THE IMMENSE POTENTIAL OF BIG DATA

Santosh Venkatraman, Tennessee State University, Nashville

ABSTRACT

The collection of data about human activities and machine operations is increasing exponentially every day. This collection of data, often referred to as Big Data, is also not necessarily like the traditional data, as it is largely unstructured, and hence cannot be managed by traditional databases and analytics platforms. NoSQL data stores such as MongoDB, along with platforms like Hadoop and Spark are far more suited for storing and analyzing Big Data.

The analysis of Big Data has immense potential for increasing revenue, profits, customer satisfaction and competitive advantages for modern organizations. The emergence of artificial intelligence is also very dependent on the availability of large volumes of clean data – so Big Data is also becoming the lifeblood of AI-powered systems. This paper describes the nature of Big Data and discusses the vast potential it offers to organizations.

INTRODUCTION

The relentless collection of large volumes of data from all kinds of sources, especially from machine sensors and websites have introduced both a high level of complexity, as well as a great opportunity for businesses. Most of the world's big organizations such as Apple, GE, Walmart, Toyota, Exxon and Samsung have global operations (factories, warehouses, transporters, and customers) and serve several customers with a wide variety of products and services. It is often hard for humans to unravel the complex problems (where and why) arising from these vast and highly sophisticated networks. The ever increasing collection of data, also known as "Big Data," will only be useful if it can be analyzed to give useful insights into business problems, and perhaps even make suggestions as to when and where future problems will occur (predictive analytics) so that the problems can be avoided or at least mitigated. Predictive analytics can also unravel positive trends and opportunities, and allow organizations to proactively allocate resources to take advantage of those future opportunities.

Entire supply chains, for example, are managed efficiently by collecting data points all along the supply chain. The data is then analyzed by analytics software to enhance the efficiency and effectiveness of supply chain management. Efficient supply chain management offers company's competitive advantages in terms of improvement in service and quality, lowering costs, and the ability to compete successfully in global marketplace.

Another example is that of the industrial giant GE, which is rapidly getting into the Industrial Internet and Internet of Things (IoT) space. On any given day, 24,000 locomotive engines are travelling about 140,000 miles, and GE estimates that if its new Big Data tools (Industrial Internet Software Suite) could even improve efficiency of its engines by 1%, that would translate into a savings of $2.8 billion annually for its customers [Gertner 2014]. GE's Trip Optimizer, for instance, is a type of cruise control that combs through piles of data and synthesizes them for the driver in a way that allows him to steer the locomotive to maintain the most efficient speed at all times, and reduce fuel burn.

Clearly the collection and analysis of Big Data can potentially be a massive advantage to many organizations. The purpose of this paper is to examine the exciting field of Big Data, and examine its role in benefitting organizations. The trend of connecting people and machines to the Internet, and then collecting data via websites and sensors is creating an unimaginably large repository of data. This Big Data can then be analyzed (often, in near-real time) for useful information. Specifically, we illustrate the many ways in which Big Data is collected and analyzed for solving business problems and its immense potential for providing competitive advantages.

The paper initially describes the nature of Big Data and details four important dimensions to describe it - volume, velocity, variety, and veracity. It then briefly discusses the ways in which Big Data is stored and analyzed. The next section describes the immense potential of Big Data to make organizations function more effectively and efficiently. Finally, we summarize and conclude the paper in the last section.

BIG DATA

Big data is different from traditional data stored in relational databases, which also can be big (in terms of storage requirements), in many significant ways. Traditional databases are collections of data that are well structured – each record has a specific number of fields, and all data records conform to that structure. Much data currently, however, is collected from websites and machine sensors on a continuous basis. Unlike traditional data stores, these often do not conform to a predefined structure and make it harder to analyze due to extra-large volumes.

IBM defines Big Data in fairly simplistic terms: managing huge amounts of data, and being able to process it quickly [Lo 2018]. The data is too big in terms of volume, moves too fast, or doesn't fit the structures of most company's database architectures [Wilder 2012]. To gain value from this Big Data, organizations need an alternative way to store and process it. Since 2012, Big Data has become a buzz word in the business world. With the advancement of hardware, networking, and software platforms. it has also become viable, as cost-effective approaches have emerged to tame the volume, velocity, variety and veracity of data.

Within this data lie valuable patterns and information, which were previously hidden because of the inability to extract insights from them. To modern, successful corporations, such as Walmart, Amazon or Google, this power has been in reach for some time, but came at a very high cost. A delay in the processing time of Big Data can have detrimental effects, such as revenue loss, customer dissatisfaction and competitive disadvantage. For instance, Google reported a 20% revenue loss with the increased time to display search results by as little as 500 milliseconds and Amazon reported a 1% sales decrease for an additional delay of as little as 100 milliseconds [Cogn1 2012]. In order to better understand the nature and complexities of Big Data, we next look at the various dimensions of Big Data

BIG DATA DIMENSIONS – THE 4 V'S

To better understand Big data, it is often described in terms of four basic dimensions, often referred to as the 4V's of Big Data: Volume, Velocity, Variety, and Veracity [IBM 2019] as shown in Figure 1. We describe the details of each of these dimensions next.

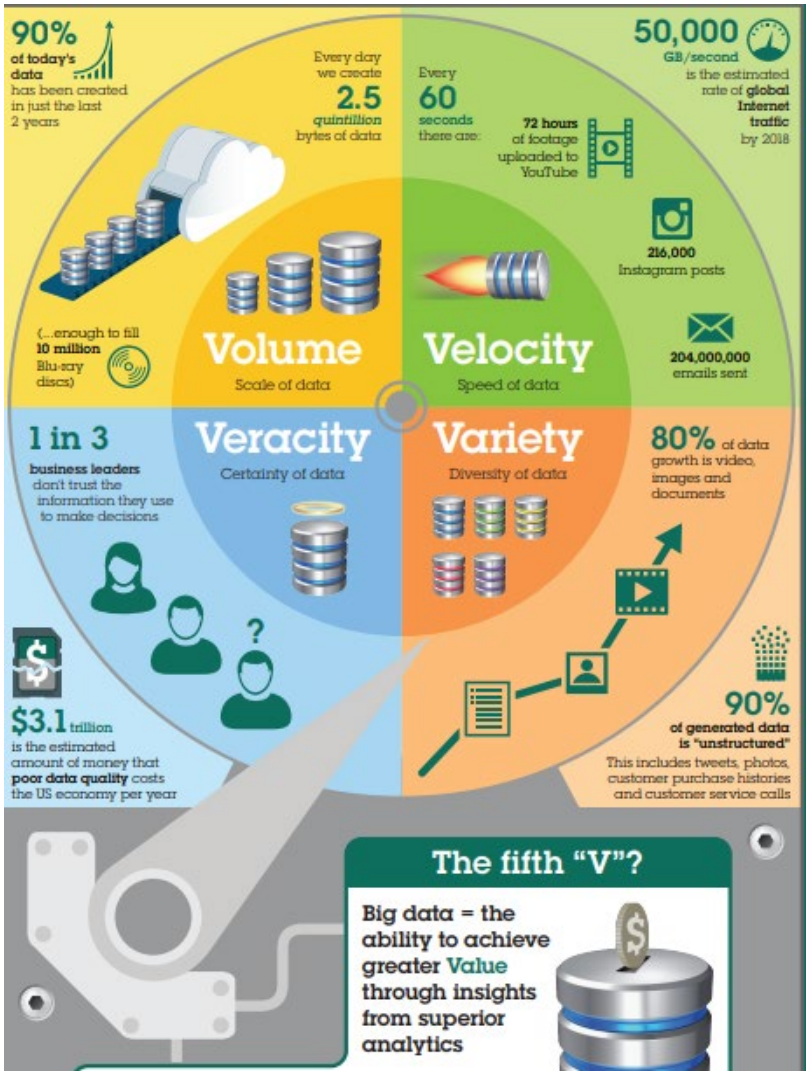

Figure 1: The Four V's of Big Data and Its Value [IBM 2019]

Volume

The sheer volume of data we create currently is perhaps unimaginable. We generated about 2.5 quintillion bytes of data just in 2018 [Marr 2018], and 90% of the data created in the world was created in the last 2 years. Data has always been big, but never nearly as massive as it is today, and never growing at this rate. With the exponential growth of IoT (Inter of Things), and high bandwidth applications such as Virtual Reality, Augmented Reality and Ultra High-Definition videos, the amount of data generated will only accelerate, and some estimates are as high as 175 Zettabytes by 2023 [Coughlin 2018]. Figure 2 shows the projected growth of Big Data in terms of its value in dollars [Columbus 2018].

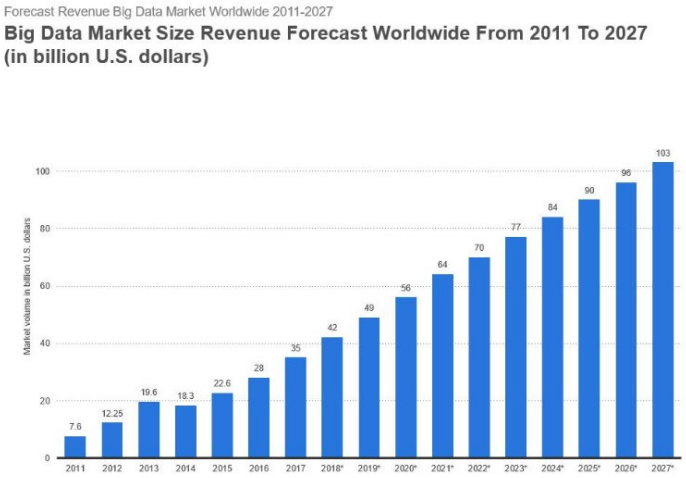

Figure 2: Big Data Growth Projection

Figure 3 shows the estimate of the data bombarding the Internet in one minute in 2017 [Domo 2017]. Additionally, there is an explosive growth in sensor based data generators in Hospital Intensive Care Units, Radio Frequency IDs tracking products and assets, GPS systems, smart meters, factory production lines, satellites and meteorology- and the list continues to grow rapidly. Just considering IoT growth, a recent Gartner report [Liton 2018] estimates that we will have more than 20 billion such sensors by 2020. These sensors are expected to generate more than 500 zettabytes of data per year just in 2019 – and continue to grow exponentially. The focus on the volume of Big Data is important, as it will determine the technologies used to store and retrieve these massive data stores effectively– and more importantly analyze them in a timely manner, to make them meaningful to the decision makers in the business.

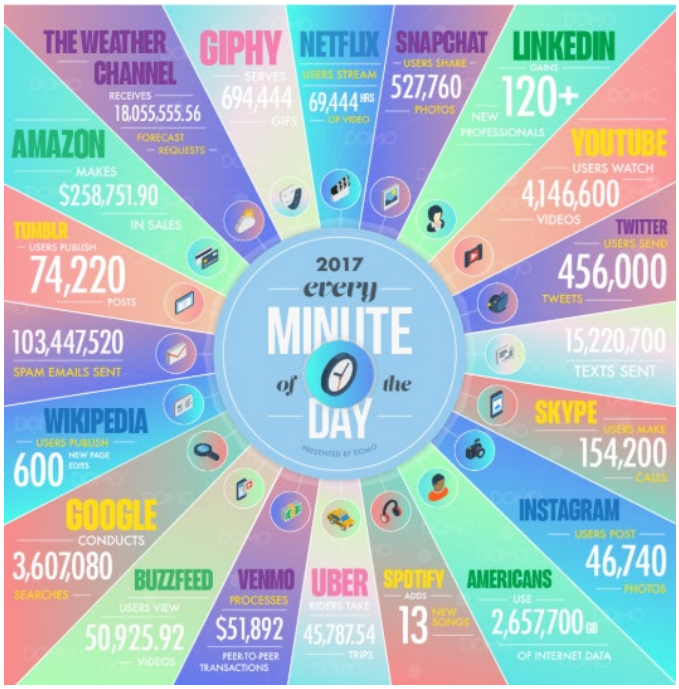

Figure 3: Data Generation Sources

Velocity

Velocity is the second dimension of Big Data. Globally, data is being generated at an ever-increasing rate. There are several aspects of the speed of data, so we have to go beyond just looking at the rate at which data is generated or received. No doubt, the "fire hose" sources like IoT and social media generate a lot of data very rapidly, however, the focus here is the frequency of data, and the degree of real-time response that is needed for obtaining true value. That velocity of data has to be processed rapidly too, if organizations want to make effective real time decisions, and then make course corrections along the way. There are many situations in which the data needs to be processed rapidly and immediately to gain value, or else the data might just lose its value and become stale or obsolete. Take for instance, the case of retailing seasonal items or perishable items. It is extremely critical to know which items are moving fast from which stores to minimize the perish rate, and, on the other hand, if the fast moving items are not restocked promptly, it would mean lost sales. So, near-real-time processing of the data can result in lower waste and losses, and simultaneously in increased sales and profits. Similarly, it would be a waste of capital and shelf space if excess products of a slow selling item is overstocked. Perishable items will have to be discarded, and unsold seasonal items must be discounted sharply to get them cleared.

Speed and agility is, hence, crucial for many organizations. Volatility (another potential V) is also a related term to velocity, as is it involves a temporal aspect. Data at high velocities can be volatile due to rapid rates of change, and the small window of time in which it could prove valuable (small lifetime). The ability to rapidly process and utilize the stream of data, to gain actionable insights for immediate execution, is indeed a much-required ability. For example, the barrage of feeds from social media sites can indicate sentiments and trends that can materialize rapidly, and dissipate equally quickly. On the other hand, trends for preferences for vehicles may be much slower to emerge, and stay around for a longer time.

Amazon takes velocity very seriously and strongly believes in high-velocity decision making [Dykes 2017]. Amazon realizes that it may have to make a sub-optimal decision on incomplete information using this approach, but is also confident that they can rapidly course correct as new data comes in later. For Amazon, making rapid decisions, with course corrections, has proved more beneficial than slow decision making.

Another example of velocity is when IoT sensors in a machine are detecting potential problems; if the rapidly collecting data is simply stored, but not analyzed rapidly, then the machine cannot be preemptively serviced to prevent breakdowns. The machine could be an aircraft engine, a locomotive engine or even an air-conditioner unit. The ability to rapidly process and act on the large volumes of data is clearly advantageous.

Agile organizations must not only collect and analyze high velocity data, but must also be prepared to act rapidly. So, the technology, processes, and the organizational culture has to all be aligned for such agility. Many executives utilize dashboards to track key performance indicators in their organizations, and then use them to make effective, real-time decisions. In order to handle high-velocity, short lifespan data we need to minimize movement and storage and increase the speed of analysis. More than ever, data must be analyzed and decisions made in real-time, which precludes storing the data in intermediate repositories because every touch point costs valuable time.

Variety

Data can be human generated or machine generated. Machine generated data, for instance, can be captured via sensors, surveillance cameras, and satellites. Humans could type data on web pages or word processors, put videos and pictures on social media, or record audio/video files and so forth. In either case, the data can be classified as structured, unstructured or semi-structured. Figure 4 [Taylor 2018] shows a good summary of the types of data and sources.

	Structured Data	Unstructured Data
Characteristics	• Pre-defined data models • Usually text only • Easy to search	• No pre-defined data model • May be text, images, sound, video or other formats • Difficult to search
Resides in	• Relational databases • Data warehouses	• Applications • NoSQL databases • Data warehouses • Data lakes
Generated by	Humans or machines	Humans or machines
Typical applications	• Airline reservation systems • Inventory control • CRM systems • ERP systems	• Word processing • Presentation software • Email clients • Tools for viewing or editing media
Examples	• Dates • Phone numbers • Social security numbers • Credit card numbers • Customer names • Addresses • Product names and numbers • Transaction information	• Text files • Reports • Email messages • Audio files • Video files • Images • Surveillance imagery

Figure 4: Sample of Data Variety and Sources

Structured data has a predictable format, and can be organized as rows and columns in tables. For example, an Employee database table record might have a EmployeeId, Name, AreaCode, Phone, GrossSalary, and CityStationed for each employee. It lends itself to relatively easy storage, analysis using traditional relational database management systems. Figure 5 shows how the structured Employee data could be represented in a table.

EmployeeID	Name	Date-Joined	AreaCode	Phone	GrossSalary	CityStationed
11	Michelle Piper	12/11/2016	505	555-1616	$ 999.00	Las Cruces
12	Mick L Mouse Jr.	12/30/2012	615	555-1313	$ 1,450.00	Nashville
13	Joe Fernandez	7/15/2011	913	555-2121	$ 1,275.00	Kansas City
14	Rhonda Lam	3/3/2012	615	555-1111	$ 3,200.00	Nashville
15	Ram Sharma	12/3/2013	501	555-1919	$ 3,100.00	Little Rock
16	Tiger Forests	1/23/2017	615	555-1717	$ 7,200.00	Nashville

Figure 5: Structured Data

Big Data, on the hand, is often associated with unstructured and semi-structured data. The data source is often diverse, such as web pages, audio files, video streams from close-captioned cameras, text messages, chats, social media platforms or even data feeds from device sensors on machines. Unstructured data does not fit neatly into rows and columns like structured data, so it cannot be stored effectively in relational databases, and cannot be meaningfully analyzed using database languages such as SQL. There are new technologies that are more suitable for storing and analyzing Big Data. Nonrelational databases like NoSQL (Not Only SQL) databases are less constrained than relational databases, and more suited to Big Data. MongoDB, Couchbase, Google's BigTable, and Amazon's DynamoDB are some examples of NoSQL databases.

Some data can also be semi-structured, and hence contains internal tags and separators identifying some distinct data elements and hierarchies - but not as clearly defined as structured data. Examples of semi-structured data include XML documents, and Open JSON (Javascript Object Notation). Email also is a common example of semi-structured data as its native metadata enables classification and keyword searching. Many of the NoSQL databases also are useful for storing semi-structured data, as unlike relational databases, the schema and the data are not separated. MongoDB, for instance, can store semi-structured documents in native JSON format. Similarly, MarkLogic is especially suited to store and take full advantage of XML documents. So

Unstructured data makes up about 80% to 90% of enterprise data, and growing at a rate of about 60% annually. It is, therefore, critical to have appropriate infrastructure to efficiently store and analyze these data streams for maximal return on investment. In comparison, structured data makes up about 10 to 15% of enterprise data, while semi-structured takes about 5% to 10% [Taylor 2018].

Veracity

The veracity aspect of Big Data deals with the conformance of data with truth and accuracy, and is perhaps the hardest to achieve. Veracity determines the level of trust in the data. Due to the velocity of the variety of large volumes of data (the other 3 V's), maintaining and verifying veracity is indeed a great challenge for Big Data. Many things can cause us to question the veracity of data, such as inconsistencies, model approximations, ambiguities, deception, fraud, duplication, spam and latency [Emani 2015]. The real purpose of Big Data, after all, is to

use it for making meaningful and effective decisions, therefore bad quality, and irrelevant data will often lead to undesirable decisions.

Data accuracy depends on many factors such as the data collection methods, the quality of the data sources, and the very methods used in processing the data. Factors such as data-bias, variability, inconsistencies and duplication can also significantly affect the quality of the data. Fortunately, data need not be perfectly accurate all the time, for all applications - so there maybe tradeoffs, when dealing with Big Data. If the data will be used for exploratory or experimental purposes, there may be some tolerance for inaccuracy in the sample (especially if it can be obtained quickly at a low cost).

In our current era of rising artificial intelligence (AI), the veracity of data is rather crucial. AI systems are often trained by Big Data, and the fair use of AI systems that affect us profoundly, depends heavily on the veracity of the training data. The use of "biased" AI systems is a topic of great interest currently, because the implications for organizations and society are profound. For example, if an AI system for detecting cancer is trained by data from just China, then it may prove very reliable in predicting cancer in people of Chinese ethnicity, but may lead to misleading results when diagnosing cancer in people of Caucasian descent. The ineffectiveness of the system will not necessarily minimize the case for using AI in healthcare – just that the training was based on a biased sample, whose veracity is in question when applying to all humans, in general. So, veracity is one of the most important dimensions of Big Data analysis. There is need to understand the allowable level of uncertainty or lack of veracity in the data, and re-define trust in the context of the questions that organizations are attempting to answer. There is also a need to weigh the cost of that uncertainty against the value the data brings to the problem.

BIG DATA STORAGE AND ANALYSIS

The four V's of Big Data, described above, present big problems for traditional data storage and analytics platforms. Despite the many advancements in database management and executive level support, most companies are still badly behind the curve, when it comes to analyzing Big Data, and reaping all the potential benefits. Surprisingly, less than 50% of structured data is actually used in decision making, and worse still, less than 1% of the unstructured data is analyzed or used at all [Davenport 2017].

As the purpose of Big Data analytics is more for predicting trends and future behavior, it is not necessary, nor realistic, to expect 100% accuracy. That is unlike traditional data analysis, as it is essential for a value, such as bank account balance, to be 100% accurate all the time. Due to the heavy volumes of wide varieties of data, of questionable veracity, arriving at high velocities, it is not easy or essential for Big Data to be neatly structured like relational databases. As traditional Relational Database technologies and methods of loading, storing and retrieving data were not really designed to process Big Data, newer technologies such as Spark, Hadoop, MapR, Cloudera, Teradata Aster, IBM Neteeza, NoSql, NuoDb, MongoDB, CouchDB, and HBase have made it easier and more efficient to handle these large volumes of data.

We next briefly describe how Big Data is better handled by Apache Hadoop, which is an Open source, free implementation of MapReduce (originally a Google Technology, but Open now). Hadoop is a programming model for processing large data sets with a parallel, distributed algorithm on a cluster. It utilizes a scale-out architecture that makes use of

commodity servers configured as a cluster, where each server possesses inexpensive internal disk drives. The HDFS (Hadoop Distributed File System) creates multiple replicas of data blocks and distributes them on compute nodes throughout a cluster to enable reliable, extremely rapid computations. The divide-and-conquer strategy of processing data is not really new, but the combination of HDFS being an open source software (which overcomes the need for high-priced specialized storage solutions), and its ability to carry out some degree of automatic redundancy and failover make it popular for modern businesses looking for Big Data analytics solutions. Hadoop is not only a receptacle for Big Data with its distributed file system, but it is also an engine that brings incredible potential to process data, and extract meaningful information in a timely manner.

NoSQL databases are often used in Hadoop environments to store Big Data, and analyze them expeditiously. Apache Spark also is another popular Open source, distributed computing platform for real time Big Data analytics. Let us now briefly study a non-traditional, popular, open source, NoSQL data store for Big Data known as MongoDB. MongoDB provides a flexible document storage system and analysis platform. It stores documents in a JSON-like format, so that the fields in each document can vary (unstructured data) and the data structure itself can be changed over time. It is a distributed database at its core, so it is designed for horizontal scaling, high availability and easy geographic distribution. MongoDB has a document model that allows software applications to easily use the stored documents (data). It is a powerful and useful platform for Big Data due to its ability to index, do real-time data aggregation and write ad-hoc queries – and it also provides end-to-end data security. Massive users like Amazon, Cisco, Comcast, eBay, eHarmony and Splunk are therefore using MongoDB, and adding to the credibility of this new technology. Table 1 [MongDB 2016] shows various ways in which NoSQL data stores like MongoDB powers Big Data applications, along with Hadoop and Spark.

Table 1: MongoDB Powered Big Data Apps using Hadoop/Spark

	MongoDB	Hadoop or Spark
eBay	User data and metadata management for product catalog	User analysis for personalized search & recommendations
China Eastern Airlines	Data supporting flight search application	Calculate fares based on permutations of rules stored in MongoDB
Orbitz	Management of hotel data and pricing	Hotel segmentation to support building search facets
Pearson	Student identity and access control, content management of course materials	Student analytics to create adaptive learning programs
Foursquare	User data, check-ins, reviews, venue content management	User analysis, segmentation and personalization
Tier 1 Investment Bank	Tick data, quant analysis, distribution of reference data	Risk modeling, security and fraud detection
Industrial Machinery Manufacturer	Storage and real-time analytics of sensor data collected from connected vehicles	Preventive maintenance programs for fleet optimization. Monitoring of manufactured components in the field
SFR	Customer service applications accessed via online portals and call centers	Analysis of customer usage, devices & pricing to optimize plans

BIG DATA POTENTIAL

The collection of Big Data, globalization and online real time business transactions has introduced a new level of complexity to the business world – but it also has opened up vast opportunities for new markets and efficiencies. Now that we have discussed the nature of Big Data, and briefly studied the way in which it is stored and analyzed, we next discuss the major benefits of implementing Big Data.

Many current businesses have global operations that serve several geographically dispersed customers with a wide variety of products and services, and use global networks of suppliers, and also utilize service providers such as Cloud vendors to do so. The complexity of such networks is hard to unravel, and makes it difficult to find where and why problems and opportunities occur. Fortunately, there is also a rapid increase in the volume of data available at various touch points, and smart organizations analyze them, and act swiftly using the insights gained. For example, an average Fortune 1000 company could just increase data usability by just 10% and gain about $2 billion a year [Crossover 2018].

In general, organizations strive to use Big Data to achieve advantages such as cost reductions; increased revenue and profits; enhanced customer satisfaction; higher employee productivity; increasing agility; more targeted marketing; risk/fraud mitigation and to effectively enter new markets with less uncertainty – essentially to gain competitive advantages, in a fast moving, hyper efficient business environment. Unlike most of the prior technologies, Big Data uses massive amounts of relevant data to make better and deeper levels of analysis to give actionable insights. Big Data analytics is almost an essential activity for modern enterprises, as it

offers several tangible advantages. About 97% of the executives in the Big Data Executive's Survey, reported that they were investing in Big Data and AI initiatives [Harvey 2018].

The rapidly increasing interest in the use of artificial intelligence (AI) is also another driver of Big Data. Machine learning is very dependent on the availability of large volumes of domain specific data. Without the availability of large data sets, it would be virtually impossible to have an effective AI system. So, we can readily see how the growth of AI and Big Data feed off each other, and perfectly complement each other – resulting in a virtuous circle that encourages both fields to grow rapidly. China's biggest fast food operation Yum China, for example, uses Big Data and AI very effectively. Yum China, which owns more than 8,400 KFC, Pizza Hut and Taco Bell restaurants, uses Big Data from their 180 million loyalty program members to drive its AI-powered menu that customizes the menu for each diner, based on preferences and local tastes. Since January 2019, these new systems have boosted the per-order spending by 1% , which amounts to about $840 million worth of fried chicken and pan pizzas per year [Ajello 2019].

Big Data also is very useful for medical studies. Big Data uses in healthcare include predictive modeling and clinical decision support; disease surveillance, public health, and research. Big Data analytics uses analytic methods developed in data mining, including classification, clustering, and regression, but are often complicated by many technical issues, such as missing values, curse of dimensionality, and bias control [Lee 2017]. As Table 2 [Lee 2017] depicts, there are significant differences between traditional medical analysis using statistics and medical big data analytics.

	Medical big data analysis	Classical statistical analysis
Application	Hypothesis-generating	Hypothesis-testing
Questions of interest	Overcoming the limitation of locally or temporally stable association with continually updating the data and algorithm	Trying to prove causal relationships
Domain knowledge	More important in interpretation of the results	Important both in collection of data and interpretation of the results
Sources of data	Any kind of sources; frequently multiple sources	Carefully specified collection of data; usually single source
Data collection	Recording without the direct supervision of a human	Human-based measurement recording
Coverage of data to be analyzed	Substantial fraction of entire population	Small data samples from a specific population with some assumptions of their distribution
Data size	Frequently huge	Relatively small
Nature of data	Unstructured and structured	Mainly structured
Data quality	Rarely clean	Quality controlled
Research questions of data analysis	May be different from those of data collection	Same as those of data collection
Underlying assumption of the model	Frequently absent	Based on various underlying probability distribution function
Analytic tools	Frequently automated with data mining algorithm	Manually by expert with classical statistics
Main outputs of analysis	Prediction, models, patterns identified	Statistical score contrasted against random chance
Privacy & ethics	Concerns about privacy and ethical issues	Data collection according to the pre-approved protocol; informed consent from the participants

Table 2: Statistical Versus Big Data Analytics in Medical Research

Modern successful enterprises also collect increasing amounts of data regarding all aspects of their supply and demand chains. Examples of data include logistics measures, vendor compliancy/lead times, POS data, inventory levels, prices, consumer behavior, demand forecasts, weather forecasts, and social media comments. Analyzing Big Data has helped businesses reduce inventory costs by up to 40 percent. [Krupnik 2013]. The ability to monitor and track real-time data sounds great, but making effective decisions quickly is likely more important. This is readily apparent for companies that are actually doing it, as they saw an increase in revenues and profits [Crossover 2018]. Predictive data analytics is fast becoming a tool to recognize key trends, patterns, and potential disruptions within supply chains, and a means to protect the enterprise's most valuable assets [Scott 2019].

In closing this section, it should be noted that decisions made from the analysis of Big Data can only be of high quality, if the underlying data itself is if good quality. The veracity dimension discussed before is very relevant to his aspect. High-quality data is a prime differentiator and is a valuable competitive asset that increases decision quality, efficiency, enhances customer service and drives profitability. Sadly, the bigger the data, the higher the chances of poor quality, and the cost of poor quality data is between 15% to 25% of revenue for many organizations [Leopold 2017].

Traditionally, companies have been shortsighted, when it comes to data quality by not having a full lifecycle view. They have implemented source system quality controls that only

address the point of origin, but that alone is not enough. Data quality initiatives have been one-off affairs at an IT level rather than collective efforts of both IT and the business side of the house. Failure to implement comprehensive and automated data cleansing processes that identify data quality issues on an ongoing basis results in organizations overspending on data quality and its related cleansing. A flexible data quality strategy is potentially required to tackle a broad range of generic and specific business rules and adhere to a variety of data quality standards. Data quality as a service (DQaaS) should be an integral part of data quality as it allows for a centralized approach. With a single update and entry point for all data controlled by data services, quality of data automatically improves, as there is a single best version of data.

SUMMARY AND CONCLUSIONS

This paper has provided a broad and useful discussion of Big Data for both practitioners and academics. The tremendous advantages of collecting Big Data, and analyzing it to gain insights and create competitive advantages is clearly getting a lot of attention in many modern successful organizations. Realizing this tremendous potential, and asking the right questions in a timely manner will help organizations collect the right type of data, and conduct the right type of analysis.

To better understand Big Data, it is useful to view it using the 4 dimensions – volume, velocity, variety and veracity, known as the 4 V's of Big Data. As organizations increase adoption rates and types of Big Data, they will need to pay careful attention to the 4 V's to maximize the benefits. As pointed out earlier, the collection of Big Data also allows organizations to initiate the adoption of artificial intelligence as well. The use of Big Data and AI opens up many new areas for research, as well as the need for identifying best practices. The use of these powerful technologies also opens up very important aspects like impacts on privacy, changes in society, and ethical uses of technology. Due to the tremendous potential it offers, the era of Big Data is here to stay for a long time.

REFERENCES

[Ajello 2019] Ajello, Robin and Kessenides, Dimitra; "KFC Aims to Keep Its China Edge With AI Menu, Robot Ice Cream Maker," Bloomberg News, March 2019. https://www.bloombergquint.com/businessweek/yum-china-is-building-the-kfc-of-the-future#gs.392c4z

[Cogn1 2012] Cognizant, "20-20 Insights, Big Data's Impact on the Data Supply Chain," White paper, May 2012. https://www.cognizant.com/InsightsWhitepapers/Big-Datas-Impact-on-the-Data-Supply-Chain.pdf

[Columbus 2018] Columbus, Louis, "10 Charts That Will Change Your Perspective Of Big Data's Growth," Forbes May 2018. https://www.forbes.com/sites/louiscolumbus/2018/05/23/10-charts-that-will-change-your-perspective-of-big-datas-growth/#23445d192926

[Coughlin 2018] Coughlin, Tom ,"Digital Storage Projections for 2019, Part 3," Forbes, Dec 2018. https://www.forbes.com/sites/tomcoughlin/2018/12/27/digital-storage-projections-for-2019-part-3/#53262abd29f8

[Crossover 2018] Crossover, "5 Ways Big Data Analytics Power Productivity & Profits," Jan 2018. https://medium.com/the-crossover-cast/5-ways-big-data-analytics-power-productivity-profits-41a17d802132

[Davenport 2017] Davenport, Thomas; "What's Your Data Strategy," Harvard Business Review; April 2017. https://hbr.org/webinar/2017/04/whats-your-data-strategy

[Domo 2017] Domo.com, "Data Never Sleeps 5.0," July 2017. https://web-assets.domo.com/blog/wp-content/uploads/2017/07/17_domo_data-never-sleeps-5-01.png

Dykes 2017] Dykes, Brent, "Big Data: Forget Volume and Variety, Focus On Velocity," Jun 2017. https://www.forbes.com/sites/brentdykes/2017/06/28/big-data-forget-volume-and-variety-focus-on-velocity/#6879b6646f7d

[Emani 2015] Emani, Cheikh; Cullot, Nadine; Nicolle, Christoph; "Understandable Big Data: A Survey," Computer Science Review, Vol. 17, 70-81, Aug 2015. https://www.sciencedirect.com/science/article/pii/S1574013715000064

[Gertner 2014] Gertner, Jon; "Behind GE's Vision For The Industrial Internet Of Things," FastCompany, June 2014. https://www.fastcompany.com/3031272/can-jeff-immelt-really-make-the-world-1-better

[Harvey 2018] Harvey, Cynthia; "Big Data Pros and Cons," Datamation August 2018; https://www.datamation.com/big-data/big-data-pros-and-cons.html

[IBM, 2019] IBM, "Extracting business value from the 4 V's of Big Data," https://www.ibmbigdatahub.com/infographic/extracting-business-value-4-vs-big-data

[Krupnik 2013] Krupnik, Yan, "Predictive Analytics—The Five Things You Need to Know," Supply and Demand Chain Executive, Feb 2013. https://www.sdcexec.com/sourcing-procurement/blog/10876460/predictive-analyticsthe-five-things-you-need-to-know

[Lee 2017] Lee, Choong Ho and Yun, Hyung-Jin, "Medical Big Data: Promise and Challenges," Kidney Research and Medical Practice, p 3-11, Vol 36, No. 1; Mar 2017. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5331970/

[Leopold 2017] Leopold, George; "As Data Quality Declines, Costs Soar," Datanami, Dec. 2017, https://www.datanami.com/2017/12/13/data-quality-declines-costs-soar/

[Liton 2018] Liton, Melissa; "How Much Data Comes From The Internet Of Things?;" Sumo Logic, Feb 2018. https://www.sumologic.com/blog/machine-data-analytics/iot-devices-data-volume/

[Lo 2018] Lo, Frank, "What is Big Data?," Datajobs 2018. https://datajobs.com/what-is-big-data

[Marr, 2018] Marr, Bernard, "How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read," Forbes May 20189. **https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#79aa612b60ba**

[McNeil, 2018] McNeil, Cassandra; "Veracity: The Most Important 'V' of Big Data," Gutcheck, Jan 2018. **https://www.gutcheckit.com/blog/veracity-big-data-v/**

[MongoDB 2016] MongoDB; "Big Data: Examples and Guidelines for the Enterprise Decision Maker," MongoDB Whitepaper, 2016; **http://s3.amazonaws.com/info-mongodb-com/10gen_Big_Data_White_Paper.pdf**

[O'Sullivan 2013] O'Sullivan, F.; "Smarter Supply Chain: Using Predictive Analytics to Combat Risk," An IBM Case Study September 24, 2013, IBM Corporation, April 14th, 2014, http://www. apics-triangle.org/backoffice/documents/FrancisOSullivan%20Presention.pdf

[Scott 2019] Scott, G. Stephenson; "Analytics and Risk Management for Supply Chain Efficiency," March 2019 http://www.verisk.com/Verisk-Review/Articles/Analytics-and-Risk-Management-for-Supply-Chain-Efficiency.html

[Taylor, 2018] Taylor, Christine; "Structured Versus Unstructured Data," Datamation, March 2018. **https://www.datamation.com/big-data/structured-vs-unstructured-data.html**

[Wilder 2012] Wilder, Edd, "An Introduction To The Big Data Landscape," Oreilly 2012, **https://www.oreilly.com/ideas/what-is-big-data**