The 51 V's Of Big Data*

Survey, Technologies, Characteristics, Opportunities, Issues and Challenges

Nawsher Khan†, Arshi Naim
College of Computer Science
King Khalid University
Abha, Saudi Arabia
nawsher@kku.edu.sa
arshi@kku.edu.sa

Mohammad Rashid Hussain,
Quadri Noorulhasan Naveed
College of Computer Science
King Khalid University Abha, KSA
humohammad@kku.edu.sa,
qnaveed@kku.edu.sa

Naim Ahmad,
Shamimul Qamar,
College of Computer Science
King Khalid University Abha, KSA
nagqadir@kku.edu.sa,
drsqamar@rediffmail.com

ABSTRACT

Currently Big Data is the biggest buzzword, and definitely, we believe that Big Data is changing the world. Some researchers say Big Data will be even bigger buzzword than the Internet. With fast-growing computing resources, information and knowledge a new digital globe has emerged. Information is being created and stored at a fast rate and is being accessed by a vast range of applications through scientific computing, commercial workloads, and social media. In 2018, over 28 billion devices globally, are connected to the internet. In 2020, more than 50 billion smart appliances will be connected worldwide and internet traffic flow will be 92 times greater than it was in 2005. The usage of such a massive number of connected devices not only increase the data volume but also the velocity of data addition with speed of light on fiber optic and various wireless networks. This fast generation of enormous data creates numerous threats and challenges. There exist various approaches that are addressing issues and challenges of Big Data with the theory of Vs such as 3 V's, 5 V's, 7 V's etc. The objective of this work is to explore and investigate the status of the current Big Data domain. Further, a comprehensive overview of Big Data, its characteristics, opportunities, issues, and challenges have been explored and described with the help of 51 V's. The outcome of this research will help in understanding the Big Data in a systematic way.

CCS CONCEPTS

• Information systems → Database Design and Models

KEYWORDS

Big Data, data generation, data characteristics, data storage

ACM Reference format:

† Nawsher Khan, Arshi Naim, Mohammad Rashid Hussain, Quadri Noorulhasan Naveed, Naim Ahmad, Shamimul Qamar. 2019.

• The 51 V's Of Big Data: Survey, Technologies, Characteristics, Opportunities, Issues and Challenges. In *Proceedings of ACM Omni-layer Intelligent Systems Conference (COINS'19). ACM,* Heraklion, Crete, Greece *6 pages.* DOI: 10.1145/3312614.3312623

1 INTRODUCTION

According to a recent estimation, 2.5 quintillion bytes of data are being generated daily [1]; and this big generation pushes the world towards the big ocean of Big Data. Huge amounts of data are being collected by many and various organizations on an ongoing continuous basis. These datasets are being stored from many sources, including but not limited to the World Wide Web (WWW), Sensor Networks and Social Networks [2]. IoE (Internet of Everything) is the main source, which combines products, people and processes to generate even more connectivity.

Big data considered the key to exposing the next big waves of growth in production. The recent amount of collected data in our world has exploded due to a number of new applications and technologies, which permeate our day-to-day lives. These applications and technologies include mobile and social networking applications, and IoE based smart-world's smart systems (smart grid, smart transportation, smart cities etc.) [3], [4]. These smart every-things generate a data each second which becomes the part of this huge flood of Big Data. Big Data has big potential and this is reality as long as the amount of data itself does not become part of the issue [1], [2]. Many writers and researchers define that Big Data is a shape of data that excels the processing capability of traditional database infrastructure or data engines. For every business to implement Big Data technologies and tools, it is essential to know what makes the Big Data's ocean. Nowadays, the research has gotten attention to direct, support, encourage, understand, process, clean, analyze and to utilize Big Data for any specific purposes. It was believed, in the infancy stage of Big Data, that the main contributors to Big Data were the scientists, simulations, physicists, military experiments, NASA and supercomputers but now Oracle talks about three various types of Big Data. A certain number of new technologies which are integrated into our daily lives, such as mobile and social networking applications, and Internet-of-Everything (IoE) based smart-world systems, massive amounts of data will be collected [3], [4]. Great research has done with Big Data, but still there are a number of major challenges, including data collection, storage, updating, analysis, sharing and others but this work, which is the extended version of [5] and [6], explores the 51 main "V's", describing the Characteristics, Opportunities, and Challenges of the Big Data.

Table 1: Types of Data Generation Sources

Data Source Types		Examples
1	Traditional Enterprise Data	CRM System, Finance & Accounting, ERP Data, Human Resources, Management Accounting, Engineering, Manufacturing, Order Processing, Supply Chain Management, Data Services etc.
2	Machine/ Sensor Generated Data	Calls Records, Manufacturing Sensors, Weblogs, Trading Systems Data, Financial Instrument Trades, Web Server Logs, Network Event Logs, SEIM logs, Web Logs, Telemetry, Computer, Network, and other equipment Logs, RFID chip Readings, Satellite, GPS System, Temperature and other Environmental Sensor Readings, Sensor Readings from Factories, Medical Devices etc.
3	Social Data	Facebook, Emails, Twitter, Blogs, Customer Feedback, Messenger, Reviews, Instagram, Google+, WhatsApp, LinkedIn, Myspace, Pinterest, Tumblr, Snapchat, Viber, VK, WeChat, Wikia, Baidu Tieba, QQ, QZone, Tik Tok, Sina Weibo, Reddit, Skype, Telegram, Tinder etc.

Big Data refers to datasets whose size exceeds the ability of typical database software tools to capture, manage, store, and analyze [7]. Gartner, 2012, describes Big Data as "High Volume, High Velocity, and/or High Variety information assets that need new forms of processing to enable enhanced decision-making, insight discovery, and process optimization".

2 BIG DATA TECHNOLOGIES

According to the Kirk Borne, Principal Data Scientist, Big Data Definition is Everything, Quantified, and Tracked [9].

Everything: Every aspect of life i.e. work, entertainment, consumerism, and play is now recognized as a source of digital information about us, our world, & everything we may come across.

Quantified: "Everything" somewhere, mostly in digital form, often as numbers, but not always in such formats. The quantification of features, characteristics, patterns, and trends in all things is enabling Data Mining, Machine Learning, Statistics, and Discovery at an unprecedented scale on an unprecedented number of things. The IoT is just one example, but the IoE is even more inspiring.

Tracked: We do not directly quantify and measure everything just once, but we do so continuously. It includes tracking sentiment, purchase logs, web clicks, social media history, geolocation, etc. or tracking every car on the road, or every motor in a manufacturing plant or every moving part on an airplane, etc.

3 BIG DATA ADVANTAGES

Table 2: BIG DATA ADVANTAGES

• Smarter Decisions	• Customer Centric Products
• Greater Knowledge	• Increased Customer Loyalty
• Better Products	• Data-to-Discovery,
• Optimal Solutions	• Data-to-Decisions
• Deeper Insights	• Data-to-Dollars

4 BIG DATA DEFINITIONS AND V'S

Big data has the following characteristics:

- Very large, Petabytes/Exabyte/Zettabytes of Data
- Millions/Billions of people/users using and generating
- Billions/Trillions of records every-day
- Flat Schemas with few complex interrelationships
- Often involving time-stamped events
- Loosely Structured, distributed, incomplete & inaccessible
- Often including connections between data elements that must be probabilistically inferred

The motivation to write this paper is a fact that Big Data offers raw components to build tomorrow's great machines. We have created 90 % of all data in the last 2 years and every day we are adding 2.5 quintillion bytes of data to the ocean of data [8]. According to Bonnie Feldman [8] exact wordings are as "Big Data is the fuel and oil. If you leave it in the ground, it does not have a lot of value. But when we need to find ways to ingest, curate, and analyze the data in new and various ways, i.e. such as in Watson, Big Data becomes very interesting". Hence, as a researcher, we should not leave Big Data on the ground but it is the time to analyze, explore, discuss, and understand all V's of Big Data in this paper, which will open doors towards finding the true value of the Big Data.

4.1. VOLUME: Size, Scalability, Vast Amount

Data volume refers to the large data set generated by science and education, business, technology and human interaction records. Volume refers to the size of data being generated from all the sources including text, audio, video, research studies, social networking, space images, medical data, weather forecasting, crime reports, natural disasters etc. However, data volume plays a significant role in storage and processing. Although, the storage capacity introduces less challenge in comparison with processing, due to rapidly developing storage technologies on one hand and decreasing the storage price on the other hand. Therefore, cost-effective storage solutions, Cloud technologies and now Edge technologies offer opportunities for organizations to store data. Data volume, however, has a fundamental influence on data processing, data management and decision-making. Because the size of data has been growing faster in comparison with the computational power of processing system [10], [11].

4.2. VELOCITY: Generation Speed

Velocity deals with the speed at which data flows in from different sources like networks, machines, human interaction, social media sites etc. The movement of data can be enormous or continuous. There are two ways that data can be imported: batch data and streaming data. Stream processing is key for choosing Big Data analysis platform, because real time process most often is time sensitive and demands faster and near-instant analytics results. i.e. Hadoop performance is good for processing archive data in a batch model, while Apache Spark is more efficient for real-time and interactive job analysis [12]. Sometimes 5 seconds is too late. For time-sensitive processes like catching fraud etc., Big Data must be used as it streams into the enterprise to increase its value. Five

million trade events and actions are scrutinized to find potential fraud daily. Fraud detection and credit card are good examples where millions of transactions are checked for unusual patterns in almost real-time. Five hundred million daily call detail records are being analyzed in real-time to predict customer churn quickly. Velocity has the categories of Batch, Near Time & Real Time [5], [6].

4.3. VARIETY: Heterogeneity, Multiple Data Sets

Data Variety refers to the degree of data organization. Unstructured data lacks enough degree of organization, while the structure organization has a high degree [13]. Data Variety is to measure the richness and fruitfulness of the data representation, which are in text, audio, video, images etc. forms. From an analytic viewpoint, it is probably the main obstacle to effectively using huge volumes of data. The appearance of Data in many shapes brings the real complexity. However, unstructured and semi-structured data are more difficult to analyze and make a decision. RDBMS has been used by traditional data analysis systems. These RDBMSs have required expensive hardware and only apply to structured data [14].

4.4. VERACITY: Quality, Truthfulness, Meaningfulness

Data Veracity refers to the truthfulness and reliability of data. IBM and Microsoft have introduced Veracity and Variability [16] & [17] which adds value as another dimension. Almost we are talking and taking the beneficial results from data for a given problem space, on which we are working to explore. Data veracity focuses on data quality and accuracy and defines how data can be trusted when important decision needs to be made regarding the collected data. Data is categorized in good, bad, or undefined data, which could be because of data inconsistency, incompleteness, ambiguity, latency, deception, and approximations [18].

4.5. VALUE: Usefulness, Mining

Everyone always has interest to mine & maximize the Value of any Big Data set. Data Value refers to data usefulness in decision-making and is one of the most significant factors in Big Data, because it has direct impact on business profits. For instance, McKinsey [19] estimated a range of health-care initiatives and asserted the 53 potential impacts could account for 300 billion to 450 billion in reduced health-care spending, or 12 to 17 percent of the 2.6 trillion baselines in US health-care costs if the early successes were scaled up to create system-wide impact. In addition, Steve pointed out in [20] that organizations agree to use business information and analytics differentiates in industry as top performers and low performers. Thus, the value lies in the meticulous analysis of accurate data.

Big Data is huge information assets that demand cost-effective and innovative forms of information processing to enhanced insight for decision making [21] ''. Although, this definition is not so perfect to give us an explicit distinction and scale from high to low value. By this definition, we cannot extract the value of a dataset. Thus, an evaluation-based definition is necessary for capturing the Value of data. Obviously, capturing and storing huge amounts of data is not the goal for all organizations and companies. However, all of them have the interest to analyze the data to extract and generate real commercial value [22]. Davenport [23] listed some real and anecdotal examples of how organizations design strategies for using collected datasets and mine value from those datasets. Moreover, a comprehensive research from [24] indicated that data driven decision-making can produce better performance over other decision making methods in terms of profitability and productivity. A number of researchers [25], [26] have identified the issues that how Big Data analytics extract and obtained commercial value. Some organization can take high cost with storage associated at higher tiers as protection is better guaranteed at those levels and as a result the value to cost ratio will be higher [28], [29].

4.6. VALIDITY: Excellency, Governance, Understandability

The validity of data might be having the same ideas with Veracity of data. However, they don't have the same concepts and theories. When the status of data changes from exploratory to actionable, data should be valid. In other words, a data set may not have veracity problems but may not also be valid if not properly understood and accepted. The validity of Big Data is essential to find the presence of hidden relationships among elements within huge Big Data generation sources.

4.7. VOLATILITY: Durability, Lifetime, Availability

Big Data Volatility refers to the life duration that for how long time data is Valid and for how long time it should be stored [30]. In this domain, for real time data, we need to determine at what point and when is data no longer relevant and applicable to the current analysis. For some sources, the data should always be present there, but for others, may be this is not the case. Therefore, this is need to understand the requirements, availability and lifetime of data. In a data standard setting, data are kept for decades to build an understanding of importance of data.

4.8. VIABILITY: Activeness

The meaning of Viability is that the Big Data should have the capability to be live and active forever, and able for developing, and to produce more data when needed. However, we need more to analyze huge data sets in real time, for which should carefully select the attributes and factors that are most likely to predict outcomes that matter most to businesses. Using Big Data, we are collecting multidimensional data, not simply collecting a large number of records, but spans a broadening array of variables.

4.9. VISUALIZATION: Attractive, Smartness

Data Visualization and Interpretation refers to its attractiveness. Visualization not only enables users or decision-makers to gain better insights into Big Data, but also is important to understand and analyze Big Data in remote sensing to bring out data details relevant for the current aims or objectives. In order to effectively use Visualization, remote sensing Big Data should be aggregated from diverse sources in a huge Volume, and imported to a model, which allows decision making in minutes rather than weeks or months. This is a big challenge for Pete Bytes level or larger volume of data inputs, for instance, in applications related with hazard monitoring. Therefore, Visualization of big remote sensing data should deal with challenges of large data Visualization as well as interactive exploration of data for an improved understanding. There are two categories, Data Visualization tools for presentations and data Visualization tools for developers. The 20 best tools for Big Data visualization. Tableau, Infogram, ChartBlocks, Datawrapper, RAW, Visual.ly, Processing.js, Plotly, NVD3, Leaflet, D3.js Ember, Charts, FusionCharts, Polymaps, Highcharts, Chartist.js, Chart.js, n3-charts, Google Charts, and Sigma JS.

4.10. VERIFICATION: Authenticity, Desired Outcome

Verification of Big Data refers to the authenticity and desired results of data processing. Everyone always has the interest to extract the highest value from any dataset, to have true and real value of the data. Furthermore, data value must exceed its cost, management or ownership. Stakeholder should pay attention to the investment of data storage. Storage may be cost-effective and relatively cheaper at time of purchase but such under-investment may harm highly valuable data, i.e. storing clinical data for a new drug on cheap and unreliable storage may save money today but can put data on risk tomorrow [31].

4.11. VERBOSITY: Loquacity, Garrulity, Volubility

Big data is a huge data, which comes from various sources they may be structured or unstructured data, and good or bad data. Bad data refer to the information, which is wrong, incomplete or out of date. The effects of storing these types of information and data may be risky sometimes. Therefore, it is recommended to check that the stored data is secured, complete, relevant, and trustworthy.

4.12. VERSATILITY: Adaptable, Alterable, Polytrophic

Big data is evolving to satisfy the desired requirements of various organizations, researchers, and Government as well. It facilitate the urban planning, visualization, environment modelling, quality classification, analysis, computational analysis, securing environment and manufacturing process required by organizations with cost-effective models and sophisticated exploration of the result.

4.13. VOLUNTARINESS: Permissive, Discretional

Big data is a big ocean of big amount of data. This big ocean should be free for volunteer usage by different organizations without any restrictions and without any interference. Voluntaries of Big Data refers to its assistance retailers by giving them knowledge of customer preferences, urban planning by visualization of environment. Big Data voluntarily help numerous enterprises. It modelling and traffic patterns, manufacturers by predicting product issues to optimize their productivity and to improve the equipment and customers performance. Those companies who producing energy improving efficiency by reducing the losses, healthcare's professionals to prevent diseases and improving patient health [32]. Research organizations improving efficiency to obtain quality of research and revolutionize life science, medical science, physical science and scientific research [33], [34], financial service organizations improving efficiency to identify and prevent fraud, government agencies to expand services in their respective fields. All these collaborators need the voluntary behavior of the Big Data.

4.14. VIRTUALIZATION: Effective, Implied, Implicit

Data Virtualization refers to an approach of data management, which allows an application to retrieve and manipulate data with no technical details about the data i.e. how data is formatted at source, or where it is physically located [35]. This approach should also be able to provide a single customer view of the overall data [36]. Unlike the traditional extract, transform, load process, the data remains in place, and real-time access is given to the source system for the data. This approach reduces the risk of data errors in the workload moving data around that may never be used, and it does not attempt to impose a single data model on the data. The technology also supports the writing of transaction data updates back to the source systems [37].

4.15. VARIABILITY: Dynamic, Evolving Behavior in Sources

Data variability refers to inconsistent data flow, and can be shown at times. This property has become challenge by increasingly usage of digital media, which is the main root of peak in data loads [38].

4.16. VISCOSITY: Complexity

Complex data management can be very complex for huge data sets especially when they come from different sources, because linking, matching and transformation are significant activities in Big Data management [39]. Complexity of Big Data deals with the degree of correlation and interdependencies in big data structures such that small changes can have significant effect in system's behavior or might not apply any changes at all [40].

4.17. VOCABULARY: Models, Semantics, Structures

Big Data is everywhere, spanning the entire range of academic research. It will be from humanities to natural sciences, from social sciences to engineering and medicine. We are bound to come across copious amount of data; this phenomenon is the outcome of modern technology, which allows us to collect, measure and sampling. The challenge is to transform Big Data into actionable, useful information. This requires a range of tools from mathematics, statistics and computer science, methodologies that might appear intimidating to the uninitiated. Bernard Marr has given a short presentations [41] introducing the Vocabulary of Big Data, the range of concepts and ideas, which underlie modern analysis of large data sets.

4.18. VENUE: Distributed, Heterogeneous, Multiple

Online Social Networks (OSN) data will soon include geographic data along with OSN interaction, i.e. geo-tag real-time geographic data [42]. Location-based data will soon expand beyond terrain. One study addresses the gauntlets of major forms of technology for three-dimensional (3-D) interaction and volume rendering technology based on graphics processing unit (GPU) technology. This work explores visual software for the hydrological environment based on data orientation. In addition, it produces ocean plans, contour mapping of surfaces, element field mapping, and dynamic simulation of the existing field [43], [44].

4.19. VIOLATION (In privacy): Crimes, Terrorist Activities

Big data applications have helped analyze and solve many data-science problems for businesses and governments as well. Governments have used Big Data applications to identify criminals, detect terrorist activities, and enhance citizen services. For example, in a smart city, vehicle movement can be tracked through sensors to determine volumes and patterns of traffic [45]. This information can then be linked with vehicle owner information to determine the relationships between age groups & their travel times and locations. This analysis can then be used for improved city planning [46].

4.20. VERSIONING: Version Control System

Data science and software development both are involve in writing codes. Data science tends to be more iterative and cyclical, where one cycle often starts with some initial understanding of the data. This cycle moves to collect, explore, clean, transform the data, and finally to build, validate, and deploy machine-learning models.

4.21. VAGUENESS: Confusion in Meaning and Tools used
The meaning of found data is often very unclear, regardless of how much data is available.

4.22. VITALITY: Stamina, Bouncing, Vigorousness, Growing
The quality or state of Big Data being vital; the principle of life; vital force; animation; as, the vitality of eggs or vegetable seeds; the vitality of an enterprise.

4.23. VIRALITY: Spreading Speed
Virility refers to the rate at which the data spreads; how often it has picked up repeatedly and used by other users.

All these above 23 Characteristics, Opportunities, Issues and Challenges explored in this research and may prove to be a milestone for the purpose of further exploring in this area. It may help to differentiate and discover more the Big Data nature. Now it is 2019, we operate in an ever more sophisticated world of smart analytics and processing. To keep up with the times, Tom Shafer has present another list [47] of Big Data's Vs. Here, this work is presenting just to introduce further V's of Big Data and in future work, author will explore more about all these 51 Vs.

4.24. VALOR: Tackling Big Problems
Valor refers to the approaches through which we must gamely tackle the big problems in Big Data.

4.25. VANE: Right direction
Vane in Data science and Big Data is point out the skills of taking right direction for right decision-making in right time.

4.26. VANILLA: Instantly value
Vanilla refers to the simplest models of Big Data created with rigor, which are able to provide value instantly.

4.27. VANTAGE: Composite systems provision
Vantage means the ability of Big Data, through which it allows us a privileged views of composite systems.

4.28. VARIFOCAL: Vast Understanding
Varifocal refers to the skills of Big Data, which allow us together to see and understand both the forest and the trees.

4.29. VARMINT: Bugs generation
Varmint refers to the speed of bugs' generation in software when Big Data gets bigger with high speed.

4.30. VARNISH: Burnish
Varnish means the approach of end-users through which they interact with our work matters, and burnish counts.

4.31. VASTNESS: Bigness acceleration
Vastness refers to the acceleration with the initiation of the internet of things; the 'bigness' of big data is also accelerating.

4.32. VATICINATION: Predictive analytics

Vaticination refers to the predictive analytics, which provides the ability to forecast. These forecasts can be more or less accurate depending on rigor and problem complexity.

4.33. VAULT: Increase in security
Data security is increasingly important, with many data science applications & software based on large & often sensitive data.

4.34. VEER: Requirements navigation
Veer enables us to have ability to navigate the customer's requirements and change directions quickly when called upon with the rise of agile data science.

4.35. VEIL: Latent variables observation
Veil is the ability in which Data science offers the capability to peer behind the curtain and observe the effects of latent variables in data.

4.36. VERDICT: Affection
As an increasing number of people are affected by models' decisions, Veracity and Validity become ever more important and getting more maturity.

4.37. VERSED: Need more knowledge
Versed refers to a new window of Big Data through which Data scientists generally need to have a little knowledge about many things i.e. programming, mathematics, statistics etc.

4.38. VET: Assumptions sophistication
Vet of Big Data refers to the approach in which Data science allows us to vet our assumptions with evidence.

4.39. VEXED: Unfolding Complicated Problems
Vexed refers to the potential of Big Data to shed light on complicated and huge problems in data science.

4.40. VIBRANT: Amativeness, Dynamism, Energetic
Vibrant express the amativeness, dynamism, energetic, lively and spirited behaviors of Big Data. These properties provide insights, ideas, and support in all aspects of our endeavors in data science.

4.41. VICTUAL: Fuels, Food
Victual refers to the fuels and food provision to data science in the form of Big Data.

4.42. VIRTUOSITY: Need more knowledge
Data scientists always need to know a little about many things regarding the data science, but common user also should grow to know a lot about this fast bouncing flood of data.

4.43. VISIBILITY: Visible for all
Big Data should be visible to all freely. This V give same meaning with voluntariness. Big Data provides visibility into complex Big Data problems in data science.

4.44. VIVIFICATION: All methods animation
Big Data in big ocean of Data science has the capabilities to animate all means of decision-making and business processes, from marketing to fraud detection.

4.45. VOGUE: Influence, fashion
Big Data has a big influence on the life of common people. However, tends in Big Data are changing continuously i.e. from Machine Learning to Artificial Intelligence and from Grid computing; now it is Cloud and Edge Computing, and from IoT is changing now to IoE.

4.46. VOICE: Speaks Loudly, Brings Knowledge
Data science provides the ability to speak loudly and present knowledge (though not all knowledge) in the form of Big Data, on a diverse and various range of topics.

4.47. VOODOOISM: Not Voodoo, Not Sorcerer, Real
As a user, we should convince potential customers of data science's value to bring results with real-world impact. Because, Big Data and

data science has no capabilities to be voodoo. Hence, Big Data was real, is real and will be real.

4.48. VOYAGE: Voyager, Keep Learning
We tackle various complicated problems that Big Data and data science provides. The quality of this try to be a Voyager always gives an environment to keep learning data science.

4.49. VULPINE: Every on Crafty
Every Big Data user by using various social network, trying to act as crafty. However, Big Data and data science should have the capabilities to handle these type of users.

4.50. VERVE: Spirit, Excitement
Verve is refers to the spirit, energy, liveness, excitement of Big Data, always anywhere for all.

4.51. VENTURESOMENESS: Venturesome
Big Data has quality or state of being Venturesome, which always brings some new adventurousness.

CONCLUSION

Big Data is a collection of Data Sets, which is growing day by day every second. Data is being produced by everyone, from everywhere, for everything by using various devices connected 24/7. In 2018, over 28 billion devices globally were connected to the internet. In 2020, more than 50 billion smart appliances will be connected worldwide and internet traffic flow will be 92 times greater than it was in 2005. This paper revolves around the Big Data definitions and its survey, technologies, characteristics, opportunities, issues and challenges in terms of V's i.e. Volume, Velocity, Variety, Veracity, Validity, Virility, Viscosity, and other 44 Vs of Big Data. Existing research reports problems and reflects that available technologies and techniques are not enough to manage and process Big Data. Hence, the present research has explored Big Data further and more Vs have been added with brief discussion on them. Big Data handing needs are growing into all sectors such as security, healthcare, education, manufacturing and government services. The discipline of Big Data is evolving around Vs. In our future work, we will try to discover more Vs and there is a chance to reach this list up to 100 Vs. This list of Vs identified will offer simple and effective practice and research environment for the Big Data professionals and scientists.

REFERENCES

[1] X. Wu, X. Zhu, and G.Wu, "Data mining with big data," IEEE transactions on knowledge and data engineering, 2014, V. 26, pp. 97-107.

[2] Y. Liu, J. Yang, Y. Huang, L. Xu, S. Li, and M. Qi, "MapReduce based parallel neural networks in enabling large scale machine learning," Computational intelligence and neuroscience, vol. 2015, p. 1, 2015.

[3] J. Lin, W. Yu, N. Zhang, X. Yang, and W. Zhao, ''A survey on Internet of things: Architecture, enabling technologies, security and privacy, and applications,'' IEEE Internet Things, vol. 4, no. 5, pp. 1125–1142, 2017.

[4] S. Mallapuram, N. Ngwum, and W. Yu, ''Smart city: The state of the art, datasets, and evaluation platforms,'' IEEE/ACIS 2017, pp. 447–452.

[5] N. Khan, Ibrar Yaqoob, Zakira Inayat, "Big Data: Survey, Technologies, Opportunities, and Challenges". The Scientific World Journal.pp. 1-18.

[6]. N. Khan, Mohammed Alsaqer, Soulmaz Salehian. "The 10 Vs, Issues and Challenges of Big Data". ICBDE 2018, USA. Pages 52-56.

[7] The McKinsey Global Institute, 2012.

[8] Big Data in Healthcare - Hype and Hope, Bonnie Feldman, Ellen M. Martin, 2012.

[9] Kirk Borne. https://mapr.com/blog/big-data-everything-quantified-and-tracked-what-means-you/. Accessed on November 21, 2018.

[10] Peter Geczy. Big data characteristics. The Macrotheme Review, 3(6):94-104, 2014.

[11] CL Philip Chen. Data-intensive applications, challenges, techniques and technologies: A survey on big data. Information Sciences, 275:314–347, 2014.

[12] Soulmaz Salehian and Yonghong Yan. Comparison of spark resource managers and distributed file systems. In IEEE Conference (BDCloud), pages 567–572.

[13] Peter Geczy. Big data characteristics. The Macrotheme Review, 3(6):94-104, 2014.

[14] Min Chen, Shiwen Mao, and Yunhao Liu. Big data: A survey. Mobile Networks and Applications, 19(2):171–209, 2014.

[15] Stephen Kaisler. Big data: Issues and challenges moving forward. 2013 IEEE. 46th Hawaii International Conference (HICSS)on, pages 995–1004.

[16] Ibrar Yaqoob, Ibrahim Abaker Targio Hashem. Big data: From beginning to future. International Journal of Information Management, 36(6):1231–1247, 2016.

[17] CL Philip Chen. Data-intensive applications, challenges, techniques and technologies: A survey on big data. Information Sciences, 275:314–347, 2014.

[18] S Mills, S Lucas. Demystifying big data: a practical guide to transforming the business of government. TechAmerica Foundation, Washington, 2012.

[19] Basel Kayyali, David Knott, and Steve Van Kuiken. The big-data revolution in us health care: Accelerating value and innovation. Mc Kinsey, 2(8):1–13, 2013.

[20] Steve LaValle, Eric Lesser. Big data, analytics and the path from insights to value. MIT sloan management review, 52(2):21, 2011.

[21] J. S. Ward and A. Barker. (2013). ''Undefined by data: A survey of big data definitions.'' [Online]. Available: https://arxiv.org/abs/1309.5821.

[22] S. F. Wamba. ''How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study,'' Int. J. Prod. Econ., V.165, 234–246, 2015.

[23] T. H. Davenport, ''Competing on analytics'' Harvard Bus. vol. 84, p. 98, 2006.

[24] A. McAfee et al., ''Big data: The management revolution,'' Harvard Bus. Rev., vol. 90, no. 10, pp. 60–68, 2012.

[25] M. Schroeck, R. Shockley, J. Smart, D. Romero-Morales, and P. Tufano, ''Analytics: The real-world use of big data,'' IBM Global Bus. vol. 12, pp. 1–20.

[26] P. M. Hartmann, M. Zaki, N. Feldmann, and A. Neely, ''Big data for big business? A taxonomy of data-driven business models used by start-up firms,'' in A Taxon.

[27] P. Tallon, Lyala Universtiy Maryland. Corporate Governance of Big Data: Perspectives on Value, Risk, and Cost, IEEE Computer Society 2013.

[28] P. Weill, S.L. Woerner, and H.A. Rubin, "Managing the IT Portfolio (Update Circa 2008): It's All about What's New," CISR, vol. 8, no. 2B, 2008, pp. 1-4.

[29] P.P. Tallon, "The Information Artifact in IT Governance: Towards a Theory of Information Governance," to appear in J. MIS, Loyola Univ. Maryland, 2013.

[30] Nawsher Khan, Ibrar Yaqoob, et al., "Big Data: Survey, Technologies, Opportunities, and Challenges", The Scientific World Journal, vol. 2014, 18 pages.

[31] Judith Hurwitz, Alan Nugent, Fern Halper, Marcia Kaufman. Big Data for Dummies. ISBN: 978-1-118-50422-2

[32] Oguntimilehin A., Ademola E.O. ''A Review of Big Data Management, Benefits and Challenges,'' Journal of Emerging Trends in Computing and Information Sciences, vol-5, pp-433-437, June 2014.

[33] http://www.ibmbigdatahub.com/presentation/current challenges – and – opportunities - bigdata- and analytics emergency management.

[34] http://www.unglobalpulse.org /sites/default/files /Big Data for Development–UNGlobalPulse June2012.pdf.

[35] Margaret Rouse. TechTarget.com. "What is Data Virtualization?", 23 Nov, 2018.

[36] Streamlining Customer Data. http://www.ardentisys.com/stories/streamlining-customer-data. Retrieved 23 November 2018.

[37] Gareth Morgan, Computer Weekly. Retrieved 23 November 2018. https://www.computerweekly.com/feature/Data-virtualisation-on-rise-as-ETL-alternative-for-data-integration.

[38] Avita Katal, Mohammad Wazid. Big data: issues, challenges, tools and good practices. IEEE Contemporary Computing (IC3), 2013. Pages: 404–409.

[39] Abdullah Gani. A survey on indexing techniques for Big Data: Taxonomy and performance evaluation. Knowledge & IS. 46(2):241-284, 2016.

[40] Stephen Kaisler. Big data: Issues and challenges moving forward. 2013. 46th Hawaii International Conference HICSS, pages 995–1004. IEEE.

[41] Big Data: The Key Vocabulary. Everyone Should Understand https://www.linkedin.com/pulse/20141203075716-64875646-big-data-the-key-vocabulary-everyone-should-understand. Retrieved 11/ 2018.

[42] R. Kitchin, "The real-time city? Big data and smart urbanism". Geo J., vol. 79, pp. 1–14, 2014.

[43] S. H. Thiago, "Large-scale study of city dynamics and urban social behavior using participatory sensing" IEEE Wireless Communication, vol. 21, no. 1, 42–51, 2014.

[44] H. Guo, X. Li, W. Wang, and W. Xu, "An event-driven dynamic updating method for 3D geo-databases," Geo-Spatial Inf. Sci., vol. 19, pp. 140–147, 2016.

[45] P.A. Laplante, "Who's Afraid of Big Data?" IT Professional, vol. 15, 2013, 6–7.

[46] Jawwad A. Shamsi. Understanding Privacy Violations in Big Data Systems. IEEE Computer Society. IT Professional. 2018 1520-9202/18.

[47] Tom Shafer. Elder Research. https://www.kdnuggets.com/2017/04/42-vs-big-data-data-science.html. Retrieved on November 26, 2018.