

Big Data Analytics in King Khalid University

¹Mrs. Arshiya Begum, ²Mrs. Asfia Sabahath, ³Mrs. Syeda Meraj Bilfaqih
¹Assistant Professor, ²Assistant Professor, ³Assistant professor
¹Computer Science Department,
¹College of Science, King Khalid University, Abha, KSA.

Abstract: As we know big data becomes the important part of big organization where computing, storing and analyzing data comes from a variety of new sources, including social media, machines, log files, video, text, images RFID and GPS. These sources have changed the traditional data base management system and login into new applications.

Big data has become the new concept in the world of economic and societal changes. The world's data collection has reached the pinnacle for major changes in the technology to bring the new ways in decision making, managing health, finance and education. It has also put the challenges on the design of highly scalable algorithms and to collect the data and uncover large hidden data from the datasets that are complex and massive. Big data can be considered as a new generation of decision support data management. If the organization wants to work on big data analytics, it is necessary to have clear idea about business and the relationship between business and its strategies, also a good data infrastructure and the people who are skilled in analytic and analytical tools are required.

This paper aims to show how big data is used on E-learning platform in King Khalid University also it discusses various characteristics and application of big data.

Keywords: Big data; Analytics; decision making; Blackboard.

I. INTRODUCTION

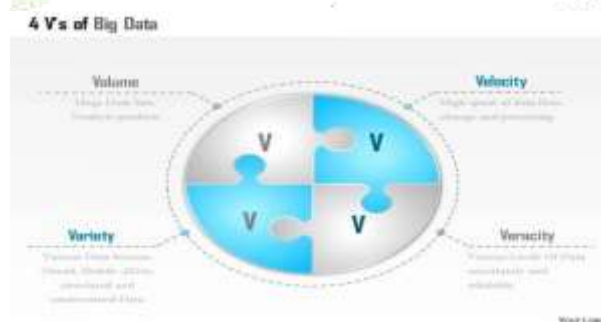
Today most of the companies are collecting; storing, analyzing and transferring huge amount of data, this data is known as “**Big Data**” as it has more data in volume and different forms of data. Once, the data has been stored then it must be analyzed, the analyzed data then used by decision maker and organizations for further processing [1].

Term Bid Data is applied to datasets that grow large and becomes difficult to work with traditional database management system.

Additionally data sets that has being capture, store and manage size is beyond the ability of commonly used software tools and storage system within a tolerable elapsed time [9].

The main feature of big data is volume, variety and velocity or three V's. Where volume is the size of data, Velocity refers to rate at which the data is changing or how often it is created and finally variety include different types and formats of data as well as different ways of analyzing data[10].

Some of the researcher and organization such as IBM scientist [2] break big data into 4 V's where 4th V is referred as Veracity which focuses on the quality of data [11].



.Figure 1: 4V's of big data

The primary attribute of big data is Data volume where it becomes really big when it is collected from variety of sources, using these sources for analytics means structured data is joined with unstructured data. Furthermore multidimensional data is drawn from data ware house. Thus big data variety is as big as its volume [12].

Finally Velocity or speed of big data is the frequency of data delivery or streaming data which is collected real-time from websites [12].

With the evolution of technology and increase in flow of data in and out of the organization, needs faster and efficient way of analyzing data, for such data it no longer suits traditional data management. Therefore new tools and specialized big data analytics is required for storing and managing data.

As the organization grows there will be necessity of expanding the database. The market will demand a platform that helps data to be secured and provide the information to end users so that the end users can analyze the data.

II. BIG DATA LIFE CYCLE

In order to provide a frame work to organize the work needed by organization and deliver insides from big data, it is useful to think of it as a cycle with different stages.

As shown in the figure 2, big data analytics Life Cycle describe the following stages [3]:



Figure 2:Big Data Life Cycle

A. Data Identification:

The data identification stage identifies the datasets required for the project analysis. Depending on scope of analysis project, the datasets can be internal or external to the enterprise [3].

Internal datasets is a list of datasets from internal sources such as data marts and operational system, Whereas External datasets is a list of datasets that are compiled from the third party datasets.

B. Data Acquisition and Filtering:

Data is gathered from different resources. The acquired data is then filtered for the removal of corrupted data or the data that has no values to the analysis objectives. In many of the cases external or unstructured data, is discarded in the filtering process [3].

The data which is filtered may possibly use for different types of analysis.

C. Data Extraction:

Some of the input data as it arrives is in incompatible format with big data solution. In Data Extraction lifecycle stage, the data is dedicated to extract and transform into the format that the big data can use [3].

D. Data Validation and Cleansing:

The data validation and cleansing defines complex validation rules and remove invalid data. Big data may have redundant data, this can be exploited to assemble parameter and fill missing data.

E. Data Aggregation and Representation:

Data is spread over multiple datasets, stages and have common fields. These common fields used in different data sets. In this stage multiple data sets is integrated together to a unified view.

F. Data Analysis:

The data analysis dedicated to carry out analysis task which involve one or more analytics. This stage is iterative in nature where analysis is repeated until appropriate pattern is achieved.

G. Data Visualizations:

It is dedicated to use visualization techniques and tools to graphically communicate the analysis result. As a result of completing data visualization provides user visual analysis.

H. Utilization of Analysis Results:

This determines how and where analytic data can be furthered leveraged.

III. BIG DATA ANALYTICS

Big data sizes are constantly increasing, ranging from terabytes to petabytes in a single data set, as the data grows becomes difficult to capture, store, search, share and visualize data.

Big data analytics is the use of advanced analytics techniques against datasets that are structured, semi structured and unstructured from different sources and sizes [4, 8]. Analyzing big data allows researcher and decision makers to make better and faster decision of data that were previously not in use [8].

Data Analytic is the process of applying algorithm in order to analyze and extract useful information, pattern and relationship from data. Also it detects hidden pattern and important relationship among the variables [12]. Big data is important for decision maker as the data is collected from multiple sources allowing the decision makers to use.

Consequently [13] developed the frame work which includes Big-Data, Analytics and Decision with tools and techniques into decision making. Dealing with such framework enhances the decision making process.

First phase of decision maker is intelligent phase where it collect, store, organized, prepared and processed the data. Next phase is design where possible actions are developed and analyzed. Last phase is implementation phase where proposed solutions from design phased are implemented [13].

In addition to phases mentioned above there are different types of analytics technique which are used on data set such as data mining, natural language processing, predictive analytics, statistics and machine learning.

There are four types of big data analytics [5].

A. Prescriptive: What action should be taken reveals by this analysis and also results in rules and recommendation for next steps.

B. Predictive: Use big data to identify the past pattern to predict the future.

C. Diagnostic: It determines what happened and why.

D. Descriptive: Based on the incoming data it determines what is happening.

Some of the industries which use big data analytics are

Public Sector Services

Health Contribution

Learning Services

Insurance Services

Industrialized and Natural Resources

Transportation Services

Banking Sectors and Fraud detection

IV. REAL TIME APPLICATION OF BIG DATA ANALYTICS

Big Data Contribution on E-Learning platform in King Khalid University:

Big data has a great influence in education world. In this paper we are going to discuss about the usage of big data in King Khalid University for E-learning. In King Khalid University most of the courses present online. Application named as Blackboard allows the teacher to post announcements, Lectures and formative assessment through blackboard. Black board is a segment which is designed to store different types of data to have easy access. Figure 3 shows overview of blackboard used in KKU.



Figure 3: Blackboard

The main feature of blackboard is virtual classroom, where the sessions can be recorded or can be taken online.

There are many course tools available such as blogs, journal, safe assignments, tasks, surveys and pools. Another tool of blackboard is customization, where faculty can customize and styles the blackboard.

Another feature of blackboard is Grade Centre which consist marks of homework and tests. The benefits of blackboard analytics just not end for few features but also elaborate the learning management system and track student and faculty activity in real time.

Tracking student activity is like checking their attendance or getting the information regarding grades, subsequently this tracking make the work easier for the faculty as well as student.

Big Educational data where big data is treated in different ways to bring light on decision-maker, academic faculty, researcher, student to encourage data driven activities and improve quality improvement[6].

As Big data is context specific and contains different sizes and types of data. The Educational data are of different variety such as learning objectives, syllabus, learning material, examination, result and course evaluation. To this data need a special techniques to be applied which manipulates big educational data and helps in student performance and learning approach.

Further big data and analytics helps in promoting actions in educational world, which prevent risk of student failure by identifying them in a very early stage. Also it helps in redesign or improves curriculum and its contents.

Big data in Quality Improvement:

Quality improvement is an effort to make changes which leads to better outcome. It involves quality of student, instructor, curriculum, performance and evaluation of students. Quality is a base line in educational institution by which certain set of standards can be achieved [7].

Big data analytics helps in quality to attain skills of students, behavior and their standards. Also helps in the creation of digital resources like digital library which help both students and teacher to access material from any place.

V. CONCLUSION

In many ways data analytics and distributed learning depends on each other. In data analytics more data is collected from different sources and distributed learning generate more data on student learning achievement which can be added to analytics. In this way data which is collected can be tracked and measured for success.

Money matters are an important aspect in investing in learning management system. As the number of institution grows creating hybrid or online class is innovative technique.

Big data help in gathering information from the students to predict which student need assistance. Most of the institution has made big data as priority in using technological system.

Innovation have speed up the process in today's world where data is combined from a wide variety of sources as a result convergence of cloud and big data create new opportunity for self-service analytics, which helps in visualize any type of data stored anywhere.

References:

- [1] <https://aisel.aisnet.org/cgi/viewcontent.cgi?article=3785&context=cais>
- [2] <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>.
- [3] Big data adoption and planning Consideration By Thomas Erl, Paul Buhler, Wajid Khattak chapter Big Data Fundamentals: Concepts, Drivers and Techniques. Pearson Publication Feb 8, 2016.
- [4] Elgendy, N.: Big Data Analytics in Support of the Decision Making Process. MSc Thesis, German University in Cairo, p. 164 (2013)
- [5] <http://www.ingrammicroadvisor.com/data-center/four-types-of-big-data-analytics-and-examples-of-th-eir-use>.
- [6] Big Data on Real world Application Edited by Sebastian Ventura Soto, José M. Luna and Alberto Cano, ISBN 978-953-51-2490-0, Print ISBN 978-953-51-2489-4, 122 pages, Publisher: InTech, Chapters published July 20, 2016 under CC BY 3.0 license Introduction to big data in educational and its

contribution.

- [7] Introduction to bigdata in Educationn and its contribution to the quality improvement processes
- [8] <https://www.ibm.com/analytics/hadoop/big-data-analytics>
- [9] Kubick, W.R.: Big Data, Information and Meaning. In: Clinical Trial Insights, pp. 26–28 (2012)
- [10] EMC: Data Science and Big Data Analytics. In: EMC Education Services, pp. 1–508 (2012)
- [11] TechAmerica: Demystifying Big Data: A Practical Guide to Transforming the Business of Government.

In:

TechAmerica Reports, pp. 1–40 (2012)

- [12] Advances in Data Mining. Applications and Theoretical Aspects pp 214-22

- [13] Elgendy, N.: Big Data Analytics in Support of the Decision Making Process. MSc Thesis, German University in Cairo, p. 164 (2013)

FirstAuthor Arshiya Begum completed M.Tech (CS) from JNTUH, India.

(email:arshiyabegum@ymail.com)

Second Author : Asfia Sabahath completed M.Sc. (CS) from Osmania University, India

(email:asfia.saba@gmail.com)

Third Author : Syeda Meraj Bilfagih completed MCA from Osmania University, India

(email:syedameraj@gmail.com)

