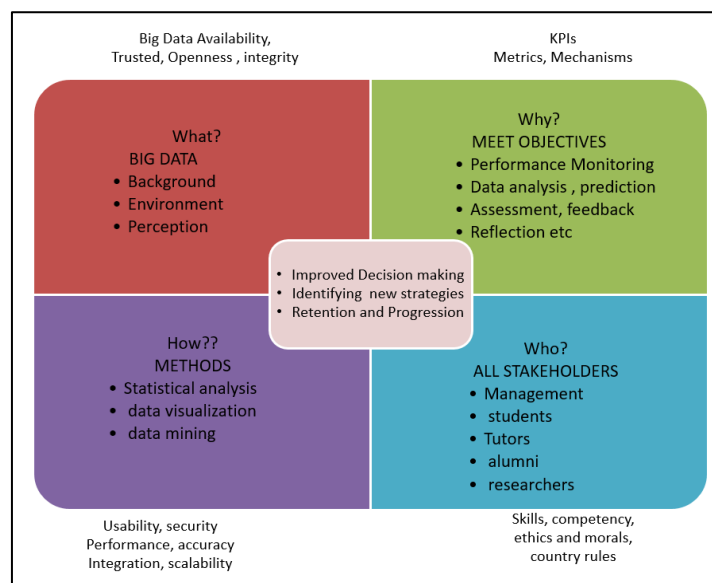


Fig 2: Factors for predicative Analysis using big data

Data Acquisition can be achieved from various educational and administrative activities happening regularly. The data will be collected from student registration, student admission results, Student Information System, Learning management System, E-Library access, Quality Assurance Survey, Social media communications, Performance of assessments, online quizzes (web data), Assessment Results, appeal /complaint inputs and results etc. The overall management of Big Data involves acquisition, storage, processing and analyzing it for various purposes and we can visualize the infrastructure, to handle Big Data related tasks, as architecture as shown in Figure 4 and a the suggested information technology infrastructure is presented in Figure 5.

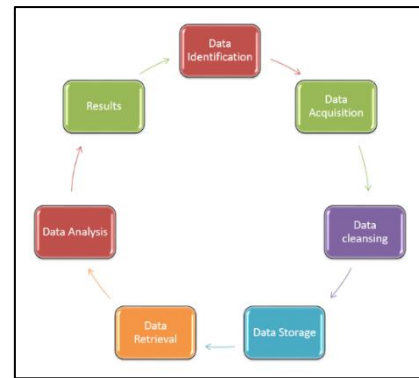


Fig 3: Big Data Life Cycle

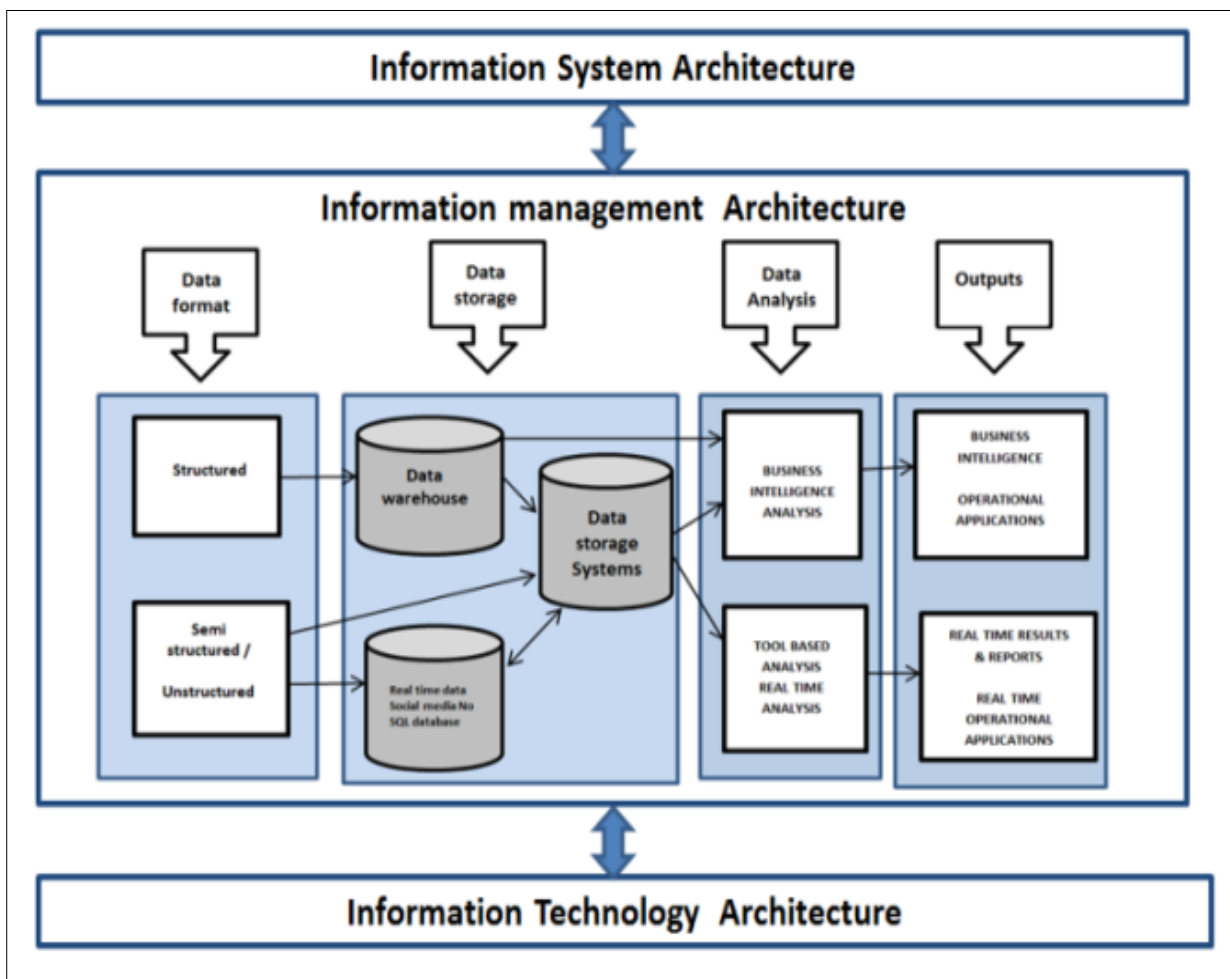


Fig 4: Suggested Big Data Architecture

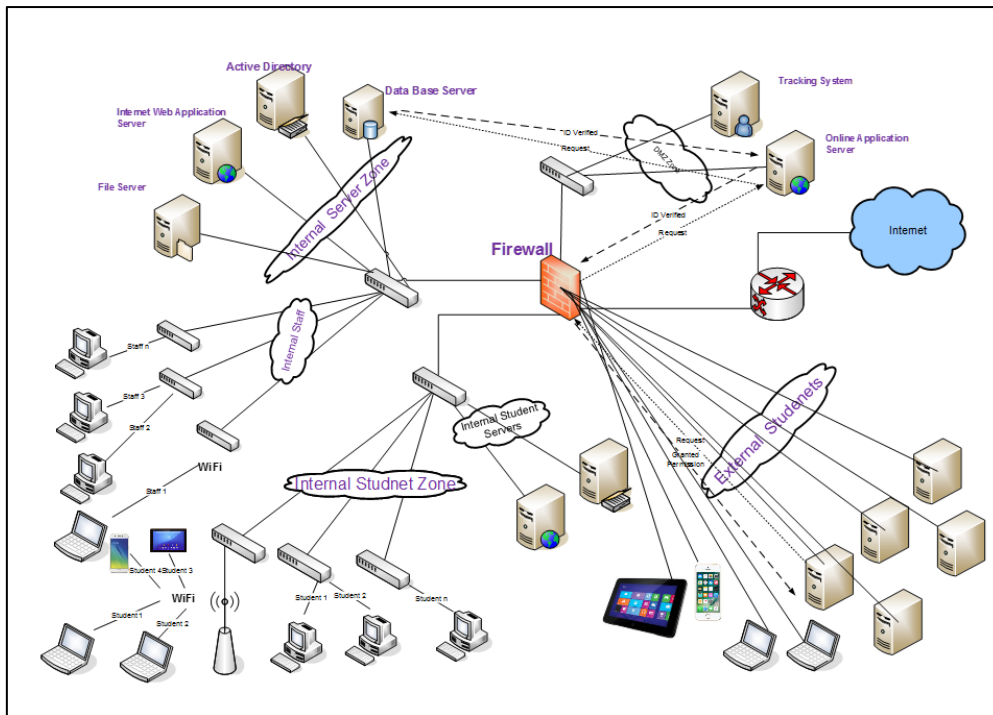


Fig 5: Suggested Information Technology Infrastructure

The information technology architecture shows the server confirmation, networks and various types of internal users and external students. The data is accessed and communications are through personal computers, laptops or mobile devices. In the suggested big data architecture, information management shows the data collection, storage, retrieval of data, analysis and output generation life cycle. Various information systems are used in higher education institutions. Learning Management systems (LMS) is used for learning and teaching, Student Information System (SIS) used for registration of courses, online payment etc. There are systems designed with mobile interfaces also. The collected data can

be in the form of structured, semi structured or unstructured data. The collected data will be stored using appropriate data storage mechanisms. These collected data, big data, will be stored as real time data for immediate processing and results and historical data will be stored in data warehouses for future uses. Data analytical tools can be used to analyze the stored data. For rapidly growing datasets, there can be a need to analyze data as it arrives to get the maximum value due to its time sensitivity. Lastly, the validity of the data needs to be considered whether the derived information from the analysis can be trusted [23].

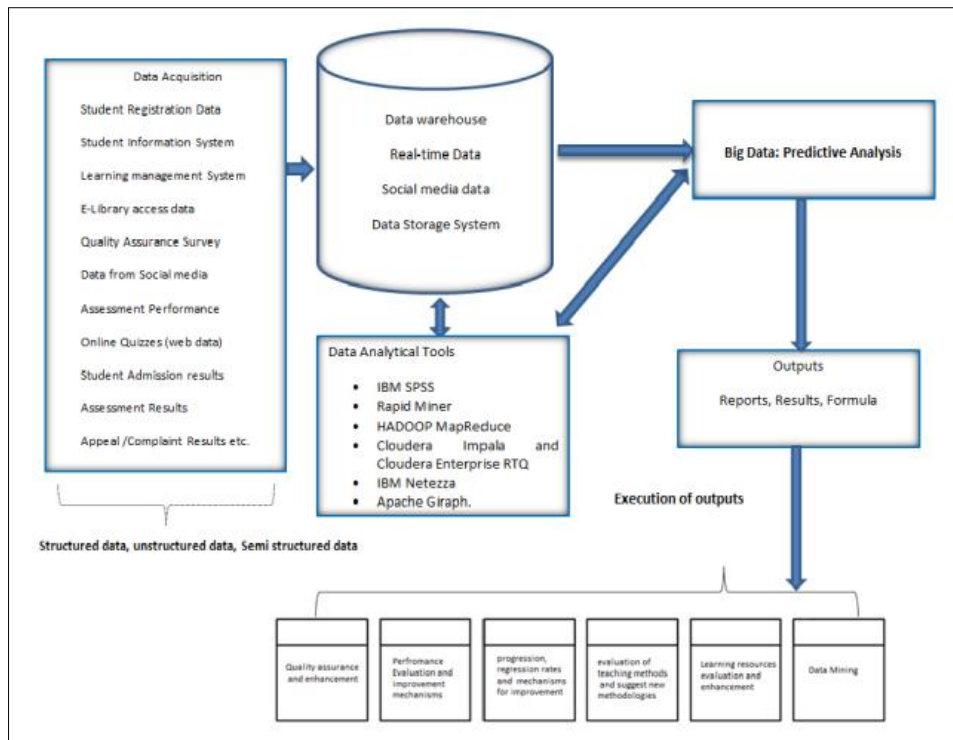


Fig 6: Suggested Framework for Big Data Analysis in Higher Education Sector

Some of the suggested Data Analytical Tools are IBM SPSS, Rapid Miner, HADOOP MapReduce, Cloudera Impala and Cloudera Enterprise RTQ, IBM Netezza and Apache Giraph. Since the most appropriate type in this context is the predictive analysis the big data can be analysed through predictive analysis. The outputs can be in the form of Reports, Results or Formula. The execution of the outputs can be used in various areas. Quality assurance and enhancement, evaluate performance, identifying the progression and regression rates and also to identify improvement mechanisms. The analyzed data can be used to study the impact of the adopted teaching methodologies and based on that new methods can be suggested for further enhancement.

Currently adopted learning resources can be evaluated and further suggestion for improvement can be identified, if necessary. Since big data concerns large-volume, complex, growing data sets from various sources as identified above, big data can be regarded as an emerging trend and the need for big data mining is evolving in all business domains. [5] With the support of high performance computing platforms big data can be used for data mining purpose as well. The major objective of data mining is to build a model and the model builders will extract knowledge from historic data and represent it in such a form that the developed model can be applied to new situations.

7. SUMMARY AND CONCLUSION

After comparing attributes of Higher Education sectors' data with the four Big Data attributes volume, velocity, variety and veracity it is concluded that the higher education data can be recognized as big data. A framework for Big Data Analysis in higher education sector has been suggested after modeling the big data management architecture. Big Data Analytics is a trend that will increase substantially in the coming period and will have a large impact on any education institution due to the many advantages. Traditional face-to-face instruction can support traditional data-driven decision-making processes, however, to move into the more extensive and especially

time-sensitive learning analytics applications, it is important that instructional transactions are collected as they occur. Institutions should identify appropriate big data analytics tool to harvest the benefits from this huge amount of data and how these tools can be used in a way to take decisions and drive the institution towards benefiting from the data.

8. REFERENCES

- [1] Ali, L., Adasi, M., Gasevic, D., Jovanovic, J. & Hatala, M. (2013). Factors influencing beliefs for adoption of a learning analytics tool: an empirical study. *Computers & Education*, 62, 130–148.
- [2] Agrawal, D., Bernstein, P., Bertino, E., Davidson, S., Dayal, U., Franklin, M., Widom, J. (2012). *Challenges and Opportunities with Big Data*. Retrieved from <http://cra.org/ccc/docs/init/bigdatawhitepaper.pdf>
- [3] Boyd, D. & Crawford, K. (2012). Critical questions for Big Data. *Communication & Society*, 15, 5, 662–679. doi: 10.1080/1369118X.2012.678878.
- [4] Bichsel, J., *Analytics in higher education: Benefits, barriers, progress*, 2012, and recommendations
- [5] D. Nettleton, *Commercial Data Mining, Processing, Analyzing and Modeling for Predictive Analytics Projects*, Amsterdam: Elsevier, Morgan Kaufmann, 2014.
- [6] Davenport, Thomas. 2012. "Enterprise Analytics Optimize Performance, Process and Decisions through Big Data." FT Press: 30-45.
- [7] E. Slack, "Storage infrastructures for big data workflows," Storage Switchland, LLC, Tech. Rep., 2012.
- [8] Heudecker, Nick. 2013. "Hype Cycle for Big Data." Gartner G00252431
- [9] <http://www.articlesbase.com/strategic-planning-articles/predictiveanalytics-1704860.html>

- [10] IBM Software Group. Analytics for achievement. Ottawa, Ontario, 2001.<http://public.dhe.ibm.com/common/ssi/ecm/en/ytw03149caen/YTW03149CAEN.PDF>.
- [11] Koedinger, K.R., Baker, R.S.J.d., Cunningham, K.,Skogsholm, A., Leber, B., Stamper, J. (2010) A DataRepository for the EDM community: The PSLCDataShop.In Romero, C., Ventura, S., Pechenizkiy, M., Baker, R.S.J.d.(Eds.)Handbook of Educational Data Mining. Boca Raton,FL: CRC Press, pp.43-56.
- [12] Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013). Big Data: Issues and Challenges Moving Forward.
- [13] M. R. Wigan and R. Clarke, —Big Data’s Big Unintended Consequences!, IEEE Computer Society, DOI:<http://dx.doi.org/10.1109/MC.2013.195>, vol. 46, no. 6, (2013), pp. 46-53.
- [14] M. A. Beyer and D. Laney, “The importance of ’big data’: A definition,” Gartner, Tech. Rep., 2012.
- [15] Mayer, M. (2009) Innovation at Google: The physics of data [PARC forum] (11 August, 2009: 3:59 mark). Available from<http://www.slideshare.net/PARCInc/innovation-at-googlethe-physics-of-data>
- [16] Macfadyen, L. P. & Dawson, S. (2012). Numbers are not enough. Why e-learning analytics failed to inform an institutional strategic plan. *Educational Technology & Society*, 15, 3, 149–163.
- [17] M Zaman, Predictive analytics; the future of business intelligencewww.mahmoudyoussef.com
- [18] S. Lohr, “The age of big data,” New York Times, vol. 11, 2012
- [19] *System Sciences (HICSS), 2013 46th Hawaii International Conference on* (pp. 995-1004). <http://dx.doi.org/10.1109/HICSS.2013.645>
- [20] S. Erevelles, N. Fukawa, L. Swayne, Big Data consumer analytics and the transformation of marketing, *Journal of Business Research*, Vol. 69, No. 2, 2016, pp. 897-904.
- [21] Russom, P.: Big Data Analytics. In: TDWI Best Practices Report, pp. 1–40 (2011)
- [22] Singh, S., & Singh, N. (2012). Big Data analytics. *Communication, Information & Computing Technology (ICCICT), 2012 International Conference on* (pp. 1-4). <http://dx.doi.org/10.1109/ICCICT.2012.6398180>
- [23] T. Erl, W. Khattak, P. Buhler, Big Data Fundamentals: Concepts, Drivers & Techniques, Boston: Prentice Hall, 2016.