Mesopotamian journal of Big Data
Vol. **(2021)** 2021, **pp**. 21–24
DOI: https://doi.org/10.58496/MJBD/2021/004 ISSN: 2958-6453
https://mesopotamian.press/journals/index.php/BigData

Review Article

Big data analysis

Rana Talib Rasheed,*[1], Mostafa Abdulgafoor Mohammed[2] , , Nicolae Tapus,[2],

[1] *Alsalam University college, Iraq*

[2] *Department of Computer Science, University Politehnica of Bucharest, Romania*

ARTICLE INFO

Article History

Received 30 Dec 2020
Accepted 22 Feb 2021
Published 02 May 2021

Keywords

Big data

Big data analysis

Cloud computing

Hadoop

ABSTRACT

"Big Data" can be considered as an expression which has spread suddenly. It could be depicted as a pioneering mechanism for saving, distributing, managing, visualizing and analyzing large data along with very high velocity and approaches for managing incompetent quantity of data. Here, the state-of-the-art and background of big data are reviewed. Every review-related technology and general background of big data are firstly reviewed, for example IoT, cloud computing, Hadoop and data centers. Then, we concentrate on every value chain phase belongs to big data such as data acquisition, data generation, data analysis and data storage. For the four phases, the general background is presented, the technicalities are discussed, and most recent advances are reviewed. Eventually, the various big data representative applications are examined, containing IoT, enterprise management, medial applications, online social networks, smart grid and collective intelligence. Those argumentations pursue to supply an overall overview for readers interested in this exciting area. The survey here is finalized with an argumentation of future directions and open problems.

1. RELATIONSHIP BETWEEN CLOUD COMPUTING AND BIG DATA

Cloud computing can be connected with big data, and its key components can be illustrated in Figure 1. Big data could be the computation-intensive operation's target, and intensify the capacity of storing for cloud systems. The major goal for such type of computing can be the using of enormous computing and storing resources under intensified management, for providing applications of big data with a capacity of computing that is fine-grained. The enhancing of cloud computing gives answers for the processing and storing of big data. Also, the growth of big data quickens the enhancement of cloud computing. The technology of distributed storing built upon cloud computing has the ability to run big data in an effective way; the parallel capacity of computing according to cloud computing has the ability to develop the acquisition's efficiency as well as the analysis of big data [1].

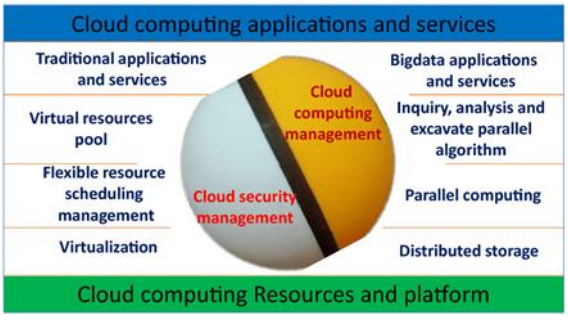

Fig.1. key components of big data

1.1 Relationship between IoT and big data

In the paradigm of the Internet of Things, a huge quantity of networking sensors can be embedded in several appliances. Sensors of this type spread in various fields might gather different types of data, for example geographical data,

*Corresponding author. Email: rana.talib@alsalam.edu.iq

public facilities are equipment of data acquisition in the Internet of Things as shown in Figure 2. The big data produced by the Internet of Things possesses various qualities if we compare it with the general big data since the various kinds of data gathered, of which most of the classical qualities including variety, heterogeneity, noise, high redundancy and unstructured feature. Even though the recent data of the Internet of Things does not dominate big data, by the year 2030, the sensor amount will have reached a trillion, and the data of the Internet of Things will have been the most significant part in big data, in view of the HP forecast [2]. Intel reported that big data in The Internet of Things possesses three characteristics conforming to the paradigm of big data:

(i) abundant terminals which produce masses of data.

(ii) data produced by the Internet of Things can be semi-structured or unstructured usually.

(iii) data of the Internet of Things can be of use just as it is under analysis.

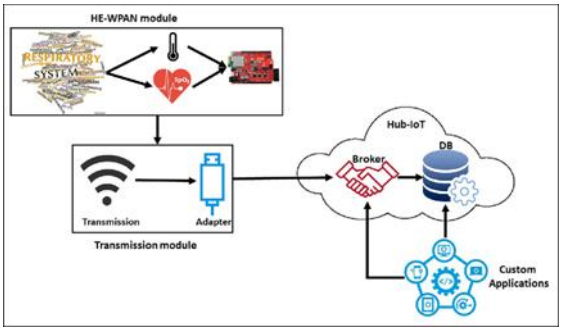

Fig.2. IoT for data acquisition equipment

1.2 Relationship between Hadoop and Big Data

Nowadays, Hadoop can be commonly utilized in every big data application such as network searching, spam filtering, social recommendation and clickstream analysis. Furthermore, Important scientific researches are presently built upon Hadoop. A few representative cases are mentioned below. As announced in 2012, Yahoo operates Hadoop in forty-two thousand servers in four data centers for supporting its services and products such as spam filtering and searching. Nowadays, the largest Hadoop cluster possesses four thousand nodes, but the total is going to reach ten thousand when Hadoop 2.0 is released. Within the same thirty days, Facebook made an announcement that their Hadoop cluster had the ability to process one hundred PB data, that increased by 0.5 PB a day in November 2012. A few famous agencies which employ Hadoop for conducting distributed computation have been listed in. Besides, various firms supply Hadoop commercial executing and/or support such as IBM, Cloudera, and Oracle [3].

Amongst contemporary industrial systems and machinery, sensors can be commonly spread for collecting information for failure forecasting and environment monitoring. A framework for data organizing and cloud computing infrastructure, which utilizes local nodes, remote clusters and mixed architectures built upon Hadoop for analyzing machine-generated data. Local nodes can be utilized for forecasting real-time failures; clusters built upon Hadoop can be utilized to analyze offline such as data analysis and case-driven.

2. OBSTETRICS AND ACQUISITION

Many key technologies are presented relating to big data such as the Internet of Things, cloud computing, Hadoop and data center. After that, it is going to be focused on the big data value chain, that is divided into four phases in general: data acquisition, data generation, data analysis and data storage. In case data is taken as a raw material, data acquisition and data generation can be an exploiting process, data storage can be a storage process, and data analysis can be a producing process which uses the raw material for creating a novel value [4].

Data generation can be considered as the step number one for big data. Considering Internet data as an example, a big quantity of data respecting Internet for upmost, microblog messages and chatting records can be produced. These data can nearly relevant to individuals' lives, and acquire resemblant high value and low-density characteristics. Internet data like this might be without any value, but through exploiting the accumulated big data, handy information i.e., users' hobbies and habits are identified, and we can forecast their emotional mood and behavior. Big data acquisition being the phase number

two in the system of big data, it contains data transmission, data pre-processing and data collection. Through this phase, as soon the raw data is collected, an efficient transmitting technique is going to be used for sending it towards an appropriate system of storage management for supporting various analytical applications. The gathered datasets might occasionally contain a lot of useless or redundant data increasing storage space and affecting the analysis of subsequent data. For instance, high redundancy can be considered widespread within datasets gathered by sensors to monitor environments. The technology of Data compression is applied to decrease the redundancy. So, the processes of data pre-processing can be indispensable for ensuring storing and exploiting efficient data [5].

3. BIG DATA STORAGE

The enormous increase in data possesses more needs on management and storage. Here, we concentrate on storing big data, which indicates storing and managing big datasets as accomplishing the availability and reliability of accessing data. We are going to make a revision to significant issues such as distributed storage systems, massive storage systems and big data storage techniques. On one hand, the infrastructure of storing requires for providing a service of information storage with a storage space; also, it has to supply a strong accessing interface for query and analyzing a big quantity of data.

According to traditions, as helping equipment of server, data storing appliance can be utilized for storing, managing, looking up, and analyzing data with structured RDBMSs. With the big increase of data, data storing appliances are being more significant, and various Internet firms aim to a huge capacity of storing to be under competition. So, there can be a huge requirement for a study concerning data storing [6][7][8].

4. GROWTH OF BIG DATA

Presently, the analyzing of big data is suggested to be a technology that is advanced analytically, that contains big and complicated software under certain analytical approaches. Actually, every data driven application was released in the previous years. For instance, in the beginning of the 20th century, BI became a dominant technology for network search engines and business applications built upon the processing of huge data mining emerged in the beginning of the 2000s.

5. APPLICATION OF BIG DATA IN ENTERPRISES

Nowadays, big data basically arise from and utilized in enterprises, as OLAP and BI could be considered as the big data application predecessors. Applying big data in enterprises could develop the competitiveness and efficiency of their production in various fields. Particularly, in advertising, with big data correlation, every enterprise may predict consumers' attitude and allocate novel business modes. In planning sales, after comparing huge data, every enterprise optimizes its prices of commodity. In operating, every enterprise improves its operation satisfaction and efficiency, optimizes labor force, forecasts individuals' allocating needs, avoids excess producing capacity, and decreases the labor cost. In supply chain, utilizing big data, every enterprise conduct inventory optimizing, logistic optimizing, and supplier coordinating for mitigating the hole between supply and demand, budget controlling, and service improving.

6. CONCLUSION

Here, the state-of-the-art and background of big data is reviewed. First of all, the general background of big data is introduced, and the related technologies are reviewed, for example the Internet of Things, cloud computing, Hadoop and data centers. After that, we concentrate on every phase of the big data value chain, in another word, data acquisition, data generation, data analysis and data storage. For every phase, the general background is introduced, the technical difficulties are discussed, and the recent advances are reviewed. In the end, the many representative applications of big data are reviewed, containing the Internet of Things, enterprise management, medical applications, social networks, smart grid and collective intelligence. Those argumentations pursue to supply an overall overview and big-picture for readers interested in exciting areas. For remembrance, the study hot spots and are summarized and the possible research directions of big data are proposed. Furthermore, possible development trends and application area are discussed.

Conflicts of Interest

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

None.

Funding

Non.

References

[1] J. Srinivas, A. K. Das, and J. J. Rodrigues, "2PBDC: privacy-preserving bigdata collection in cloud environment," The Journal of Supercomputing, vol. 76, no. 7, pp. 4772-4801, 2020.

[2] L. J. Ramírez López, A. Rodriguez Garcia, and G. Puerta Aponte, "Internet of things in healthcare monitoring to enhance acquisition performance of respiratory disorder sensors," International Journal of Distributed Sensor Networks, vol. 15, no. 5, p. 1550147719847127, 2019.

[3] A. Bahga and V. K. Madisetti, "Analyzing massive machine maintenance data in a computing cloud," IEEE Transactions on Parallel and Distributed Systems, vol. 23, no. 10, pp. 1831-1843, 2012.

[4] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," Mobile networks and applications, vol. 19, no. 2, pp. 171-209, 2014.

[5] B. Abu-Salih, P. Wongthongtham, D. Zhu, K. Y. Chan, and A. Rudra, "Social Big Data Analytics," Springer Singapore, 2021.

[6] M. Z. Kastouni and A. A. Lahcen, "Big data analytics in telecommunications: Governance, architecture and use cases," Journal of King Saud University-Computer and Information Sciences, 2020.

[7] M. Beyer, "Gartner says solving big data challenge involves more than just managing volumes of data," Gartner, 2011.

[8] Y. Sun, M. Chen, B. Liu, and S. Mao, "Far: a fault-avoidant routing method for data center networks with regular topology," in Proceedings of ACM/IEEE symposium on architectures for networking and communications systems (ANCS'13), ACM, 2013.