Big Data in the Cloud for Education Institutions

Pradit Songsangyos[1]
Rajamangala University of Technology Suvarnabhumi,
Phranakhon Si Ayutthaya, Thailand
[1]pradit.s@rmutsb.ac.th

Prachyanun Nilsook[2]
King Mongkut's University of Technology North Bangkok, Thailand
[2]prachyanun@hotmail.com

Abstract - **The collection of data in details and size is increasing continuously, referred to herein as big data. Cloud computing technology provides a platform for automation of all processes in data collecting, storing and processing of the huge data. First, the definition, characteristics, and challenges of big data along with cloud computing are introduced. In addition, the analytics of big data and cloud computing in the enterprises, government, and educational institutes, also impact of them have been concerned and discussed. Followed by the emergence of big data and cloud technologies for educational institutions, to support online curriculums and resources of learning materials, and administration tasks.**

Keywords - **Big Data, Cloud Computing, Educational Institute, Higher Education**

I. INTRODUCTION

Currently, information and communications technology is broadly used in working processes, education, and daily life. It has developed rapidly and continuously. The input data and information to the system and the process outputs are increasing dramatically. It required a large data storage, high performance transmission and processing system. Cloud computing is a flexible service, capable to manage the resources to meet the demand for services and processed rapidly. The expenditure according to workload usage to deal with big data, then support for education institutions.

II. THE CLOUD COMPUTING

A. Definition of Cloud Computing

The National Institute of Standards and Technology (NIST) have provided definitions "The Cloud Computing" as follows:

1. As a service by demand from users and service users to get information upon request.

2. Access to a computer network management, network and software is available from anywhere and at anytime.

3. Many users can share data. The cost depends on location and usage time. Sustainability and the environment of the seller.

4. Allocate resources flexibly according to changes in the demand for services, and can be processed rapidly. Users do not have to worry about allocating resources to the service.

5. The service is measured by the number of users or duration of service [1].

B. Classification of Cloud Computing

The cloud computing system can be divided into four main types as follows:

1. Public Cloud / External Cloud is a public resource. It can disclose information to the public. Service providers are an administrator. Sharing resources and providing basic utilities via the Internet. Web Applications or Web services for users generally.

2. Private Cloud / Internal Cloud user management system itself. By simulating clouds up in a private network. It disclosure to the public data security and reliability levels. It commonly used in the public sector and non-disclosure requirements.

3. Community Cloud formed by a partnership between the organization or company. By agreement of common resources such as hardware, software, security measures the cost of a higher form. It either less public or lower than private

4. Hybrid Cloud occur between public and private providers to disclose some information. Users are mostly private companies [2, 3].

C. Cloud Computing Services

Cloud computing services are flexible for users. Their services can be categorized into 3 groups as follows:

1. Software as a Service (SaaS) has provided a computer application for sharing anywhere, anytime, any device with a computer program that is necessary and appropriate for multiple users, by using a browser to access. SaaS is useful for marketing management, Human Resources Management, and the Enterprise Resource Planning.

2. Platform as a Service (PaaS) is used by the operating system, and supporting which sharing through the web applications. Users can access from any device, software or operating system. In addition, users can use their own applications on the infrastructure provided by service providers.

3. Infrastructure as a Service (IaaS) provides infrastructure services in computing systems, and sharing of requests data centres on demand [1, 3, 4].

III. BIG DATA

A. Definition of Big Data

Dumbill (2012) has defined Big Data as "the amount of data that has a massive excess capacity in the processing of conventional database systems to accommodate. The data volume is very large. Data rates are increasing rapidly and be in the form of unstructured or semi-structured. Which cannot store information in the database [5].

B. Characteristics of Big Data

Big Data is composed of five parts:

1) **Features of Big Data**.

- Volume, which is an enormous quantity of data structure so data can not store in the database system.

- Velocity, refers to the increasing rate of data. The data rate enters into the database is increasing rapidly.

- Variety, refers to a variety of data formats. This could be in format of structured, unstructured, and semi-structured.

2) **A New Data Model**.

- Data types, data structures, Data Link Source of data.

- The data lifecycle, the variance/ evolution.

3) **A New Analysis**.

- Analysis of real-time/ streaming, using new statistical methods. There is interaction, and learning of analysis automatically.

4) **Infrastructure and Tools**.

- Big data infrastructure as a Cloud-based, high-performance of storage, networking, and processing.

- The integration of different services, by various service providers.

- Centralized information services (Multi-stakeholder).

- Centralized information security. The

reliability of infrastructure, data processing, and storage.

5) **The importance source and target of data types and data structures, such as raw data, data stream, data relations, and required data processing methods**.

- Fast data collection, by a various sensors and sources.

- Transmission of data to visualize and operate in the different systems and consumers.

- Import and export of data input are fully digital format (prevalent), network detection, and fully digital control [5, 6, 7].

C. Big Data Architecture Framework

The Big Data Architecture Framework (BDAF) consisting of five parts below.

1) **Data Model Structure and Types**.

- Data model, relationship / irrelevant, and file systems, etc.

2) **Management of Big Data**.

- Management of big data lifecycle.
- Transformation / state of big data.
- Source, and storage.

3) **Analysis and Tools of Big Data**.

- How to apply and analyze using big data.

- The goal of application, presentation and visualization.

4) **Big Data Infrastructure (BDI)**.

- Computing and storage infrastructure. (Cloud-based), High-performance Computer (HPC), networking.

- Network detection, target devices/ operation

- Transportation of big data, and operational support.

5) **The Security and Privacy of Big Data**.

- Data security and privacy while halt, mobile, and the reliable processing environment [6].

D. Benefits of Big Data

There are several benefits of big data, there are two reasons obviously.

1. Analytical use, is a data analysis overwhelming to see those hidden knowledge, such as weather information, from many detector on the satellite, radar, balloons, and aircraft. These enormous amounts of information, used to precisely forecast, hourly and so on.

2. Enabling new products, big data can be used to create the products or improvement of services. To meet the needs of customers or users, such as shipping to the customer when they are buying through the network, etc [5].

IV. BIG DATA IN THE CLOUD FOR EDUCATION INSTITUTIONS

The concept of big data and analytics can be applied to a variety of higher education in the management and use of teaching and learning, including the recruitment and processing, admissions, financial planning of donors, track and monitor student performance.

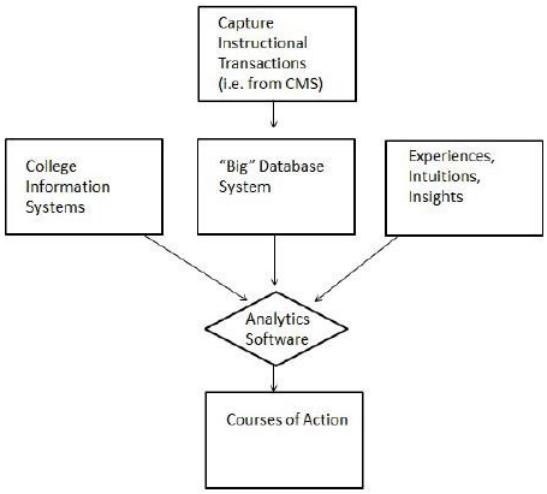

Fig 1. Learning Analytic Flow Model (Adopted from [8])

In a white paper published by IBM, explained about the types of applications for teaching as follows:

1. Monitoring the performance of individual students.

2. Student performance selected based on key characteristics such as race, class, education and so on.

3. A discussion of the unusual.

4. Predict the potential to make all students successful at best.

5. To prevent the degradation of the educational curriculum or program.

6. Identify and develop effective teaching techniques.

7. Analysis of assessment techniques, standards and tools.

8. Test and Evaluation Program.

There are also a number of concerns about this emerging technology needs to be considered as follows:

- First Big data and analytics to learn to be applied to the appropriate information. It must be accurate and precise. This will provide the best results for the courses are delivered electronically.

- Second, the lack of personnel trained to use it properly big data and analytics. Database administrators need to have the experience and capabilities in the design and integration of data warehousing and data file formats.

- Third, there is a very serious concern. Since the analysis of learning that require large amounts of data gathered from students, and integration with other databases to be careful about privacy and the privacy of student records in terms of personal behavior.

- The latest addition to the management of the university. University student data may be stored in a huge database of government at either the school district or national level.

- Finally, the theory of Herbert Simon about the scope of a rational decision. When the database is larger and more sophisticated analysis. As said, "The wealth of data may make it more interesting," [8].

The design features of the platform for the construction of information resources for teaching morals in schools and universities consist of three layers with different display shelf. Business class module and data layer floor display to display information on the functions of different modules interaction between computers and human beings. Business class module responsible for the management and operation of the information flow and the flow of business rules to meet the needs of all functions of the module. Layer responsible for the storage and management of the school district. The school effectiveness [9].

REFERENCES

(Arranged in the order of citation in the same fashion as the case of Footnotes.)

[1] Meesuwan, W. (2014). "Cloud Computing System in Education". Journal of Education, Naresuan University.

[2] Kummanee, S. "Security of Cloud Computing Technology". <http://www.resjournal.kku.ac.th/abstract/18_2_4.pdf>. Accessed 19 September 2014.

[3] CISCO. (2009). "Cloud Computing in Higher Education: A Guide to Evaluation and Adoption". White Paper.

[4] Maorapong, W. (2010). "IT Outsourcing Service on Cloud Computing". Journal of TPA News.

[5] Koochaiyasit, S. (2013). "The New World of Data Era, Big Data". Journal of Management, pp. 22-28.

[6] Demchenko, Y. and et al. (2014). "Model for Building effective Big Data Curricula for Online and Campus Education". Proceeding of 2014 IEEE 6th International Conference on Cloud Computing Technology and Science, pp. 935-941.

[7] Richard, J. (2014). "Self Governance Strategies for the Cloud, big Data and Other Technologies in Education".

Proceeding of IEEE/ACM International Conference on Unity and Cloud Computing, pp. 630-635.

[8] Anthony, G.P. (2014). "The Evolution of Big Data and Learning Analytics in American Higher Education". Journal of Asynchronous Learning Networks, Vol. 16, Issue 3, pp. 9-20.

[9] Johnson, S. and et al. (2014). "Informatization Construction of Moral Education Curriculum Resources in Schools and Universities". Proceeding of IEEE Workshop on Electronics, Computer and Applications, pp. 594-691.