International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611

Overview of Big Data on Cloud Computing

Aishwarya Karnik[1], Neha Humbal[2]

Department of Electronics and Telecommunication, Dwarkadas J. Sanghvi College of Engineering

Department of Electronics, Dwarkadas J. Sanghvi College of Engineering

Abstract: ***Big data is a popular term used to describe the exponential growth and availability of data, both structured and unstructured. A major challenge for researchers and practitioners is that the growth rate of data exceeds their ability to design appropriate platforms for data analysis and update intensive workloads. Cloud computing has become a powerful architecture to perform large-scale and complex computing. Cloud computing can not only minimize the cost and restriction for automation and computerization by individuals and enterprises but can also provide reduced infrastructure maintenance cost, efficient management, and user access[4].As a result of the above advantages cloud computing platforms are developed. This paper focuses on the relationship between big data on cloud computing.***

Keywords: Big Data, Cloud Computing, Big Data Analytics, Network-Attached Storage, Hadoop.

1. Introduction to Big Data

Big Data as the name suggests is a broad term for data sets so massive that the traditional methods render ineffective. As shown in Fig1.1. functional requirements of Big Data include Collection, Integration, Analysis and thereby taking efficient decisions. Big Data analytics focus on analyzing the data to predict the future for solving tomorrow‟s business challenges. Today data is not constrained to the older structured one, we now deal with the semi-structured or unstructured data. Research suggests that 88 percent of the data generated is unstructured and it is the fastest growing data

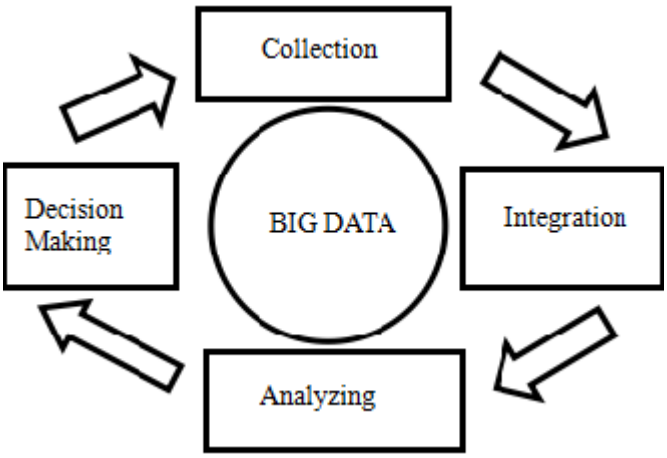

Figure 1.1: Functional requirements of Big data

Big data Stack focuses on managing the Fast Data and performing Big Data Analytics to provide focused services on top of the analytics. The term „ Big data‟ is relatively new and various sources have their respective definitions. For instance, [2] referred to big data as a large volume of scientific data for visualization. Several definitions of big data currently exist. For instance, [5] defined big data as “the amount of data just beyond technology's capability to store, manage, and process efficiently.” Meanwhile, [6] and [3] defined big data as characterized by three Vs: volume, variety, and velocity. The above three were Garner‟s terms to introduce Big Data. IDC later extended it to four Vs namely: Volume, Variety, Velocity and Value. Here Volume referred to the wide data generated from multiple sources. The size of the data determines the value and potential of the data under consideration, and whether it can actually be considered big data or not. Variety refers to the different kinds of data collected for example, video, audio, image and text from varied sources like sensors, social media, smart phones, etc. Today, the major data generated from the internet is unstructured including, messages, blogs, online games, etc. This data received has a certain speed or Velocity. Thus, the speed at which the data is collected and processed to meet the challenges comes under Velocity. Value is nothing but discovering huge hidden values from the data set. Today's data comes from multiple sources. And it is still an undertaking to link, match, cleanse and transform data across systems. However, it is necessary to connect and correlate relationships, hierarchies and multiple data linkages or your data can quickly spiral out of control. Thus, Complexity comes into picture.

2. Need For Cloud Computing

This extensively increasing data has bought up multiple challenges. The traditional structured data used relational database management system (RDBMS). Additionally they are now using more and more expensive hardware, this was the reason RDBMS could not handle this big data. The big data requires huge storage space. A typical solution for the same is based on clustered Network- attached storage (NAS) [7]. This infrastructure made use of NAS “pods” with each pod comprising of several storage devices connected on a NAS device [7]. These NAS devices are interconnected to allow massive sharing and searching of Data. However, this approach can be very expensive for a small or medium scaled business.

Thus, data storage on Cloud was a viable solution. A number of architectures and deployment models exist for cloud computing. Businesses who are unable to afford Network Attached storage techniques can consider a number of cloud computing models for Big Data [1]. A cloud service provider can prove to be an inexpensive solution for the storage problems.

3. Cloud Computing and Its Services

Common deployment models for cloud computing include platform as a service (PaaS), software as a service (SaaS), infrastructure as a service (IaaS), and hardware as a service (HaaS) [8]. Cloud deployment solutions can provide services that businesses would otherwise not be able to afford. Businesses can also use cloud deployment solutions as a test measure before adopting a new application or technology company-wide.

There are a wide number of alternatives for businesses using the cloud for PaaS .Platform as a Service is the use of cloud computing to provide platforms for the development and use of custom applications .The PaaS solutions include application design and development tools, application testing, versioning, integration, deployment, and hosting, state management, and other related development tools. As related to big data, PaaS provides companies a platform for developing and using custom applications needed to analyze large quantities of unstructured data at a low cost and low risk in a secure environment [8].

Software as a service provides businesses with applications that are stored and run on virtual servers –in the cloud. The business is not charged for hardware, only for the bandwidth for the time and number of users necessary [1] .The main advantage of SaaS, Is that the solution allows businesses to shift the risks associated with software acquisition while moving IT from being reactive to proactive. Software as a Service provides companies analyzing big data proven software solutions for data analysis. The difference between SaaS and PaaS in this case is that SaaS is not going to provide a customized solution whereas PaaS will allow the company to develop a solution tailored to the company‟s needs[8] .

In the IaaS model, a client business will pay on a per-use basis for use of equipment to support computing operations including storage, hardware, servers, and networking equipment. Infrastructure as a service is the cloud computing model receiving the most attention from the market, with an expectation of 25% of enterprises planning to adopt a service provider for IaaS [1] . Services available to businesses through the IaaS model include disaster recovery, compute as a service, storage as a service, data center as a service, virtual desktop infrastructure, and cloud bursting, which is providing peak load capacity for variable processes

While not as yet being used as extensively as PaaS, SaaS, or IaaS, HaaS is a cloud service based upon the model of time sharing on minicomputers and mainframes from the 1960s and 1970s [1]. Time sharing developed into the practice of managed services. In a managed services situation, the managed service provider (MSP) would remotely monitor and administer hardware located at a client‟s site as contracted. A problem with managed services was the necessity for some MSPs to provide hardware on-site for clients, the cost of which needed to be built into the MSP‟s cost. The HaaS model allows the customer to license the hardware directly from the service provider which alleviates the associated costs.

4. Types of Clouds

Three types of clouds exist -the public cloud, the private cloud, and the hybrid cloud. A public cloud is the pay-as-you-go services available to the general public [1]. In a public cloud configuration, a business does not own the core technology resources and services but outsources these. A public cloud is considered to be an external cloud.

A private cloud, unlike public cloud, is internal data center of a business that is not available to the general public but uses cloud structure. In a private cloud configuration, resources and services are owned by the business, with the services accessible within the business through the intranet. Since the technology is owned and operated by the business, private cloud is more expensive than a public cloud, but is also more secure. A private cloud is an internal cloud, residing inside the company‟s firewall and managed by the company.

When a company uses a hybrid cloud , as its name suggests it uses a public cloud for some tasks and a private cloud for other tasks. When using a hybrid cloud model, public cloud will use by a company to expedite extra tasks that are not able to be easily run in the company‟s data center or on its private cloud .A hybrid cloud allows a company to maintain critical, confidential data and information within it firewall while leveraging the public cloud for non-confidential. The private cloud portion of the hybrid cloud is accessed by company employees, both in the company and on the road, and is maintained by the internal technology group. The private cloud part of the hybrid cloud is also accessed by the company employees but is maintained by external service providers. Each portion of the hybrid cloud can connect to the other portion.

The type of cloud a company uses varies depending upon the company‟s needs and resources. Where The public cloud is considered the least secure of the three types, with services and resources able to be accessed over the Internet through protocols adopted by the provider (Géczy, Izumi, & Hasida, 2012) [1]. The communications protocols adopted by the provider are not necessarily secure; the choice of using secure or non-secure protocols is up to the providers. The public cloud is also the least costly of the cloud types, with cost savings in the areas of information technology deployment, management, and maintenance.

The private cloud provides services to company employees through an intranet. If mobile employees are able to access the private cloud, the access is typically through secure communication protocols .All services and resources provided are tailored to the needs of the business, and the business has total control over the services and resources. Due to the financial and human resources needed to deploy, manage, and maintain the information technology resources and services provided, the private cloud is the most expensive type of cloud .

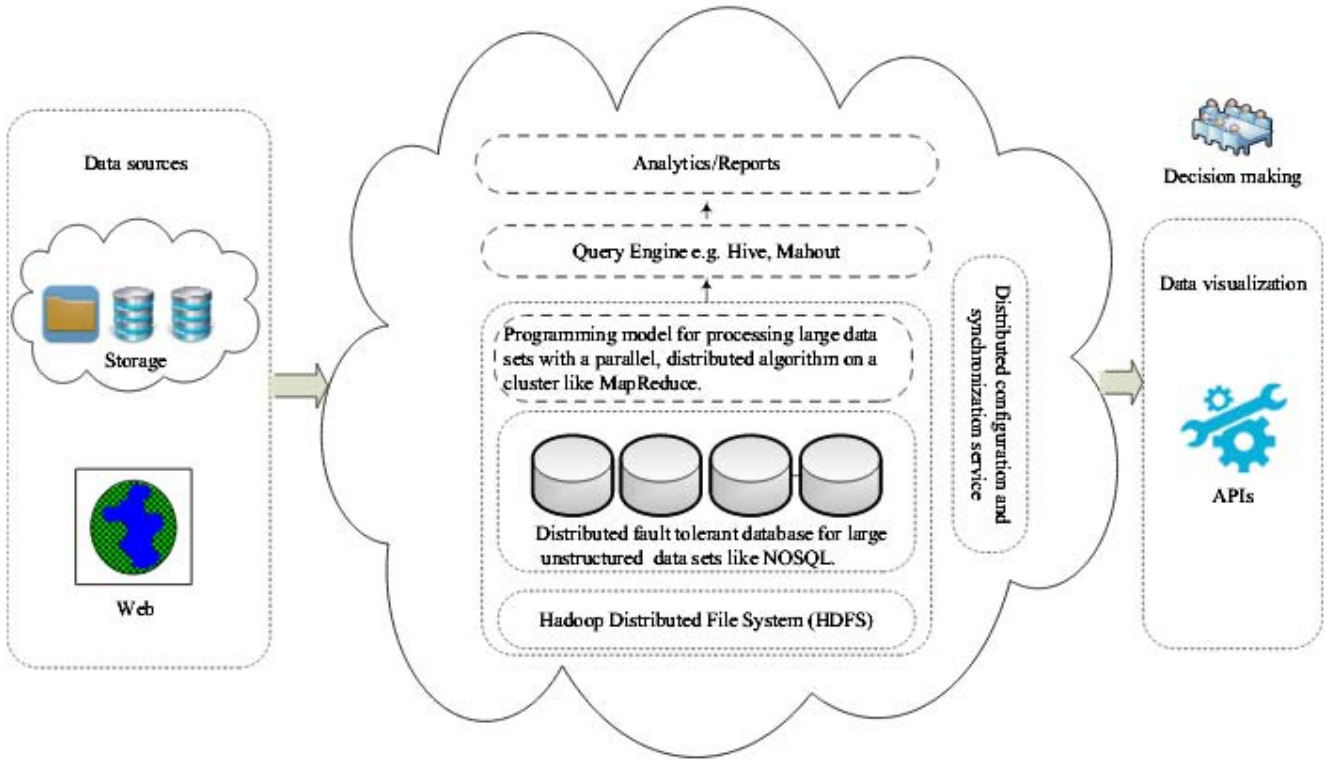

Figure 4.1: Cloud Computing on Big Data

When a business uses a hybrid cloud, the business owns its core information technology resources and services and will host and provide the resources and services inhouse. Noncritical services are outsourced and maintained on a public cloud. Typically, core information technology resources and services are mission critical and are often confidential. Therefore, resources and services that need to be secure are hosted and maintained on the private cloud, with the public cloud used for other services as a cost saving measure

5. Relationship Between Big Data and Cloud Computing

Big data and cloud computing are closely associated with each other. Big data provides users the platform to use commodity computing to process distributed queries across multiple datasets and return resultant sets in a timely manner. Cloud computing provides the underlying engine through the use of Hadoop, a class of distributed data-processing platforms. The use of cloud computing in big data is shown in Fig.4.1.

Large data sources from the cloud and Web are stored in distributed fault-tolerant database and processed through a programming model for large datasets with a parallel distributed algorithm in a cluster [9].

The main purpose of data visualization, as shown in Fig.4.1, is to view analytical results presented visually through different graphs for decision making [9]. Big data utilizes distributed storage technology based on cloud computing rather than local storage attached to a computer or electronic device. Big data evaluation is driven by fast-growing cloud-based applications developed using virtualized technologies. Therefore, cloud computing not only provides facilities for the computation and processing of big data but also serves as a service model.

6. Conclusion

Cloud computing enables 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. Several models of cloud computing services are available to the businesses to consider, with each model having tradeoffs between the benefit of cost savings and the concerns data security and loss of control.

References

[1] Bernice M. Purcell, Big data using cloud computing, Journal of Technology Research, Holy Family University
[2] J.Manyika,M.Chui,.B.Brown,.J.Bughin,R.Dobbs,C.Roxburgh,A.H.Byers,Bigdata: The next frontier for innovation, competition and productivity, (2011).
[3] J.JBerman,Introduction,in:Principlesof BigData,Morgan Kaufmann,Boston,2013,xix–xxvi(pp).
[4] L.Chih-Wei,H.Chih-Ming,C.Chih-Hung,Y.Chao-Tung, AnImprovementtoDataServiceinCloudComputingwithContentSensitiveTransactionAnalysisandAdaptation,ComputerSoftwareandApplicationsConferenceWorkshops(COMPSACW),2013IEEE37thAnnual, 2013,pp.463–468.
[5] MCox,D.Ellsworth,ManagingBigDataForScientificVisu

alization,ACMSiggraph,MRJ/NASAAmesResearchCenter,1997.
[6] P.Zikopoulos,K.Parasuraman,T.Deutsch,J.Giles,D.Corrigan,HarnessthePowerofBigDataTheIBMBigDataPlatform,McGrawHillProfessional,2012.
[7] White, C. (2011). Data Communications and Computer Networks: A business user‟s approach, (6th ed.). Boston: Cengage Learning.
[8] Domenico Talia, cloud for scalable Big data analytics, University of Calabria, Italy.
[9] Ibrahim Abaker Targio Hashem, Ibrar Yaqoob, Nor Badrul Anuar, Salimah Mokhtar, Abdullah Gani, The rise of“big data”on cloud computing: Review and open research issues, faculty of Computer Science and information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia

Author Profile

Aishwarya Karnik, Department of Electronics and Telecommunication, Dwarkadas J. Sanghvi College of Engineering.

Neha Humbal, Department of Electronics, Dwarkadas J. Sanghvi College of Engineering