Sketch of Big Data Real-Time Analytics Model

Bakhtiar M. Amen
School of Computing and Engineering
The University of Huddersfield
Huddersfield, UK
e-mail: bakhtiar.amen@hud.ac.uk

Joan Lu
School of Computing and Engineering
The University of Huddersfield
Huddersfield, UK
e-mail: j.lu@hud.ac.uk

Abstract— **Big Data has drawn huge attention from researchers in information sciences, decision makers in governments and enterprises. However, there is a lot of potential and highly useful value hidden in the huge volume of data. Data is the new oil, but unlike oil data can be refined further to create even more value. Therefore, a new scientific paradigm is born as data-intensive scientific discovery, also known as Big Data. The growth volume of real-time data requires new techniques and technologies to discover insight value. In this paper we introduce the Big Data real-time analytics model as a new technique. We discuss and compare several Big Data technologies for real-time processing along with various challenges and issues in adapting Big Data. Real-time Big Data analysis based on cloud computing approach is our future research direction.**

Keywords - Big Data Analytics; Real-time Analytics; Big Data state-of-the-art

I. INTRODUCTION

The term "Big Data" is universal and has gained popularity within the domain of scientist, bioinformatics, geophysics, astronomy and meteorology [1]. In fact, all Big Data has blind spot areas in which data are missing, scarce, or otherwise unrepresentative of the data domain [2]. Big Data analytics enable enterprise and scientists to extract usable information out of enormous, complex, interconnected and varied datasets. However, from 2.8 Zettabytes of global data only 0.5% of these data was analyzed in 2012 [3]. In addition to this, current Big Data techniques and technologies are incapable of storing, processing or analyzing data, as data is not extracted by particular scientific disciplines (e.g., bioinformatics, geophysics, astronomy and meteorology). The way in which people think about data and data analysis will gradually change as well, in addition to the technological possibilities. Thanks to the latest internet technologies, the potential for harnessing all that can be measured and analysed using solid data, intelligent sensors and filtering has never been as promising and lucrative as today, at the dawn of the digital era [4].

The Big Data paradigm consists of batch and real-time processing [5]. The batch process focuses entirely on structured and semi-structured data. Likewise, the goal of real-time processing paradigm is to deal with velocity of Big Data such as processing streaming data but with low latency. This paper aims to review the background of Big Data and compare several Big Data real-time processing technologies, as well as introduce the new real-time Big Data analytics model.

The rest of this paper is organized as follows: In Section 2 we briefly overview some concepts of Big Data including its definition, characteristics and size. Section 3 presents the Big Data domains. Section 4 presents related work. Section 5 presents an evaluation and discussion. The article culminates in a conclusion and recommendation for future work.

II. BACKGROUND

In this section, we present Big Data definitions, its characteristics, followed by the Big Data revenue and the size of global data. Next, we present a Big Data technology map.

A. Big Data Definition

Big Data is one of the key buzzwords in the current technological landscape, but there is no agreed definition by either academia or industry. Chen et al. [6] defined Big Data as "Datasets which could not be captured, managed, and processed by general computers within an acceptable scope". Hashem et al. [7] also defined Big Data as "a term utilized to refer to the increase in the Volume of data that is difficult to store, process, and analyze through traditional database technologies". However, these definitions basically state the most obvious dimensions of Big Data Volume, Variety, Velocity and Veracity. Whereas, the data flows in today's digital era are being produced around the clock and all over the world.

B. Big Data Characteristics

The conjunction of these four dimensions helps both to define and distinguish Big Data. Volume refers to the amount of data from Terabyte to Petabyte and Exabyte to Zettabyte [6]. Variety refers to various data sources collected from web logs, social networks, machines, sensors, transactions and the internet of things, in different formats of semi-structured and unstructured [8]. Velocity refers to the speed at which data is generated and the speed of data transfer [7]. Data has become an extremely valuable factor in business productivity and the opportunity to discover new value from it. The 4V's of Big Data are shown in Figure 1.

 ISBN: 978-1-61208-415-2

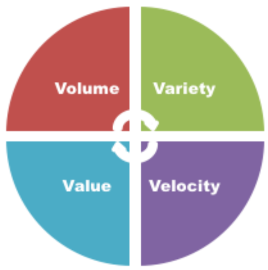

Figure 1. The characteristics of Big Data

C. Size of Global Data

The size of digital data has been growing at an increasing rate. Figure 2 depicts the size of created data volumes in percentages across the USA, West Europe, India and the rest of the world. According to the International Data Corporation (IDC) study, the size of global data in 2009 was 1.8 Zettabyte, it increased to 8 Zettabyte in 2015 [9]. It is doubling in size every two years, and by 2020 the digital universe data is estimated to be 44 Zettabytes [10].

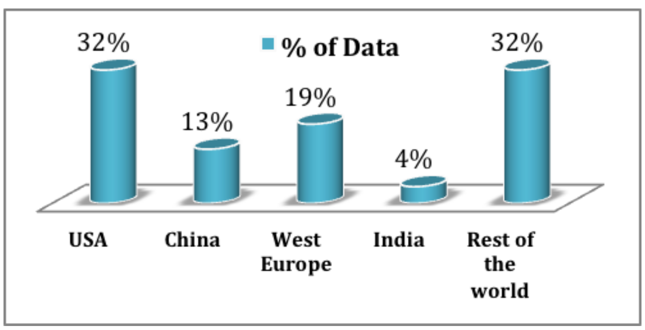

Figure 2. The scale of global data generates in percentages.

Furthermore, the explosive growth of global data increased rapidly. In fact, 90% of the world's entire data was created since 2012 [11], whereas only 10% of all these data is structured data compare to 80% is unstructured data [11]. Figure 3 depicts the scale of global structured data versus unstructured data.

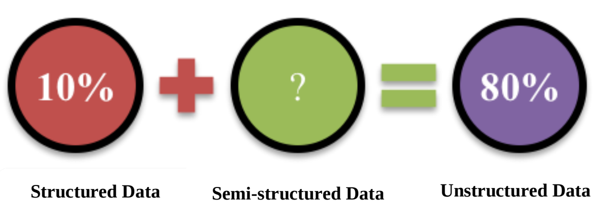

Figure 3. The scale of universe data format

D. Big Data Revenue

Manyika et al. [12] estimated that the power of Big Data analytics guaranteed 60% of potential revenue through new opportunities from location-aware and location-based services. In reality, the ambiguous demand in the Big Data era is more related to business insights since the 4Vth Value of Big Data has been introduced. Big Data revenue increased from $3.2 billion to $16.9 billion between 2010 and 2015 [6]. However, the potential value to consumers, business and users are estimated to be $700 billion in the next ten years [7].

E. Big Data Real-time State- of-the-art

Hadoop is known as innovative in Big Data analytics, since Hadoop has the ability to touch 50% of the global data by 2015 [13]. In fact, Hadoop and MapReduce have been criticized by both academia and enterprises for their real-time limitations. The MapReduce programming model is an open-source version of Hadoop [13][14]. Fan et al. [14] stated that Hadoop made a world record in sorting one petabyte of data within 16.25 hours and one terabyte of data in only 62 seconds. Furthermore, the Hadoop ecosystem consists of several projects as introduced only the real-time applications in Figure 7.

Twitter has developed storm in 2011 [10][15] for data streaming processing. Storm is an open source and it has been improved in scalability while maintaining a low latency for real-time data stream processing, which integrates with other queuing and bandwidth systems. Storm consists of several moving parts, including the coordinator (ZooKeeper), state manager (Nimbus) and processing nodes (Supervisor). Yahoo has developed S4 in 2010 [13][14][16] for data stream parallel distributing processing. Kafka also developed LinkedIn in 2011 [16] for the purposes of messaging processing. Spark [9] Stream is an extension of Spark that supports continuous stream processing. In practice, some other new computing models have recently been introduced for stream data processing (e.g., GraphLab and Dryad), which are suitable for machine learning and data mining programming models [16] .

III. BIG DATA DOMAINS

In this section, we describe some of the challenges and issues of Big Data in several disciplines from both industry and scientific perspectives.

A. Big Data in the Bio-Medical Sector

Kambatla et al. [17] highlighted that "healthcare and human welfare is one of the most convincing applications of Big Data analytics, it is a fastest growing datasets". In fact, a large amount of medical data sources like RMI scans, bioscience data, and genomic data are becoming more complex and difficult to be captured, storage, and analyzed [18]. Although, China attempts to collect and store 30 million of traceable biological samples by 2015 [19]. Manyika et al. [12] stated that every year the USA has wasted more than $2 trillion in healthcare sectors. Implementing Big Data analytics technique has helped the US to save $300 billion as well as helping Europe to save over $149 billion. In addition, bio-informatics requires new advanced computational techniques to support efficient knowledge discovery.

B. Big Data in Enterprise

Facebook, Google, Yahoo and Falcon are creating large scale of data. As an example, Wal-Mart produced over 1 million customer transactions per hour across 6000 stores [6]. Amazon Web Services (AWS) has also been successful in IaaS services with 70% of their market share including the most popular Elastic Compute Cloud (EC2). A Simple Storage Service (S3) enables the processing of 500,000 queries over millions of terminal operations from third party sellers each day [6][20]. Akamai also managed to analyze 75 million events per day. However, the most observable domain in Big Data analytics is value [20]. Hence, Data has become extremely valuable in enterprise to produce productivity and business predictions.

C. Big Data in Scientific Research

Every day NASA solar observatory and telescopes are capturing more than 1.6 TB of high quality images and collecting 140 TB data from the large synoptic survey telescope [21]. Likewise, one space satellite generates over 800 gigabytes of data on a daily bases. In 2012, the Earth Observing System Data and Information System (EOSDIS) also succeeded in distributing more than 4.5 million gigabytes of data per day [20]. However, physicists and astronomists have made numerous efforts to engage with massive crowded data for many years to test the novelty of our universe.

D. Big Data in Engineering

The key challenge in the area of engineering is the discovery of techniques that are able to process machinery and the internet of things data. These sources are creating massive amounts of data through embedded networking and real-time approaches. The size of the internet of things data is estimated to be one trillion by 2030 [20]; this includes 350 billion annual meter readings. The volume of data generated in engineering is by a wide range of sensors, through power plants, machinery data and GPS as well as electronic devices [22].

IV. RELATED WORK

Patel [23] highlighted several issues and challenges in storing, processing and analyzing data in real-time. The author argued that highly efficient algorithm and technology will enhance the accuracy of valuable information [24]. Ranjan [24] investigated in different Big Data applications and discussed their differences from traditional analytics, and he also described the new solutions for real-time Big Data analytics. Kambatla et al. [17] implemented several projects in a real-time data caching and processing graph in one of Google's distributing systems. The author also highlighted that current technology such as Hadoop incapable of processing large-scale of graphs. Hadoop mainly consists of two components; Hadoop File System (HDFS) and Programming model (Map Reduce) [25]. HDFS stores huge data set reliably and streams it to user application at high bandwidth and MapReduce is a framework that is used for processing massive data sets in a distributed fashion over a several machines. It has two parts- job tracker and task tracker [26].

Hu [27] proposed HACE theorem which characterizes the features of the Big Data revolution, and recommends the Big Data processing model, from a data mining perspective. Moreover, the rapid growth of complex diversity and dimensionality of the Remote Sensor (RS) lies in collected metadata to analyze as stated in [28]. The recent lower level parallel programming was comprehensively engaged with RS image processing along with a multi-level hierarchical cluster. Hence, parallel programming is required for RS applications to predict accurate results. According to Ma et al. [28], the current Big Data analytics model is beyond the capabilities of processing and analyzing real-time Satellite Data.

The scale of remote satellite data is depicted in Figure 4; this scale demonstrates the volumes of satellite Data per day as well as per year across the world.

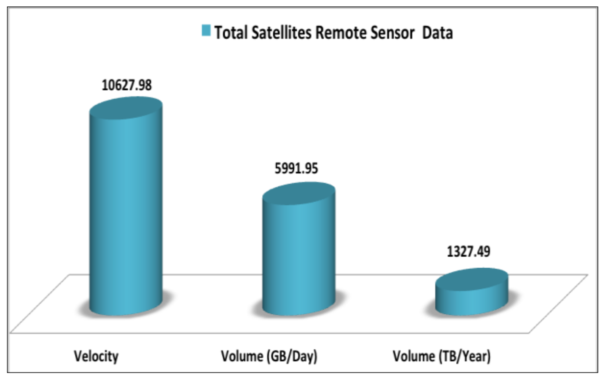

Figure 5. Scale of Global Satellite Remote Sensor Data.

Two Big Data real-time/stream analytics model were found in our literature, known as Simith's Big Data real-time analytics Model [29] and Big Data life cycle management model [30]. Khan et al. [16] proposed Big Data life cycle management model using the technologies and terminologies of Big Data. The author's proposed data life cycle consists of: Data Acquisition/Generation, Data Collection, Data storing (temporarily/permanently), and Data Analysis. Likewise, Barlow [29] presented Smith's five phases of the real-time Big Data analytics model which includes: Data extraction, development model, validation and deployment, real-time scoring, and model refresh.

Barlow also stated that the correct analytics model is necessary to process heterogeneous data in real time. Furthermore, this model is utilized from the high-performance of data mining, predictive analytics, text mining, and data optimization to enhance the decision makers [1][31]. In fact, the heart of any prediction system is the Model, for instance, a credit card fraud prediction system could leverage a model built using previous credit card transaction data over a period of time.

 ISBN: 978-1-61208-415-2

TABLE I. SMITH'S BIG DATA ANALYTIC MODEL

Analytic Model	Descriptions
Data Extraction/Distillation	Like unrefined oil, heterogeneous data types are messy and complex. Emerging new extracting models and performing accurate analysis are necessary and challenging to handle unstructured data [11][18].
Development Model	In this phase, the model process consists of speed, flexibility productivity, and reproducibility.
Validation and Deployment	Extracting fresh data and running against the model and comparing the results with other existing models leading into productivity [13].
Real-time Scoring	Data in real-time scoring is triggered by actions at the decision layer. At this phase of the process, the deployed scoring rules are "divorced" from the data in the data layer or data mart [21].
Model Refresh	Data is always changing, it is necessary to refresh the data and refresh the model built on the original data. Simple exploratory data analysis is also recommended.

Hu et al. [1] categorized Big Data analytics into Descriptive, Predictive and Prescriptive. Descriptive analytics focuses on historical data and the description of what occurred previously from Data visualization results. Predictive analytics focuses on future probabilities and describes the business value outcome. Advanced analytics [31] is known as Prescriptive analytics which address the decision making efficiently. For example, simulation is used to analyze complex systems to gain insight into system behavior and identify issues and optimization techniques are used to find optimal solutions under given constraints. However, only about 3% of companies are utilizing prescriptive analytics to predict future events according to a recent Gardner Research survey [32].

V. EVALUATION AND DISCUSSION

Despite the availability of new technologies for handling massive amounts of data at incredible speeds, the real promise of advanced data analytics lies beyond the area of pure technology. The existing Big Data analytics appears to be suffering from a lack of effectiveness compared to the speed of real-time data volume. Thus, Big Data real-time analytics has been proposed to describe the advanced analysis methods or mechanisms for massive data [11]. In fact, increasing the heterogeneous data in the real-time monition from various data sources (e.g., The Internet of Things, multimedia, social networking) plays a significant role in Big Data. In addition to these, the new real-time model shown in Figure 6 is required. Banerjee [33] highlighted the traditional analytics versus real-time analytics in Figure 5. Banerjee also compared several parameters in each feature from storage cost to support cost. It shows that Big Data analytics is more reliable in terms of data speed, time and velocity compared to traditional analytics. This highlights the key differences between the realities of yesterday's analytics and the predictions for today's Big Data analytics.

	Traditional Analytics	Big Data Analytics
Storage Cost	High	Low
Analytics	Offline	Real-time
Utilizing Hadoop	No	Yes
Data Loading Speed	Low	High
Data Loading Time	Long	Average 50%-60% faster
Data Discovery	Minimal	Critical
Data Variety	Structured	Unstructured
Volume	Gigabyte, terabyte	Petabyte, Exabyte, Zettabyte
Velocity	Batch	Real-time
Administration Time	Long	Average 60% faster
Complex Query Response Time	Hours/days	Minutes
Data Compression Technique	Not matured	Average 40%-60% more data compression
Support Cos	High	Low

Figure 7. The Traditional Analytics versus Big Data Analytics [33].

In addition to these, we implemented the new real-time analytics model from Smith's model as depicted in Figure 6, because Smith's model was precisely based on data mining and text mining [28]. As shown in Figure 6, this model consists of five phases: Data Extraction, Data Cleaning/Filtering, Data Analysis, Data Visualization, and Decision-making.

In the Data extraction phase, Data is required to be processed by one of the real-time technologies such as Storm and S4 as highlighted in Figure 7. Data must be cleaned before being transformed for analyzing to unlock the hidden potential value from it. Therefore, the second phase of filtering technique is required for two reasons. Firstly, data intent to lies in the extracting stage as indicated in [28]. Secondly, processing data without filtering means invalid results. Data visualization has to communicate and predict data through graphics to aid decision-making through sophisticated analytics results. In addition to this, advanced analytics for massive data is required as a new solution to effectively improving decision making in the final phase. As a result, the process of Real-time data analytics is still a challenging task and the model requires an advanced computational and robust real-time algorithm to predict it efficiently.

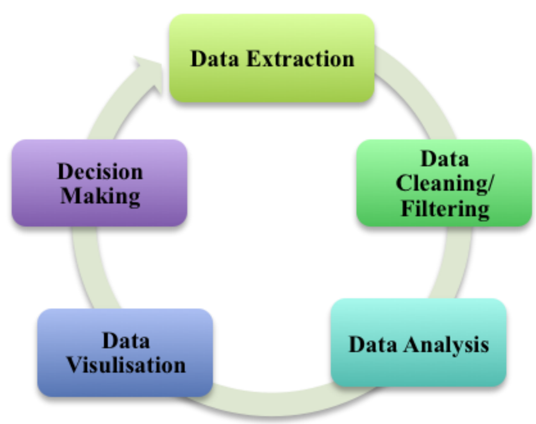

Figure 8. The Big Data Real-time analytics model.

Furthermore, selecting an appropriate real-time analytics model and technology depends on data objects. As depicted in Figure 7, we highlighted the Big Data real-time processing state-of-the-art in terms of its developers, programming model, capabilities, and limitations of data structure types. We highlighted each application's advantages and disadvantages as well as their architectures. The results show that, first, real-time processing is becoming more important in real-time analytics, likewise batch processing remains the most common data processing paradigm [28, 34].Second, most of the systems adopted a graph programming model, because the graph processing model can express more complex tasks. Third, all the systems support concurrent execution to accelerate the processing speed. Fourth, data stream processing models use memory as the data column storage to achieve higher access and processing rates, whereas batch-processing models employ a file system or disk to store massive data and support multiple visiting. Fifth, some of these real-time technologies were backed by partially fault tolerant and have limitations in their node backup as highlighted in Storm and S4 [1][13][14][15][16].

(e.g., semi-structured or unstructured), and this helped us to highlight their advantage and disadvantages.

In the second part of our study, we discussed Smith's five phases of the Big Data real-time analytics model as depicted in Table 1. Furthermore, we introduced our real-time Big Data analytics model as shown in Figure 6. Throughout our investigation, the real-time analytics appears to play a key role in Big Data and enrich the potential revenue. Likewise, it needs further research and collaborations between the scientists and industries to improve the real-time analytics bottleneck. In fact, a different storage mechanism is required, because all of the data cannot fit in a single type of storage area [35]. Hence, Cloud computing is playing an important role as it gives organizations the ability to store and analyze revolutionized data economically and offers extensive computing resources [16].

As result, the motivation for undertaken this research was an attempt to develop a real-time Big Data analytics framework which enable to enrich the decision-makers in the real-time monition. This research allowed us to identify the weaknesses of existing systems, and to design a roadmap of contributions to the state of the art.

	Big Data Streaming Analysis State-Of-The-Art							
	Developer	**Application**	**Programming Model**	**Specified Use**	**Structure Type**	**Advantages**	**Disadvantages**	**Architecture**
Storm - 2011 [3] [18] [28]	**Twitter**	Kafka, HBase (Storm-HBase) Twitter [23]	Directed Acyclic graph [3] [18]	Distribute real-time computation system for processing fast, large streams of data [18]	Un-structured, Real time Streaming process [18] [28][27]	Embeddable networking library API [1][3] Scalable, fault-tolerant, and is easy to set up and operate [3][18]	Partial fault tolerance [3] Lack of dedicated backup nodes [10]	Parallel-Distributed [23] Master-workers [3]
S4 – 2010 [3] [18] [25] [28]	**Yahoo**	Distributed Streaming process [3][18]	Directed Acyclic graph [3][18] [30]	Worker processes and execution, [3] Graph of Processing Elements [27]	Un-structured Real time streaming processing [3][18] [21] [27]	Distributed, Scalable, Fault-tolerant [21] [25] Pluggable platform [18] Easy develop applications [21]	Node fail data lose, Partial fault tolerance[3] [25] Lack of dedicated backup nodes [10]	Decentralised and systematic [23]
Kafka - 2011 [2][18][28]	**LinkedIn**	Messaging system Tool [28]	Distributed Messaging system [28]	Distributed, partitioned and replicated commit log services tools [18]	Real-time streaming process [3] Un-structured [18]	High-throughput stream of unchallengeable activity data [18][28]		Column store [22]

Figure 9. Big Data stream processing state-of-the-art

VI. CONCLUSION AND FUTURE WORK

In this research, we investigated on two research areas: in the first part of study, we presented the concepts of Big Data as well as some of the challenges and issues in both industry and scientific domains, followed by a comparison of several Big Data real-time processing technologies in terms of their capabilities and limitations as shown in Figure 7. However, each of these technologies were compared in terms of their architecture, programming model, data structure capabilities

Our future plan is to investigate on real-time analytics based on cloud computing and attempts to answer the following questions:

- Investigate on existing cloud paradigms and highlight their limitations in real-time analytics aspects.
- Develop an algorithm for the real-time analytics based on cloud computing.
- Determine how to test, implement and compare our results with other existing cloud computing technologies.

 ISBN: 978-1-61208-415-2

REFERENCES

[1] Y. W. Han Hu, Tat-Seng Chua, Xuelong Li, "Toward Scalable Systems for Big Data Analytics: A Technology Tutorial," vol. 2, pp. 652-687, 2014.

[2] J. J. Berman, Principles of big data: preparing, sharing, and analyzing complex information: Newnes, 2013.

[3] N. Khan, I. Yaqoob, I. A. T. Hashem, Z. Inayat, W. K. M. Ali, M. Alam, M. Shiraz, and A. Gani, "Big data: survey, technologies, opportunities, and challenges," TheScientificWorldJournal, vol. 2014, p. 712826, 2014.

[4] G. Aydin, I. R. Hallac, and B. Karakus, "Architecture and Implementation of a Scalable Sensor Data Storage and Analysis System Using Cloud Computing and Big Data Technologies," Journal of Sensors, vol. 2015, pp. 1-11, 2015.

[5] K. Kolomvatsos, C. Anagnostopoulos, and S. Hadjiefthymiades, "An Efficient Time Optimized Scheme for Progressive Analytics in Big Data," Big Data Research, 2015.

[6] M. Chen, S. Mao, and Y. Liu, "Big Data: A Survey," Mobile Networks and Applications, vol. 19, pp. 171-209, 2014.

[7] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. Ullah Khan, "The rise of “big data” on cloud computing: Review and open research issues," Information Systems, vol. 47, pp. 98-115, 2015.

[8] N. Khan, I. Yaqoob, I. A. T. Hashem, Z. Inayat, W. K. M. Ali, M. Alam, M. Shiraz, and A. Gani, "Big data: survey, technologies, opportunities, and challenges," vol. 2014, p. 712826, 2014.

[9] J. J. Berman, Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information, 1st ed.: Morgan Kaufmann, 2013.

[10] D. R. John Gantz, "The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East," 2013.

[11] C. L. Philip Chen and C.-Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data," Information Sciences, vol. 275, pp. 314-347, 2014.

[12] B. Brown, M. Chui, and J. Manyika, "Are you ready for the era of ‘big data’," McKinsey Quarterly, vol. 4, pp. 24-35, 2011.

[13] C. Dobre and F. Xhafa, "Parallel Programming Paradigms and Frameworks in Big Data Era," International Journal of Parallel Programming, vol. 42, pp. 710-738, 2014.

[14] F. H. Jianqing Fan, and Han Liu, "Challenges of Big Data Analysis," National science review, vol. 1, pp. 293-314, Jun 2014.

[15] B. D. a. Zachary Miller a, William Deitrick a, Wei Hua, Alex Hai Wang, "Twitter spammer detection using data stream clustering," 2013.

[16] N. Khan, I. Yaqoob, I. A. Hashem, Z. Inayat, W. K. Ali, M. Alam, M. Shiraz, and A. Gani, "Big data: survey, technologies, opportunities, and challenges," TheScientificWorldJournal, vol. 2014, p. 712826, 2014.

[17] K. Kambatla, G. Kollias, V. Kumar, and A. Grama, "Trends in big data analytics," Journal of Parallel and Distributed Computing, vol. 74, pp. 2561-2573, 2014.

[18] M. C. James Manyika, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, Angela Hung Byers, "Big data: The next frontier for innovation, competition, and productivity," 2011.

[19] V. C. M. Leung, M. Chen, Y. Zhang, and S. Mao, Big Data: Related Technologies, Challenges and Future Prospects. DE: Springer Verlag, 2014.

[20] C. Esposito, M. Ficco, F. Palmieri, and A. Castiglione, "A knowledge-based platform for Big Data analytics based on publish/subscribe services and stream processing," Knowledge-Based Systems, vol. 79, pp. 3-17, 2014.

[21] J. Green, P. Schechter, C. Baltay, R. Bean, D. Bennett, R. Brown, C. Conselice, M. Donahue, X. Fan, and B. Gaudi, "Wide-field infrared survey telescope (WFIRST) final report," arXiv preprint arXiv:1208.4012, 2012.

[22] H. J. Watson, "Tutorial_ Big Data Analytics_ Concepts Technologies and Applica," Journals at AIS Electronic Library, vol. 34, pp. 1247-1268, 2014.

[23] A. B. Patel, M. Birla, and U. Nair, "Addressing big data problem using Hadoop and Map Reduce," 2012, pp. 1-5.

[24] B. W. Dan Vesset, Henry D. Morris, Richard L. Villars, Gard Little, Jean S. Bozman, Lucinda Borovick, Carl W. Olofson, Susan Feldman, Steve Conway, Matthew Eastwood, Natalya Yezhkova, "Market Analysis Worldwide Big Data Technology and Service " IDC Analyse Future, vol. 1, 2012.

[25] A. Katal, M. Wazid, and R. H. Goudar, "Big data: Issues, challenges, tools and Good practices," 2013, pp. 404-409.

[26] J. Nandimath, E. Banerjee, A. Patil, P. Kakade, and S. Vaidya, "Big data analysis using Apache Hadoop," pp. 700-703.

[27] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, "Data Mining with Big Data," IEEE Transactions on Knowledge and Data Engineering, vol. 26, pp. 97-107, 2014.

[28] Y. Ma, H. Wu, L. Wang, and B. Huang, "Remote sensing big data computing: Challenges and opportunities," Future Generation Computer Systems, vol. in press, 2014.

[29] M. Barlow, "Real-Time Big Data Analytics: Emerging Architecture," pp. 24-25, 2013.

[30] R. Casado and M. Younas, "Emerging trends and technologies in big data processing," Concurrency and Computation: Practice and Experience, vol. 27, pp. 2078-2091, 2015.

[31] Vijay Srinivas Agneeswaran, "Big Data Analytics Beyond Hadoop " pp. 22-23, 2014.

[32] S. P. Soumya Sree Laxmi P. , "Impact of Big Data Analytics on Business Intelligence-Scope of Predictive Analytics," vol. Vol.5, 2015.

[33] A. Banerjee, "Big Data & Advanced Analytics in Telecom: A Multi-Billion-Dollar Revenue Opportunity" pp. 7-22, 2013.

[34] Y. Ma, H. Wu, L. Wang, B. Huang, R. Ranjan, A. Zomaya, and W. Jie, "Remote sensing big data computing: Challenges and opportunities," Future Generation Computer Systems, 2014.

[35] E. B. Dudin and Y. G. Smetanin, "A review of cloud computing," Scientific and Technical Information Processing, vol. 38, pp. 280-284, 2011.