

Big Data On Cloud Computing

Mubeena Shaik, Naseema Shaik

Research Scholars, Acharya Nagarjuna University, India

ABSTRACT: Big Data and Cloud Computing are primarily different, Big data is all about dealing with the massive scale of data whereas Cloud computing is about infrastructure. 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 technologies results the organizational benefits. These technologies lead to cost effective. In this paper we are discussing about the combined benefits of big data and cloud computing.

Token words: Big Data, Cloud Computing, IAAS, PAAS, SAAS, Security, Availability.

I. INTRODUCTION

The combination of both yields beneficial outcome for the organizations. Not to mention, both the technologies are in the stage of evolution but their combination leverages scalable and cost-effective solution in big data analytics. Big data and Cloud computing both the technologies are valuable individually. Additionally, many businesses are targeting to combine the two techniques to obtain more business benefits. Both the technologies aim to increase the revenue of the company while reducing the investment cost. While Cloud manages the local software, Big data helps in business decisions.

Big data involves manipulating petabytes of data, and the cloud's scalable environment makes it possible to deploy data-intensive applications that power business analytics. 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 and cloud pairing because it gives executives a better view of the business and boosts data-driven decision making [1].

Cloud computing has allowed the usage of the most advanced technological resources available on the market in any domain, from individuals to small and medium enterprises, due to the open gate to hardware and software resources available at low costs. Meanwhile, this opportunity has led to a significant increase in the produced data volume and to the cornering of the Big Data concept. Not only the business environment is interested in collecting information from unconventional data sources, but also government agencies, national institutions and other organizations analyse and extract meaningful insight from this maze of data, be it security related or simply behavioural patterns of consumers [2]. It could be said that there is a strong correlation and symbiosis between these two technologies, as any Cloud Computing implementation includes a high-capacity storage solution and any Big Data platform uses distributed information collection and processing, as in Cloud architectures. As mentioned, there is no limit in the storage space allocated for this kind of data, and thus Cloud platforms are faced with problems related to the efficient scaling in storage capacity. Another challenge comes from the evolution from structured data in relational databases to fast processing of large, unstructured data sets [2]. Hybrid Clouds are often the preferred option for the companies, which may use Private Clouds to manage internal structured data, while Public Clouds allow the extension of their resources and the addition of model services.

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 6, Special Issue , August 2019

International Conference on Recent Advances in Science, Engineering, Technology and Management at Sree Vahini Institute of Science and Technology-Tiruvuru, Krishna Dist, A.P

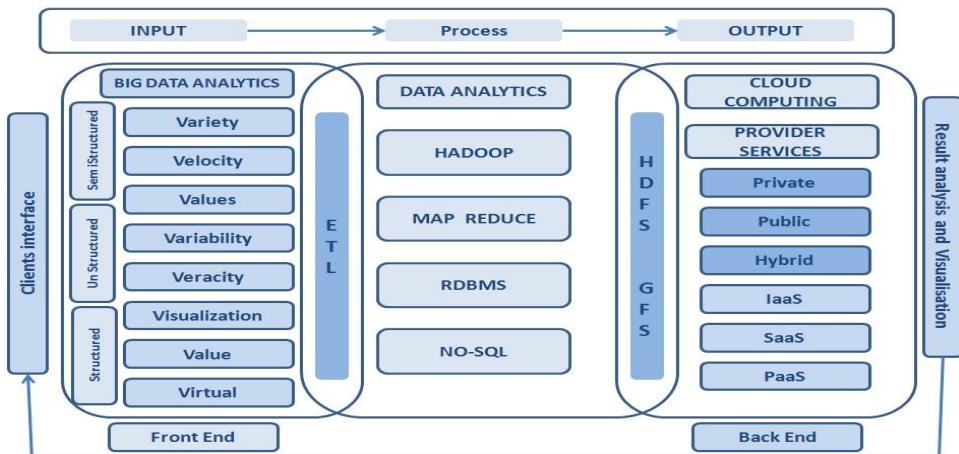


Fig 1: Relationship between Big Data and Cloud Computing Architecture

IBM researchers come with a new approach in cloud services for business, called Analytics as a Service AaaS [3]. Starting from their experience of working with large companies, they observed that enterprises often keep their most sensitive data in-house, while the volumes of external data or archives (Big Data) may be located in a Public Cloud environment. To extract value from these collections is necessary to implement a new cloud service, called Analytics as a Service (AaaS), having the following key capabilities [4]:

- capturing and extracting structured and unstructured data from different sources;
- managing and controlling data in accordance with company policy and specific requirements;
- performing data integration, analysis, transformation in order to deliver the required information.

From the studies that were published and briefly mentioned in this paper, that both technologies are evolving constantly and their interconnection will lead to a new, hybrid concept that will bring together the best of what each has to offer, with the aim of cloud-based data gaining and processing for very large volumes of information, available anywhere on demand[5].

II. RELATION BETWEEN BIG DATA AND CLOUD COMPUTING

Big data and Cloud computing relationship can be categorized based on service types:

IAAS in Public Cloud

IaaS is a cost-effective solution and utilizing this Cloud service, Big Data services enable people to access unlimited storage and compute power. It is a very cost-effective solution for enterprises where the Cloud provider bears all the expenses of managing underlying hardware.

PAAS in Private Cloud

PaaS vendors incorporate Big Data technologies into their offered service. Hence, they eliminate the need for dealing with the complexities of managing single software and hardware elements which is a real concern while dealing with terabytes of data.

SAAS in Hybrid Cloud

Analyzing social media data is nowadays an essential parameter for companies for business analysis. In this context, SaaS vendors provide an excellent platform for conducting the analysis.

we can see from the above discussion that Cloud enables “As-a-Service” pattern by abstracting the challenges and complexity through a scalable and elastic self-service application. Big data requirement is same where distributed processing of massive data is abstracted from the end users[9].



ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 6, Special Issue , August 2019

International Conference on Recent Advances in Science, Engineering, Technology and Management at Sree Vahini Institute of Science and Technology-Tiruvuru, Krishna Dist, A.P

III. BENEFITS OF USING BIG DATA ON CLOUD COMPUTING

There are multiple benefits of Big data analysis in Cloud. Which are discussing below:

1) Improved analysis

With the advancement of Cloud technology, big data analysis has become more improved causing better results. Hence, companies prefer to perform big data analysis in the Cloud. Moreover, Cloud helps to integrate data from numerous sources[6].

2) Simplified Infrastructure

Big Data analysis is a tremendous strenuous job on infrastructure as the data comes in large volumes with varying speeds, and types which traditional infrastructures usually cannot keep up with. As the Cloud computing provides flexible infrastructure, which we can scale according to the needs at the time, it is easy to manage workloads.

3) Lowering the cost

Both Big data and Cloud technology delivers value to organizations by reducing the ownership. The Pay-per-user model of Cloud turns CAPEX into OPEX. On the other hand, Apache cut down the licensing cost of Big data which is supposed to be cost millions to build and buy. Cloud enables customers for big data processing without large-scale big data resources. Hence, both Big Data and Cloud technology are driving the cost down for enterprise purposes and bringing value to the enterprise.

4) Security and Privacy

Data security and privacy are two major concerns when dealing with enterprise data. Moreover, when your application is hosted on a Cloud platform due to its open environment and limited user control security becomes a primary concern[7]. On the other hand, being an open source application, Big data solution like Hadoop uses a lot of third-party services and infrastructure. Hence, nowadays system integrators bring in Private Cloud Solution that is Elastic and Scalable. Furthermore, it also leverages Scalable Distributed Processing.

Besides that, Cloud data is stored and processed in a central location commonly known as Cloud storage server. Along with it the service provider and the customer signs a service level agreement (SLA) to gain the trust between them. If require the provider also leverages required advanced level of security control. This enables the security of big data in Cloud computing covering the following issues:

- Protecting big data from advanced threats.
- How Cloud service providers maintain storage and data.

On the other hand in many organizations, big data analytics is utilized to detect and prevent advanced threats and malicious hackers.

5) Virtualization

Infrastructure plays a crucial role to support any application. Virtualization technology is the ideal platform for big data. Virtualized big data applications like Hadoop provide multiple benefits which are not accessible on physical infrastructure, but it simplifies big data Management[8]. Big data and Cloud computing point to the convergence of various technologies and trends that makes IT infrastructure and related applications more dynamic, more expendable and more modular and. Hence, Big data and Cloud computing projects rely heavily on virtualization.

IV. CHALLENGES OF BIG DATA ON CLOUD COMPUTING

V.

Bringing big data to the cloud presents huge opportunities, but there are some challenges that need to be overcome[10].Concerning the existing problems, we define some of the possible advances in the next few years:

- Data governance and data recovery plans are difficult to manage and implement, but as Big Data become a de facto technology, companies are starting to understand the need of such plans.
- Big Data variety can be addressed by using data standardization. This, we believe, is the next step to minimize the impact of heterogeneity.
- Security and Privacy can be resolved using data encryption. However, a new generation of systems must ensure that data is accessed quickly and that encryption does not affect processing times so badly.

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 6, Special Issue , August 2019

International Conference on Recent Advances in Science, Engineering, Technology and Management at Sree Vahini Institute of Science and Technology-Tiruvuru, Krishna Dist, A.P

- New and secure QoS (quality of service) based data uploading mechanisms may be the answer to ease data uploading onto the cloud.
- Exaflop computing is a major challenge that involves governments funding and which is in its best interest. The best solutions so far use HPCs and GPUs.
- Scalability and elasticity techniques exist and are broadly used by several Big Data vendors such as Amazon and Microsoft. The major concern relies upon developing fully automatic reactive and proactive systems that are capable of dealing with load requirements automatically[9].

VI. CONCLUSION

Big Data and Cloud Computing are two emerging technologies in todays business environment. Benefits of implementing big data technology on cloud computing are cost effective in hardware and processing, as well as the ability to experiment with big data technology before making a substantial commitment of company resources. Though few challenges exist there like data storage capabilities, however, these are negligible before the offered beneficial outcomes. So, we can conclude that Big Data and Cloud Computing is the perfect combination.

REFERENCES

- [1] Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G...Zaharia, M. (2010, April). A view of cloud computing. Communications of the ACM, 53(4), 50-58. DOI: 10.1145/1721654.1721672.
- [2] Ferkoun, M., 2014. Cloud computing and big data: An ideal combination, available at <http://thoughtsoncloud.com/2014/02/cloud-computing-and-big-data-an-ideal-combination>
- [3]Fattah, A., 2014. Cloud Analytics: Selecting Patterns of Integration, Source: IBM Data magazine, available at <http://ibmdatamag.com/2014/09/cloud-analytics-selecting-patterns-of-integration/>
- [4] Intel IT Center, 2013. Big Data in the Cloud: Converging Technologies, available at <http://online.ip expo.co.uk/index.php/layout/set/print/content/download/81982/1679544/file/big-data-cloud-technologies-brief.pdf>
- [5] Banica, L., Stefan, C., Rosca, D. & Enescu, F., 2013. Moving from Learning Management Systems to the e-Learning Cloud, AWERProcedia Information Technology & Computer Science. [Online], pp 865-874, available from: www.awer-center.org/pitcs
- [6] CompuerWeekly.com. (2009, March). Hardware as a service. Retrieved from <http://www.computerweekly.com/feature/Hardware-as-a-Service>
- [7] IOS Press. (2011). Guidelines on security and privacy in public cloud computing. Journal of EGovernance, 34 149-151. DOI: 10.3233/GOV-2011-0271
- [8] Sliwa, C. (2011, June 16). Scale-out NAS, object storage, cloud gateways replacing traditional NAS. Retrieved from <http://searchstorage.techtarget.com/feature/Scale-out-NAS-objectstorage-cloud-gateways-replacing-file-storage>
- [9] Salesforce.com. (2012). The end of software: Building and running applications in the cloud. Retrieved from <http://www.salesforce.com/paas/>
- [10] Eckerson, W., 2001. Meta Data Management in the Data Warehouse Environment - A Logical Approach to Meaningful Data Analysis, <http://www.olap.it/Articoli/Ascential -meta data management in the dwh environment.pdf>/