An Analytical Review on Big Data in a Diversified Approach

Dr.B.Umadevi[1], Dr.D.Sundar[2], B.Sakthi[3]

[1]Assistant Professor, P.G.Department of Computer Science, Government Arts College, Melur, Madurai District, TamilNadu, India
[2]Assistant Professor, P.G.Department of Computer Science, Government Arts College, Melur, Madurai District, TamilNadu, India
[3]Assistant Professor, Department of MCA, V.V.Vanniapeurmal College for Women, Virudhunagar, Tamilnadu, India

---

Abstract - *The rapid climb of the latest technologies, modernization of application and therefore the dimensionless growth of communication industries produces multivolume of data in every year. Data comes from various sources at different rates. Big data is often Structured, Unstructured or Semi-structured, leading to the incapability of conventional data management methods. This provides rise to the new era of big data. This paper focuses on the analytical view made by various researchers within the domain of massive data. Hadoop framework is employed to process large data. It primarily focuses the main components of the Framework, Characteristics, Technologies, Solutions and a few of the Challenges on Big Data.*

Key Words: Data Mining, Big Data, Data analytics, Hadoop, Data Science.

1. INTRODUCTION

The digital world produces various data in several formats either in min or max structure, which can extend up to some Exabyte. Handling the info of mega size is a tedious in storage and as well as in manages. With the rise in storage capabilities and methods of data collection, huge amounts of data became easily available. Additionally, thereto data has become cheaper to store, so organizations got to get the maximum amount value as possible from large amounts of stored data. The size, variety, and rapid change of such data require a replacement sort of big data analytics also as different storage and analysis methods. Such complete amounts of massive data got to be properly analyzed, and pertaining information should be extracted. These technologies, methodologies, and skills can help organizations gain additional insight [1] about customers and operations.

2. BACKGROUND AND RELATED WORKS

Big data is a collection of data sets or a combination of data sets. The concept of big data has been widespread within digital communication and information science since the earliest days of computing. Big data is growing day by day because data is made by everyone and for everything from mobile devices, call centers, web servers, and social networking sites. But the challenge is that it's overlarge, too fast and hard to handle for traditional database and existing technologies.

Many organizations gather the huge amounts of data generated from high-volume transactions like call centers, sensors, web logs, and digital images. The completion of their large business depends on meeting big data challenges while continually improving operational efficiency. Most of the researchers have contributed their research to effectively manage large data. The study associated with big data, and its characteristics are analyse at various levels, and therefore, the comprehensive literature review of varied researchers' works are stated below:

2.1 Literature Review

Shan Ren, Yingfeng Zhang, Yang Liu, Tomohiko sakao, Donald Huisingh [2] and Cecilia M.V.B. Almeida were proposed a framework which analyzed potential applications and key advantages. They also discussed the present challenges and future research direction provided valuable insights for academia and industry.

G. Devateja, P. V. B. Kashyap, C. Suraj, C. Harshavardhan and Impana Appaji were analyzed and discussed about the Hadoop [3] and big data in the organizations that too data present in different machines at different locations.

M. Ahmed and Shahat Osman has offered a completely unique big data analytics framework for smart cities called "Smart City Data Analytics Panel (SCDAP)". They adopted a scientific literature review on big data analytics frameworks in smart cities. In their research study [4] they proposed new functionalities to big data analytics frameworks represented in data model management and aggregation to form comparison between lore discovery approaches.

S. Arockia Panimalar, S. Varnekha Shree and A. Veneshia Kathrine were identified [5] and defined the fourteen characteristics of big data, and a new three characteristics of massive data, and a replacement three characteristics of massive data has been explored further to handle big data efficiently.

Ritu Agarwal and Vasant Dhar addressed the key questions associated with the explosion of interest within the emerging fields of big data, analytics, and data science [6]. They also discussed the novelty of the fields, the strengths of the information systems (IS), the interesting research questions for IS scholars, the role of predictive and explanatory modeling, and therefore, the emerging area for the contribution and significance in their study.

Yichuan Wang, LeeAnn Kung, William Yu Chung Wang and Casey G. Cegielski examined the historical development, architectural design and component functionalities of big data analytics. They analyze different big data implementation cases in healthcare [7]. They also mapped the advantages driven by the big data analytics concerning information technology (IT) infrastructure, operational, organizational, managerial, and strategic areas to formulate simpler data-driven analytics strategies.

Sahar Sohangir, Dingding Wang, Anna Pomeranets and Taghi M. Khoshgoftaar were formulated inquiries close to determine whether the deep learning [8] models are often adapted to enhance the performance of sentiment analysis for StockTwits. They used variety of neural network models like long short-term memory, doc2vec, and convolutional neural networks to stock exchange opinions posted in StockTwits to predict sentiment of authors in StockTwits dataset.

3. TRADITIONAL APPROACH TO DATA MINING

The process of data mining [9] includes the operations like selection, preprocessing, transformation and evaluation of data within the discovery of knowledge. The primary task in data mining is data input which incorporates collecting, selecting, preprocessing the data. Preprocessing includes [10] cleaning and filtering the data to form it useful for further activities. After the data is cleaned and reduced from various data mining [11] methods like clustering, classification, association rules and sequential patterns are often applied for data analysis.

Most of the methods can't be applied to big data due to the subsequent reasons:

(i) Designed to work with a single machine with all the data within the memory.

(ii) Some methods aren't for huge and complex data.

(iii) The methods cannot produce the analysis dynamically supported the input.

(iv) It works with an equivalent format of input.

The output is going to be generated by applying these methods, evaluation, and interpretation and also to measure the results. The output are often measured regarding the number of errors, accuracy of results, computation speed, computation cost, response time, utilization of memory, etc. Hence the knowledge generation becomes complex and need to be more versatile for handling the big data.

4. BIG DATA ANALYTICS

Big data analysis emphasizes to uncover hidden patterns, correlations and other insights from the large volume of data. [12]. within the midst of today's technology, it's possible to analyze your data and obtain answers from it soon. An attempt that's slower and fewer efficient with more traditional business intelligence solutions.

The advanced analytic techniques against very large, diverse data sets such as structured, semi-structured and unstructured data from diverse sources are being used by Big data analytics, which falls in the range from terabytes to zeta bytes.

4.1 Big Data

Big data is a term applied to data sets whose size or type is beyond the power of traditional relational databases to capture, manage and process the data with low latency [13]. Big data has one or more of the subsequent characteristics: high volume, high velocity or high variety. Artificial intelligence (AI), mobile, social and therefore, the Internet of Things (IoT) are driving data complexity through new forms and sources of knowledge. In favor of, big data [13] come from

sensors, devices, video or audio, networks, log files, transactional applications, web, and social media are generated in real time, and at a really large scale.

Analysis of massive data allows analysts, researchers and business users to form better and faster decisions using data that were previously inaccessible or unusable. Businesses can use advanced analytic techniques [12] like text analytics, machine learning, predictive analytics, data mining, deep learning, statistics, and natural language processing to realize new insights from previously untapped data sources independently or along together with existing enterprise data.

4.2 Big Data Characteristics

Big Data is important, because it enables many organizations to collect, store, manage, and manipulate vast amounts of data at the right speed, at the right time, to gain the right insights. Characteristics of Big Data by what is usually referred to as a multi V model which is discussed below:

(i) 3 V's of Big Data (Volume, Velocity, and Variety)

(ii) 4 V's of Big Data (Volume, Velocity, Variety and Veracity)

(iii) 5 V's of Big Data (Volume, Velocity, Variety, Variability and Value)

(iv) 10 V's of Big Data (Volume, Variety, Velocity, Veracity, Validity, Value, Variability,

Venue, Vocabulary, and Vagueness).

(v) 15 V's of Big Data (Volume, Velocity, Value, Variety, Veracity, Validity, Volatility, Visualization, Virality, Viscosity, Variability, Venue, Vocabulary, Vagueness, Complexity).

(vi) 18 V's of Big Data (Volume, Velocity, Value, Variety, Veracity, Validity, Volatility, Visualization, Virality, Viscosity, Variability, Venue, Vocabulary, Vagueness, Complexity, Verbosity, Voluntariness, Versatility).

4.3 Technologies used for Big Data

Big Data technology are often defined as a Software-Utility that's designed to Analyze, Process and Extract the knowledge from a particularly complex and enormous data sets which the normal, or traditional data processing software could never deal with. The technology [14] is a combination of several techniques, and processing methods of big data analytics.

The fastest growth of data volume within banking, healthcare, insurance, securities and investment services, and telecommunications are happening in big data technologies. It's noteworthy that three of these industries lie within the financial sector which has many particularly strong use cases for giant data analytics, such as fraud detection, risk management and customer service optimization. The list of technology vendors offering big data solutions [15] is seemingly infinite. Many of the big data solutions that are particularly popular right now fit into one of the following categories:

(i) The Hadoop Ecosystem

Hadoop Ecosystem is a platform or a suite which provides various services to solve the big data problems. It includes Apache projects and various commercial tools and solutions. There are four major elements of Hadoop such as HDFS, MapReduce, YARN, and Hadoop Common.

(ii) Spark

Apache Spark is a lightning-fast cluster computing technology, designed for fast computation. It is based on Hadoop MapReduce and it extends the MapReduce model to efficiently use it for more types of computations, which includes interactive queries, and stream processing. The main feature of Spark is its in-memory cluster computing that increases the processing speed of an application.

Spark is designed to cover a wide range of workloads such as batch applications, iterative algorithms, interactive queries and streaming. Apart from supporting all these workloads in a respective system, it reduces the management burden of maintaining separate tools.

(iii) R

R is a leading programming language of data science, consisting of powerful functions to tackle all problems related to Big Data processing. It will further expand to include Big Data tools such as Apache Hadoop ecosystem, HDFS and MapReduce frameworks.

(iv) Python

Python is a general-purpose programming language which enables programmers to write fewer lines of codes and make it more readable. It has scripting features and besides that uses many advanced libraries such as NumPy, Matplotlib, and SciPy which makes it useful for scientific computing.

The Python big data is a compatible, similarly Hadoop and big data are synonymous with each other. Hence, Python has been made inherently compatible with Hadoop to work with big data. Python consists of Pydoop package which help in accessing HDFS API and also writing Hadoop MapReduce programming.

(v) Data Lakes

A data lake is architecture for storing high-volume, high-velocity, high-variety, as-is data in a centralized repository for Big Data, and real-time analytics. Companies can pull in vast amounts of data such as structured, semi structured and unstructured. In real time into a data lake, from anywhere.

(vi) NoSQL Databases

NoSQL database is for distributed data store with humongous data storage needs. NoSQL is used for Big data and real-time web apps. Instead, a NoSQL database system encompasses a wide range of database technologies that can store structured, semi-structured, unstructured and polymorphic data.

(vii) Predictive Analytics

Predictive Analytics is an enabler of big data: Businesses collect vast amounts of real-time customer data, and predictive analytic uses this historical data, combined with customer insight, to predict future events.

(viii) In-Memory Databases

An in-memory database (sometimes abbreviated to db) is based on a database management system that stores its data collections directly in the working memory of one or more computers. Using RAM has a key advantage in that in-memory databases have significantly faster accessed speeds.

(ix) Big Data Security Solutions

Big data security solutions let customers secure their big data deployments, whether it's a Hadoop infrastructure or a non-relational (NoSQL) database such as MongoDB or Couchbase without getting on the way of the analytics tools that make these solutions important.

(x) Big Data Governance Solutions

Data Governance is a centralized control mechanism to manage data availability, security, usability, and integrity.

(xi) Self-Service Capabilities

Self-Service Analytics or ad hoc reporting gives users the ability to develop rapid reports, empowering users to analyze their data. With JReport's self-service BI capability, end users can quickly get started by building their reports.

(xii) Artificial Intelligence

Artificial Intelligence (AI) algorithms enhance the ability for big data analytics and IoT platforms to provide value to each of these market segments. AI coupled with advanced big data analytics provides the ability to make raw data meaningful and useful as information for decision-making purposes.

(xiii) Streaming Analytics

Streaming analytics involves knowing and acting upon events happening in your business at any given moment. Since Streaming Analytics occurs immediately, companies must act on the analytics data quickly within a small window of opportunity before the data lose its value.

(xiv) Edge Computing

An edge computing application utilizes the IoT (Internet of Things) devices' processing power to score, aggregate, pre-process or filter IoT data. It utilizes the flexibility and power of cloud services to execute complex analytics on that data and, in a feedback loop, support actions and decisions on and about the physical world.

Edge Computing solution, a computer connected to a camera can automatically strip out demographic information, sending that to the cloud for storage and processing. This dramatically cuts down on the amount of data collected, providing purely useful information.

(xv) Blockchain

Blockchain is a distributed database system that acts as an "open ledger" to store, and manage transactions. Each record in the database is called a block and contains details such as the transaction timestamp as well as a link to the previous block.

(xvi) Prescriptive Analytics

Prescriptive analytics uses the latest technologies such as machine learning and artificial intelligence to understand what the impact is of future decisions and uses those scenarios to determine the best outcome.

5. HADOOP SOLUTIONS FOR BIG DATA ANALYTICS

Hadoop is one of the most highly flexible storage platforms, and also a programming framework developed for distributed storage and distributed processing of large sets of data in a distributed computing environment. Hadoop was developed by Google's MapReduce that is a software framework where applications break downs into various parts. The Apache Hadoop ecosystem consists of the Hadoop Kernel, MapReduce, HDFS and numbers of various components such as Apache Hive, Base and Zookeeper. Hadoop primarily consists of two Parts which is mentioned in Fig - 1:

(i) File System (The Hadoop File System) and
(ii) Programming Paradigm (Map Reduce)

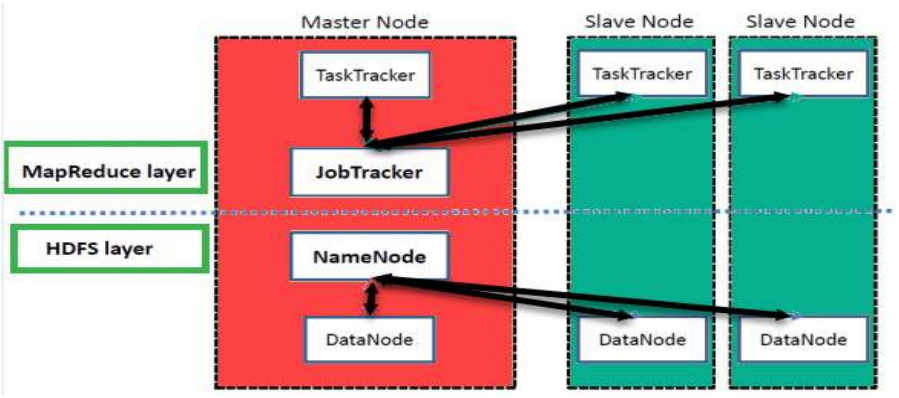

Fig - 1: Hadoop Architecture

5.1 HDFS Architecture

Hadoop includes a fault-tolerant storage system called the Hadoop Distributed File System, or HDFS which is given in Fig – 2. [16]. HDFS is able to store huge amounts [17] of information, scale up incrementally, and survive the failure of significant parts of the storage infrastructure without losing the data. Hadoop creates clusters of machines and coordinates work among them. Clusters can be built with inexpensive computers. If one fails, Hadoop continues to operate the cluster without losing the data or interrupting work, by shifting work to the remaining machines in the cluster. HDFS manages storage on the cluster by breaking incoming files into pieces, called "blocks", and storing each of the blocks redundantly across the pool of servers [17]. In the common case, HDFS stores three complete copies of each file by copying each piece to three different servers.

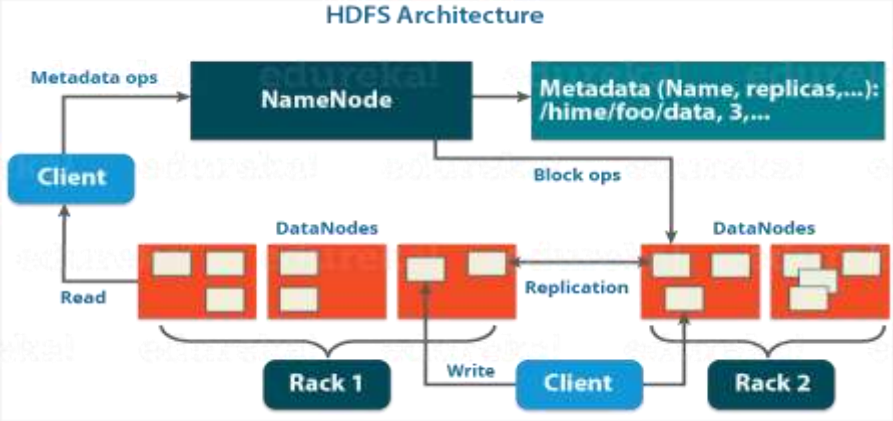

Fig - 2: HDFS Architecture

5.2 Map Reduce Architecture

Mapreduce is a software framework and programming model which is specified in Fig – 3 that allows users to perform distribute and parallel processing in large data sets in a distributed environment. It has two phases, namely, Map and Reduce. Map tasks deal with splitting and mapping of data while Reduce tasks shuffle and reduce the data. The name MapReduce suggests, the reducer phase takes place after the mapper phase has been completed.

The first is the map job, where a block of data is read and processed to produce key-value pairs as intermediate outputs [15]. The output of a Mapper or map job (key-value pairs) is input to the Reducer. The reducer receives the key-value pair from multiple maps jobs. Then, the reducer aggregates those intermediate data tuples (intermediate key-value pair) into a smaller set of tuples or key-value pairs which is the final output.

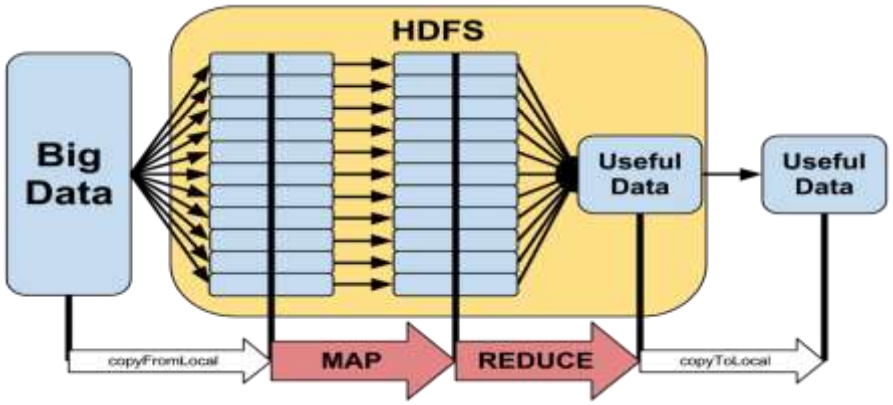

Fig - 3: Map Reduce Architecture

6. CHALLENGES OF BIG DATA ANALYTICS

The traditional data processing applications are inadequate for handling Big Data of large and complex datasets. The integration of the huge data sets is quite complex. There are several challenges [18] one can face during this integration such as analysis, data curation, capture, sharing, search, visualization, information privacy and storage. Accuracy in managing big data will lead to more confident decision-making [18]. The integration of big data, and its copious challenges that can be faced during the process.

(i) The Uncertainty of Data Management

The innovative data management tools and frameworks whose designs are dedicated to supporting operational and analytical processing. The NoSQL called as Not only SQL frameworks are designed to fulfill performance demands of big data applications. There are a variety of NoSQL approaches such as hierarchical object representation such as JSON, XML and BSON and the concept of a key-value storage. The wide range of NoSQL tools, developers, and the status of the market are creating uncertainty with the data management.

(ii) Talent Gap in the Big Data

The new tools evolved in this sector can range from traditional relational database tools with some alternative data layouts designed to maximize to access speed while reducing the storage footprints, NoSQL data management frameworks, in-memory analytics, and as well as the broad Hadoop ecosystem. The reality is there are a lack of skills available in the market for big data technologies.

(iii) Getting Data into Big Data Structure

The intention of a big data management involves analyzing and processing a large amount of data. The complexity behind the transmission, access, and delivery of data and information from a wide range of resources and then loading these data into a big data platform. The intricate aspects of data transmission, access and loading are only part of the challenge. The requirement to navigate transformation and extraction is not limited to conventional relational data sets.

(iv) Syncing across Data Sources

Import data into big data platforms and migrated from a wide range of sources on different rates can rapidly get out of the synchronization with the originating system. It means that, the commonality of data definitions, concepts, metadata, and the like. The traditional data management and data warehouses, the sequence of data transformation, extraction and migrations all arise the situation in which there are risks for data to become unsynchronized.

(v) Insufficient Awareness of the Possibilities of Big Data

Enterprises are managed by people with managerial skills and IT takes a backseat. Managers are content with traditional methods of analytics and business intelligence based on data warehouses. They simply are not sufficiently informed as to how much big data can benefit them by widening the scope and bringing in predictive capabilities.

The solution is, the top management needs to lead the way, and seriously focus on how to leverage big data. It pays for the top management to learn and also initiate programs for everyone down in the hierarchy to get a good grasp on the fundamentals of big data and how useful it is for their organization. A good starting point would be to engage a big data development company to act as their consultant and show them how they benefit now and in the future.

(vi) Volume and Complexity

The enterprises are aware of all the positive benefits that big data analytics can bring, the sheer volume and complexity seem a wall too high to climb. Data flows from a variety of sources, and in varieties of types. Apart from the usual company generated data, there is a data from external sources such as IoT devices, searches and social media to name only a few. The diversity in sources also creates a diversity of data types. Data are not just text these days. It also covers audio, video and graphics as well as images. What makes it even more complex is that data is unstructured. Data is there for the taking but making sense of it means streamlining data into a structure that analytics software can understand. The solution is to use Hadoop MapReduce or Cassandra or HBase. All these tools are complex and need the services of trained professionals. The shortcut solution is to retain specialists offering custom big data application development solutions.

(vii) Stupendous Infrastructure and a High Cost

Whichever way one looks at it, big data adoption is expensive. One must have superior IT hardware infrastructure and that alone is expensive. Then, it needs power, space and people to manage big data. You could be looking at a team of data scientists and analysts who do not come cheap. You need custom big data solution, and this also represents significant costs. Businesses can opt for an on-premise solution at tremendous costs or opt for cloud-based solutions at a slightly lower cost. Costs may be driven down by the use of data lakes for data that do not need immediate analysis. The solution to bypass all this infrastructure costs is to simply assign the entire task of custom big data solution and analytics to big data Development Company. They do the work; you enjoy the benefit of predictive analysis.

(viii) Trained Manpower

Even assuming that an organization is willing to spend a lot of money on the hardware and software side to collect and analyze big data, you still need experts with qualifications to use the software. It is a simple fact of life there are thousands of IT professionals with some knowledge and qualifications in handling big data, but not expertise at the data scientist level. Scientists are hard to source and harder to retain. The best big data analytics software simply gives you numbers. It takes a data scientist to make sense of the analytics and give results that are useful. The simple solution is to hire big data development company and you do not have to worry about any aspect of big data analytics and deriving actionable insights.

(ix) Big Data Is About Real-Time Insights and Predictive Capabilities

Big data benefits can be felt only when you can derive real time insights and predictive capabilities. This means high veracity, high volume, and a high precision handling of data inflow using tools such as visualization, computation, libraries, ETL engines and frameworks that give fast results. Real-time analytics give insights into developing trends and helps businesses be ready for demand or to change product specs to suit evolving needs. If you cannot get these capabilities then the big data solution is not giving 100% to your organization. The solution is to let a big data development company with its capabilities take care of real-time analytics and reports that give you insights into developing trends.

(x) Data Security

Data brings in its wake the issue of governance and security. Big data, by its very nature, means that data flows from different sources. The more nodes there are, the more the system is vulnerable to exploits that could lead to losses. Managing such sources and ensuring integrity as well as security call for expert governance measures. The problems arise should you opt for big data operations in the house. Leave it to an outsourced big data service provider like Smart Sight Innovations and remain free from such worries.

(xi) Up Scaling

An enterprise that has put in place a system to handle big data will find that up scaling needs to be built in. This does not just refer to storage capabilities but also to processing facility and performance to be able to handle increasing loads. The solution could be to go for upgrades to the system architecture but this is a never-ending process. The right answer is to simply let custom big data application development solutions provide you with the fruits of their labor while they take care of up scaling and upgrading systems.

Big data is different. The challenges can be overwhelmed from the financial perspective and operational contexts as well as human resources. It is cheaper and smarter to let someone else handle big data, and simply use their custom big data solution to drive insights you can apply in real time, and thus, go one up on the competition.

7. CONCLUSION

The new evolution in data management is Big data, which evolves to satisfy the needs of many organisations, researchers and the Government. The Various researchers' outcomes, and their predictions were helps in urban planning, environment modeling, visualization, analysis, quality classification, securing environment, computational analysis, biological understanding, designing, and manufacturing process required by organisations and cost-effective models. The traditional approach to Data Mining is discussed. The different characteristics and Challenges of Big data are anlysed. These technical challenges must be addressed for efficient and fast processing of Big Data. Hadoop offers a proven solution to the modern challenges facing legacy systems.

REFERENCES

[1] https://dataanalytics.report/whitePapers/data-science-and-big-data-enterprise-paths-to-success.

[2] Shan Ren, Yingfeng Zhang, Yang Liu, Tomohiko Sakao, Donald Huisingh, Cecilia M. V. B. Almeida, "A comprehensive review of big data analytics throughout product lifecycle to support sustainable smart manufacturing: A framework, challenges and future research directions", Journal of Cleaner Production, doi: 10.1016/j.jclepro. (2018).

[3] G. Devateja, P. V. B. Kashyap, C. Suraj, C. Harshavardhan and Impana Appaji , "A Study: Hadoop Framework", International Journal of Advance Engineering and Research Development, Vol. 3, Elsevier, Issue 2, , e-ISSN: 2348 - 4470, print-ISSN: 2348-6406, Feb - 2016.

[4] M. Ahmed and Shahat Osman, "A novel big data analytics framework for smart cities", Future Generation Computer Systems, Vol. 91, pp. 620-633, Feb 2019.

[5] S.Arockia Panimalar, S.Varnekha Shree and A.Veneshia Kathrine, "The 17 V's Of Big Data", International Research Journal of Engineering and Technology (IRJET), e-ISSN: 2395-0056, Vol. 04 Issue: 09, Sep-2017.

[6] Ritu Agarwal and Vasant Dhar, "Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research", Information Systems Research, Vol. 25, No. 3, pp. 443–448, ISSN 1047-7047 (print), ISSN 1526-5536 (online), Sep 2014.

[7] Yichuan Wang, LeeAnn Kung, William Yu Chung Wang and Casey G. Cegielski, "An Integrated Big Data Analytics-Enabled Transformation Model: Application to Health Care", Information & Management, DOI: 10.1016/j.im.2017.04.001, Jan 2018.

[8] Sahar Sohangir, Dingding Wang, Anna Pomeranets and Taghi M. Khoshgoftaar, "Big Data: Deep Learning for financial sentiment analysis", Journal of Big Data, Springer, doi.org/10.1186/s40537-017-0111-6.

[9] Dr. B. Umadevi and A. Kanimozhi," A Modern Approach to Predict the Stock Market Behaviour for Naïve Investors", International Journal of Computer Science and Mobile Applications, Vol. 6 Issue. 7, pg. 34-42, ISSN: 2321-8363, July- 2018.

[10] Dr.B. Umadevi and C. Sujitha, "An Analytical Approach For Customer Retention Through Business Intelligence", International Journal of Computer Science and Mobile Applications, Vol. 6 Issue. 7, pg. 23-33, ISSN: 2321-8363, July- 2018.

[11] B.Umadevi, D.Sundar and Dr.P.Alli, "Novel Framework For The Portfolio Determination Using PSO Adopted Clustering Technique", Journal of Theoretical and Applied Information Technology, , Vol. 64 No.1, ISSN: 1992-8645, E-ISSN: 1817-3195, 10th June 2014.

[12] www.sas.com/en_in/insights/analytics/big-data-analytics.html.

[13] www.ibm.com/ Analytics/Hadoop.

[14] www.towardsdatascience.com.

[15] www.quora.com/What-are-the-new-big-data-technologies.

[16] S.M. Kawale and Dr. A.N.Holambe, "A Survey on Big Data and Hadoop", International Journal of Engineering Development and Research, Vol. 6, Issue 1, ISSN: 2321-9939, 2018.

[17] S. Bhosale Harshawardhan, Prof. P.Devendra Gadekar, "A Review Paper on Big Data and Hadoop", International Journal of Scientific and Research Publications, Vol. 4, Issue 10, ISSN 2250-3153, Oct 2014.

[18] www.flydata.com/the-6-challenges-of-big-data-integration.

[19] Dr. B. Umadevi and P. Saranya Devi, "An Overview Of Organizational Growth By Human Resource Based On Knowledge Management", International Journal of Creative Research Thoughts (IJCRT), Vol. 6, Issue 2, ISSN: 2320-2882, April 2018.

Authors' Biography

	**Dr. B. UMADEVI** has received her Doctoral degree in Computer Science from Manonmaniam Sundaranar University, Tirunelveli, India. Currently, Working as Assistant Professor in P.G. Department of Computer Science, Government Arts College, Melur, Tamilnadu, India. She has over 25 years of Teaching Experience and published her research papers in various International, National Journals and Conferences. She has been recognized as a reviewer in reputed national and international journals. During her academic journey she guided and produced more scholars in Master of Philosophy. Her research interests include Data Mining, Soft Computing and Evolutionary Computing. She also extends her research in the applications of Data Science, IoT and Big Data Analytics.
	**Dr. D. SUNDAR** has received his Doctoral degree in Computer Science from Madurai Kamaraj University. He has over 25 years of teaching and 2 years industrial experience. He had worked as Associate Professor in the Computer Application Department of PES University Bangalore. Currently, working as Assistant Professor in P.G. Department of Computer Science, Government Arts College, Melur, Madurai District, and Tamilnadu, India. He also holds Master's Degree in Business Administration. He published many research papers in various International, National Journals and Conferences. His research areas include Software Engineering and Data Mining. He also very much interested to make research in Data Science and Big data Analytics. He possessed industrial experience in IT sector across Singapore and Malaysia.

B. SAKTHI has received her Master Degree and Master of Philosophy at Madurai Kamaraj University, Madurai, India. She has over 14 years of Teaching experience in the field of Computer Science. Currently, She is working as Assistant Professor in MCA Department, V.V.Vanniapeurmal College for Women, Virudhunagar, Tamilnadu, India. Her research interest includes in Big data analytics and Applications of Data Science.

.