American Journal of Information Science and Technology
2018; 2(1): 1-8
http://www.sciencepublishinggroup.com/j/ajist
doi: 10.11648/j.ajist.20180201.11

Review Article

Big Data: Myths, Realities and Perspectives - A Remote Look

Djamel Ghernaout[1, 2, 3, *], Mohamed Aichouni[2, 4], Abdulaziz Alghamdi[2, 5], Noureddine Ait Messaoudene[2, 5]

[1]Chemical Engineering Department, College of Engineering, University of Ha'il, Ha'il, Saudi Arabia

[2]National Initiative on Creativity and Innovation Project, College of Engineering, University of Ha'il, Ha'il, Saudi Arabia

[3]Chemical Engineering Department, Faculty of Engineering, University of Blida, Blida, Algeria

[4]Industrial Engineering Department, College of Engineering, University of Ha'il, Ha'il, Saudi Arabia

[5]Mechanical Engineering Department, College of Engineering, University of Ha'il, Ha'il, Saudi Arabia

Email address:
djamel_andalus@hotmail.com (D. Ghernaout)

[*]Corresponding author

To cite this article:
Djamel Ghernaout, Mohamed Aichouni, Abdulaziz Alghamdi, Noureddine Ait Messaoudene. Big Data: Myths, Realities and Perspectives - A Remote Look. *American Journal of Information Science and Technology*. Vol. 2, No. 1, 2018, pp. 1-8. doi: 10.11648/j.ajist.20180201.11

Received: April 13, 2018; **Accepted**: April 27, 2018; **Published**: May 14, 2018

Abstract: In a world where data is gathered in ever - increasing quantities, summing more of what persons and organizations perform, and catching smallest detail of their comportment. There are three fashions to distinguish data occasionally reported as volume, variety, and velocity—the meaning of Big Data. This review aims to focus on defining Big Data and describing some of its myths and realities. The significance of big data does not focus on how much data is possessed, but what things may be performed with it. Data may be extracted from any origin and examined to detect replies that let 1) cost decreases, 2) time decreases, 3) fresh product expansion and studied offerings, and 4) smart decision making. As a magic, charming, and mysterious noun, Big Data remains an attractive novel field in both science and technology. Despite of the developed technology and open knowledge, Big Data still needs more familiarization and demystification. More developed computer skills will be needed to understand and touch its practical extent.

Keywords: Big Data, Descriptive Analytics, Predictive Analytics, Prescriptive Analytics, Brontobytes Period, Internet

1. Introduction

The lightning blooming of the Internet and World Wide Web has conducted to huge quantities of data obtainable online [1]. Moreover, business and government institutions generate enormous quantities of both structured and unstructured data that require treating, examining and connecting [2, 3]. In 2007, it is evaluated that the quantity of data stocked in a digital configuration was 281 Exabytes (10^{18} bytes) (Table 1), and the global compound expansion rate has been 57% with data in institutions expanding at an even quicker rate (Figure 1) [4, 5]. It is as well evaluated that 95% of all present-day data is in unstructured configuration with augmented information treating needs comparatively with structured data. Stocking, controlling, obtaining, and treating this large quantity of information constitute a vital requirement and a gigantic defy with a view to meet the expectation to discover, examine, extract, and envisage these data as information. This flood of data, in conjunction with rising methods and engineering employed to manage it, is frequently mentioned nowadays as *Big Data computing* [1, 6].

Big Data may be described as bulks of data obtainable in changing levels of intricacy, produced at various rates and changing levels of confusion, that cannot be treated employing conventional engineering, treating techniques, algorithms, or any commercial off-the-shelf solutions [7, 8]. These data comprise weather, geo-spatial and Geographic Information Systems (GIS) data, consumer-driven data from

social media, enterprise-generated data from legal, sales, marketing, procurement, finance and human-resources departments, and device-generated data from sensor networks, nuclear plants, X-ray and scanning devices, and airplane engines [1].

This review describes the characteristics, challenges, and solutions offered by Big Data computing.

Table 1. *Measurement of storage capacity [3].*

Unit	Power of 2	Approximate Number
Kilobytes	2^10 = 1024	Thousands
Megabytes	2^20 = 1048576	Millions
Gigabytes	2^30 = 1073741824	Billions
Terabytes	2^40 = 1099511627776	Trillions
Petabytes	2^50 = 1125899906842624	Quadrillions
Exabytes	2^60 = 1152921504606846976	Sextillions
Zettabytes	2^70 = 1180591620717411303424	Septillions

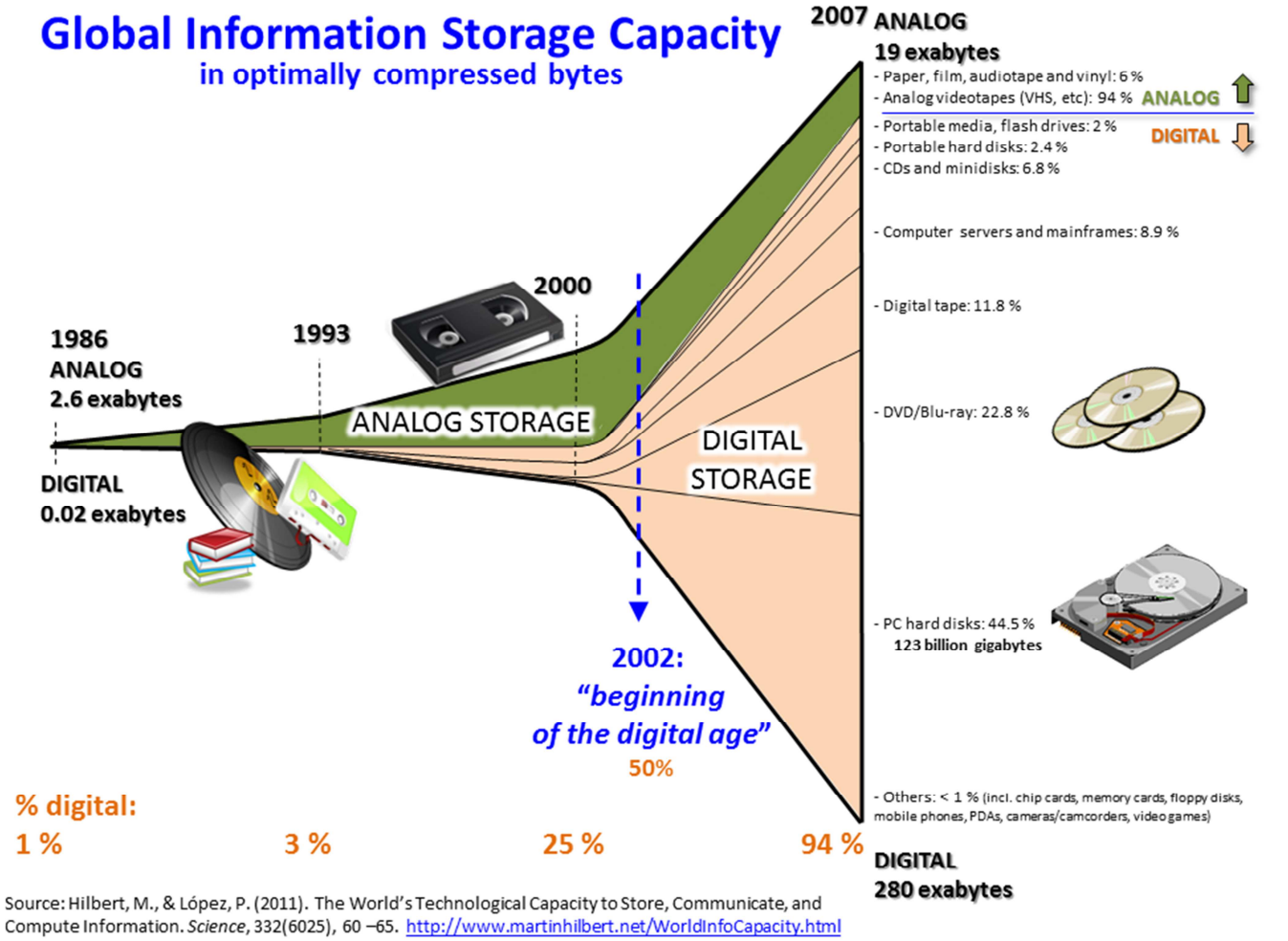

Figure 1. *Growth of and digitization of global information-storage capacity [4, 5].*

2. Big Data

An executive is supposed to work like a handyman [9]. He is invited to be an industrial/organizational psychologist, a logician, a bean counter, and a representative of his firm to the exterior world. In a different way, he is slightly a specialist who may go deeply into details. Since the distinct techniques run across are beginning to be more complicated, the dissimilarities between them and their antecedents are beginning to more confused [10].

A manager may have previously conducted his company's passage to different novel techniques such as the Internet [9]. Since 2000, Internet existence moved from being not required to being compulsory for most companies. Since 2005, Internet existence moved from being operating in a single direction to chatty. The moment that a company may publicize its online board with either data concerning its physical location, hours, and offerings if it were a brick-and-mortar business or else its offerings and an automated payment system if it were an online business, it is launched [11]. Companies extending from Barnes & Noble to the corner pizza chain traversed these universes.

A novel departure came: Web 2.0 [9]. Notwithstanding much exaggerated way with words, this naming defined the actual event of a mutual online universe.

At the time that the Internet open out and developed, it begun to be a required meeting for connecting, in addition to a menacing implement whose capacity for usefulness or badness may affect others by abruptness or produce

self-inflicted damage. At the moment, there is an additional novel departure: Big Data [9].

Consequently, what is Big Data? Is it a craze (Figure 2)? Is it vacant specialized language? Is it just a novel label for expanding capability of the identical databases that have been a portion of our lives for ten years? Or, is it matter qualitatively various? What are the guarantees of Big Data? From which trend should an executive expect menaces [9]?

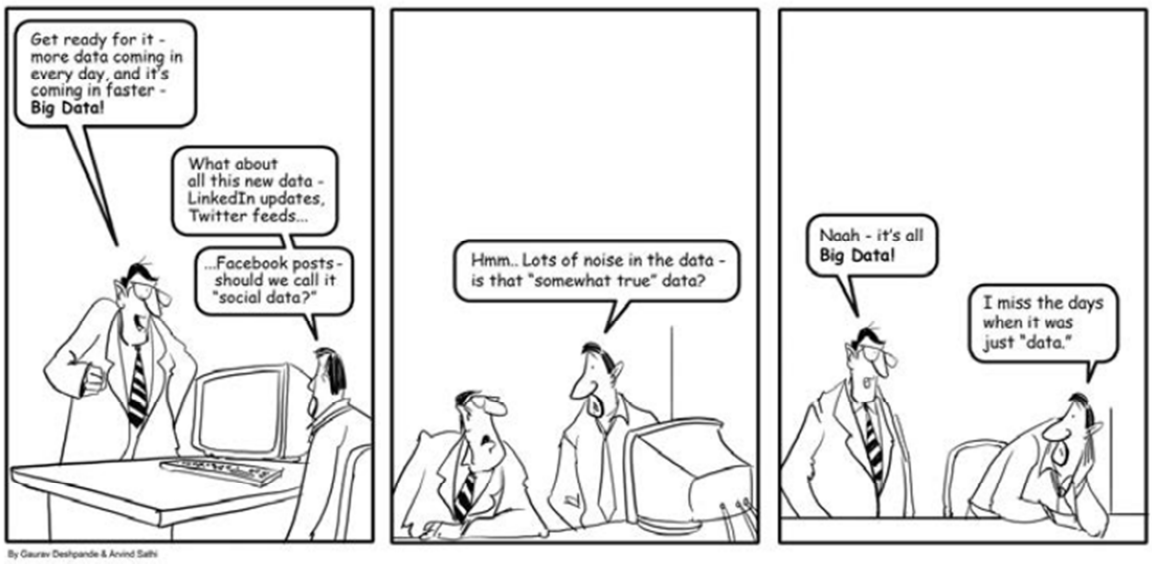

Figure 2. *From "data" to "Big Data" [10].*

The inclination of the press to publicize novel and hardly comprehend events renders it arduous to assess novel techniques, beside the kind and degree of their importance. Authors, such as Pries and Dunnigan [9], insisted that Big Data is novel and has strategic importance. The disagreement Pries and Dunnigan [9] made concerning Big Data is regarding how it rests on comprehensible expansions in engineering and is itself understandable. In spite of the fact that it is understandable, it is not simple to employ and it may bring ambiguous or false findings. Nevertheless, such wrong results are not frequently unexpected. They arise from particular statistical and data-related events. Taking into account that such events are actual and comprehending how they work let you as an executive to be a better user of your Big Data method (Figure 3) [12].

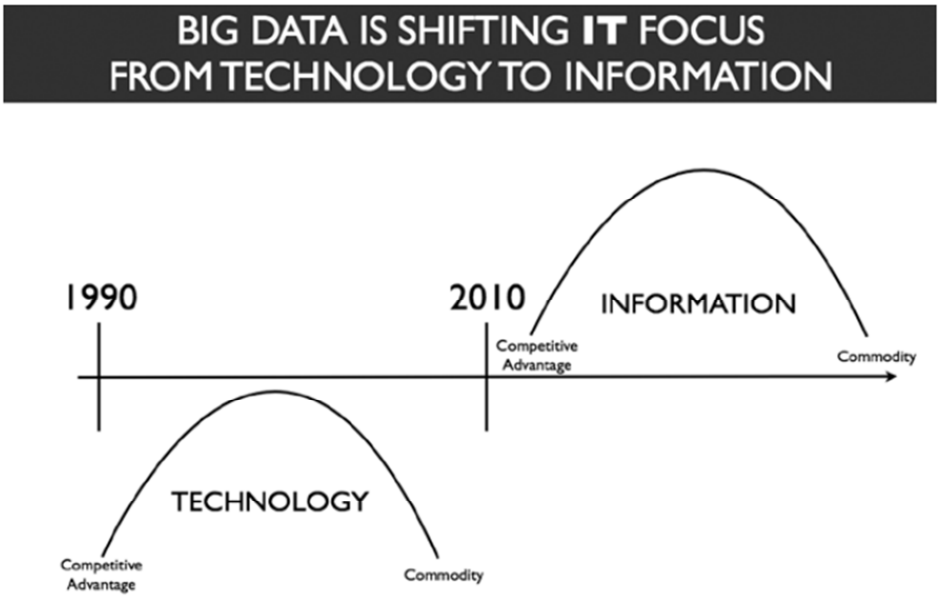

Figure 3. *Information is becoming the critical asset that technology once was [12].*

Such as cell phones and e-mail, Big Data is a new event that has appeared as a piece of the scene of our everyday existences. While purchasing through the Internet, discussing with connection on Facebook, performing web seeking, browsing papers mentioning database searches, and obtaining unwelcome vouchers, is interfacing with Big Data. As Pries

and Dunnigan [9] wrote: *"Many readers, as participants in a store's loyalty program, possess a key fob featuring a bar code on one side and the logo of a favorite store on the other. One of the primary rationales of these programs, aside from decreasing your incentive to shop elsewhere, is to gather data on the company's most important customers. Every time you swipe your key fob or enter your phone number into the keypad of the credit card machine while you are checking out at the cash register, you are tying a piece of identifying data (who you are) with which items you purchased, how many items you purchased, what time of day you were shopping, and other data. From these, analysts can determine whether you shop by brand or buy whatever is on sale, whether you are purchasing different items from before (suggesting a life change), and whether you have stopped making your large purchases in the store and now only drop in for quick items such as milk or sugar. In the latter case, that is a sign you switched to another retailer for the bulk of your shopping and coupons or some other intervention may be in order. Stores have long collected customer data, long before the age of Big Data, but they now possess the ability to pull in a greater variety of data and conduct more powerful analyses of the data"* [9].

Big Data affects us less visibly—it communicates to the hidden organizations of our community, like industrialization, transport, and power [9]. Any manufacturing expanding huge amounts of varying data is prepared for Big Data. Indeed, such productions in all likelihood employ Big Data before now. The technical dramatic change taking place in data analytics lets more exact allotment of materials in our advancing economy—at the same level the development in navigational engineering, from the superseded sextant to modern Global Positioning System (GPS) gadgets, authorized ships to navigate open seas [9].

To a great extent, Big Data is comparable to the Internet—it possesses disadvantages; however, its pure merit is certain. Discussing Big Data, such as political discussion, is inclined to illusory absolutes and erroneous contrasts. Like the situation of political discussion, the veracity most likely not consists in such absolutes. As Pries and Dunnigan [9] wrote: *"Like a car, you do not start up a Big Data solution and let it motor along unguided—you drive it, you guide it, and you extract value from it"*.

In contrast with the past, data itself is a plus. Organizations put money into practices to arrange and take out plus from their data. Several organizations construct entire businesses to collect, refine, and sell data [9].

3. Driving Forces for Big Data

The generation of Big Data is mounting at an alarming rate [10]. This section discusses the factors beyond this surge of Big Data. There are three participating elements: consumers, automation, and monetization (Table 2). In addition to these participating elements, their reciprocal influence is accelerating the formation of Big Data. With rising automation, it is simpler to provide Big Data formation and use chances to the clients, and the monetization operation is more and more offering a dynamic marketplace for Big Data [10].

Table 2. *Three participating elements to Big Data formation [10].*

Contributing element	Description
Advanced Consumers	Augmenting the information degree and the linked instruments has generated a novel type of advanced consumers. Such consumers are more checker, more perceptive at employing statistics, and more connected, employing social media to quickly gather and compare viewpoint from others. Today, the world is filled of selling communications. During the time that most of the selling remains transmitted employing newspaper, magazine, network TV, radio, and display advertising, surprisingly in the classical press, tight casting is progressively beginning to be more outstanding. The Internet world may begin to be greatly individualized. Search engines, social network sites, and electronic yellow pages enter ads fixed to a person or to a micro-segment. Internet cookies are more and more employed to pursue customer comportment and to cut amount following this comportment.
Automation	Interactive Voice Response (IVR), kiosks, mobile devices, email, chat, corporate Websites, third-party applications, and social networks have produced an acceptable quantity of event data concerning the clients. Moreover, client reciprocal actions through classical press like call centers may at the present time be examined and arranged. The greatest variation is in our capacity to change the client skill employing software policies, procedures, and personalization, making self-service progressively customer friendly.
Monetization	From a Big Data Analytics standpoint, a "data bazaar" is the greatest enabler to form an outer market space, where customer information is gathered, traded and put up for sale. A novel movement is observed in the market space, in which client practice from one sector is made anonymous, covered, and put on sale to different sectors.

4. Big Data Perspectives

The semantic Web, or Web 3.0, is frequently mentioned as the following stage of the Internet. Conducted by the World Wide Web Consortium (W3C), the aspiration is to metamorphose the contemporary web of unstructured and semi-structured data into a "web of data." Following W3C, the semantic web will render it simpler to share and reuse data through application, community, and enterprise interfaces [13].

There are two decades, the inventor of the web, Tim Berners-Lee, already named the semantic web "a web of data, in some ways like a global database" [13]. This database will be comprised of all unstructured, semi-structured, and structured data currently online but still residing in silos.

The semantic web will let all persons and Internet linked implements (think: the Internet of Things) to interact with each other as well as share and reuse data in various forms through changing applications and institutions in real time. Apparently this has all to do with Big Data [13].

The tomorrow of Big Data will as well convert the fashion by which the huge quantities of fresh data are examined [13].

At some time in the past, the focus was on comprehending how the institutions or the world surrounding us acted by examining the obtainable data employing descriptive analytics. This replied to the interrogation: "What happened in the past with the business?" With the handiness of Big Data, the novel period of predictive analytics embarked, which concentrates on replying the interrogation: "What is probably going to happen in the future?" Nevertheless, the actual benefit of analytics arrives with its ultimate phase, which can be named the tomorrow of business analytics and is named prescriptive analytics. This kind of analytics attempts to reply to the interrogation: "Now what?" or "so what?" It tends to give a recommendation for fundamental resolutions founded on tomorrow results. To comprehend what this will significate for an institution, it would be better to focus on what the divergences are within these three various kinds of analytics and how they touch an institution?

In fact, these three kinds of analytics must be present together—business, predictive, and prescriptive (Table 3). As Van Rijmenam [13] wrote: *"One is not better than the other; they are just different. All, however, are necessary to obtain a complete overview of your organization. In fact, they should be implemented consecutively. All of them contribute to the objective of improved decision making"*.

Table 3. *Three kinds of analytics [13].*

Three kinds of analytics	
Descriptive analytics concerns the past	Descriptive analytics lets institutions comprehend the past. The past may be one minute ago or a few years back. Descriptive analytics lets to comprehend the link among clients and products. The aim is to understand what way to adopt tomorrow, or comprehend from previous comportment to touch tomorrow results. Frequent cases of descriptive analytics are management reports that offer data about sales, clients, procedures, and finance, and then seek to determine relationships between the parameters. Descriptive analytics is consequently a crucial reservoir for finding what to perform tomorrow. Using predictive analytics, such information may be transformed into data concerning the probable next result of a phenomenon.
Predictive analytics concerns the future	Predictive analytics offers institutions practicable intuitions founded on information. It gives an evaluation concerning the probability of a tomorrow result. To perform this, a set of methods are employed, comprising machine learning, data mining, modeling, and game theory. Predictive analytics may assist to determine tomorrow hazards and occasions. Historical and transactional data are employed to detect models, at the same time statistical patterns and algorithms are employed to get links in different datasets. Predictive analytics has genuinely gained maturity in the Big Data period, and there are numerous instruments accessible for companies to predict tomorrow results. Using predictive analytics, it is fundamental to possess as much information as possible. More data significates better predictions.
Prescriptive analytics offers advice founded on predictions	Prescriptive analytics is the ultimate phase in comprehending your commerce; however, it remains in its debut. Five years ago, prescriptive analytics was cited as an "Innovation Trigger" that will require additional five or ten years to attain its maximum of efficiency. Prescriptive analytics not only anticipates what will arrive and when, but as well why it will arrive. It gives advices for doing on this data to profit from the predictions. It employs a group of several various methods and instruments, like mathematical sciences, business rule algorithms, machine learning, and computational modeling techniques, and also several various data sets ranging from historical and transactional data to public and social datasets. Prescriptive analytics tends to predict what the impact of tomorrow decisions will be in view to adapt the decisions before they are really taken. This will greatly enhance decision making, since future results are considered in predictions. Prescriptive analytics is extremely fresh; it has only been around since 2003, and is so complicated that there are very few best practices on the market. Only three percent of companies employ this method.

5. The Brontobytes Period

As Van Rijmenam [13] concluded: "Big Data Scientists will be in very high demand in the coming decades, but the real winners in the startup field will be those companies that can make Big Data so easy to understand, implement, and use that Big Data Scientists are no longer necessary. Large corporations will always hire Big Data Scientists, but the much larger market of Small and Medium Enterprises (SMEs) cannot afford to do this. Those startups that enable Big Data for SMEs without the need to hire experts will have a competitive advantage."

The algorithms matured by those Big Data startups will be ever smarter, smartphones will become better, and everyone will be capable to have a supercomputer in a pocket that may do daunting computing tasks in real time and visualize them on the small screen in a hand (Table 4) [14]. And, with the Internet of Things and the Industrial Internet and trillions of sensors, the quantity of data that requires to be treated by these gadgets will expand exponentially [13].

Table 4. *Big Data timeline [14].*

Items illustrating effective phenomena that arranged the field for the Big Data period	
1991	The Internet, or World Wide Web, is born. The protocol Hypertext Transfer Protocol (HTTP) becomes the standard means for sharing information in this new medium.
1995	Sun releases the Java platform. Java, invented in 1991, has become the second most popular language behind C. It dominates the Web applications space and is the de facto standard for middle‐tier applications. These applications are the source for recording and storing web traffic. Global Positioning System (GPS) becomes fully operational. GPS was originally developed by DARPA (Defense Advanced Research Projects Agency) for military applications in the early 1970s. This technology has become omnipresent in applications for car and airline navigation and finding a missing iPhone.
1998	Carlo Strozzi develops an open‐source relational database and calls it NoSQL. Ten years later, a movement to develop NoSQL databases to work with large, unstructured data sets gains momentum. Google is founded by Larry Page and Sergey Brin, who worked for about a year on a Stanford search engine project called BackRub.

Items illustrating effective phenomena that arranged the field for the Big Data period	
1999	Kevin Ashton, cofounder of the Auto‐ID Center at the Massachusetts Institute of Technology (MIT), invents the term "the Internet of Things."
2001	Wikipedia is launched. The crowd-sourced encyclopedia revolutionized the way people reference information.
2002	Version 1.1 of the Bluetooth specification is released by the Institute of Electrical and Electronics Engineers (IEEE). Bluetooth is a wireless technology standard for the transfer of data over short distances. The advancement of this specification and its adoption lead to a whole host of wearable devices that communicate between the device and another computer. Today nearly every portable device has a Bluetooth receiver.
2003	According to studies by IDC and EMC, the amount of data created in 2003 surpasses the amount of data created in all of human history before then. It is estimated that 1.8 zettabytes (ZB) was created in 2011 alone (1.8 ZB is the equivalent of 200 billion high‐definition movies, each two hours long, or 47 million years of footage with no bathroom breaks). LinkedIn, the popular social networking website for professionals, launches. In 2013, the site had about 260 million users. Wikipedia reaches 500,000 articles in February; seven months later it tops 1 million articles.
2004	Facebook, the social networking service, is founded by Mark Zuckerberg and others in Cambridge, Massachusetts. In 2013, the site had more than 1.15 billion users.
2005	The Apache Hadoop project is created by Doug Cutting and Mike Caferella. The name for the project came from the toy elephant of Cutting's young son. The now‐famous yellow elephant becomes a household word just a few years later and a foundational part of almost all Big Data strategies. The National Science Board recommends that the National Science Foundation (NSF) create a career path for "a sufficient number of high‐quality data scientists" to manage the growing collection of digital information.
2007	Apple releases the iPhone and creates a strong consumer market for smartphones.
2008	The number of devices connected to the Internet exceeds the world's population.
2011	IBM's Watson computer scans and analyzes 4 terabytes (200 million pages) of data in seconds to defeat two human players on the television show *Jeopardy!* (There is more about the show in Part Two.) Work begins in UnQL, a query language for NoSQL databases. The available pools in the IPv4 address space have all been assigned. IPv4 is a standard for assigning an Internet protocol (IP) address. The IPv4 protocol was based on a 32‐bit number, meaning there are 232 or 4.5 billion unique addresses available. This event shows the real demand and quantity of Internet connected devices. The Obama administration announces the Big Data Research and Development Initiative, consisting of 84 programs in six departments. The NSF publishes "Core Techniques and Technologies for Advancing Big Data Science & Engineering."
2012	IDC and EMC estimate that 2.8 ZB of data will be created in 2012 but that only 3% of what could be usable for Big Data is tagged and less is analyzed. The report predicts that the digital world will by 2020 hold 40 ZB, 57 times the number of grains of sand on all the beaches in the world. The *Harvard Business Review* calls the job of data scientist "the sexiest job of the 21st century."
2013	The democratization of data begins. With smartphones, tablets, and Wi‐Fi, everyone generates data at prodigious rates. More individuals access large volumes of public data and put data to creative use.

Big Data will only be bigger, and those brontobytes (Figure 4) will be frequent language in the boardroom. Fortunately, data storage will as well be more largely accessible and less costly. Brontobytes will be so frequent that probably the words "Big Data" will vanish. Big Data will be only data once more [13].

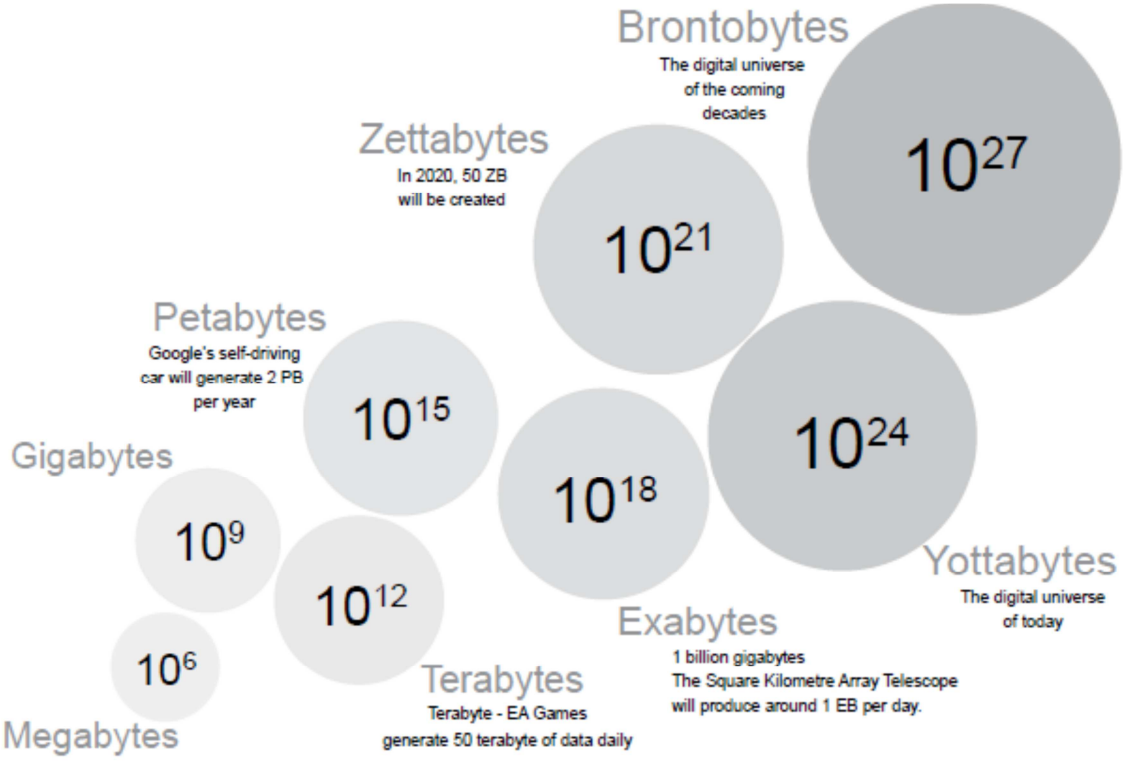

Figure 4. *Brontobytes infographic [13].*

Nevertheless, before this step may be attained, the increasing quantity of information [15] that is treated by institutions and governments will generate privacy problems (Figure 5). Those companies that adhere to moral regulations will remain alive; other companies that consider freedom from interference carelessly will recede from view, since freedom from interference will be self-regulating. Nevertheless, the issue will not disappear with governments, since citizens have not the ability to choose not to treat with them. Significant public discussions concerning the impacts of Big Data on client privacy are unavoidable [13, 16-18].

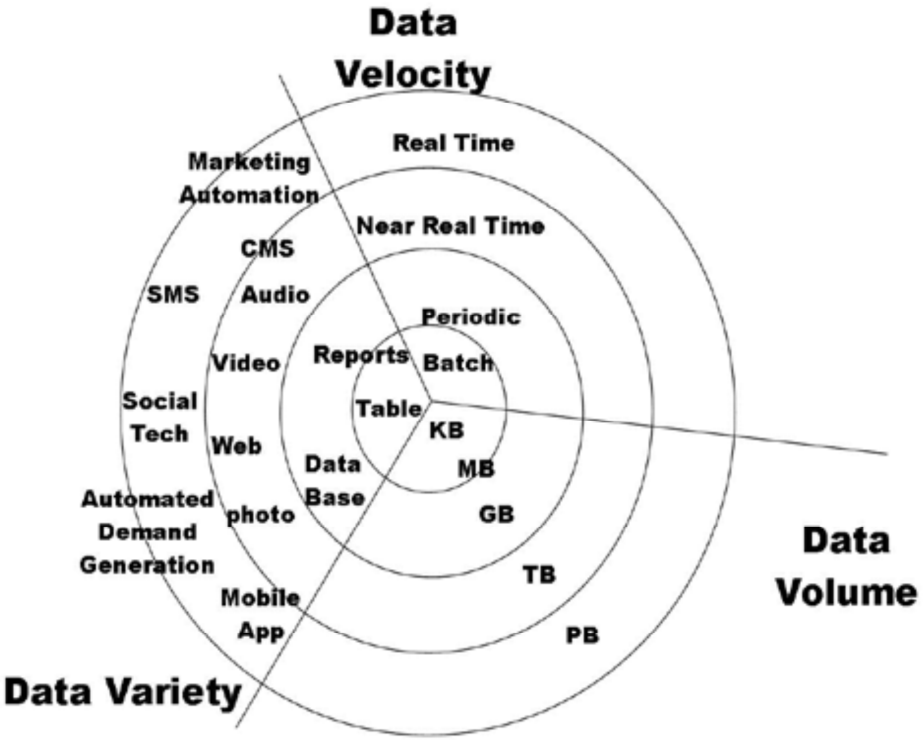

Figure 5. *Increasing data volume through applications [15].*

6. Conclusions

The main important points drawn from this review may be listed as:

The phenomena of the last 25 years have basically modified the fashion data is processed. More of it is generated each day; it is not a useless matter but a entombed mine requiring to be examined by inquisitive, inspired scientists and engineers who sense such tendencies and are arriving at satisfying the actual defies.

The significance of big data does not focus on how much data is possessed, but what things may be performed with it. Data may be extracted from any origin and examined to detect replies that let 1) cost decreases, 2) time decreases, 3) fresh product expansion and studied offerings, and 4) smart decision making.

As a magic, charming, and mysterious noun, Big Data remains an attractive novel field of both science and technology. Despite of the developed technology and open knowledge, Big Data still needs more familiarization and demystification. More developed computer skills, more than Excel, would be required to understand and touch its practical extent

References

[1] V. Kale, Big data computing, A guide for business and technology managers, Big Data Series, S. Ranka (Series Ed.), CRC Press, Taylor & Francis Group, A Chapman & Hale Book, Boca Raton, Florida, 2017.

[2] R. Boire, Data mining for managers, How to use data (big and small) to solve business challenges, Palgrave Macmillan, New York, 2014.

[3] N. Dunlop, Beginning big data with Power BI and Excel 2013, Apress, Springer Science+Business Media, New York, 2015.

[4] The world's technological capacity to store, communicate, and compute information, http://www.martinhilbert.net/WorldInfoCapacity.html/ (Accessed on 11/04/18).

[5] Wikipedia, Big data, https://en.wikipedia.org/wiki/Big_data#cite_note-1 (Accessed on 11/04/18).

[6] M. Paluszek, S. Thomas, MATLAB machine learning, Apress, Springer Science+Business Media, New York, 2017.

[7] D. Ghernaout, M. Aichouni, A. Alghamdi, Applying big data in water treatment industry: A new era of advance, Intern. J. Adv. Appl. Sci. 5 (2018) 89-97.

[8] K. B. Carter, D. Farmer, C. Siegel, Actionable intelligence, a guide to delivering business results with big data fast!, John Wiley & Sons, Inc., Hoboken, New Jersey, 2014.

[9] K. H. Pries, R. Dunnigan, Big Data analytics, A practical guide for managers, CRC Press, Taylor & Francis Group, An Auerbach Book, Boca Raton, Florida, 2015.

[10] A. Sathi, Big data analytics: Disruptive technologies for changing the game, MC Press Online, LLC, IBM Corporation, Boise, USA, 2012.

[11] A. Birkbak, H. B. Carlsen, The public and its algorithms, comparing and experimenting with calculated publics (Ch. 1), Algorithmic life, Calculative devices in the age of big data, L. Amoore, V. Piotukh (Eds.), Routledge, Taylor & Francis Group, New York, 2016.

[12] D. Feinleib, Big data bootcamp, what managers need to know to profit from the big data revolution, Apress, Springer Science+Business Media, New York, 2014.

[13] M. Van Rijmenam, Think bigger, developing a successful big data strategy for your business, AMACOM, American Management Association, New York, 2014.

[14] J. Dean, Big Data, data mining, and machine learning: Value creation for Business Leaders and Practitioners, John Wiley & Sons, Inc., Hoboken, New Jersey, 2014.

[15] V. Prajapati, Big Data analytics with R and Hadoop, Set up an integrated infrastructure of R and Hadoop to turn your data analytics into Big Data analytics, Packt Publishing, Birmingham, UK, 2013.

[16] T. Gurtzick, S. Al-Sefou, How we implemented (and secured) a big data microservices infrastructure, https://insidebigdata.com/2018/04/16/implemented-secured-big-data-microservices-infrastructure/(Accessed on 24/04/18).

[17] Big Data, What it is and why it matters, https://www.sas.com/en_us/insights/big-data/what-is-big-data.html (Accessed on 24/04/18).

[18] How big data helps you, https://www.microsoft.com/en-us/sql-server/big-data (Accessed on 24/04/18).