Volume-4, Issue-12, Dec-2015 • ISSN No 2277 - 8160

Research Paper | Computer Science

Use and Impact of Big Data in Cloud Computing

Gunjan Sharma Assistant Professor, Department of Computer Science, D.A.V College Chandigarh.

Neha Arora Assistant Professor, Department of Computer Science, D.A.V College Chandigarh.

Anita Rai Assistant Professor, Department of Computer Science, D.A.V College Chandigarh.

ABSTRACT

Big data is an all-encompassing term for any collection of data sets which are large and complex it contain structured and unstructured both type of data. As big data comes from everywhere, sensors used to gather climate information, digital pictures and videos etc. We all know that big data is a new term used to identify the datasets that due to their large size and complexity, we cannot manage them with our current methodologies. Cloud computing offers the promise of big data implementation to small and medium sized businesses. Big Data processing is performed through a programming paradigm known as MapReduce. Typically, implementation of the MapReduce paradigm requires networked attached storage and parallel processing. The computing needs of MapReduce programming are often beyond what small and medium sized business are able to commit.

Cloud computing is on-demand network access to computing resources, provided by an outside entity. Common deployment models for cloud computing include platform as a service (PaaS), software as a service (SaaS), infrastructure as a service (IaaS).The three types of cloud computing are the public cloud, the private cloud, and the hybrid cloud. A public cloud is the pay- as-you-go services. A private cloud is internal data center of a business not available to the general public but based on cloud structure. The hybrid cloud is a combination of the public cloud and private cloud.

Three major reasons for small to medium sized businesses to use cloud computing for big data technology implementation are hardware cost reduction, processing cost reduction, and ability to test the value of big data.The major concerns regarding cloud computing are security and loss of control.

KEYWORDS : Big data, cloud computing, private cloud, public cloud, hybrid cloud

INTRODUCTION:

Enormous amounts of data are collected every year from various sensors, devices in different formats, from independent or connected applications. Let us consider the Internet data. The web pages indexed by Google were around one million in 1998, but quickly reached 1 billion in 2000 and have already exceeded 1 trillion in 2008. This rapid expansion is accelerated by the dramatic increase in acceptance of social networking applications, such as Facebook, Twitter, etc., that allow users to create contents freely and amplify the already huge Web volume. As in today world mobile phones becoming the sensory gateway to get real time data on people from different aspects, the vast amount of data that mobile carries can potentially improve our daily life has significantly outpaced our past CDR (call data record)-based processing for billing purposes only. It can be foreseen that by having internet applications it will raise the scale of data to an unprecedented level. People and devices from home, to cars, to buses, railway station and airports are all loosely connected. The trillions of such connected components will generate a huge data and valuable information must be discovered from the data which help us to improve the quality of life and make our world a better place. Scalability is at the core of the expected new technologies to meet the challenges coming along with big data. The simultaneously emerging and fast maturing cloud computing technology delivers the most promising platforms to realize the needed Scalability with demonstrated elasticity and parallelism capacities. Numerous notable attempts have been initiated to exploit massive parallel processing architectures (Berkovich, S., Liao, D. (2012)). Google's novel programming model, Map Reduce (Beyer, M.A., Laney, D. (2012)).and its distributed file system, GFS (Google File System) (Beyer, M.A., Laney, D. (2012)).represent the early groundbreaking efforts made in this line.

Background Information

CLOUD COMPUTING:

Cloud computing can be defined as a new style of computing in which dynamically scalable and often virtualized resources over internet. As computing services initially offered by commercial providers, such as Amazon, Google, and Microsoft (Rajkumar Buyya, 2011). The key concept of cloud computing is to deliver computing resources through a global network when and where the customer requests. With the cloud computing technology, users use a variety of devices, including laptops, smartphones, PCs.

As cloud based apps can be up and running in days or weeks, and they cost very less. With this app, user just open a browser, log in, customize the app and start using it.

Categories/ Classification of cloud computing services

From a service point of view, cloud computing includes 3 models: software, platform, and infrastructure

- *Software as a service (SaaS):* It includes a complete software offering on the cloud. As this application are hosted by a cloud service provider and made available to customer over a network, typically the internet.
- *Infrastructure as a service (IaaS):* As this service offering hardware related services using the principles of cloud computing the equipment used to support operations, including storage, hardware, servers, and networking components.
- *Platform as a service (PaaS):* With PaaS, developers can build Web applications without installing any tools on their computer, and then deploy those applications without any specialized administrative skills.

Benefits of Cloud Computing:

- **Shared Infrastructure: -** As shared infrastructure uses a virtualized software model, enabling the sharing of physical services, networking and storage capabilities. The cloud infrastructure, regardless of deployment model, seeks to make the most of the available infrastructure across a number of users.
- **Economical: -** Cloud computing is an approach to IT, in which a low initial investment is required to get going.
- **Network Access: -** Network access needs to be accessed across the internet from a broad range of devices such as PCs, laptops, and mobile devices. Deployments of services in the cloud include

everything from using business applications to the latest application on the smart phones.

- **Increased Effectiveness: -** Cloud computing frees the user from the finer detail of IT system configuration and maintenance, enabling them to spend more time on mission-critical tasks and less time on IT operation and maintenance.
- **Energy Efficient: -** Because resources are pooled, each user community does not need to have its own dedicated IT infrastructure.

Models of Cloud Computing

To deploy cloud computing, the US National Institute of Standards and Technology (NIST) listed 4 models:-

(1) *Public cloud:* In public cloud the computing infrastructure is hosted by the cloud vendor at the vendor premises. The customer has no visibility and control over where the computing infrastructure is hosted.

(2) *Private cloud:* A cloud infrastructure is operated solely for a single organization. In other words, the proprietary network or the data center supplies hosted services to a certain group of people.

(3) *Community cloud:* The cloud infrastructure is shared by several organizations with common concerns (e.g., mission, security requirements, and policy). For example, the Google Gov Cloud provides the Los Angeles City Council with a segregated data environment to store its applications and data that are accessible only to the city's agencies.

(4) *Hybrid cloud:* The cloud infrastructure comprises 2 or more clouds (private, public, or community). In this infrastructure, an organization provides and manages some resources within its own data center and has others provided externally.

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 (Géczy, Izumi, & Hasida, 2012).

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 (Rouse, 2007).

In-house processing of big data typically requires use of the MapReduce programming paradigm (Eaton et al., 2012). The parallel processing needs of MapReduce 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 (Géczy, Izumi, & Hasida, 2012). 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 (Géczy, Izumi, & Hasida, 2012). 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.

The Future State

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 trade-offs between the benefit of cost savings and the concerns data security and loss of control.

REFERENCES

1. Berkovich, S., Liao, D.: On Clusterization of big data Streams. In: 3rd International Conference on Computing for Geospatial Research and Applications, article no. 26. ACM Press, New York (2012) 2. Beyer, M.A., Laney, D.: The Importance of 'Big Data': A Definition. Gartner (2012) 3. 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 4. Eaton, Deroos, Deutsch, Lapis, & Zikopoulos. (2012). Understanding big data: Analytics for enterprise class Hadoop and streaming data. New York: McGraw-Hill. 5. Géczy, P., Izumi, N., & Hasida, K. (2012). Cloudsourcing: Managing cloud adoption. Global Journal of Business Research, 6(2), 57-70. 6. 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. 7. Oracle. (2012). Oracle platform as a service. Retrieved from http://www.oracle.com/ us/ technologies/cloud/oracle-platform-as-a-service-408171.html 8. Rouse, M. (2007, December). Hardware as a service. Retrieved from http://searchitchannel .techtarget.com/ definition/Hardware-as-a-Service-in-managed-services