

A Survey on Big Data Traffic Management with the Help of Cloud computing (for Biometric systems)

Chetana D.Tukkoji¹, Dr. Seetharam K²

¹Department Of CSE Gitam University Bangalore Campus, India

²Under The Guidance

Abstract: The concept of the big data (BD) in modern days has provided many significant advantages for both social and business aspects. The Big term data can be considered as the collection of huge data sets. The big data may be of structured, semi-structure, unstructured type. The adaptability of BD is widespread in various areas like the education system, industries, Political, governance system, etc. Today the use of BD is also known with traffic management system by the cloud computing mechanism. However, there exist many traditional methods for traffic management system, but they lack with the handling of huge data over the system. This survey paper provides the essential concept of the Big Data covering various research carried by the many authors on Big Data traffic management by cloud mechanism. The paper given an idea of improving traffic management in big data by cloud for future research.

Keywords: component; formatting; style; styling; insert (key words)

II. INTRODUCTION

Today every public and private sector are adopting the big data concept. Big data (BD) is the large amount of data which cannot store, managed or processed significantly by existing way of data management or in general we can say a huge amount of data that goes beyond the analysis [1]. The data may in structured, unstructured or semi-structure. In the last few decades, around 80% of the data is generated, and the growth of the unstructured data is still increasing. As per survey around 3.0 Z terabytes of data is exist now and around 3.0 Exabyte of data is generating every day. It is predicted that around 40 Zettabyte of data may be generated in next half decade [1,2].

The importance of big data is wide spread all over the world. Every minute million of data will be transferred over the internet. The top industries of the world are adopting BD because this is the future aspect of the world. Even various countries have initiated the use of BD. But the data privacy is the major issue that needs to be secured. The big data can be identified with some of its characteristics like data volume, data values, data variety, and data velocity and more interestingly by data complexity. Considering the data volume is the volume of data created by some operation while the data variety is the different kinds of data types that are composed of audio, video, and other digital data types. The data variety can be structured, semi-structured or unstructured. The data value is the type of data that holds definite value. The data velocity is data creation speed. Presently all over the world over millions of emails, twits are sending over the internet. Mainly in BD, the data complexity having various is needed to be considered for every data operation [3, 4].

With all the significant characteristics of BD also face some of the issues including, data storage, data processing, and data management. The data created by many social networks, mails, etc., is increasing day to day as the users of these services are increasing, and the storage of this data is become quite tough. Moreover, the transportation of this much of data over the internet has become biggest issue. Managing the larger volume of data is always being a major issue because many kinds of protocols are used to maintain data validity. The processing of voluminous data is also a biggest issue. The big data challenge like input data processing. It is known fact that an input data can be easily processed to get the output. However, the existing data processing tools handle only relational databases and are not capable of handling the huge volume of data. BD includes various unstructured data that needs to be analyzed to for filtering the better quality of data for analysis. The time variance of data is also the biggest challenge in selection data in real time. The data accuracy, data privacy of the particular user, is also a challenge in BD [4, 5].

The solution for the storage and data transmission is done by transmitting the critical data. However, there is no solution exist for data management. For processing of BD, there must need an algorithm that can handle huge data and perform parallel processing in real time. The cloud computing offers the better solution for all these issues even in real time applications by its scalability nature and processing ability. This survey paper presents the concept of big data traffic along with solution by cloud computing. The sectional organization of the paper is provided as Section II gives the big data, characteristics, issues, and its solution. Section III gives the role of cloud computing in big data traffic management. The research carried by various authors is given in Section IV. By the survey analysis of the research work the research gap is stated in Section V. The future flow

of work to overcome the big data traffic issue by the cloud is given in Section VI. The conclusion of this survey paper is given in section VII.

I. BIG DATA CONCEPTS

In this section, various existing concepts of big data (BD) are presented. The section also includes a various characteristic of BD. The issues related to the big data and solutions are also expressed in this section.

A. Big data:

The term "Big data" reveals that it contains the huge volume of data (it may structure, semi-structured or unstructured). This volume of data cannot be stored, processed or managed by the traditional data centers. Due increased use of digital media for data transfer is creating a vast amount of data in these days. The concept of the big data got buzz since the year 2000. Basically, among the globally created data, only a few or a small amount of data can be analyzed with traditional data tool [6]. However, the analysis of entire data makes the difference in business prospective. The big data concept does not restrict with the large volume of data; how the data can be processed. The BD is related with some of the common terms like data veracity, data capture, data duration, data storage, data variety, data volume, data velocity, data analysis, data visualization, data sharing, etc. [7].

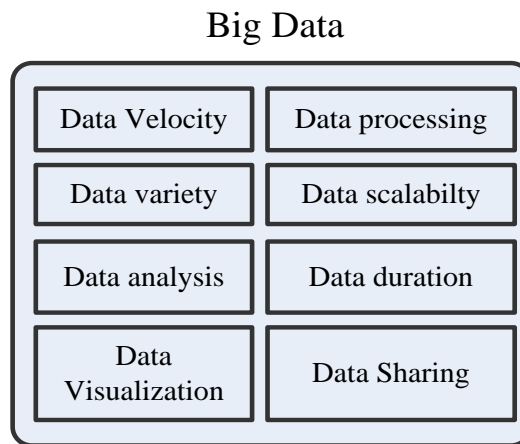


Figure1. Important terms associated with Big Data

B. Big data charecteristics:

The characteristics of the big data can be explained as below.

- **Data Volume (DVl):** This is the volume of data which completely measures the organizational data. The measurement of data is necessary for accessing of important data. The increase in data volume may lead to the reduction of data age, quantity, richness, etc. [8].
- **Data Velocity (DVt):** This functions for measurement of data generation, data streaming and data aggregation speed. Today the e-commerce has gained popularity in different business transactions. The management of data velocity is like bandwidth issue [4].
- **Data Variety (DVr):** The different kinds of data including audio, image, and video can be considered as data variety in big data. These generated data may be of structured or unstructured or semi-structured type.
- **Data Value (DVu):** The measurement of this data value will be helpful in decision making. In big data the value of data, helps in computing [8].
- **Data Complexity (DCt):** The complexity of data is nothing but the interconnection level of data with other elements. The data interconnection may be large or small ripple across the system. This will affect the system behavior in big data. [9].

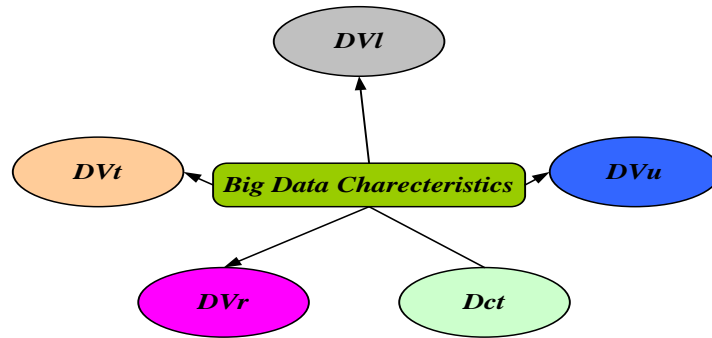


Figure.2. Big data Characteristics

C. Issues in Big data:

In this part the happening issues of big data are described.

- **Transmission and storage:-**

In every time if the data is increased, most of the time the latest storage unit was invented. In recent days the use of social medias or sites is increasing rapidly by which huge of amount of data is generating every day. Also, the data generated by journalist, researchers, and writers is also becoming very large. Thus the storage and transmission of this huge amount of data are been an active research issue in real time application [10].

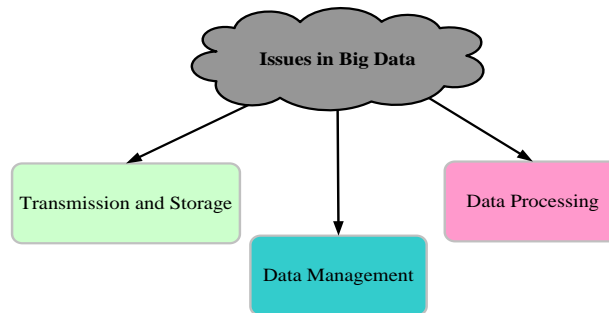


Figure.2. Issues related to big data

- **Data management:-**

This is the known fact that the managing huge amount of data in real time is biggest issues. This problem is addressed since many years. The many concepts like metadata, updating, governance and access are still the unresolved issues. [11].

- **Data processing:-**

The processing issue can be represented with a following addressed example. Consider a data, which is a need to be processed is about 1Exabyte. This much of data needs 635years to process; if 1petabyte of data is processed in 20nanoseconds at 5GHz. The Exabyte of data processing needs parallel processing and some new algorithms to perform actions.

D. Challenges in Big data:

There exist many challenges [12] in big data and are considered as long term challenges in big data research. With increased data volume in recent years are addressing some of the unknown challenges. In this, some of the challenges are briefly stated.

- **Input data and output data processing:**

Recently the challenge arises with the big data output process. The existing data storage can be managed with the processes for the relational databases. The issue is that the tools that exist for transaction processing will add, retrieve, update and search the small data that are helpful in huge data extraction but takes huge time.

- **Ql Vs Qn:-**

The latest challenge is Quality (Ql) versus quantity (Qn). If the user achieves more data and still needs more data. The some user's data acquisition is a necessary thing.

▪ **Data Growth (Dg) Vs Data Expansion (De):-**

The data of an organization will grow if the services for clients are increased. Some of the business unit is intending to expand data when they evolve over time with additional information as new techniques, processes and information demands evolve. Most data is time-varying – the same data items can be collected in the different time domain.

▪ **Speed(Sp) Vs Scale (Sc):-**

If the data volume increases the processing speed will decrease even the scalability will become tougher. In real-time the data will be continuously generating and hence the analysis will become more difficult.

▪ **Structured data (Sd) Vs Unstructured data(Ud):-**

The data of any organization may be structured data (Sd) or Unstructured data (Ud) and for storing in the relational database will Sd and Ud for analytics purpose is required.

▪ **Compliance and Security (C & O)**

For each data in the different domain needs a proper security for important data in every organization. The critical data of an organization may include personal data that can be hacked by the unauthorized person, so privacy promised algorithm is required.

▪ **Any data Vs Total data:-**

The value of each data will be different will also effects on the total data. The data may be temporary hence there need a proper data filtering algorithm to select the desired data.

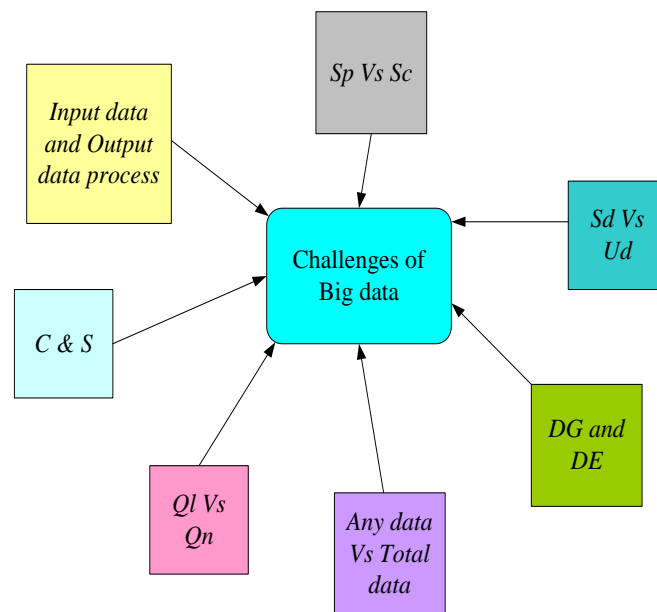


Figure.3. Challenges associated with big data

II. EXISTING WORK IN BIG DATA TRAFFIC MANNAGEMENT AND CLOUD COMPUTING

This section explains the some of the recent work that has performed towards the big data security and traffic management in it by cloud computing.

The mechanism of securing the big data applications over the cloud is presented in the work of Bhargav et al. [13]. As the traditional mechanisms cannot meet the issues of huge data scalability and security is always the biggest issue. In these, work advantages of the cloud computing in big data security are addressed.

The work of Zhang et al. [14] gives the idea of cloud computing and big data. Both the concepts are given promising move towards the today’s information technology. The fact is that the cloud computing offers greater service in various business units. The big data requires more computing power by cloud computing mechanism. The secure, efficient, data sharing and synchronization for big data are discussed in Chard et al. [15].

The concept of big data bringing into the cloud computing is addressed in Khurana [16]. The study has expressed how the integration of cloud and big data together will bring the effectiveness in a public cloud environment.

The work of Nicola et al. [17] has extensively explained the real-time data center traffic and its performance analysis in large scale data. The simulation results for the method analysis is considered with the plot of average input load and latency, average input load and packet loss, average input load and throughput. The analysis shows that better performance is achieved.

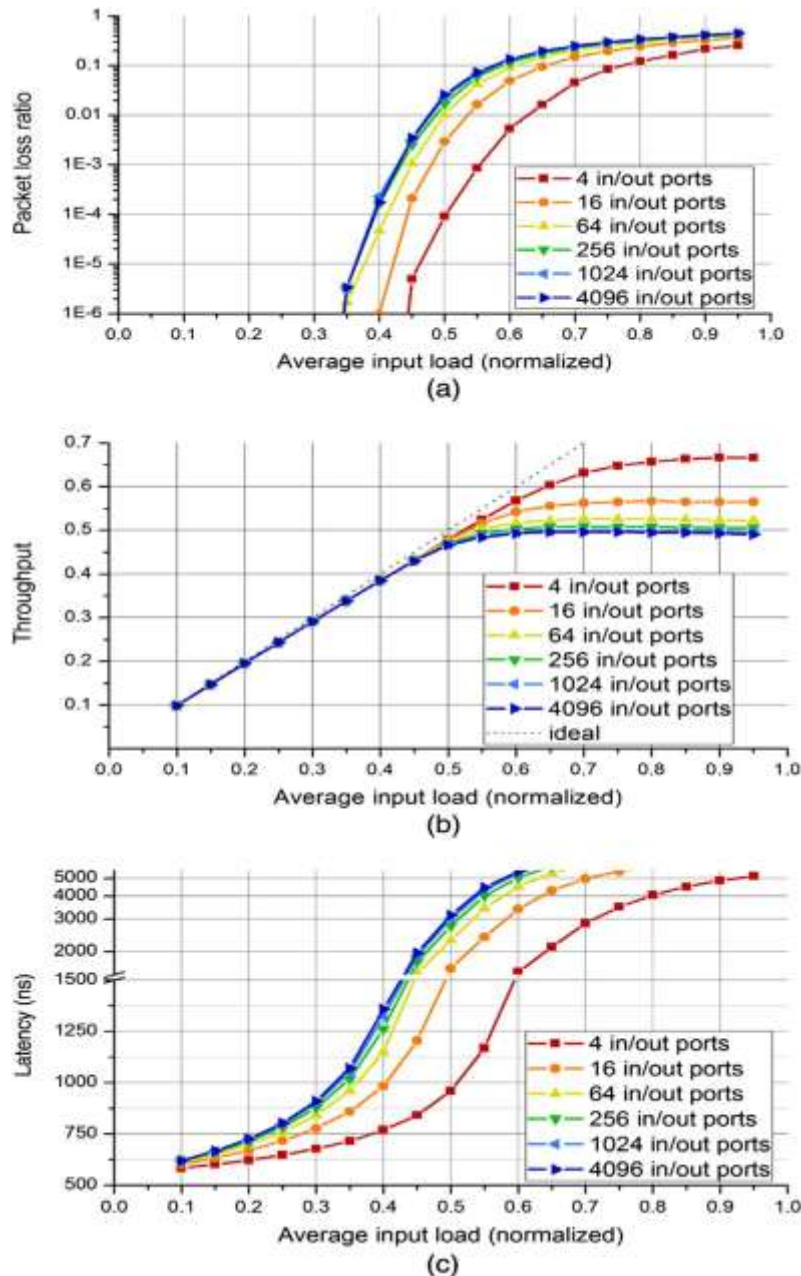


Figure.4. The performance results of Nicola et al. [17]

The evolutionary study on scheduling loads for the big data over the wide spread cloud is explained in Zhang et al. [18]. The study intended to find a better solution for multi-tasking workload in big data and in shortest scheduling time. The experiment carried to analyze the effectiveness of the method with performance factors related to scheduling, throughput, and outcomes with promising results. The work of Rajan [19] has expressed the view of streaming big data processing within the cloud based datacenters. The study was intended to provide a future scope for processing of big data over the cloud with quality.

Author Chen et al. [20] has introduced the multi-label based control over the applications of big data. This study analysis gives a future research idea for healthcare system scalability and access control.

A view of Collins [21] has expressed the interfacing the big data concepts and cloud. The concept described in this work gives an idea how the big data and cloud can be used together. Also in another work author has expressed the impact of big data over the public cloud [22].

A cost minimizing concept for big data cloud is explained in Zhang et al. [23]. The study aims at cost minimizing during the cloud data upload, data migration. The author's research towards the online algorithm is compared with the offline algorithms performance.

The concept of adopting cloud-based big data for enterprises is explained in Liu [24]. In this conceptual work, it has addressed that the functionality of cloud (storing, processing and analyzing) had made a great impact on enterprise usage with required scalability. The best example is Amazon and Google.

The author Talia [25] stated the concept of the cloud for big data. The advanced usage of cloud computing for data management is a major focused topic in this work.

The application platform of cloud computing for big data in the smart grid is expressed in Simmhan et al. [26]. The platform helps to manage large scale of massive grid dataset and helps in expanding the micro grid units of the power grid.

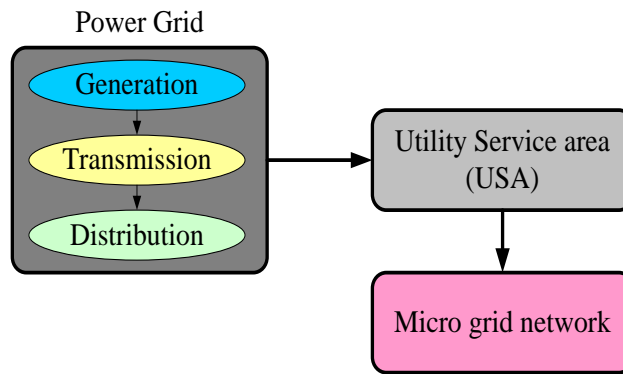


Figure.5. The block view of Simmhan et al. [26].

In the above work [26] the smart meters are placed to analyze the real-time energy usage and communicated over the micro-grid network.

The study towards the carrier cloud and advantages, issues & methodology to overcome these issues are explained in Taleb [27]. This study was meant to scale the mobile internet traffic by using cloud computing.

The mechanism for modeling and optimizing the big data is presented in Slavakis et al. [28]. The work helps to solve the issues like network visualization, prediction, load traffic, decentralization, etc.

III. EXISTING STUDY SURVEY

The survey of recent work that has been published in IEEE transaction for big data traffic and cloud computing are cited at 7.00pm (29/8/2016). The details of each conference, journals, articles, books, standards and courses are given in following table.1.

Type	Big data traffic	Big data traffic + Cloud
Conference Publications	27661	84
Journals & Magazines	2947	5
Early Access Articles	518	0
Books and e-Books	102	0
Standards	37	0
Courses	02	0

Table.1.Statistic of collected from IEEE Xplore.

From the above table.1, we can see that both Big data and cloud computing are the trending research topic. The table above details are represented in the following the plot as figure.6.

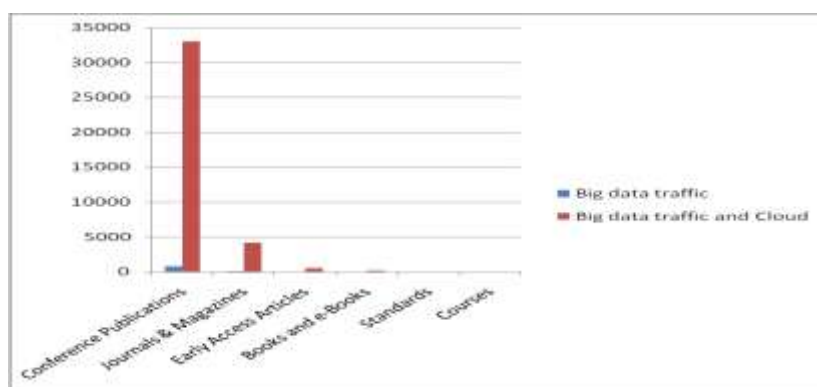


Figure.6. Plot of Study statistics

IV. RESEARCH GAP

Over the decades number of works has been carried to overcome the issues related with increasing data size. There were some traditional data scaling tools which manage or process the small volume of data. Since in last decade the digitalization in every field is creating huge volume of data and is considers as Big data. The research in big data offers many significance with is some of the characteristics like *Dvt, Dct, DVu, Dvl and Dvr*. The big data is also having data security and data traffic issue in real time applications in that sense many of the researchers have given solutions, but they are not up to the expectations.

V. FUTURE RESEARCH SCOPE

The future work can be performed with following methods.

- A novel algorithm can be designed which integrates both cloud computing and big data.
- MapReduce can be used to generate a large volume of data set and process the novel algorithm.
- Hadoop platform can be used to predict some peak traffic.
- A framework can be designed to minimize the big data traffic using cloud computing.
-

VI. CONCLUSION

This survey paper gives an idea to minimize the data traffic of big data with the help of cloud computing. As the big data refers the large volume of data and cloud computing can offer the better scalability for the large data. In this paper top and recent journal papers from IEEE Xplore are surveyed to understand the existing research work. A section Existing study survey is described that gives details of work done in big data traffic and cloud computing. A research gap is addressed that gives an overview of existing research drawbacks. The future research scope hints an idea for better data traffic management with big data.

REFERENCES

- [1]. Rubinstein, Ira. "Big data: the end of privacy or a new beginning?." International Data Privacy Law (2013 Forthcoming) (2012): 12-56.
- [2]. Chen, Hsinchun, Roger HL Chiang, and Veda C. Storey. "Business Intelligence and Analytics: From Big Data to Big Impact." MIS quarterly 36.4 (2012): 1165-1188.
- [3]. Géczy, Peter. "Big data characteristics." The Macrotheme Review 3.6 (2014): 94-104.
- [4]. Kaisler, Stephen, et al. "Big data: issues and challenges moving forward." System Sciences (HICSS), 2013 46th Hawaii International Conference on. IEEE, 2013.
- [5]. Hashem, Ibrahim Abaker Targio, et al. "The rise of "big data" on cloud computing: Review and open research issues." Information Systems 47 (2015): 98-115.
- [6]. Lohr, Steve. "The age of big data." New York Times 11 (2012).
- [7]. Chen, Hsinchun, Roger HL Chiang, and Veda C. Storey. "Business Intelligence and Analytics: From Big Data to Big Impact." MIS quarterly 36.4 (2012): 1165-1188.
- [8]. Jagadish, H. V., et al. "Big data and its technical challenges." Communications of the ACM 57.7 (2014): 86-94.
- [9]. Suthaharan, Shan. "Big data classification: Problems and challenges in network intrusion prediction with machine learning." ACM SIGMETRICS Performance Evaluation Review 41.4 (2014): 70-73.
- [10]. Jacobs, Adam. "The pathologies of big data." Communications of the ACM 52.8 (2009): 36-44.
- [11]. Gandomi, Amir, and Murtaza Haider. "Beyond the hype: Big data concepts, methods, and analytics." International Journal of Information Management 35.2 (2015): 137-144.

- [12]. Bizer, Christian, et al. "The meaningful use of big data: four perspectives--four challenges." *ACM SIGMOD Record* 40.4 (2012): 56-60.
- [13]. Angin, Pelin, et al. "An entity-centric approach for privacy and identity management in cloud computing." *Reliable Distributed Systems, 2010 29th IEEE Symposium on*. IEEE, 2010.
- [14]. Bharat Bhargava; Ibrahim Khalil; Ravi Sandhu,"Securing Big Data Applications in the Cloud [Guest editors' introduction]", *IEEE Cloud Computing*, Volume: 1, Issue: 3,Year: 2014.
- [15]. Zhang, Liang-Jie. "Editorial: Big services era: Global trends of cloud computing and big data." *IEEE Transactions on Services Computing* 5.4 (2012): 467-468.
- [16]. Chard, Kyle, Steven Tuecke, and Ian Foster. "Efficient and secure transfer, synchronization, and sharing of big data." *IEEE Cloud Computing* 1.3 (2014): 46-55.
- [17]. Amandeep Khurana "Bringing Big Data Systems to the Cloud"*IEEE Cloud Computing*, Volume: 1, Issue: 3,Year-2014.
- [18]. Calabretta, Nicola, et al. "On the performance of a large-scale optical packet switch under realistic data center traffic." *Journal of Optical Communications and Networking* 5.6 (2013): 565-573.
- [19]. Ranjan, Rajiv. "Streaming big data processing in datacenter clouds." *IEEE Cloud Computing* 1.1 (2014): 78-83.
- [20]. Chen, Hongsong, Bharat Bhargava, and Fu Zhongchuan. "Multilabels-Based Scalable Access Control for Big Data Applications." *IEEE Cloud Computing* 1.3 (2014): 65-71.
- [21]. Collins, Eli. "Intersection of the cloud and big data." *IEEE Cloud Computing* 1.1 (2014): 84-85.
- [22]. Collins, Eli. "Big data in the public cloud." *IEEE Cloud Computing* 2.1 (2014): 13-15.
- [23]. Zhang, Linqun, et al. "Moving big data to the cloud: an online cost-minimizing approach." *IEEE Journal on Selected Areas in Communications* 31.12 (2013): 2710-2721.
- [24]. Liu, Huan. "Big Data Drives Cloud Adoption in Enterprise." *IEEE internet computing* 17.4 (2013).
- [25]. Talia, Domenico. "Toward cloud-based big-data analytics." *IEEE Computer Science* (2013): 98-101.
- [26]. Simmhan, Yogesh, et al. "Cloud-based software platform for big data analytics in smart grids." *Computing in Science & Engineering* 15.4 (2013): 38-47.
- [27]. Taleb, Tarik. "Toward carrier cloud: Potential, challenges, and solutions." *IEEE Wireless Communications* 21.3 (2014): 80-91.
- [28]. Slavakis, Konstantinos, Georgios B. Giannakis, and Gonzalo Mateos. "Modeling and optimization for big data analytics:(statistical) learning tools for our era of data deluge." *IEEE Signal Processing Magazine* 31.5 (2014): 18-31.