ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 4, Issue 4, pp: (34-38), Month: October - December 2016, Available at: www.researchpublish.com

Big Data with Cloud Computing

Shubhavi Arya

University of Minnesota, USA

Abstract: **Big Data is a data investigation method empowered by late advances in technologies and engineering. Big data involves an enormous amount of hardware equipment and preparing resources, making adoption expenses of big data methodology restrictive to small and medium organizations. Cloud Computing offers the potential of big data execution to small and medium sized organizations.**

Big Data processing is performed through a programming methodology known as MapReduce. Often, execution of MapReduce methodology requires networked appended storage and parallel processing. The computing requirements of MapReduce programming are frequently past what small and medium sized organizations are able to commit.

Cloud Computing is on-request network access to figuring assets, given by an outside entity. Basic deployment models for cloud computing incorporate platform as a service (PaaS), hardware as a service (HaaS), software as a service (SaaS), and infrastructure as a service (IaaS).

The three sorts of cloud computing are the private cloud, hybrid cloud and public cloud. A private cloud is internal data center of an organization not accessible to the general population but rather takes into account cloud structure. A public cloud is the pay-as-you-go service. The hybrid cloud is a mix of the private cloud and general public cloud.

Three noteworthy explanations behind small to medium sized organizations to utilize cloud computing for big data methodology execution are reduction in hardware cost, processing cost reduction, and capacity to test the estimation of big data. The significant concerns with respect to cloud computing are security and loss of control.

Keywords: **Cloud computing, big data, private cloud, hybrid cloud, public cloud.**

1. INTRODUCTION

Big data is a data analysis technology enabled by a new era of advancements and engineering which bolster high-speed information capture, storage, and analysis (Villars, Olofson, and Eastwood 2011). Information sources reach out past the customary corporate database to incorporate email, sensor-generated data, mobile device output and social media output (Villars, Olofson & Eastwood, 2011). Information is no more confined to organized database records but incorporates unstructured information – information having no standard formatting (Coronel, Morris & Rob, 2013)

Big Data requires gigantic amounts of storage room. While the cost of storage keep on declining, the assets expected to influence big data can in any case posture monetary troubles for little to medium sized organizations. A normal big data storage and examination infrastructure will be based on clustered network-attached storage(NAS) (White, 2011). Clustered NAS framework requires setup of a few NAS "pods" with every NAS "pod" incorporated of a few storage devices associated with a NAS device (White, 2011). The series of NAS devices are then interconnected to permit enormous sharing and searching of information (White, 2011).

Data stockpiling utilizing cloud computing is a practical alternative for small to medium sized organizations considering the utilization of Big Data analytic methods. Cloud computing is on-request network access to computing assets which are frequently provided by an outside entity and requires little administration effort by the business (IOS Press, 2011). Various structures and deployment models exist for cloud computing, and these designs and models can be utilized with different technologies and to plan approaches (IOS Press, 2011). Proprietors of small to medium sized organizations who

can't manage the cost of adoption of clustered NAS technology can consider various cloud computing models to meet their big data needs. Small to medium sized organization entrepreneurs need to consider the right cloud computing model with a specific end goal to stay both competitive, productive and profitable.

2. BIG DATA AND THE CLOUD

The term big data is derived from the fact that the datasets are large to the point that common database frameworks are not able to store and dissect the datasets (Manyika et al., 2011). The datasets are extensive because the information is no more traditional structured information, but information from numerous new sources, including email, web-based social networking, and Internet-available sensors (Manyika et al., 2011). The qualities of big data present data storage and data examination difficulties to organizations.

A typical model for in-house storage of big data is clustered Network-Attached Storage (Sliwa, 2011). The design would start with a network-attached storage (NAS) pod comprising of a few PCs joined to a PC utilized as the (NAS) device. A few NAS pods would be connected to each other through the PC utilized as the NAS device. Clustered NAS storage is a costly prospect for small to medium sized organizations. A cloud services supplier can furnish the essential storage space for significantly lower expenses.

Analyzing big data is done by utilizing a programming paradigm called MapReduce (Eaton, Deroos, Deutsch, Lapis, and Zikopoulos, 2012). In the MapReduce paradigm, a question is made and information is mapped to discover key values considered to identify with the inquiry; the outcomes are then reduced to a dataset answering the question (Eaton, Deroos, Deutsch, Lapis, and Zikopoulos, 2012). The MapReduce paradigm requires that big amounts of information be analyzed. The mapping is done simultaneously by every different NAS device; the mapping requires parallel processing. The parallel processing needs of MapReduce are expensive, and require the configuration noted already for storage. The processing needs can be met by cloud-service suppliers.

3. CLOUD COMPUTING SERVICE MODELS

Typical deployment models for cloud computing incorporate platform as a service (PaaS), infrastructure as a service (IaaS), hardware as a service (HaaS) software as a service (SaaS). Cloud deployment models can provide services that organizations would somehow not have the capacity to afford. Organizations can likewise utilize cloud deployment models as a test measure before embracing another application or innovation company-wide.

There are a wide number of choices for organizations utilizing the cloud for PaaS (Géczy, Izumi, and Hasida, 2012). Platform as a Service is the utilization of cloud computing to provide platforms for the advancement and utilization of custom applications (Salesforce.com, 2012). The PaaS models incorporate application design and development tools, application testing, versioning, deployment, integration, hosting, state management and other related development tools(Géczy, Izumi, and Hasida, 2012). Organizations accomplish cost savings by utilizing PaaS through institutionalization and high use of the cloud-based platform over various applications (Oracle, 2012). Other benefits of utilizing PaaS incorporate bringing down risks by utilizing pretested technologies, advancing shared services, enhancing software security, and bringing down skill necessities required for new systems development (Jackson, 2012). As identified with big data, PaaS gives organizations a platform to create and utilize custom applications needed to investigate large amounts of unstructured information at a low cost and less risk in a secure environment.

Software as a service furnishes organizations with applications that are stored and running on virtual servers – in the cloud (Cole, 2012). The business is not charged for equipment, just for the transmission capacity for the time and number of clients necessary(Cole, 2012). The principle benefit of SaaS is that the technology permits organizations to move the risks connected with software acquisition while moving IT from being responsive to proactive (Carraro and Chong, 2006). Advantages of utilizing SaaS are less demanding software administration, patch management and automatic updates, easier collaboration, global accessibility and software compatibility across the business(Rouse, 2010a). Software as a Service gives organizations examining big data proven software models for data analysis. The distinction amongst SaaS and PaaS for this situation is that SaaS is not going to give a customized arrangement whereas PaaS will permit the organization to build up an solution custom-made to the organization's needs.

In the IaaS model, a client business will pay on a per-use basis for use of hardware to bolster computing operations including storage, equipment, servers, and networking and systems administration hardware (Rouse, 2010b).

Infrastructure as a service is the cloud computing model receiving the most consideration from the market, with an expectation of 25% of endeavors wanting to adopt a service provider for IaaS (Cisco, 2009). Services accessible to organizations through the IaaS model incorporate computer as a service, disaster recovery, data center as a service, cloud bursting, virtual desktop infrastructure, which is providing peak load capacity for variable procedures (Cisco, 2009). Advantages of IaaS incorporate increased monetary flexibility, selection of services, business agility, increased security and cost-effective scalability (Cisco, 2009).

While not used as extensively as PaaS, SaaS, or IaaS, HaaS is a cloud service based upon the model of time sharing on minicomputers and centralized servers from the 1960s and 1970s (ComputerWeekly.com, 2009). Time sharing developed into the practice of managed services (ComputerWeekly.com, 2009). In a managed services situation, the managed service provider(MSP) would remotely monitor and oversee equipment located at a client's site as contracted (Rouse, 2007). An issue with managed services was the need for some MSPs to give equipment on location to customers, the cost of which needed to be incorporated with the MSP's cost (Rouse, 2007). The HaaS model permits the client to license the equipment directly from the service supplier which lightens the related costs (Rouse, 2007). Sellers in the HaaS field incorporate Google with its Chromebooks for Business, CharTec, and Equus (Panettieri, 2011).

4. TYPES OF CLOUDS

Three sorts of cliuds exist – the hyrbid cloud, the public cloud, the private cloud. A public cloud is the pay-as-you-go service about accessible to the general population (Armbrust et al., 2010). In a public cloud design, a business does not possess the core technology assets and services but rather outsources these (Géczy, Izumi, and Hasida, 2012). A public cloud is considered to be an external cloud (Aslam, Ullah, and Ansara, 2010).

A private cloud is internal data center of a business that is not accessible to the general population but rather utilizes cloud structure (Armbrust et al., 2010). In a private cloud arrangement, assets and services are possessed by the business, with the services accesible within the business through the intranet (Géczy, Izumi, and Hasida, 2012). Since the technology is possessed and operated by the business, this kind of cloud is more costly than a public cloud, but at the same time is more secure (Géczy, Izumi, and Hasida, 2012). A private cloud is an internal cloud, living inside the organization's firewall and oversaw by the organization (Aslam, Ullah, and Ansara, 2010).

At the point when an organization uses a hybrid cloud, it utilizes a public cloud for a few tasks and a private cloud for different assignments. At the point when utilizing a hybrid cloud design, an organization will utilize the public cloud to expedite additional tasks that are not ready to be effectively keep running in the organization's data center or on its private cloud (Armbrust et al., 2010). A hybrid cloud permits an organization to maintain important, private information and data within its firewall while utilizing the public cloud for non-classified information (Aslam, Ullah, and Ansara, 2010). Figure 1 represents a hybrid cloud. The private cloud part of the hybrid cloud is accessed by organization's workers, both in the organization and on the road, and is kept up by the internal technology group. The private cloud part of the hybrid cloud is likewise accessed by the organization's workers yet is maintained by external service providers. Every part of the hybrid cloud can connect with the other part.

5. WHICH KIND OF CLOUD FOR YOUR DATA

The kind of cloud an organization utilizes depends on the organization's needs and assets. The public cloud is viewed as the least secure of the three sorts, with administrations and assets ready to be accessed over the Internet through protocols embraced by the provider (Géczy, Izumi, and Hasida, 2012). The communications protocols adopted by the provider are not really secure; the decision of utilizing secure or non-secure protocols is up to the provider (Géczy, Izumi, and Hasida, 2012). The public cloud is likewise the least costly of the cloud types, with cost savings in the areas of data innovation deployment, administration, and support (Géczy, Izumi, and Hasida, 2012).

The private cloud gives services to organization employees through an intranet (Géczy, Izumi, and Hasida, 2012). If mobile employees can acess the private cloud, the access is typically through secure communication protocols (Géczy, Izumi, and Hasida, 2012). All services and assets provided are custom fitted to the necessities of the business, and the business has total control over the services and assets (Géczy, Izumi, and Hasida, 2012). Due to the monetary related and HR required to send, oversee, and maintain the data innovation assets and services provided, the private cloud is the most costly kind of cloud (Géczy, Izumi, and Hasida, 2012).

When a business uses a hybrid cloud, the business owns its core information technology assets and services and will host and provide the assets and services in-house (Géczy, Izumi, and Hasida, 2012). Non-critical services are outsourced and maintained on a public cloud (Géczy, Izumi, and Hasida, 2012). Typically, core information technology assets and services are mission-critical and are frequently classified (Géczy, Izumi, and Hasida, 2012). Subsequently, assets and services that should be secure are facilitated and maintained on the private cloud, with public cloud utilized for different services as a cost saving measure (Géczy, Izumi, and Hasida, 2012).

6. CONCLUSION

Cloud computing enables small to medium sized organizations to actualize big data technology with a reduced commitment of organization assets. The handling capacities of the big data model could provide new insights to the business pertaining to performance improvement, basic leadership support, and advancement in business models, products, and services. Advantages of executing big data methodology through cloud computing are cost savings in equipment and processing, and additionally the capacity to explore different avenues regarding big data before making a significant commitment of organization assets. A few models of cloud computing services are accessible to the organizations to consider, with every model having trade-offs between the benefit of cost savings and the concerns of information 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.

[3] 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

[4] 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

[5] Cole, B. (2012). Looking at business size, budget when choosing between SaaS and hosted ERP. E-guide: Evaluating SaaS vs. on premise for ERP systems. Retrieved from http://docs.media.bitpipe.com/io_10x/io_104515/item_548729/SAP_sManERP_IO%23104515_EGuide_061212.pdf

[6] ComputureWeekly.com. (2009, March). Hardware as a service. Retrieved from http://www. computerweekly.com/feature/Hardware-as-a-Service

[7] Coronel, C., Morris, S., & Rob, P. (2013). Database Systems: Design, Implementation, and Management, (10th Ed.). Boston: Cengage Learning.

[8] Eaton, Deroos, Deutsch, Lapis, & Zikopoulos. (2012). Understanding big data: Analytics for enterprise class Hadoop and streaming data. New York: McGraw-Hill.

[9] Géczy, P., Izumi, N., & Hasida, K. (2012). Cloud sourcing: Managing cloud adoption. Global Journal of Business Research, 6(2), 57-70.

[10] IOS Press. (2011). Guidelines on security and privacy in public cloud computing. Journal of E-Governance, 34 149-151. DOI: 10.3233/GOV-2011-0271

[11] Jackson, K. L. (2012). Platform-as-a-service: The game changer. Retrieved from http://www.forbes.com/sites/kevinjackson/2012/01/25/platform-as-a-service-the-game-changer/

[12] Manyika, J., Chui. M. Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011, June). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. Retrieved from http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_ innovation

[13] Oracle. (2012). Oracle platform as a service. Retrieved from http://www.oracle.com/ us/technologies/cloud/oracle-platform-as-a-service-408171.html

[14] Panettieri, J. (2011, June 13). Can Google take hardware as a service (HaaS) mainstream? MSP Mentor. Retrieved from http://www.mspmentor.net/2011/06/13/can-google-take-hardware-as-a-service-haas-mainstream/

[15] Rouse, M. (2010a, August). Software as a service. Retrieved from http://searchcloudcomputing. techtarget.com/ definition/Software-as-a-Service

[16] Rouse, M. (2010b, August). Infrastructure as a Service. Retrieved from http://searchcloudcomputing.Techtarget. com/definition /Infrastructure-as-a -Service-IaaS

[17] Rouse, M. (2007, December). Hardware as a service. Retrieved from http://searchitchannel.techtarget.com/ definition/Hardware-as-a-Service-in-managed-services

[18] Salesforce.com. (2012). The end of software: Building and running applications in the cloud. Retrieved from http://www.salesforce.com/paas/

[19] 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-object-storage-cloud-gateways-replacing-file-storage

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

APPENDIX - A