The Rise of Big Data and Cloud Computing

Mohaiminul Islam[1], Shamim Reza[2]

[1]School of Intelligent Technology and Engineering, Chongqing University of Science and Technology, Chongqing, China

[2]Department of Information and Communication Technology, Mawlana Bhasani Science and Technology University, Tangail, Bangladesh

Email address:
mohaiminul100@hotmail.com (M. Islam), shamim.doict@gmail.com (S. Reza)

Abstract: Big Data has emerged in the past few years as a new paradigm providing abundant data and opportunities to improve and/or enable research and decision-support applications with unprecedented value for digital earth applications including business, sciences and engineering. At the same time, Big Data presents challenges for digital earth to store, transport, process, mine and serve the data. Cloud computing provides fundamental support to address the challenges with shared computing resources including computing, storage, networking and analytical software; the application of these resources has fostered impressive Big Data advancements. This paper surveys the two frontiers – Big Data and cloud computing – and reviews the advantages and consequences of utilizing cloud computing to tackling Big Data in the digital earth and relevant science domains. While Big Data is responsible for data storage and processing, the cloud provides a reliable, accessible, and scalable environment for Big Data systems to function. Big Data is defined as the quantity of digital data produced from different sources of technology, for example, sensors, digitizers, scanners, numerical modeling, mobile phones, Internet, videos, social networks. Cloud Computing and Big Data are complementary to each other. Rapid growth in Big Data is regarded as a problem. Clouds are evolving and providing solutions for the appropriate environment of Big Data while traditional storage cannot meet the requirements for dealing with Big Data, in addition to the need for data exchange between various distributed storage locations. Cloud Computing provides solutions and addresses problems with Big Data. Big data and Cloud computing both the technologies are valuable on its own. Furthermore, many businesses are targeting to combine the two techniques to reap more business benefits. Both the technologies aim to enhance the revenue of the company while reducing the investment cost. While Cloud manages the local software, Big data helps in business decisions. In paper introduces the relationship between Big Data and Cloud Computing, Cloud Computing role of Big Data, advantages of Big Data and Cloud computing, cloud architecture, importance of Cloud Computing.

Keywords: Big Data, Cloud Computing, Hadoop, Analytics, NoSQL

1. Introduction

Big data basically refers to sets of data that are large in volume and cannot be processed through traditional application software. The term big data is not new, as it has been around for a long time and there are many new concepts related to the term. Even though the concept is not new in the industry, there is a lot of confusion around the true meaning of what big data actually is. When you work on a particular principle and start collecting knowledge on it, you start generating data that will be useful for you in the future to analyze and get further insights. Before computers and the rise of the internet, transactions were recorded on paper and archive files that were fundamentally data [1]. Today, computers allow us to save whatever data we have on spreadsheets and organize them in the most efficient way. Cloud computing offers the best technology with a wide range of applications for various purposes in the most cost-effective way. Big data and cloud computing are a match made in heaven because there is a lot of data — and only cloud computing can provide that kind of compute power to process the data. Whatever we do almost leaves a digital trail, as we generate data whenever we are on the internet. As cloud computing is transforming IT, huge amounts of compute power are needed with the help of the internet to store and analyze this data. Cloud computing has reshaped the way computers are being used to process data. Cloud has made it

very simple for data storage compared to traditional data storage [2]. Cloud computing provides scalable resources on demand and has changed the way data is stored and processed. Cloud computing is a powerful approach to analyze data provided and has become vital in the growth of big data in multiple industries.

Today, two mainstream technologies are the center of concern in IT – Big Data and Cloud Computing. Fundamentally 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 yields beneficial outcome for the organizations [3]. 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.

Since its inception, information technology has been exclusively available for technology companies, large organizations, government and educational institutions. That was until the emergence of cloud computing in a process many call the "democratization" of information technology [4]. With an ever-expanding reach to the masses, a significant reduction in cost and an abundant choice of applications available, you can be truly empowered to leverage the best of existing technology, often without spending a penny on initial investment. The democratization of information technology has not only affected the cloud space, but big data as well. Adoption of open source Hadoop is growing at a fast pace and the ability to perform analytics on non-proprietary and affordable hardware is becoming more ubiquitous. Along with this phenomenon, we are now witnessing an explosion of information generated through social media, messaging, emails and more. Organizations and individuals are navigating a maze of ever-increasing data that can be difficult to roam through, let alone control and dissect [5]. This surge in the volume of data is now presenting a challenge to the cloud. Organizations have built their data architecture, storage policies and best practices mainly working with structured data, whereas the unstructured data does not fit the traditional relational database management system (RDBMS) framework. The issue is how to manipulate and extract the essence of the data rather than simply storing and retrieving it. As pointed out during IBM InterConnect 2013, businesses can get increased value from data insights gained through big data analytics supported by a cloud infrastructure. The explosion in unstructured data means that ways to harness the benefits of hybrid cloud and big data are more important than ever. The hybrid cloud model can assist organizations in addressing security concerns in their private cloud, while leveraging the public cloud infrastructure for analytics services. It comes as no surprise that there has been a keen interest from government agencies, government bodies and other

organizations to try and extract meaningful insight from this maze of data, whether it be security related or simply about patterns of consumers. On average, 2.5 billion gigabytes of data is created daily, consisting of 200 millions tweets and 30 billion pieces of content shared on Facebook each month. Looking at the available projections, the amount of data created by the year 2020 will reach a staggering 43 trillion gigabytes, with six billion people in possession of cell phones [6-7].

Cloud computing and big data, while still in constant evolution, are proving to be the ideal combination. Together, they provide a cost effective and scalable infrastructure to support big data and business analytics.

2. Big Data Evaluation

"You can have data without information, but you cannot have information without data." – Daniel Keys Moran

The above quote defines the importance of data. Ignoring the importance of big data can lead to be a very costly mistake for any kind of business in today's world. If data is that important then using effective analytics or big data tools to unlock the hidden power of data becomes imperative. Here we will discuss the benefits of using cloud computing for big data. If you have followed our earlier blogs, we have discussed at length the value of big data and here we will explore it even further [8].

Today, every organization, government, IT firm and political party considers data as a new and extremely useful currency. They willingly invest resources to unlock insights from collected data in their respective fields which can be profitable if it is adequately mined, stored and analyzed [9]. The early stages of using big data were mostly based around storing the data and applying some basic analytics modules. Now, as the practice has evolved, we have adopted more advanced methods of modeling, transforming, and extracting on a much larger scale. The field of big data now has the capacity for a globalized infrastructure. Internet and social media giants such as Google and Facebook were the pioneers of big data when they began uncovering, collecting and analyzing information collected by their users. Back then companies and researchers entered worked with externally sourced data, which was basically drawn from the "internet" or "public data sources". The term "big data" wasn't coined until 2010 approximately when they realized the power, need and importance of this information. Given the scope of information, the term "big data" come into the picture. And with that, the arrival of newly developed technologies and processes to help companies to turn the data into insight and profit [10, 11].

The term Big Data is being rapidly used almost everywhere across the planet – online and offline. Before that, information stored on your servers or computers was only sorted and filed. But today, all data becomes big data no matter where you have stored it or in which format [12].

Figure 1. *Big Data Infographic.*

3. The Relationship Between Big Data and Cloud Computing

With the generation of an enormous amount of data, cloud computing is playing a significant role in the storage and management of that data. It's not only about the growth of big data but also the expansion of data analytics platforms like Hadoop. As a result, it is creating new opportunities in Cloud computing. Hence, the service providers like AWS, Google and Microsoft are offering their own big data systems in a cost-efficient manner which is scalable for businesses of all sizes.

This, in turn, has led to a new service model which is known as *Analytics as a Service* (*AaaS*). This will provide a faster and scalable way to integrate different types of structured, semi-structured and unstructured data, analyze them, transform and visualize them in real time [13].

Additionally, Big data cloud computing relationship can be assessed from below perspectives and benefits –

1. A cloud computing environment usually has several user terminals and service providers. From the collection terminals, the user collects the data using the big data tools. On the other hand, from the service provider end it saves, stores and processes the big data. Hence, cloud computing provides a big data infrastructure. The infrastructure must provide on-demand resources and services to ensure uninterrupted service.

2. Since the cloud environment is scalable, hence it can provide adequate data management solution irrespective of the volume of the data. If the necessary cloud computing service provider can also offer security policies as per the user demands.

3. Identity management and access control are two major concerns while dealing with confidential company data. Cloud computing can meet this security requirement using a simple software interface by abstracting internal details of the information. Additionally, this guarantees complete confidentiality of user data and only provides access to the authorized users.

4. Big data for data processing can be located across the global locations and maintaining such huge servers in different locations is a costly measure for an organization. As cloud computing can store and process data through geographically dispersed and as well as virtual servers it reduces the cost of big data processing significantly.

5. Cloud computing uses high-level software and applications which do not depend on the efficiency of the user devices. Furthermore, it depends on the network servers and their strength. On the contrary, if we use personal resources for big data that will be dependent on the user device. Hence, big data cloud computing service is beneficial.

6. Cloud computing enables high-speed data flow over the network. As a result, it causes faster big data processing [14-15].
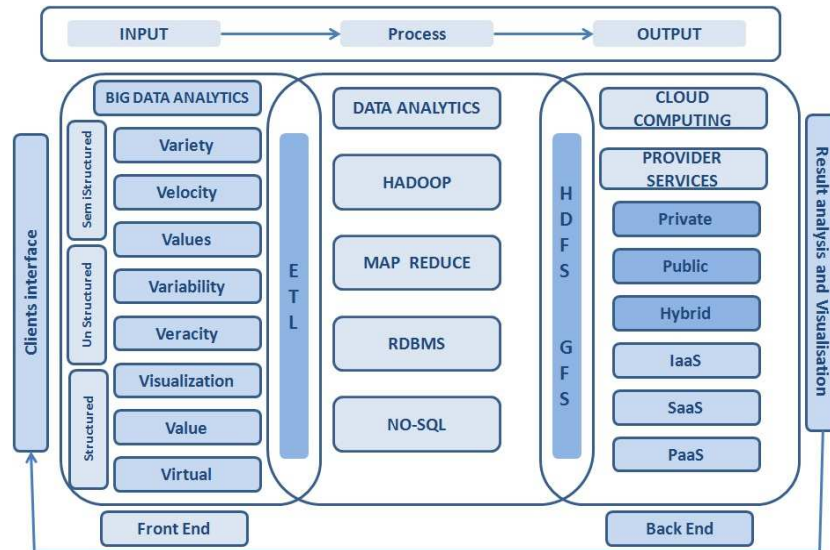

Figure 2. Relationship between Big Data and Cloud Computing.

4. Cloud Computing Role for Big Data

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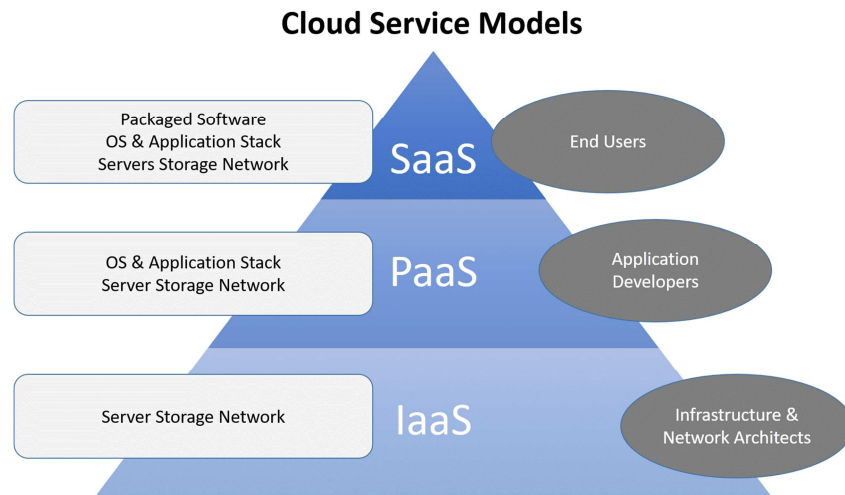

Figure 3. Cloud Computing Services and Models.

4.1. 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.

4.2. 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.

4.3. 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 [16].

5. Benefits of Big Data Analysis in Cloud

Hence, from the above description, we can see 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.

There are multiple benefits of Big data analysis in Cloud.

5.1. 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.

5.2. 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 [17].

5.3. 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.

5.4. 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. 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:

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

There are rules associated with service level agreements for protecting.

a) data
b) capacity
c) scalability
d) security
e) privacy
f) availability of data storage and data growth

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

5.5. 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 [18]. 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.
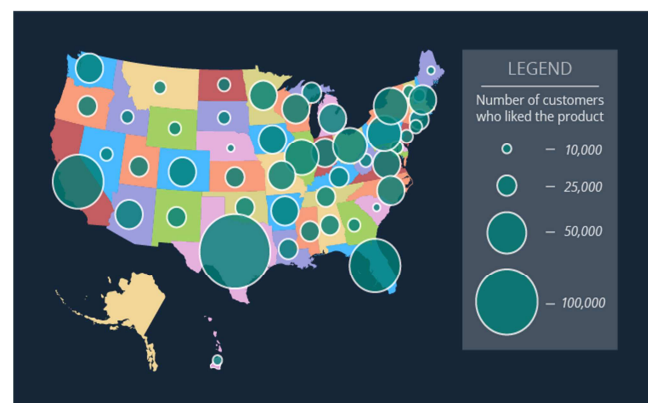

Figure 4. *A picture of Data Visualization.*

6. Advantages of Big Data and Cloud Computing

Cloud computing and big data is an ideal combination as it provides a solution which is both scalable and accommodating for big data and business analytics. Imagine a world where all the information resources are easily accessible, and every aspect of life can benefit from this information. Let's take a look at these advantages in detail:

6.1. Agility

The traditional method of storing and managing data is becoming outdated quickly. Setting up an infrastructure is not only expensive but also time-consuming, as installing and running a server can take weeks. With cloud computing, it's possible to provide any infrastructure with all the required resources almost instantly. A good cloud provider will companies to ensure that their work is always on the go without any hitches.

6.2. Elasticity

A cloud platform can dynamically expand to provide storage for ever increasing data. Once the company gets the necessary insight from the data, storage space can be increased or reduced to accommodate the data as per the requirement.

6.3. Data Processing

A large volume of data leads to the issue of how to process it efficiently. Social media alone generate a massive amount of unstructured data in various forms. With Big Data platforms, cloud computing makes the whole process easier and accessible to small, medium and large enterprises.

6.4. Cutting Cost with Big Data in the Cloud

Cloud computing is a terrific solution for enterprises that wish to have state of the art technology running their operations under a limited budget. Maintaining a big data center to perform Big Data analytics can quickly drain an IT budget. Nowadays, companies have the option to avoid investing heavily in setting up the IT department and maintaining hardware infrastructure. With the cloud computing, the responsibility shifts to the cloud providers and the company only have to pay for the storage space and power consumption.

6.5. Reduced Complexity

Any implementation of big data solution requires several components and integrations. Cloud computing provides the option to automate these components, thus reducing complexity and enhancing the productivity of the Big Data analysis team [19-20-21].

7. Cloud Big Data Challenges

Vertical scaling achieves elasticity by adding additional instances with each of them serving a part of the demand.

Software like Hadoop are specifically designed as distributed systems to take advantage of vertical scaling. They process small independent tasks in massive parallel scale. Distributed systems can also serve as data stores like NoSQL databases, e.g. Cassandra or HBase, or filesystems like Hadoop's HDFS. Alternatives like Storm provide coordinated stream data processes in near real-time through a cluster of machines with complex workflows.

The interchangeability of the resources together with distributed software design absorbs failure and equivalently scaling of virtual computing instances unperturbed. Spiking or bursting demands can be accommodated just as well as personalities or continued growth. Renting practically unlimited resources for short periods allows one-off or periodical projects at a modest expense. Data mining and web crawling are great examples [22]. It is conceivable to crawl huge web sites with millions of pages in days or hours for a few hundred dollars or less. Inexpensive tiny virtual instances with minimal CPU resources are ideal for this purpose since the majority of crawling the web is spent waiting for IO resources. Instantiating thousands of these machines to achieve millions of requests per day is easy and often costs less than a fraction of a cent per instance hour [23].

Of course, such mining operations should be mindful of the resources of the web sites or application interfaces they mine, respect their terms, and not impede their service. A poorly planned data mining operation is equivalent to a denial of service attack. Lastly, cloud computing is naturally a good fit for storing and processing the big data accumulated form such operations [24-25].

8. Cloud Architecture

Three main cloud architecture models have developed over time; private, public and hybrid cloud. They all share the idea of resource commodification and to that end usually virtualize computing and abstract storage layers.
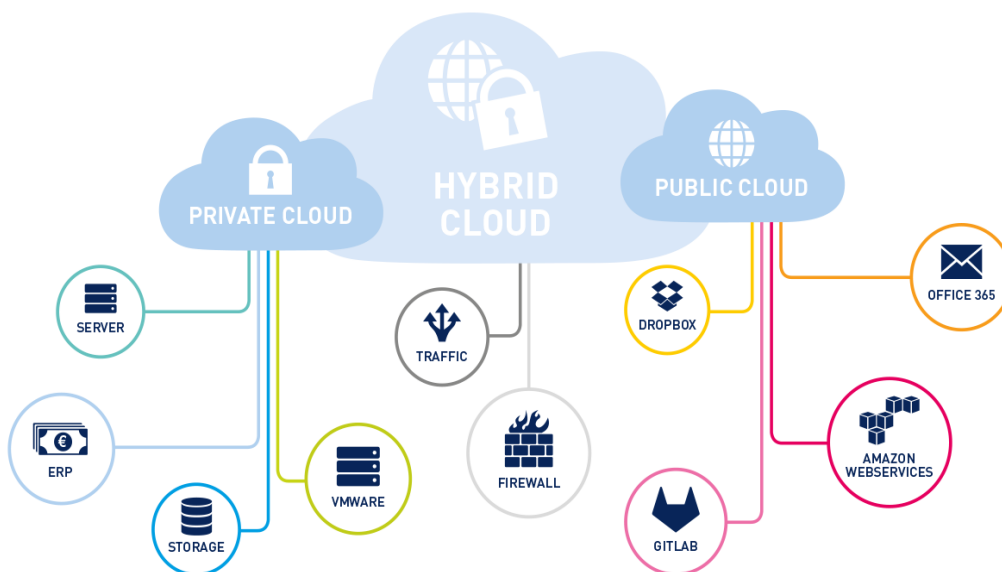

Figure 5. *Cloud Computing Architecture Diagrams.*

8.1. Private Cloud

Private clouds are dedicated to one organization and do not share physical resources. The resource can be provided in-house or externally. A typical underlying requirement of private cloud deployments are security requirements and regulations that need a strict separation of an organization's data storage and processing from accidental or malicious access through shared resources [26]. Private cloud setups are challenging since the economical advantages of scale are usually not achievable within most projects and organizations despite the utilization of industry standards. The return of investment compared to public cloud offerings is rarely obtained and the operational overhead and risk of failure is significant.

Additionally, cloud providers have captured the trend for increased security and provide special environments, i.e. dedicated hardware to rent and encrypt virtual private networks as well as encrypted storage to address most security concerns. Cloud providers may also offer data storage, transfer, and processing restricted to specific geographic regions to ensure compliance with local privacy laws and regulations [27].

Another reason for private cloud deployments are legacy systems with special hardware needs or exceptional resource demand, e.g. extreme memory or computing instances which are not available in public clouds. These are valid concerns however if these demands are extraordinary the question if a cloud architecture is the correct solution has to be raised. One reason can be to establish a private cloud for a transitionary period to run legacy and demanding systems in parallel while their services are ported to a cloud environment culminating in a switch to a cheaper public or hybrid cloud.

8.2. Public Cloud

Public clouds share physical resources for data transfers, storage, and processing. However, customers have private visualized computing environments and isolated storage. Security concerns, which entice a few to adopt private clouds or custom deployments, are for the vast majority of customers and projects irrelevant. Visualization makes access to other customers' data extremely difficult.

Real-world problems around public cloud computing are more mundane like data lock-in and fluctuating performance of individual instances. The data lock-in is a soft measure and works by making data inflow to the cloud provider free or very cheap. The copying of data out to local systems or other providers is often more expensive. This is not an insurmountable problem and in practice encourages to utilize more services from a cloud provider instead of moving data in and out for different services or processes. Usually this is not sensible anyway due to network speed and complexities around dealing with multiple platforms.

The varying performance of instances stems typically from the dependency on what kind of load other customers generate on the shared physical infrastructure. Secondly, over time the physical infrastructure providing the virtual resources changes and is updated. The available resources for each customer on a physical machine are usually throttled to ensure that each customer receives a guaranteed level of performance. Larger resources generally deliver very predictable performance since they are much closer aligned with the physical instance's performance. Horizontally scaling projects with small instance should not rely on an exact performance of each instance but be adaptive and focus on the average performance required and scale according to need.

8.3. Hybrid Cloud

The hybrid cloud architecture merges private and public cloud deployments. This is often an attempt to achieve security and elasticity, or provide cheaper base load and burst capabilities. Some organizations experience short periods of extremely high loads, e.g. as a result of seasonality like black Friday for retail, or marketing events like sponsoring a popular TV event. These events can have huge economic impact to organizations if they are serviced poorly.

The hybrid cloud provides the opportunity to serve the base load with in-house services and rent for a short period a multiple of the resources to service the extreme demand. This requires a great deal of operational ability in the organization to seamlessly scale between the private and public cloud. Tools for hybrid or private cloud deployments exist like Eucalyptus for Amazon Web Services. On the long-term the additional expense of the hybrid approach often is not justifiable since cloud providers offer major discounts for multi-year commitments. This makes moving base load services to the public cloud attractive since it is accompanied by a simpler deployment strategy [28].

9. Importance of Cloud Computing for Big Data

There are several reasons for having a big data on cloud. Some of them are discussed below:

9.1. Instant Infrastructure

One of the key benefits of a cloud-based approach to big data analytics is the ability to establish big data infrastructure as quickly as possible with a scalable environment. A big data cloud service provides the infrastructure that companies would otherwise have to build up themselves from scratch.

Big data offers all analytics needs in a single roof. It is important to note that cloud-based big data analytics success is dependent on many key factors. Most significant of these is the quality and reliability of the solution provider. The vendor must combine robust, extensive expertise in both the big data and cloud computing sectors.

9.2. Cutting Costs with Big Data in the Cloud

This offers major financial advantages to participating

companies, but how? Performing big data analytics in the house requires companies to attain and maintain big data centres, and maintain the big data centres is more about, that budget can be used in other companies' expansion plans and policies.

Shifting the big data analytics on the cloud, allows firms to cut costs in terms of purchasing equipment, cooling machines and ensuring security, while also allowing them to keep the most sensitive data on-premise and have the full control on it.

9.3. Fast Time to Value

A modern data-management platform brings together master data management and big data analytics capabilities in the cloud so that business can create data-driven applications using the reliable data with relevant insights. The principal advantage of this unified cloud platform is faster time-to-value, keeping up with the pace of business [29]. Whenever there is a need for a new, data-driven decision management application, you can create one in the cloud quickly. There is no need to set up infrastructure (hardware, operating systems, databases, application servers, analytics), create new integrations, or define data models or data uploads. In the cloud, everything is already set up and available [30].

10. Conclusion

Big Data has emerged in the past few years as a new paradigm providing abundant data and opportunities to improve and/or enable research and decision-support applications with unprecedented value for digital earth applications including business, sciences and engineering. At the same time, Big Data presents challenges for digital earth to store, transport, process, mine and serve the data. Cloud computing provides fundamental support to address the challenges with shared computing resources including computing, storage, networking and analytical software; the application of these resources has fostered impressive Big Data advancements. Cloud computing is a powerful technology to perform massive-scale and complex computing. It eliminates the need to maintain expensive computing hardware, dedicated space, and software. Massive growth in the scale of data or big data generated through cloud computing has been observed. Addressing big data is a challenging and time-demanding task that requires a large computational infrastructure to ensure successful data processing and analysis. Cloud computing provides enterprises a cost-effective & flexible way to access a vast volume of information we call the Big Data. Because of Big Data and cloud computing, it is now much easier to start an IT company than ever before. However, it is important to note that cloud-based big data analytics success depends on many factors. An important factor is a reliable cloud provider with extensive expertise, offering highly robust services. VEXXHOST cloud services allowing enterprises an excellent opportunity to take advantage of both Big Data analytics and cloud computing simultaneously.

References

[1] Villars, R. L., Olofson, C. W., & Eastwood, M (June, 2011). Big data: What it is and why you should care. IDC White Paper. Framingham, MA: IDC.

[2] Shakil, K. A.; Sethi, S.; Alam, M., (2015). An effective framework for managing university data using a cloud based environment, Computing for Sustainable Global Development (INDIACom), 2nd International Conference on, vol., no., pp. 1262, 1266, 11-13.

[3] K. Das and P. M. Kumar, Big data analytics: A framework forum structured data analysis, International Journal of Engineering and Technology, 5 (1) (2013), pp. 153-156.

[4] T. K. Das, D. P. Acharjya and M. R. Patra, Opinion mining about aproduct by analyzing public tweets in twitter, International Conference on Computer Communication and Informatics, 2014.

[5] M. K. Kakhani, S. Kakhani and S. R. Biradar, Research issues in bigdata analytics, International Journal of Application or Innovation in Engineering & Management, 2 (8) (2015), pp. 228-232.

[6] A. Gandomi and M. Haider, Beyond the hype: Big data concepts, meth-ods, and analytics, International Journal of Information Management, 35 (2) (2015), pp. 137-144.

[7] C. Lynch, Big data: How do your data grow?, Nature, 455 (2008), pp. 28-29.

[8] X. Jin, B. W. Wah, X. Cheng and Y. Wang, Significance and challenges of big data research, Big Data Research, 2 (2) (2015), pp. 59-64.

[9] R. Kitchin, Big Data, new epistemologies and paradigm shifts, Big Data Society, 1 (1) (2014), pp. 1-12.

[10] C. L. Philip, Q. Chen and C. Y. Zhang, Data-intensive applications, challenges, techniques and technologies: A survey on big data, Infor-mation Sciences, 275 (2014), pp. 314-347.

[11] K. Kambatla, G. Kollias, V. Kumar and A. Gram, Trends in big data analytics, Journal of Parallel and Distributed Computing, 74 (7) (2014), pp. 2561-2573.

[12] S. Del. Rio, V. Lopez, J. M. Bentez and F. Herrera, On the use of mapreduce for imbalanced big data using random forest, Information Sciences, 285 (2014), pp. 112-137.

[13] MH. Kuo, T. Sahama, A. W. Kushniruk, E. M. Borycki and D. K. Grunwell, Health big data analytics: current perspectives, challenges and potential solutions, International Journal of Big Data Intelligence, 1 (2014), pp. 114-126.

[14] R. Nambiar, A. Sethi, R. Bhardwaj and R. Vargheese, A look at challenges and opportunities of big data analytics in healthcare, IEEE International Conference on Big Data, 2013, pp. 17-22.

[15] Z. Huang, A fast clustering algorithm to cluster very large categorical data sets in data mining, SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, 1997.

[16] Chang, J. Dean, S. Ghemawat, W. C. Hsieh, D. A. Wallach, M. Burrows, T. Chandra, A. Fikes, and R. E. Gruber. Bigtable: A Distributed Storage System for Structured Data. In OSDI, pages 205–218, 2006.

[17] Rajeev Gupta, Himanshu Gupta, and Mukesh Mohania, "Cloud Computing and Big Data Analytics: What Is New from Databases Perspective?". S. Srinivasa and V. Bhatnagar (Eds.): BDA 2012, LNCS 7678, pp. Springer-Verlag Berlin Heidelberg 42–61, 2012.

[18] Curino, C., Jones, E. P. C., Popa, R. A., Malviya, N., Wu, E., Madden, S., Balakrishnan, H., Zeldovich, N.: Realtional Cloud: A Database-as-a-Service for the Cloud. In: Proceedings of Conference on Innovative Data Systems Research, CIDR-2011.

[19] Alberto Ferandez, Sara del R, Victoria opez, Abdullah Bawakid, Maria J. del Jesus, Jose M. Benitez, and Francisco Herrera. "Big Data with Cloud Computing: an insight on the computing environment, Map Reduce, and programming frameworks". doi: 10.1002/widm.1134. WIREs Data Mining Knowl Discov, 4:380–409, 2014.

[20] Lu, Huang, Ting-tin Hu, and Hai-shan Chen. "Research on Hadoop Cloud Computing Model and its Applications". Hangzhou, China: 2012, pp. 59 – 63, 21-24 Oct. 2012.

[21] Wie, Jiang, Ravi V. T, and Agrawal G. "A Map-Reduce System with an Alternate API for Multi-core Environments.". Melbourne, VIC: 2010, pp. 84-93, 17-20 May. 2010.

[22] K, Chitharanjan, and Kala Karun A. "A review on hadoop - HDFS infrastructure exten-sions.". JeJu Island: 2013, pp. 132-137, 11-12 Apr. 2013.

[23] F. C. P, Muhtaroglu, Demir S, Obali M, and Girgin C. "Business model canvas perspective on big data applications." Big Data, 2013 IEEE International Conference, Silicon Valley, CA, Oct 6-9, p. 32 – 37, 2013.

[24] Castelino, C., Gandhi, D., Narula, H. G., & Chokshi, N. H. (2014). Integration of Big Data and Cloud Computing. International Journal of Engineering Trends and Technology (IJETT), 100-102.

[25] Chandrashekar, R., Kala, M., & Mane, D. (2015). Integration of Big Data in Cloud compu-ting environments for enhanced data processing capabilities. International Journal of Engineering Research and General Science, 240- 245.

[26] James Kobielus, I., & Bob Marcus, E. S. (2014). Deploying Big Data Analytics Applica-tions to the Cloud: Roadmap for Success. Cloud Standards Customer Council.

[27] M. Herland, T. M. Khoshgoftaar and R. Wald, A review of data mining using big data in health informatics, Journal of Big Data, 1 (2) (2014), pp. 1-35.

[28] X. Y. Chen and Z. G. Jin, Research on key technology and applicationsfor internet of things, Physics Procedia, 33, (2012), pp. 561-566.

[29] N. Mishra, C. Lin and H. Chang, A cognitive adopted frameworkfor iot big data management and knowledge discovery prospective, International Journal of Distributed Sensor Networks, 2015, (2015), pp. 1-13.

[30] Z. Hongjun, H. Wenning, H. Dengchao and M. Yuxing, Survey of research on information security in big data, Congresso da sociedada Brasileira de Computacao, 2014, pp. 1-6.