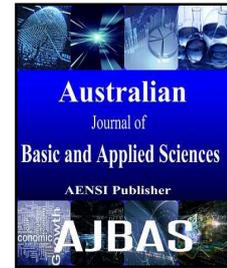




AUSTRALIAN JOURNAL OF BASIC AND APPLIED SCIENCES

ISSN:1991-8178 EISSN: 2309-8414
Journal home page: www.ajbasweb.com



The Convergence of Big Data and Cloud Computing: A Study from Research Perspective

¹Arunkumar Panneerselvam and ²Bhuvaneshwari Subbaraman

¹Pondicherry University, Karaikal Campus, Department of Computer Science, Engineering and Technology Faculty, Pondicherry, India.

²Central University of Tamilnadu, Thiruvavur, Tamil nadu, India.

Address For Correspondence:

Nancy H.J. Mandey, Department of Business Administration, Politeknik Negeri Manado, Jalan Kampus, Desa Buha. Manado. North Sulawesi. Indonesia. Post Code 95252

ARTICLE INFO

Article history:

Received 26 April 2016

Accepted 21 July 2016

Published 30 July 2016

Keywords:

Cloud, Big data, Analytics as a service, Hadoop, Migration, Parallel Processing

ABSTRACT

The convergence of cloud and big data is inevitable in the business and industry in the forthcoming years. The cloud is a more suitable platform to store, process and analyze big data since the cloud is providing extensive on-demand resources including high computing power and storage which is an essential driving factor for the big data processing. Also the cloud is capable of providing Analytics as a Service to a wide range of business. This paper explores the Challenges in storing, processing and analyzing big data in the cloud environment.

INTRODUCTION

This paper depicts the relationship between cloud and big data followed by the challenges in storing the big data in cloud environment. The next discussion centre around the research issues in processing the big data in cloud. The final discussion throws spot light on Analytics as a service and the problems related to it. The power of Hadoop in processing big data in cloud also explored.

Cloud Computing And Big Data:

Although number of definitions explains about the cloud the most appropriate and exact definition of cloud with its characteristics is given by National Institute of Standards and Technology's (NIST). According to NIST (2011), "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction".

Big data simply refers to the data which requires more storage and processing power. The Gartner definition says "Big data high volume, high velocity and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization".

Relationship between Cloud and Big data:

The IDC predicted that by the year 2020, 40% of the world data would be in cloud. The cloud players like Amazon provider of AWS, Microsoft the provider of Azure and IBM deployed their big data in cloud using Hadoop clusters. Ibrahim Abaker *et al.* (2015) describes the relation of big data with cloud as in Fig.1

Open Access Journal

Published BY AENSI Publication

© 2016 AENSI Publisher All rights reserved

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

To Cite This Article: Arunkumar Panneerselvam and Bhuvaneshwari Subbaraman., The Convergence of Big Data and Cloud Computing: A Study from Research Perspective. *Aust. J. Basic & Appl. Sci.*, 10(12): 98-101, 2016

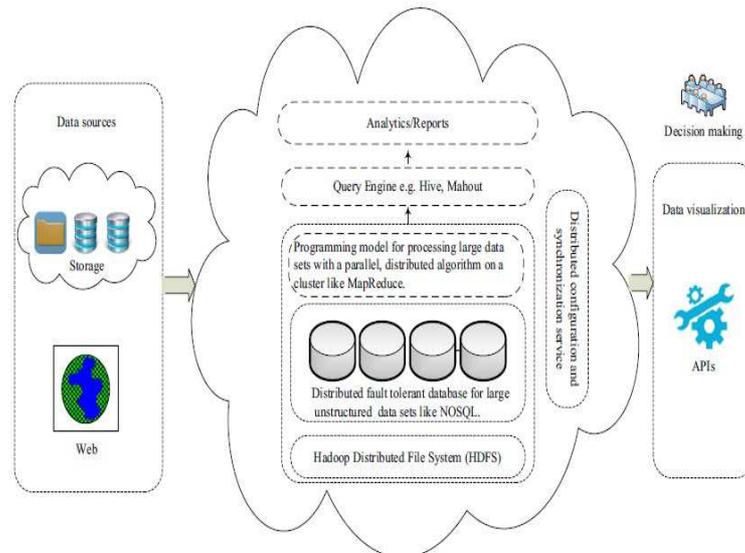


Fig. 1: Relation of Big Data with Cloud

Big data Analytics requires more storage and processing power and ideally the cloud provides the platform for it. Big data uses large datasets to perform the analytics and return useful information. Cloud allows distributed and parallel processing of queries in their large datasets through Hadoop clusters. Hadoop tool allow the query program to move into datasets in clusters rather than moving the dataset for analysing. The program can be processed in distributed and parallel manner and the returning results are map reduced.

Challenges In Storing Big Data In Cloud:

Though cloud provides extensive and scalable storage for big data there are several challenges and issues needs to be addressed for efficient storage of big data in cloud environment. The following are the key issues:

Data Security and Privacy:

The data are in third party cloud servers which are out of the owner control. Data privacy is affected and the user has less control. Entire datasets depends on the security policies and governance of the cloud owner. Venkata Narashima Inukollu *et al.* (2014) reported that issues in cloud can be at Network level, Data level and Authentication level.

Data Migration:

Sophisticated Data Migration techniques and algorithms are yet to mature. The Organization needs to know which data are to be migrated for deriving data patterns useful to business decision. If there is no clear idea then the cloud servers hired will be populated with uninteresting datasets.

Cost:

The cost of moving data to and fro from the big data systems to cloud involves network utilization, network time and other related resources. Moving high beneficial dataset with less cost is a challenging issue.

Research issues in processing big data in cloud:

The Processing of Big data through cloud brings in several advantages like Distributed parallel processing, High computing power etc. However the following challenges are to be addressed:

Dynamic Processing:

The data stored in cloud are scalable and dynamic. Distributed Parallel Processing mechanisms on fault tolerant datasets in multi cloud environment are to be improved for adapting elasticity nature of cloud and for better cost effective processing. Algorithms that support dynamic processing of data are to be developed.

Programming Models and software Architecture:

To process big data on cloud the traditional programming models will not able to cater the needs. New programming models with optimized performance are vital importance for big data processing on cloud. Sophisticated Software Architecture for processing Big data are in early stages.

Stephen Kaisler and Frank Armour (2013) says that design for the big data system and components need well understanding of the user and tools used for processing. Not all the big data users have the same requirement. User interfaces should designed considering this dynamic feature of big data processing.

Performance issues:

When number of parallel processing increases the resultant data returned by query program operating on the datasets also increases utilizing the network bandwidth, speed and other related resources. The performance of the system should be optimized for low cost. The time taken for processing the data should be minimized allowing increased number of parallel processing in the datasets. Latency and Bandwidth are important considerations.

Adapting Analytics As A Service:

Cloud basically provides Software as a Service (SaaS), Infrastructure as a Service (IaaS), and Platform as a Service (PaaS). As per the White paper published by the AtoS inc. on Data Analytics (2013), Analytics as a Service (AaaS) is an extensible analytical platform provided using a cloud-based delivery model, where various tools for data analytics are available and can be configured by the user to efficiently process and analyse huge quantities of heterogeneous data.

Analytical Life Cycle Management:

The Big data analytics includes a series of process like Data acquisition, Data Modelling, Data Mining and Data Visualization. Managing these aspects of Analytics in cloud environment, which is dynamic in nature need flexible methods because the traditional Analytics techniques are fixed and confined to a particular environment.

Diversity in Data model:

There are variety of data model exists for business with different needs. Each data model is tied strongly with a particular type of analytics. Availability and choosing the right analytical and data model from the cloud is a challenging task.

Need of Scalable Analytical Approaches:

Stephen Kaisler and Frank Armour (2013) reported that every Algorithm has a knee point at which the algorithm performance ceases to increase because most of the analytical algorithms are designed for specific implementation on specific machines on specific scale. The computational resources provided by the cloud are highly elastic in nature. Development of new analytical algorithms which are scalable in nature is very important to business for adapting analytics as a service.

Lack of required Skills:

Big data Analytics requires extensive knowledge about the data and the ability to classify which data are useful for the business decision. Data scientists with insight about analytics and the analytics tools are very less in the market which will make the adoption of big data process.

According to Evans Data Corporation, Big Data and Advanced Analytics Survey 2015, the following figure Fig. 2 shows the problem related to big data working.

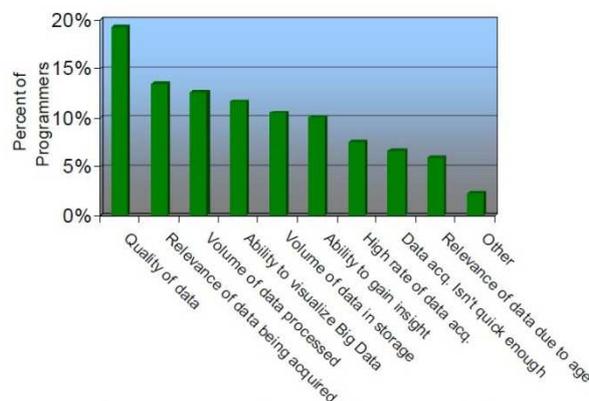


Fig. 2: Problems with the Programmers of Big data Analytics

Hadoop and the cloud:

There are number of tools available to work with big data in cloud. However the Apache Hadoop is widely popular because of the following features:

- Scalable – Hadoop Clusters are used to distribute and store very large datasets in cloud servers which operates in parallel.
- Less cost – Using Hadoop and doing analytics in cloud is cost effective since it is open source and the cloud providers charge less for pay per use.
- More Flexible – Hadoop can process both structured and unstructured data and generate value of the data.
- Processing Speed – The Hadoop architecture supports moving program to work on the data sets rather than moving data. Hence the speed of data processing for analytics is significantly fast.
- Fault tolerance – one of the key advantages of using Hadoop in cloud is its fault tolerance nature. When a data is stored in a node in the cluster a copy of the same data is stored in another node in the cluster. In case of failure on one node data for analytics is available in another node. Using map reduce distributed processing of data are done efficiently.

According to Harvard Business Review Analytics Services Survey, July 2015, the following figure Fig. 3 shows the usefulness of the analytical tools in general.

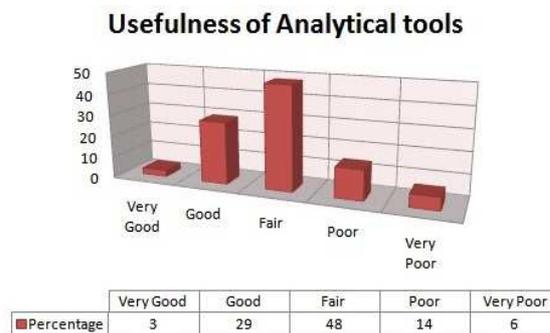


Fig. 3: Analytical tool usage survey

The survey throws a positive note on the usage of analytical tools in business.

Conclusions:

The union of the cloud and big data is at the earlier stage. It needs standardization, Reliable software architectures and right tools. The integration of big data and cloud will change the way of the current organizations business decisions and fetch more business advantages. The issues in terms of storage, processing and analyzing big data in cloud environment should be address with applicable solutions for better long term benefits.

REFERENCES

- AtoS Inc., 2013. Data Analytics as a Service: unleashing the power of Cloud and Big Data, White paper published.
- Ibrahim Abaker, Targio Hashem, Ibrar Yaqoob. 2015. The rise of big data on cloud computing: Review and open research issue, Information Systems, 47: 98-115, Elsevier Publications.
- Laney, Douglas, 2012. The Importance of Big Data: A definition, Gartner. Retrieved.
- NIST special publication, 2011. The NIST definition of Cloud Computing, 800-145.
- Stephen Kaisler, Frank Armour, 2013. Big Data: Issues and Challenges Moving Forward, 46th Hawaii International Conference on System Sciences.
- Venkata Narashima Inukollu, Sailaja Arsi, Srinivasa Rao Ravuri, 2014. Security Issues associated with Big data in Cloud Computing, International Journal of Network Security and its Applications (IJNSA), 6: 3.