Higher Education Studies; Vol. 10, No. 4; 2020
ISSN 1925-4741 E-ISSN 1925-475X
Published by Canadian Center of Science and Education

Design on Big data Platform-based in Higher Education Institute

Sajeewan Pratsri[1] & Prachyanun Nilsook[2]

[1] Thepsatri Rajabhat University, Lopburi, Thailand

[2] King Mongkut's University of Technology North Bangkok, Thailand

Correspondence: Sajeewan Pratsri, Thepsatri Rajabhat University, Lopburi, Thailand. E-mail: sajeewan.p@lawasri.tru.ac.th

Received: August 15, 2020 Accepted: September 30, 2020 Online Published: October 8, 2020

doi:10.5539/hes.v10n4p36 URL: https://doi.org/10.5539/hes.v10n4p36

Abstract

According to a continuously increasing amount of information in all aspects whether the sources are retrieved from an internal or external organization, a platform should be provided for the automation of whole processes in the collection, storage, and processing of Big Data. The tool for creating Big Data is a Big Data challenge. Furthermore, the security and privacy of Big Data and Big Data analysis in organizations, government agencies, and educational institutions also have an impact on the aspect of designing a Big Data platform for higher education institute (HEi). It is a digital learning platform that is an online instruction and the use of digital media for educational reform including a module provides information on functions of various modules between computers and humans. 1) Big Data architecture is a framework for an architecture of numerous data which consisting of Big Data Infrastructure (BDI), Data Storage (Cloud-based), processing of a computer system that uses all parts of computer resources for optimal efficiency (High-Performance Computing: HPC), a network system to detect the target device network. Thereafter, according to Hadoop's tools and techniques, when Big Data was introduced with Hadoop's tools and techniques, the benefits of the Big Data platform would provide desired data analysis by retrieving existing information, to illustrate, student information and teaching information that is large amounts of information to adopt for accurate forecasting.

Keywords: big data, platform-based, educational institute, higher education institute

1. Introduction

As in the current situation, Covid-19 pandemic affects all educational societies as well. To proceed the education, educational institutions must adapt to the situation instantly. This current situation has resulted in an unprecedented drive for online learning that many people, including the digital learning platform providers, offer support for the education, namely, to formulate additional social problems for education to resolve. Hence, this is a critical moment to reflect on how alternative education institutions today are impacted in this circumstance of Covid-19 and in term of online learning. Whether they complement the capitalist's view as an educational tool or promote human growth or not.

Use of information, the capacity of information, and content generated by an individual organization, the development of communication technology such as internet technology and information technology for example various types of electronic services while the Big Data industry evolved from these phenomena. (Beyadar et al., 2017) Stated information is so extensive that operations such as collection, storage, and analysis cannot be performed by conventional software and data management tools. A comprehensive problem in higher education institutions around the world is academic success and student retention. As higher education institutions collect cumulative students' data moreover, as the database of a student record becomes more complex and accessible thus, we are entering a new era of information that used to improve student success, improve processes, and use resources more efficiently, better data analysis and student selection processes, more accurate enrollment predictions, and an early warning system that identifies and helps students at risk from studying.

The framework and architecture of data analysis have been used and applied in many information systems research studies (IS). The causal mechanism remains unspecified through the use of scientific research methods. The study design combines various analytical frameworks to develop more comprehensive Big Data architecture for analytical learning. The survey was conducted to review similar and different opinions from different people with different views and interests. This practice is to coordinate the accuracy and reliability of the research

results.

In this Research, we will describe the nature and architecture of Big Data on higher education which presented the tools and techniques of Hadoop that will be useful to organize and analyze the data that need to apply the effective tools for Big Data processing.

2. Big Data

2.1 Definition of Big Data

Big Data is the enormous amount of data that exists in a single type of organization whether the source is from an internal or external organization. Songsangyos & Nilsook, (2015) said that the processing of a common database system to support the enormous amount of data, data rates increase rapidly and is in the form of structure or semi-structure which cannot store data in the database. This is in accordance with (Cravero, 2018) found that a new generation of technology and architecture which designed to extract the value of the gigantic volume of data from a wide variety of sources by enabling high-speed discovery or analysis. Murumba & Micheni, (2017) also mentioned that Big Data in higher education has technological innovation and development of the data storage analysis and cloud computing, combined with the growing ownership of digital devices, education users can collect, manage, and maintain massive amounts of data for the benefit of driving future institutional strategies and policy determination to be available in a complex platform (Moreno et al., 2019) Moreover, there is a wide range of technologies related to privacy and security (Altaye & Nixon, 2019) In summary, Big Data is a large amount of data that is collected and managed as databases, such as business, social media, events, photos, videos, sensors, emails, text files, and applications which all are kept as a huge source of Big Data that can be explained by a variety of speeds, volumes, and nature. To extract useful information from Big Data, excellent processing with analytical capability is necessary. Big Data is divided into structured, unstructured, and semi-structured data such as content, photos, and comments.

2.2 Characteristics of Big Data

(Daniel, 2014; Songsangyos & Nilsook, 2015) stated that the 5V characteristics of Big Data consist of Volume characteristic which is the amount of data should be sufficient when it is analyzed, it will gain insights that correspond to reality, for example, we have information about the age and gender of most customers that enabling us t accurately find information of the general demographic of customers. On the other hand, if we have only a small portion of customer information, the result value may not be accurate. Velocity characteristic, the information is generated rapidly, continuously, and modernly, it allows us to analyze results for decision making and respond promptly, furthermore, we will be able to manage Big Data in order to successfully manage Knowledge Management (KM). Variety characteristics, the variety of information, and types of information. Both structured and unstructured data are table data stored in the database in which structured data consists of numeric data, letter data, date, etc. and unstructured data includes information, images, audio, video, and comments via social media.

(Kumar & Singh, 2019; Hariri et al., 2019; Altaye & Nixon, 2019; Rizk et al., 2019; Al-Barashdi & Al-Karousi, 2019) presented that Big Data requires high velocity and variety so the Veracity of data should have the reliability of the data source and the accuracy of the dataset with a process for checking and confirming the accuracy of the information which is directly related to the results of data analysis. To maintain accurate data without any duplication of the dataset is certainly the hardest and the most time-consuming. Thus, it is considered as the most important feature in generating Big Data. Value characteristic is the valuable data that can be useful or important to business use such as data analysis for summary and data analysis for business planning to create product value or increase the competitiveness in the market of the target product. Summarized as shown in Figure 1.

Figure 1. Characteristics of Big Data

2.3 Big Data Architecture Framework

To emphasize about Big Data architecture framework, (Songsangyos & Nilsook, 2015) suggested that the components of Big Data Architecture, consisting of first, data model structure and type which is the relationship / non-related data model and the file system. Later, the management of Big Data is to conduct a Big Data lifecycle, Big Data transition, or Big Data state, sources, and storage. Afterward, Big Data analysis and tool is a method of application and consideration by using Big Data The objective of presentation and visualization is to achieve Big Data Infrastructure (BDI) and data storage (Cloud-based), processing of a computer system that uses all parts of computer resources for optimal efficiency (High-Performance Computing: HPC), a network system to detect the target device network or Big Data transferred operation, and support operation. More importantly, Big Data should be secure and private namely data security and privacy while stopping mobile and reliable processing environments. Summarized as shown in Figure 2.

Figure 2. Big Data Architecture Framework

2.4 Security and Privacy in Big Data

To describe security and privacy in Big Data, (Jain et al., 2019) said that initially, the security assessment in the distribution framework which results in the optimal safety practices for unrelated resources. Then, the security of data sources and records of changes must be prepared in order to actualize real-time performance check whether how the data endpoint is and analyze various privacy information. Besides, the accessibility of information should be regulated by determining access rights for secured communication. Finally, data sources and performance should be reviewed in detail.

2.5 Big Data in Higher Education

To review Big Data in higher education, (Altaye & Nixon, 2019; Murumba & Micheni, 2017) said that first, it is Predictive Analytics, using Data Mining techniques can benefit a predication of student behavior, analysis of activities performed by students as they interact with the Learning Management System, predicting student performance based on the activities or the measure to be taken to improve student performance. Next, Behavior Detection is to describe students' faces, expressions after school by participation activities, based on player movement in games, and modeling knowledge and understanding. He said that this type of modeling helps to understand the learning process of users who interact with the system and adapting the learning environment for users to adopt the results of the behavior examination to predict risks (Risk Prediction). A technique that uses Big Data to predict the risks involved in students. Some students abandoned the course, their activity was monitored and the engagement score was predicted and using historical data to create a student behavior model and a model used to calculate risk by applying a students' skill estimation to transform the education system to suit students'

skills.

Skills are calculated based on students' interaction with the system or on a message board or forums in order to carry out financial planning, check student performance. Finally, an intelligent teaching system will be formed with a teaching model and games that can be used to generate opportunities to collect and analyze student data from discoverable patterns and trends to support the interaction between humans and information technology environment. From the above, it can be summarized as in Figure 3.

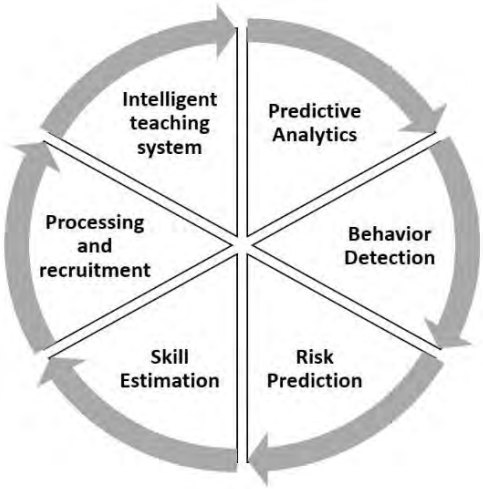

Figure 3. Big Data in Higher Education

3. Hadoop's Tools and Techniques for Big Data

Hadoop is open-source software that is built as a data storage platform. It provides a framework for storing and processing enormous data called Big Data. Hadoop is scalable and able to handle large amounts of data. (Kumar & Singh, 2019) showed that the Hadoop platform was developed to record, organize, and analyze data, effective tools are needed to separate meaningful output from Big Data. Various tools used in the processing of Big Data are detailed in Table 1.

Table 1. Hadoop's Tools and Techniques

Hadoop's Tools and Techniques (Rizk et al., 2019; Kumar & Singh, 2019; Cravero, 2018; Cravero et al., 2018)	
Apache Hive	Data storage's built on Hadoop to simplify a Big Data framework to make it easier to access the information you need. Due to Hadoop's search capability is limited and the complexity of the MapReduce framework, developers must write a complex program that may be difficult to maintain and use even apply it for simple analysis. (Rizk et al., 2019)
Apache Hadoop	An open-source software project for building a highly stable and extendable distributed computing system.
Hadoop Distributed File System (HDFS)	The primary storage system used in the Hadoop HDFS software creates a data block model in a cluster for reliability and fast calculation.
MapReduce	Programming framework helps processing data with multiple datasets running simultaneously which will have to rely on multiple computers to cooperate.
Apache Pig	Tool likes Hive that allows data processing without Map / Reduce Pig programming, using a simple scripting program called Pig Latin instead of Pig which is suitable for ETL for data conversion in various formats.
Apache HBase	A tool that allows Hadoop to read and write data in Real-Time Random Access. It will be a large table that can store unlimited data in rows or columns. HBase is compared to making Hadoop be a NoSQL database.
Apache Oozie	A workflow tool that will allow us to integrate Hadoop system processing instructions such as Map / Reduce, Hive, or Pig into a workflow.
Apache Avro	A framework for permanent data serialization and remote procedure calls between Hadoop nodes and between client programs and Hadoop services.
Apache Zookeeper	A centralized system used by applications to maintain systems including organizing other elements between nodes.
Apache Yarn	The YARN distributed application has two components; the Resource Manager (RM) that manages all the resources within the cluster needed for the work, and the Node Manager (NM) that resides on every host in the cluster and manages the available resources independently.
Apache Sqoop	Tool for transferring data between tables in an RDBMS database format such as SQL Server, Oracle, or MySQL and HDFS data by Hadoop.
Apache Flume	Tool for retrieving data from other systems in real-time into HDFS, such as retrieving logs from a web server. To retrieve these data, the Agent must be installed on the server.

4. Benefits of Big Data

Due to the adoption of Big Data, (Murumba & Micheni, 2017; Songsangyos & Nilsook, 2015) mentioned that Big Data is helpful on-demand data analysis by retrieving existing data such as student data, teaching data. This huge amount of data is adopted for an accurate prediction. Big Data adaptation can create products or improve services to meet customer or users' satisfaction, such as delivering to customers when purchased through the network, moreover, it is a collection of different data sources and types since the data comes from a variety of sources.

5. Platform-Based in Higher Education

The platform used in higher education includes; 1) Social Learning Platforms that use sophisticated and costly data analytics and (Xi 2018) said that it is a traditional analytic tool to provide useful insights while continuing to pressure educational establishments to deal with the generated information. 2) E-learning Platform is applied to facilitate data processing in parallel, which (Dahdouh et al., 2019) purposed that to help to process data in cost calculation by using Spark-based and Hadoop in decentralized computing and data analytics. For validation, the system can handle large amounts of data and scalable compute capability. 3) Digital Learning Platform, the web-based teaching and learning platform, offers a wide range of functions teaching and training design for organizations, and on-site learning space. Additionally, they can provide works, various educational materials, and media for solving learning problems. Furthermore, there is an opportunity to exchange and learn together to participate. (Hartmann et al., 2019; Sousa & Rocha, 2018) added that it is an approach to online learning and

using digital media for educational reform which consists of flexibility, individuality, quality, learning analytics, cost-effectiveness, and flipped learning.

6. Big Data Platform-Based in Higher Education

According to the concept of Big Data and analysis, this can be applied to a variety of higher education in the management and instruction plus improve services to meet the needs of teachers or learners including seeking and processing donor financial planning, and student performance checking and monitoring as shown in Figure 4.

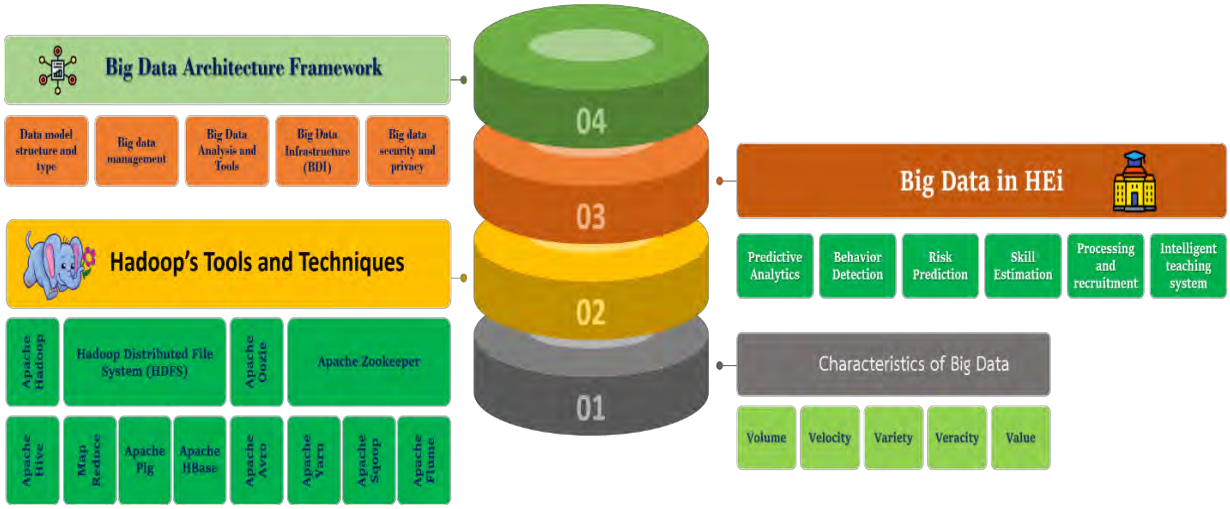

Figure 4. Big Data Platform-Based in Higher Education

According to Figure 4, it shows an overview of the big data platform design in higher education which is a digital learning platform adopted an online learning management and digital media adaptation for a business class educational reform and an information class display module providing information about the functions of the modules between computer and human. The big data architecture is a framework for a large data architecture consisting of; 1) Characteristics of Big Data, 2) Hadoop Tools and Techniques, 3) Higher Education Platforms, and 4) Big Data Architecture Framework.

Characteristics of Big Data (Daniel, 2014; Songsangyos & Nilsook, 2015; Kumar & Singh, 2019; Hariri et al., 2019; Altaye & Nixon, 2019; Rizk et al., 2019; Al-Barashdi & Al-Karousi, 2019) states that the 5V characteristics of big data consists of Volume which is a quantity of accurate data for an analysis, Velocity is the ability of consequence analysis for making a decision and responding promptly to a variety of circumstances. Variety is diversified structured and unstructured data in the form of either RDBMS, text, XML, JSON or Image, Veracity is data with multiple quality and value levels for an analysis, and Value is worthwhile data adoption when receive an accurate data leads to development that can measure success concretely.

Hadoop is an open source software created as a data storage platform which provides a framework for storing and processing enormous data namely Big Data. Hadoop is scalable and flexible to support the enormous amount of data due to Hadoop contains a data processing distributed across computers with clustered format leading to unlimited and reliable data management capability. (Kumar & Singh, 2019) said that the Hadoop platform was developed to record, organize, and analyze data whereupon an effective tool is necessary to distinguish meaningful output from big data which includes Apache Hive, Apache Hadoop, Hadoop Distributed File System (HDFS), MapReduce, Apache Pig, Apache HBase, Apache Oozie, Apache Avro, Apache Zookeeper, Apache Yarn, Apache Sqoop and Apache Flume. Hadoop is not suitable for small and real-time data as Hadoop is a batch processing and structured data storage furthermore, another alternative is much more interesting such as SQL. Hadoop is eligible for universities and companies where utilize Hadoop for web-based data collection since the current data analysis requires external data as consideration variables moreover, the external data is unstructured and has a high-velocity expansion rate, Hadoop is a favorable selection.

Big data in higher education (Altaye & Nixon, 2019; Murumba & Micheni, 2017) is a predictive analysis which assists divide students into groups for actualizing reports or grouping them by categorizing according to the students' characteristics with similar behavior, learning styles or preferences in the same group to enhance the

highest learning efficiency**.** Likewise, Behavior Detection is behavior while studying, such as data source retrieved for work, and the successful percentage of work submitting which expedites to understand the learning process of the user interacting with the system and improving the learning environment for the user causing the results of the behavior detection for risk prediction. Big data technique to predict the risk involved with some students abandoned their course were monitored and the engagement score was predicted. Historical data was applied to create student behavior model and risk measurement model by using student skill estimation to lead the educational information system can assist in arranging a course that is suitable for individual learners. Especially the college and university learning plans which the information can facilitate students to organize and enroll the course that superbly suits them. Finally, an intelligent teaching system will be conducted to provide students insights about how their learning is at each level which each student will have a different learning style. The different learning style affects the academic performance of the course. It has teaching model and valuable games used to create opportunities for analyzing students' data from the human interaction model.

The big data architecture framework (Songsangyos & Nilsook, 2015) addressed that the elements of the big data architecture, including model data and type framework which is a relationship/non-related data model and a file system. Big data management is to manage big data cycle, big data transformation, state of big data, source, and storage. Big data analysis and tool is a method of using and analyzing big data with the goal of presentation and visualization to receive Big Data Infrastructure (BDI), cloud-based storage, and High-Performance Computing (HPC). Network system detects the target device network or operates big data transfer, supportive operation, and most importantly, big data security and privacy should be provided to authorize the person who has an access right, and designate laws and regulations controls on collecting, using, sharing, storing, and transferring information infallibly.

7. Conclusions

Due to an information technology including various tools and techniques at present that are available for the development of Big Data in higher education institutions which enable to update information, teaching and learning process or analyze data processing greatly by applying predictive analytics using Data Mining technique, it can perform advanced and real-time investigations swiftly and cost-effectively. Then, using an intelligent teaching system that can analyze students' data about the interaction during teaching and learning. However, implementing higher education platforms have many challenges, especially it must hold a low level in term of investment and network management systems to provide a wide range of communication and access to information extensively. In this article, we proposed architecture and implementation for the Big Data platform. The concepts that underpinning analysis can be applied to a variety of higher education in term of management and instruction including improvement of services to meet the needs of teachers or learners also enable to receive all the services required for efficient adaptation and deployment in educational institutions.

References

Cravero, A. (2018). *Big Data Architectures and the Internet of Things: A Systematic Mapping Study.* pp. 1219-1226.

Al-Barashdi, H., & Al-Karousi, R. (2019). Big Data in academic libraries: literature review and future research directions. *Journal of Information Studies & Technology, 2018*(2), 1-16. https://doi.org/10.5339/jist.2018.13

Altaye, A. A., & Nixon, J. S. (2019). A Comparative Study on Big Data Applications in Higher Education. *International Journal of Emerging Trends in Engineering Research, 7*(12), 739-745. https://doi.org/10.30534/ijeter/2019/027122019

Beyadar, H., Askari, M., & Askari, A. (2017). A Network platform for creating digital entrepreneurship in cloud environment based on big data. In *Application of Information and Communication Technologies, AICT 2016 - Conference Proceedings*. https://doi.org/10.1109/ICAICT.2016.7991766

Cravero, A., Saldaña, O., Espinosa, R., & Antileo, C. (2018). Big Data Architecture for Water Resources Management: A Systematic Mapping Study. *IEEE LATIN AMERICA TRANSACTIONS, 16*(3), 902-908.

Dahdouh, K., Dakkak, A., Oughdir, L., & Ibriz, A. (2019). Large-scale e-learning recommender system based on Spark and Hadoop. *Journal of Big Data, 6*(1). https://doi.org/10.1186/s40537-019-0169-4

Daniel, B. (2014). Big Data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology, 46*(5), 904-920. https://doi.org/10.1111/bjet.12230

Hariri, R. H., Fredericks, E. M., & Bowers, K. M. (2019). Uncertainty in big data analytics: survey, opportunities, and challenges. *Journal of Big Data, 6*, 44. https://doi.org/10.1186/s40537-019-0206-3

Hartmann, M., Nestler, A., Wohlrabe, D., Arnold2, F., Hoffmann, J., & Wenkler, E. (2019). *Development of a digital learning platform for the planning of manufacturing processes*. IOP Conference Series: Materials Science and Engineering PAPER. https://doi.org/10.1088/1757-899X/564/1/012085

Jain, P., Gyanchandani, M., & Khare, N. (2019). Enhanced Secured Map Reduce layer for Big Data privacy and security. In *Journal of Big Data, 6*, 30. https://doi.org/10.1186/s40537-019-0193-4

Kumar, S., & Singh, M. (2019). Big data analytics for healthcare industry: impact, applications, and tools. *Big Data Mining and Analytics, 2*(1), 48-57. https://doi.org/10.26599/BDMA.2018.9020031

Moreno, J., Fernandez, E. B., Serranoand, M. A., & Fernández-Medina, E. (2019). Secure Development of Big Data Ecosystems. *IEEE Access, 7*, 96604-96619. https://doi.org/10.1109/ACCESS.2019.2929330

Murumba, J., & Micheni, E. (2017). Big Data Analytics in Higher Education: A Review. *The International Journal of Engineering and Science, 6*(6), 14-21. https://doi.org/10.9790/1813-0606021421

Rizk, R., McKeever, S., Petrini, J., & Zeitler, E. (2019). Diftong: a tool for validating big data workflows. *Journal of Big Data, 6*, 41. https://doi.org/10.1186/s40537-019-0204-5

Songsangyos, P., & Nilsook, P. (2015). Big Data in the Cloud for Education Institutions, 32(1).

Sousa, M. J., & Rocha, A. (2018). *DIGITAL LEARNING IN AN OPEN EDUCATION PLATFORM FOR HIGHER EDUCATION STUDENTS*. Proceedings of EDULEARN18 Conference, Palma, Mallorca, Spain. pp. 11194-11198. https://doi.org/10.21125/edulearn.2018.2770

Xi, Y. (2018). *Research on the Construction of Library Data Integration System in Big Data Era*. Proceedings of the 2018 International Conference on Transportation & Logistics, Information & Communication, Smart City. https://doi.org/10.2991/tlicsc-18.2018.18