

CLOUD AND BIG DATA: A COMPELLING COMBINATION

¹T. Suneetha, ²T. Kavitha,

^{1,2}Lecturer,

^{1,2}Department of Computer Science And Engineering,

^{1,2}Loyola Academy Degree And PG College, ^{1,2}Secunderabad, ^{1,2}TS, ^{1,2}INDIA.

Abstract: *With the rapid growth of emerging applications like social network analysis, semantic Web analysis and bio informatics network analysis, a variety of data to be processed continues to witness a quick increase. Effective management and analysis of large-scale data poses an interesting but critical challenge. Recently, big data has attracted a lot of attention from academia, industry as well as government. This paper introduces several big data processing techniques from system and application aspects. First, from the view of cloud data management and big data processing mechanisms, we present the key issues of big data processing, including cloud computing platform, cloud architecture, cloud database and data storage scheme. Following the Map Reduce parallel processing frame work, we then introduce Map Reduce optimization strategies and applications reported in the literature. Finally, we discuss the open issues and challenges, and deeply explore the research directions in the future on big data processing in cloud computing environments.*

Keywords - *Big Data, cloud Computing, cloud architecture, private cloud, public cloud, hybrid cloud.*

I. INTRODUCTION

One essential quality of cloud computing is in aggregation of resources and data into data centers on the internet. The present cloud services (IaaS, PaaS and SaaS) realize improved execution efficiency by aggregating application execution environments at various levels including server, OS and middle ware levels for sharing them. Meanwhile, another approach of aggregating data into clouds has also been launched, and it is to analyze such data with the powerful computational capacity of clouds. In this way cloud is now in the phase of expanding from application aggregation and sharing to data aggregation and utilization. To make full use of data, tens of terabytes (TBs) or tens of Terabytes (PBs) of data need to be handled and a new type of technology different from ordinary information and communications technology (ICT) is required. This paper presents distributed data store and complex event processing, which are basic technology for big data processing in cloud environments, and the research by Fujitsu Laboratories of Europe on workflow description for data processing. This should provide a perspective on the direction in which data processing technology will develop in the future.

II. RELATED WORK

A. Big Data

Big data is defined as a large amount of data which requires new technologies and architecture to make possible to extract value from it by capturing and analysis process. New sources of big data include location specific data arising from traffic management and from the tracking of personal devices and smart phones. Big Data has emerged because we are living in a society which makes increasing use of data intensive technologies. Due to such large size of data it becomes very difficult to perform effective analysis using the existing traditional techniques. Since, Big Data is a recent upcoming technology in the market which can bring huge benefits to the business organization, it becomes necessary that the challenges and issues associated in bringing and adapting to this technology are need to be understood. Big Data concept means a datasets which continues to grow so much that it becomes difficult to manage it using existing database management concepts and tools. The difficulties can be related to data capture, storage, search, sharing, analytics and visualization, etc.

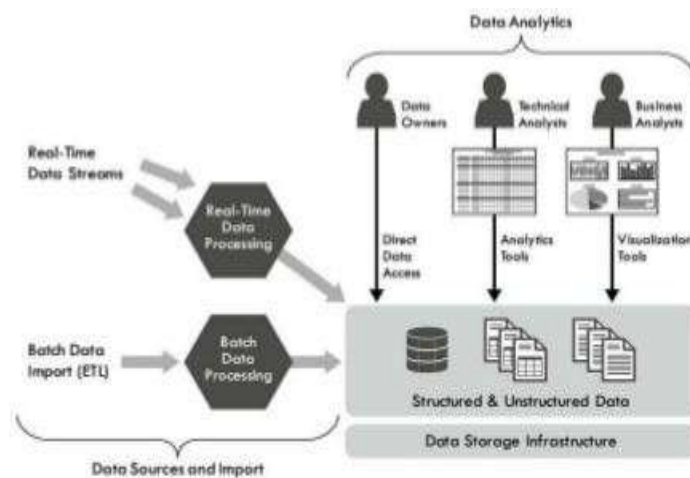


Figure 1 : Example of Big Data Architecture (Avekxa Inc., 2013)

Figure.1. Example of Big Data Architecture (Avekxa Inc., 2013)

Big Data due to its properties like volume, velocity, variety, value and complexity put forward many challenges. The various faced in large data management include scalability, unstructured data, accessibility, real time analytics, fault tolerance and many more. In addition to variations in the amount of data stored in different sectors, the types of data generated and stored -i.e. encoded video, image, audio, or text/numeric information; also differ marked from industry to industry.

B. Cloud Computing

Cloud computing is a paradigm of distributed computing to provide the customers on-demand, utility based computing services. Cloud users can provide more reliable, available and updated services to their clients in turn. Cloud itself consists of physical machines in the data centers of cloud providers. Virtualizations provided on top of these physical machines. These virtual machines are provided to the cloud users. Different cloud provider provides cloud services of different abstraction level. So the cloud services are divided into many types like Software as a Service, Platform as a Service or Infrastructure as a Service. These services are available over the Internet in the whole world where the cloud acts as the single point of access for serving all customers. Cloud computing architecture addresses difficulties of large scale data processing.

1) Types of Cloud:

Cloud can be of three types:

Private Cloud: This type of cloud is maintained within an organization and used solely for their internal purpose. So the utility model is not a big term in this scenario. Many companies are moving towards this setting and experts consider this is the 1st step for an organization to move into cloud. Security, network bandwidth are not critical issues for private cloud.

Public Cloud: In this type an organization rents cloud services from cloud provider's on-demand basis. Services provided to the users using utility computing model.

Hybrid Cloud: This type of cloud is composed of multiple internal or external cloud. This is the scenario when an organization moves to public cloud computing domain from its internal private cloud.

III. METHODOLOGY

A. Cloud as an Enabler for Big Data analytics

Two IT initiatives are currently top of mind for organizations across the globe: big data analytics and cloud computing. Big data analytics offers the promise of providing valuable insights that can create competitive advantage, spark new innovations, and drive increased revenues. As a delivery model for IT services, cloud computing has the potential to enhance business agility and productivity while enabling greater efficiencies and reducing costs. Both technologies continue to evolve. Organizations are moving beyond questions of what and how to store big data to addressing how to derive meaningful analytics that respond to real business needs. As cloud computing continues to mature, a growing number of enterprises are building efficient and agile cloud environments, and cloud providers continue to expand service offerings. It makes sense, then, that IT organizations should look to cloud computing as the structure to support their big data projects. Big data environments require clusters of servers to support the tools that process the large volumes, high velocity, and varied formats of big data. Clouds are already deployed on pools of server, storage, and networking resources and can scale up or down as needed. Cloud computing offers a cost-effective way to support big data technologies and the advanced analytics applications that can drive business value.

B. Cloud and Big Data: A Compelling Combination

Cloud delivery models offer exceptional flexibility, enabling IT to evaluate the best approach to each business user's request. For example, organizations that already support an internal private cloud environment can add big data analytics to their in-house offerings, use a cloud services provider, or build a hybrid cloud that protects certain sensitive data in a private cloud, but takes advantage of valuable external data sources and applications provided in public clouds. Using cloud infrastructure to analyze big data makes sense because:

1) Investments in big data analysis can be significant and drive a need for efficient, cost-effective infrastructure: The resources to support distributed computing models in-house typically reside in large and midsize data centers. Private clouds can offer a more efficient, cost-effective model to implement analysis of big data in-house, while augmenting internal resources with public cloud services. This hybrid cloud option enables companies to use on-demand storage space and computing power via public cloud services for certain analytics initiatives (for example, short-term projects), and provide added capacity and scale as needed.

2) Big data may mix internal and external sources: While enterprises often keep their most sensitive data in-house, huge volumes of big data (owned by the organization or generated by third-party and public providers) may be located externally—some of it already in a cloud environment. Moving relevant data sources behind your firewall can be a significant commitment of resources. Analyzing the data where it resides—either in internal or public cloud data centers or in edge systems and client devices—often makes more sense.

C. **Unlocking the Potential of Big Data in Clouds**

Cloud computing models can help accelerate the potential for scalable analytics solutions. Clouds offer flexibility and efficiencies for accessing data, delivering insights, and driving value. However, cloud-based big data analytics is not a one-size-fits-all solution. Organizations using cloud infrastructure to provide IaaS have multiple options. By weighing factors of workload, cost, security, and data interoperability, IT can choose to utilize their private cloud to mitigate risk and maintain control; use public cloud infrastructure, platform, or analytics services to further enhance scalability; or implement a hybrid model that combines private and public cloud resources and services. No matter which cloud delivery model makes the most sense, businesses with varying needs and budgets can unlock the potential of big data in cloud environments.

D. **Which cloud for your data**

The type of cloud a company uses depends upon the company's needs and resources. 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. 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. 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 in-house. Non-critical 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.

E. **Cloud computing for Big data in a small to medium sized business**

Cloud computing provides an environment for small to medium sized businesses to implement big data technology. Benefits that businesses can realize from big data include performance improvement, decision making support, and innovation in business models, products, and services (Manyika et al., 2011). Three major reasons for small to medium sized businesses to use cloud computing for big data technology implementation are the ability to reduce hardware costs, reduce processing costs, and to test the value of big data before committing significant company resources. The major concerns regarding cloud computing are security and loss of control. Platform as a Service is a cloud computing model that provides hardware cost savings. Hardware cost savings are accrued using PaaS through standardization and high utilization of the cloud-based platform across a number of applications (Oracle, 2012). Businesses can also realize hardware cost savings from the SaaS model since the business incurs no additional hardware costs for implementation; the only costs are for bandwidth based on the time and number of users (Cole, 2012). Hardware as a Service is not currently used as often as other models, but businesses can derive hardware cost savings through the model since HaaS allows customers to license the hardware directly from the service provider. In-house processing of big data typically requires use of the Map Reduce programming paradigm. The parallel processing needs of Map Reduce entails a huge commitment of processing power. Use of cloud computing for big data implementation lowers the in-house processing power commitment by shifting the data processing to the cloud. The use of big data could provide sufficient benefit to a small to medium sized company to the extent that the business would be willing to commit resources to implement big data technology in-house. However, the level of benefit is difficult to determine without some experience. Cloud computing implementation of big data processing could provide the business with justification to adopt the technology in-house. If the benefit accrued from big data use on the cloud is significant, the business has established a reason to adopt the technology in house. Otherwise, the business can continue cloud computing use of big data or rely on its current data processing environment. The advantages of cloud computing are tempered by two major concerns – security and loss of control. While the public cloud

provides the greatest costs savings, it also incurs the greatest security risk and loss of control, since all of the company's big data is transferred to the cloud service provider. If the data being processed is considered mission critical to the company, the more expensive private cloud, implemented in - house, would provide a more secure environment with the company keeping the mission critical data in-house.

IV. 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. Use of cloud computing for big data implementation lowers the in - house processing power commitment by shifting the data processing to the cloud .Use of big data could provide sufficient benefit to a small to medium sized company to the extent that the business would be willing to commit resources to implement big data technology in-house. 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 . Also the public cloud provides the greatest costs savings, it also incurs the greatest security risk and loss of control, since all of the company's big data is transferred to the cloud service provider.

V. REFERENCES

- [1]. Big Data using Cloud Computing-Bernice M.Purce ll<http://www.aabri.com/OC2013Manuscripts/OC13030.pdf>
- [2]. IEEE Xplore Abstract. Big Data processing in Cloud Environment Satoshi Tsuchiya, Vivian lee.[ieeexplore .ieee.org](http://ieeexplore.ieee.org)
- [3]. Geczy.P.Izuni, N., & Hazida, K.(2012). Cloud sourcing:Managing cloud Adoption. Global Journal of Business Research, 6(2), 57-70.
- [4]. Big Data in the Cloud: Converging Technologies Intel IT center (2015).
- [5]. Cloud computing for e-governance. White paper, IIIT-Hyderabad, January2010
- [6]. Bhaskar Prasad Rimal, Eunmi Choi, and Ian Lumb. A taxonomy and survey of cloud computing systems. Networked Computing and Advanced Information Management, International Conference on, 0:44{51, 2009}.
- [7]. K.I. Juster. Cloud Computing Can Close the Development Gap. <http://www.salesforce.com/assets/pdf/misc/IT-development-paper.pdf>.
- [8]. 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.
- [9]. Cisco. (2009). Infrastructure as a Service: Accelerating time to profitable new revenue streams. Retrieved from http://www.cisco.com/en/US/solutions/collateral/ns341/ns991/ns995/1aaS_BDM_WP.pdf.
- [10]. Eaton, Deroos, Deutsch, Lapis, & Zikopoulos. (2012). *Understanding big data: Analytics forenterprise class Hadoop and streaming data*. New York: Mc Graw-Hill.
- [11]. IOS Press. (2011). Guidelines on security and privacy in public cloud computing. *Journal of EGovernance*, 34 149-151. DOI: 10.3233/GOV-2011-0271.