International Journal of Application or Innovation in Engineering & Management (IJAIEM)
Web Site: www.ijaiem.org Email: editor@ijaiem.org
Volume 9, Issue 9, September 2020 **ISSN 2319 - 4847**

How Many Old and New Big Data V's Characteristics, Processing Technology, And Applications (BD1)

Abou_el_ela Abdou Hussein

Computer Science Department, Modern Academy-Maddi, ARE,

ABSTRACT

With enormous growth of data volume that is transported across the internet links today, there has been a huge amount of complex data generated. Very large sets of data including universities companies system, Healthcare, institution gas & petroleum sector, that have so big and so complex data with more varied structure. The big challenge is how to operate this notable volume of data, which referred to as Big Data. Although big data has to be securely flying and delivered through the internet, it cannot be manipulated with conventional methods that fail to handle it, so there is a need for newer developed tools to handle it. The big data have frequently broken down into V's characteristics starting from three V's: volume, velocity and variety. The starting three V's have been extended during time through researches to reach fifty six V's till now Among them are two newly discovered by the author that means it multiplied near twenty times. We had to dive to search for all of these characteristics in many researches to detect and build comparisons to answer the old, modern and renewed vital question, "how many V's dimensions (characteristics) in big data". This paper presents the precise clarification of all big data V's (characteristics) and what are its strengths or weaknesses and why. Also we conducted several in-depth studies, some comparisons and measurements to cover the widest range areas of big data application activities. We also don't miss through this paper institutes absolute description of big data management cycle (BDMC) and Big Data processing Technology (BDPT) through four order steps that introduce conceptual View of Big Data.

KEYWORDS: BIG DATA, HEALTHCARE, VARIED STRUCTURE, TRADITIONAL METHODS, BIG DATA PROCESSING TECHNOLOGY.

1. INTRODUCTION

The Big data term is often used correspondent with other relevant terms, as Business Intelligence (BI) and Data mining (DM).All these terms is about analyzing data, but big data concept is different when data volumes, origins, and transactions are so big in size, longitudinal, heterogeneous noisy and more complex to be managed, consequently it needs special techniques to deal with this complex data. Also, there are big differences between Big Data (BD) and Business Intelligence (BI) as explained in table 1,[1]. "Big Data" was introduced by "Roguer Magoulas" in order to define collection of enormous and complex data, which ordinary and traditional management techniques fail to handle and also couldn't manipulate by existing tools due to their differences. That failure comes from more complex points of cost, long query times, reliability, and their disability to handle new sources of big data with different types. Big data encompasses everything from money transaction to tweets to images to audio. These differences can be totally described using four properties as data complexity, variety, too fast processing and too big size [2]. **First** the expression "**complexity**" is the size and intricacy of data, which comes from different types of big data as explained in figure 1.

From figure 1 big data types could be clarified as structured, semi-structured, Multi-structured, and unstructured data which greatly increases the complexity of both storing and analyzing it. **Structured data** is highly-organized and formatted in a way so it's easily searchable as in relational databases, also structured data could be seen as information that is organized or easily interpreted by traditional data models or databases. **Semi-structured** data is a form of structured data that does not obey the formal structure of data models associated with relational databases or other forms of data tables, but even so it contains tags or other signs to separate semantic features and apply hierarchies of records and fields within the data. **Multi-structured** data have different types of data which produced by different sources and applications, assimilating information from this various collection presents huge challenges and could also represented by a variety of data types and formats that came from interaction between peoples and machines, such as social network services and web applications including text and multimedia configuration like photos and videos [4].**Unstructured data** has no pre-defined format or organization, making it much more difficult to collect, process, and analyze, also unstructured data can be defined as information that is not organized or easily interpreted by traditional data models or databases, and usually is text-heavy, such as posts from twitter, LinkedIn and other social media services. Big Data has the possible to generate more income, while decreasing risk and forecast future outcomes

with more confidence at little cost [2]. Figure 2 illustrates the Big Data Management Cycle. Big Data Management is arranged around finding and organizing possible data. **Second** the expression "**variety**" emphasis that data varying very much in their shape , type and characters so data may not be allow to store, manage and analyze taking in account many dimensions as time, speed, and size. **Third** the expression "**Too fast**" implies operate data as quickly as possible to be benefit on time. **Fourth** the expression "**Too big**" tightening that organization needs to deal with large amount of terabytes, petabytes or zeta bytes.

Table 1.Business Intelligence VS Big Data.

Subject	Business Intelligence (BI)	Big Data (BD)
Type	Documents	Photos
Contents	Finances	Audio and Video
Dimensions	Stock Records(one dimension)	3D models
Relation	Personal file	Simulation Location data
Tools	Reporting tool like Cognos, SAS,SSIS, SSAS	Visualization tool like QlikView or Tableau
Data Shape	Sample data or specific historical data	Huge volume of data
Data Source	Data from the enterprise	Data from external sources like social media apart from enterprise data
Based on	Based on statistics	Based on statistics and social sentiment analysis or other data sources
Data Shape	Data warehouse and data mart	OLTP, real-time as well as offline data
Processing Method	Sequential Computing	Parallel Computing using multiple machines
Language Used	Query languages SQL, TSQL	Scripting Languages Java script, Python, Ruby and SQL
Source Data Type	Specific type of data : txt, xml, xls, etc	Multiple kinds of data : pictures, sounds, text, map coordinates

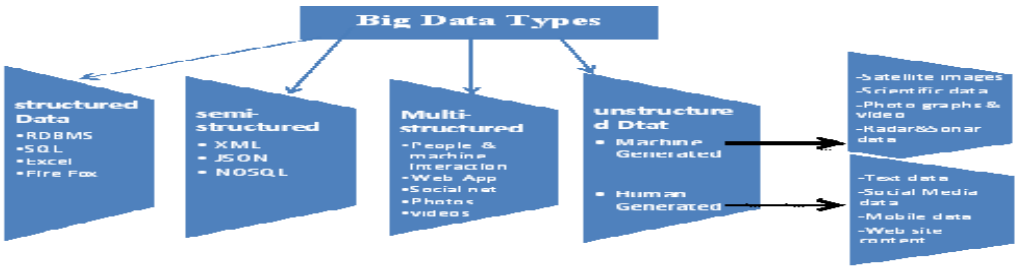


Figure 1 Types of Big Data.

In this paper we present Big Data processing Technology in section 2, Big Data fifty six V's Dimensions (characteristics) including old and new V's by author in section 3, Big Data Applications in section 4and finally conclusion are provided in section 5.

2. Big Data Processing Technology:

A practical sight of Big Data can be illustrated through next four orders as in figure3. Figure 3 introduces conceptual View of Big Data in four parts as, data collection from different sources, data storing, data analysis, and data presentation [5].

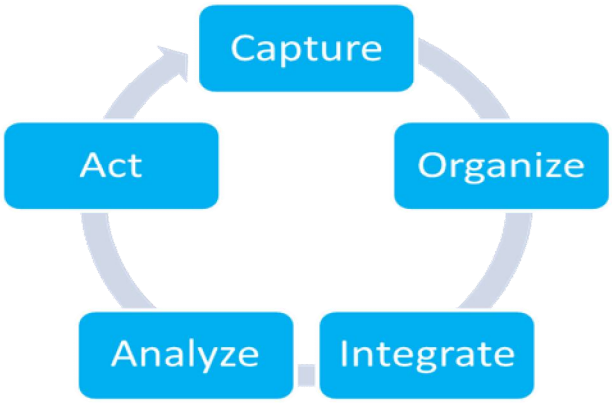


Figure 2 The Big Data Management Cycle.

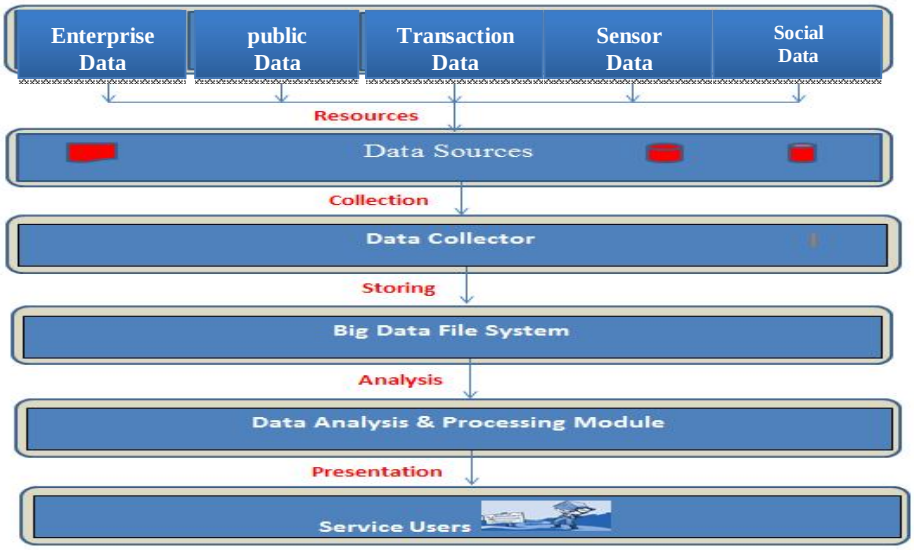


Figure 3 conceptual View of Big Data.

2.1 Big Data Collection from data sources:

This data technique includes all data gathered from internal and/or external origin in any shapes and structures. For example, origins of data could be Social Media Chats and Message, huge data of log data, and Internet Web Robots/Crawlers, Discussion Forums, Remote Sensing Networks, Transactions, etc.

2.2 Big Data Storing using distributed computing domain:

In the big data structure, the amount of extensive heave data that collected from different origins is too great to be processed on a unique computing device, which requires dealing with distributed computing devices. Consequently, all kinds of spread computing techniques need to be developed for the structure of big data. Hadoop Distributed File System (HDFS) could be seen as official file system that can memorizes these extensive heave data into a distributed storage system. HDFS can control all distributed storage data based on HTTP protocol and has the power to hold all kinds of data – JSON, XML, video, images, and plain text over the same file system. HTTP is used for moving, replications, and rebalancing of storage data.

2.3 Big Data Analysis using (SNS) analysis:

All recent developed data analysis technologies have the ability function to analyze unstructured and informal text mining that leading to use easily the Social Networking Sites (SNS) analyses. Such analysis technologies depend on Natural Language Processing (NLP), in order to find meaningful keywords, relations, or patterns from both the flat and complicated texts. There are a lot of tools for data analysis as (Apache Hadoop, CDH, Cassandra, Knime, Datawrapper, MongoDB, Lumify, HPCC, Storm, Apache SAMOA, Talend, Rapidminer, Qubole, Tableau, and R), with their figures in table 2.

Table 2: Top 15 Big Data Analytics Tools.

APACHE hadoop	cloudera Ask Bigger Questions	Apache CASSANDRA
Open for Innovation KNIME	Datawrapper	mongoDB.
lumify	HPCC SYSTEMS	APACHE STORM
Apache SAMOA	talend	rapidminer
Qubole ELEMENTAL TO BIG DATA	tableau	R

2.4 Big Data Presentation:

Big data presentation technologies could be represented by both data manipulating tools and data visualization tools and techniques, as mentioned in figure 4[6], which give an easy way to traverse data from different formats ranging from graphs or pie charts. Data visualization technique also could combine multiple graphs for easy and insightful analysis of data.

Figure 4 Top 10 Data Visualization Techniques.

3. Big Data Fifty Six V's Characteristics (Dimensions):

After the first big data 3 V's characteristics (Volume, Velocity, and Variety) began getting a lot of attention for many people to add some more V's to the characterization of big data. Other authors used the term pillars instead of big data V's characteristics [7]. Quickly the three V's with "veracity", "Value", "Variability" reached Fifty Six V's characteristics till now, two of them were added by the author as explained in figure 5. This presented both advantage in some cases and disadvantage in other cases. It was disadvantage in the feeling that many people criticize so many significant discussions were disregarded or ignored. On the other hand, the addition of more V's was advantage, in the feeling that big data's "first responders" were facing these additional new V's to the list of big data characteristics, were providing treasures lessons learning people and best exercise for them. Several V's are awarded by figure 5.Different declarations for each "V" characteristic (dimension) are presented as follows:

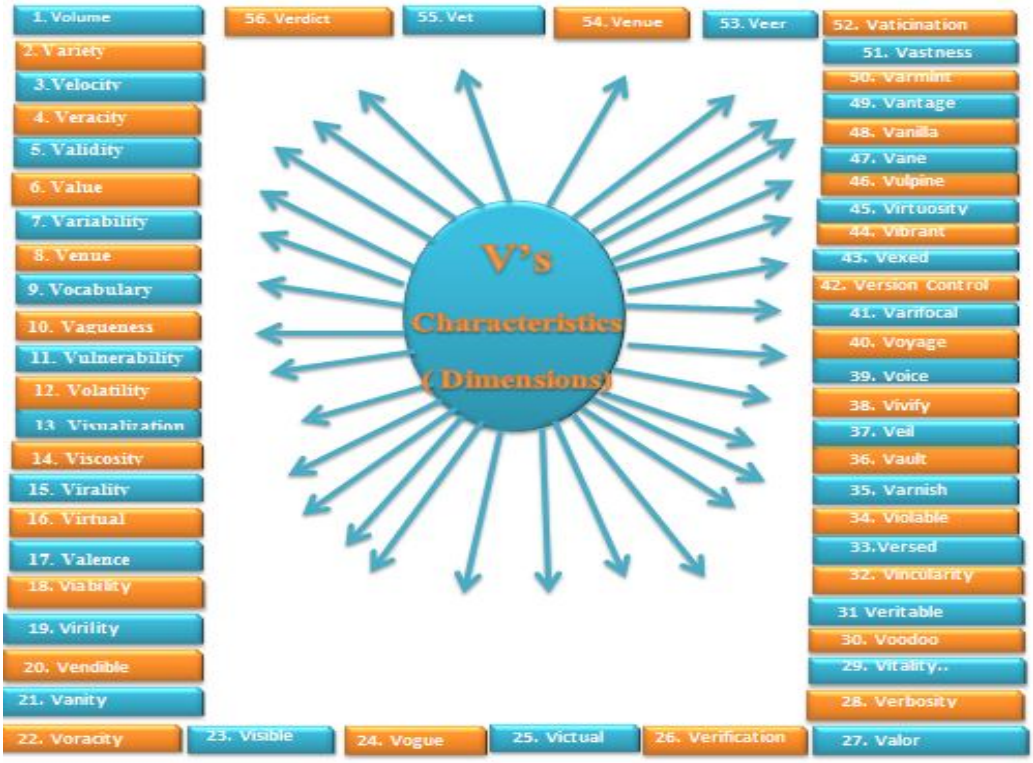


Figure 5 Big Data V's Characteristics.

3.1 Volume (too big):

The best known characteristic that presents the most immediate challenge to traditional IT structures. Many companies already amount of archived "Ocean of data" in the form data or of information that can came from every possible sensor, logs, Hundreds hours of YouTube uploaded videos, billions gigabytes from global mobile traffic, Trillions photos were taken everywhere or document sharing services through social media services. There is a big benefit gained from the facility of analyzing huge amounts of big data information which gives better results and more efficiency in various fields.

3.2 Variety (diversity):

Big Data is represented by different formats and varied types of data between structured, semi-structured, multi-structured and mostly unstructured data as well that came from many types of data resources, so it is heterogeneous in both size and type, consequently cannot be putted together into a relational database. This means that Big Data category needs to be known by the data analysts dealing with all of these varieties of data shapes because it is very complicated to analyze text, images and sounds in the same context which leads to two primary challenges here. First, saving and retrieving these data types in the presence of both advantages, quickly and cost efficiently. Second, during analysis, combining or arranging data types from different sources so that all types of data describing a single event can be extracted and analyzed together which leads to more difficulties of both saving and analyzing them to avoid problems of dark data that sits in the organization and is unused and also is not free [1]. More than 90% of data generated is data is in unstructured form [3], not only video files, social media updates, audio, image, and other text formats but also click data, log files, machine and sensor data, etc [8]. Analyzing varied types of data leads to new and valuable insights not previously available.

3.3 Velocity (faster arrival):

Could be defined as the speed of data traveling from one side to another or moves around and the speed of processing it with high rate of receiving data and information or increasing rate at which this data is produced so, the rising speed at which the data can be processed, stored and analyzed. Velocity also mentioned to the rate at which data arrived. Every daily seconds, hundreds hours of video on YouTube are propagates, over 200 million emails through Gmail's are sent [3]. A challenge raised is the characteristics of timing or abeyance the data being captured at a rate or with a lag

time that makes it useful. Also how long the data will be valuable, is it permanently valuable or dose it rapidly age and lose its meaning and importance.

3.4 Veracity (confidence or trust in the data):

We need clear and definite answer for a very important question, does data comes from a reliable source. Because it is not possible that all of the data is going to be 100% precise, accurate, complete, correct, could be redundant, and could be unreliable so there will be dirty data. There is a great mistake comes from "dirty data" because it can cause very easy to an erroneous of errors, improper results. Veracity refers more to the origin or accuracy of the data source, its context, and it's significant to analysis based on it. Knowledge of the data's veracity helps us better appreciate the risks related with analysis and decisions based on specific data set.

3.5 Validity (quality):

How quality consistence, preciseness, reasonableness and correctness the data for its intentional use. Validity in data gathering means that your detections accurately represent the incident you are declared to measure. According to Forbes, estimated data scientist's spent 60 percent of their time in cleansing their data before being able to do any analysis.

3.6 Value (merit):

One of the important and influencing features of big data is value characteristic. Unless turning the enormous amount of data in big data into value, it could be useless and unusable. To bay a lot of money to implement IT framework systems to store big data, requires a return on capital employed [3]. Data and information have no meaning if you don't acquire actual value from them. Considerable value can be found in big data, including understanding your customer's habits, preferences, and how to attract them. Also important data value could enhance machine or business execution.

3.7 Variability (quick changes & variations):

Includes dynamic, progress, time series, spatiotemporal data, periodic & seasonal, interval of reporting, and any other type of dynamic (non-static) behavior in your data sources, customers, etc. Variability in big data's circumstances means variability in the data, which required to be found by deviation and aberration detection methods leading for any relevant analytics to occur. Big data is also variable because of the multiple data dimensions resulting from multiple different data types and sources. Variability can also refer to the incompatible speed at which big data is packed into your database [8].

3.8 Venue (place):

Big data is distinguished by its distributed heterogeneous data from various platforms, from numerous owners' systems, with different formatting and access needs, private or popular[9].

3.9 Vocabulary (lexicon):

All metadata shapes like data models, schema, semantics, ontologies, taxonomies, and other contents that describe the data's structure, syntax, content, and origin[9].

3.10 Vagueness (obscurity or, blurriness):

we've always had?, what are the techniques we could use?,etc [9].

3.11 Vulnerability (unguarded):

This means that no The meaning of found data is often very unclear, not only has how much data been available but also how much it is not obscure. Misunderstand over the meaning of big data, what's new about it? Is it Hadoop?, is it something that system is perfect, which means it's probable there is a way for its hardware or software to be agreement, successively meaning that any associated data can be tacked or manipulated,[8]. Big data systems are more unguarded than anyone might think. Big data running new security issues.

3.12 Volatility (changeable):

What time does remain data valid and should be stored. How old dose data need to be before it is considered irrelevant (neither here nor there), historic, or not useful any longer? How long dose data require being remains? Prior to big data, organizations trend to store data forever – little terabytes of data might not create high storage expenses. In

traditional data setting, there not might even be data record strategies in place. There is a need now to display rules for data currency and availability as well as ensure rapid retrieval of information when required, [8].

3.13 Visualization (envision):

Refers to the application of more recent visualization techniques to explain the relationships between data. Visualization strategies include applications that can display real-time changes and more illustrative graphics, thus going beyond pie, bar and other charts. Also could indicate how thrilling it is to visualize. Current big data visualization tools face technical challenges according to constraints of in-memory technology and poor scalability, functionality, and response time. We can't rely on normal traditional graphs when trying to plot a billion data coordinates, so you need different techniques of constituting data such as data clustering or utilizing tree maps, parallel coordinates, sunbursts, circular network diagrams, or cone trees, [8].

3.14 Viscosity (thickness):

Refers to the inaction when traverse through a data collection. For example according to the differences of data origins, the velocity of data glides and the complexity of the needed processing. Inertia measures the speed at which data can spread through a network. This expression is occasionally used to express the delay, latency or lost time in the data relative to the phenomenon being described, [10].

3.15 Virality:

Measures the rate at which data can propagates through a network. Could be defined also as the speed at which the data propagates, [10]. Could be defined also as the rate at which the data spreads; how often it is picked up and repeated by other users or events [30].

3.16 Virtual (essential):

Based on building imaginary structures for big data organizations. Enterprises and other groups can benefit from big data virtualization because it authorize them to use all the data assets they gather to accomplish various goals and objectives.

3.17 Valences (equivalent):

While big data could be massive, if the connections between the data items are not established, we would have pockets or islands of disparate data whose interrelationships may not be fully understood nor utilized. Any direct connections could be established when data gets collected as they get streamed. However, discovering indirect connections between data items is more difficult and they add value to the organization. These interconnections, similar to the bonding between atoms in a molecule, result in the tenth dimension, namely Valence of big data [18]. It is a measure indicating how dense the data is, and a measure of Valence is determined as the ratio between the actually connected data items and the number of connections that could possibly be established within the data collection.

3.18 Viability:

Viability could be seen as carefully choosing those attributes in the data that are most likely to forecast outcomes that matter most to organizations. Carefully select the attributes and elements that are most likely to forecast outcomes what matters most to companies; the secret is uncovering the latent, hidden relationships among these variables" [22,23].

3.19 Virility:

In its basic meaning the property of having the characteristics of an adult male like Strong, Strength, agility, and a healthy physique. But with big Data it means that it creates itself. The more Big Data you have, the more Big Data gets strength and forceful [23, 24, 25, 26].

3.20 Vendible:

The sale ability of Big Data proves its existence. The very existence of client's for Big Data shows crucially that it is appreciable – this is evident from the communication of some known means of trading with subscribers data [23, 27,28].

3.21 Vanity:

Vain of data means that it is glad with the effect it produces on other individuals. Conceited data is satisfied with the effect it produces on itself [23, 27, 28].

3.22 Voracity (Strength):

Big Data is huge and it has a pervasive and insistent self-achievement desire. Big Data is potentially so insatiable that it may achieve the influence, manage and the possibility to consume itself [23, 27, 28].

3.23 Visible:

The data state that implies data could be seen. Information giant warehouses have always existed within companies and have been one of the major barriers in the effort to extract utility from data. Not only pertinent information should exist, but also should be evident to the intended person at the proper time [29].

3.24 Vitality (Verve):

Vitality of the data is an important perception that is vital and is included in the concept of Value. We need to put priorities for information that has more meaningful or critical to the underlying business objective [23, 29].

3.25 Vincularity:

It implies in its exact meaning connectivity or linkage. This idea is very pertinent in today's interconnected world through the internet. There is important value budget possible by connecting various information groups [23, 32].

3.26 Verification:

The process of initiate the fact, precision, or validity of data. The verification of official documents so that the data follows the standardization in many aspects to show its reliability [23, 29].

3.27 Valor:

The specific data that has the possibility to produce value and guiding how this can be accomplished. In the face of big data, we must valiantly treat the big problems [23, 31].

3.28 Verbosity:

Based on a lot of redundancy that appears in all data shapes whether it is unstructured, semi-structured, multi-structured, and structured data, there is, often the majority of raw volume. Understanding how to quickly separate the meaning you keep about from its repetition is important for efficiency of processing [23, 29].

3.29 Versality:

Versatility of data shows to what extent the data is useful, in different scenarios. The key element is understands its quality, origin, meaning, and context [23, 29].

3.30 Varnish: Interaction of end-users with our work matters, and polish counts.
3.31 Vogue: Artificial intelligence are become? [33].
3.32 Vault: Importance of data security [33].
3.33 Voodoo: Deliver results with real-world impact [33].
3.34 by author, Violable data capable of being or likely to be violated. Also capable of being transgressed or dishonored. Store data in an unstructured way but then it is still just raw data and not necessarily protected. **Veil:** Examine latent variables from behind the curtain [33].
3.35 Vulpine: Data leads to a new technology [33].
3.36 Verdict: People affected by model's decision [33].
3.37 Vet: Vetting the assumptions with evidence [33].
3.38 Vane: Unclear direction of decision-making [33].
3.39 Vanilla: Simple methods if tackled with care, can provide value [33].
3.40 Victual: Big Data fuel of data science [33].
3.41 Vantage: Privileged view of complex systems [33].
3.42 Varmint: As data gets bigger, so do software bugs [33].
3.43 Vivify: Ability of data science to cope with every real-life aspect [33].
3.44 Vastness: Bigness of Big Data [33].
3.45 Voice: Ability to speak with knowledge [33].
3.46 Vaticination: Ability to forecast [33].
3.47 Veer: Change direction according to customer need [33].
3.48 Voyage: Increasing knowledge [33].
3.49 Varifocal: It is about trees and forest [33].
3.50 Version control: You are using it right? [33].
3.51 Vexed: Potential of data science to handle complicated problems [33].
3.52 Vibrant: Provision of insight by data science [33].
3.53 Vogue: Artificial intelligence will become? [33].
3.54 Virtuosity: Craze to get more knowledge about Big Data [33].

3.55 Veritable (real):

This is a new characteristic by author, Data being in fact the thing named and not false, unreal, or imaginary. The problem here lies in making sure that the data are real and represent the reality of things and not misleading data and are not made by reality in any way. This kind of data spreads today to all paths of the Internet whether through individuals or fake companies wanting to erect on individuals and entities.

3.56 Violable:

This is also a new characteristic

4. Big data Applications:

There are several big data application activities. Let's introduce some application activates of big data as explained in figure 6:

Figure 6 Big Data Applications:

The applications of big data could be introduced as follows:

4.1 Healthcare Applications:

Significant progress in the healthcare section has already occurs using Big Data. The health care field generates enormous data every day, so there is a need, and probability, to mine this data and send it to the researchers and professionals of medical sector who can put it to work in real life, to help real patient people. With the aid of medical predictive analytics, professionals, and Health Care Providers (Professionals), HCPs are now able to provide individualized healthcare services to individual patients [11]. Apart from that, fitness wearable's, remote monitoring telemedicine– all powered by Artificial Intelligence (AI) and Big Data are helping change healthcare sector for the better. Many health care participants have under invested in information technology because of unknown returns. The nature of health care manufacturing itself also creates challenges: despite of there are many participants, there is no way to easily participate data among them because of privacy issues. Healthcare participants now have access to encouraging new themes of knowledge [12]. Researchers can mine the big data to see what nursing are more effective for particular conditions, finding related drug.

4.2 Academia Applications:

Big Data contributes to the continuity of education today specially on line learning which solving the problem that comes from covid-19 virus that requires distance education after closing all the Education Sector places. It motivates the interplay and connection between educated and their lecturers and/or faculty member. Education is not limited to the physical bounds of the classrooms and lectures–there is various online educational courses to learn from, which becomes clear in current covid-19 virus as mentioned before. Using Big Data technologies aid academic institutions are investing in digital courses all-round development of budding learners.

4.3 Banking Applications:

The banking sector relies on Big Data for trickery (fraud) detection. Big Data tools and techniques can clearly detect trickery and fraudulent transactions in real-time such as records of searching tracks, damaged alteration in customer stats, misuse of credit/debit cards, etc [11]. The use of customer data without exception raises privacy issues. By reveal hidden connections between apparently separate pieces of data, big data analytics could possibly disclose sensitive personal information [12]. Further, security risks are rising by outsourcing of data analysis activities or distribution of customer data across departments for wealthy understanding.

4.4 Manufacturing Applications:

Tata Consultancy Services TCS Global Trend Study has proven that the most important superiority of Big Data in manufacturing is enhancing the supply strategy and product quality and value. In the manufacturing sector, big data helps create a clear infrastructure, consequently, forecast uncertainties and inadequacy that can affect the business disadvantageous [11].

4.5 Information Technology (IT)

One of the important portion users of Big Data is IT companies around the world which using Big Data to enhance their performance, enhance employee capacity, and reduce risks in business operations. By combining Big Data technologies with machine learning (ML) and Artificial Intelligence (AI), the IT sector is frequently powering innovation to find solutions even for the most complex problems [11].

4.6 Big Data Analytics Applications (BDA):

Big Data Analytics Applications (BDA Apps) are a new kind of software applications, which analyze big data using massive parallel processing frameworks (e.g., Hadoop) [12]. Developers of such applications progress them using a small sample of data in a pseudo-cloud domain. There is a wide diversity of analytic tools that can be used to perform BDA, among others on the basis of SQL queries, text analytics, data mining, fast clustering, statistical analysis, natural language processing, data visualization and artificial intelligence (AI). These techniques and tools provide easily and rapidly exploitation of big data. Big Data Analytics Applications (BDA) are a new class of software applications that benefit wide-reaching data, which is specific too large to suited in memory or even on unique hard disk, which need large scale parallel-processing infrastructures.

4.7 Clustering Applications:

Users can directly find groups within data based on particular data dimensions using clustering (K-means algorithm) through a simple point and click dialog [12].Clustering algorithms including k-means use distance-based calculation to determine the similarity between data points. K-means clustering is one of the easiest and famous unsupervised machine learning algorithms. K-means algorithm determines "k" number of centroids, and then regards every data point to the closest cluster, while keeping the centroids as small as possible. With clustering, it is then simple to recognize and label groups by customer type, products, purchasing patterns, text documents, click path, patient records, behavior, etc.

4.8 Data Mining Applications:

Building classification models in the form of a tree-like structure, exactly like its name, is considered one significant type of data mining technique using Decision Trees. This kind of mining returns to manage class learning. While the categorical data represent gender, marital status, etc, the numerical data could represent age, temperature, etc. Both categorical data and numerical data could be represented using Decision Trees that systematically helps users realize what mixture of data attributes outcome in a desired outcome [12]. Decision trees illustrate the strengths of relationships and dependencies within data and are often used to determine what common attributes effect outcomes such as fraud risk, disease risk, online signups and purchases.

4.9 Finance Applications:

Today the bank accomplishes its own credit score analysis for existing customers using a large range of data, including many processes like savings, credit cards, checking, mortgages, and investment data. Using big data Companies claim to help institutions with at least one of the following applications [12]:

- Anticipating the achievement of portfolios & financial markets.
- Improving cyber security staff skills which enabling forgery reveal.
- Helping organizations using predicting and verification to find out their compatibility with all regulations.

4. 10 Economy Applications:

Big data allows for better prediction of economic phenomena and improves casual inference. Also big data allows for the complex modeling of relationships that predict well beyond the sample [12]. Also could applying big data approaches (using web data mining and machine learning) to the spreading information. Designed from deal intelligently with hardware, Hadoop can help organizations changing to low-cost servers.

4.11 Conservation Applications:

Only big data applications can enable users to manage social conversations. Keeping data in a merged, isolated system provides business intelligence advantages and is both financially and ecologically sound [12].

4.12 Marketing Applications:

There are several applications in marketing using big data as follows:

- Enhancing forecasting and planning activities.
- Online advertising progress.

Advertisers have begun to utilize facial recognition software to learn how well their advertising succeeds or fails at bracing interest in their products [12].

4.13 Sybase (now SAP) Applications:

With huge expansion of video images, audio, xml documents, applications should treats with quickly larger volumes of data & information that is not similar or even unstructured [12]. These unstructured data needs techniques that could extract value and benefits from these variant types and shapes data. Sybase adaptive Server enterprise has been updated to realize the benefits from these huge volumes of data. Sybase utilizes these data to achieve profitability, productivity, and customer satisfaction. Sybase (now SAP), IBM, Oracle, and HP tacked the opportunity to go into the market, either by constructing or even buying technology, to provide purpose-built analytical systems to meet desires and needs of new and existing customers.

4.14 Stock Applications:

With rapid expansion in the stock market, investors use a huge amount of stored data consisting big stock data base with complex mathematical formulas which need computer applications that used to feed these tremendous data which considered backbones to play a significant role in taking trading decisions tacking in account database analytics to establish a wide-ranging system to detect rude trading patterns to detect cheating and fraud [12].

4.15 Credit Cards Applications:

Different Credit card companies like Visa are relying on the speed and accuracy of in-database analytics to discover possible fraudulent transactions using big data and machine learning [12]. Also Big data technology can join with advanced block chain technology for securing users data.

4.16 Enterprise Applications:

Processing variant big data with traditional methods is a path of imagination. It originated with web search firms who had the problem of querying a huge distributed unstructured data [12]. For enterprises industries, in-database analysis is providing an excellent advantage.

4.17 Consumer Goods Applications:

Consumer goods industry trends are how firms are leading product sales using big data [12]. Big data analytics leads to increasing demand and purchases Basket. It develops relation between customer and goods engagement, loyalty and confidence. It helps in determining optimal advertising means by collecting consumer preference and purchasing data extracted from purchases, surveys, product reviews from online retailers, web logs, mobile conversations with customer call centers, etc. By doing this, the firms develops could understand why certain products succeed and why others Not. They can spot reasons that can help them determine the right products in the right marketing advertising media.

4.18 Hadoop Applications:

Hadoop is open origin structure used for storing and processing big data. The tremendous volume of big data is stored in cheap servers and its spread file system enables mutual processing [12]. There are data tasks as advanced analytics or data processing, with which Hadoop can assist.

4.19 Agriculture Applications:

There are a lot of big data applications in agriculture which include sensor deployment used by advanced analytics or data processing, bench marking, and introducing analytics and predictive modeling, and use better methods to optimize crop efficiency [12]. It plants test crops and performs simulations to measure how plants affected and react to various changes in condition. These simulation processes allow discovering the optimal environmental conditions for specific gene types.

4.20 Smart Phones Applications:

Phones applications now have improved the method that people all over the world use their smart phones [12]. Users of Android smart phones and I-Phone have applications at their hands that use facial recognition technology for different tasks. I-Phone users can unlock their mobiles with recognizes me, an app that uses facial recognition in instead of a password.

4.21 Telecom Applications:

Most telecom use cases categories as network service optimization, customer acquisition, retention and security [12]. Multiple telecom operators applied big data for their applications starting from internal business ending with external applications. Operators believe that big data and advanced analytics will play a very important role in helping them to achieve their business objectives.

5. Conclusion:

Big data is a collection of huge and composite data sets that comes from different types as structured, semi-structured, Multi-structured, and unstructured data. These variant in data types increases the complexity of both storing

and analyzing it. This paper introduces first, big data management cycle with five main phases, namely Capture, Organize, integrate, Analyze, and Act. Secondly, we introduce clear and definite answer for the important and puzzling question, "how many V's Characteristics of big data" for both old and new (discovered by author) characteristics. From evolution of various characteristics of big data, we could answer this important question by identifying fifty six characteristics, two from them are new characteristics defined by author. Also, we introduce Big Data processing Technology with four parts, namely Collection, Storing, Analysis, and Presentation. Finally this paper covers big data multiple applications in various fields that accompany the rise of big data. Future work, which is already in progress, is the challenges that object big data, and how to overcome these challenges in order to exceed the maximum benefit from big data technologies which will take title named (BD2) referring to complete searching in the same field of Big Data because, as it is clear from the title we called this paper (BD1).

References

[1] Mircea Raducu TRIFU, Mihaela laura IVAN "Big Data: present and future", Database Systems Journal, Vol. 1, no.1, 2014.

[2] Hiba Jasim, Ammar Hameed Shnain, Sarah Hadishaheed, Azizahbt Haji Ahmad "Big Data and Five V'S Characteristics", International Journal of Advances in Electronics and Computer Science, ISSN:2393-2835, Vol. 2, Issue-1, 2015.

[3] Ishwarappa, Anuradha J, "A Brief Introduction on 5Vs Characteristics and Hadoop Technology, Science Direct, Procedia Computer Science 48, p319-324, 2015.

[4] Ishan Verma, Lokendra Singh, "Multi-structured Data Analytics using Interactive Visualization to aid Business Decision Making", ACM COMPUTE, Bhopal, India, November 16–17,2017.

[5] Owen O'Malley, Kan Zhang, Sanjay Radia, Ram Marti, and Christopher Harrell, "Hadoop Security Design" ,Yahoo, Inc., Technical Report, October,2009.

[6] https://www.datapine.com/blog/data-visualization-techniques-concepts-and-methods/.

[7] Dr Satyam Priyadarshy, "The 7 pillars of Big Data", Petrolium Review, 14 january, 2015.

[8] George Firican, "The 10 V'S BIG DATA", Work Paper, February 8, 2017.

[9] Dr Kirk Borne, " Top 10 Big Data challenges-A serious look at 10 big Data V's", blog post , April 11, 2014.

[10] William Vorhies,View Blog "How Many V'S in Big Data", Work Paper, October 31, 2014.

[11] Abhinav Rai, "What is Big Data – Characteristics, Types, Benefits & Examples", Work paper, June 2019.

[12] Kuchipudi, Tatireddy Subba Reddy," Applications of Big data in Various Fields", Kuchipudi Sravanthi et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (5) , 2015.

[13] Sitalakshmi Venkatraman,Ramanathan Venkatraman, "Big data security challenges and strategies", AIMS Mathematics, http://www.aimspress.com/journal/Math, ,4(3): 860–879. DOI: 10.3934/math.2019.3.860, 2019.

[14] C. Liu, C. Yang, X. Zhang, et al." External integrity verification for outsourced big data in cloud and IoT: a big picture", Future Gener. Comp. Sy., 49, 58–67, 2015.

[15] M.Benjamin, S. B. Michelle and T. B. Nadya, "Eigenspace Analysis for Threat Detection in Social Networks", 14th International Conference on Information Fusion, IEEE, pp. 1–7, 2011.

[16] José María Cavanillas , Edward Curry Wolfgang Wahlster," New Horizons for a Data-Driven Economy A Roadmap for Usage and Exploitat tion of Big Data in Europe", springer open, ISBN 978-3-319-21569-3 (eBook) DOI 10.1007/978-3-319-21569-3 Library of Congress Control Number: 2015951834,2017.

[17] Rajeev Agrawal, Christopher Nyamful," Challenges of big data storage and management", Global Journal of Information Technology, http://sproc.org/ojs/index.php/gjit, Volume 06, Issue 1,p. 01-10, 2016.

[18] Policroniades, C., & Pratt, I, Alternatives for Detecting Redundancy in Storage Systems Data. Paper presented at the USENIX Annual Technical Conference, General Track, 2004.

[19] G.B. Tarekegn, Y. Y. Munaye, "Big Data: Security Issues, Challenges and Future Scope", International Journal of Computer Engineering & Technology, volume 7, issue4, July-august 2016.

[20] Z. Yan, W. Ding, V. Niemi, et al.," Two schemes of privacy-preserving trust evaluation" , Future Gener. Comput. Sy., 62, 175–189, 2016.

[21] Rebecca Webb, "12 Challenges of Data Analytics and How to Fix Them", Risk management Blog-Clearrisk, https://www.clearrisk.com/risk-management-blog/challenges-of-data-analytics, Dec, 2018.

[22] Caesar Wu, R. B. (n.d.)," Big Data Analytics = Machine Learning + Cloud Computing. BDA= ML+CC ", p.1-25,2016.

[23] s. Dhamodharavadhani, Gowri Rajasekaran, "Unlock Different V's of Big Data for Analytics", International Journal of Computer Sciences and Engineering, Open Access Research Paper , Vol.6, Special Issue-4, May 2018.

[24] Dr.Darrin, Retrieved https://educationalresearchtechniques.wordpress.com /2016/05/02/"characteristics-of-big-data"/, May 2, 2016.

[25] Gartner. Retrieved,http://www.forbes.com/sites/gartnergroup/2013/03/27/ gartners "-big-data-definition -consists-of-three-parts-not-to-be-confused- with-three-vs", /#5040d1cc3bf6, March 27, 2013.

[26] Retrieved from https://hbr.org/2012/10/making-advanced-analytics-work-for-you/ar/1, 2012.

[27] Gewirtz,D, Retrieved http://www.zdnet.com /article/volume-velocity-and-variety-understanding-the-three-vs-of-big-data/ ,April 20, 2016.

[28] GoodStratTweet. Retrieved from http://www.informationweek.com /big-data/big-data-analytics/big-data-avoid-wanna-v-confusion/d/d-id/1111077?page _number=1, Feb 22,2015.

[29] Laney, D,Retrieved from http://blogs.gartner.com /doug-laney/deja-vvvue- others, claiming-gartners-volume-velocity-variety-construct-for-big-data/, January14, 2012.

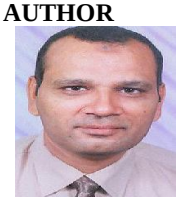
[30] Retrieved from http://www.infoivy.com /2013/11/comparison-table-of-hive-pig-shark.html, June 10, 2015.

[31] Borne, D, Retrieved from https://www.mapr.com /blog/top-10-big-data-challenges-serious-look-10-big-data-vs , April 11, 2014.

[32] Chuck Cartledge, "How Many Vs are there in Big Data?" ,working paper, 18 February, 2016.

[33] Muhammad Mashab Farooqi, Munam Ali Shah, Abdul Wahid, Adnan Akhunzada, Faheem Khan, Noor ul Amin, and Ihsan Ali, "Big Data in Healthcare: A Survey", Springer Nature Switzerland AG, Khan et al. (eds.), https://www.researchgate.net/publication/328859276, Applications of Intelligent Technologies in Healthcare, EAI/Springer Innovations in Communication and Computing, 2019.

AUTHOR



Abouelela Abdou received the Hight Diploma, M.S., and PHD degrees in Computer Science from Faculty of graduate studies for Statistical Research, Cairo University, and Teaching Computer Science subjects in modern Academy – maddi – Cairo – Egypt since 2001 Till Now. My research gate address is https://www.researchgate.net/profile/Abou_El_Ela_Hussien/achievement/5f6213423568de05d2ca4634, I have published Seven Papers before, all of them are included in my page in research gate.