Asian Journal of Information Technology 15 (14): 2516-2525, 2016
ISSN: 1682-3915

A Review on Relationship and Challenges of Cloud Computing And Big Data: Methods of Analysis and Data Transfer

[1]Parvin Ahmadi Doval Amiri and [2]Mina Rahbari Gavgani
[1]Department of Engineering, Islamic
Azad University, Isfahan (Khorasgan) Branch, Isfahan, Iran
[2]Sama Technical and Vocational Training College, Islamic
Azad University, Isfahan (Khorasgan) Branch, Isfahan, Iran

Abstract: Continuous increase in volume and details of the data recorded by organizations, after emerging the social media, the Internet of Things (IoT) and public broadcasting has led to the impending explosion of data flow, either in structured or unstructured format. The data that we read it as big data have been created with unprecedented speed and its processing and analysis is a challenging task that requires powerful computing infrastructure. Meanwhile, cloud computing is a powerful technology that has been provided for massive and complex calculations on a large scale. Due to the large volumes of research resources such as study and books that are related to these two concepts, i.e., cloud computing and Big Data, we decided to provide a comprehensive and integrated study which aims at investigating the relationship between these two concepts from the conceptual and engineering perspectives. In this study, the concepts and the relationship between big data and cloud computing will be discussed. Moreover, challenges, opportunities and different methods of analyzing and transferring big data to the cloud will be expressed.

Key words: Big data, cloud computing, big data analysis, analysis, challenges

INTRODUCTION

Big data have attracted public attention, including universities, government and industry. Big data with three main characteristics are distinct from normal data: Data are countless. The data cannot be categorized in regular relational databases. Data can be quickly created, recorded and processed. Big Data are reshaping to health care, science, engineering, finance, business and the community. Now the rate of data creation is confusing. The main challenge of researchers and practitioners in the field is that, this growth rates over lapped its ability to design appropriate platforms for cloud computing to analyze data and to update the intensive workload (Zhang *et al*., 2013).

Definition and characteristics of big data: Big data, a term that refers to the increasing volume of data that storage, processing and analysis of by using traditional database technologies is hard. Big data have none unique name. The term "Big Data" is relatively new term in the business and IT. But researchers and workers in this area used this term in their studies before. For example, we can refer to big data as large amount of scientific data to make them more visible. There are various definitions of big data. For example big data were defined as: "The amount of data that is more than storage, management and efficient processing technology capacity. The definition of big data is subject to three features at the same time (That all of them start with letter V and they call it V3): Volume, variety and velocity. The term volume, variety and velocity were first introduced by Gartner. He used this term to describe the challenges of big data elements. IDC research firm defined big data technologies as this: "A new generation of technologies and architectures that are designed that by enabling high-speed registration, discovery and analysis to extract the amounts to more affordable way of large volumes of a wide range of data". Big data features not only are not limited to V3 but also can be expanded to V4 that their name is: Volume, variety, velocity and value. This V4definition is known globally because shows meaning and importance of big data very well. The following definition was suggested based on the definitions listed above and our observation and analysis of big data. Big data is a collection of techniques and technologies that require new forms of integration to extract hidden large amount from several big datasets, complex and large-scale. (Purcell, 2014) (Manyika *et al*., 2011; Zikopoulos *et al*., 2012.

Corresponding Author: Parvin Ahmadi Doval Amiri, Department of Engineering, Islamic Azad University, Isfahan (Khorasgan) Branch, Isfahan, Iran

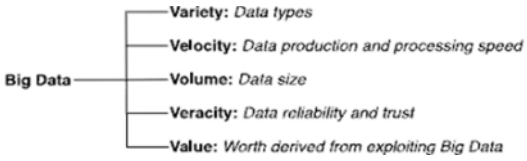

Fig. 1: Some of the V of big data

Volume: Volume relates to the production of all kinds of data from different sources that are still widespread. The advantage to collect massive amounts of data includes creation of information and hidden patterns by analyzing the data. Laurila presented a unique collection of longitudinal data from smart mobile devices and made it available for research community use (Hashe *et al*., 2015)

Variety: Variety is related to different types of data collected by sensors, smart phones and social networking sites. This kind of data, including video, image, text, audio and other data types are found whether in the form of structured or non-structured form. The majority of data created by applications of mobile phones, the majority of data created by applications of mobile phones are in structured format. For example, text messages, online games, blogs and social media, create various types of non-structured data, through devices and sensors. Internet users also produce highly diverse set of structured and unstructured data (Berman, 2013; Gantz Reinsel, 2011) (Fig. 1).

Velocity: Velocity refers to the speed of data transmission. Content of data have already been archived due to attraction of additional data collection. Rule sets and data that arrive from various sources (Cox and Ellsworth, 1997).

Amount is the most important aspect of big data that is related to process of massive amounts of secret discovery of massive datasets with different types and fast generation (Manyika *et al*., 2011).

Also 3 other V was also presented as three other dimensions of big data. These dimensions include the following:

Validity: IBM made validity as 4th V that says Uncertainty in some sources of information is inseparable. For example, naturally, customer sentiment on social media is unreliable. However it has valuable information. Hence the need to address the misinformation and other big data is unreliable procedure that considered the use of analysis tools developed to manage and extract unreliable data. The variability and complexity of big data is defined as two extra dimensions. Variability refers to variety in the flow of information. Speed of big data is not

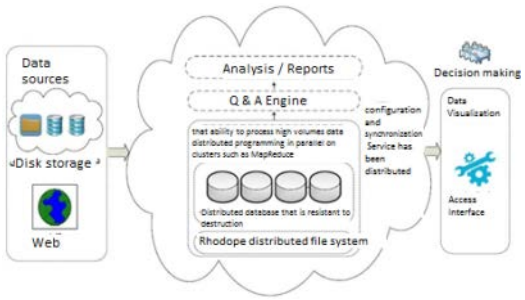

Fig. 2: Cloud computing on big data

stable and often has periodic up and down. Complexity refers to the fact that Big Data have been created through countless resources (Hashem *et al*., 2015), (Assuncao *et al*., 2015.

Value: Oracle enterprise defined value as defining characteristic of big data. According to the Oracle, big data has often been described as a low-density value and that's how, information is received in original form and usually depends on the size of its low value. However, high-value can be achieved by analyzing large volumes of such information. Therefore, there are no international standard for value, velocity and variety. Defining definitions depends on the limits to the size, sector and location on the economic unit and these limitations evolve over time. Also important is the fact that these dimensions are not independent of each other. So that if a later change increases the likelihood that change will result in another dimension. Therefore, the value of the expected future economic unit should be in front of big data technology implementation costs weigh (Leary, 2013) (Gandom and Haider, 2015).

The relationship between cloud computing and big data: Cloud computing by adding flexibility to the IT industry has created a revolution that enables organizations to use them only for the resources and services that are paid (Fig. 2).

Cloud computing is one of the most notable changes in modern ICT and services for enterprise applications and becomes powerful architecture at the macro level to run complex calculations. Clouds are very diverse in performances and specific technologies. But most of the infrastructure and resources provide software as service. The benefits of cloud computing virtualized resources, parallel processing, security and integrity of data services with data storage is scalable. Not only can the cost and

limitations of cloud computing automation and computerization by individuals and companies to a minimum but also it can infrastructure maintenance costs, provide efficient management and user access. As a result of the benefits mentioned, a number of programs that multiple cloud platforms enable the creation and led to a significant increase in the scale data created and is consumed by such applications. Some of the first people that have used cloud computing of big data, there are customers that had developed Hadoop clusters in highly scalable and flexible computing environments by vendors such as IBM and Microsoft Azure and Amazon AWS (Chen *et al*., 2014).

Cloud computing and big data link together. Big data provide the ability to use the appropriate calculations to process queries across disparate data sets and multiple result sets returned, timed manner for its users. Cloud computing, motor guaranteed through the use of Hadoop, provides a class of distributed data processing platforms. The use of cloud computing on big data presented in Fig. 2. Massive cloud and Web data sources in a distributed database of reserve depletion has been resistant and through Map reduce programming model for big data sets with a parallel distributed algorithm is processed in a cluster (Chen *et al*., 2012). Big data uses dispersed storage technology that instead of local storage attached to the computer or electronically, based on cloud computing. Assessment of big data, cloud-based applications developed using virtual technology has been carried out. Therefore, cloud computing not only provides facilities for computing and big data processing but also acts as a service model.

The main reasons for the small to medium sized businesses using cloud computing for big data cloud computing promises a big data reduction commitment of resources for small to medium businesses data. Processing of big data through a programming pattern is known as MapReduce. Typically mapping algorithm reduces the need for network attached storage and parallel processing.

These mapping calculations-reduction programming often requires more than just something small to medium sized businesses are able to do so.

The main reason and advantages of using cloud computing for small to medium businesses reduce hardware costs and reduce processing costs and the ability to test the value of the data in big data. A major concern about cloud computing is security and loss of control. There are several models of cloud computing services for businesses save money by taking the balance between security concerns and the loss of control access (Purcell, 2014).

Migration to cloud applications and big data transfers: Cloud computing provides quick access to scalable resources which is especially important for big data processing in cloud computing is an open issue. How to move data from different geographic locations over time is for efficient processing of data. The transition from hard disk is not reliable and flexible approach because it requires time study, the cost for a massive load and dynamically generates data regarding the geographic distribution of data in the cloud. As a result, we have challenge of expensive data migration. Two online algorithms was proposed for taking the time to choose the optimized data centers for collecting and processing and the Online Migration of data transmission with Lazy (OLM) and algorithms to achieve a competitive ratio 1+1 (RFHC) (Zhang *et al*., 2013).

In recent years we have seen significant interest in the various applications on the cloud platform migration. Hajjat *et al*. (2010) did development and optimization model to migrate its applications to a hybrid cloud. Wu *et al*. (2012) are defenders of deploy applications to the cloud, social media, rich in natural resources and are pay-as-you-go pricing. This project is immigration focus on workflow optimization and application performance carefully modules decide to move to the cloud storage and data replication strategy in the cloud. A little tasks in relation to the transfer of huge amounts of data to the cloud has been done, including Cho and Gupta (2010) by Pandva plan created a system aware of cost to transfer data that with design of Hajj *et al*. (2010) of these direction of that plan on a static scenario the value is fixed several similar but the data center is intended (Hajjat *et al*., 2010; Miller, 2013; Bollier and Firstone, 2010).

MATERIALS AND METHODS

Methods in relation to analysis of data transmission on the cloud for big data: Analysis of data on the cloud for big data, there are many different ways that went on to mention these methods have been considered:

Traditional data analysis: Figure 3 shows the current phase of work flow analysis to data from various sources includes traditional databases and shows current and Martz and data warehouses are used to build these model. Large quantities and different types of data

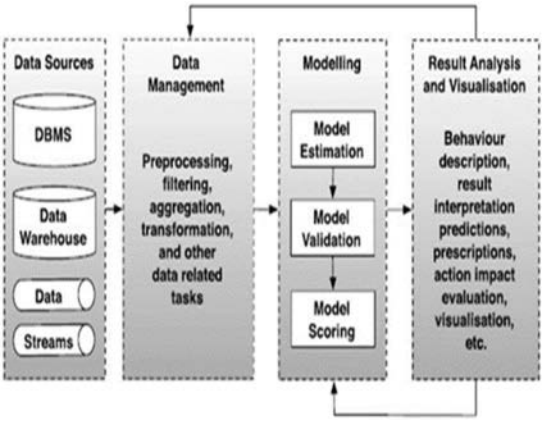

Fig. 3: Traditional analysis of current phases

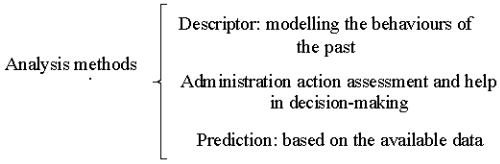

Fig. 4: Analysis method categories

pre-processing caused demand for data integration and cleansing and filtering them is required. Data ready for training model and its parameters are used. Once validated the model estimates should be treated before consumption. These phases normally require the use of specific methods for validation of the original input data and the model. Finally, the model was built as the data received and used. These phase that called simply the fruition to anticipate and production can result. The estimated results are interpreted and scaled to produce new or existing models or aggregated data pre-processing of used or pre-processed data are integrated (Ji *et al*., 2012; Talia, 2013).

Segmentation analysis solutions: Analytical solutions as shown in Fig. 4 can be predicted, divided descriptor and prescription. Descriptive analysis of historical data used for pattern recognition and preparation of management reports that mobile-model of the past. These analyses predict the future by analyzing current and historic data. Solutions predictability in decision-making by determining and evaluating the business requirements and constraints helps analysts. Despite the use of this approach, using this analysis still requires more effort. This is because the current solution suite based on the Software and public purposes. That's why this effort for special-purpose solution is to organize which includes the integration of different sources and deploy the software on the hardware these company or even household appliances integrate with the rest of the system should be performed by the company. These solutions are usually spread at the customer site and the time it needs to operate for several hours. Cloud brings a model to analyze where customers can pay to pay-as-you-go method.

This model has several techniques to become a reality must be addressed. That can be configured models and techniques such as data management and privacy and data management and data quality and currency named. This technique is outstanding and accomplished solutions for big data analysis capacity on the cloud. Checking of workflow traditional analysis is given in Fig. 3. We're focused on the phases of analysis solutions. Big data analytics cloud with most of the challenges are concerned about data management and integration and processing are obvious. Now we analyze models to calculate data on the cloud.

One of the most important and most time-consuming tasks of analyzing is preparing data for analysis. This problem is exacerbated when big data infrastructure along with all its limitations is larger. Performance analysis of large amounts of data requires effective

Analytical solutions based on type of cloud: Analysis of cloud solutions require multiple cloud is expanded model has been adopted by the company.

Private cloud: In a state where private cloud deployment will be managed by the organization or a third party. A private cloud is suitable for businesses that require the highest level of security and privacy control data. This type of cloud in a similar situation to share services and data that can be used across a large company-will be divided

Public cloud: These clouds are accessible to the public to be deployed over the Internet. Public cloud offers high efficiency and shares low-cost resources.

Analysis and data management services and quality assurance by producers such as the availability and security and privacy control is specified in the contract. An organization can share these clouds to reduce costs or reduce results (Thusoo *et al*., 2010).

Hybrid cloud: The cloud is a combination of both private and public cloud. Where additional resources of a public cloud can be transferred to a private cloud in their time of need customers can use applications with analytics applications using a private setting. So has the advantages of higher security of a public cloud is when used alone. Due to cloud deployment following scenario with regard to the availability of data and analysis will be presented prediction model (Ananthanarayanan *et al*., 2009).

Solutions to save and retrieve data amount by big data: Several solutions to save and retrieve large amounts of data have been proposed that some of them are commonly used in the cloud. Internet file-scale systems such as SEO are trying to robustness, scalability and reliability required specific Internet refer to the storage capacity is subject to mobile solutions where files can be replicated across multiple sites as well as data redundancy is avoided with these method. A sample storage service cloud storage on Amazon and Azure Microsoft Nirovanix and store large binary goodies that most cloud applications need them. One of the key aspects in developing great performance in applications is data analysis. This is because the amount of data involved in the analysis will be involved for data transfer and processing heavy. These options are also preferred in high-performance computing systems (Thusoo *et al*., 2010).

In systems that are often concerned about the performance of CPU intensive calculations practical work is transferred to the computing unit because the data transfer processing is small. Big data computing nodes in the data transfer rate of data transmission time devoted to the processing time so another method is preferred where calculations are moved to where the data is there. Similar way to discover the location that was used in the workflow and data network also applies here.

MapReduce analytics on big data provides an interesting model where local data to improve the performance of applications are searched. Hadoop MapReduce is an open source implementation of the main clusters allows of system to use Hadoop distributed file partitioning and replication of data for nodes Mapper are done by the consumer. The operation is carried out a large number of nodes as well as the impact HDFS data replication error for setting minimizes the number of nodes (Assuncao *et al*., 2015).

In order to process big data sets facebook platform is also developed. The report also stable platform in order to integrate the web server then sends them to HDFS files and of Hive-Hadoop cluster analysis in order to carry out tasks. The platform includes replication and compression and a compression technique to store large amounts of data is a column Hive. Including bugs and implements MapReduce cloud storage techniques is that they require the customer has to learn a new set of API to build cloud solution for analysis.

To minimize these barrier previous works POSIX system was developed to analyze the data. For example, in a study POSIX-based cluster file systems store data in the cloud Add as they used to. By using the concept of meta-blocks they recognized that IBM GPFS parallel system or public files can be read performance HDFS to the cause. A meta-block series of consecutive blocks the data that have been in the same disk location, therefore the proximity of support. Suggested way trade between the blocks of different sizes checks so that these blocks MapReduce applications to minimize overhead Small blocks reduce overhead prefetching and improve cash management for normal applications to (Tantisiriroj *et al*., 2011).

Parallel Virtual File System or HDFS compared with PVFS briefly and observed that PVFS significant improvements in completion time and cannot be compared to HDFS. Although, a large part of unstructured data is generated but make choices based on DBMS interface, customers and sales and its products are classified. When the data to be managed according to traditional DBMS for data retrieval and analysis of data to be transmitted to the storage location. Models like MapReduce generally not suitable for relational data analysis. Efforts are also on to create a hybrid solution for display and data processing is done in MapReduce to some commands and requirements of data processing carried out by the DBMS. SQL and MapReduce to analyze such support in the DBMS to integrate multiple data sources are located. Some providers of analysis and data mining solutions by models such as MapReduce cluster processing tasks have moved to places where data is stored. So trying to minimize surpluses and have been stored and processed. Data processing and analysis capabilities have to move forward EDW briefly that these practice is said to facilitate the re-use or multiple data sets-are deployed in data centers (Cohen *et al*., 2009; Assuncao *et al*., 2015).

According to some producers cloud EDW offered solutions that Peta bytes of a given scale or more was also supported. For example, Amazon Redshift columnar storage and data compression is recommended that write performance is searchable by specification series that includes parallel processing architecture using high-performance hardware and networks complete graph and map regional and local storage is done that reduces requirements of I/O by inquiry.

Amazon user and customer data pipelines will allow to analysis Data in various web services on Amazon Elastic MapReduce or EMR such as dynamic database that can be used in combination with a requirements. Another specific requirements increased use of databases in the cloud is NO SQL that is used to store and retrieve information. NOSQL a non-relational model for data storage has approved (Cohen *et al*., 2009).

Leavitt argues that non-relational models over 50 year that is hierarchical and object-oriented and database diagrams are available but just to get more attention such as model-based key storage and storage pedestals are based on the evidence presented.

Levitt favorite causes increased performance and capacity control unstructured data and is suitable for distributed environment. In a later study without NOSQL provided some facts about the database that is

Table 1: Analysis models for big data

Model Name	Performance	Cons and pros	References
sMap reduce	Search local data to improve the performance of applications	*APL* Yadbgyrd new set of customer needs and to provide a solution for the analysis of claude bsazdv not suitable relational data	Thusoo *et al.* (2010)
POSIX	Meta-blocks consecutive blocks of data that have been in the same disk location	Minimize application overhead and improve cache Mapreduce	Ananthanarayanan *et al.* (2009)
Parallel Virtual File System) (PVFS	*SQL* and *Mapreduce* on top *DBMS* to integrate multiple data sources	Significant improvements in completion time and not be compared with *HDFS*	Tantisiriroj *et al* (2012)
EDW	Data processing and analysis capabilities have that this to move forward practice is said *EDW* briefly to facilitate the re-use or multiple data sets are deployed in the data center	Data processing capacity and better analysis	Cohen *et al.* (2009)
Amazon redshift	Columnar storage and data compression	Peta-scale data bytes or more will also be supported.I/O by query easing	-
NoSQL databases	Non-relational model for data storage	Better performance and capacity control of unstructured data and is suitable for distributed environment	Han *et al.* (2011)

emphasized on the advantages and limitations of cloud. NOSQL systems are classified into three parts according to its capacity, consistency, availability and partition split. It also has a model that is NOSQL support systems (Hashem *et al*., 2015). The researchers of the study, according to data models support different compared NOSQL systems as well as a variety of support orders and support for concurrency and stability compared Repeat and came to the conclusion that there is a big difference between different technologies and no system that is perfect for all of them. It is therefore important that merchants understand the requirements of applications and systems based upon the capacity and the most suitable by taking into account differences in these choices (Han *et al*., 2011) (Table 1).

RESULTS AND DISCUSSION

Existing models to calculate the data model on the cloud: Here discussed modeling for cloud providers a way to parallelize algorithms machine learning will be discussed. Guazzelli et al used amazon EC2 as a platform for hosting ADAPA. Predictive models expressed in the language or PMML model forecasts have been deployed in the cloud and are exposed to service location relations. Users can achieve search the web through modeling techniques and data mining solutions. By using the PMML is also available as a language for exchanging information on the models was predicted.

Zementis provide technical analysis and modeling that are implemented at the customer site or as a service provided Amazon EC2 and IBM smart cloud. Google is forecast API allows users to degree that machine learning to build models and use them to predict the value of a new element based on valid data handled previously used or a new classification that describes if an element is present. Prediction API allows users to education provided as separate files with comma and models provided under the special convention share their models or allow others to use of shared models by anticipating the Google API, users can use the software to perform logical tasks such as traditional analysis and forecasting of purchase and apply detected spam.

Mahut Apache project aims to provide scalable learning tool to build a library on top of Hadoop by using MapReduce paradigm. The proposed library can be deployed on a cloud and solutions required for the construction and extract clustering description and classification of documents to be used. By trying to simplify the complexity of systems such as Apple's Siri and Google's Knowledge Graph Hazy plan focused on diagnosis and approved two sets of abstracts in the construction of educational systems and infrastructure programming that is abstract. The plan is argued that by providing a summary of installation solutions and training makes system development easier. To achieve a small programming interface and data model uses a combination of Hazy. Communication and language model based on the combination possibilities. Hazy infrastructure for abstracts can be seen that more analytical algorithms defined by the user in relational database management systems are integrated. Then, the characteristics of the underlying infrastructure it uses to improve performance (Assuncao *et al*., 2015).

Challenges of research on big data in the cloud: Although, cloud computing is widely accepted by many organizations but Research on big data in the cloud remains in its early stages. Organizations using applications and social media, leading to high data growth rates and as a result have been the emergence of new challenges. Therefore, in the case of some of the key

Table 2: Characteristics of scalable data source in cloud environment

Characteristic	Advantage	Flaw
DBMS	Faster data access faster processing	Less appealing for the development of large-scale datait's limited
Key amount	Scale with very large size unlimited	
Google file system	Scalable distributed file system for large distributed data applications. Very big data files stored on a server	Collect garbage can become a problemIf the number of writers and researchers be more there is likely to degrade performance
Distributed File System Hadoop (HDFS)	Stores Large amounts of data uses big data cluster	

research challenges such as scalable, high availability, data integrity, data conversion, data quality, heterogeneous data, privacy and surveillance data in the cloud are discussed in this study (Chen *et al*., 2012).

The scalability: Scalability means the ability to control increasing amounts of data, in the proper manner. The scalable distributed data storage systems, cloud computing has become a vital part of infrastructure. The lack of charm RDBMS characteristics of cloud computing to support them to develop large-scale applications in the cloud is less. These defects are caused by the popularity of NoSQL.

A mechanism NoSQL database to store and retrieve large volumes of distributed data will be presented. Various types of database NoSQL such as Amount keyboard, leaning columns and bowed to the document, of Big Data support. The following Table shows characteristics of scalable data source is displayed in a cloud environment (Hashem *et al*., 2015) (Table 2).

Accessibility: Accessibility refers to the available resources permitted by the system. In a cloud environment, one of the main issues in relation to cloud service providers, Accessibility data stored in the cloud. For example, one of the pressing needs for cloud service providers effectively in order to meet the needs of mobile users that the one or more data are needed in a short time. In addition, a growing number of cloud users, cloud service providers should be subject to the availability of data needed to provide high quality services to users. Lee rain clouds called cloud model to support the using of Big Data have been introduced. Rain clouds includes partnerships between individual clouds is to provide available resources quickly. Schroeck and colleagues predicted that the need for better and faster access to data can continue such as growing or evolving business models and organizations on technologies required for faster data access and smart phones investment (Table 3).

Data integrity: Data security is a key aspect of integration. Data integrity means that an only person authorized by the person manipulating data to potential misuse of data is prevented. The proliferation of cloud-based applications will allow users to take their data in the cloud to store and manage data centers. Such programs should ensure data integrity. However, one of the main challenges is to ensure the integrity of user data in the cloud is to be directed and managed. Given that users may physically be able to access data, cloud should provide a mechanism for the user to ensure data retention (Talia, 2013).

Conversions (change): We can convert the data into a form suitable for analysis allows Denial of Big Data. With the variety of data formats, to analyze Big Data can be converted to two way shown in Fig. 5. In the case of structured data, data pre-processing before being stored in local databases have been set up restrictions were lifted. The data can then be retrieved for analysis. However, unstructured data, data must be distributed databases such as HBase stored and this must be done before processing them for analysis. Structured data of databases distributed after the resolution of Reading Recovery have limitations (Ji *et al*., 2012).

Data quality: In the past, data processing usually specified in the data collection sources were identified and restricted. Thus, the results were accurate. But with the advent of massive data, the data are derived of many different sources; all these sources are not known or verifiable.

Because the data is often collected from various sources of poor data quality has become a serious matter for cloud service providers. Voted instance, large amounts of data from smart phones have been produced in which conflicting data formats can be generated and these result is a heterogeneous sources. The issue of data quality is often a problem with one or more qualitative dimension has encountered. Therefore providing quality data of extensive collections of data sources is a challenge. Data quality and data consistency are described in the cloud. If the data obtained from new sources compatible with data from other sources, the raw data are of high quality.

Heterogeneity: Diversity is one of the main aspects of big data as a result of growth in resources and unlimited data. These growths would be great in homogeneities in nature.

Table. 3: Solutions for major challenges of big data in the cloud

Research challenge	Definition	Provided models and algorithms	Performance model alogrithm
Scalability	Remember ability to control increased amounts of data in a proper manner	Hadoop (HDFS) Hadoop Distributed file system	Great for large amounts of data by using clusters for data collection stores
Accessibility	Available system resources by authorized person in the cloud	Rain clouds, cloud model	Rain clouds includes partnerships between individual clouds to provide the resources are available quickly
Privacy policy	To prevent the spread of private data	Privacy algorithms within the framework layer map reduce	The retention of data pre-processing latency to ensure

Table. 4: Evaluation of privacy in the cloud

Suggested solution	Technique	Description	Limitation
Reconstruction algorithms for exploring data latency	Algorithms to maximize expectations	Maintaining latency measurement	Random performance making
Three-class data protection architecture	Portable data binding	Subject wearing caused by indexing data guides	Protection against malicious attempts (spying)
Privacy act (ppl) in the MapReduce frame work	Genetic algorithms to reduce cost of maintaining latency	Makes sure of retention latency data before further processing by the MapReduce tasks	Integration with data processing installations
Privacy leakage requirement based on the upper bound		Specifying which data collection interface code must be abandonment	The effectiveness of the proposed techniques

Data from different sources and in different types and forms of incompatibility and inconsistency displayed. In the development and application of computer technology based on internet users can take the form of structured data, store them partially structured or unstructured. Structured data for systems based on appropriate database while a somewhat structured data are only slightly. Unsuitable because of complex unstructured data in rows and columns that display it is difficult according to Kocarev and Jakimoski and challenges over how to handle different types of data sources there.

Privacy: The importance of privacy related to users that their private data on the Internet broadcast. These concerns with the expansion of extract Big Data analysis that requires personal information to produce relevant results are more serious. Keeping private information increases the importance of saving, theft and loss of control. Currently researchers use lot of hiding to take care of their private data.

Xu Yun have made the problem of maintaining the privacy of cloud computing data on average. They stated that the encryption of data in the cloud computing and cost-effective and not effective, because much time is required to encrypt or decrypt data. The researchers also studied cloaks to cut costs by researching on the part of the average data sets are displayed (Which sections should be and what not). Table 4 shows the techniques and their limitations.

Legal issues-settings: Specific rules and regulations should be established to prevent the spread of important information and personal users. Different countries have different rules and regulations for data privacy and support it. In several countries, a control relations

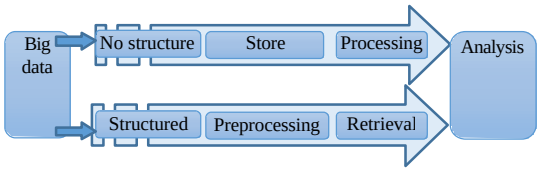

Fig. 5: Analyzing process of big data

employee is not allowed. However, the electronic control is permitted under certain conditions. So the question is that such laws and regulations, effective support for data subjects when using big data to these extent in society offers?

Supervision: Monitoring data includes exercises control and authority over legal rules related to data, transparency and the comptroller and auditor general of people and information systems to achieve business goals. The key issue relating to the use big data in the online space is the range of data streams from external sources. So a correct policy and acceptable data regarding the type of data that needs to be stored and the person with the required speed data access should be defined. Monitor big data involves the method of applying data set goals and tasks are manifold. The communications operator access to customer information such as details and search for information come in on the market to sell as money to. In addition, big data creating significant opportunities for service providers to provide more valuable information. Although the policies, principles and frameworks that move between risk and increased value to form to get better and faster data management technology can create greater challenges. So the acceptance of regulatory practices that provide balance between risk and increased value; could be a new

challenge for opening up competitive advantages and increase the value through the use of big data on the internet.

CONCLUSION

The production data is growing dynamically and with high volumes. The challenge arises as to what, how to manage and analyze large amounts of data which if not properly utilizes its existence, will become useless that only for the organization/business will be the cost of maintenance and storage. So if trade is able to search for information available in the data will be able to attract more customers and better business. In this study we have tried to analyze the nature of cloud computing and distributed processing power and storage technology, as it admitted the occasion to host introduce the concept of big data. Big data is a term that refers to the very high volume of data, the data from different sources such as computers, mobile devices, sensors and even household appliances are produced and various fields such as industry, trade, health, etc., in the covers. These data features, specifications and special requirements should be considered As big data. In this study we have tried to express the most important features and effort in order to manage and resolve each of these requirements also discussed. In discussing methods of analysis and data on the cloud, it is important that providers of cloud services and big data users (organizations) understand the requirements and with regard to capacity and differences to choose the most appropriate method. But in the research challenges related to big data cloud platform, it can be concluded, that integrity of the data that makes the data manipulated only by persons authorized products (To avoid possible misuse) a challenge and opportunity is open for research. Moreover, given that the data are produced of various sources, the resources are very heterogeneous and difficult to identify, obtain quality data is very valuable given that the data quality is poor and yet the challenge is also an important issue for research.

REFERENCES

Ananthanarayanan, R., K. Gupta, P. Pandey, H. Pucha, P. Sarkar, M. Shah and R. Tewari, 2009. Cloud analytics: Do we really need to reinvent the storage stack? Proceedings of the Workshop on Hot Topics in Cloud Computing, June 15, 2009, San Diego, CA., USA., pp: 1-5.

Assuncao, M.D., R.N. Calheiros, S. Bianchi, M.A.S. Netto and R. Buyya, 2015. Big data computing and clouds: Trends and future directions. J. Parallel Distrib. Comput., 79-80: 3-15.

Berman, J.J., 2013. Introduction. In: Principles of Big Data: Preparing, Sharing and Analyzing Complex Information, Berman, J.J. (Ed.). Morgan Kaufmann, Boston, MA., USA., ISBN-13: 9780124047242.

Bollier, D. and C.M. Firestone, 2010. The promise and peril of big data. The Aspen Institute, Communications and Society Program, Washington, DC., USA.

Chen, H., R.H.L. Chiang and V.C. Storey, 2012. Business intelligence and analytics: From big data to big impact. MIS Quart., 36: 1165-1188.

Chen, M., S. Mao and Y. Liu, 2014. Big data: A survey. Mobile Networks Applic., 19: 171-209.

Cho, B. and I. Gupta, 2010. New algorithms for planning bulk transfer via internet and shipping networks. Proceedings of the IEEE 30th International Conference on Distributed Computing Systems, June 21-25, 2010, Genoa, Italy, pp: 305-314.

Cohen, J., B. Dolan, M. Dunlap, J.M. Hellerstein and C. Welton, 2009. MAD skills: New analysis practices for big data. Proc. VLDB Endowment, 2: 1481-1492.

Cox, M. and D. Ellsworth, 1997. Managing big data for scientific visualization. ACM Siggraph, 97: 146-162.

Gandomi, A. and M. Haider, 2015. Beyond the hype: Big data concepts, methods and analytics. Int. J. Inform. Manage., 35: 137-144.

Gantz, J. and D. Reinsel, 2011. Extracting value from chaos. IDC iView 1142, IDC Research Inc., Framingham, MA., USA., pp: 1-12.

Hajjat, M., X. Sun, Y. Sung, D. Maltz, S. Rao, K. Sripanidkulchai and M. Tawarmalani, 2010. Cloudward bound: Planning for beneficial migration of enterprise applications to the cloud. Proceedings of the ACM SIGCOMM 2010 Conference, August 30-September 3, 2010, ACM, New Delhi, India, pp: 243-254.

Han, J., E. Haihong, G. Le and J. Du, 2011. Survey on NoSQL database. Proceedings of the 6th International Conference on Pervasive Computing and Applications, October 26-28, 2011, Port Elizabeth, pp: 363-366.

Hashem, I.A.T., I. Yaqoob, N.B. Anuar, S. Mokhtar, A. Gani and S.U. Khan, 2015. The rise of big data on cloud computing: Review and open research issues. Inform. Syst., 47: 98-115.

Ji, C., Y. Li, W. Qiu, U. Awada and K. Li, 2012. Big data processing in cloud computing environments. Proceedings of the 12th International Symposium on Pervasive Systems, Algorithms and Networks, December 13-15, 2012, San Marcos, TX., pp: 17-23.

Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh and A.H. Byers, 2011. Big data: The next frontier for innovation, competition and productivity. Report, McKinsey Global Institute (MGI), USA., May 2011.

Miller, H.E., 2013. Big-data in cloud computing: A taxonomy of risks. Inform. Res., Vol. 18.

O'Leary, D.E., 2013. Artificial intelligence and big data. IEEE. Intell. Syst., 2: 96-99.

Purcell, B.M., 2014. Big data using cloud computing. J. Technol. Res., 5: 1-8.

Talia, D., 2013. Clouds for scalable big data analytics. Computer, 46: 98-101.

Tantisiriroj, W., S.W. Son, S. Patil, S.J. Lang, G. Gibson and R.B. Ross, 2011. On the duality of data-intensive file system design: Reconciling HDFS and PVFS. Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, November 12-18, 2011, Seattle, WA., USA

Thusoo, A., Z. Shao, S. Anthony, D. Borthakur and N. Jain et al., 2010. Data warehousing and analytics infrastructure at Facebook. Proceedings of the ACM SIGMOD International Conference on Management of Data, June 6-11, 2010, Indianapolis, IN., USA., pp: 1013-1020.

Wu, Y., C. Wu, B. Li, L. Zhang, Z. Li and F.C.M. Lau, 2012. Scaling social media applications into geo-distributed clouds. Proceedings of the 31st Annual IEEE International Conference on Computer Communications, March 25-30, 2012, Orlando, FL., USA., pp: 684-692.

Zhang, L., C. Wu, Z. Li, C. Guo, M. Chen and F.C. Lau, 2013. Moving big data to the cloud. Proceedings of the 32nd IEEE International Conference on Computer Communications, April 14-19, 2013, Turin, Italy, pp: 405-409.

Zikopoulos, P., D. deRoos, K. Parasuraman, T. Deutsch, J. Giles and D. Corrigan, 2012. Harness the Power of Big Data The IBM Big Data Platform. McGraw Hill Professional, USA., ISBN-13: 978-0071808170, Pages: 280.