Volume 5, Issue 8, August 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

The Big Data Revolution, Issues and Applications

Azzeddine RIAHI
IMC Laboratory, Department of Physics
Chouaib Doukkali University
Faculty of Sciences, El JADIDA, Morocco

Sara RIAHI
Department of Mathematics and Computer Science
Chouaib Doukkali University
Faculty of Sciences, El JADIDA, Morocco

Abstract - Mega-Data represent massive data that are very large and therefore very difficult to treat. This treatment depends on the area where these data are involved. One may want to visualize, analyze, extract measurements, make comparisons or explorations, store them if necessary ... We find the big data everywhere and in recent years it becomes a priority area of research that comes to the forefront whatever the specialties are: energy networks, the Internet and the Internet of Things, artificial intelligence, security, biology, ecology, health. This is not to mention the traditional areas where we have always tended to manipulate these variables, such as in quantum physics with the infinitely small and the infinitely large, which are mathematically related in nature. For example, in one nuclear physics, the European Organization for Nuclear Research announced in 2013 that it had crossed the threshold of 100 IT petabytes of data, or a million of millions bytes. Massive data is contained in the most of the fields. It's identification and collection is increasing for strategic purposes or for business risk measurement for governement.An industrial revolution just started .New challenges and risks are always related to the new opportunities.

Keywords- Big data, privacy, variety, velocity, volume, 3V's

I. INTRODUCTION

The large volumes of data, "Big Data" represent a significant technological evolution model of the company. The Big Data phenomenon radically changes the data management procedures since it introduce new issues concerning the volume, the transfer speed and data type [1]. It allows adapting the technologies and business strategies by providing critical information for targeted and optimized actions and provides the ability to access new business opportunities and better control the inherent risks. Moreover, it is likely to transform the modern enterprise as we know it today. Here are the obvious conclusions that we can draw:

- Big Data covers the real market needs, based on a new technological breakthrough, while most companies are in the research phase, many Big Data usage patterns have emerged, though the integration of data is essential to the processing of big data, quality and governance of data are nevertheless major concerns [2].
- Big Data projects leave the sphere of experimentation to become a strategic asset for the company, development tools are needed to increase the adoption of new technologies and thus reduce the mandatory use of highly skilled developers. All major vendors of infrastructure and databases begin to launch Big Data solutions on the market [3].

\Big Data is a very difficult concept to define precisely, since the very notion of big in terms of volume of data varies from one area to another. It is not defined by a set of technologies, on the contrary, it defines a category of techniques and technologies. This is an emerging field, and as we seek to learn how to implement this new paradigm and harness the value, the definition is changing. However, if it can be ambiguous, many experts believe that whole sectors of industry and market will be achieved and other created, as when these capabilities will enable the realization of new products and features that were previously unimaginable [4].

As the term suggests, Big Data is characterized by the size or volume information. But other attributes, including the speed and the type of data, must also be considered. Regarding the type, Big Data is often attached to the informal or semi-structured content, which can be challenging for conventional relational environments storage and computing. Unstructured and semi-structured data are everywhere: web content, twitter posts or comments customer free format [5]. For speed means the speed with which information is created. With these new technologies, it is now possible to analyze and use the large amount of data provided by the log files of websites, analysis of opinions of social networks, and even streaming videos and environmental sensors. We can take advantage of a possible strategic vision until today.

To understand the impact of this new paradigm, it is important to have basic knowledge on new concepts, as well as terms and key technologies that define the Big Data. At the heart of this revolution, known as the architecture MapReduce: it consists of a powerful massively parallel environment that performs advanced functions in very little time [3]. Introduced in 2004 by Google, MapReduce allows a programmer to run a data transformation, which are then delegated and handled by a "cluster" architecture which can include thousands of computers operating simultaneously. Basically, it consists of two steps: "map", where the problems are divided and distributed to several parallel servers, and "Reduce", where answers are consolidated from each map and solve the initial problem [6].

II. LITERATURE REVIEW

Massive data are the result of the meeting of three elements: Internet, social networks and smart devices (computers, phones, tablets, etc.). The Internet allows the transmission of information regardless of its form created on smart devices and shared on social networks [7]. The data creation tool is the intelligent device, the consumer is the user of social networks and the transmission vector is the internet. The English term "Big Data" was proposed by John Mashey, then Scientist at Silicon Graphics (Diebold, 2012) [8]. The acronym VVV, that stands for, volume, velocity and variety (Laney, 2001) is used often to define "Big Data" .The massive amounts of data that are available, is "volume" ,"Velocity refers to the speed needed for processing, analysing and using data .The both types of data available: structured or unstructured one is reffered to by "variety". [9].

- **Structured data:**

Structured data is the data that can clearly be identified and codified. The spreadsheet data are typically structured data. One can understand their meaning by crossing the tracks of the line and the column in which the data resides [10]. These data correspond to a codification which classifies and draw information. Algorithmic analysis systems have traditionally been developed to handle such data. The era of Big Data processing allows especially large-scale, real-time data.

- **Unstructured data:**

As their name suggests, unstructured data does not meet a codification that allows to draw information mechanically. There is no other way than reading posts to extract the meaning. It is then by analyzing the content of messages that can structure the information. Examples of types of unstructured data are text files, audio or video files and other information from an analog signal. Unstructured data now represent the vast majority of information that we interact and especially the least exploited part [10]. By implementing analytical tools for effective massive data, this information becomes exploitable.

Other definitions of "Big Data" are proposed in the literature (Warin & Sanger, 2014), we will limit ourselves here to the massive data set as having the following characteristics:

- they are either structured or unstructured,
- they are available in large quantities almost infinite,
- they are available in real time,
- they are mostly longitudinal.

This is not the first time that the information sciences, statisticians and econometricians are dealing with data in large quantities. For example, in finance, it is clear that information volumes traded on the servers of financial institutions correspond to the characteristics of massive data set out above [11]. Indeed, the financial data are identified (structured), they are available in extremely large amount, in real time, often for long periods. However, with the "variety" of data, other information - unstructured - are now available and are circulating on social networks. For example, twittering give qualitative information about what users think, their analysis of structured financial data, their reactions and interpretations after the publication of an annual report of a company or its financial data, etc. These unstructured data, once they are "restructured", can help to better explain certain phenomenon and reactions in the financial markets. They are therefore a new source of mass information.

Certainly computerization and digitization of information about our lives, our perceptions and our feelings will impact our societies. This new source of data a bit less neutral - as qualitative - will add to the massive data we already have [12]. Crossing these data give us new information. In the language of econometricians, interactions between these variables within most comprehensive database will allow the genesis of extremely important information that was not available before.

- **Metadata**

Metadata is structured data that primarily accompany unstructured data. Metadata partially characterize some unstructured data but does not actually help to process information from unstructured data [11]. So it's all of these data, structured and unstructured, which gives better information today. In terms of implications, this means for example that we could do a better risk management. Nevertheless, the stakes are not well defined [12].

III. THE BIG DATA AND 3VS

The concept of "big data" (massive data or mega-data) refers to a set of data so large it is difficult to treat with conventional tools. Often these data from various sources and are registered to permit their operation and without predetermined goal analysis and without time limit [13].

Two factors were decisive in the appearance of Big Data. There is firstly the development of Internet and the increasing number of connected objects that contribute to the creation of large volumes of data and secondly the development of storage capacity and computing that enable their treatment at ever lower costs [14].

Big Data in principle meet three characteristics: volume, velocity, variety.

- **Volume:** Big amount of data is represented by Big Data .In the last two years, 90%of data available today was created.
- **Velocity:** the data are generated, captured and shared in an ever increasing speed Deadlines updating and analyzing data are always shorter and are usually processed in real or near real time.
- **Variety** (or heterogeneity): the data analyzed are not necessarily structured. They can come from different sources (and have a different format such as text, images, digital traces, multimedia content, etc.) and be

combined. Data stored in an internal customer database can be combined with external data from social networks, search engines, official notice leaves or open data portals managed by public authorities.

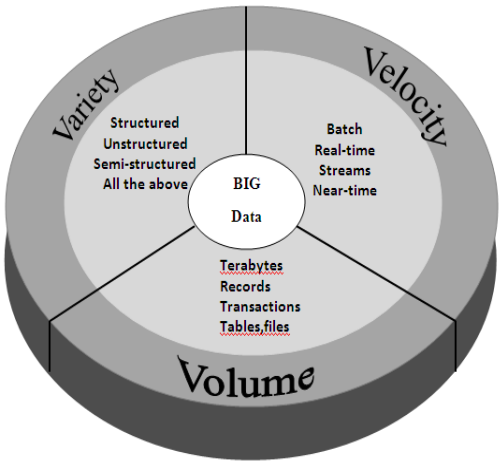

Fig 1: The 3 Vs of big data: volume, variety, velocity.

III.A- Quantities of data in information systems continuously increasing:

In effect, the world is creating more and more data each day and this data is stored on history much more greater. Their level of granularity is finer and many carriers are dematerialized.

Current data and data processing database management tools can not withstand such an explosion of volume. They can not meet the new requirements in a reasonable time and at reasonable costs [15].

Current Datamining tools allow for example to analyze a lot of data but on a sample considered representative; Big Data can treat all data.

The volumes of data to process and store which are more and more important in the information systems can quickly become a liability for companies (exponential costs, inability to analyze and access all the data in a reasonable time delay compared to competitors). They may become a force for the creation of solutions able to manage and develop these data volumes [16].

III.B- New data formats, more varied, integrate and analyze:

The trend was previously structured as possible to integrate the data in the information for easy storage and treatment systems. Tools such as ETL (Extract Transform and Load) enable standardized, structured data before integrating them into Information Systems [15]. But new data formats knocking at the door of companies, determined to prove their importance in business strategy and decision-making.

These new unstructured data themselves are varied: photos, emails (with the semantic content analysis), data from social networks, or data from GPS sensors, meteorological sensors [16].

Current information systems are not able to receive and analyze these new data formats. Big Data can integrate and analyze this unstructured data, collected from various sources [17].

III.C- Data to be collected and analyzed in real time:

These new data are generated streaming [15], and must therefore be treated in near real time as they are integrated in certain time-sensitive process: analysis of meteorological data in real time to anticipate natural disasters, such fraud management [17].

IV. THE TYPE OF THE DATA AND METHODS OF ANALYSIS

The analytical models developed greatly depend on the quality of access to relevant data .The internal optimization problems can often be resolved by analyzing data specific to the company [18]. However, once the analyzes are beyond the scope of the company (customers, suppliers, new markets ...), many barriers to access to information and data can stand.

IV.A-The characterization of data:

There are four variables that characterize the data on their use (McKinsey & Company, 2014) [18]:

- The degree of access to data;
- Computer data compatibility;
- The cost of data;
- The right to use and disseminate data.

These four characteristics are used to determine how these data are easily accessible and fully usable. Many organizations make their data available to all ("open data") using their website and contribute to the generation of massive data. Another feature that could be added is the type of data and the possible or not, direct use (structured data or not).

IV.B- The legal nature of the data:

Massive data is generated virtually, but also have other characteristics that must be considered [18]. It is important for example to define the legal nature of data: owner, personal or opened [17].

IV.B.1- The owner or personal data:

Company-specific, proprietary data are those obtained directly from the activity of a company. Here we find customer data, suppliers, transactional data, production data, employees, financial results [18]. This data is not generally available for external actors to the company, unless setting up a contract for an exchange. The personal data of individuals such as the number of bank accounts owned by them and are not accessible. Some data simply must not end up on sites open to all [19].

IV.B.2- The Open data:

In contrast, some data can be shared freely by governments, some companies or non-governmental organizations, but also the data from the research [19]. For example, open government data is available to everyone and in particular to allow access to relevant market information in the country.

IV.B.3- The massive data analysis methods:

The big question is whether our quantitative tools are sufficient to analyze all the data. Are there any additional problem with unstructured data begin with a tentative definition of unstructured data. These are actually the data that originally rather a qualitative nature which are nevertheless translated into computer code to use on our electronic devices. The fact that qualitative information be translated into computer code will allow to quantitatively analyze this information. This will structure these data. For example, accepting someone as a friend on a social network falls under the emotional. Yet choosing someone as a friend will create a field in the social network database that will combine these two [20].

Models of analysis of quantitative data already existed since the early work on probability and stochastic analysis, and have never ceased to be improved over time. Analysis of very large databases tools with a powerful capacity calculations have been developed. In addition, to explore and look for data, data mining methods are an obvious starting point, but we must also take account of specific data. The machine learning methods are another obvious starting point. The availability of a high computing power with the development of new hosts allowed this development. Today, considering the nature of the data and their widespread availability, the methods from machine learning are interesting application areas, such as trees or regressions factor analyzes are sometimes more effective than econometric models first generation [21].

V. THE MASSIVE DATA

Big data is a term used to describe data sets that become so large that they become difficult to work with conventional tools database management .In these new orders of magnitude, capture, storage, search, sharing, analysis and visualization of data must be redefined. "So this is a set of technologies, architecture, tools and procedures for an organization to quickly catch process and analyze large quantities and heterogeneous and changing content and extract relevant information at an affordable cost [22].

V.A- Innovation and Revolution:

Massive data correspond to the concept of innovation in its great definition: they are the result of product innovations and are a process innovation. They have an impact on operations in companies of business processes on business strategies, but also government institutions and the choices of individuals [22].

The massive data are used in a number of sectors increasingly important. It is also important to understand that massive data create new markets for companies that manage these massive data but also for existing businesses that can use this data to better define their strategies, target their marketing campaign, know their markets, etc.

V.A.1- Opportunities for companies:

Massive data are in themselves a market. Silicon Valley is one of the examples, and certainly the most convincing, convergence between high technology products and services around these products. Today, initially different companies like Apple and Google find themselves actually in the same markets. Originally, Apple manufactures computers and Google makes an Internet search service. Nothing pitted. Today, their business model relies increasingly on their ability to generate, attract, collect and analyze data to income or advertising purposes or towards the sale of products.

Massive data have also a strategic dimension in the analysis of industrial sectors and their evolution. Certainly data on users great value and have long been at the heart of the major issues on the risks associated with the development of online services. Users are often aware of the arbitration they do when they give access to their information: they compare this indirect cost of accessing their information with the profit made by the website or online service [23].

But beyond product innovations, keep in mind that mass data is a radical innovation process. Indeed, their use changes the sharpness analysis and organizational processes. This new tool allows to direct more efficiently the organizational processes within a company, but also about acquisitions of companies that will deal with new partners that have been identified thanks to the conclusions from these data.

V.A.2- The field of Finance:

The financial industry relies heavily on access and analysis of information. In this sense, the massive data enable financial actors to access more data, and after analysis, more information. There are specific properties to each type of data and a use that is different. For example, financial data are used and will remain so for contextual information they produce. This information is critical for two important phases in finance: the analysis phase and the phase of decision-making [24]. Massive data can be used in the analysis phase to complete financial data and participate in improving the

contextual information about the intrinsic characteristics of financial products, but they will be primarily used in the decision-making phase to inform analysts that have drawn conclusions other analysts in the markets. In this example, mass data are used to measure the performance of others.

V.A.3 -The field of Health:

In view of the quantity and also the diversity of data that a person produces daily, the use of massive data is needed as an essential strategy for the health sector and more specifically public health [25].

Google's vision of the challenges related to health is an agreement with his main vision of business, that is to say, the processing and archiving of massive data. Government initiatives have gained momentum in recent years by setting up structures to facilitate the emergence of massive use of health data (such as genome sequencing, imaging, etc.) [26].The Big Data Institute of the University of Oxford in the UK, with 30 million pounds, will focus on the analysis of healthcare records of patients, and clinical trials of DNA sequences through the work of 600 researchers. Still in the UK, the research center dedicated to DNA sequencing, has partnered with DataDirect Networks company to treat 10 pentabytes data generated daily by DNA sequencing of the Institute [27].

V.A.4- The field of education and knowledge:

One of the Internet's promises of universal access to information. The massive data contribute to the spread of knowledge by revealing new possibilities previously nonexistent: attend university pushed online feedback of tens of thousands of users, setting light exclusion zones and ensure information of the cultural content are part of the new reality offered by massive data. A blog can not alone change the information within this ocean of data, but the social fabric built on the Internet can transmit information faster and more comprehensively [28].

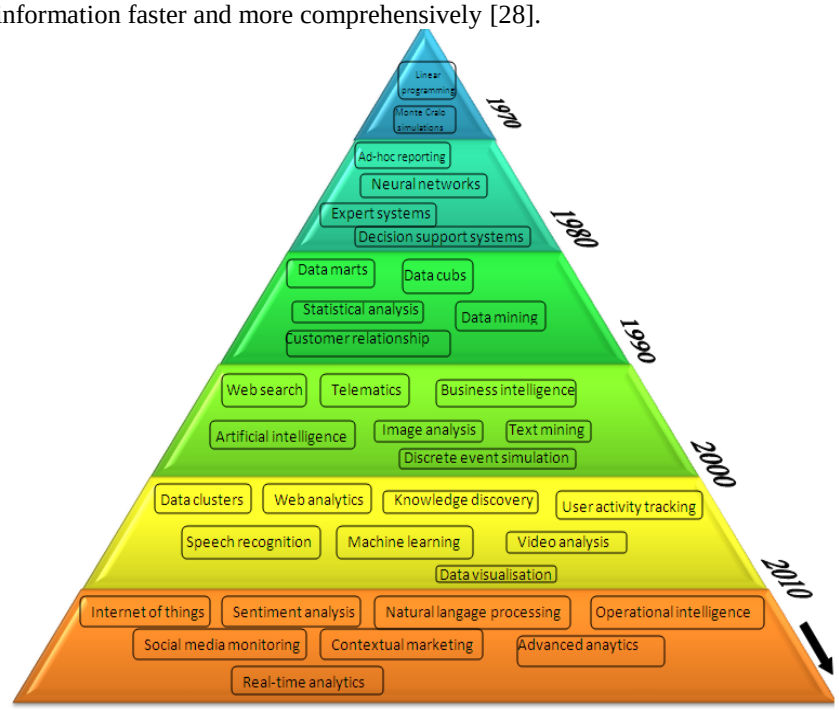

Fig 2: Evolution of the degree of sophistication depending on the complexity of the data

V.B- Opportunities and risks:

The opportunities, issues, challenges and risks are manifold. As with any radical innovation, uses can both create immense opportunities but also accompanied by significant risks. Yet in the past, there have been few innovations that have generated mismanaged risk by corporations. In theory, the only reason that we would cross the border between risk and uncertainty [29] .The risk is defined as a closed set of events probabilisables, while uncertainty is an open set of events by definition can not be calculated. The ambiguity associated with situations of uncertainty would make us into areas of "risk" not calculable. And it is true that the nature of massive data is to be a radical innovation so that it is difficult now to imagine all the opportunities and thus all risks.

V.B.1- The issues related to data ownership:

It is often understood that the data items are virtual and even when they are stored. The reality is of course more complex. The data certainly represent a somewhat vague concept a priori, but they have a real material existence conferred on them by the law of intellectual property and the jurisdiction where are the headquarters of companies or institutions collecting this data or the courts subsidiaries, necessary and sufficient condition [30].

V.B.2- The protection of privacy:

The literature on the protection of privacy and the risk to the spread of massive data is still in its infancy. The first articles are intended to emphasize the existence of risks and the need for self discipline. Some authors have started to make a review of the existing literature. Under these conditions, we understand that the data have a different citizenship as a legal person of the physical personality that generates the data [31]. More and more studies show the risks to the protection of privacy. For example, user contracts including protection clauses of online privacy that users accept when registering to a website are often questionable.

V.B.3- Strategic challenges for companies:

V.B.3. a- Massive data and competitive advantages:

Companies that have access to data and have the capacity to analyze and highlight the important strategic elements for they will have significant competitive advantages over their competitors. Combined with earlier notions on geostrategic issues, it appears that States have an even more important role to play in the future and that the concept of competitive advantage of nations will be even closer to the competitive advantages of companies [32].

This competitive advantage of companies benefiting from this location and all the positive externalities pose serious pressures and threats on other companies. The latter will have to revise their business processes, their geographical location, etc. There will be significant adjustment costs for businesses follower. But overall, it is certain that the principle of creative destruction will be very beneficial to the economies [33].

V.B.3.b- Massive data and business reputation:

The company's intangible value will be his priority is a real source of value creation. The issue of reputation is a real issue for companies. The issue of credibility is important, but the virality of a rumor, true or false, is a major risk. Social networks can make and break reputations .The messages broadcast on social networks can become viral and damage for a very long time the reputation of a company. Virality is a real risk, and the authors suggest that these risks are taken very seriously by management and boards of directors [34].

VI. CONCLUSION AND PROSPECTS:

The convergence between the hardware and the software that we live a real radical innovation process. This convergence allows to collect, analyze and form new answers to the problems facing the different stakeholders. Intuition would like the services sectors are most affected by the massive data, but as we have seen, the behavior of consumers of products is generating more and more data. The massive data is both from service users as product consumers. This is only the beginning. We may know in the coming years, new developments, new uses and new integrations that will open new possibilities. With these new opportunities, we will also face new challenges and new risks. While this seems alarming, it's the nature of all the innovations that bring new questions. So far, the human being rather taken advantage of these innovations. Massive data is in the process of establishing itself as a revolution of which we can not do without. They will become a starting point for the analysis of competitiveness, the level of competition and growth of a country and its companies .The massive data represents a radical innovation process, and like all of them, they are one of the main vectors of the industrial revolutions.

Our next work focuses on medical imaging since data increase and multiply in the field of brain research, which also seems to get into the big data era. The increase in volume generated images neuroimaging, the size of the cohort followed patients and the number of analyzes are factors behind this explosion. Added to this is the regular monitoring of patients, which requires the continuous updating of the analyzes.

REFERENCES

[1] Riccardo Torlone,'Big Data: An introduction ',Università Roma Tre,2012.

[2] Jules J. Berman,' Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information', ISBN 0124047246, 9780124047242, Newnes, 2013.

[3] Juliana Freire,' Principles for Working with Big Data', Visualization and Data Analysis (ViDA) Lab ,Computer Science & Engineering, Center for Urban Science & Progress (CUSP), Center for Data Science, New York University, http://sites.nationalacademies.org/cs/groups/depssite/documents/webpage/deps_087672.pdf.

[4] Big data and data protection,20140728, Version: 1.0, https://ico.org.uk/media/for-organisations/documents/1541/big-data-and-data-protection.pdf.

[5] Edmon Begoli, James Horey,' Design Principles for Effective Knowledge Discovery from Big Data', Computational Sciences & Engineering Division,Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, 2012.

[6] Hong-Linh Truong, Schahram Dustdar,' Principles of Software-defined Elastic Systems for Big Data Analytics', IEEE International Conference on Cloud Engineering, DOI 10.1109/IC2E.2014.67 ,2014.

[7] Kenneth Cukier,' Big Data and the Future of Business', Reinventing the Company in the Digital Age, 2013.

[8] Gopalakrishna Palem,' Formulating an Executive Strategy for Big Data Analytics', Technology Innovation Management Review, www.timreview.ca, March 2014.

[9] The 7 pillars of Big Data, Petroleum Review,January2015,https://www.landmarksoftware.com/Lists/DocumentRepository/The7pillarsofBigDataWhitepaper.pdf

[10] Big Data Next-Generation Analytics, Business Analytics, https:// www.at .capgemini. com /resource-file access /resource /pdf /Big Data Next Generation Analytics.pdf.

[11] Unstructured data: A big deal in big data, DEEP ANALYTICS, GAME CHANGING TECHNOLOGY TO MEET AGENCY MISSIONS ,Game Changer, http://www.digitalreasoning.com/resources/Holistic-Analytics.pdf.

[12] Big Data 101: Unstructured Data Analytics, a Crash Course on the IT Landscape for Big Data and Emerging Technologies, JUNE 2012, www.triforce.com.au/resources/big-data-101-brief.pdf .

[13] Bernice Purcell,' the emergence of "big data" technology and analytics', the emergence of "big data", Journal of Technology Research, http:/ /www. aabri.com /manuscripts /121219. Pdf.

[14] Doug Laney, ''Application Delivery Strategies'', Copyright © 2001 META Group Inc. APPLICATION DELIVERY STRATEGIES is published by META Group Inc., 208 Harbor Drive, P.O. Box 120061, Stamford, CT 06912-0061.
[15] Big Data:The next frontier for innovation ,competition and productivity ,McKinsey &Company ,May 2011.
[16] Scott Brueggeman,'Big Data and the 3 V's:Volume, Variety and Velocity', http:// www. cyrusone.com/pdf/c1-exec-report-big-data-and-the-vs-volume-variety-and-velocity.pdf.
[17] Big Data beyond the hip, White paper by Datastax Corporation, October 2013, http:/ /www.datastax.com/wp-content/uploads/2011/10/WP-DataStax-BigData.pdf.
[18] Thomas F. Dapp, Veronika Heine, ''big data the untamed force '', May 5, 2014. https:// www. dbresearch. com/PROD/DBR_INTERNET_EN-PROD/ PROD0000000000334 340 /Big data+-+the+untamed+force.PDF.
[19] Helena Forest & Evelyn Foo,' Big Data How it can become a differentiator', GTB Product Management and Donya Rose & Dmitriy Berenzon, Product Strategy & Business Development, Copyright © October 2014 Deutsche Bank AG.
[20] Chun Wang, Ming-Hui Chen, Elizabeth Schifano, Jing Wu, and Jun Yan,'' A Survey of Statistical Methods and Computing for Big Data'', arXiv:1502.07989v1 [stat.CO] 27 Feb 2015 ,March 2, 2015.
[21] Zeynep Tufekci,'' BIG DATA: PITFALLS, METHODS AND CONCEPTS FOR AN EMERGENT FIELD'', Electronic copy available at: http://ssrn.com/abstract=2229952,2013. K2
[22] BIG DATA AND PRIVACY: A TECHNOLOGICAL PERSPECTIVE, Executive Office of the President President's Council of Advisors on Science and Technology, https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy.pdf ,May 2014.
[23] Thomas H. Davenport, Jill Dyché,'' Big Data in Big Companies'', Copyright © Thomas H. Davenport and SAS Institute Inc. All Rights Reserved. Used with permission, May 2013.
[24] Steve Lavalle, Eric Lesser, Rebecca Shockley, Michael S.Hopkins and Nina Kruschwitz,'' Big Data, Analytics and the path from insights to value'', Vol 52 N 2, winter 2011.
[25] Wullianallur Raghupathi, and Viju Raghupathi,'' Big data analytics in healthcare: promise and potential'', Raghupathi and Raghupathi Health Information Science and Systems 2014, 2:3http://www.hissjournal.com/content/2/1/3, 2014.
[26] Azzeddine Riahi,'' MRI Image Segmentation by K-Means Clustering Method and Detection of Lesions'', International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064, Volume 4 Issue 6, June 2015.
[27] Matthew Herland, Taghi M Khoshgoftaar and Randall Wald,'' A review of data mining using big data in health informatics'', Herland et al. Journal of Big Data 2014, 1:2 http://www.journalofbigdata.com/content/1/1/2,Springer, 2014.
[28] Oliver Marsh, Lajos Maurovich-Horvat, Dr Olivia Stevenson,'' Big Data and Education: What's the Big Idea'', UCL POLICY BRIEFING, SEPTEMBER 2014.
[29] Nathaniel Schenker,'' Big Data: Uses and Limitations'', Presentation for discussion at the meeting of the NCHS Board of Scientific Counselors, September 19, 2013.
[30] AAPOR Report on Big Data, AAPOR Big Data Task Force, http://www .aapor.org/ AAPOR Kentico/AAPOR_Main/media/Task-Force-Reports /BigData_TaskForc eReport_ FINAL_2_12_15.pdf,February 12, 2015.
[31] Securing the Big Data Life Cycle, http://www.oracle.com/us/technologies/big data /securing-big-data-life-cycle-2543085.pdf, Copyright © 2015, MIT Technology Review.
[32] The power of combining Big Data analytics with business process workflow ,white paper ,cgi.com,© 2013 CGI GROUP INC.
[33] Michael Schroeck, Rebecca Shockley, Dr. Janet Smart, Professor Dolores Romero-Morales and Professor Peter Tufano,''Analytics: The real-world use of big data'', IBM Institute for Business Value, GBE03519-USEN-00, Copyright IBM Corporation 2012.
[34] Travis Pearson, Rasmus Wegener,'' Big Data: The organizational challenge '', Copyright © 2013 Bain & Company, Inc.

