

Big Data Analytics: A Literature Review Paper

Nikhil Madaan¹

¹Department of Computer Science and Engineering,
Mangalmai institute of Engineering and Technology,
Greater Noida,

Umang Kumar²

²Department of Computer Science and Engineering,
Mangalmai institute of Engineering and Technology,
Greater Noida,

Suman Kr Jha³

³Department of Computer Science and Engineering,
Mangalmai institute of Engineering and Technology,
Greater Noida,

Abstract – In this modern era of computer's, a large amount of data is available to decision makers. Big data doesn't only refer to datasets that are big, but also high in velocity and variety, which is hard to handle using traditional tools and techniques. Due to speedy growth of such data, some ways are necessary to found to get important knowledge and values from these data sets. Also, decision makers need to gain some valuable vision from such big and continuously changing data, ranging from daily transactions to customer interactions and data of social network. Such vision can be given using Big Data Analytics, which is the application of Advanced Analytics Technique on big data. This paper aims to study some of the dis-similar analytics methods and tools which can be applied to big data, as well as the charge provided by the applications of big data analytics in different decision domain.

Keywords: big data, data processing, analytics, higher cognitive process.

1 INTRODUCTION

Imagine a world without data storage; an area where every detail a couple of persons or organization, every transaction performed, or every aspect which may be documented is lost directly after use. Organizations would thus lose the power to extract valuable information and knowledge, perform detailed analyses, furthermore as provide new opportunities and advantages. Anything starting from customer names and addresses, to products available, to purchases made, to employees hired, etc. has become essential for day-to-day continuity. Data is that the building block upon which any organization thrives.

Now think about the extent of details and therefore the surge of knowledge and knowledge provided nowadays through the advancements in technologies and therefore the internet. With the rise in storage capabilities and methods of knowledge collection, huge amounts of knowledge have become easily available. Every second, more and more data are being created and needs to be stored and analyzed so as to extract value. Furthermore, data has become cheaper to store, so

organizations have to get the maximum amount value as possible from the huge amounts of stored data.

The size, variety, and rapid change of such data require a replacement kind of big data analytics, furthermore as different storage and analysis methods. Such sheer amounts of huge data have to be properly analyzed, and pertaining information should be extracted.

The contribution of this paper is to produce an analysis of the available literature on big data analytics. Accordingly, a number of the varied big data tools, methods, and technologies which might be applied are discussed, and their applications and opportunities provided in several decision domains are portrayed. Our corpus mostly includes research from a number of the highest journals, conferences, and white papers by leading corporations within the industry. because of long review process of journals, most of the papers discussing big data analytics, its tools and methods, and its applications were found to be conference papers, and white papers. While big data analytics is being researched in academia, several of the economic advancements and new technologies provided were mostly discussed in industry papers.

2 BIG DATA ANALYTICS

The term "Big Data" has recently been applied to datasets that grow so large that they become awkward to figure with using traditional management systems. They are data sets whose size is beyond the flexibility of commonly used software tools and storage systems to capture, store, manage, similarly as process the info within a tolerable period [12]. Big data sizes are constantly increasing, currently starting from some dozen terabytes (TB) to several petabytes (PB) of knowledge in an exceedingly single data set. Consequently, some of the difficulties associated with big data include capture, storage, search, sharing, analytics, and visualizing. Today, enterprises are exploring large volumes of highly detailed data so as to discover facts they didn't know before [17].

Hence, big data analytics is where advanced analytic techniques are applied on big data sets. Analytics supported large data samples reveals

and leverages business change. However, the larger the set of knowledge, the harder it becomes to manage [17].

In this section, we'll start by discussing the characteristics of massive data, as well as its importance. Naturally, business benefit can commonly be derived from analyzing larger and more complex data sets that need real time or near-real time capabilities; however, this ends up in a desire for brand new data architectures, analytical methods, and tools.

Therefore, the successive section will elaborate the massive data analytics tools and methods, particularly, starting with the massive data storage and management, then moving on to the massive data analytic processing. It then concludes with a number of the assorted big data analyses which have grown in usage with big data.

2.1 Characteristics of Big Data

Big data is data whose scale, distribution, diversity, and/or timeliness require the employment of new technical architectures, analytics, and tools so as to enable insights that

unlock new sources of business value. Three main features characterize big data: volume, variety, and velocity, or the three V's. The quantity of the info is its size, and how enormous it's. Velocity refers to the speed with which data is changing, or how often it's created. Finally, variety includes the various formats and kinds of knowledge, as well because the different sorts of uses and ways of analyzing the information [9].

Data volume is that the primary attribute of huge data. Big data are often quantified by size in TBs or PBs, in addition as even the amount of records, transactions, tables, or files.

Additionally, one in every of the items that make big data really big is that it's coming from a greater form of sources than ever before, including logs, clickstreams, and social media.

Using these sources for analytics implies that common structured data is now joined by unstructured data, like text and human language, and semi-structured data, like extensible nomenclature (XML) or Rich Site Summary (RSS) feeds. There's also data, which is difficult to categorize since it comes from audio, video, and other devices.

Furthermore, multi-dimensional data are often drawn from an information warehouse to feature historic context to big data. Thus, with big data, variety is simply as big as volume.

Moreover, big data are often described by its velocity or speed. This can be basically the frequency of knowledge generation or the frequency of knowledge delivery.

The vanguard of big data is streaming data, which is collected in real-time from the websites [17]. Some researchers and organizations have discussed the addition of a fourth V, or veracity. Veracity focuses on the standard of the information. This characterizes big data quality pretty much as good, bad, or undefined because of data inconsistency, incompleteness,

ambiguity, latency, deception, and approximations [22].

2.2 Big Data Analytics Tools and Methods

With the evolution of technology and therefore the increased multitudes of knowledge flowing in and out of organizations daily, there has become a necessity for faster and more efficient ways of analyzing such data. Having piles of knowledge available is not any longer enough to create efficient decisions at the proper time.

Such data sets cannot be easily analyzed with traditional data management and analysis techniques and infrastructures. Therefore, there arises a necessity for brand spanking new tools and methods specialized for giant data analytics, in addition because the required architectures for storing and managing such data. Accordingly, the emergence of huge data has an effect on everything from the information itself and its collection, to the processing, to the final extracted decisions.

Consequently, [8] proposed the large – Data, Analytics, and Decisions (B-DAD) framework which includes the large data analytics tools and methods into the choice making process [8]. The framework maps the various big data storage, management, and processing tools, analytics tools and methods, and visualization and evaluation tools to the various phases of the choice making process. Hence, the changes related to big data analytics are reflected in three main areas: big data storage and architecture, data and analytics processing, and, finally, the large data analyses which may be applied for knowledge discovery and informed deciding. Each area is further discussed during this section. However, since big data continues to be evolving as a crucial field of research, and new findings and tools are constantly developing, this section isn't exhaustive of all the chances, and focuses on providing a general idea, instead of an inventory of all potential opportunities and technologies.

Big Data Storage and Management

One of the primary things organizations need to manage when handling big data, is where and the way this data is stored once it's acquired. The standard methods of structured data storage and retrieval include relational databases, data marts, and data warehouses. The info is uploaded to the storage from operational data stores using Extract, Transform, Load (ETL), or Extract, Load, Transform (ELT), tools which extract the info from outside sources, transform the info to suit operational needs, and finally load the info into the database or data warehouse. Thus, the info is cleaned, transformed, and catalogued before being made available for data processing and online analytical functions [3].

However, the large data environment necessitates Magnetic, Agile, Deep (MAD) analysis skills, which differ from the aspects of a standard Enterprise Data Warehouse (EDW) environment. First of all, traditional EDW approaches discourage the incorporation of new data sources until they're cleansed and

integrated. because of the ubiquity of knowledge nowadays, big data environments have to be magnetic, thus attracting all the info sources, regardless of the info quality [5]. Furthermore, given the growing numbers of knowledge sources, similarly because the sophistication of the info analyses, big data storage should allow analysts to simply produce and adapt data rapidly. this needs an agile database, whose logical and physical contents can adapt in sync with rapid data evolution [11]. Finally, since current data analyses use complex statistical methods, and analysts have to be able to study enormous datasets by drilling up and down, a giant data repository also has to be deep, and function a complicated algorithmic runtime engine [5].

Accordingly, several solutions, starting from distributed systems and big data processing (MPP) databases for providing high query performance and platform scalability, to non-relational or in-memory databases, are used for large data.

Non-relational databases, like Not Only SQL (NoSQL), were developed for storing and managing unstructured, or non-relational, data. NoSQL databases aim for massive scaling, data model flexibility, and simplified application development and deployment. Contrary to relational databases, NoSQL databases separate data management and data storage. Such databases rather concentrate on the high-performance scalable data storage, and permit data management tasks to be written within the application layer rather than having it written in database specific languages [3].

On the opposite hand, in-memory databases manage the info in server memory, thus eliminating disk input/output (I/O) and enabling real-time responses from the database. rather than using mechanical disk drives, it's possible to store the first database in silicon-based main memory. This leads to orders of magnitude of improvement within the performance, and allows entirely new applications to be developed [16].

Furthermore, in-memory databases are now being employed for advanced analytics on big data, especially to hurry the access to and scoring of analytic models for analysis. This provides scalability for large data, and speed for discovery analytics [17].

Alternatively, Hadoop may be a framework for performing big data analytics which provides reliability, scalability, and manageability by providing an implementation for the MapReduce paradigm, which is discussed within the following section, similarly as gluing the storage and analytics together. Hadoop consists of two main components: the HDFS for the large data storage, and MapReduce for large data analytics [9]. The HDFS storage function provides a redundant and reliable distributed filing system, which is optimized for big files, where one file is split into blocks and distributed across cluster nodes. Additionally, the info is protected among the

nodes by a replication mechanism, which ensures availability and reliability despite any node failures [3]. There are two forms of HDFS nodes: the info Nodes and also the Name Nodes. Data is stored in replicated file blocks across the multiple Data Nodes, and also the Name Node acts as a regulator between the client and also the Data Node, directing the client to the particular Data Node which contains the requested data [3].

Big Data Analytic Processing

After the massive data storage, comes the analytic processing. in line with [10], there are four critical requirements for large processing. the primary requirement is fast data loading. Since the disk and network traffic interferes with the query executions during data loading, it's necessary to cut back the info loading time. The second requirement is fast query processing. so as to satisfy the necessities of heavy workloads and real-time requests, many queries are response-time critical. Thus, the info placement structure must be capable of retaining high query processing speeds because the amounts of queries rapidly increase. Additionally, the third requirement for large processing is the highly efficient utilization of cupboard space. Since the rapid climb in user activities can demand scalable storage capacity and computing power, limited space necessitates that data storage be managed during processing, and issues on how to store the info in order that space utilization is maximized be addressed. Finally, the fourth requirement is that the strong adaptivity to highly dynamic workload patterns. As big data sets are analyzed by different applications and users, for various purposes, and in various ways, the underlying system should be highly adaptive to unexpected dynamics in processing, and not specific to certain workload patterns [10].

Map Reduce could be a parallel programming model, inspired by the "Map" and "Reduce" of functional languages, which is suitable for large processing. it's the core of Hadoop, and performs the info processing and analytics functions [6]. According to EMC, the MapReduce paradigm relies on adding more computers or resources, rather than increasing the facility or storage capacity of one computer; in other words, scaling out instead of scaling up [9]. the elemental idea of MapReduce is breaking a task down into stages and executing the stages in parallel so as to cut back the time needed to complete the task [6].

The first phase of the MapReduce job is to map input values to a group of key/value pairs as output. The "Map" function accordingly partitions large computational tasks into smaller tasks, and assigns them to the acceptable key/value pairs [6]. Thus, unstructured data, like text, will be mapped to a structured key/value pair, where, for instance, the key can be the word within the text and also the value is that the number

of occurrences of the word. This output is then the input to the “Reduce” function [9]. Reduce then performs the collection and combination of this output, by combining all values which share the identical key value, to supply the ultimate results of the computational task [6].

The MapReduce function within Hadoop depends on two different nodes: the work Tracker and also the Task Tracker nodes. The work Tracker nodes are those which are responsible for distributing the mapper and reducer functions to the available Task Trackers, moreover as monitoring the results [9]. The MapReduce job starts by the JobTracker assigning some of an input data on the HDFS to a map task, running on a node [13]. On the opposite hand, the Task Tracker nodes actually run the roles and communicate results back to the work Tracker. That communication between nodes is usually through files and directories in HDFS, so inter-node communication is minimized [9].

Figure 1 shows how the MapReduce nodes and therefore the HDFS work together. At step 1, there's an awfully large dataset including log files, sensor data, or anything of the types. The HDFS stores replicas of the info, represented by the blue, yellow, beige, and pink icons, across the info Nodes. In step 2, the client defines and executes a map job and a reduce job on a selected data set, and sends them both to the task Tracker. The task Tracker then distributes the roles across the Task Trackers in step 3. The Task Tracker runs the mapper, and therefore the mapper produces output that's then stored within the HDFS filing system. Finally, in step 4, the reduce job runs across the mapped data so as to supply the result.

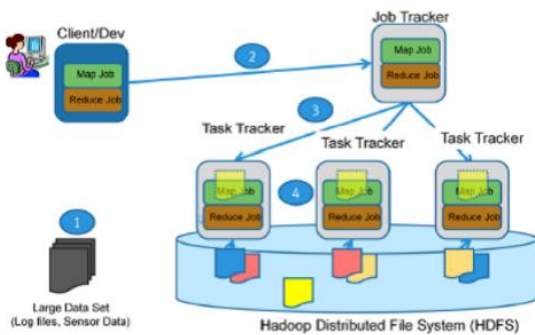


Fig. 1. MapReduce and HDFS

Hadoop could be a MAD system, thus making it popular for giant data analytics by loading data as files into the distributed classification system, and running parallel MapReduce computation on the info. Hadoop gets its magnetism and agility from the very fact that data is loaded into Hadoop just by copying files into the distributed classification system, and Map Reduce interprets the info at interval instead of loading time [11]. Thus, it's capable of attracting all data sources, in addition as adapting its engines to any evolutions that will occur in such big data sources [6].

After big data is stored managed, and processed, decision makers must extract useful insights by performing

big data analyses. within the subsections below, various big data analyses are going to be discussed, starting with selected traditional advanced data analytics methods, and followed by samples of a number of the extra, applicable big data analyses.

Big Data Analytics

Nowadays, people don't just want to gather data, they need to grasp the meaning and importance of the info, and use it to help them in making decisions. Data analytics are wont to extract previously unknown, useful, valid, and hidden patterns and data from large data sets, in addition on detect important relationships among the stored variables. Therefore, analytics have had a big impact on research and technologies, since decision makers became more and more curious about learning from previous data, thus gaining competitive advantage [21].

Along with a number of the foremost common advanced data analytics methods, such as association rules, clustering, classification and decision trees, and regression some additional analyses became common with big data.

For example, social media has recently become important for social networking and content sharing. Yet, the content that's generated from social media websites is enormous and remains largely unexploited. However, social media analytics may be used to analyze such data and extract useful information and predictions [2]. Social media analytics is predicated on developing and evaluating informatics frameworks and tools so as to gather, monitor, summarize, analyze, in addition as visualize social media data. Furthermore, social media analytics facilitates understanding the reactions and conversations between people in online communities, in addition as extracting useful patterns and intelligence from their interactions, additionally to what they share on social media websites [24].

On the opposite hand, Social Network Analysis (SNA) focuses on the relationships among social entities, in addition because the patterns and implications of such relationships [23]. An SNA maps and measures both formal and informal relationships so as to comprehend what facilitates the flow of data between interacting parties, such as who knows who, and who shares what knowledge or information with who and using what [19].

However, SNA differs from social media analysis, therein SNA tries to capture the social relationships and patterns between networks of individuals. On the opposite hand, social media analysis aims to investigate what social media users are saying so as to

uncover useful patterns, information about the users, and sentiments. this is often traditionally done using text mining or sentiment analysis, which are discussed below.

On the opposite hand, text mining is employed to investigate a document or set of documents in order to grasp the content within and also the meaning of the data contained. Text mining has become important nowadays since most of the data stored, not including audio, video, and images, consists of text. While data mining deals with structured data, text presents special characteristics which basically follow a non-relational form [18].

Moreover, sentiment analysis, or opinion mining, is becoming more and more important as online opinion data, like blogs, product reviews, forums, and social data from social media sites like Twitter and Facebook, grow tremendously. Sentiment analysis focuses on analyzing and understanding emotions from subjective text patterns, and is enabled through text mining. It identifies opinions and attitudes of people towards certain topics, and is beneficial in classifying viewpoints as positive or negative. Sentiment analysis uses linguistic communication processing and text analytics in order to spot and extract information by finding words that are indicative of a sentiment, in addition as relationships between words, in order that sentiments may be accurately identified [15].

Finally, from the strongest potential growths among big data analytics options is Advanced Data Visualization (ADV) and visual discovery [17]. Presenting information in order that people can consume it effectively could be a key challenge that must be met, in order for decision makers to be ready to properly analyze data during a thanks to result in concrete actions [14].

ADV has emerged as a robust technique to get knowledge from data. ADV combines data analysis methods with interactive visualization to enable comprehensive data exploration. It's a knowledge driven exploratory approach that matches well in situations where analysts have little knowledge about the info [20]. With the generation of more and more data of high volume and complexity, an increasing demand has arisen for ADV solutions from many application domains [25]. Additionally, such visualization analyses profit of human perceptual and reasoning abilities, which enables them to thoroughly analyze data at both the overview and also the detailed levels. Along with the dimensions and complexity of huge data, intuitive visual representation and interaction is required to facilitate the analyst's perception and reasoning [20].

ADV can enable faster analysis, better deciding, and simpler presentation and comprehension of results by providing interactive statistical graphics and a point-and-click interface [4]. Furthermore, ADV could be a natural suited big data since it can scale its visualizations to represent thousands or a lot of data points, unlike standard pie, bar, and line charts. Moreover, it can handle diverse data types, as well as present analytic data

structures that aren't easily flattened onto a monitor, such as hierarchies and neural nets. Additionally, most ADV tools and functions can support interfaces to all or any the leading data sources, thus enabling business analysts to explore data widely across a range of sources in search of the proper analytics dataset, usually in real-time [17].

3 BIG DATA ANALYTICS AND DECIDING

From the choice maker's perspective, the importance of huge data lies in its ability to provide information and knowledge important, upon which to base decisions. The managerial deciding process has been a vital and thoroughly covered topic in research throughout the years.

Big data is becoming an increasingly important asset for decision makers. Large volumes of highly detailed data from various sources like scanners, mobilephones, loyalty cards, the web, and social media platforms provide the chance to deliver significant benefits to organizations. This is often possible given that the info is correctly analyzed to reveal valuable insights, with decision makers to capitalize upon the resulting opportunities from the wealth of historic and real-time data generated through supply chains, production processes, customer behaviours, etc. [4].

Moreover, organizations are currently acquainted with analyzing internal data, such as sales, shipments, and inventory. However, the necessity for analyzing external data, such as customer markets and provide chains, has arisen, and also the use of huge data can provide cumulative value and knowledge. With the increasing sizes and kinds of unstructured data there, it becomes necessary to form more informed decisions based on drawing meaningful inferences from the info [7].

Accordingly, [8] developed the B-DAD framework which maps big data tools and techniques, into the choice making process [8]. Such a framework is meant to enhance the standard of the choice making process with reference to addressing big data. The first phase of the choice making process is that the intelligence phase, where data which can be wont to identify problems and opportunities is collected from internal and external data sources. During this phase, the sources of huge data must be identified, and also the data must be gathered from different sources, processed, stored, and migrated to the top user. Such big data must be treated accordingly, so after the info sources and kinds of information required for the analysis are defined, the chosen data is acquired and stored in any of the massive data storage and management tools previously discussed. After the massive data is acquired and stored, it's then organized, prepared, and

processed, this is often achieved across a high-speed network using ETL/ELT or big data processing tools, which are covered within the previous sections.

The next introduce the choice making process is that the design phase, where possible courses of action are developed and analyzed through a conceptualization, or a representative model of the matter. The framework divides this phase into three steps, model planning, data analytics, and analyzing. Here, a model for data analytics, such as those previously discussed, is chosen and planned, then applied, and at last analyzed.

Consequently, the subsequent introduce the choice making process is that the choice phase, where methods are wont to evaluate the impacts of the proposed solutions, or courses of action, from the planning phase. Finally, the last introduce the choice making process is that the implementation phase, where the proposed solution from the previous phase is implemented [8].

As the amount of huge data continues to exponentially grow, organizations throughout the various sectors are getting more curious about a way to manage and analyze such data. Thus, they're rushing to seize the opportunities offered by big data, and gain the foremost benefit and insight possible, consequently adopting big data analytics in order to unlock quantity and make better and faster decisions. Therefore, organizations are turning towards big data analytics so as to investigate huge amounts of data faster, and reveal previously unseen patterns, sentiments, and customer intelligence. This section focuses on a number of the various applications, both proposed and implemented, of huge data analytics, and the way these applications can aid organizations across different sectors to realize valuable insights and enhance deciding.

According to Manyika et al.'s research, big data can enable companies to make new products and services, enhance existing ones, in addition as invent entirely new business models. Such benefits may be gained by applying big data analytics in several areas, like customer intelligence, supply chain intelligence, performance, quality and risk management and fraud detection [14]. Furthermore, Cebr's study highlighted the main industries which will like big data analytics, like the manufacturing, retail, central government, healthcare, telecom, and banking industries [4].

3.1 Customer Intelligence

Big data analytics holds much potential for customer intelligence, and might highly benefit industries like retail, banking, and telecommunications. Big data can create transparency, and make relevant data more easily

accessible to stakeholders during a timely manner [14]. Big data analytics can provide organizations with the flexibility to profile and segment customers supported different socioeconomic characteristics, as well as increase levels of customer satisfaction and retention [4]. this could allow them to make more informed marketing decisions, and market to different segments supported their preferences together with the popularity of sales and marketing opportunities [17]. Moreover, social media may be wont to inform companies what their customers like, in addition as what they don't like. By performing sentiment analysis on this data, firms can be alerted beforehand when customers are turning against them or shifting to different products, and accordingly take action [7].

Additionally, using SNAs to observe customer sentiments towards brands, and identify influential individuals, can help organizations react to trends and perform direct marketing. Big data analytics may enable the development of predictive models for customer behaviour and buy patterns, therefore raising overall profitability [4]. Even organizations which have used segmentation for several years are beginning to deploy more sophisticated big data techniques, like real-time micro segmentation of shoppers, so as to focus on promotions and advertising [14]. Consequently, big data analytics can benefit organizations by enabling better targeted social influencer marketing, defining and predicting trends from market sentiments, as well as analyzing and understanding churn and other customer behaviours [17].

3.2 Supply Chain and Performance Management

As for supply chain management, big data analytics may be wont to forecast demand changes, and accordingly match their supply. this could increasingly benefit the manufacturing, retail, in addition as transport and logistics industries. By analyzing stock utilization and geospatial data on deliveries, organizations can automate replenishment decisions, which is able to reduce lead times and minimize costs and delays, as well as process interruptions. Additionally, decisions on changing suppliers, supported quality or price competitiveness, may be taken by analyzing supplier data to observe performance. Furthermore, alternate pricing scenarios may be run instantly, which might enable a discount in inventories and a rise in profit margins [4]. Accordingly, big data can result in the identification of the foundation causes of cost, and supply for better planning and forecasting [17].

Another area where big data analytics may be important is performance management, where the governmental and healthcare industries can easily benefit. With the increasing must improve productivity, staff performance information may be monitored and forecasted by using predictive analytics

tools. this could allow departments to link their strategic objectives with the service or user outcomes, thus resulting in increased efficiencies. Additionally, with the provision of huge data and performance information, in addition as its accessibility to operations managers, the utilization of predictive KPIs, balanced scorecards, and dashboards within the organization can introduce operational benefits by enabling the monitoring of performance, in addition as improving transparency, objectives setting, and planning and management functions [4].

3.3 Quality Management and Improvement

Especially for the manufacturing, energy and utilities, and telecommunications industries, big data may be used for quality management, so as to extend profitability and reduce costs by improving the standard of products and services provided. as an example, within the manufacturing process, predictive analytics on big data may be used to minimize the performance variability, in addition as prevent quality issues by providing early warning alerts. this could reduce scrap rates, and reduce the time to promote, since identifying any disruptions to the assembly process before they occur can save significant expenditures [4]. Additionally, big data analytics may result in manufacturing lead improvements [17]. Furthermore, real-time data analyses and monitoring of machine logs can enable managers to create swifter decisions for quality management. Also, big data analytics can yield the real-time monitoring of network demand, in addition to the forecasting of bandwidth in response to customer behaviour.

Moreover, healthcare IT systems can improve the efficiency and quality of care, by communicating and integrating patient data across different departments and institutions, while retaining privacy controls [4]. Analyzing electronic health records can improve the continuity of care of individuals, similarly as creating a large dataset through which treatments and outcomes will be predicted and compared. Therefore, with the increasing use of electronic health records, together with the advancements in analytics tools, there arises a chance to mine the available de-identified patient information for assessing the standard of healthcare, similarly as managing diseases and health services [22].

Additionally, the standard of citizens' lives will be improved through the employment of big data. For healthcare, sensors will be employed in hospitals and homes to produce the continuous monitoring of patients, and perform real-time analyses on the patient data streaming in. this could be used to alert individuals and their health care providers if any health anomalies are detected within the analysis, requiring the patient to seek medical help [22]. Patients also can be monitored remotely to investigate their adherence to their prescriptions, and improve drug and treatment options [14].

Moreover, by analyzing information from distributed sensors on handheld devices, roads, and vehicles, which offer real-time traffic information, transportation will be transformed and improved. Traffic jams will be predicted and prevented, and drivers can operate more safely and with less disruption to the traffic flow. Such a replacement type of traffic ecosystem, with "intelligent" connected cars, can potentially renovate transportation and the way roadways are used [22]. Accordingly, big data applications can provide smart routing, in keeping with real-time traffic information supported personal location data. Furthermore, such applications can automatically demand help when trouble is detected by the sensors, and inform users about accidents, scheduled roadwork, and congested areas in real-time [14].

Furthermore, big data will be used for better understanding changes within the location, frequency, and intensity of weather and climate. this could benefit citizens and businesses that rely on weather, like farmers, similarly as tourism and transportation companies. Also, with new sensors and analysis techniques for developing future climate models and nearer weather forecasts, weather related natural disasters will be predicted, and preventive or adaptive measures will be taken beforehand [22].

3.4 Risk Management and Fraud Detection

Industries like investment or retail banking, similarly as insurance, can use big data analytics within the area of risk management. Since the evaluation and bearing of risk could be a critical aspect for the financial services sector, big data analytics can help in selecting investments by analyzing the likelihood of gains against the likelihood of losses. Additionally, internal and external big data will be analyzed for the complete and dynamic appraisal of risk exposures [4]. Accordingly, big data can benefit organizations by enabling the quantification of risks [17]. High-performance analytics also can be used to integrate the danger profiles managed in isolation across separate departments, into enterprise wide risk profiles. this will aid in risk mitigation, since a comprehensive view of the various risk types and their interrelations is provided to decisionmakers [4].

Furthermore, new big data tools and technologies can provide for managing the exponential growth in network produced data, furthermore reduce database performance problems by increasing the power to scale and capture the desired data. Along with the enhancement in cyber analytics and data intensive computing solutions, organizations can incorporate multiple streams of knowledge and automatic analyses to guard themselves against cyber and network attacks [22].

As for fraud detection, especially within the government, banking, and insurance industries, big data analytics may be wont to detect and forestall fraud [17]. Analytics are already commonly employed in automated fraud detection, but organizations and sectors are looking towards harnessing the potentials of huge data so as to enhance their systems. Big data can allow them to match electronic data across several sources, between both public and personal sectors, and perform faster analytics [4].

In addition, customer intelligence may be wont to model normal customer behaviour, and detect suspicious or divergent activities through the accurate flagging of outlier occurrences. Furthermore, providing systems with big data about prevailing fraud patterns can allow these systems to find out the new sorts of frauds and act accordingly, as the fraudsters adapt to the old systems designed to detect them. Also, SNAs may be used to identify the networks of collaborating fraudsters, furthermore as discover evidence of fraudulent insurance or benefits claims, which is able to result in less fraudulent activity going undiscovered [4]. Thus, big data tools, techniques, and governance processes can increase the prevention and recovery of fraudulent transactions by dramatically increasing the speed of identification and detection of compliance patterns within all available data sets [22].

4 CONCLUSION

In this research, we've got examined the innovative topic of huge data, which has recently gained many interests thanks to its perceived unprecedented opportunities and benefits. within the information era we are currently living in, voluminous sorts 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 may be applied to leverage business change and enhance deciding, by applying advanced analytic techniques on big data, and revealing hidden insights and valuable knowledge.

Accordingly, the literature was reviewed so as to produce an analysis of the large data analytics concepts which are being researched, furthermore as their importance to decision making. Consequently, big data was discussed, furthermore as its characteristics and importance. Moreover, a number of the large data analytics tools and methods specifically were examined. Thus, big data storage and management, furthermore as big data analytics processing were detailed. additionally, a number of the various advanced data analytics techniques were further discussed.

By applying such analytics to big data, valuable information will be extracted and exploited to reinforce higher cognitive process and support informed decisions. Consequently, some of the various areas where big data analytics can support and aid in decisionmaking were examined. it had been found that big data analytics can provide vast horizon of opportunities in various applications and areas, like customer intelligence, fraud detection, and provide chain management. Additionally, its benefits can serve different sectors and industries, like healthcare, retail, telecom, manufacturing, etc.

Accordingly, this research has provided the people and therefore the organizations with samples of the assorted big data tools, methods, and technologies which may be applied. This gives users a concept of the mandatory technologies required, further as develops an idea of what they will do to supply more enhanced solutions for giant data analytics in support of higher cognitive process. Thus, the support of massive data analytics to decisionmaking was depicted.

Finally, any new technology, if applied correctly can bring with it several potential benefits and innovations, including big data, which could be a remarkable field with a bright future, if approached correctly. However, big data is incredibly difficult to accommodate. It requires proper storage, management, integration, federation, cleansing, processing, analyzing, etc. With all the issues faced with traditional data management, big data exponentially increases these difficulties because of additional volumes, velocities, and varieties of data and sources which must be restrained. Therefore, future research can specialize in providing a roadmap or framework for giant data management which may encompass the previously stated difficulties.

We believe that big data analytics is of great significance during this era of information overflow, and might provide unforeseen insights and benefits to decision makers in various areas. If properly exploited and applied, big data analytics has the potential to supply a basis for advancements, on the scientific, technological, and humanitarian levels.

REFERENCES

- [1] Adams, M.N.: Perspectives on Data Mining. International Journal of Market Research 52(1), 11–19 (2010)
- [2] Asur, S., Huberman, B.A.: Predicting the Future with Social Media. In: ACM International Conference on Web Intelligence and Intelligent Agent Technology, vol. 1, pp. 492–499 (2010)
- [3] Bakshi, K.: Considerations for Big Data: Architecture and Approaches. In: Proceedings of the IEEE Aerospace Conference, pp. 1–7 (2012)
- [4] Cebr: Data equity, Unlocking the value of big data. in: SAS Reports, pp. 1–44 (2012)
- [5] Cohen, J., Dolan, B., Dunlap, M., Hellerstein, J.M., Welton, C.: MAD Skills: New Analysis Practices for Big Data. Proceedings of the ACM VLDB Endowment 2(2), 1481–1492 (2009)
- [6] Cuzzocrea, A., Song, I., Davis, K.C.: Analytics over Large-Scale Multidimensional Data: The Big Data Revolution! In: Proceedings of the ACM International Workshop on Data Warehousing and OLAP, pp. 101–104 (2011)

-
- [7] Economist Intelligence Unit: The Deciding Factor: Big Data & Decision Making. In: Capgemini Reports, pp. 1–24 (2012)
- [8] Elgendy, N.: Big Data Analytics in Support of the Decision-Making Process. MSc Thesis, German University in Cairo, p. 164 (2013)
- [9] EMC: Data Science and Big Data Analytics. In: EMC Education Services, pp. 1–508 (2012)
- [10] He, Y., Lee, R., Huai, Y., Shao, Z., Jain, N., Zhang, X., Xu, Z.: RCFile: A Fast and Spaceefficient Data Placement Structure in MapReduce-based Warehouse Systems. In: IEEE International Conference on Data Engineering (ICDE), pp. 1199–1208 (2011)
- [11] Herodotou, H., Lim, H., Luo, G., Borisov, N., Dong, L., Cetin, F.B., Babu, S.: Starfish: A Self-tuning System for Big Data Analytics. In: Proceedings of the Conference on Innovative Data Systems Research, pp. 261–272 (2011)
- [12] Kubick, W.R.: Big Data, Information and Meaning. In: Clinical Trial Insights, pp. 26–28 (2012)
- [13] Lee, R., Luo, T., Huai, Y., Wang, F., He, Y., Zhang, X.: Ysmart: Yet Another SQL-toMapReduce Translator. In: IEEE International Conference on Distributed Computing Systems (ICDCS), pp. 25–36 (2011)
- [14] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A.H.: Big Data: The Next Frontier for Innovation, Competition, and Productivity. In: McKinsey Global Institute Reports, pp. 1–156 (2011)
- [15] Mouthami, K., Devi, K.N., Bhaskaran, V.M.: Sentiment Analysis and Classification Based on Textual Reviews. In: International Conference on Information Communication and Embedded Systems (ICICES), pp. 271–276 (2013)
- [16] Plattner, H., Zeier, A.: In-Memory Data Management: An Inflection Point for Enterprise Applications. Springer, Heidelberg (2011)
- [17] Russom, P.: Big Data Analytics. In: TDWI Best Practices Report, pp. 1–40 (2011)
- [18] Sanchez, D., Martin-Bautista, M.J., Blanco, I., Torre, C.: Text Knowledge Mining: An Alternative to Text Data Mining. In: IEEE International Conference on Data Mining Workshops, pp. 664–672 (2008)
- [19] Serrat, O.: Social Network Analysis. Knowledge Network Solutions 28, 1–4 (2009)
- [20] Shen, Z., Wei, J., Sundaresan, N., Ma, K.L.: Visual Analysis of Massive Web Session Data. In: Large Data Analysis and Visualization (LDAV), pp. 65–72 (2012)
- [21] Song, Z., Kusiak, A.: Optimizing Product Configurations with a Data Mining Approach. International Journal of Production Research 47(7), 1733–1751 (2009)
- [22] TechAmerica: Demystifying Big Data: A Practical Guide to Transforming the Business of Government. In: TechAmerica Reports, pp. 1–40 (2012)
- [23] Van der Valk, T., Gijsbers, G.: The Use of Social Network Analysis in Innovation Studies: Mapping Actors and Technologies. Innovation: Management, Policy & Practice 12(1), 5–17 (2010)
- [24] Zeng, D., Hsinchun, C., Lusch, R., Li, S.H.: Social Media Analytics and Intelligence. IEEE Intelligent Systems 25(6), 13–16 (2010)
- [25] Zhang, L., Stoffel, A., Behrisch, M., Mittelstadt, S., Schreck, T., Pompl, R., Weber, S., Last, H., Keim, D.: Visual Analytics for the Big Data Era—A Comparative Review of State-of-the-Art Commercial Systems. In: IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 173–182 (2012)