BIG DATA AND FIVE V'S CHARACTERISTICS

[1]HIBA JASIM HADI, [2]AMMAR HAMEED SHNAIN, [3]SARAH HADISHAHEED, [4]AZIZAHBT HAJI AHMAD

[1]Ministry of Education, Islamic University College, Third Author Affiliation
E-mail: nassirfarhan@yahoo.com, [s802371, s802370, s93456]@student.uum.edu.my

Abstract- Big Datais used to refer tovery large data sets having large, more varied and complex structure with the difficulties of storing, analyzing and visualizing for further processes or results. The research into large amounts of data in order to reveal hidden patterns and secret correlations named as Big Data analytics. This isuseful information for companies or organizations tohelp gain richer and deeper insights and achieving an advantage over the competition. Due to this, Big Data implementation must be accurately carried out. Therefore, Big Datais an importanttechnology trend,and it has the potential for dramatically changing the way organizations use information to enhance the customer experience and transform their business models. As well as, Big Data is often about doing things that weren't widely possible because the technology was not advanced enough or the cost of doing so was prohibitive. This paper presents an overview of Big Data's content, types, architecture, technologies, and characteristics of Big Datasuch as Volume, Velocity, Variety, Value, and Veracity.

Keywords- Big Data, Healthcare, Architecture, Big Data technologies, Structure data

I. INTRODUCTION

The term "Big Data" was first introduced to the computing 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. Madden define the Big Data as: "data that's too big, too fast, or too hard for existing tools to process."

"Too big" means that organizations must increasingly deal with petabyte-scale collections of data that come from click streams, transaction histories, sensors, and elsewhere. "Too fast" means that not only is the data large, but must be processed quickly, such as carrying out fraud detection or to find an ad to display. "Too hard",is a phrase which means that such data may not be easily processed by existing tools, or that needs some more analysis not suited to existing tools Big Datadoes not refer to a single market. Rather, the term is used to refer to data management technologies which have evolved over time. Big Dataallows interested parties to store, manage, and analyze large amounts of data at both the proper speed and time to gain real insights.

The key to understanding Big Data is that data must be used in such a way that it actually supports real-life profitable or beneficial outcomes. Most have just begun exploiting Big Data. Many companies have beenexperimenting with techniques that allow them to collect massive amounts of data in order to determine whether hidden patterns exist within that data that might be an early indication of an important change. Data mightshow, for example, that customer buying patterns are changing or that new factors affecting the business must be considered. A study on the Evolution of Big Data as a Research and Scientific Topic shows that the term "Big Data" was present in research beginning in the 1970s. Nowadays, the Big Data concept is addressed from various angles, demonstrating its importance. Big Data is important from many perspectives.

The significant amount of data generated allows stakeholdersto make decisions in a timely manner where money can be saved and operations maybe betteroptimized in both public and private sector. For example, in the retail business, consumer behavior and preferences maybe understand throughanalysis ofBig Data,such ascustomer movement in a store, navigation of a website, product searches, and so on. Also, the United States Healthcare Big Data World, which comprises records of over 50 million patients, uses data-driven concept to determine challenges in the healthcare sector. Aslavsky, Perera and Georgakopoulos have statedthat the queries used to process Big Data are very complex. Figure 1 below showsthe amount of new data stored across geographical regions:

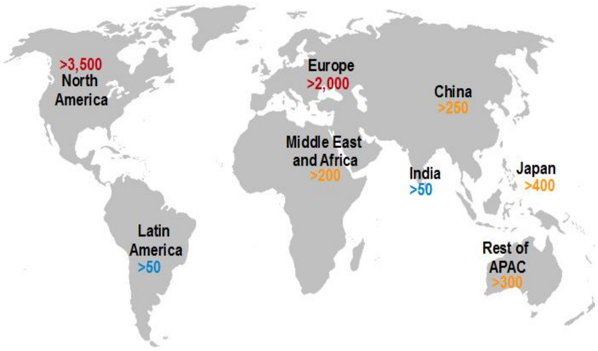

Figure 1: Amount of new data stored varies across geography (Source)

Big Data has the potential to generate more revenue, while reducing risk and predicting future outcomes with greater confidence atlow cost. The figure below depicts the Big Data managementcycle:

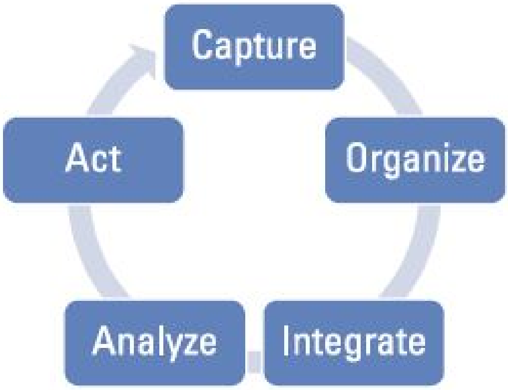

Figure 2: Big Data Management

Big Data Management is organized aroundfinding and organizing relevant data. Per the figure below:

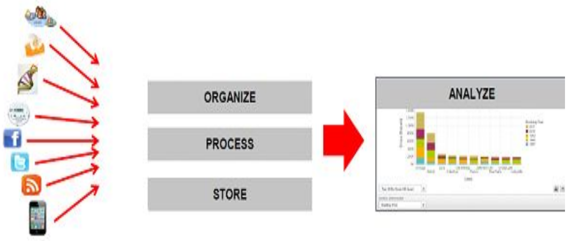

Figure 3: Big Data Management

II. BIG DATA TYPES

Big Data encompasses everything, from dollar transactions to tweets to images to audio. Therefore, taking advantage of Big Data requires that all this information to be integrated for analysis and data management. This is more difficult than it appears. There are two main types of data concerned here: structured and unstructured. Structured data is like a data warehouse, in which data is tagged and sortable, while unstructured data is random and difficult to analyze. TheFigure below depicts these types,along with examples:

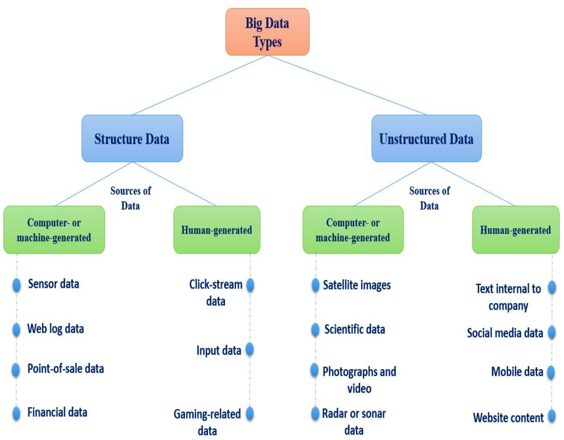

Figure 4: Big Data Types

III. THE ARCHITECTURE FOR BIG DATA

The figure below depicts the Big Data architecture:

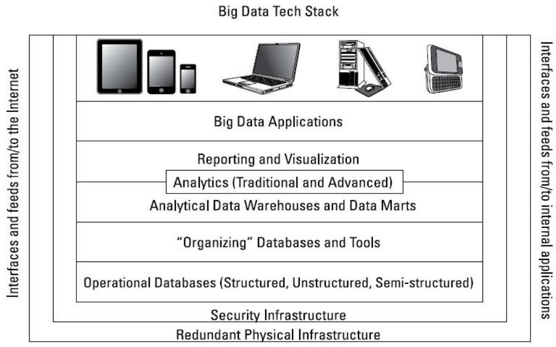

Figure 5: Big Data architecture (Source:)

A. Interfaces and feeds

Before we get into the nitty-gritty of the Big Data technology stack itself, must we understand how Big Data works in the real world, therefore, it is important to start by understanding this necessity. In fact, what makes Big Data big is the fact that it relies on picking up lots of data from lots of sources. Therefore, open application programming interfaces (APIs) are a core part ofany Big Data architecture.

In addition, interfaces exist at every level and between every layer of the stack. Without integration services, Big Datacannot happen. Other important operational database approaches include columnar databases that store information efficiently in columns,and notrows. This approach leads to faster performance,asinput/output is extremely fast. When geographic data storage is part of the equation, a spatial database is optimized to store and query data based on how objects are related in real terms.

B. Redundant physical infrastructure

The supporting physical infrastructure is fundamental to the operation and scalability of a Big Data architecture. In fact, without the availability of robust physical infrastructures, Big Data would likely not have become such a strong trend. To support an unanticipated or unpredictable volume of data, a physical infrastructure for Big Data has to be different than that for traditional data. The physical infrastructure has beenbased on a distributed computing model. This means that data may be physically stored in many different locations, allowing it to be linked through networks, the use of a distributed file system, and various Big Data analytic tools and applications.

Redundancy is important, as companies must handlea great deal of data from many sources. Redundancy comes in many forms. For instance, if the company has created a private cloud, company may want create redundancy within private areas so that it can scale out to support changing workloads. If a company needs to limit internal IT growth, it may use external cloud services to add to its own resources. In some cases, this redundancy may come in the form of a Software as a Service (SaaS), allowing companies to carry out advanced data analysis as a service. The SaaS approach allows for afaster start, lowering costs.

C. Security infrastructure

AsBig Data analysis becomes part of workflow, it becomes vital to secure that data. For example, a healthcare companyprobably wants to use Big Data applications to determine changes in demographics or shifts in patient needs. This data about patients needs to be protected, both to meet compliance requirements and to protect patient privacy. The company needs toconsider who is allowed to see the data and when they may see it. Also, the company need to be able to verify the identity of users, as well as protect the identity of patients. These types of security requirements must be part of the Big Data fabric from the outset, and not an afterthought.

D. Operational data sources

ConcerningBig Data, a company mustensure that all sources of data will provide a better viewpoint about the business and allow it to understand how data effects the operational methods of that company. Traditionally, an operational data source consisted of highly structured data, managed by the line of business in a relational database. However, operational data now has to considera broader set of data sources, including unstructured sources suchas social media or customer data.

E. Performance matters

Data architecture also must work to perform in concert with supporting infrastructure of organization or company. For instance, the company might be interested in running models to determine whether it is safe to drill for oil in an offshore area,provided with real-time data of temperature, salinity, sediment resuspension, and many other biological, chemical, and physical properties of the water column. It might take days to run this model using a traditional server configuration. However, using a distributed computing model, a days' long task may take minutes. Performance might also determine the kind of database that company would use. Under certain circumstances, stakeholders may want to understand how two very distinct data elements are related, or the relationship between social network activity and growth in sales. This is not the typical query the company could ask of a structured, relational database. A graphing database might be a better choice, as it may betailored to separate the "nodes" or entities from its "properties" or the information that defines that entity, and the "edge" or relationship between nodes and properties. Using the right database mayalso improve performance. Typically,agraph database maybe used in scientific and technical applications.

F. Organizing data services and tools

Indeed, not all the data that organizations use is operational. A growing amount of data comes from a numberof sources that are notquite as organized or straightforward, including data that comes from machines or sensors, and massive public and private data sources. In the past, most companies are notable to either capture or store this vast amount of data. It was simply too expensive or too overwhelming. Even if companies areable to capture the data, they donot have the tools to do anything about it. Very few tools canmake sense of these vast amounts of data. The tools that did exist were complex to use and did not produce results within a reasonable time frame. In the end, companies who really wanted to go to the enormous effort of analyzing this data were forced to work with snapshots of data. This means that stakeholders may miss out on relevant events as they may not have been captured in a certain snapshot.

G. Analytical data warehouses and data marts

After a company sorts through the massive amounts of data available, it is often importantto take the subset of data that reveals patterns and put it into a form that's available to the business. Such so-called'warehouses' provide compression, multilevel partitioning, and a massively parallel processing architecture.

H. Reporting and visualization

Companies have always relied on the capability to create reports to give them an understanding of what the data tells them about everything from monthly sales figures to projections of growth. Big Data changes the way that data is managed and used. If a company is able tocollect, manage, and analyze enough data, it mayuse a new generation of tools to help management truly understand the impact not just of a collection of data elements, but also how these data elements offer context based on the business problem being addressed. With Big Data, reporting and data visualization have become tools for looking at the context of how data is related and the impact of those relationships on the future.

I. Big Data Applications

Traditionally, business has anticipated that data would be used to answer questions about what to do and when to do it. Data hasoften been integrated into general-purpose business applications.

With the advent of Big Data, this is changing. Now, the companies are seeing the development of applications that are designed specifically to take advantage of the unique characteristics of Big Data. Specific emerging applications include areas such as healthcare, manufacturing management, and traffic management. All of these applications rely on huge volumes, velocities, and varieties of data to transform the behavior of a market. For example, in healthcare, a Big Data application might be able to monitor premature infants to determine ifdata indicates when intervention is needed. In manufacturing, a Big Data application can be used to prevent a machine from shutting down during a production run. A Big Data traffic management application mayreduce the number of traffic jams on busy city highways, decreasing the number of accidents while saving fuel and reducing pollution.

IV. BIG DATA TECHNOLOGIES

With the evolution of computing technology, businesses may now manage immense volumes of data previously could dealt with using expensive supercomputers. These are now much cheaper. As a result, new techniques for distributed computing are mainstream. Big Databecame paramount as companies like Yahoo!, Google, and Facebook came to the realization that they needed help in monetizing the massive amounts of data their offerings were creating. Thus, these newcompanies mustsearch fornew technologies to store, access, and analyze huge amounts of data in near real time. Such real-time analysis is required in order to profit from so much data from users. Their resulting solutions haveaffected the larger data management market. In particular, the innovations MapReduce, Hadoop, and Big Table have provenlead to a new generation of data management.These technologies will allow businesses to address one of the most fundamental problems, namely the capability to process massive amounts of data efficiently, cost-effectively, and quickly.

J. MapReduce

MapReducewasdesigned by Google to efficiently carry outa set of functions against a large amount of data in batch mode. The “map” component distributes the programming problem or task across a large number of systems while managing placement to balance the load and allow recovery from failures. After the distributed computation is complete, another function called “reduce” aggregates all the elements back together to provide a result. An example of MapReduce would be determining the number of pages in a book that are written in each of 50 different languages.

K. Big Table

Big Table was developed by Google to be a distributed storage system to manage highly scalable structured data. Data is organized into tables with rows and columns. Unlike typical relational database models, Big Table is a sparse, distributed, persistent multidimensional sorted map. It has been designed to keep large volumes of data across commodity servers.

L. Hadoop

Hadoop is an Apache-managed software framework created usingMapReduce and Big Table. Hadoop allows applications based on MapReduce to run on large clusters of commodity hardware. The project has become the basis for the computing architecture underlying Yahoo!’s business. Hadoop is designed to parallelize data processing across computing nodes to speed computations and diminishlatency. Two major components of Hadoop exist: a massively scalable distributed file system that can support petabytes of data, and a massively scalable MapReduce engine that computes results in batches.

V. BIG DATA AND HEALTHCARE

Healthcare is one an important area of investment. It generates more data than most industries. Therefore, healthcare is likely to greatly benefit by new forms of Big Data. Healthcare providers, insurers, researchers, and healthcare practitioners often make decisions about treatment options with data that is incomplete or not relevant to specific illnesses. This is because gathering relevant data from patients is difficult. Data elements are often stored and managed in different places by different organizations. In addition, clinical research conducted all over the world may be helpful in determining how a specific disease or illness might be approached. Big Data can help change this problem. Also, healthcare companies strive to use Big Data applications to determine changes in demographics or shifts in patient needs. In healthcare, a Big Data application might be able to monitor premature infants to determine when data indicates whetherintervention is needed. The Table below describesanumber of research papers about Big Datainvariousdisciplines:

Table 1: Big Data with different disciplines

Author (s)	Year	discipline	Title
Bertot and Choi	2013	Electronic government	Big Data and e-Government: issues, policies, and recommendation s
Bertot, Gorham, Jaeger, Sarin and Choi	2014	Electronic government	Big Data, open government and e-Government: Issues, policies and recommendation s
Rajagopal an and Vellaipand iyan	2013	Electronic government	Big Data framework for national E-governance plan
Mahmood, Rashad and El-Dosuky	2014	Communicat ion	Efficient Implementation of Big Data Switch for Iraqi Cellular Phone Service Providers

Hansen, Miron-Shatz, Lau and Paton	2014	Healthcare	Big Data in Science and Healthcare: A Review of Recent Literature and Perspectives
Liu and Park	2014	Healthcare	Big Data as an e-Health Service
Raghupathi and Raghupathi	2014	Healthcare	Big Data analytics in healthcare: promise and potential
Kim, Trimi and Chung	2014	Public Sector	Big-data applications in the government sector
Fasel	2014	Public Sector	Potentials of Big Data for governmental services
Yu, Jiang and Zhu	2013	Public Sector	RTIC-C: A Big Data System for Massive Traffic Information Mining
Kim, Oh, Lee and Jung	2013	Electronic Commerce	Discovery of Travel Patterns in Seoul Metropolitan Subway Using Big Data of Smart Card Transaction Systems
Smith, Szongott, Henne and von Voigt	2012	Social Media	Big Data privacy issues in public social media

Also, McKinsey Global Institute has specified the potential of Big Data in five main topics:

- Healthcare: clinical decision support systems, individual analytics applied for patient profile, personalized medicine, performance based pricing for personnel, analyze disease patterns, improve public health.

- Public sector: creating transparency by accessible related data, discover needs, improve performance, customize actions for suitable products and services, decision making with automated systems to decrease risks, innovating new products and services.

- Retail: in store behavior analysis, variety and price optimization, product placement design, improve performance, labor inputs optimization, distribution and logistics optimization, and web-based markets.
- Manufacturing: improved demand forecasting, supply chain planning, sales support, developed production operations, websearch-based applications.
- Personal location data: smart routing, geo targeted advertising or emergency response, urban planning, and new business models.

In addition, web provides kind of opportunities for Big Data too. For example, social network analysis may include understanding user intelligence for more targeted advertising, marketing campaigns, and capacity planning, customer behavior and buying patterns, includingsentiment analytics.

VI. FIVE V'S BIG DATA CHARACTERISTICS

Big Data is important because it enables organizations to gather, store, manage, and manipulate vast amounts data at the right speed, at the right time, to gain the right insights. In addition, Big Data generators must create scalable data (Volume) of different types (Variety) under controllable generation rates (Velocity), while maintainingthe important characteristics of the raw data (Veracity), the collected data can bring to the intended process, activity or predictive analysis/hypothesis. Indeed, there is no clear definition for 'Big Data'. It has beendefined based on some of its characteristics. Therefore, these five characteristics have been used to define Big Data, also known as 4V's (Volume, Variety, Velocity and Veracity), as illustrated in Figure 6 below:

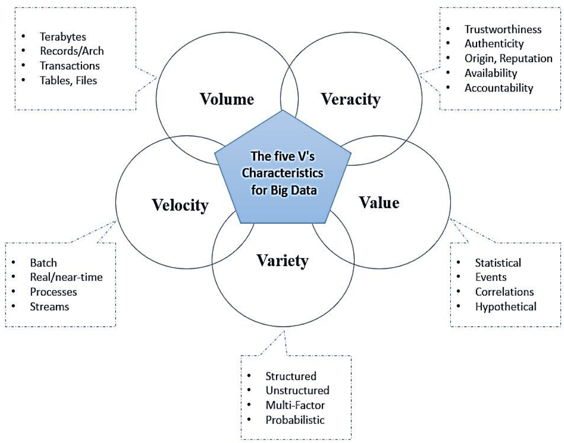

Figure 6: FiveVsBig Data Characteristics

Volume: refers to the quantity of data gathered by a company. This data must be used further to gainimportant knowledge. Enterprises are awash with ever-growing data of all types, easily amassing terabytes even petabytes of information(e.g. turning 12 terabytes of Tweets per day into improved product sentiment analysis; or converting 350 billion annual meter readings to better predict power consumption).

Moreover, Demchenko, Grosso, de Laat and Membrey statedthat volume is the most important and distinctive feature of Big Data, imposing specific requirements to all traditional technologies and tools currently used.

Velocity: refers to the time in which Big Data can be processed. Some activities are very important and need immediate responses, which is why fast processing maximizes efficiency. For time-sensitive processes such fraud detection, Big Data flows must be analyzed and used as they stream into the organizations in order to maximize the value of the information (e.g. scrutinize 5 million trade events created each day to identify potential fraud; analyze 500 million daily call detail records in real-time to predict customer churn faster).

Variety:refers to the type of data that Big Data can comprise. This data maybe structured or unstructured. Big data consists in any type of data, including structured and unstructured data such as text, sensor data, audio, video, click streams, log files and so on. The analysis of combined data types brings new problems, situations, and so on, such as monitoringhundreds of live video feeds from surveillance cameras to target points of interest, exploiting the 80% data growth in images, video and documents to improve customer satisfaction));

Value:refers to the important feature of the data which is defined by the added-value that the collected data can bring to the intended process, activity or predictive analysis/hypothesis. Data value will depend on the events or processes they represent such as stochastic, probabilistic, regular or random. Depending on this the requirements may be imposed to collect all data, store for longer period (for some possible event of interest), etc. In this respect data value is closely related to the data volume and variety.

Veracity: refers to the degree in which a leader trusts information in order to make adecision. Therefore, finding the right correlations in Big Data is very important for the business future. However, asone inthree business leaders do nottrust the information used to reach decisions, generating trust in Big Data presents a huge challenge as the number and type of sources grows.

VII. CHALLENGES IN BIG DATA MANAGEMENT

The challenges in Big Data can be broadly divided in to two categories: engineering and semantic. Engineering challenges include data management activities such as query, and storage efficiently. The semantic challenge is determining the meaning of information from large volumes of unstructured data. Several other challenges in Big Data are presented below: Jones et al., found a number of major challenges toBig Data management.

- High volume of processing using a low-power consumed digital processing architecture.
- Discovery of data-adaptive machine learning techniques that are able toanalyze data in real-time.
- Design scalable data storages that provide efficient data mining.

On the other hand, Patidar, Rane and Jain have identified a number of key challenges in Big Data management related to the cloud as follows:

- Data security and privacy;
- Approximate results;
- Data exploration to enable deep analytics;
- Enterprise data enrichment with web and social media;
- Query optimization; and
- Performance isolation for multi-tenancy.

CONCLUSION

Big Data is new and requires investigation and understanding of both technical and business requirements. Indeed, Big Data is not a stand-alone technology; rather, it is a combination of the last 50 years of technological evolution. The big advantage ofBig Data is its ability to leverage massive amounts of data without all the complex programming that was required in the past.

On the other hand, based on previous work, Big Data initiatives have shown significant promise for policy and decision-making, as well as fostering collaboration between governments and citizens and businesses, and for ushering in a new era of digital government services. Unfortunately, there have beenfew studies that concentrate on Big Datain terms ofe-Government.

Therefore, there is a need for future research to understanding the critical issues for using Big Data with e-Government. In addition, there is substantial requirement that aBig Data governance model better address the policies and practices surrounding Big Data.

REFERENCES

[1] S. Madden, "From databases to Big Data," IEEE Internet Computing, vol. 16, pp. 0004-6, 2012.

[2] D. Che, M. Safran, and Z. Peng, "From Big Data to Big Data Mining: Challenges, Issues, and Opportunities," in Database Systems for Advanced Applications, 2013, pp. 1-15.

[3] O. Kwon and J. M. Sim, "Effects of data set features on the performances of classification algorithms," Expert Systems with Applications, vol. 40, pp. 1847-1857, 2013.

[4] M. D. K. Jadhav, "Big Data: The New Challenges in Data Mining," International Journal of Innovative Research in Computer Science & Technology, vol. 1, pp. 39-42, 2013.

[5] A. Nugent, F. Halper, and M. Kaufman, Big DataFor Dummies: John Wiley & Sons, 2013.

[6] K. Michael and K. W. Miller, "Big Data: New opportunities and new challenges [guest editors' introduction]," Computer, vol. 46, pp. 22-24, 2013.

[7] P. Helland, "Condos and clouds," Communications of the ACM, vol. 56, pp. 50-59, 2013.

[8] A.-R. Bologa, R. Bologa, and A. Florea, "Big Data and Specific Analysis Methods for Insurance Fraud Detection," Database Systems Journal, vol. 4, pp. 30-39, 2013.

[9] G. Halevi and H. Moed, "The evolution of Big Data as a research and scientific topic: overview of the literature," Research Trends, Special Issue on Big Data, vol. 30, pp. 3-6, 2012.

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

[11] S. Sagiroglu and D. Sinanc, "Big Data: A review," in Collaboration Technologies and Systems (CTS), 2013 International Conference on, 2013, pp. 42-47

[12] C. Bizer, P. Boncz, M. L. Brodie, and O. Erling, "The meaningful use of Big Data: four perspectives--four challenges," ACM SIGMOD Record, vol. 40, pp. 56-60, 2012.

[13] A. Zaslavsky, C. Perera, and D. Georgakopoulos, "Sensing as a service and Big Data," arXiv preprint arXiv:1301.0159, 2013.

[14] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, "Big Data: The next frontier for innovation, competition," and productivity. Technical report, McKinsey Global Institute2011.

[15] B. Yunan. (2012). What is Big Data's role in helping companies achieve competitive advantage through analytics.

[16] E. G. Ularu, F. C. Puican, A. Apostu, and M. Velicanu, "Perspectives on Big Data and Big Data Analytics," Database Systems Journal, vol. 3, pp. 3-14, 2012.

[17] T. H. Davenport, P. Barth, and R. Bean, "How ‘Big Data’is different," MIT Sloan Management Review, vol. 54, 2013.

[18] M. K. Dagli and B. B. Mehta, "Big Data and Hadoop: A Review," 2014.

[19] R. L. Villars, C. W. Olofson, and M. Eastwood, "Big Data: What it is and why you should care," White Paper, IDC, 2011.

[20] O. Tene and J. Polonetsky, "Big Data for all: Privacy and user control in the age of analytics," 2013.

[21] R. Jain, "Big Data Fundamentals," 2013.

[22] W. Xu, "Research on Traffic Management-Oriented “Big Data” and its Application," Applied Mechanics and Materials, vol. 427, pp. 2743-2747, 2013.

[23] R. Cumbley and P. Church, "Is “Big Data” creepy?," Computer Law & Security Review, vol. 29, pp. 601-609, 2013.

[24] V. Mayer-Schönberger and K. Cukier, Big Data: A revolution that will transform how we live, work, and think: Houghton Mifflin Harcourt, 2013.

[25] H. G. Miller and P. Mork, "From data to decisions: a value chain for Big Data," IT Professional, vol. 15, pp. 57-59, 2013.

[26] S. Humbetov, "Data-intensive computing with map-reduce and hadoop," in Application of Information and Communication Technologies (AICT), 2012 6th International Conference on, 2012, pp. 1-5.

[27] C. J. Tauro, S. Aravindh, and A. Shreeharsha, "Comparative study of the new generation, agile, scalable, high performance NOSQL databases," International Journal of Computer Applications, vol. 48, pp. 1-4, 2012.

[28] T. B. Murdoch and A. S. Detsky, "The inevitable application of Big Data to health care," JAMA, vol. 309, pp. 1351-1352, 2013.

[29] J. C. Bertot and H. Choi, "Big Data and e-Government: issues, policies, and recommendations," in Proceedings of the 14th Annual International Conference on Digital Government Research, 2013, pp. 1-10.

[30] J. C. Bertot, U. Gorham, P. T. Jaeger, L. C. Sarin, and H. Choi, "Big Data, open government and e-Government: Issues, policies and recommendations," Information Polity, vol. 19, pp. 5-16, 2014.

[31] M. Rajagopalan and S. Vellaipandiyan, "Big Data framework for national E-governance plan," in ICT and Knowledge Engineering (ICT&KE), 2013 11th International Conference on, 2013, pp. 1-5.

[32] R. Mahmood, M. Rashad, and M. El-Dosuky, "Efficient Implementation of Big Data Switch for Iraqi Cellular Phone Service Providers," International Journal of Information Science and Intelligent System, vol. 3, pp. 1-8, 2014.

[33] M. Hansen, T. Miron-Shatz, A. Lau, and C. Paton, "Big Data in Science and Healthcare: A Review of Recent Literature and Perspectives," Yearb Med Inform, vol. 9, pp. 21-26, 2014.

[34] W. Liu and E. Park, "Big Data as an e-Health Service," in Computing, Networking and Communications (ICNC), 2014 International Conference on, 2014, pp. 982-988.

[35] W. Raghupathi and V. Raghupathi, "Big Data analytics in healthcare: promise and potential," Health Information Science and Systems, vol. 2, p. 3, 2014.

[36] G.-H. Kim, S. Trimi, and J.-H. Chung, "Big-data applications in the government sector," Communications of the ACM, vol. 57, pp. 78-85, 2014.

[37] D. Fasel, "Potentials of Big Data for governmental services," in eDemocracy&eGovernment (ICEDEG), 2014 First International Conference on, 2014, pp. 17-18.

[38] J. Yu, F. Jiang, and T. Zhu, "RTIC-C: A Big Data System for Massive Traffic Information Mining," in Cloud Computing and Big Data (CloudCom-Asia), 2013 International Conference on, 2013, pp. 395-402.

[39] K. Kim, K. Oh, Y. K. Lee, and J.-Y. Jung, "Discovery of Travel Patterns in Seoul Metropolitan Subway Using Big Data of Smart Card Transaction Systems," Journal of Society for e-Business Studies, vol. 18, 2013.

[40] M. Smith, C. Szongott, B. Henne, and G. von Voigt, "Big Data privacy issues in public social media," in Digital Ecosystems Technologies (DEST), 2012 6th IEEE International Conference on, 2012, pp. 1-6.

[41] A. Sheth, "Transforming Big Data into Smart Data: Deriving value via harnessing Volume, Variety, and Velocity using semantic techniques and technologies," in Data Engineering (ICDE), 2014 IEEE 30th International Conference on, 2014, pp. 2-2.

[42] Z. Ming, C. Luo, W. Gao, R. Han, Q. Yang, L. Wang, and J. Zhan, "Bdgs: A scalable Big Data generator suite in Big Data benchmarking," arXiv preprint arXiv:1401.5465, 2014.

[43] Y. Demchenko, C. Ngo, C. de Laat, P. Membrey, and D. Gordijenko, "Big Security for Big Data: Addressing Security Challenges for the Big Data Infrastructure," in Secure Data Management, ed: Springer, 2014, pp. 76-94.

[44] Y. Demchenko, P. Grosso, C. de Laat, and P. Membrey, "Addressing Big Data issues in scientific data infrastructure,"

in Collaboration Technologies and Systems (CTS), 2013 International Conference on, 2013, pp. 48-55.

[45] P. Zikopoulos and C. Eaton, Understanding Big Data: Analytics for enterprise class hadoop and streaming data: McGraw-Hill Osborne Media, 2011.

[46] J. M. Young, "An Epidemiology of Big Data," 2014.

[47] H. Demirkan and D. Delen, "Leveraging the capabilities of service-oriented decision support systems: Putting analytics and Big Data in cloud," Decision Support Systems, vol. 55, pp. 412-421, 2013.

[48] W. Fan and A. Bifet, "Mining Big Data: current status, and forecast to the future," ACM SIGKDD Explorations Newsletter, vol. 14, pp. 1-5, 2013.

[49] M. M. G. Huddar and M. M. Ramannavar, "A Survey on Big Data Analytical Tools," International Journal of Latest Trends in Engineering and Technology, 2013.

[50] P.-T. Chung and S. H. Chung, "On data integration and data mining for developing business intelligence," in Systems, Applications and Technology Conference (LISAT), 2013 IEEE Long Island, 2013, pp. 1-6.

[51] D. L. Jones, K. Wagstaff, D. R. Thompson, L. D'Addario, R. Navarro, C. Mattmann, W. Majid, J. Lazio, R. Preston, and U. Rebbapragada, "Big Data challenges for large radio arrays," in Aerospace Conference, 2012 IEEE, 2012, pp. 1-6.

[52] S. Patidar, D. Rane, and P. Jain, "A survey paper on cloud computing," in Advanced Computing & Communication Technologies (ACCT), 2012 Second International Conference on, 2012, pp. 394-398.

★★★