BIG DATA AND CLOUD COMPUTING

[1]Selvakumar, [2]S. Subathra, [3]Chinnasamy

[1][2][3]Assistant Professor in Computer Science

Dhanalakshmi Srinivasan College of Arts and Science for Women (Autonomous)

Perambalur

ABSTRACT

Big Data may be a data analysis methodology enabled by recent advances in technologies and architecture. However, big data leads to enormous commitment of hardware and processing resources, which prevents the adoption costs of massive data technology for small and medium-sized businesses.Cloud computing offers the promise of massive data implementation to small and medium sized businesses.Big processing is performed through a programming paradigm referred to as MapReduce. Typically, implementation of the MapReduce paradigm requires networked attached storage and multiprocessing . The computing needs of MapReduce programming are often beyond what small and medium sized business are ready to commit. Cloud computing is on-demand network access to computing resources, provided by an outdoor entity. 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). The three sorts of cloud computing are the general public cloud, the private cloud, and therefore the hybrid cloud. A public cloud is that the pay- as-you-go services. a personal cloud is internal data center of a business not available to the overall public but supported cloud structure. The hybrid cloud may be a combination of the general public cloud and personal cloud. Three major reasons for little to medium sized businesses to use cloud computing for giant data technology implementation are hardware cost reduction, processing cost reduction, and skill to check the worth of massive data. the main concerns regarding cloud computing are security and loss ofcontrol.

Keywords: Big Data, Cloud Computing, Private Cloud, Public Cloud, Hybrid Cloud

INTRODUCTION

Big Data can be a data-analysis method implemented by an alternative generation of technologies and architectures that support high-speed data capture, storage and analysis (Willers, Olofsson, & Eastwood, 2011).Data sources that connect to email, mobile device output, sensor-generated data and social media output (Willers, Olofson, & Eastwood, 2011) extend beyond the normal corporate database.Data are not any longer restricted to structured database records but include unstructured data – data having no standard formatting (Coronel, Morris, & Rob, 2013). Big Data requires huge amounts of space for storing . While the worth of storage continued to say no , the resources needed to leverage big data can still pose financial difficulties for little to medium sized businesses. A typical big data storage and analysis infrastructure are going to be supported clustered network-attached storage (NAS) (White, 2011). The clustered NAS infrastructure requires the configuration of multiple NAS "bots" with each NAS "bot" with multiple storage devices connected to one NAS device (White, 2011).

The series of NAS devices are then interconnected to permit massive sharing and searching of knowledge(White,2011).Data storage using cloud computing may be a viable option for little to medium sized businesses considering the utilization of massive Data analytic techniques. Cloud computing is on-demand network access to computing resources which are often provided by an outdoor entity and need little management effort by the business (IOS Press, 2011). variety of architectures and deployment models exist for cloud computing, and these architectures and models are ready to be used with other technologies and style approaches (IOS Press, 2011). Owners of small and medium businesses that do not accept clustered NAS technology can consider different types of cloud computing models to meet their big data needs.Small to medium sized business owners got to consider the right cloud computing soastostaybothcompetitiveandprofitable.

BIG DATA AND THEREFORE THE CLOUD

The term big data springs from the very fact that the datasets are so large that typical database systems aren't ready to store and analyze the datasets (Manyika et al., 2011). The datasets are large because the info is not any longer traditional structured data, but data from many New resources including sensors for email, social media and Internet access (Manica et al., 2011).

The characteristics of massive data present data storage and data analysis challenges to businesses.A typical model for in-house storage of massive data is clustered Network-Attached Storage(Sliwa, 2011). The configuration would begin with a network-attached storage (NAS) pod consisting of several computers attached to a computer used because the (NAS) device. Several NAS pods

would be attached to every other through the pc used because the NAS device.

Clustered NAS storage is an upscale prospect for alittle to medium size business. A cloud services provider can furnish the required space for storing for substantially lower costs.Analyzing big data is completed employing a programming paradigm called MapReduce (Eaton,Deroos, Deutsch, Lapis, & Zikopoulos, 2012). within the MapReduce paradigm, a question is formed anddata are mapped to seek out key values considered to relate to the query; the results are then reducedto a dataset Responding to query (Eden, Terus, Deutsche, Lobis, & Zygopoulos, 2012).

TheMapReduce paradigm requires that massive amounts of knowledge be analyzed. The mapping is doneconcurrently by each separate NAS device; the mapping requires multiprocessing .The multiprocessing needs of MapReduce are costly, and need the configuration noted previously for storage. The processing needs are often met by cloud-service providers.

CLOUD COMPUTNG SERVICE MODELS

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).Cloud deployment solutions can provide services that companies would otherwise not be ready to afford. Businesses also can use cloud deployment solutions as a test measure before adopting a replacement application or technology company-wide.

There are a good number of alternatives for businesses using the cloud for PaaS (Géczy,Izumi, & Hasida, 2012). Platform as a Service is that the use of cloud computing to supply platforms for the event and use of custom applications (Salesforce.com, 2012).Businesses attain cost savings using PaaS through standardization and high utilization of the cloud-based platform across variety of applications (Oracle,2012).

Other advantages of using PaaS include lowering risks by using pretested technologies, promoting shared services, improving software security, and lowering skill requirements needed for brand spanking new systems development (Jackson, 2012). As associated with big data, PaaSprovides companies a platform for developing and using custom applications needed to research large quantities of unstructured data at a coffee cost and low risk during a secure environment. Software as a service provides businesses with applications that are stored and run on virtual servers – within the cloud (Cole, 2012).

The business isn't charged for hardware, just for the bandwidth for the time and number of users necessary (Cole, 2012).

The most advantage of SaaS is that the answer allows businesses to shift the risks

related to software acquisition while moving IT from being reactive to proactive (Carraro & Chong, 2006). The advantages of using SaaS are easy software management, automated updates and connection management, software compatibility across the business, easy collaboration and global access (ROS, 2010a).

Provides companies that analyze big data proven software solutions for software data analysis as a service.The difference between SaaS and PaaS during this case is that SaaS isn't getting to provide a customized solution whereas PaaS will allow the corporate to develop an answer tailored to the company'sneeds.

In the IaaS model, a client business can pay on a per-use basis to be used of kit to support computing operations including storage, hardware, servers, and networking equipment(Rouse, 2010b). Infrastructure as a service is that the cloud computing model receiving the mostattention from the market, with an expectation of 25% of enterprises getting to adopt a service provider for IaaS (Cisco, 2009). Among the services available to businesses through the IaaS model are disaster recovery, accounting as service, storage as service, data center as a service, virtual desktop infrastructure and cloudExplosion, which provides peak load capacity for variable processes (Cisco, 2009).The benefits of IaaS include increased financial flexibility, choice of services, business agility, lower cost scaling and increased security (Cisco, 2009).

While not so far getting used as extensively as PaaS, SaaS, or IaaS, HaaS may be a cloud service based upon the model of your time sharing on minicomputers and mainframes from the 1960s and 1970s (ComputerWeekly.com, 2009). Partitioning was created in the practice of managed services (ComputerWeekly.com, 2009).

In a managed services situation, the managed service provider (MSP) would remotely monitor and administer hardware located at a client's site as contracted (Rouse, 2007). a drag with managed services was the need for a few MSPs to supply hardware on-site for clients, the value of which needed to be built into the MSP's cost (Rouse, 2007). The HaaS model allows the customer to license the hardware directly from theservice provider which alleviates the associated costs (Rouse, 2007). Vendors within the HaaS arena include Google with its Chromebooks for Business, CharTec, and Equus(Panettieri,2011).

TYPES OF CLOUDS

Three sorts of clouds exist – the general public cloud, the private cloud, and therefore the hybrid cloud. Apublic cloud is that the pay- as-you-go services previously discussed available to the overall public(Armbrust et al., 2010). during a public cloud configuration, a business

doesn't own the coretechnology resources and services but outsources these (Géczy, Izumi, & Hasida, 2012). Apublic cloud is taken into account to be an external cloud (Aslam, Ullah, & Ansara, 2010).

A private cloud is internal data center of a business that's not available to the overall public but uses cloud structure (Armbrust et al., 2010). Since the technology is owned and operated by the business, this sort of cloud is costlier than a public cloud, but is additionally safer (Géczy, Izumi, & Hasida, 2012).a personal cloud is an indoor cloud, residing inside the corporate ’s firewall and managed Bythe company (Aslam, Ullah, & Ansara, 2010).

When a corporation uses a hybrid cloud,it uses a public cloud for a few tasks and apersonal cloud for other tasks.When employing a hybrid cloud model, a corporation will use the general public cloud to expedite extra tasks that aren't ready to be easily run within the company’s data In the center or in its individual cloud (Ambrost et al., 2010).

A hybrid cloud allows a corporation to take care of critical, confidential data and knowledge within it firewall while leveraging the general public cloud for nonconfidential data (Aslam, Ullah, & Ansara, 2010). Figure 1 illustrates a hybrid cloud.

The private cloud portion of the hybrid cloud is accessed by company employees, both within the company and on the road, and is maintained by the interior technology group. The private cloud a part of the hybrid cloud is additionally accessed by the corporate employees but is maintained by external service providers. Each portion of the hybrid cloud can hook up with the oppositeportion.

WHICHCLOUDFORYOURDATA?

The type of cloud a corporation uses depends upon the company’s needs and resources. the general public cloud is taken into account the smallest amount secure of the three types, with services and resources ready to be accessed over the web through protocols Accepted by the Provider (Géczy, Izumi, & Hasida, 2012).

The communications protocols adopted by the provider aren't necessarily secure; the selection of using secure or non-secure protocols is up to the provides (Géczy, Izumi, & Hasida, 2012). the general public cloud is additionally the smallest amount costly of the cloud types, with cost savings within the areas of data technology deployment, management, and maintenance (Géczy, Izumi, & Hasida, 2012).

The private cloud company provides services to its employees through an internship (Géczy, Izumi, & Hasida, 2012).If mobile employees are ready to access the private cloud, the access is usually through secure communication protocols (Géczy, Izumi, & Hasida, 2012). All services and resources provided are tailored to the requirements of the business, control over the services and resources

(Géczy, Izumi, & Hasida, 2012). thanks to the financial and human resources needed to deploy, manage, and maintain the knowledge technology resources and services provided, the private cloud is that the costliest sort of cloud (Géczy, Izumi, & Hasida, 2012).

When a business uses a hybrid cloud, the business owns its core information technology resources and services and can host and supply the resources and services in-house (Géczy, Izumi, & Hasida, 2012). Non-critical services are outsourced and maintained in the public cloud (Géczy, Izumi, & Hasida, 2012).In general, the work of important information technology resources and services is important and often confidential (Géczy, Izumi, & Hasida, 2012).Therefore, resources and services that require to be secure are hosted and maintained on the private cloud, with the general public cloud used for other services as a price saving measure (Géczy, Izumi, & Hasida, 2012).

CLOUD COMPUTING FOR GIANT DATA DURING A SMALL TO MEDIUM SIZED BUSINESS

Cloud computing provides an environment for little to medium sized businesses to implement big data technology. Benefits that companies can realize from big data include performance improvement, deciding support, and innovation in business models, products, and services (Manyika et al., 2011). Three major reasons for little to medium sized businesses to use cloud computing for giant data technology implementation are the power toreduce hardware costs, reduce processing costs, and to check the worth of massive data before committing significant company resources. the main concerns regarding cloud computing are security and loss of control (Géczy, Izumi, & Hasida, 2012).

Platform as a Service may be a cloud computing model that gives hardware cost savings.Hardware cost savings are accrued using PaaS through standardization and high utilization of the cloud-based platform across variety of applications (Oracle, 2012). Businesses also can realize hardware cost savings from the SaaS model since the business incurs no additional hardware costs for implementation; the sole costs are for bandwidth supported the time and number of users (Cole, 2012). Hardware as a Service isn't 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 (Rouse, 2007).

In-house processing of massive data typically requires use of the MapReduce programming paradigm (Eaton et al., 2012). The multiprocessing needs of MapReduce entails an enormous commitment of processing power. Use of cloud computing for giant data implementation lowers the in-house processing power commitment by shifting the info processing to the cloud.

The use of massive data could provide sufficient benefit to alittle 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 extent of benefit is difficult to work out without some experience. Cloud computing implementation of massive 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 critical , the business has established a reason to adopt the technology in house. Otherwise, the business can continue cloud computing use of massive data or believe its current processing environment.

The benefits of cloud computing are affected by two main concerns - security and loss of control (Géczy, Izumi, & Hasida, 2012).While the general public cloud provides the best costs savings, it also incurs the best security risk and loss of control company's big data is transferred to the cloud service provider (Géczy, Izumi, & Hasida, 2012).If the info being processed is taken into account mission critical to the corporate , the costlier private cloud, implemented in-house, would offer a safer environment with the corporate keeping the mission critical data in-house.

CONCLUSION

Cloud computing enables small and medium businesses to implement big data technology with less commitment to enterprise resources. The processing capabilities of the large data model could provide new insights to the business concerning performance improvement, deciding support, and innovation in business models, products, and services. The benefits of implementing big data technology through cloud computing are cost savings in hardware and processing because of the ability to test big data technologyconsiderable commitment of company resources. Several models of cloud computing services are available to the companies to think about , with each model having trade-offs between the advantage of cost savings and therefore the concerns data security and loss of control.

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. Aslam, U., Ullah, I, & Ansara, S. (2010, November). Open source private cloud computing.*Interdisciplinary Journal of Contemporary Research in Business. 2*(7), 399-407.Carraro, G., & Chong, F. (2006, October). Software as a service: An enterprise perspective.Retrieved from http://msdn.microsoft.com/en-us/library/aa905332.aspx#enterprisertw_topic3
3. 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/IaaS_BDM_WP.pdf
4. Cole, B. (2012). Looking at business size, budget when choosing between SaaS and hostedERP. *E-guide: Evaluating SaaS vs. on premise for ERP systems.* Retrieved fromhttp://docs.media.bitpipe.com/io_10x/io_104515/item_548729/SAP_sManERP_IO%23104515_EGuide_061212.pdf
5. ComputureWeekly.com. (2009, March). Hardware as a service. Retrieved from http://www.computerweekly.com/feature/Hardware-as-a-Service
6. Coronel, C., Morris, S., & Rob, P. (2013). *Database Systems: Design, Implementation, andManagement*, (10th Ed.). Boston: Cengage Learning.