Big Data: The V's of the Game Changer Paradigm

Conference Paper · December 2016

DOI: 10.1109/HPCC-SmartCity-DSS.2016.0014

CITATIONS
56

READS
20,274

2 authors:

Ripon Patgiri
National Institute of Technology, Silchar
76 PUBLICATIONS **346** CITATIONS
SEE PROFILE

Arif Ahmed
Université de Rennes 1
24 PUBLICATIONS **1,121** CITATIONS
SEE PROFILE

2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems

Big Data: The V's of the Game Changer Paradigm

Ripon Patgiri, and Arif Ahmed
Department of Computer Science & Engineering
National Institute of Technology Silchar
Assam, India-788010
{*ripon, arif*}*@cse.nits.ac.in*
http://cse.nits.ac.in/rp/

***Abstract*—The Big Data is the most prominent paradigm now-a-days. The Big Data starts rule slowly from 2003, and expected to rule and dominate the IT industries at least up to 2030. Furthermore, the Big Data conquer the technological war and easily capture the entire market since 2009. The Big Data is blasting everywhere around the World in every domain. The Big Data, a massive amount of data, able to generate billions of revenue. The secret behind of these billions of revenue is ever growing volume.**

This paper presents the redefinition of volume of Big Data. The volume is redefined by engaging three other V's, namely, voluminosity, vacuum, and vitality. Furthermore, this paper augments two new V's to the Big Data paradigm, namely, vendee and vase. This paper explores all V's of Big Data. There are lots of controversy and confusion regarding V's of Big Data. This paper uncovers the confusions of the V family of the Big Data.

***Keywords*-Big Data, V family of Big Data, All V's of Big Data, Trends of Big Data, Redefinition of Big Data, Redefinition of Volume.**

I. Introduction

The Big Data technology, we say it as a game changer technology and it is the most popular buzzword. The Big Data is a game changer paradigm in any field and it has almost nothing untouched area, for instance, earth science, Genome, Oceanology, Aeronautical, Physics, and almost all fields where massive data are generated. Many researchers would like to develop low cost commodity hardware to perform High Performance Computation. Nevertheless, the data-intensive and task-intensive computation differs. Albeit, the aim and objective of both computations may be same, but Big Data refers about data-intensive tasks. However, the Big Data capable to engage thousands of researchers, capable to attract millions of audience, able to generate billions of revenue in a few years. What is the mystery of this magic? The answer is ever growing volume. Furthermore, the Big Data technology has always dealt with gigantic data size to store and process. For instance, the most popular Big Data technology, Hadoop can scale petabytes of data. Notwithstanding, the Hadoop technology has a limitation of scalability and the most of the researchers have already started working on the issue of infinite scalability. Dr. Hadoop [4], Xu et. al. [5], DROP [6], etc. are the few examples of this journey. However, there are many research challenges in Big Data technology and still, the Big Data technology is not mature enough to serve infinite scale of data. Notably, the Big Data technology can scale up to petabytes of data. Many technology has been introduced to solve the dilemma of Big Data, for instance, Hadoop stack (MapReduce, Cassandra, Hive, Mahout, HBase), Google File System, BigTable, CephFS, NoSQL (Giraph, Pregel, Mizan, MongoDB, CouchDB, Berkeley DB, DynamoDb, MemcachedDB, etc.), etc.

Everywhere data are growing. e-health and wearable technology, for instance. There are huge volume of data which is collected in the form of sensor data, weather data, video surveillance data, road traffic data, e-health, earthquake data, oil and natural gas data, atmospheric data and many more. According to IDC [27], the digital universal data will reach 44 zettabytes in 2020. Further, every year the size of data is doubled. In addition, data from embedded system contributes 10% of digital universal data in 2020 [27].

On the other hand, the volume is analogous to the Big Data. The Big Data concerns mostly on volume. However, there is lots of controversy and confusion about V's of Big Data. This paper exposes all the V family of Big Data. The V family of Big Data is abbreviated as $V_3^{11} + C$ where V_3 denotes voluminosity, vacuum, and vitality of volume. The V^{11} denotes all other V's of Big Data and C denotes complexity.

A. Motivation

The volume is major part of Big Data, and one can definitely state that if we remove volume from Big Data then the Big Data never be the big enough. It becomes a small set of data which is well fitted with conventional system to store and process.

"All V's of Big Data - Volume $\neq$ Big Data"

However, there many V's in Big Data paradigm. The 3V's of Big Data is defined by Doug Laney [1] in 2001 to describe the data management in 3-dimension. The 3V's where Volume, Velocity and Variety. Nowadays, many V's are added to the Big Data, that is, 6V's and 1C. The 6V's are Veracity[16], [21], [18], [23], [20],

978-1-5090-4297-5/16 $31.00 © 2016 IEEE
DOI 10.1109/HPCC-SmartCity-DSS.2016.8

Value [16], [21], [18], [23], [20], Validity [16], Variability [2], [18], [21], [23]/Volatility [16], Virtual [21], [2], Visualization[23]/Visibility, and Complexity[24]. There are lots of controversy and confusion among the V's of Big Data. There are many blog and website published about V's which are different from each other. Some confusing questions are listed below:

- **Controversy.** How many V's are there?
- **Information.** What are the widely accepted V's of Big Data?
- **Confusion.** Which are the correct V's of Big Data?

The most widely accepted V's are volume, velocity, variety, veracity and value. However, the other V's are also important for Big Data paradigm.

B. Contribution

The article present unique contribution to the Big Data architecture which is enlisted below.

- V_3**.** This article presents three more V to define the characteristics of volume in 3V's of Big Data defined by Doug Laney[1]. The three more V's of Volume are voluminosity, Vacuum, and vitality. Correspondingly, we emphasize on necessity vs demands of volume of Big Data.
- **Confusions.** There are many confusions of V's of Big Data. For instance, whether variability is correct or volatility. This article clear the all confusion with proper justification.
- **Two more V's.** This paper adds two more V's to the Big Data paradigm, namely, vendee and vase.
- **Insight.** The article presents in-depth insight on 9V+C of Big Data.

C. Organization

This article is organized as follows. The section II introduces the bigness of Big Data. The section III illustrate the volume of Big Data and redefine the volume in XIV. The section IV illustrate the velocity in terms of growth and transmission. In addition, the section V demonstrates the variety in terms of structured, unstructured and semi-structured data. The section VI, VII, X and VIII discuss about veracity, validity, virtual and value respectively. The section IX and XI justify the definition of confusion among visualization vs. visibility, and variability vs. volatility respectively. The section XV provides authors future vision on these new V's. And finally, the section XVI concludes the article.

II. Overview on Big Data

Cloud Computing and Big Data are associated with each other. Big data processing involves handling and processing of a petabytes of data or beyond. Big Data helps the user to use utility computing to leverage distributed queries over gigantic data size and in return gives the intended result. The integration of cloud computing and Big Data is appealing, and many research work has been conducted. Cloud computing is an enabler of Big Data technology. Moreover, the Big Data as a Service is the most prominent research field now-a-days. In addition, Internet of Things(IoT) is a new paradigm of computing where every devices are intelligent and connected to each other through sophisticated networking infrastructure. IoT is very popular both in research and academia in the recent years. The applications of IoT are ranges from Home appliance to military equipment and sensor devices. It allows to connect devices with low computing capacity to the central data center. Millions of such smart devices creates huge data daily. As a consequence, the data are growing very fast as well as the IoT. The features of IoT have impacted many applications, and of course, day to day life of people. The realization of computing outside the desktop environment can be achieved in the IoT paradigm [34]. In the year 2008, US National Intelligence Council reported that IoT will be a potential technology on the US interests 2025 [32]. In the year 2011, the number of devices interconnected had overcome the total population of the world [34]. In a survey reported in 2012, the number of IoT devices in the world was estimated to be 9 billion, and the growth was expected to reach around 21 billions by 2020 [33]. The growing number of IoT devices will be the main source of data in the future and it is growing as shown in the figure 4. The data collected from the IoT devices are used for data analysis and complex analysis using Big Data. But, the Big Data was defined by Doug Laney with 3D [1] in 2001.

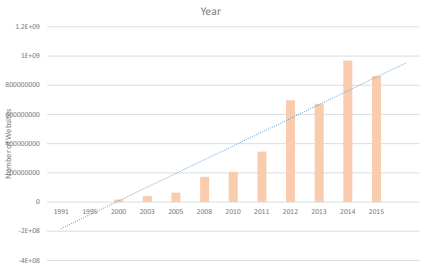

Figure 1. Growth of new websites [12]

On the other hand, the revenue is the key objective of a technology, from the perspective of IT industries. As shown in the figure 3, the revenue is growing linearly and it is a good indication of Big Data technology. The revenue of Big Data technology impact on the researchers and able to draw the attention of millions of audience. As per report [35], the Big Data software revenue will exceed $7 billion over hardware revenue by 2020. However, the report [35] expects compound annual growth rate (CAGR) of 14% growth in the Big Data investment by 2020. On the other hand, the Hadoop technology is forecast to grow at CAGR of 58% with worth

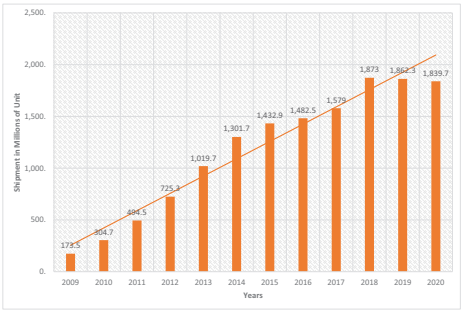

Figure 2. Shipment of smart phone [30]

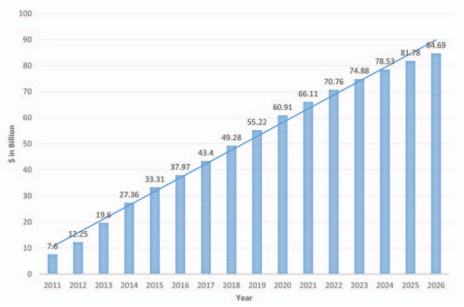

Figure 3. Big Data market forecast [26]

of $16 billion alone by 2020 [36]. This $ in billions able to attract more audiences. Therefore, the Big Data has close relation with $ and GB. The Big Data deals with conversion of GB to $. Consequently, there is a strong quest for defining Big Data more accurately. The characteristics and dilemmas of Big Data are addressed using these V's. Let us explore the most confusion and controversial terminologies of Big Data.

III. Volume

The volume is a huge set of data to be stored and processed [1]. The volume grows exponentially and it does not have any bound. The volume in the Big Data is large set of data which are very perplex to manage. There are lots of technology arise to manage these huge volumes of data set, but can the technology process beyond Exabytes?

Figure 4. Number of connected objects World wide [39]

The volume can span to Zettabytes, Yottabytes or beyond in the future. The technology must ensure to cope up with the growing size of data. The data are collected manually and automatically in the databases and data warehouse. These data are to be managed, processed, and served. The every bit of those data is very important for clients as well as to the companies. It is very difficult to store and manage 1 MB as well as 1 TB because those service providers have to ensure fault-tolerance, high-availability, disaster management, and security. Any time a machine can be faulty or crash, a link may fail and there is no certainty of natural calamities. Moreover, security takes vital a role in involvement of any kind of data. It is a very difficult task to provide service in an unsecured network environment. Security can be broken at any time, and therefore, invent a new security system that is to be broken. This phenomenon will continue. That's why, a bit of data also important in security system as well as 1TB of data or beyond. Now, let us come back to the volume of Big Data, where the data are unmanageable huge set of data. The volume is a state of being gigantic amount of data, and it is termed as big size, understood as unmanageable data and beyond the processing capacity of traditional system. It is really very hard to store and process in a conventional system. However, we further emphasize on volume in section XIII.

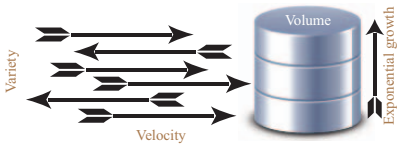

Figure 5. Relation among Volume, Velocity and Variety

IV. Velocity

The size of data is growing exponentially, and this velocity contributes a bigger database [1]. Data creates another data. The data always increases even if we use compressing technology. The velocity is always defined with respect to volume in Big Data. The velocity in Big Data concerns mainly two things, namely, speed of growth and speed of transfer. These two velocity requirements differ each other.

A. Growth

In 2009, the Facebook Inc. announced that the company has 1 Petabytes of data, and more interestingly, Google has 15 Exabytes of data in present day. The reason for data growing are enlisted below:

- **Users.** The internet users are growing daily basis and one user creates much amount of data. The internet users were 1000 million in 2005, and it was 3000 million in 2015 [3]. The users of internet are linearly growing. This implies exponential growth of data volume.

- **IoT.** Emerging of Internet of Things (IoT) is the prominent contributor to the Big Data growth. For instance, sensor devices, surveillance camera, RFID, etc. are the prominent data generator for Big Data. Moreover, the number of connected objects are increasing as shown in the figure 4. In addition, the number of smart phones are increasing as shown in the figure 2.
- **Cloud Computing.** Evolving of cloud computing generates huge data to store, process and manage.
- **Website.** Every day many new websites are launched. As shown in the figure 1, the number of websites are rising exponentially. The growth rate of the new website was highest in 2011 and 2014. The new spawned websites generate huge data and thus, the volume is the prominent field to do research.
- **Scientific Data.** The scientific data are naturally huge in size. For instance, seismic data, Ocean, Weather data etc.

B. Transmission

There is Big Data, then transferring data become a prominent issue. That is, when massive amount of data flies! The transmission of small volume does not have any problem, except security concern. The movement of large volume creates problem, even if there is a fibre-optics medium to transfer those data. If the distance increases, then the latency becomes a prominent barrier to transfer [8]. The bandwidth is also a barrier in the field of Big Data. The user requirements are always minimal latency with the highest bandwidth, which is not possible due to cost involvement. Therefore, Big Data technology also concerns about $ per GB.

V. Variety

Variety [1], [16], [21], [23] is the assorted forms of data, this includes structured, semi-structured, and unstructured data. The structured data comprises a tabular form of data, semi-structured data includes log data, XML data and unstructured data includes videos, images, scientific information. The big database comprises a variety of data [1]. Therefore, the volume grows at an exponential pace with a variety of data as shown in the figure 5. We know that everything is file. No variety. The data concerns with file only. Again, there are varieties of structured, semi-structured and unstructured data. For instance, there are .docx, .doc, .odt, .text etc. for textual data. There are a variety of data formats. However, these varieties of data comprise high volume of Big Data. The unstructured data has higher varieties of data than other two categories. The unstructured data contributes 90% of Big Data [13]. The unstructured data is raising daily and it become massive varieties in Big Data. The massive amount of varieties of data cause a big issue, either with low volume or high volume.

VI. Veracity

The veracity is accuracy, truthfulness, and meaningfulness [16], [21], [18], [23], [20]. The Big Data with huge volume becomes problematic when we would like to perform some operation on these data. The question is- how do we believe that the operation is successful? Is it accurate? The veracity is also a large problem in big data as it's rather impossible to spell check the huge quantity of information. Any data is worthless and meaningless if it is not accurate [23]. Inaccurate data may lead to wrong direction or decision. For instance, we would like to recommend some product to the users and accuracy matter a lot to revenue, but revenue is affected by an inaccurate results.

VII. Validity

On the other hand, though the process can perform a task accurately, but the data may not be valid [16]. For example, the very old data in e-commerce becomes obsolete and it can be truncated. However, some data never obsoletes. For instance, a transaction log record of a financial bank. Sometimes, the input of very old data is not valid for a process. The validity may differ from time to time. The validity refers to the data those have worthiness. The correct data may not be valid for certain processing. The huge volume of data, all are not valid at all. It depends on time as well. $Volume - Validity = Worthlessness$ [16]

VIII. Value

However, the Big Data is concerned mainly on extracting value from enormous stored data [16], [21], [18], [23], [20]. The Big Data extract values from data, reveal hidden truth from the data, uncover useful message from the data, creates value of data. Data in itself has no value. The massive sets of data are processed to give a worth, for instance, Big Data Mining. The Big Data Mining is nothing but a bigger dimension of Data Mining. The dump data are mined to search the hidden jewels. Therefore, the Big Data is a platform to provide worthiness of unworthy data. $BigData = Data + Value$ [16]

IX. Visibility or Visualization?

The visibility is the state of being able to see or be seen. The disparate data to be stitched together to show, either the data source is disparate or the data itself. There is no point to archive the data, if data cannot be seen or shown. The visibility defines the data to be viewed from a point of reference. The data may be visible from an administrator or other side, but data are made visible. On the other hand, visualization [23] is the process to show the hidden data of Big Data. This term is more precise to describe Big Data, because visualization is making visible. The Visualization is the most key process to enhance the performance of the data and business processes/decisions. In this investigation, we found that the key to success of

the business is the analysis using their own data and the visualization process gives the growth rate of the company's revenues. Thus, we are excluding the "visibility" term from Big Data "V" family. The visualization of large volume of data is a dilemma in Big Data [23]. For instance, how would we like to view the followers of Barack Obama in Twitter? The conventional technology never able to answer this query, because Barack Obama has $75,705,896$ followers [14]. In addition, the visualization of Big Graph is very big issue and it is very complex process.

X. Virtual

The virtual is found in the article [21], [2]. The data management itself a virtual process, and delineates the management of data. The virtual is a process to manage the data effectively and efficiently as per the demand of users. If we recall conventional operating system, then the virtual is the management of resources, for instance, demand paging. It also applies to Big Data. Moreover, the Big Data analytics visualize the required data which are purely virtual process. The role of Cloud Computing in Big Data is a tremendous contribution to growth of Big Data and this Cloud Computing evolves on the basis of virtualization. Therefore, the term "virtual" is accurate to describe the characteristics of Big Data.

XI. Variability or Volatility?

The variability [2], [23] and Volatility [16] are conflicting terminology that is wholly in the same boat. But, they conflict each other. The volatility is the nature of changing abruptly, sudden change, instability, and changed by unintentionally or anonymously. The volatile nature is superfluous in Big Data. On the other hand, the variability is the nature of changing, shifting, mutation, and modification with good intention. Therefore, the "variability" is more weighted and accurate to describe the characteristics of Big Data. The changes in the data may happen due to time, modification of user, obsolete data. For instance, Katy Perry has the highest followed person in twitter [14], but, after some time it may change. Moreover, suddenly one video on YouTube has got popular and access rate is increasing exponentially. After some time, access rate of this video falls down. This is the nature of variability.

XII. Complexity

The complexity is the pure form of computational term. There is a very high complexity in the case of a bulk volume of data [24]. The large volume of data is stored, processed and managed to serve to their intended clients. The handling huge volume of data always associates with high complexity.

XIII. Volume Redefinition

Time flies! Wait for none! Therefore, it is the right time to re-look the volume of 3V's of Big Data. The volume is defined in cube in general science, namely, height, width and length. As shown in the figure 6, the volume is 3D in Big Data. Another 3V's are added to the Big Data. These 3V's defines the characteristics of volume in Big Data. The volume is cog in the machine of Big Data. The Big Data and volume are analogous meaning, and thus the volume is integral part of Big Data. Further, the volume is the real backbone of the Big Data. However, the technology must support volume of the Big Data to store, process and manage. This is 100% sure that size of volume in the Big Data paradigm never going to decrease in near future. That is why, this the time to experiment the behavior of volume. Further, the 360-degree investigation of volume is always called for. Since, the volume is the main dilemma of Big Data to conquer. From the perspective of hardware, changing of the floppy disk to DVD and right now, the DVD is almost obsolete. Moreover, the SSD is emerging to replace the HDD. Different sort of RAM has been introduced. For instance, the RAMCloud [8], [9] has been introduced to overcome the latency issue [8], [9]. This technology transfer occurs only when the growth of volume in Big Data. Even, there are many issues in the data warehouse due to rising of volume [11]. On the other hand, the software perspective is also promoting the technology transfer by innovating new technology, and devising new algorithm. For instance, erasure coding creates more empty storage spaces [25].

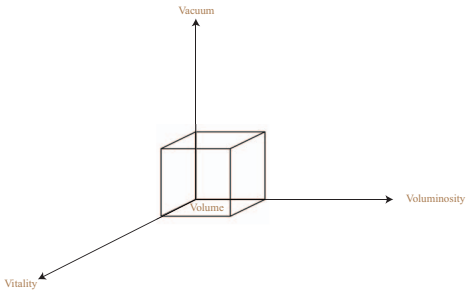

Figure 6. Re-definition of Volume of 3V's of Big Data.

A. Voluminosity

The voluminosity is the greatness of volume. The voluminosity in volume state that there is a very large set of data collected so far and to be collected. The data bank becomes unmanageable if data is extensive in size and hence, this is the time to do research on how to store, process and manage the mammoth size of data such that these data results revenue. Of course, the volume collected so far and to be collected has a significant gap. Surprisingly, the Google and NSA claims that they have 15 and 10 exabytes of data warehouse respectively [37], [38]. These data are collected

not to dump, but it is collected to generate revenue from these data. After a few years, many companies will bypass exabytes and this is the ground truth.

Let us explore the current trends of voluminosity. There are 7,238 Tweets sent in 1 second, 54,708 Google searches in 1 second, 124,563 YouTube videos viewed in 1 second, 2,497,667 Emails sent in 1 second, 2,158 Skype calls in 1 second, 35,651 GB of Internet traffic in 1 second, 726 Instagram photos uploaded in 1 second, and 1,127 Tumblr posts in 1 second [28]. Moreover, there are 1.09 billion daily active Facebook users on average for March 2016, 989 million mobile daily active Facebook users on average for March 2016, 1.65 billion monthly active Facebook users as of March 31, 2016, and 1.51 billion mobile monthly active Facebook users as of March 31, 2016 [29]. Another voluminosity is shown in the figure 2, where the millions of smart phones are shipped. These smart phones create not only data. Putting it other way, smart phones creates billions of photos. Moreover, connected object will increase and these object generates tremendous amount of data. The voluminous data is acquired, stored, processed and managed, and it is onerous task. However, the technological advent can partially solve the problem. Albeit, the real time processing of the Big Data is strenuous in the current state-of-the-art technology. By the same token, if data meet voluminosity, then it become more arduous to process in real-time.

B. Vacuum

The Big Data, without volume, it is just a small set of data, those can be processed by conventional way. The vacuum in volume state that there is a strong requirement of empty space to store large volume of data. Furthermore, vacuum also refers to creation of room for store, process and mange a tremendous data set from the existing data sets. This is the time to think about how much storage space available for incoming data rather than how much data we have stored. The process of creating storage space for incoming data is very challenging as well as managing these vast sets of data. The vacuum concerns with creating empty space, either augmenting storage devices or other techniques to reduce the size of data. For example, storage space can be created using erasure code in unused data. Moreover, we can apply compressing algorithm in infrequently used data as well as frequently used data[10]. The article [10], shows that processing of uncompressed data is more faster that compressed data. This de-duplication process improves the storage performance of the system. The opposite, the replication is the mechanism to duplicate a data block in multiple blocks to maximize the parallelism. Both, de-duplication and replication can also be combined to enhance the performance of a storage system [25]. However, big problems always come along with humongous set of data. The key point of research is that how to enhance a storage system to process gigantic amount of data. The exabytes of data is no farther from today. Let us step into the future today by concentrating on vacuum.

C. Vitality

The vitality is power of enduring, survive and grow. The vitality of a volume state that there is a massive amount of data actively served and remaining are not. Further, the vitality state about the data survival in big data in storage environment. That's reliability of data. In a large data bank, there is some data which are actively used and some are not. However, the company generates revenue from the actively used data only and the rest are stored in hope for future uses. Let us assume, a farm house having exabytes data and there is no vitality. Consequently, risk becomes higher and hence, anything can happen to those data. Anywhere anything can happen. Putting it other word, fire, earthquake, flood, war, and terrorist are the prominent reason for data loss. There is no big issue in a small amount of data. The more issues are augmented when the volume increases and reach it limit. Then, the vitality is the most prominent question. For instance, how do we rely on a system when there is no disaster management system? If not, then how do we implement on the colossal volume of data (beyond exabytes)? It is literally a billion dollar question. Undoubtedly, the disaster management system is always a prominent research issue. Apart from the disaster management system, fault-tolerance system also play a vital role in Big Data. However, ensuring vitality is a very big deal in tremendous data size. Moreover, the vitality describes about reliability, flexibility, dependability and security. Therefore, the vitality is an integral component of volume of Big Data.

XIV. Adding Two more V's

A. Vendee

The one more new V in Big Data is "Vendee" to define the client size associated with the Big Data to deal with. The Vendee is the most significant component to define the Big Data, where the 9V's are made only for clients to conform as their requirements. Moreover, the data and clients/users are ubiquitous. In short, business is client-driven process. Let's presume, the users per months in Facebook, and it is 1.01 billion. But, the access log of that billions of users becomes very large to handle and store. Not just that, the action performed by this user is enormous in data size. Furthermore, the Google search is 100 billion per month [12]. The users/clients aggrandize the companies. The crystal clear matter of fact is that Big Data involves big user/customer management. Unquestionably, the "Vendee" is very easy to define, interpret as a component of Big Data and the 9V's are under the umbrella of "Vendee". The trillions of user to deal with, exponential growth of users, and day-to-day basis requirement of the clients are also a crucial part of Big Data. Without "Vendee", does Big Data has a significance?

B. Vase

According to Merriam-Webster dictionary, the vase is a container, that is used for holding flowers or for decoration. The Big Data is high volume, and stored in the datacenter. The datacenter requires farmhouse, land, huge electric power, thousands of hardware, thousands of manpower, varieties of hardware, and many more small products to enable Big Data. In the Big Data paradigm, the flower refers to Big Data and the vase refers to underlying requirements to enable Big Data. The Google data centers alone consume 1.1% and 1.5% of Global electricity [22]. What happen to the companies which deals with the gigantic size of data? The exabytes of data warehouse requires not only many hardware to store, but also electric power, manpower, disaster recovery system, lands, farmhouse etc [38]. The Big Data cannot obviate from the vase. The vase is root of Big Data.

XV. Future Direction

As shown in the figure 1, 2, 3, and 4, we can easily guess the future of the Big Data without doing research on future of Big Data. The Big Data will at least dominate the market till 2030 and proven to be crackerjack technology. Since, new technologies are introduced day by day, which will generate data by any means. Therefore, data will continue to grow. It will never be stopped. Thus, the volume of data increases and will create a dilemma for the industries. This is the time to do the research on the $V_3^{11} + C$ of Big Data to develop new technology, since it is the most current demands of Big Data, like IoT. The future, we can create it, but not always. In the technological sense, most of the future, we predict it first, and then create the future. Nevertheless, the volume is growing exponentially, need not to imagine today, and it is clearly "visible future" from today. The voluminosity, vacuum and vitality will be the most prominent research area in the near future as well as vendee and vase. Right now, most of the researchers have begun the journey in this direction.

XVI. Conclusions

The volume of Big Data takes lion's share in revenue without whom the Big Data does not exist. However, we have demonstrated 9V's of Big Data and engage new more V's to enhance the Big Data technology. The V family ($V_3^{11} + C$) is worthy in the sense of future technology and stepping towards the future right now. And finally, we conclude that the $V_3^{11} + C$ is a milestone in developing future technology of Big Data and cannot be imagined a new technology without this V's of Big Data.

References

[1] Doug Laney, "3D Data Management: Controlling Data Volume, Velocity, and Variety", Gartner, file No. 949. 6 February 2001, http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

[2] C.L. Philip Chen, and Chun-Yang Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data", Information Science, 275(2014), pages 314347, DOI: http://dx.doi.org/10.1016/j.ins.2014.01.015

[3] Number of internet users worldwide from 2005 to 2015 (in millions), [online], Retrieved on 11 June 2016 from http://www.statista.com/statistics/273018/number-of-internet-users-worldwide/

[4] Dipayan Dev, and Ripon Patgiri, "Dr. Hadoop: an infinite scalable metadata management for HadoopHow the baby elephant becomes immortal", Frontiers of Information Technology & Electronic Engineering, January 2016, 17(1), pp 15-31, DOI: http://dx.doi.org/10.1631/FITEE.1500015

[5] Quanqing Xu, Rajesh Vellore Arumugam, Khai Leong Yong, and Sridhar Mahadevan, "Efficient and Scalable Metadata Management in EB-Scale File Systems", IEEE Transactions on Parallel and Distributed Systems, 25(11), pages 2840 - 2850, 2014, DOI: http://dx.doi.org/10.1109/TPDS.2013.293

[6] Quanqing Xu, Rajesh Vellore Arumugam, Khai Leong Yang, Sridhar Mahadevan, "DROP: Facilitating distributed metadata management in EB-scale storage systems", In 2013 IEEE 29th Symposium on Mass Storage Systems and Technologies (MSST), 6-10 May 2013, pages 1-10, DOI: http://dx.doi.org/10.1109/MSST.2013.6558422

[7] Jeffrey Dean and Sanjay Ghemawat, "MapReduce: simplified data processing on large clusters", In OSDI'04: Sixth Symposium on Operating System Design and Implementation, San Francisco, CA, December, 2004.

[8] John Ousterhout, Parag Agrawal, David Erickson, Christos Kozyrakis, Jacob Leverich, David Mazires, Subhasish Mitra, Aravind Narayanan, Diego Ongaro, Guru Parulkar, Mendel Rosenblum, Stephen M. Rumble, Eric Stratmann, and Ryan Stutsman, "The case for RAMCloud" Communications of the ACM, 54(7), July 2011, DOI: http://dx.doi.org/10.1145/1965724.1965751

[9] John Ousterhout, Arjun Gopalan, Ashish Gupta, Ankita Kejriwal, Collin Lee, Behnam Montazeri, Diego Ongaro, Seo Jin Park, Henry Qin, Mendel Rosenblum, Stephen Rumble, Ryan Stutsman, and Stephen Yang, "The RAMCloud Storage System", ACM Transactions on Computer Systems (TOCS), 33(3), September 2015, Article No. 7, DOI: http://dx.doi.org/0.1145/2806887

[10] Adnan Haider, Xi Yang, Ning Liu, Xian-He Sun, and Shuibing He, "IC-Data: Improving Compressed Data Processing in Hadoop", 2015 IEEE 22nd International Conference on High Performance Computing (HiPC), 16-19 Dec. 2015, pages 356 - 365, DOI: http://dx.doi.org/10.1109/HiPC.2015.28

[11] Alfredo Cuzzocrea, "Warehousing and Protecting Big Data: State-Of-The-Art-Analysis, Methodologies, Future Challenges", In ICC '16: Proceedings of the International Conference on Internet of things and Cloud Computing, Artcile nmber. 14, 2016, DOI: http://dx.doi.org/10.1145/2896387.2900335

[12] "Internet Live Stats", [online], Retrieved on 11 June 2016 from http://www.internetlivestats.com/total-number-of-websites/

[13] Mark van Rijmenam , "Why The 3V's Are Not Sufficient To Describe Big Data", [Online], Accessed on 20 June 2016, from https://datafloq.com/read/3vs-sufficient-describe-big-data/166.

[14] Twitter Counter, "Twitter Top 100 Most Followers", [online], Accessed on 27 June 2016 from http://twittercounter.com/pages/100

[15] Gema Bello-Orgaz, Jason J. Jung, and David Camacho, "Social big data: Recent achievements and new challenge", Information Fusion, 28(March), pages 45-59, 2016 DOI: http://dx.doi.org/10.1016/j.inffus.2015.08.005

[16] M Ali-ud-din Khan, M F Uddin, and N Gupta, "Seven Vs of Big Data: Understanding Big Data to extract Value", In 2014 Zone 1 Conference of the American Society for Engineering Education (ASEE Zone 1), pages 3-5, April, 2014, DOI: http://dx.doi.org/10.1109/ASEEZone1.2014.6820689

[17] Cheikh Kacfah Emani, Nadine Cullot, and Christophe Nicolle, "Understandable Big Data: A survey", Computer Science Review, 17, pages 70-81, 2015, DOI: http://dx.doi.org/10.1016/j.cosrev.2015.05.002

[18] Yuri Demchenko, Cees de Laat, and Peter Membrey, "Defining Architecture Components of the Big Data Ecosystem", In 2014 International Conference on Collaboration Technologies and Systems (CTS), pages 104 - 112, 2014, DOI: http://dx.doi.org/10.1109/CTS.2014.6867550

[19] Jianzheng Liu, Jie Li, Weifeng Li and Jiansheng Wu, "Rethinking big data: A review on the data quality and usage issues", ISPRS Journal of Photogrammetry and Remote Sensing, In Press, DOI: http://dx.doi.org/10.1016/j.isprsjprs.2015.11.006

[20] Xiaolong Jin, Benjamin W. Wah, Xueqi Cheng, and Yuanzhuo Wang, "Significance and Challenges of Big Data Research", Big Data Research, 2(2), pages 59-64, 2015, DOI: http://dx.doi.org/10.1016/j.bdr.2015.01.006

[21] landmark.solutions, "The 7 pillars of Big Data", A White Paper of Landmark Solutions, Retrieved on 10, June, 2016 from https://www.landmark.solutions/Portals/0/LMSDocs/Whitepapers/The_7_pillars_of_Big_Data_Whitepaper.pdf

[22] Inhibitant, "INFOGRAPHIC: How Much Energy Does Google Use? ", [online], Accessed on 28 June 2016 from http://inhabitat.com/infographic-how-much-energy-does-google-use/

[23] Eileen McNulty, "Understanding Big Data: The Seven Vs", [online] Retrieved 10, June, 2016 from http://dataconomy.com/seven-vs-big-data/

[24] Monica Bulger, Greg Taylor, and Ralph Schroeder, "Engaging Complexity: Challenges and Opportunities of Big Data", In London: NEMDOE, 2014.

[25] Mingyuan Xia, Mohit Saxena, Mario Blaum, and David A. Pease, "A tale of two erasure codes in HDFS", In Proceedings of the 13th USENIX Conference on File and Storage Technologies, Pages 213-226, 2015.

[26] Louis Columbus, "Roundup Of Analytics, Big Data & Business Intelligence Forecasts And Market Estimates, 2015", [online], Retrieved on 11 June 2016 http://www.forbes.com/sites/louiscolumbus/2015/05/25/roundup-of-analytics-big-data-business-intelligence-forecasts-and-market-estimates-2015/#16f1c4834869

[27] "The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things", [online], Retrieved on 11 June 2016 from http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm

[28] ILS, "Internet Live Stats", Retrieved on 12 June 2016, at 2:31AM from http://www.internetlivestats.com/one-second/#tweets-band

[29] Facebook, "Facebook Newsroom", [online], Retrieved on 12 June 2016 from http://newsroom.fb.com/company-info/

[30] IDC, "Global smartphone shipments forecast from 2010 to 2020 (in million units)", [online], Retrieved on 12 June 2016 from http://www.statista.com/statistics/263441/global-smartphone-shipments-forecast/

[31] Rajkumar Buyya, Chee Shin Yeo, and Srikumar Venugopal, "Market-oriented cloud computing: Vision, hype, and reality for delivering it services as computing utilities", In 10th IEEE International Conference on High Performance Computing and Communications, pages 5 - 13, 2008, DOI: http://dx.doi.org/10.1109/HPCC.2008.172

[32] SRI Consulting Business Intelligence, "Disruptive Civil Technologies: Six Technologies with Potential Impacts on US Interests out to 2025", [online], Retrieved on 12 June 2016 from https://fas.org/irp/nic/disruptive.pdf

[33] Nathan Eddy, "Gartner: 21 Billion IoT Devices To Invade By 2020", [online], Retrieved on 12 June 2016 from http://www.informationweek.com/mobile/mobile-devices/gartner-21-billion-iot-devices-to-invade-by-2020/d/d-id/1323081

[34] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions", Future Gener. Comput. Syst., 29(7), pages 1645-1660, 2013, DOI: http://dx.doi.org/10.1016/j.future.2013.01.010

[35] "Big Data Market 2016 to 2030 - Company Profiles and Strategies of 150 Big Data Ecosystem Players", Retrieved on 12 June 2016 from http://finance.yahoo.com/news/big-data-market-2016-2030-201700635.html

[36] "Hadoop Market Forecast 2017-2022", [online], Retrieved on 12 June 2016 from http://www.marketanalysis.com/?p=279

[37] Follow the Data, "Data Size Estimates", Retrieved on 10 June 2016 from https://followthedata.wordpress.com/2014/06/24/data-size-estimates/

[38] James Zetlen, "Google's datacenters on punch cadrs", [online], Retrieved on 10 June 2016 from https://what-if.xkcd.com/63/

[39] Charles McLellan, "The internet of things and big data: Unlocking the power", [online], Retrieved on 10 jun 2016 from http://www.zdnet.com/article/the-internet-of-things-and-big-data-unlocking-the-power/