Conference Proceeding Issue Published in International Journal of Trend in Research and Development (IJTRD), ISSN: 2394-9333, www.ijtrd.com

Services Provided by Integration of Big Data and Cloud Computing: A Review

Rakesh. A
Student, M.Sc Computer Science, Indian Academy Degree College, Bangalore, India

Abstract: Big Data and Cloud Computing are the most trending terms in the everlasting IT (information technology) sector today. Big Data represents a large volume of data which is virtually impossible to process by just one machine regardless if the data is structured or unstructured. The structured or unstructured Big Data can be processed for analysis using Data Analysis. Where as Cloud Computing is more than just an application that systematically not only stores data and programs using a network of remote servers over the internet but also provides services like SaaS(Software as a Service), PaaS(Platform as a Service) and IaaS(Infrastructure as a Service). It is a technology used to store data and information on a remote server rather than physical hard drive. Big Data exists without Cloud Computing but Cloud Computing requires Big Data for computing resources.

Keywords*: Big data, cloud computing*

I. INTRODUCTION

1.1 Big data is a term utilized to refer to the increase in the volume of data that are difficult to store, process, and analyze through traditional database technologies. The nature of big data is indistinct and involves considerable processes to identify and translate the data into new insights. The term “big data” is relatively new in IT and business. However, several researchers and practitioners have utilized the term in previous literature. For instance, [1] referred to big data as a large volume of scientific data for visualization. Several definitions of big data currently exist. For instance, [2] defined big data as “the amount of data just beyond technology's capability to store, manage, and process efficiently.” Meanwhile[3][4] defined big data as characterized by three V’s: volume, variety, and velocity.

Volume: Volume refers to the amount of data generated through websites, portals and online applications.

Variety: Variety in Big Data refers to all the structured and unstructured data that has the possibility of getting generated either by humans or by machines.

Velocity: Velocity refers to the speed with which data are being generated[5].

1.2 Cloud computing is the use of various services, such as software development platforms, servers, storage and software, over the internet, often referred to as the "cloud."

In general, there are three cloud computing characteristics that are common among all cloud-computing vendors:

1. The back-end of the application (especially hardware) is completely managed by a cloud vendor.
2. A user only pays for services used (memory, processing time and bandwidth, etc.).
3. Services are scalable

Many cloud computing advancements are closely related to virtualization. The ability to pay on demand and scale quickly is largely a result of cloud computing vendors being able to pool resources that may be divided among multiple clients.

II. REVIEW OF LITERATURE

Gunasekaran Manogaran et al.,[6], had proposed MetaCloud Data Storage Architecture for protecting Big Data in Cloud Computing Environment. This proposed approach requires high implementation effort, it provides valuable information for cloud computing environment that can have high impact on the next generation systems[6]. They have also proposed to do MetaCloud Data Storage Architecture for real time processing of streaming data in their future work.

Yibin Li et al.,[7], has focused on the problem of the cloud data storage and aimed to provide an approach that could avoid the cloud operators reaching user’s sensitive data. Our experimental evaluations had proved that our proposed scheme could effectively defend major threats from cloud-side[7]. Future work would address securing data duplications in order to increase the level of data availability when data center’s down will cause the failure of data retrievals.

Hongming Cai et al.,[8], has proposed a timely research which overviews the current and potential IoT big data storage systems in cloud computing and at the same time surveys the state-of-art in literature from the view of data processing process. The IoT storage system enables tracking of essential information about items as they move through cloud platforms. It shows significant value for IoT applications by providing an accurate knowledge of the current IoT data processing, which results in higher availability and flexible resource provision[8].

Mehdi Bahrami et al.,[9], have discussed a definition of big data, the importance of big data, and major big data challenges and issues. We understand that, if we analyze big data with business intelligence tools, we may provide a catalyst to change an organization to a smart organization. We discussed the importance of cloud computing technology as a solution to handle big data for both computing and storage[9]. Finally, they have discussed major cloud computing system issues that need to be addressed for cloud computing to become a solution for handling big data.

Marcos D. Assun et al.,[10], has discussed about the amount of data currently generated by the various activities of the society has never been so big, and is being generated in an ever increasing speed. This Big Data trend is being seen by industries as a way of obtaining advantage over their competitors: if one business is able to make sense of the information contained in the data reasonably quicker, it will be able to get more customers, increase the revenue per customer, optimise its operation, and reduce its costs. Cloud computing helps in alleviating these problems by providing resources on-demand with costs proportional to the actual usage. Cloud computing plays a key role for Big Data, not only because it provides infrastructure and tools, but also because it is a business model that Big Data analytics can follow (e.g. Analytics as a Service (AaaS) or Big Data as a Service (BDaaS))[10]. Big data can have many levels of Volume,

Variety, and Velocity. Thus, it is important to choose the appropriate Big Data tools.

Alberto Fernández et al.,[11], has explained about the Big Data problem in DM(Data Mining) and BI (Business Intelligence)has been established as one of the hot topics in the areas of business and research. In particular, emphasizing the importance of this new field of work with respect to its application in BI tasks, and exposing the use of Cloud Computing as the right tool when compared to more traditional solutions. This new paradigm of work is based on the use of elastic computing resources. It allows users to execute requests dynamically, such that it can achieve a high degree of scalability even with low-end hardware resources[11]. The main purpose is to provide an opportunity for any interested reader in the field of Big Data, to become familiar with the most relevant information of this topic.

Joonsang Baek et al.,[12], have introduced the Smart-Frame, a general framework for big data information management in smart grids based on cloud computing technology.Our basic idea is to set up cloud computing centers at three hierarchical levels to manage information: top, regional and end-user levels. While each regional cloud center is in charge of processing and managing regional data, the top cloud level provides a global view of the framework.Additionally, in order to support security for the framework, we have presented a solution based on identity-based cryptography and identity-based proxy re-encryption[12]. This proposed framework achieves not only scalability and flexibility but also security features.We implemented a proof-of-concept for our framework with a simple identity-based management for data confidentiality.

Bernice M. Purcell et al.,[13], has explained about Cloud computing enabling small to medium sized business to implement big data technology with a reduced commitment of company resources. The processing capabilities of the big data model could provide new insights to the business pertaining to performance improvement, decision making support, and innovation in business models, products, and services. Benefits of implementing big data technology through cloud computing are cost savings in hardware and processing, as well as the ability to experiment with big data technology before making a substantial commitment of company resources[13]. There are several models of cloud computing services available to the businesses to consider, with each model having trade-offs between the benefit of cost savings and the concerns data security and loss of control.

Jitendra Kumar Jaiswal[14], has proposed that Data is a new capital for enterprises and organisations. The analysis process ensures that hidden facts are highlighted for better decision making. Storing and processing RDBMS data was good enough solution in the previous decade. With the evolution of data types and its volume and velocity created a new phenomena named as Big data. Traditional servers and databases have limitations to process these Big data sets and thus evolution to Cloud computing begun. Cloud computing is a platform, which is flexible, efficient and scalable and adds strategic value to organisations of all scale. Executing Big data analytics projects require scalable computing as well as storage to process a huge amount of data. As data is becoming the key business driver, it's important to have high availability of data as without it the credibility can be on stake should there be a downtime due to production failures[14]. Thus Cloud computing is offering disaster recovery capability to handle major accidents and recovering faster as stated in agreed SLA's(Service Level Agreement).

Pedro Caldeira Neves et al.,[15], has discussed that data increasing on a daily base, big data systems and in particular, analytic tools, have become a major force of innovation that provides a way to store, process and get information over petabyte datasets. Cloud environments strongly leverage big data solutions by providing fault-tolerant, scalable and available environments to big data systems. Although big data systems are powerful systems that enable both enterprises and science to get insights over data, there are some concerns that need further investigation. Additional effort must be employed in developing security mechanisms and standardizing data types. Another crucial element of Big Data is scalability, which in commercial techniques are mostly manual, instead of automatic. Regarding this, we are planning to use adaptable mechanisms in order to develop a solution for implementing elasticity at several dimensions of big data systems running on cloud environments[15]. In this we provide an overview of big data in cloud environments, by highlighting its advantages and showing that both technologies work very well together but also presenting the challenges faced by the two technologies.

CONCLUSION

Big data is one of the most critical emerging technologies. The velocity, variety and volume makes the data management and analytics easier. Cloud computing has been a perfect vehicle for hosting big data workloads. However, working on big data in the cloud brings two different principles together for a challenging technology. The integration of big data with cloud computing technologies, businesses and education institutes can have a better direction to the future. The ability to store large amounts of data in different forms and process it all at very large speeds will result in data that can guide businesses and education institutes and many other fields in developing fast. This review paper has also discussed about the characteristics, trends and challenges of both Big Data and Cloud Computing. It also shows the benefits and risk that may arise by integrating Big data and cloud computing. The main advantage of integrating big data and cloud computing is the data storage and processing power availability. The cloud has access to a large pool of resources and various forms of infrastructure. With minimum effort the environment can be set up and managed to allow an excellent work space for all the big data needs(Data Analytics).

References

[1] M. Cox, D. Ellsworth, Managing Big Data For Scientific Visualization, ACM Siggraph, MRJ/NASA Ames Research Center, 1997.

[2] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A.H. Byers, Big data: The next frontier for innovation, competition, and productivity, (2011).

[3] P. Zikopoulos, K. Parasuraman, T. Deutsch, J. Giles, D. Corrigan, Harness the Power of Big Data The IBM Big Data Platform, McGraw Hill Professional, 2012.

[4] J.J. Berman, Introduction, in: Principles of Big Data, Morgan Kaufmann, Boston, 2013, xix–xxvi (pp).

[5] https://www.whishworks.com/blog/big-data/understanding-the-3-vs-of-big-data-volume-velocity-and-variety

[6] Gunasekaran Manogaran, Chandu Thota , M. Vijay Kumar, MetaCloud Data Storage Architecture for Big Data Security in Cloud Computing , Elsevier , 2016 , pp.128 – 133.

[7] Yibin Li , Keke Gai , Longfei Qiu , Meikang Qiu , Hui Zhao, Intelligent cryptography approach for secure distributed big data storage in cloud computing, Elsevier, 2016, pp.1-13.

[8] Hongming Cai, Boyi Xu, Lihong Jiang, Athanasios V. Vasilakos, IoT-based Big Data Storage Systems in Cloud Computing: Perspectives and Challenges, IEEE Internet of Things Journal, 2015.

[9] Mehdi Bahrami, Mukesh Singhal, The Role of Cloud Computing Architecture in Big Data, Vol. 8, pp. 275-295, Chapter 13, Pedrycz and S.-M. Chen (eds.), Springer, 2015.

[10] Marcos D. Assun, Rodrigo N. Calheiros , Silvia Bianchi , Marco A. S. Nettoc , Rajkumar Buyya, Big Data Computing and Clouds: Trends and Future Directions, Journal of Parallel and Distributed Computing, 2015.

[11] Alberto Fernández, Sara del Río, Victoria López, Abdullah Bawakid, María J. del Jesus, José M. Benítez and Francisco Herrera, Big Data with Cloud Computing: an insight on the computing environment, MapReduce, and programming frameworks, WIREs Data Mining Knowl Discov 2014, 4:380–409.

[12] Joonsang Baek, Quang Hieu Vu, Joseph K. Liu, Xinyi Huang, Yang Xiang, A secure cloud computing based framework for big data information management of smart grid, 2012.

[13] Bernice M. Purcell, Big data using cloud computing, Journal of Technology Research, 2014.

[14] Jitendra Kumar Jaiswal, Cloud Computing for Big Data Analytics Projects, Espoo, Apr 2018.

[15] Pedro Caldeira Neves, Bradley Schmerl, Jorge Bernardino, Javier Cámara, Big Data in Cloud Computing: features and issues, International Journal of Computer Applications, 2016.