JOURNAL OF CONTEMPORARY TRENDS IN BUSINESS AND INFORMATION TECHNOLOGY (JCTBIT)
Vol.1, No.1, pp.70-80, December 2016

Impact of Cloud Computing and Big Data Analytics on Business

Sonia V. Soans
Lecturer, Information Technology Department
Nizwa College of Technology
Nizwa, Sultanate of Oman

Abstract

Business organisations are working more effectively on producing smarter outputs which can support financial benefits, and revenue generation simultaneously lowering the cost of the business risks. In order to do this it is necessary to have an appropriate perception of the business and this is very well done through business analytics which helps in increasing the business accomplishments. Business Analytics using the substantial data generated by big data and cloud is preferably used for achieving appropriate results. This paper highlights the use of cloud computing and the big data which can be used for analytics to generate useful inference.

Keywords: Big Data, Cloud Computing, Business Analytics.

I. Introduction:

Perpetually, it is coming to notice that both big data and cloud computing are very important tools for analytics. Big data analytics analyzes structured and unstructured data to generate necessary results.

With big data analytics, companies are able to gain perception and a great deal of insight from unstructured and semi structured data even without directly communicating with the stakeholders. Companies are discovering better ways to improve the internal operations, stakeholder outreach and a host of other critical processes.

Cloud Computing has a lot of advantages. A company can upgrade its data storage and data warehousing capabilities in a scalable, cost-efficient manner within the organization and store in the Cloud.

Computing possibilities are significantly improved with the use of cloud services. (Rittinghouse, John W., Ransome James F., 2010, p. 6-7)

ISSN: 2521-3288

Cloud based services and big data analytics can have a great impact on the business organization growth.

Big Data providers also sell solutions based on cloud which help the business organisations to access processing power and the software applications installed on the networks.

Big data solutions derived from Cloud are leading to higher profits in the business world. These solutions include connecting the increasing abilities of data warehouse which is needed for effective data analysis, ensuring the availability of big data analytics tools to all related and saving time when firms strive to monetize differentiated data. Cloud computing and big data used separately gives a substantial profit but when they are used together they become a very convenient and versatile solution for the future generation.

In order to reach these objectives of better performance, a major change is needed in the analytics technology. Big data analytics opens a wide scope of analytical tools which are used by the organisations. Big data presents a lot of new sources of information that enriches and intensifies the knowledge about situations, issues, customers and organisations to enhance analytic results and create new opportunities. The need for quality customer interactions is also on the rise. Consumers have a multitude of choices related to who they do business with and how they choose to interact with each vendor. Virtualisation of data and services is done in Cloud which impacts how they are both managed. The pressure to embrace these shifts and to adapt analytical strategies around them is vital to deliver accurate, timely and competitive insight.

II. Big Data and Analytics of Data:

"Big Data" was first introduced to the internet and IT world by *Roger Magoulas*, 2005 in order to define a great amount of data that traditional data management techniques could not manage and process due to the complexity and size of the data. Big Data is classified as:

Volume: refers to the huge amount of data that is gathered over the time.

Velocity: refers to the speed in which Big Data can be processed. Some activities are very important and need immediate action or response that is why fast processing maximizes efficiency.

Variety: refers to the different types of data that Big Data can be made of. This data may be structured or unstructured type.

ISSN: 2521-3288

Veracity: refers to the degree in which an Analytic trusts the used information in order to take decisions.

In other words Big Data is considered to be a technology that consists of different types of data sets which may be very large, highly complex, unstructured or structured. This Big Data is organised, stored and processed using specific methods and techniques which maybe further used for business processes.

III. Cloud Computing:

Cloud Computing is also a known term. It is considered as a technology which provides virtualized third-party hosting. (Soans, Sonia V., 2012, p.1-3) That is, rather than renting part or all of an actual physical server from a hosting company, the client rents a certain amount of server resources. The server runs inside a virtual container which can be moved from one physical server to another without interruption of service. Such a container is also capable of spanning multiple physical machines, giving it potentially limitless resources. A web server typically has three tiers to it: The physical infrastructure, the operating system platform, and the web application software being run. A cloud container may contain one, two or all of these layers. (Vadalasetty, 2003, p. 10-12)

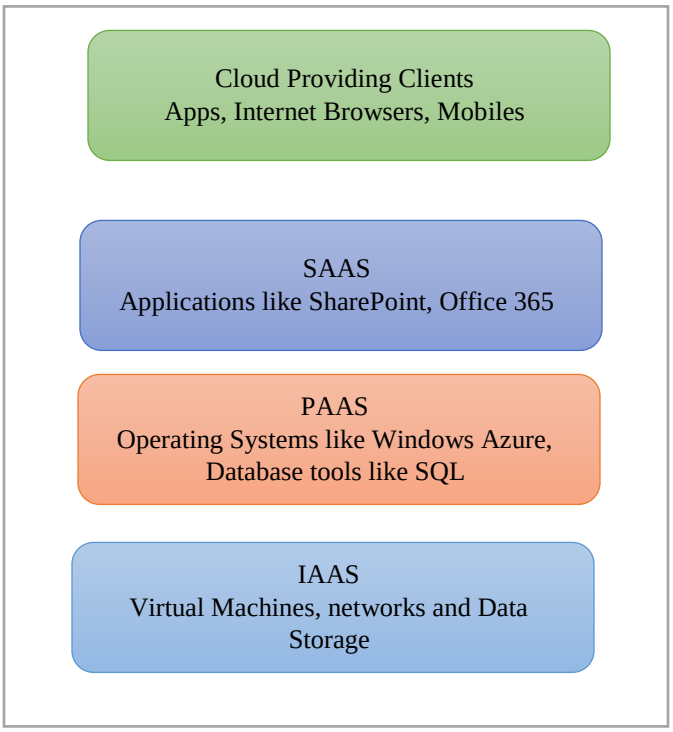
A fully operational server with total control over the operating system and configuration, can mean Infrastructure as a Service, while something in between, with a pre-configured operating system but a reasonable amount of control over the rest of the configuration is Platform as a Service and only uploading the files and leaving the website to run, indicates Software as a Service. In Cloud Computing we also find that it allows accessing the hardware, virtualization, large data storage, various platforms for software development, software usage, a number of applications in a series of services.

Thus Cloud Computing can also be reframed as a layered structure of its services. Public clouds are available from Google.com, Amazon.com, Microsoft, Oracle/Sun, Canonical/Eucalyptus and many other vendors. Private cloud technologies, where the cloud software is loaded on local or in-house server hardware are available from VMware, Eucalyptus, Citrix, Microsoft, and there are thousands of vendors offering cloud solutions of all sorts.

ISSN: 2521-3288

IV. Cloud Models:

Cloud Models can be segmented into Software as a Service (SaaS), Platform as a service (PaaS) and Infrastructure as a Service (IaaS) [5]. When an organisation is considering Cloud security (Krutz, Ronald L., Vines, Russell Dean 2010 p 8-10), it should consider both the differences and similarities between these three segments of Cloud Models. (Rani, Dimpi, Ranjan, Rajiv Kumar, 2014, p.2-4)



A. SAAS

SaaS focuses on managing access to applications. For example, policy controls may dictate that a salesperson can only download particular information from sales CRM applications since they are only permitted to download certain leads, within certain geographies or during local office working hours. In effect, the security officer needs to focus on establishing controls regarding users' access to applications. (Vadalasetty, Sreenivasa Rao., 2003, p.10-12)

ISSN: 2521-3288

B. PAAS

PaaS focuses on protecting data. This is especially important in the case of storage as a service. An important element to consider within PaaS is the ability to plan against the possibility of an outage from a Cloud provider. The security operation needs to consider providing for the ability to load balance across providers to ensure fail over a series of services in the event of an outage. Another key consideration should be the ability to encrypt the data whilst stored on a third-party platform and to be aware of the regulatory issues that may apply to data availability in different geographies. (Armbrust, M., et.all, p.1-4)

C. IAAS

IaaS focuses on doing the engineering, investing in the infrastructure, deploying it and maintaining it in such a way such that the clients can get on with business quickly and effectively without any initial investment or tiresome installations. As the business grows they know that there is capacity available to them on demand. Private cloud deployments are making cloud computing a reality for many businesses apart from public cloud. Similarly IT organisations are evolving to support cloud services, which further enhances big data analytics a believable effort. As a result business organisations are developing the growing trust in the cloud delivery models. Cloud models offer flexible approaches to each business user's request. Further using this cloud infrastructure to analyze the big data makes a lot of sense because organisations investing huge amounts of money on data analytics, must be convinced about the cost effectiveness, and its significance.

In this regard, private clouds offer typically more efficient and less expensive model in order to implement analysis of big data as internal resources are augmented with public cloud services. The hybrid cloud which is a mix of Public and Private clouds, helps the business organisations to use on-demand storage space and computing power in order to perform certain data analytics. Big data forms a combination of internal and external data. Big organisations do not like to share the sensitive data, yet large amount of data is shared with the external storage due to third party ownership, this external storage could be the cloud. (Feng, J., Chen, Y., Liu, P., 2010, p. 1-2) Analyzing this data wherever it resides, is the task to be performed. A lot of data services are needed to extract sensible data from the much collected big data. Thus it is now proposed to have AaaS that is Analytics As A Service.

ISSN: 2521-3288

The AaaS framework usually possesses the capabilities like obtaining all forms of data from the trusted sources, which can be filtered for use. This data has to be managed following all the rules and regulations, pertaining to various ecommerce policies. And finally integrating the data, analyzing the data, transforming and obtaining the right outcomes is the result.

New and innovative methods to examine how most organisations can be transformed using Big Data Analytics are being defined. (http://www.oracle.com/technetwork/database...) In this way, the organisations are learning that Big Data is not a single technology and it is rather a trend across many areas of business and technology. It is also being realized that the data which is being termed as Big Data is not only diverse in nature but also changing very fast. Thus new technologies have to be adopted to realize the value of this data and also utilize it for any behavioral trends or to improve various campaigns, pricing and stock age.

Not only business, Governments can detect and track the emergence of any disease outbreaks via social media signals. ***Oil and gas*** companies can use the output obtained from sensors in their drilling equipment to make more efficient and safer drilling decisions. ***Logistics*** can keep track of their vehicle movement and maps to generate shorter routes to their destination which can save huge amount of money.

The features of Big Data being so large and complex it is literally impractical to manage with the traditional software tools. A growing number of companies are using the technology to store and analyze petabytes of data including click stream data, web logs and data gathered from social media to gain better insights about their stakeholders and their businesses. As a result, information classification becomes even more critical; and information ownership must be addressed to facilitate any reasonable classification.

ISSN: 2521-3288

Most organisations are already struggling by implementing these concepts, thus making it a significant challenge. It is necessary to identify owners for the outputs of Big Data processes, as well as the raw data. Since raw data is processed to give valuable information, data ownership is distinct from information ownership – where IT owns the raw data and business units take the responsibility of the outputs.

Organisations do not wish to build a Big Data environment in-house since it would be cost provoking and work taxing, so cloud and Big Data are inextricably linked. Here an issue rises, as many businesses are aware that, storing all the big data in the cloud does not remove their responsibility for protecting it - from both a regulatory and a commercial perspective. This leads to the thought that the security is also an important issue here. Various security techniques are implemented in order to safeguard the huge data. Some of the techniques such as attribute based encryption to protect sensitive data are used.

Taking the idea a step further, the challenge of detecting and preventing advanced persistent threats may be answered by using Big Data style analysis. These techniques could play a key role in helping to detect threats at an early stage, using more sophisticated pattern analysis, and combining and analyzing multiple data sources. There is also the potential for anomaly identification using feature extraction. Today logs are often ignored unless an incident occurs. Big Data provides the opportunity to consolidate and analyze logs automatically from multiple sources rather than in isolation. This could provide insight to individual logs that can be used to potentially enhance intrusion detection system (IDS) and intrusion prevention system adjustments and effectively help in learning “good” and “bad” behaviors.

V. Big Data Security Issues:

A lot of security threats occur in Big Data concept also. It is so much related to large storage of incoming data that more or less all the security threats of Cloud Computing are related to Big Data as well. Big Data is collected from Cloud, real devices, social media, and virtual devices for various analytical purpose. All the security and the privacy issues are growing with the growth of Big Data. The collection and aggregation of huge volumes of data in terms of Big Data is a source of rich

ISSN: 2521-3288

information to various analysts from scientists, businesses, Government agencies, local citizens to clinicians. Yet the tools and technologies that are needed to develop and manage these voluminous data are often not adequate to support security or privacy. Large scale data security is a threat until now. Adequate policies are necessary to secure the data from all points of view. Furthermore, existing technological approaches to security and privacy are increasingly being breached, whether accidentally or intentionally, thus necessitating the continual reassessment and updating of current approaches to prevent data leakage.

VI. Big Data Analytics:

Big data analytics is the use of advanced analytic techniques against very large, diverse data sets that include different types such as structured/unstructured and streaming/batch, and different sizes from terabytes to zettabytes. Using advanced analytics techniques such as text analytics, machine learning, predictive analytics, data mining, statistics, and natural language processing, businesses can analyze previously untapped data sources independent or together with their existing enterprise data to gain new insights, resulting in significantly better and faster decisions. (Kallahalla, M., et. all, 2003, p.29-42)

Big Data Analytics Tools:

Apache Hadoop (http://azure.microsoft.com/), a nine-year-old open-source data processing platform first used by Internet giants including Yahoo and Facebook, leads the big-data revolution. (Juels, A., Pors Jr, B. S. K. 2007, p. 584 - 497) Cloudera introduced commercial support for enterprises in 2008, and MapReduce (Budgen, D., Turner, M., Brereton, P., Kitchenham, 2008 p 195 - 204) (Petersen, K., Feldt, R., Mujtaba, S., Mattsson, M., 2008, p 7 - 10) and Hortonworks piled on in 2009 and 2011, respectively. Among data-management incumbents, IBM and EMC-spinout Pivotal each has introduced its own Hadoop distribution. Microsoft and Teradata offer complementary software and first-line support for Hortonworks' platform. Oracle resells and supports Cloudera, while HP, SAP, and others act more like Switzerland, working with multiple Hadoop software providers.

ISSN: 2521-3288

SAP has been the biggest champion of the in-memory approach with its Hana platform, but Microsoft and Oracle are now poised to introduce in-memory options for their flagship databases. Focused analytical database vendors including Actian, HP Vertica, and Teradata (Vadalasetty, Sreenivasa Rao., 2003, p.10-12) have introduced options for high-RAM-to-disk ratios, along with tools to place specific data into memory for ultra-fast analysis.

Where volume and variety are extreme, Hadoop (https://hadoop.apache.org) has proven its utility and cost advantages. Cloudera, Hortonworks, and MapReduce are doing everything they can to move Hadoop beyond high-scale storage and MapReduce processing into the world of analytics.
The new vendors like Actian, InfiniDB/Calpont, HP Vertica, Infobright, and Kognitio, all of which have centered their big-data stories around database management systems have also focused entirely on analytics rather than transaction processing.

Hadoop MapReduce is a software framework for easily writing applications which process vast amount of data (multiterabyte data-sets) in-parallel on large clusters (thousands of nodes) of commodity hardware in a reliable, fault-tolerant manner. (https://hadoop.apache.org)

A Map Reduce job usually splits the input data-set into independent chunks which are processed by the map tasks in a completely parallel manner. The framework sorts the outputs of the maps, which are then input to the reduce tasks. Typically both the input and the output of the job are stored in a file-system. The framework takes care of scheduling tasks, monitoring them and re-executes the failed tasks. Typically the compute nodes and the storage nodes are the same, that is, the MapReduce framework and the Hadoop Distributed File System are running on the same set of nodes. This configuration allows the framework to effectively schedule tasks on the nodes where data is already present, resulting in very high aggregate bandwidth across the cluster.

The MapReduce framework (Feng, J., Chen, Y., Liu, P., 2010, p. 1-2) consists of a single master Job tracker and one slave TaskTracker per cluster node. The master is responsible for scheduling the jobs' component tasks on the slaves, monitoring them and re-executing the failed tasks. The slaves execute the tasks as directed by the master.

ISSN: 2521-3288

Minimally, applications specify the input/output locations and supply map and reduce functions via implementations of appropriate interfaces and/or abstract-classes. These, and other job parameters, comprise the job configuration. The Hadoop job client then submits the job (jar/executable etc.) and configuration to the Job Tracker which then assumes the responsibility of distributing the software/configuration to the slaves, scheduling tasks and monitoring them, providing status and diagnostic information to the job-client. Although the Hadoop framework is implemented in JavaTM, MapReduce applications need not be written in Java.

VII. Impact of Cloud and Big Data Analytics:

The combined effective power of cloud computing and big data analytics can be an extremely powerful force for all business organisations of all sizes and capacities. By applying these resources, business organisations can position themselves to leap to the forefront of their competitors.

Yet in order to achieve this goal a lot of input is needed. In future Cloud based big data tools may not guarantee success. Thus to effectively control the resources, business organisations need to invest on critical supplementary solutions.

Business organisations must have a place in house for moving data in and out from cloud environment for safer and better analytics. Some tools are not designed to adapt to this need, thus firms relying on such resources will experience serious bottlenecks, inadequacies in the firm and delays bearing heavy losses.

Fortunately, tools exist which can help business organisations overcome these issues. One of the most useful of these is Attunity CloudBeam. With CloudBeam, firms can easily and quickly transfer data to the cloud environments for data analytics, storage or any other requirements. This allows business organisations make the most of both their cloud and big data resources, and provides a major boost to their bottom lines.

ISSN: 2521-3288

VIII. Conclusion:

The usage of cloud environment with big data analytics proves to be an effective way to handle the growth of business organisations. A lot of tools are available for data and service storage in the cloud and further analytics of this data is handled by big data analytics tools. Thus the potential of Big Data in Cloud computing models is tremendous for have analytical solutions. Cloud provides a lot of flexibility and efficiency in data access however Big Data Analytics based on cloud is not a one size all solution. Business organisations use cloud infrastructure to provide AaaS so that a multiple service facility can be provided on the cloud. Thus the bottom line is that whatever maybe the cloud services business organisations are solely dependent on the cloud environment and the Big Data tools for proper analytics. A comprehensive cloud based big data strategy can be framed and optimized for the enterprise data.

References:

1. Rittinghouse, John W., Ransome James F., (2010) *Cloud Computing: Implementation, Management, and Security*, CRC Press, Taylor & Francis Group, Boca Raton, FL p 6-7.
2. Soans, Sonia V. (2012) *Security Issues of Infrastructure as a Service [IAAS] in Cloud Computing*, NCT 3rd Symposium p 1-5.
3. Vadalasetty, Sreenivasa Rao., (2003) *Security Concerns in Using Open Source Software for Enterprise Requirements*, SANS Institute 2009 p 10-12 .
4. Krutz, Ronald L., Vines, Russell Dean (2010) *Cloud Security: A Comprehensive Guide to Secure Cloud Computing*, ISBN: 978-0-470-58987-8 p 8-10 .
5. Cloud Security Alliance, *Top Threats to Cloud Computing V1.0*, (2010), https://cloudsecurityalliance.org/topthreats/csathreats.v1.0.
6. Rani, Dimpi, Ranjan, Rajiv Kumar, (2014) *A Comparative Study of SAAS, PAAS and IAAS in Cloud Computing*, Volume 4, Issue 6, June 2014 ISSN: 2277 128X p 2-4.
7. Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R. H., Konwinski, A., Lee, G., Patterson, D. A., Rabkin, A., Stoica, I., Zaharia, M.,(2009) *"Above the Clouds: A Berkeley View of Cloud"*, Electrical Engineering and Computer Sciences, University of California, Berkeley, February 10, 2009 p 1-4.
8. Feng, J., Chen, Y., Liu, P., (2010) *Bridging the Missing Link of Cloud Data Storage Security in AWS*, the 7th IEEE Consumer Communications and Networking Conference-Security for CE Communications (CCNC 10), Las Vegas, Nevada, USA, January 9 - 12, 2010 p 1-2.
9. Kallahalla, M., Riedel, E., Swaminathan, R., Wang, Q., Fu, K. (2003) *Plutus: Scalable Secure File Sharing on Untrusted Storage*, USENIX Conference on File and Storage Technologies (FAST), pages 29-42 .
10. http://azure.microsoft.com/
11. Juels, A., Pors Jr, B. S. K. (2007) *Proofs of retrievability for large files*, Proc. ACM CCS, pp. 584–597, 2007.
12. Budgen, D., Turner, M., Brereton, P., Kitchenham, B., (2008) *Using Mapping Studies in Software Engineering*, Proceedings of PPIG 2008, Lancaster University, pp.195-204.
13. Petersen, K., Feldt, R., Mujtaba, S., Mattsson, M., (2008) *Systematic Mapping Studies in Software Engineering*, 12th International Conference on Evaluation and Assessment in Software Engineering, Australia p 7-10.
14. https://hadoop.apache.org.
15. Roy, I., Setty, S. T. V., Kilzer, A., Shmatikov, V., Witchel, E. (2010) *Airavat: security and privacy for Map Reduce*" USENIX conference on Networked systems design and implementation, pp 20-20.
16. http://www.oracle.com/technetwork/database/options/advanced-analytics/bigdataanalyticswpoaa-1930891.pdf

 ISSN: 2521-3288