RESEARCH ARTICLE OPEN ACCESS

The Study of Resource Management in Big Data Using Cloud Computing

ChamandeepKaur [1], Fatima Farhan Ali Almaliki [2], WejdanMohommedTherwi [2], AlaaMohommed Alhassan Tomihe [1], Samar Mansour Hassen [1]

Lecturer Jazan University [1], tudent Jazan University [2]

ABSTRACT

Resource management is a essential design issue for big data processing systems in the cloud. Many resource allocation strategies can have suggestively different impacts on performance on the system. Resource sharing is a typical and classic approach to resource efficiency. Resource management and task scheduling play an essential role, in cases where one is concerned with optimized use of resources. Additionally, a recently emerging research trend focuses on the possible convergence of Big Data Analytics and High-Performance Computing. As more and more big data applications with expanding and various resource requirements tend to deploy in the cloud. In this paper we present several types of resource management techniques for big data processing. We discuss designing, developing, and analysing mechanisms for resource management in cloud computing systems and data centres.

Keywords — Big data, Cloud computing, Resource sharing, Resource management

I. INTRODUCTION

In the innovative world, big data became powerful and utilizing the cloud makes it easier to track, analyze and ultimately act on insights. Big data and Cloud computing play an crucial role in business development. They are the center of concern in business as this technology helps to enhance the growth and provide more productivity. It plays an important role in growth of business and decision making. Cloud Computing is a technology which provides virtual services[1]. These services are safe, reliable, and we can access with the help of the internet. It provides a major advantage to small businesses as it provides storage facility at affordable costs and uses their data efficiently for analysis.

The combination of big data and the cloud computing can results tremendous value to all kinds of companies. The most general definition of Big Data is that it's a large volume of data, it might be terabyte or pet byte or even more than that. Data can be either structured or unstructured. This data can be so extensive that it cannot be processed through traditional database and software techniques. As for cloud computing, it means storing and accessing data, files, and programs over the Internet instead of the local computer's hard drive. The cloud is a metaphor for the Internet. Through hardware virtualization, cloud computing provides the option of storing significant amounts of data with the help of scalability, fault tolerance and availability[2]. This allows Big Data to be available, scalable and fault tolerant through cloud computing. Teams often found it difficult to share insights widely. Sharing the information was difficult and transferring data, especially large amounts of it, was slow. The cloud has reduced many of these limitations, making it easy for teams to coordinate across any distance and to widely share data, ideas and information. Big data processing used to be difficult and expensive. This also meant that big data efforts were reactionary, providing insights from out-of-date data. Businesses, however, need to be proactive and able to access, analyze and act upon the most recent data.

II. BIG DATA AND CLOUD COMPUTING

Two mainstream technologies are set the trend in IT sector - Big Data and Cloud Computing. Big data is all about dealing with the massive scale of data whereas Cloud computing is about infrastructure [3][4]. However, the simplification offered by Big data and Cloud technology is the main reason for their huge enterprise adoption. For example Amazon "Elastic Map Reduce" demonstrates how the power of Cloud Elastic Computes is leveraged for Big Data processing.

The combination of both results in beneficial outcome for the organizations. These technologies are in the stage of evolution but their combination leverages scalable and cost-effective solution in big data analytics.

So, can we say Big data and Cloud computing a perfect combination. there are data points in support of it. Besides that, there are also some real-time challenges to deal with.

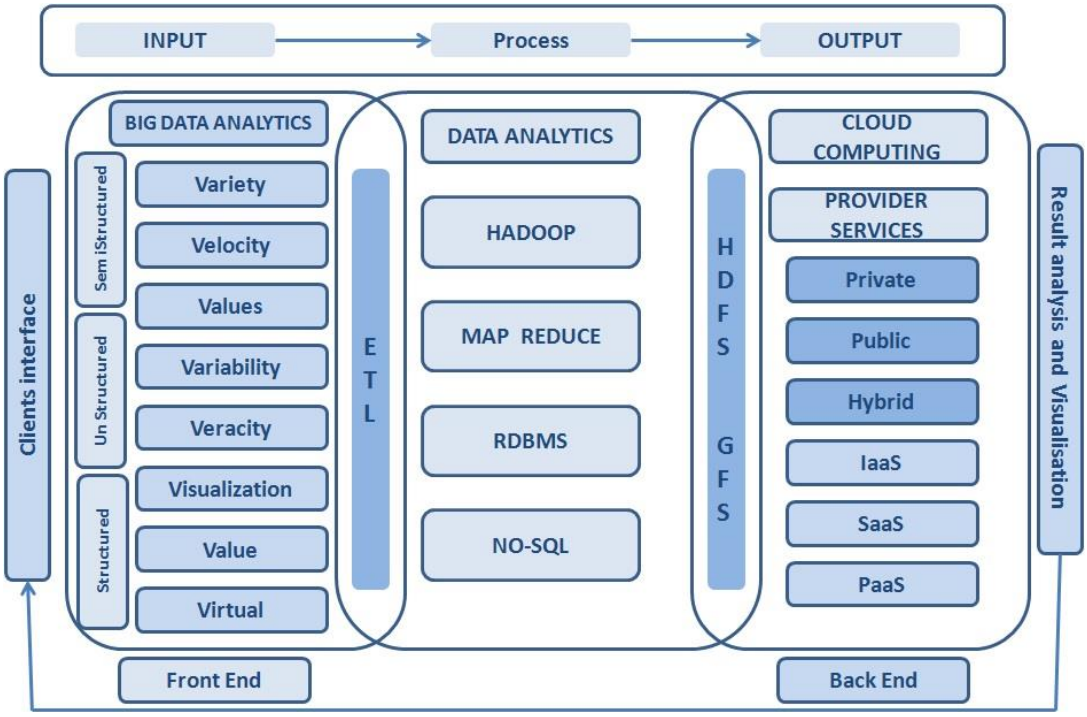

Fig 1: The relation between cloud and Big data model

Big data involves manipulating petabytes (soon it may exabytes and zettabytes) of data, and the cloud's scalable environment makes it possible to deploy data-intensive applications that power business analytics[5]. The cloud also simplifies connectivity and collaboration within an organization, which gives more employees access to relevant analytics and streamlines data sharing.

While it's easy for IT leaders to recognize the advantages of putting big data in the cloud, it may not be as simple to get C-suite executives and other primary stakeholders on board. But there's a business case to be made for the big data + cloud pairing because it gives executives a better view of the business and boosts data-driven decision making.

Cloud computing offers access to data storage, processing, and analytics on a more scalable, flexible, cost-effective, and even secure basis than can be achieved with an on-premises deployment[10]. These characteristics are essential for customers when data volumes are growing exponentiallyto make storage and processing resources available as needed, as well as to get value from that data. Also for those organizations that are just starting on the journey toward doing big data analytics and machine learning, and that want to avoid the potential complexities of on-premises big data systems, the cloud offers a way to experiment with managed services. Big data projects typically get started with data storage and application of basic analytics modules[6]. However, as you discover ways to extract data at a much larger scale, you will need to find better methods to process and analyze this data, which will likely require infrastructure upgrades.

You may add more capacity to your in-house data warehouse or power up more servers to cater to the rapidly-increasing analytics requirements. But even with the boost of your on-premise systems, your infrastructure eventually may not be able to keep up.

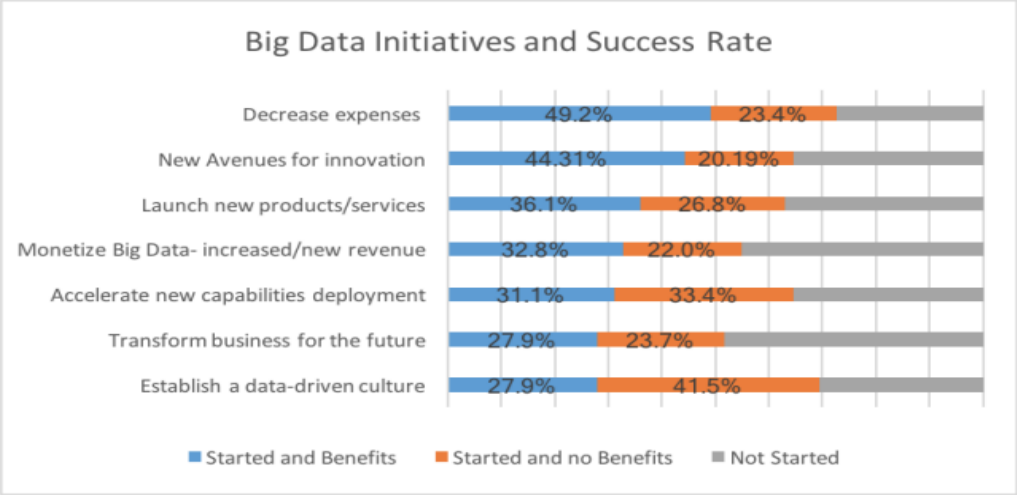

Fig 2: Big data success rate

III. CLOUD STRUCTURE FOR BIG DATA

In Cloud computing the data is gathered in data centers and distributed to data end users. For the business continuity, the data will be backup and also ensures of recovery of the data, all resources will be available in cloud. Resources will not be available in physical form. The usage will be done in the end terminals like laptops, desktops, mobile phones etc with the internet connection[7].

IAAS in a public cloud: Using a cloud provider's infrastructure for Big Data services, gives access to almost limitless storage and compute power. IaaS can be utilised by enterprise customers to create cost effective and easily scalable IT solutions where cloud providers bear the complexities and expenses of managing the underlying hardware. If the scale of a business customer's operations fluctuate, or they are looking to expand, they can tap into the cloud resource as and when they need it rather than purchase, install and integrate hardware themselves.

PAAS in a private cloud: PaaS vendors are beginning to incorporate Big Data technologies such as Hadoop and MapReduce into their PaaS offerings, which eliminate the dealing with the complexities of managing individual software and hardware elements. For example, web developers can use individual PaaS environments at every stage of development, testing and ultimately hosting their websites. However, businesses that are developing their own internal software can also utilise Platform as a Service, particularly to create distinct ring-fenced development and testing environments.

SAAS in a hybrid cloud: Many organizations feel the need to analyse the customer's voice, especially on social media. SaaS vendors provide the platform for the analysis as well as the social media data. Office software is the best example of businesses utilising SaaS. Tasks related to accounting, sales, invoicing and planning can all be performed through SAAS. Businesses may wish to use one piece of software that performs all of these tasks or several that each perform different tasks. The software can be subscribed through internet and then accessed online via any computer in the office using a username and password. If needed, they can switch to software that fulfills their requirements in better manner. Everyone who needs access to a particular piece of software can be set up as a user, whether it is one or two people or every employee in a corporation that employs hundreds.

IV. RESOURCE MANAGEMENT OF BIG DATA IN CLOUD COMPUTING

In cloud computing, the data is stored in the servers which are maintained by different service providers**.** We can access the data with the help of internet. However, big data break a large amount of data and distribute it across different computer systems, where the data will analyze and process.

The big data solutions can be deployed with the help of platform as a service or software as a service in software as a service various components or applications which are running on Hadoop are accessible[8]. In the

platform as a service the platform is Hadoop which is provided to the customers. The main focus of cloud computing is to provide computer resources and services with the help of network connection. While big data is about solving problems when a huge amount of data generating and processing[9]. Big data utilizes the data which we can use to generate before buy an organization and provides the insights which can benefit the business in the future. On the other hand, Cloud computing is flexible and quick service with respect to it deployments, so that the steam line operations of an organization go successfully.

V. CONCLUSION

Cloud computing provides enterprises a cost-effective & flexible way to access a vast volume of information we call the Big Data. In this paper we shown how the big data and cloud computing are related to each other in order to make the business successful. Cloud Computing and Big Data Analytics have truly impacted the way organizations function and humans operate. Cloud Computing provides benefits which are applicable to all sizes of businesses and all kinds of individuals. However, it is important to note that cloud-based big data analytics success depends on many factors. It has seen that with the help of technologies like Big data and Cloud Computing there has been a gradual growth of the company. So, both plays an important role for an organization.

REFERENCES

[1] "Big Data in the Cloud: Converging Technologies, (August), " https://www.intel.com/content/dam/www/public/emea/de/de/documents/product-briefs/big-data-cloud-technologies-brief.pdf.

[2] R. Arora, "An introduction to big data, high performance computing, high-throughput computing, and Hadoop," Conquering Big Data with High Performance Computing, pp. 1–12, 2016.

[3] G. Hesse and M. Lorenz, "Conceptual survey on data stream processing systems," in Proceedings of the 2015 IEEE 21st International Conference on Parallel and Distributed Systems (ICPADS), pp. 797–802, Melbourne, Australia, December 2015.

[4] F. Bajaber, R. Elshawi, O. Batarfi, A. Altalhi, A. Barnawi, and S. Sakr, "Big data 2.0 processing systems: taxonomy and open challenges," Journal of Grid Computing, vol. 14, no. 3, pp. 379–405, 2016.

[5] A. Beloglazov, R. Buyya, Y. C. Lee, A. Zomaya, et al. A taxonomy and survey of energy-efficient data centers and cloud computing systems. Advances in computers, 82(2):47-111,2011.

[6] Dong, Z., Liu, N., & Rojas-Cessa, R. (2015). Greedy scheduling of tasks with time constraints for energy-efficient cloud-computing data centers. Journal of Cloud Computing,

[7] Thaman, J., & Singh, M. (2016). Current perspective in task scheduling techniques in cloud computing: A review. International Journal in Foundations of Computer Science & Technology.

[8] Bechini, A., Marcelloni, F., and Segatori, A. (2016). A MapReduce solution for associative classification of big data. Information Sciences, 332, 33-55.

[9] Sowmya, T. S. R. (2016). Cost minimization for big data processing in geo-distributed data centers. Asia-Pacific Journal of Convergent Research Interchange, 2(4), 33-41.

[10] M. Bertoni, S. Ceri, A. Kaitoua, and P. Pinoli, "Evaluating cloud frameworks on genomic applications," in Proceedings of the 3rd IEEE International Conference on Big Data, IEEE Big Data 2015, pp. 193–202, November 2015.

[11] Satish, Karuturi S R V, and M Swamy Das. "Quantum Leap in Cluster Efficiency by Analyzing Cost-Benefits in Cloud Computing." In Computer Science and Engineering by Auroras Scientific Technological & Research Academy Hyderabad, vol. 17, no. 2, pp. 58-71. Accessed 2018.