A REVIEW PAPER ON BIG DATA ANALYTICS WITH ITS APPLICATIONS

Srikanth Gangadhara[1]
Associate Professor
CSE Department, VITS
Karimnagar,Telangana
gangadharasrikanth@gmail.com

Puram Srinivas[2]
Assistant Professor
CSE Department, VITS
Karimnagar,Telangana
puramvasu7@gmail.com

Podhili Thulasi[3]
Assistant Professor
CSE Department, VITS
Karimnagar,Telangana
thulasi.podhili@gmail.com

Shashi Rekha[4]
Assistant Professor
CSE Department, VITS
Karimnagar,Telangana
shashi.sharva@gmail.com

ABSTRACT:

Big data is data sets that are so voluminous and complex that traditional data processing application software are inadequate to deal with them. Big data challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating, information privacy and data source. There are five dimensions to big data known as Volume, Variety, Velocity and the recently added Veracity and Value. Lately, the term "big data" tends to refer to the use of predictive analytics, user behavior analytics, or certain other advanced data analytics methods that extract value from data, and seldom to a particular size of data set. "There is little doubt that the quantities of data now available are indeed large, but that's not the most relevant characteristic of this new data ecosystem." Analysis of data sets can find new correlations to "spot business trends, prevent diseases, and combat crime and so on." Scientists, business executives, practitioners of medicine, advertising and governments alike regularly meet difficulties with large data-sets in areas including Internet search, fintech, urban informatics, and business informatics. Big Data Analytics poses a grand challenge on the design of highly scalable algorithms and systems to integrate the data and uncover large hidden values from datasets that are diverse, complex, and of a massive scale.

Keywords*:* Data, Big Data, Analytics, Variety, Velocity, Volume, Visualization

1. INTRODUCTION: Big data is a blanket term for the non-traditional strategies and technologies needed to gather, organize, process, and gather insights from large datasets. While the problem of working with data that exceeds the computing power or storage of a single computer is not new, the pervasiveness, scale, and value of this type of computing has greatly expanded in recent years.

In this article, we will talk about big data on a fundamental level and define common concepts you might come across while researching the subject. We will also take a high-level look at some of the processes and technologies currently being used in this space. An exact definition of "big data" is difficult to nail down because projects, vendors, practitioners, and business professionals use it quite differently. With that in mind, generally speaking, big data is:

- large datasets
- the category of computing strategies and technologies that are used to handle large datasets

In this context, "large dataset" means a dataset too large to reasonably process or store with traditional tooling or on a single computer. This means that the common

scale of big datasets is constantly shifting and may vary significantly from organization to organization.

2. UNIQUENESS: The basic requirements for working with big data are the same as the requirements for working with datasets of any size. However, the massive scale, the speed of ingesting and processing, and the characteristics of the data that must be dealt with at each stage of the process present significant new challenges when designing solutions. The goal of most big data systems is to surface insights and connections from large volumes of heterogeneous data that would not be possible using conventional methods.

In 2001, Gartner's Doug Laney first presented what became known as the "three Vs of big data" to describe some of the characteristics that make big data different from other data processing:

2.1 Volume

The sheer scale of the information processed helps define big data systems. These datasets can be orders of magnitude larger than traditional datasets, which demands more thought at each stage of the processing and storage life cycle.

Often, because the work requirements exceed the capabilities of a single computer, this becomes a challenge of pooling, allocating, and coordinating resources from groups of computers. Cluster management and algorithms capable of breaking tasks into smaller pieces become increasingly important.

2.2 Velocity

Another way in which big data differs significantly from other data systems is the speed that information moves through the system. Data is frequently flowing into the system from multiple sources and is often expected to be processed in real time to gain insights and update the current understanding of the system.

This focus on near instant feedback has driven many big data practitioners away from a batch-oriented approach and closer to a real-time streaming system. Data is constantly being added, massaged, processed, and analyzed in order to keep up with the influx of new information and to surface valuable information early when it is most relevant. These ideas require robust systems with highly available components to guard against failures along the data pipeline.

2.3 Variety

Big data problems are often unique because of the wide range of both the sources being processed and their relative quality.

Data can be ingested from internal systems like application and server logs, from social media feeds and other external APIs, from physical device sensors, and from other providers. Big data seeks to handle potentially useful data regardless of where it's coming from by consolidating all information into a single system.

The formats and types of media can vary significantly as well. Rich media like images, video files, and audio recordings are ingested alongside text files, structured logs, etc. While more traditional data processing systems might expect data to enter the pipeline already labeled, formatted, and organized, big data systems usually accept and store data closer to its raw state. Ideally, any transformations or changes to the raw

data will happen in memory at the time of processing.

3. CHARACTERISTICS

Big data can be described by the following characteristics:

The eight (8) 'V' Dimension Characteristics of Big Data:
Part One: Volume, Velocity, Variety
Part Two: Variability, Veracity, Virality, Visualization and Value.

Volume

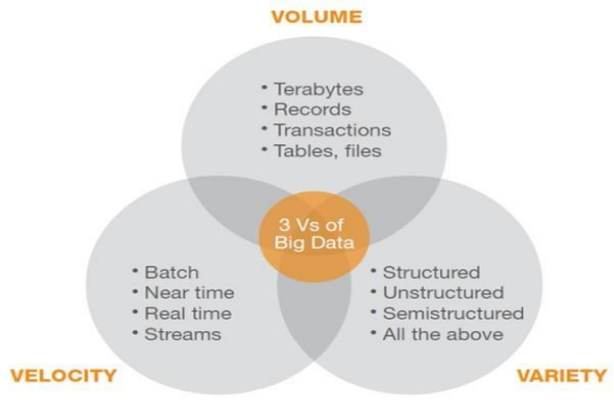
The quantity of generated and stored data. The size of the data determines the value and potential insight- and whether it can be considered big data or not.

Variety

The type and nature of the data. This helps people who analyze it to effectively use the resulting insight.

Velocity

In this context, the speed at which the data is generated and processed to meet the demands and challenges that lie in the path of growth and development.



3.1 Volume (amount of data the size of the data set)

Volume Refers to the vast amounts of data generated every second. We are not talking Terabytes but Zettabytes or Brontobytes. If we take all the data generated in the world between the beginning of time and 2008, the same amount of data will soon be generated every minute. This makes most data sets too large to store and analyze using traditional database technology. New big data tools use distributed systems so that we can store and analyze data across databases that are dotted around anywhere in the world.

3.2 Velocity (speed of data in and out or data in motion)
Velocity Refers to the speed at which new data is generated and the speed at which data moves around. Just think of social media messages going viral in seconds.
Technology allows us now to analyze the data while it is being generated (sometimes referred to as in-memory analytics), without ever putting it into databases.

3.3 Variety (range of data types, domains and sources)
Variety Refers to the different types of data we can now use. In the past we only focused on structured data that neatly fitted into tables or relational databases, such as financial data. In fact, 80% of the world's data is unstructured (text, images, video, voice, etc.) With big data technology we can now analyze and bring together data of different types such as messages, social media conversations, photos, sensor data, and video or voice recordings.

4. APPLICATIONS:

Big data has increased the demand of information management specialists so much so that Software AG, Oracle Corporation, IBM, Microsoft, SAP, EMC, HP and Dell have spent more than $15 billion on software firms specializing in data management and analytics. In 2010,

this industry was worth more than $100 billion and was growing at almost 10 percent a year: about twice as fast as the software business as a whole.

4.1 Government

The use and adoption of big data within governmental processes allows efficiencies in terms of cost, productivity, and innovation, but does not come without its flaws. Data analysis often requires multiple parts of government (central and local) to work in collaboration and create new and innovative processes to deliver the desired outcome.

4.2 International development

Research on the effective usage of information and communication technologies for development suggests that big data technology can make important contributions but also present unique challenges to International development. Advancements in big data analysis offer cost-effective opportunities to improve decision-making in critical development areas such as health care, employment, economic productivity, crime, security, and natural disaster and resource management. Additionally, user-generated data offers new opportunities to give the unheard a voice. However, longstanding challenges for developing regions such as inadequate technological infrastructure and economic and human resource scarcity exacerbate existing concerns with big data such as privacy, imperfect methodology, and interoperability issues.

4.3 Manufacturing

Based on TCS 2013 Global Trend Study, improvements in supply planning and product quality provide the greatest benefit of big data for manufacturing. Big data provides an infrastructure for transparency in manufacturing industry, which is the ability to unravel uncertainties such as inconsistent component performance and availability. Predictive manufacturing as an applicable

Approach toward near-zero downtime and transparency requires vast amount of data and advanced prediction tools for a systematic process of data into useful information. A conceptual framework of predictive manufacturing begins with data acquisition where different type of sensory data is available to acquire such as acoustics, vibration, pressure, current, voltage and controller data. Vast amount of sensory data in addition to historical data construct the big data in manufacturing. The generated big data acts as the input into predictive tools and preventive strategies such as Prognostics and Health Management (PHM).

4.4 Healthcare

Big data analytics has helped healthcare improve by providing personalized medicine and prescriptive analytics, clinical risk intervention and predictive analytics, waste and care variability reduction, automated external and internal reporting of patient data, standardized medical terms and patient registries and fragmented point solutions. Some areas of improvement are more aspirational than actually implemented. The level of data generated within healthcare systems is not trivial. With the added

adoption of mHealth, eHealth and wearable technologies the volume of data will continue to increase. This includes electronic health record data, imaging data, patient generated data, sensor data, and other forms of difficult to process data. There is now an even greater need for such environments to pay greater attention to data and information quality. "Big data very often means `dirty data' and the fraction of data inaccuracies increases with data volume growth." Human inspection at the big data scale is impossible and there is a desperate need in health service for intelligent tools for accuracy and believability control and handling of information missed. While extensive information in healthcare is now electronic, it fits under the big data umbrella as most is unstructured and difficult to use.

4.5 Education

A McKinsey Global Institute study found a shortage of 1.5 million highly trained data professionals and managers and a number of universities including University of Tennessee and UC Berkeley, have created masters programs to meet this demand. Private bootcamps have also developed programs to meet that demand, including free programs like The Data Incubator or paid programs like General Assembly.In the specific field of marketing, one of the problems stressed by Wedel and Kannan is that marketing has several subdomains (e.g., advertising, promotions, product development, branding) that all use different types of data. Because one-size-fits-all analytical solutions are not desirable, business schools should prepare marketing managers to have wide knowledge on all the different techniques used in these subdomains to get a big picture and work effectively with analysts.

4.6 Media

To understand how the media utilizes big data, it is first necessary to provide some context into the mechanism used for media process. It has been suggested by Nick Couldry and Joseph Turow that practitioners in Media and Advertising approach big data as many actionable points of information about millions of individuals. The industry appears to be moving away from the traditional approach of using specific media environments such as newspapers, magazines, or television shows and instead taps into consumers with technologies that reach targeted people at optimal times in optimal locations. The ultimate aim is to serve or convey, a message or content that is (statistically speaking) in line with the consumer's mindset. For example, publishing environments are increasingly tailoring messages (advertisements) and content (articles) to appeal to consumers that have been exclusively gleaned through various data-mining activities.

5. COMMON BENEFITS OF BIG DATA ANALYTICS TOOLS:

The main business advantages of big data generally fall into one of three categories: cost savings, competitive advantage, or new business opportunities.

5.1 Cost Savings: Big data tools like Hadoop allow businesses to store massive volumes of data at a much cheaper price tag than a traditional database. Companies utilizing big data tools for this benefit typically use Hadoop clusters to augment their current data warehouse, storing long-term data in Hadoop rather than expanding the data warehouse. Data is then moved

from Hadoop to the traditional database for production and analysis as needed. Versatile big data tools can also function as multiple tools at once, saving organizations on the cost of needing to purchase more tools for the same tasks.

5.2 Competitive Advantage: According to a survey of 540 enterprise decision makers involved in big data purchases by Webopedia's parent company QuinStreet, about half of all respondents said they were applying big data and analytics to improve customer retention, help with product development, and gain a competitive advantage. One of the major advantages of analyzing big data is that it gives businesses, particularly Data Analysts and Data Scientists, access to data that was previously unavailable or difficult to access. With increased access to data sources such as social media streams and clickstream data, businesses can better target their marketing efforts to customers, better predict demand for a certain product, and adapt marketing and advertising messaging in real-time. With these advantages, businesses are able to gain an edge on their competitors and act more quickly and decisively when compared to what rival organizations do. Needless to say, a business that effectively utilizes these analytics tools will be better prepared for the future than one that doesn't understand how important those tools are.

5.3 New Business Opportunities: The final benefit of such analytics tools is the possibility of exploring new business opportunities. Entrepreneurs have taken advantage of big data technology to offer new services in Adtech and Marketingtech. Data Analysts and Data Scientists at mature companies can also take advantage of the data they collect to offer add-on services or to create new product segments that offer additional value to their current customers. In addition to those benefits, big data analytics can pinpoint new or potential audiences that have yet to be tapped by the enterprise. Finding whole new customer segments can lead to tremendous new value.

These are just a few of the actionable insights made possible by available big data analytics tools. Big data insights help organizations boost sales and marketing results, uncover new revenue opportunity, improve customer service, optimize operational efficiency, reduce risk, and improve security.

6. REAL TIME ADVANTAGES:

There are many advantages of processing Big Data Analytics in real-time.

1. Knowing errors instantly within the organisation.
2. Implementing new strategies
3. To improve service dramatically
4. Fraud can be detected the moment it happens
5. Cost savings
6. Better sales insights
7. Keep up the customer trends

The advantages of processing Big Data in real-time are many:

6.1 Errors within the organisation are known instantly. Real-time insight into errors helps companies react quickly to mitigate the effects of an operational problem. This can save the operation from falling behind or failing completely or it can save your customers from having to stop using your products.

6.2 New strategies of your competition are noticed immediately. With Real-Time Big Data Analytics you can stay one step ahead

of the competition or get notified the moment your direct competitor is changing strategy or lowering its prices for example.

6.3 Service improves dramatically, which could lead to higher conversion rate and extra revenue. When organisations monitor the products that are used by its customers, it can pro-actively respond to upcoming failures. For example, cars with real-time sensors can notify before something is going wrong and let the driver know that the car needs maintenance.

6.4 Fraud can be detected the moment it happens and proper measures can be taken to limit the damage. The financial world is very attractive for criminals. With a real-time safeguard system, attempts to hack into your organisation are notified instantly. Your IT security department can take immediately appropriate action.

6.5 Cost savings: The implementation of a Real-Time Big Data Analytics tools may be expensive, it will eventually save a lot of money. There is no waiting time for business leaders and in-memory databases (useful for real-time analytics) also reduce the burden on a company's overall IT landscape, freeing up resources previously devoted to responding to requests for reports.

6.6 Better sales insights, which could lead to additional revenue. Real-time analytics tell exactly how your sales are doing and in case an internet retailer sees that a product is doing extremely well, it can take action to prevent missing out or losing revenue.

6.7 Keep up with customer trends: Insight into competitive offerings, promotions or your customer movements provides valuable information regarding coming and going customer trends. Faster decisions can be made with real-time analytics that better suit the (current) customer.

7. CONCLUSION:

In the information era we are currently living in, voluminous varieties of high velocity data are being produced daily, and within them lay intrinsic details and patterns of hidden knowledge which should be extracted and utilized. Hence, big data analytics can be applied to leverage business change and enhance decision making, by applying advanced analytic techniques on big data, and revealing hidden insights and valuable knowledge.By applying such analytics to big data, valuable information can be extracted and exploited to enhance decision making and support informed decisions. Consequently, some of the different areas where big data analytics can support and aid in decision making were examined. It was found that big data analytics can provide vast horizons of opportunities in various applications and areas, such as customer intelligence, fraud detection, and supply chain management.

8. REFERENCES

1. https://en.wikipedia.org/wiki/Big_data#Applications
2. https://www.bing.com/search?q=big+data+analytics&pc=MOZI&form=MOZSBR
3. http://searchbusinessanalytics.techtarget.com/definition/big-data-analytics
4. http://ieeexplore.ieee.org/document/6398180/?anchor=keywords
5. https://www.digitalocean.com/community/tutorials/an-introduction-to-big-data-concepts-and-terminology
6. https://www.greycampus.com/opencampus/big-data-developer/applications-of-big-data

7. https://en.wikipedia.org/wiki/Big_data#Case_studies
8. https://www.quora.com/What-are-the-advantages-of-Big-Data-Analytics