Syamanta Garud *et al*, International Journal of Computer Science and Mobile Applications,
National Conference on "The Things Services and Applications of Internet of Things",
Gandhi Institute for Education and Technology (GIET) Baniatangi, 23-24 March 2018, pg. 34-38
ISSN: 2321-8363
Impact Factor: 5.515

Big Data and Cloud Computing: An Improved Tool for Scientific Analytics and Data Governance

Syamanta Garud[1], Ranjita Rout[2]

[1]*Department of Computer Science and Engineering, Centurion University, BBSR,* syamantagarud@rediffmail.com
[2]*Department of Computer Science and Engineering, GIET, Baniatangi, BBSR,* rout.ranjita@gmail.com

Abstract

Big Data' as term came to limelight under the extensive use of global data as a technology that is able to store and process big and varied volumes of data, providing both enterprises and science with deep insights over its clients/experiments. A new concept called cloud computing provides a reliable, fault-tolerant, available and scalable environment to harbour big data distributed management systems. This paper given as overview of both technologies and cases of success when integrating big data and cloud frameworks. This paper also contents an overview of both cloud and big data technologies are converging to offer a cost-effective delivery model for cloud-based big data analytics. It also includes:

- How cloud computing is an enabler for advanced analytics with big data
- How IT can assume leadership for cloud-based big data analytics in the enterprise by becoming a broker of cloud services
- Analytics-as-a-service (AaaS) models for cloud-based big data analytics
- Practical next steps to get you started on your cloud-based big data analytics initiative

No doubt some of our current problem is solved, it still presents some gaps and issues that arise concern and need improvement. Security, privacy, scalability, data governance policies, data heterogeneity, disaster recovery mechanisms, and other challenges are yet to be addressed. Other concerns are related to cloud computing and its ability to deal with exabytes of information or address exaflop computing efficiently.

Keywords*:* Big Data, Cloud Computing, data heterogeneity, exabytes, exaflop, AaaS, Security, privacy, scalability

Introduction

At present, two IT initiatives are considered as significant 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.

ISSN: 2321-8363
Impact Factor: 5.515

THE CONCEPT OF BIG DATA ANALYTICS

Big data refers to huge data sets that are orders of magnitude larger *(volume)*; more diverse, including structured, semi structured, and unstructured data *(variety)*; and arriving faster *(velocity)* than you or your organization has had to deal with before. This flood of data is generated by connected devices—from PCs and smart phones to sensors such as RFID readers and traffic cams. Plus, it's heterogeneous and comes in many formats, including text, document, image, video, and more.
The real value of big data is in the insights it produces when analyzed—discovered patterns, derived meaning, indicators for decisions, and ultimately the ability to respond to the world with greater intelligence. Big data analytics is a set of advanced technologies designed to work with large volumes of heterogeneous data. It uses sophisticated quantitative methods such as machine learning, neural networks, robotics, computational mathematics, and artificial intelligence to explore the data and to discover interrelationships and patterns.

BIG DATA & CLOUD COMPUTING

Big data has evolved as major force of invotation across both academic and corporations. The paradigm is deemed as endeavor to understand and get proper knowledge from bug data set (big data analytics) providing summarized information over huge data loads. As such, this paradigm is considered by corporations as a technique to understand their clients, to get clouser to them, find pattern and predict trends. Furthermore, big data is viewed by scientists as a mean to store and process huge scientific datasets. This concept is a hot topic and is expected to continue to grow in popularity in the coming years.
Although big data is mostly associated with the storage of huge loads of data it also concerns ways to process and extract knowledge from it (Hashem et al., 2014). The five different aspects used to describe big data (commonly referred to as the five "V"s) are Volume, Variety, Velocity, Value and Veracity (Sakr & Gaber, 2014):

- *Volume* describes the size of datasets that a big data system deals with. Processing and storing big volumes of data is rather difficult, since it concerns: scalability so that the system can grow; availability, which guarantees access to data and ways to perform operations over it; and bandwidth and performance.
- *Variety* concerns the different types of data from various sources that big data frameworks have to deal with.
- *Velocity* concerns the different rates at which data streams may get in or out the system and provides an abstraction layer so that big data systems can store data independently of the incoming or outgoing rate.
- *Value* concerns the true value of data (i.e., the potential value of the data regarding the information they contain). Huge amounts of data are worthless unless they provide value.
- *Veracity* refers to the trustworthiness of the data, addressing data confidentiality, integrity, and availability. Organizations need to ensure that data as well as the analyses performed on the data are correct.

Cloud computing is another paradigm which promises theoretically unlimited on-demand services to its users. Cloud's ability to virtualize resources allows abstracting hardware, requiring little interaction with cloud service providers and enabling users to access terabytes of storage, high processing power, and high availability in a *pay-as-you-go* model (González-Martínez et al., 2015). Moreover, it transfers cost and responsibilities from the user to the cloud provider, boosting small enterprises to which getting started in the IT business represents a large endeavour, since the initial IT setup takes a big effort as the company has to consider the total cost of ownership (TCO), including hardware expenses, software licenses, IT personnel and infrastructure maintenance. Cloud computing provides an easy way to get resources on a pay-as-you-go model, offering scalability and availability, meaning that companies can easily negotiate resources with the cloud

Syamanta Garud *et al*, International Journal of Computer Science and Mobile Applications,
National Conference on "The Things Services and Applications of Internet of Things",
Gandhi Institute for Education and Technology (GIET) Baniatangi, 23-24 March 2018, pg. 34-38
ISSN: 2321-8363
Impact Factor: 5.515

provider as required. Cloud providers usually offer three different basic services: Infrastructure as a Service (IaaS); Platform as a Service (PaaS); and Software as a Service (SaaS):

- *IaaS* delivers infrastructure, which means storage, processing power, and virtual machines. The cloud provider satisfies the needs of the client by virtualizing resources according to the service level agreements (SLAs);
- *PaaS* is built atop of IaaS and allows users to deploy cloud applications created using the programming and run-time environments supported by the provider. It is at this level that big data DBMS are implemented;
- *SaaS* is one of the most known cloud models and consists of applications running directly in the cloud provider;

These three basic services are closely related: SaaS is developed over PaaS and ultimately PaaS is built atop of IaaS.

From the general cloud services other services such as Database as a Service (DBaaS) (Oracle, 2012), BigData as a Service (BDaaS) and Analytics as a Service (AaaS) arose.

Since the cloud virtualizes resources in an on-demand fashion, it is the most suitable and compliant framework for big data processing, which through hardware virtualization creates a high processing power environment for big data.

BIG DATA IN THE CLOUD

Storing and processing big volumes of data requires scalability, fault tolerance and availability. Cloud computing delivers all these through hardware virtualization. Thus, big data and cloud computing are two compatible concepts as cloud enables big data to be available, scalable and fault tolerant.

Business regards big data as a valuable business opportunity. As such, several new companies such as Cloudera, Hortonworks, Teradata and many others, have started to focus on delivering Big Data as a Service (BDaaS) or DataBase as a Service (DBaaS). Companies such as Google, IBM, Amazon and Microsoft also provide ways for consumers to consume big data on demand. Next, we present two examples, Nokia and RedBus, which discuss the successful use of big data within cloud environments.

BIG DATA TRENDS

Cloud computing is regarded as a cost effective delivery model for bug data analytics. It is the ripe time to know that whether big data and cloud technologies converging to make big data analytics in clouds a reasonable option For big data analytics:

Data is becoming more valuable. Today the conversation is shifting from the problems of storing data what data to have a proper utilities of the data Enterprises are looking to unlock data's hidden potential and deliver competitive advantage. Gartner predicts that enterprise data will grow by 800 percent from 2011 to 2015, with 80 percent unstructured (for example, e-mails, documents, video, images, and social media content) and 20 percent structured (for example, credit card transactions and contact information).1

With the potential for so much data to reveal insights that can boost competitiveness, companies must find new approaches to processing, managing, and analyzing their data—whether it's structured data typically found in traditional relational database management systems (RDBMSs) or more varied, unstructured formats. Plus, combining diverse data sources and types has the potential to uncover some of the most interesting unexplored patterns and relationships.

Data analytics is moving from batch to real time. Intel's 2012 survey of 200 IT managers in large enterprises found that while the amount of batch versus real-time processing is split evenly today, the trend is toward increasing real time to two-thirds of total data management by 2015.2 At the same time, the technology for processing real-time or near-real-time information is moving past hype to early stages of maturity.

Real time supports predictive analytics. Predictive analytics enables organizations to move to a future-oriented view of what's ahead and of fers organizations some of the most exciting opportunities for driving value from big data.

ISSN: 2321-8363
Impact Factor: 5.515

Real-time data provides the prospect for fast, accurate, and flexible predictive analytics that quickly adapt to changing business conditions. The faster you analyze your data, the more timely the results, and the greater its predictive value.
The scope of big data analytics continues to expand. Early interest in big data analytics focused primarily on business and social data sources, such as e-mail, videos, tweets, Facebook* posts, reviews, and Web behavior. The scope of interest in big data analytics is growing to include data from intelligent systems, such as in-vehicle infotainment, kiosks, smart meters, and many others, and device sensors at the edge of networks—some of the largest-volume, fastest-streaming, and most complex big data. Ubiquitous connectivity and the growth of sensors and intelligent systems have opened up a whole new storehouse of valuable information. Interest in applying big data analytics to data from sensors and intelligent systems continues to increase as businesses seek to gain faster, richer insight more cost-effectively than in the past, enhance machine-based decision making, and personalize customer experiences.

CONCLUSION

In every organization, IT is deemed as a special entity. In spite of extensive data growth, emerging technology and quick change, one can provide requisite leadership within the organization for big data analytics. First and foremost, consider how IT can evolve as the broker for big data analytics cloud-based services for your business. You can also: Partner with business owners now to help determine the utilities of big data to have a solution of organization business problems and grabbing opportunities. As a real time partner, one can assess impact the choice of technology and establish benchmarking. Explore technology options for cloud-based big data analytics, including private, public, and hybrid delivery models. Keep up to date with trends, watch the market, and understand costs. Create or update an existing big data strategy that defines the process that the big data analytics projects as an emerging IT tool, should be made and easy process eliminating the complication and should be used with a faster pace so that business units shall be not take the matters into their own hand. Consider how to organize IT to better engage with business users and collaborate and consult on big data projects.

REFERENCES

1. G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. (references)
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
4. R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.
5. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
6. M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.
7. V.V. Mayer, and K. Cukier, Big Data: A Revolution that will transform How we Live, Work and Think, John Murray Press, 2013.
8. Chervenak, I. Foster, C. Kesselman, C. Salisbury, and S. Tuecke, The Data Grid: Towards an architecture for the distributed management and analysis of large scientific datasets, Journal of Network and Computer Applications 23 (3) (2000) 187-200.

ISSN: 2321-8363
Impact Factor: 5.515

9. J. Dean, and S. Ghemat, MapReduce: Simplified Data Processing on Large Cluster, Communications of the ACM, 51(1) (2008): 107-113.
10. W. Hoschek, J. JaenMatrinez, A. Samar, H. Stockinger, and K. Stockinger, Data Management in International Data Grid Project, Proceedings of the 1st IEEE/ACM International Workshop on Grid Computing, 2000, Springer Verlag Press.
11. H. Nakada, M. Sato, and S. Sekiguchi, Design and Implementation of Ninf, Towards a Global Computing Infrastructure, 15(5-6):649–658, 1999.
12. T. H. Davenport, At the Big Data Crossroads: turning towards a smarter travel experience, 2013 Amadeus IT Group.

