IJCTA, 9(27), 2016, pp. 479-488 ISSN: 0974-5572

A Comprehensive Study on Big-Data Analytics-Tools, Techniques, Technologies and Applications

S. Arista*, A. Arun Gnanaraj and J. Gnana Jayanthi*****

ABSTRACT

Big Data is transforming science, engineering, medicine, healthcare, finance, business, and ultimately society itself. The prospective of Big Data is extensively prominent and acknowledged. Many of the organizations are seeking fortunes with Big Data today. Big Data has drawn attention from researchers in Information Sciences, policy and decision makers in governments and enterprises. On the other hand, a new scientific paradigm is born as Data Intensive Scientific Discovery (DISD), also known as Big Data problems. This paper reviews of most relevant information regarding Big Data with its salient features, Data Analytics tools, techniques, supporting technologies. The objective of this paper is intended to foster the dissemination of state-of-the-art research in the emerging fields of Big Data and Analytics which motivates improving business intelligence and business trends.

Keywords: Big Data, Data Management, Big Data Analytics

1. INTRODUCTION

The amount of data in our world has been exploding, and analyzing large data sets, called Big Data, becomes a key basis of competition, underpinning new waves of productivity growth, innovation, and consumer surplus. Leaders in every sector will have to grapple with the implications of big data, not just a few data-oriented managers. The increasing volume and detail of information captured by enterprises, the rise of multimedia, social media, and the Internet of Things (IoTs) will fuel exponential growth in data for the foreseeable future.

1.1. Big-Data

Data is the most precious asset of every organization and these data are very frequently generated by people, devices and networks [1]. Aware of these facts, People are looking for turning their data into value; by utilizing those data in an intelligent way in order to improve their business processes, minimize risks, reduce costs and increase business progress in general.

In a research report, META Group, (Gartner), Doug Laney, the analyst, defines data growth challenges and opportunities as being three-dimensional, i.e. increasing Volume (amount of data), Velocity (speed of data in and out), and variety (range of data types and sources) and others have attempted to categorize these changes by describing Big Data [Doug Laney, 2001]. Actually, Big Data refers to the data volumes in the range of exabytes (1018) and beyond. Such volumes exceed the capacity of current on-line storage systems and processing systems. Data, information, and knowledge are being created and collected at a rate that is rapidly approaching the exabyte/year range. But, its creation and aggregation are accelerating and will approach the range of zettabyte/year within a few years.

* MPhil Scholar, PG and Research Dept. of Computer Science, Rajah Serfoji Govt. College, Thanjavur, 613005, India, *Email: aristasm93@gmail.com*

** PhD Research Scholar,Dept of Computer Science, Bharathiyar University, Coimbatore, Tamilnadu, India, *Email: arun.art06@gmail.com*

*** Assistant Professor, PG and Research Dept. of Computer Science, Rajah Serfoji Govt. College, Thanjavur, 613005, Tamilnadu, India, *Email: jgnanamtcy@gmail.com*

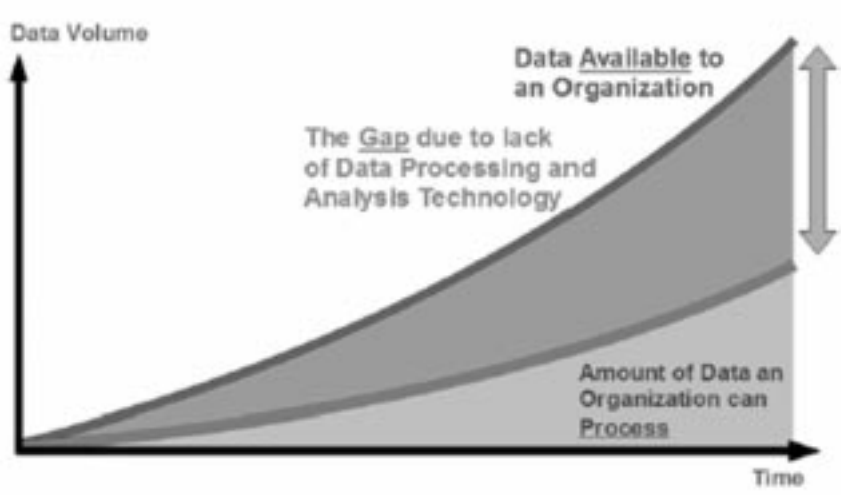

Figure 1: Gap between Big Data availability and Processing in Organizations

Due to high number of big data producers and high frequency of data generation, the gap between data available to organization and data, an organization can process is getting wider all the time. The figure Fig.(1) illustrates the volumes of data available to organizations today, rise rapidly, while the volumes of data that can be processed by an organization is slowl. This creates a gap between the two. In Other words, the percentage (%) of data an organization can analyse is on the decline [3]. The primary value from big data comes not from the data in its raw form, but from the processing and analysis of it and the insights, products, and services that emerge from analysis which in turn is referred to as Big Data Analytics.

1.2. Big Data Analytics

Big-Data Analytics is the process of collecting, organizing and analyzing large sets of data called Big Data which are archived to a data warehouse via Extract-Transform-Load (ETL) routines, in order to discover patterns and other useful information. Big Data Analytics can help organizations to better understand the information contained within the data and will also help in identifying the data that is most important to the business and future business decisions. Analysts working with Big Data are interested to acquire knowledge that comes from analyzing the data.

Big Data Analytics helps organizations to pack their data and use it to identify new opportunities; which in turn, lead to smarter business moves, more efficient operations, higher profits and happier customers. The benefits of Big Data Analytics are realized in terms of (i) Cost Reduction,(ii) Faster, Better Decision Making and (iii) New Products and Services [5].

Further, Big Data can (i) unlock significant value by making information transparent and usable at much greater frequency; (ii) with sophisticated analytics, improve decision-making substantially; (iii) be used to improve the development of the next generation of products and services; (iv) become a key basis of competition and growth for individual firms; and (v) underpin new waves of productivity growth and consumer surplus.

Further, it is said and proved through study cases that more data usually can beat better algorithms. With this statement companies started to realize that they can chose to invest more in processing larger sets of data rather than investing in expensive algorithms. The large quantity of data is better used as a whole because of the possible correlations on a larger amount, correlations that can never be found if the data is analyzed on separate sets or on a smaller set. A larger quantify of data gives a better output but also working with it can become a challenge due to processing limitations.

This paper presents brief overview of evolution, data sources & types and the characteristics of big data in section 2 under technology background secton 3 summarized the Technology that enable big data analytics. Further

supporting techniques and tools for big data analytics in concised in section 4. The challenge and issues for further research in Big data analytics are Rightlighted in section 5. Finally the paper is concluded in section 6.

2. TECHNOLOGY BACKGROUND

2.1. Evolution

Big Data has been first introduced in the digital world by Roger Magoulas from O'Reilly media in 2005 in order to define a great amount of data that traditional data management techniques cannot manage and process due to the complexity and size of this data.

Big Data is an affordable technology to quickly store, discover and analyze massive data sets, even in real-time. High speed processing enables to (i) submit queries repeatedly; (ii) improve the queries gradually and (iii) improve results incrementally. These help people, with little training, to be productive with analytics – something absolutely beyond belief in the past. However, Big Data has changed dramatically. The evolution of the Web has redefined:

- The speed at which information flows into these primary online systems
- The number of customers a company must deal with
- The acceptable interval between the time that data first enters a system, and its transformation into information that can be analyzed to make key business decisions
- The kind of data that needs to be handled and tracked

In [5], the author stated that Big Data as a revolution that will transform how we live, work and think and conveyed the following facts:

- The rest was digital. But not long ago the picture looked very different. Though the ideas of the "information revolution" and "digital age" have been around since the 1960s.
- In 1986 around 40 percent of the world's general-purpose computing power took the form of pocket calculators, which represented more processing power than all personal computers at the time.
- In 2007 only about 7 percent of the data was analog (paper, books, photographic prints. By Hilbert's reckoning, more than 300 exabytes of stored data existed in 2007.
- By 2010 the survey's archive teemed with a whopping 140 terabytes of information.
- The number of messages on Twitter grows at around 200 percent a year and by 2012 had exceeded 400 million tweets a day.
- So in 2013 the amount of stored information in the world is estimated to be around 1,200 exabytes, of which less than 2 percent is non-digital.
- Google processes more than 24 petabytes of data per day, a volume that is thousands of times the quantity of all printed material in the U.S. Library of Congress.
- Facebook, a company that didn't exist a decade ago, gets more than 10 million new photos uploaded every hour.
- Meanwhile, the 800 million monthly users of Google's YouTube service upload over an hour of video every second.

The era of big data challenges the way we live and interact with the world. Big data marks the beginning of a major transformation.

2.2. Data Sources and Types

Data originates from a wide variety sources. Radio-Frequency Identification (RFID) tags and Global Positioning System (GPS) receivers are already spread all around us. This produces petabytes of data [6].

The IBM Big Data Flood Infographic shows that 2.7 Zettabytes of data exist in the digital universe. Also according to this study there are 100 Terabytes updated daily through Facebook, and a lot of activity on social networks this leading to an estimate of 35 Zettabytes of data generated annually by 2020. Just to have an idea of the amount of data being generated, one zettabyte (ZB) eguals 1021 bytes, meaning 1012 GB [7].

According to the study of Digital Universe study of International Data Corporation (IDC), the amount of information the world generated and replicated in 2009 was 800 exabytes, and surpassed 1.8 Zettabytes (ZB) in 2011, making an increase by a factor of 9 in just five years. They estimated an expansion rate of 40 percent to 2020 [8].

This large volume of data, in terms of Big Data can be categorized into (i) Open data, (ii) Closed data and (iii) Shared data as detailed below.

Open Data can be accessed, used and shared by anyone; which is usually published on the web; be available in a machine-readable format; have a licence that permits anyone to access, use and share it. Closed Data are the data owned by data owners or people within an organisation to access, for reasons like privacy, commercial sensitivity and security. Shared Data are the data that is shared with specific people and organisations for a specific purpose: to provide services, connect information and contribute to research.

Further, Big Data is classified as (iv) Structured, (v) Unstructured, (vi) Semi-Structured and (vi) Poly-Structured Data from versatile data sources which are physically distributed.

2.3. Characteristics

Big Data can be characterized by 5 V's which are: (i) Volume, (ii) Variety, (iii) Velocity, (iv) Variability and (v) Veracity [9] and 6 C's system which are: (i) Connection (sensor and networks), (ii) Cloud (computing and data on demand), (iii) Cyber (model and memory), (iv) Content/context (meaning and correlation), (v) Community (sharing and collaboration) and (vi) Customization (personalization and value) [10].

3. ENABLING TECHNOLOGIES

3.1. Granular Computing

Granular Computing (GrC) is a general computation theory to (i) construct an efficient computational model for complex applications with huge amounts of data, information and knowledge; (ii) reduce data size into different level of granularity and hence GrC may take up with Big Data [11]. If information hidden in Big Data is lost partially because of data size reduction, GrC does not work well, and we need other solutions [12].

3.2. Cloud Computing

Cloud computing offers companies to get more value from their data enabling blazing-fast analytics at a fraction of previous costs. This, in turn drives companies to acquire and store even more data, creating more need for processing power and driving a virtuous circle [13]. Apart from its flexibility, Cloud Computing addresses one of the challenges relating to transferring and sharing data as data are shared with others. It is right to say that Big Data problems will push the cloud computing to a high level of development [14].

3.3. Storage Technologies

As the data volumes grow, there is a need for efficient and effective storage techniques which are addressed by data compression and storage virtualization through the technologies, such as Solid-State Drive (SSD) and Phase-Change Memory (PCM), Direct-Attached Storage (DAS), Network-Attached Storage (NAS), and Storage Area Network (SAN) [15, 16].

3.4. Biological Computing

Biological computing models are better appropriate for Big Data because they have mechanisms with high-efficiency to organize, access and process data in ways that are more practical for the ranging and nearly infinite inputs we deal with every day. It is true that the future constructed by bio-inspired technologies is so remarkable that a large amount of funds and human resources are poured into related research activities [17, 18].

3.5. Quantum Computing

Quantum computing is to harness and exploit the powerful laws of quantum mechanics to process information. A Quantum System encodes the zero and the one into two distinguishable quantum states. Certain problems can be solved much faster by larger-scale quantum computers more efficiently and faster than traditional ones [19].

3.6. IPv6 and IOTs

IPv6, incontrovertible protocol, a choice for the Internet of things, has a zillion IP(v6) addresses. IETF standardized and adopted two new major protocols in IPv6, for IoTs and they are Low-Power and Lossy Networks (RPL) in Routing Protocol [20] and the Constrained Application Protocol (CoAP) [21]. IPv6 Internet connects all things of a human to monitor even at his work place and results in generating Big-Data.

3.7. Wireless and Sensor Networks

Data communication and networking technologies such as optical/MPLS, IP networks, Passive Optical Networks (PONs), mobile wireless access technologies are embedded with sensing technology resulting as Wireless Sensor Networks (WSN). With the potential benefits of large-scale deployment of low-cost, energy efficient, and multi-service capabilities, WSN generates huge amounts of heterogeneous data for a wide variety of application domains. It is mandatory to ûnd novel approaches towards ensuring data transmission delay and loss guarantees. The ability to discover, store, clean, analyse, and model big sensor data still needs to be developed.

4. SUPPORTING TECHNIQUES AND TOOLS

The techniques for Big Data Analytics includes a number of areas such as data mining, statistics, neural networks, machine learning, social network analysis, pattern recognition, signal processing, optimization methods and visualization approaches which are summarized here.

4.1. Statistics

To collect, organize and interpret, the data statistics techniques are used. To exploit the casual relationship and correlation ship among distinct objectives, efficient approximate algorithm is proposed for large-scale multivariate monotonic regression. It is an approach for estimating functions that are monotonic with respect to input variables. Another trend of data-driven statistical analysis is focused on scale and parallel implementation of statistical algorithms. With the help of statistics numerical descriptions are generated [22]. Statistical learning and Statistical computing are the two hot research sub-fields.

4.2. Optimization Methods

To solve quantitative problems, Optimization methods are applied. Most of the research works are done to scale up large-scale optimization by co-evolutionary algorithms. Real-time optimization is needed in various Big Data application [23].

4.3. Data Mining

Data mining is to extract useful patterns from data such as Classification and Clustering analysis, association rule mining and regression, discriminate analysis. It involves the methods from statistics and machine learning. When

compared to conventional data mining algorithms Big Data mining is a Challenging issue. Most of the extensions usually relies on analyzing a particular amount of samples of Big Data, and vary in how the sample-based results are used to derive a partition for the overall data. To reflect the goodness Genetic algorithms are also applied to clustering as optimization criterion [24].

4.4. Social Network Analysis

Social Network Analysis (SNA) views social relationships in terms of network theory; consists of nodes and ties. Visualization Approaches are used to create diagrams, tables, images and other intuitive display ways to understand data [25].

4.5. Artificial Neural Network

Artificial Neural Network (ANN) is a mature technique and has a wide range of application coverage. Its successful applications can be found in pattern recognition, image analysis, adaptive control, and other areas. Most of the currently employed ANNs for artificial intelligence are based on statistical estimations, classification optimization and control theory [26].

4.6. Deep Learning

Deep Learning, based on neural networking, shows great potential for solving business problems; enables to recognize items of interest in large quantities of unstructured and binary data; deduce relationships without needing specific models or programming instructions [27].

4.7. Visualization

Visualization Approaches are used to create tables, images, diagrams and other intuitive display ways to understand data [28]; for large-scale data visualization, many researchers use feature extraction and geometric modeling to significantly reduce the data size before the actual data rendering [29].

4.8. Column-Oriented Databases

Since the Big Data contains unstructured, Column-oriented databases are introduced allowing for huge data compression and very fast query times. The downside to these databases is that they will generally only allow batch updates, having a much slower update time than traditional models [30].

4.9. Schema-Less Databases or NoSQL Databases

NoSQL databases systems offers key-value storage, focuses on the scalability of data storage with high-performance, provides low-level access mechanism and very flexible for data modeling, and easy to update application developments and deployments [30].

4.10. Tools

Big Data tools can be classified into three major categories [31] as (i) Tools for batch processing (ii) Tools for stream processing and (iii) Tools for Interactive analysis processing.

There are several tools available for big data using batch processing methods. Some of the popular tools are (i) Apache Hadoop, (ii) map/reduce, (iii) Dryad, (iv) Karmasphere Studio and Analyst, (v) Jasper soft BI, (vi) Sky tree Server, (vii) Pentaho Business Analytics, (viii) Apache Mahout, (ix) Tableau and (x) Talend Open Studio.

Big Data tools based on stream processing have been developed and some are under developing. One of the most famous platforms is Storm and others include S4, SQLstream, Splunk, Apache Kafka, and SAP Hana.

Interactive analysis processing tools are available through Open source. Such tools are Google's Dremel, Hive, PIG, WibiData and PLATFORA.

The following table summarizes tools with their specific use and their advantage.

Table 1
Big Data Tools based on Batch Processing

Tool Name	*Specified Use*	*Advantage*
Apache Hadoop	Infrastructure and platform	High scalability, reliability, completeness
Drad	Infrastructure and platform	High performance distributed execution engine, good programmability
Apache Mahout	Machine learning algorithms in business	Good maturity
Jaspersoft BI Suite	Business intelligence software	Cost-effective, self-service BI at scale
Pentaho Business Analytics	Business analytics platform	Roboustness, scalability, flexibility in knowledge discovery
Skytree Server	Machine learning and advanced analytics	Process massive datasets accurately at high speed
Tableau	Data Visualization, Business analytics	Faster, smart, fit, beautiful and ease of use dashboards
Karmasphere studio and analyst	Big data workspace	Collaborative and standards-based unconstrained analytics and self service
Talend open studio	Data management and application integration	Easy –to-use,edipse-based graphical environment

Table 2
Tools based on Stream Processing

Tool Name	*Specified Use*	*Advantage*
Storm	Realtime computation system	Scalable, fault-tolerant, and is easy to set up and operate
S4	Processing continuous unbounded streams of Data	Proven, distributed, scalable, fault-tolerant, pluggable platform
SQLstream s-Server	Sensor, M2M, and telematics applications	SQL-based, real-time streaming Big Data platform
Splunk	Collect and harness machine data	Fast and easy to use, dynamic environments, scales from laptop to Data center
Apache Kafka	Distributed publish-subscribe messaging system	High-throughput stream of immutable activity data
SAP Hana	Platform for real-time business	Fast in-memory computing and real time analytic

5. OBVSERVATIONS AND RESEARCH DIRECTIONS

In our literature, we infer that (i) Digital data processing capacity have increased exponentially in recent years on the average of 10 to 17 billion data points processed per day, (ii) Data is doubling in size every two years and by 2020 will multiply ten-fold from 4.4 trillion gigabytes to 44 trillion gigabytes.

In the literature, With its ever-increasing 5 V's and 6 C's of structured and unstructured data, Big Data require a set of techniques and technologies with new forms of integration to reveal insights from datasets which is otherwise is called Big-Data Analytics.

In Big-Data Analytics, a large sets of data are collected, organized, processed, analyzed in order to (i) Uncover the hidden patterns, unknown correlations, market trends, customer preferences; (ii) Identify the data that is most important to the business and future business decisions; (iii) Unlock significant value by making information transparent; (iv) improve the development of the next generation of products and services (v) Create innovative after-sales service offerings.

This literature also explored the facts such as (i) Realities for commercial use of big data is a key basis of competition and growth for individual firms, (ii) Datasets and analysis tools are now available, and there is a necessity to use, (iii) In practice, commercial use of big data requires strong analytic skills, understanding of scientific method, business expertise, exceptional programming abilities and an ability to adapt to a constantly changing, challenging environment.

Henceforth, we observe that all Organizations need to take big data seriously. Organizations need to structure workflows and incentives to optimize the use of big data. Access to data is critical—companies will increasingly need to integrate information from multiple data sources, often from third parties, and the incentives have to be in place to enable this. There will be a shortage of talent necessary for organizations to take advantage of big data and therefore, Organizations need to train people to go with Big-Data Analytics, to make effective decisions.

Business users and data analysts need to gain insight from a growing variety of data sources. Besides, several challenges will have to be addressed to experience the full potential of Big Data. These challenges in big data can be broadly divided in to two categories: (i) Engineering data (data management activities) and (ii) Semantic data (extract the meaning of the information). Some of the most essential major challenges in big data are presented here.

- Data and information origin will become a critical issue. There is no universally accepted way to store raw data, reduced data, code and parameter choices that produced the data.
- Most data is time-varying, presenting an issue of duplication of data. Time-to-information is critical when one considers Real-Time Processes.
- An effective processing on the huge volume of data will require extensive parallel processing and new analytics algorithms in order to provide timely and actionable information.
- An emerging challenge for Big Data users is “quantity vs. quality”. To access very large quantities of semi- or unstructured data, and how to utilize as yet unknown tool designs is not known. The following queries are to be addressed

✓ How do we decide which data is irrelevant versus selecting the most relevant data?

✓ How do we ensure that all data of a given type is reliable and accurate? Or, maybe just approximately accurate?

✓ How much data is enough to make an estimate or prediction of the specific probability and accuracy of a given event?

✓ How do we assess the “value” of data in decision making? Is more necessarily better?

✓ How can businesses mine the precious data they are accumulating?

- Another major challenge is data dissemination. The bottleneck is the communications middleware.
- Data source has no beginning and no end. The data streams may dynamically change and in reality, do not well-behave. There is a need to rethink data stream processing.
- Data ownership presents a critical and ongoing challenge, particularly in the social media networks unverified and introduces non-relevant and non-accuracy of data.

- Some big data, need Policies related to privacy, security, intellectual property and even liability will need to be addressed in a Big Data world.

✓ How can IT and governance teams ensure that sensitive data is protected, whilst providing non-expert users with the data access they need?

✓ Where should data reside - on-premise or in the Cloud?

✓ How do enterprises comply with international data protection regulation whilst taking advantage of the revolution in Cloud, social and mobile data?

6. CONCLUSION

To conclude, (i) Big Data is a buzzword in business progress and social science frontier; (ii) the level of acquiring information and knowledge haul from the Universe is ever growing (ii) Big Data Analytics is in its way of promising progress to deliver good predictions but still in the initial stage of development. There is no perfect big data management solution yet. This is identified as an important gap in the research literature of Big Data and needs to be filled. Big Data also means big systems, big challenges and big profits, so more research works in these sub-fields are necessary to resolve it.

REFERENCES

[1] R. Chaudhary, J. R. Pandey, P. Pandey and P. Chaudhary, "Case study of Internet of Things in area of Agriculture, AGCO's Fuse Technology's Connected Farm Services," *International Conference on Green Computing and Internet of Things, Noida*, 148-153, 2015.

[2] Elena Geanina ULARU, Florina Camelia PUICAN, Anca APOSTU, Manole VELICANU, "Perspectives on Big Data and Big Data Analytics", *Database Systems Journal*, **3(4)** , 2012.

[3] Doug Laney, "Application Delivery Strategies", META Group, File 949, Addendum, Feb-2001.

[4] Paul Zikopoulos, IBM Big Data: What is Big Data Part 1 & 2, ftp://public.dhe.ibm.com/software/pdf/at/SWP10/Harness_the_Power_of_Big_Data.pdf

[5] Thomas H. Davenport and SAS Institute Inc., "Big Data in Big Companies", Whitepaper, International Institute for Analytics, May 2013.

[6] http://www.nytimes.com/2013/06/11/books/big-data-by-viktor-mayer-schonberger-and-kenneth-cukier.html?_r=0

[7] Joseph M Hellerstein. Programming a Parallel Future. Technical Report UCB/EECS-2008-144, University of California, Berkeley, November 2008. http://www.eecs.berkeley.edu/Pubs/TechRpts/2008/EECS-2008-144.html

[8] G. Noseworthy, Infographic: Managing the Big Flood of Big Data in Digital Marketing, 2012. http://analyzingmedia.com/2012/infographic-big-flood-of-big-data-in-digital-marketing/

[9] Thibaud Chardonnens, "Big Data analytics on high velocity streams: specific use cases with Storm", Software Engineering Group, Department of Informatics, University of Fribourg, Switzerland, 2013.

[10] Lee, Jay; Lapira, Edzel; Kao, Hung-an. "Recent advances and trends in predictive manufacturing systems in big data environment", *Manufacturing Letters* **1(1)**, 38–41, 2013.

[11] Lee, Jay; Bagheri, Behrad; "Recent Advances and Trends of Cyber-Physical Systems and Big Data Analytics in Industrial Informatics". *IEEE Int. Conference on Industrial Informatics*, 2014.

[12] Witold Pedrycz, Andrzej, Vladik Kreinovich, Handbook of Granular Computing, WILEY, 2008.

[13] Michael A. Nielsen, Isaac L. Chuang, Quantum Computation and Quantum Information, Cambridge University Press, 2009.

[14] A. u. R. Khan, M. Othman, S. A. Madani and S. U. Khan, "A Survey of Mobile Cloud Computing Application Models", *IEEE Communications Surveys & Tutorials*, **16(1)** , 393-413, First Quarter 2014.

[15] Eric E. Schadt, Michael D. Linderman, Garry P. Nolan, Computational solutions to large-scale data management and analysis, *Nat. Rev. Genet*. **11(9)** , 647–657, 2010.

[16] Lee Hutchinson, Solid-state revolution: in-depth on how ssds really work, Ars Technica (2012).

[17] A. Pirovano, A.L. Lacaita, A. Benvenuti, F. Pellizzer, S. Hudgens, R. Bez, Scaling analysis of phase-change memory technology, IEEE Int. Electron Dev.Meeting (2003) 29.6.1–29.6.4.

[18] Josh, Bongard, "Biologically inspired computing", *Journal of Computer* **42(4)** ,95–98, 2009.

[19] Nakano, Biological computing based on living cells and cell communication, in: 2010 13th International Conference on Network-Based Information Systems (NBiS),42–47, 2010.

[20] Daniel R. Simon, On the power of quantum computation, **SIAM Journal on Comput**. **26**,116–123, 1994.

[21] T. Winter and P. Thubert, "RPL: IPv6 Routing Protocol for Low-Power and Lossy Networks," in RFC 6550, 2012.

[22] Z. Shelby, K. Hartke, and C. Bormann, "Constrained Application Protocol (CoAP)", In Internet Draft: Intended status: Standards Track, 2013.

[23] Philippe Pébay, David Thompson," Design and performance of a scalable, parallel statistics toolkit", *IEEE International Symposium on Parallel and Distributed Processing*, 1475–1484, 2011.

[24] Vikas C. Raykar, Ramani Duraiswami, "A fast algorithm for learning a ranking function from large-scale data sets", *IEEE Trans. Pattern Anal. Mach. Intell.* **30(7)** , 1158–1170, 2008.

[25] X. Wu, X. Zhu, G. Q. Wu and W. Ding, "Data mining with big data," *in IEEE Transactions on Knowledge and Data Engineering*, **26(1)** , 97-107, Jan. 2014.

[26] W. Tan, M. B. Blake, I. Saleh and S. Dustdar, "Social-Network-Sourced Big Data Analytics" , *IEEE Internet Computing*, **17(5)** , 62-69, Sept.-Oct. 2013.

[27] Yan-Jun Liu, Chen, Wen, "Adaptive neural output feedback tracking control for a class of uncertain discrete-time nonlinear systems", *IEEE Trans. on Neural Networks*, **22(7)**, 1162–1167, 2011.

[28] X. W. Chen and X. Lin, "Big Data Deep Learning: Challenges and Perspectives," *in IEEE Access*, **(2)**, 514-525, 2014.

[29] Simeon Simoff, Michael H. Böhlen, Arturas. Mazeika, "Visual Data Mining: Theory, Techniques and Tools for Visual Analytics", Springer, 2008.

[30] David Thompson, Joshua A. Levine, Philippe P. Pébay, "Analysis of large-scale scalar data using hixels", *IEEE Symposium on Large Data Analysis and Visualization*, 23–30, 2011.

[31] Jing Han, Haihong E, Guan Le, Jian Du," Survey on nosql database", *6th International Conference on Pervasive Computing and Applications*, 363–366, 2011.

[32] Han Hu,Yonggang Wen, Xuelong Li, "Toward Scalable Systems for Big Data Analytics: A Technology Tutorial", *IEEE Practical Innovations : Open Solutions*, (2), 2014.