

## PalArch's Journal of Archaeology of Egypt / Egyptology

### ADOPTION OF BIG DATA ANALYTICS FRAMEWORK FOR BUSINESS INTELLIGENCE AND ITS EFFECTIVENESS: AN ANALYSIS

*Farhad Khoshbakht<sup>1</sup>, Atena Shiranzai<sup>2</sup>, S. M. K. Quadri<sup>3</sup>*

<sup>1,3</sup>Department of Computer Science Jamia Millia Islamia (A Central University), New Delhi,  
India 110025

<sup>2</sup>Department of Computer Engineering, Faculty of Industry and Mining(Khash), University of  
Sistan and Baluchestan,Zahedan,Iran

Email: <sup>1</sup>f.khoshbakht630@gmail.com, <sup>2</sup>ashiranzai@eng.usb.ac.ir, <sup>3</sup>quadrisk@jmi.ac.in

**Farhad Khoshbakht, Atena Shiranzai, S. M. K. Quadri: Adoption of Big Data Analytics Framework for Business Intelligence and its Effectiveness: An Analysis -- Palarch's Journal Of Archaeology Of Egypt/Egyptology 17(9). ISSN 1567-214x**

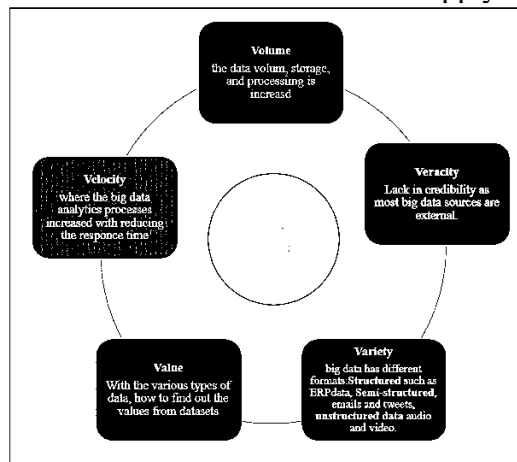
**Keywords: Adoption, big data analytics framework, business intelligence, effectiveness**

#### ABSTRACT

The paper showed the developments in big data analytics technology, applications and the effect. A certain sections were examined several significant recent developments in big data analytics, among others. Sustained growth of big data analytics was expanded with adoption of big data, software development of the Hadoop environment, with data processing models, and algorithmic accountability standards. The ability for big data analytics to be used in market forecasting is to determine future criteria based on previous research used to identify predictive techniques. Since then, big data analytics have been used in several fields some of which have covered multiple industries. The usage of big data analytics and its computer architecture helps many enterprises. The goal of this paper is to conduct a thorough analysis into strategies of big data and business analytics for better making business decisions, technical techniques, implementations and open research challenges. In addition, the research is looking to draw awareness to the tremendous advantages that big data has provided to companies in developing countries and how emerging business organizations can replicate these. In addition, the report addresses various problems facing big data analytics with an emphasis on data protection, implementation, accessibility, legislation, and enforcement. We have concluded that from our proposed framework for the sample 1(89.53%) and 3(89.49%) was more recall analysis rather than sample 2(89.44) from various samplesets. Hence we can say that the data analysis System has a major effect on market intelligence and corporate success in analytics.

## 1. Introduction

Implementation of big data analytics has a major influence on the retail sector, boosts consumer satisfaction and prevents fraud. In modern society, the shopping sector is of great significance since virtually everyone purchases their everyday needs nowadays. Predicting demand for goods helps distributors to improve customer service, and distributors may collect market analytics information using consumer billing info. To archive, maintain and review these data, a Hadoop distributed file system (HDFS) method is used to extract more data (Singh et al., 2015; Wamba et al., 2017). Big data analytics helps certain businesses to be more visible in business decisions that aid in consumer segmentation based on their functionality. To alert corporations about what their clients want, we may even use social network monitoring. Using sentiment analysis on these data offers early alerts to the company when the customer switches to other brands, encouraging the organization to take action (Elgendy, and Elragal, 2014). For several years, companies have utilized purchase behaviour of the consumers, and it is now made possible by innovative Big Data strategies including such real-time segments and sub-segmentation to provide better-targeted ads. By understanding customer tastes and predicting trends in market sentiment, companies may also accomplish stronger social network targets (Manyika, et al., 2011; Russom, 2011; Elgendy, and Elragal, 2014). Retailers therefore use data analytics to address new problems and identify opportunities based on increasing consumer desires, competition and volatility. The usage of data analytics tools can provide increased accuracy, transparency and interpretation for many firms, and such information can be extended to supply chains in the industry.

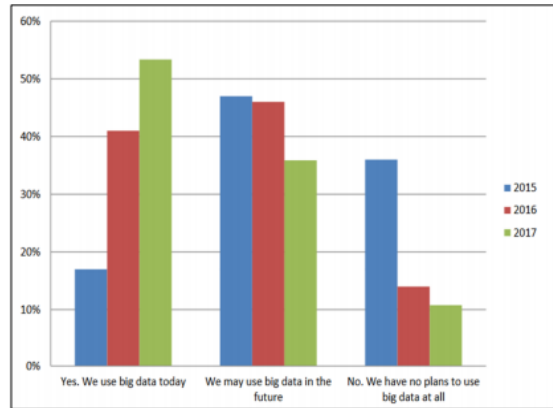


*Figure 1: Big data for the 5 Vs.*

Using multi-dimensional data to provide the historical sense of big data is essential. The diversity of big data is as critical as its length, although pace or velocity can illustrate how challenging it can be to handle big data. Velocity can be applied to the frequency of data production or the frequency of data propagation. Big data material may also be classified as ambiguous, good or

poor, based on data incoherence, inadequacy, uncertainty, delay, deceit, and estimations.

A new study (Nashua, 2017) has illustrated the increase in the usage of big data analytics among corporations. Comparative analysis was shown usage of big data by businesses during 2015-2017 in figure. In 2017 around 50 percent of businesses were utilizing big data (Figure 2) (Watson, 2019) (Figure 2).



**Figure 2:** Comparative study of big data used by business during 2015-2017 and its effect

Big data is also the potential central weapon in several businesses' business strategies, as these 'business strategies are embodiments in the approach used'. Business models support the company's potential to grow and improve when necessary. Businesses need to reconsider their use of big data with respect to their business models, utilizing technology to navigate multiple data sources and strategies to boost competition and performance. (Günther, et al., 2017). Grover, et al. (2018) argued that crucial investment between both hardware and computational capabilities is needed to promote technology research and global strategy to process sustainable competitive advantage in big data. Companies must therefore provide access to modern tools and attract data-savvy workers who enjoy the current technology. Watson (2014) wrote a piece on massive data processing written by the information management association (CAIS). At the period, the study shows the developments in technology, applications as well as the influence of big data analytics. In 2019, Watson outlined some notable recent developments, including predictive analytics. Subsequent acceptance of big data analytics, expanded deployment of big data, apache Hadoop application growth, data lakes, sophisticated mathematical analysis, and concepts of computational accountability. Big data analytics may be applied to global business forecasts, assessing potential needs based on previous analysis used during and before COVID-19 to define analytical techniques and their effects. (Dublin, 2020; Sebastian, 2020; Cabrera-Sánchez, et al., 2020; Paramita (Guha) Ghosh, 2020; Maritz, et al., 2020). In other regions, big data analytics have already been applied and have covered numerous sectors.

### **Motivations**

Motivation about this study to explore Big Data Analytics Framework for Business Intelligence and its Effectiveness after adoption reliable techniques on performance as well as its significant impact on quality of decision-making. We also review the earlier researcher who adopted various BI tools for the various purposes for several years. Hence we motivated in this way apply new way current tools frameworks which enhance our results.

### **Contributions**

The main contribution of our research is based on adoption of Big Data Analytics Framework for Business Intelligence and its Effectiveness on organization performance. In this way we have contributed our analysis to show briefly interpretation after adopting various new tools and techniques throughout research to improve our result as per earlier workers. And also we have to find out various challenges and issue of traditional business intelligence the shift to modern enterprise analytics. How are to explore this research as given below.

The first section is discussed about the preliminary discussion about Big Data Analytics and Business Intelligence and its some comparative study. Second section is discussed about objective, third section discussed about hypotheses which are going to prove. Fourth section is research method adopting though whole analysis. Fifth section discussed about analysis and results interpretations. Sixth section is discussed about the challenges and issues of traditional business intelligence the shift to modern enterprise analytics. Seven section is discussed about the framework of MapReduce and Hadoop file system and its application and recent advancements and seventh section is conclusion on whole research as per given objective.

### **2. Objectives**

- To study after adoption of Big Data Analytics Framework for Business Intelligence and its Effectiveness

### **3. Hypothesis**

- There is a significant influence of Big Data Analytics Framework for Business Intelligence and its performance
- There is significant influence of Big Data Framework on quality of decision-making and its performance.

### **4. Research Method**

In this section we would explain methodology adopted under Big Data Framework for Analytics Business experimental techniques adopted in this research as per given context. This includes the research objective, hypothesis, research design, universe and sample of the study, data Collection of the study, statistical techniques used for data analysis.

### **Research Design**

The Apache Hadoop is an open source software library that incorporates a framework that takes into consideration the dispersed preparing of huge data

sets in gatherings of PCs utilizing basic programming models. We have a variety of alternatives extending from a solitary PC to a great many PCs, every one of which offers nearby handling and capacity. Instead of depending on equipment, the library itself is intended to identify and oversee disappointments and guarantee high accessibility at the application level.

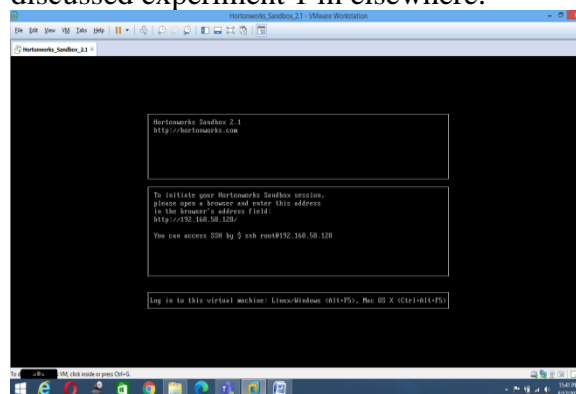
### Research Strategies

First the installation of tools Hadoop, Mapreduce, hive, python, related libraries and download datasets. Create table in hive and load data into it. The MapReduce code in python for mapping the unstructured hive database to structured form and applied the reducer for splitting the dataset in smaller chunks and process in parallel then apply preprocessing and extracting features. Now the Analysis Parameters obtained with Confusion Matrix, Accuracy, Precision, Recall, and MSE etc.

## 5. Analysis And Results

Hadoop is a popular framework in big data analytics. Implementation for the MapReduce paradigm and the combination of storage and review, Hadoop' offers reliability, scalability and manageability '(Elgendy, and Elragal, 2014). Hadoop offers HDFS through big data analysis and MapReduce through this analytics makes split into multiple block but instead defining data sets to be distributed through cluster nodes.

In this study we first Installation of tools Hadoop, Mapreduce, hive, python, and related libraries and download datasets. Create table in hive and load data into it. The MapReduce code in python for mapping the unstructured hive database to structured form and applied the reducer for splitting the dataset in smaller chunks and process in parallel then apply preprocessing and extracting features. Now, the Analysis Parameters were obtained with Confusion Matrix, Accuracy, Precision, Recall, MSE etc. in given experiments 2. Earlier we have discussed experiment 1 in elsewhere.



*Figure 3: Installation of Linux Platform*

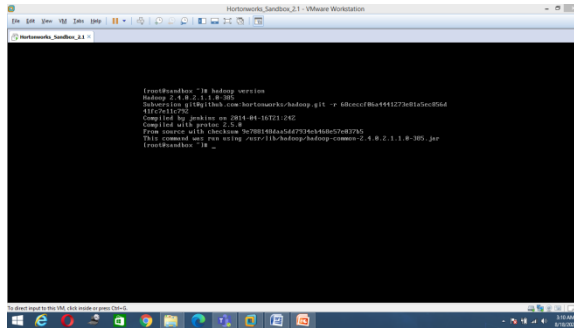


Figure 4: Check for Hadoop Version

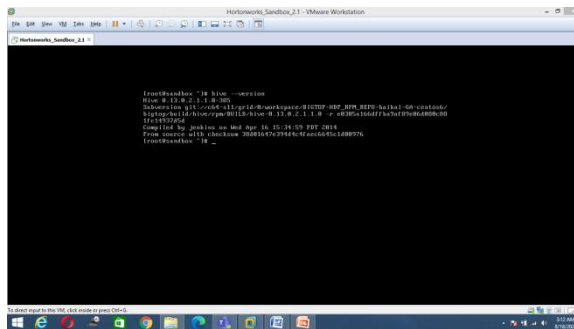


Figure 5: Check for Hive Version

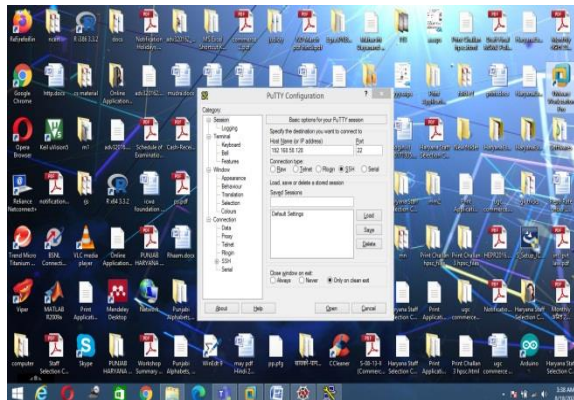


Figure 6: Putty Environment for Remote Access

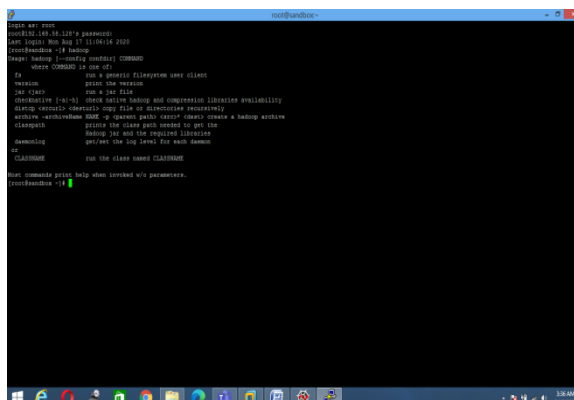
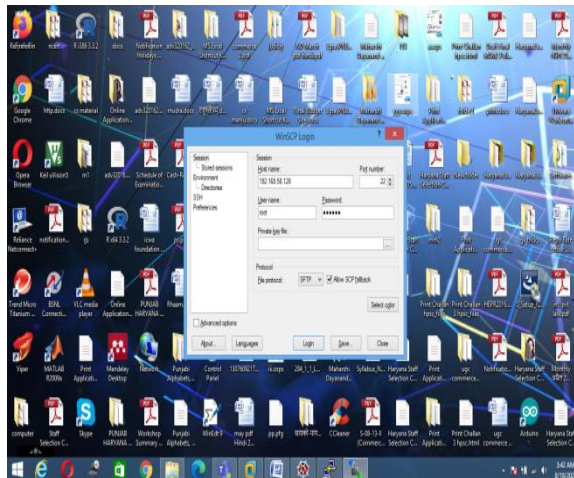
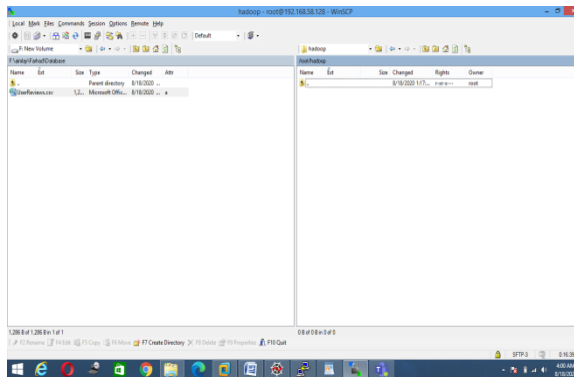


Figure 7: Putty based Linux Machine Access



**Figure 8:** WinSCP environment for Transferring Review CSV file to Hadoop Environment



**Figure 9:** Environment for Transferring CSV file in Hadoop environment

**Table 1 Properties of experiments**

Properties	Values
Size of Training Data	200000
Size of Testset	100000, 500000, 1000000
Text Features Collected	Unigrams, Bigrams, Negative Features
Preprocessing	Stop list removal, stemming
Other Features	Frequency, Weights
Number of Class	2

**Table 2** Sample 1 Confusion Matrix

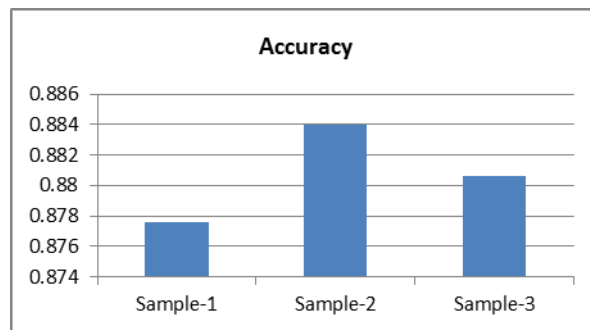
		Predicted	
		Positive	Negative
Actual	Positive	65108	10524
	Negative	1714	22654

**Table 3** Sample 2 Confusion Matrix

		Predicted	
		Positive	Negative
Actual	Positive	279093	45598
	Negative	12408	162901

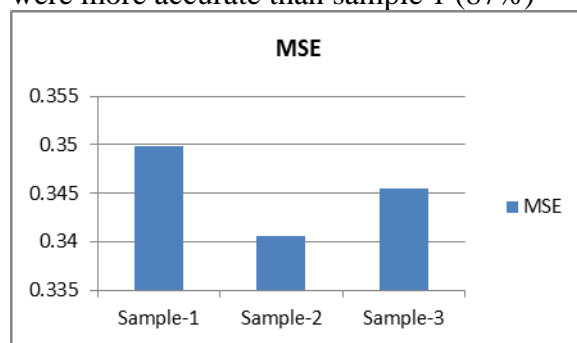
**Table 4** Sample 3 Confusion Matrix

		Predicted	
		Positive	Negative
Actual	Positive	605474	98689
	Negative	20712	275125



**Figure 10** Accuracy Analyses for Different Samplesets

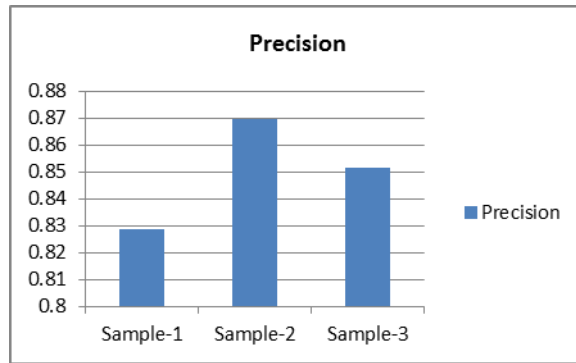
From figure 10 we have observed that the sample 2 (88%) and sample 2 (88%) were more accurate than sample 1 (87%)



**Figure 11:** MSE Analysis for Different Sample Sets

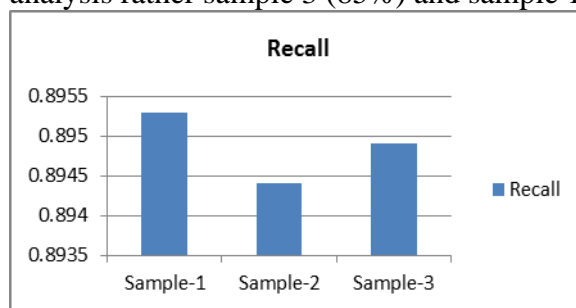
From figure 11 we have observed that the sample 1 (0.35) was more MSE analysis rather than the sample 3 (0.34) and sample 2 (0.34) from our data set.





**Figure 12:** Precision Analysis for Different Sample Sets

From figure 12 we have observed that the sample 2 (87%) was more precision analysis rather sample 3 (85%) and sample 1(83%) from the different data set.



**Figure 13:** Recall Analysis for Different Samplesets

From figure 13 we have observed that the sample 1(89.53%) and 3(89.49%) was more recall analysis rather than sample 2(89.44) from various samplesets. From all results we can say that the Big Data System has a major impact on business intelligence and corporate success in analytics.

### **Importance of Big Data Processing**

Big data research is important since, in order to obtain fresh, meaningful knowledge, it helps data scientists and statisticians to dive deeper into vast quantities of data. The retail-to-government industries still need to identify avenues to increase consumer loyalty and streamline processes.

The relevance of big data analytics has expanded, along with the range of unstructured data that can be exploited for information: social networking posts, emails, clickstream data and the multitude of Internet of Things sensors. Big data analytics is required because the influx of unstructured data that shapes the environment today cannot be addressed by conventional data warehouses and relational databases. Organized content is the perfect match for them. They can't cope with demands for real-time data either. The increasing need for real-time knowledge of big information is met by Big Data Analytics. This is especially relevant for enterprises that depend on fast-moving capital markets and the number of websites or mobile operations.

When it comes to finding potential growth sources and improving efficiencies that offer a strategic edge, big data analytics are seen by enterprises as crucial in supporting the bottom line.

As bigger businesses trust big data analytics, they profit from the opportunities of big data analytics:

- Reduction of costs: Through finding increasingly profitable forms of doing it.
- Decision-making: swift and better decisions with the opportunity to quickly interpret data and act on learning quickly.
- New goods: Leveraging knowledge to truly consider clients allows firms the opportunity to develop products and services that clients want and need.

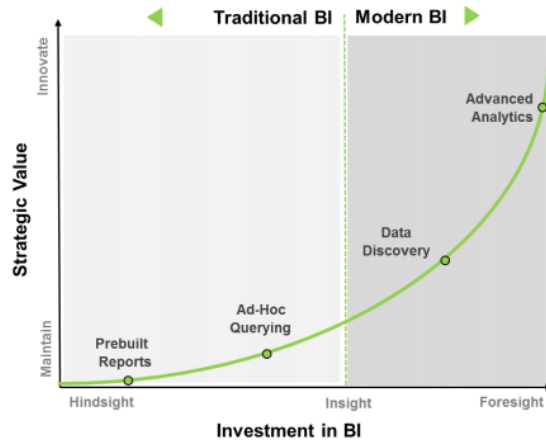
## **6. Challenges Of Traditional Business Intelligence The Shift To Modern Enterprise Analytics**

In an environment where the volume of data produced is rising exponentially, Deloitte (2018) explained that federal agencies and IT departments are experiencing growing rising demand for business data to tap into value. Many businesses are looking for fresh and imaginative ways to convert organizational data into practical information, with the ability to boost consumer performance and overall mission quality. The technology, tools, and strategies needed to derive information from these vast volumes of data may seem challenging to make sense of. A modern market analytics strategy, though, often does not entail a comprehensive redesign of previous investments. By engaging in the new business intelligence (BI) framework, which complements traditional business intelligence platforms, companies will extend their spectrum of insight-driven capabilities. Federal agencies may leverage analytics to meet project priorities more effectively by adopting a new BI system, working to protect and safeguarding Americans' health, maintaining the community safe and stable from domestic and global attacks, and preventing replication, fraud, and misuse of government programs. This was discusses conventional market intelligence and monitoring problems, the need for methods to respond to today's hardest data challenges, and the related individuals, systems and technology to make the move to new business analytics feasible.

Deloitte, (2018) outlined the following issues for federal agencies in the analytics sector surrounding conventional BI strategies:

- *Absence of the on-demand analytics tools-* Modern BI 's specialized consumers do not want to wait to get responses to their most difficult business concerns. Self-service functions are expected by additional customers, depending around their own knowledge, to link and interpret specific data sets for any purpose at any time.
- *Need for Predictive Analysis* – Only one piece of the puzzle is provided through contextual research tools: insight into the history.
- *Analysis of mixed data styles-* Conventional BI structures focus largely on standardized knowledge, but consumers now need the ability to access and analyze semi-structured, massive amounts of data and third-party data.

Many companies that have the requisite resources, processes, and technologies to extend their potential for data analytics towards the next level are prevented. These issues have a philosophical structure and strategy that, as seen in Figure 14, goes outside the limits of conventional BI frameworks.



**Figure 14:** Traditional BI vs. Modern BI investments showing strategic value changes.

This paper clarifies emerging business intelligence applications and helps explain how in a traditional BI reporting context the new paradigm can coexist in order to broaden company expertise in the data exploration and advanced analytics field.

## 7. Framework Of Mapreduce And Hadoop File System And Its Application

MapReduce has been based on a diverse body of established works using Machine Learning approaches. The following is an array of different computational algorithms used by Hadoop Distributed File System for processing large data sets. For a two-step method a predictive query model uses the least square regression. Initial phase computes the statistics related to processing data. The next step focuses on computation using data mining approaches to summaries [Andrew, et al., 2010].

MapReduce performs an unconscious operation on a large array of commodity hardware that parallelizes and executes the program. Being a data processing tool, vast amounts of data are processed using distributed system on multiple computers. MapReduce implements Hadoop an open source distributed file system for efficient storage on parallel databases. While the MapReduce programming model has notable benefits, the following section describes few limitations affecting its cluster performance.

MapReduce is primarily used for sorting, data mining, and machine learning within Google. This model of programming deals with computational fault tolerance in distributed parallel background. Network bandwidth is saved by local optimization provided by the operation of read / write data to a local disc. Redundant deployment reduces the effects of slow device errors and data loss [Burhan, 2014].

Implementing MapReduce depends on an in-house cluster management program to distribute and execute user tasks on shared machines. This cluster management is close to other such systems as Condor [Douglas, et al. 2004].

Fundamental parallels in the MapReduce programming model [Dean et al. 2004] are redundant execution that recovers data loss triggered by errors and locality-conscious scheduling to mitigate the volume of data transmitted over congested network connections.

Techniques such as active discs, under which the computation is channeled into computing elements adjacent to local discs, encourage locality improvements. The number of data transmitted via the I / O or network subsystem condenses in this method. This device supports commodity processor in which the limited numbers of disks are attached directly, instead of operating directly on the CPU of the disk controller [Soila, 2009].

MapReduce is a row-oriented fashion for scanning input data that apparently slows down the execution of analytical tasks in interactive searches compared to other advanced knowledge databases [Huston et al. 2004] within the data blocks are arranged in Trojan layout data according to the incoming workloads deployed on top of HDFS. This architecture replaces the MapReduce system due to unique features such as co-locating attributes based on query workloads, multiple replicas of data blocks and broken mirror imitators [Alek, et al. 2011].

Mastiff is an automated Big Data Analytics focused on time framework. It utilizes a systemic hybrid of a column group store and a lightweight support platform for diverse workloads to boast query results. This programme substitutes the MapReduce method for its high data loading scheme and high query performance [Sijie, et al. 2012].

Hadoop software stack consists of a MapReduce execution system, pluggable distributed storage system, and a set of Big Data analytics procedural to declarative interfaces. Starfish introduced by Herodotos specifically proves negative aspect of Hadoop results. This lifecycle analytics of the self-tuning system automatically running knob for performance tuning to maximize energy, time and money [Herodotos, et al. 2011].

Using Hadoop MapReduce to incorporate work optimization and physical data structures such as data formats and indexing. This research study addressed the state of performance improvement approach [Pavlo, et al. 2009]

ETL iterative methods are applied for the transition of analytical analysis data from warehouses to marts. The new challenges such as data mining, data quality validation, cleaning, profiling, statistical algorithms, hierarchical and drill down features are replaced with Intel [Dean, 2010].

Google developed a modular distributed file system for data-intensive applications of large scale. The feature of fault tolerance at the execution of commodity hardware provides a large number of customers with high aggregate efficiency. MapReduce enables high aggregate throughput to parallel readers and authors [Sanjay, et al. 2003].

Apache Hadoop provides a flexible environment with all the automation, scaling and built-in error torrents for data processing devices. Apache Hadoop community plays a significant role in day-to-day business and study company activities. In this article the author underlined the importance of the Hadoop

method by contrasting the relational databases due to the growing volumes of digital information [Thomas, et al. 2014].

Hadoop, MapReduce and HDFS: An article about Developers Perspective discussed the complexity behind the MapReduce programming model on a commodity server cluster. MapReduce is designed as an independent service layer for platform enabling data processing and analysis [Mohn Rehan and Durgaprasad 2015].

Hive QL is a declarative SQL language built on MapReduce using Hadoop, allowing MapReduce scripts to be inserted into queries to solve the problems of large data sets [Mohn Rehan and Durgaprasad 2015].

Cloud Computing is advanced by the handling of large and novel data sets by rising business value and computing capacity. For flexibility and easy use of data analytics algorithms, cloud vendors upload the data into their clouds. HDFS systems are designed to hold data by providing the cloud systems with high-throughput and access to this information.

## 8. Conclusion

Big data has the potential to transform organization, education and science. Improving the results of the students by optimizing the performance of the student during the course and knowing their actions can be achieved using analytics of big data. Big data analytics often allows students the ability to align their needs with the accessible programmes; they will then pick the right school or educational programme. Big data, on the other side, helps teachers recognize the degree of competence of each pupil and change the teaching approach on an individual basis to the most advantageous result. Therefore, the use of Big Data to direct instructional advice plays an important role in expanding opportunities for education through enabling students to access low-cost information groups via remote teachers. Data monitoring and research software are easily available. We have found that from our proposed framework for the sample 1(89.53%) and 3(89.49%) was more recall analysis rather than sample 2(89.44) from various samplesets. Hence we may assume that the Big Data Paradigm has a substantial effect on market intelligence and corporate success in analytics.

Many companies use data collection to promote reliable and secure market innovation that drives decision-makers. These may also evaluate performance, operating models and consumer service restoration. Trends open the path of a creative transformation in science, innovation, and marketing for business. Several firms, such as Amazon, eBay and Google, evaluate forces that influence success to regulative that raise market revenue and record customer behaviour. Big data has major impacts on investment services as they choose to change their direct debit consumer approach in the field. Big data is used by companies such as Brick and Mortar to validate the opportunity to direct user data by collecting transactional data from millions of buyers by utilizing the details obtained to recognize potential prospects such as one of the most successful promotions being discovered. Several companies have been utilizing data mining to collect insights on social networking. Ford Motor, Southwest

Airlines, and Pepsico evaluate customer mobile messaging posts such as twitter and Facebook to optimize the immediate impact of a pattern and track consumer views on their products. Big data impacts many facets of society and results in social gains.

### **Suggestions for Future Directions**

These kinds of tools will use for a variety of activities, including correlation and cluster analysis, from the viewpoint of big data analytics, to identify trends in unstructured data sets. This would be the study of data to produce insight and insights that would contribute to companies making informed decisions. This will be to plan and grow an organization's data warehouses and data marts. This will be the extraction, cleansing, aggregation, protection and transformation into a single enterprise data warehouse or any organization of heterogeneous data with accuracy, minimum time and cost. This kind of research will be improving more futuristic applications.

### **References**

- Alekh J. and Dittrich J et al. (2011) 'Trojan Data Layouts: Right Shoes for a Running Elephant' in SOCC ACM Conference.
- Andrew Y. N. G., Cheng Tao Chu. and Yi-an Lin. (2010) 'Machine Learning on Multicore' in Conference.
- Burhan U.I.Khan., Rashidah F., and Hunain A. et.al, (2014) 'Critical Insight for MapReduce Optimization in Hadoop' in International J of Computer Science and Control Engineering, Vol.2, Issue-1, pp: 1-7
- Cabrera-Sánchez, Juan-Pedro, & Villarejo-Ramos, Ángel F. (2020). 'Factors Affecting The Adoption Of Big Data Analytics In Companies.' *Revista de Administração de Empresas*, 59(6), 415-429. Epub January 10, 2020.<https://doi.org/10.1590/s0034-759020190607>
- Dean J. and Sanjay G. (2004). 'MapReduce: Simplified data processing on large clusters' in ACM Conference OSDI.
- Deloitte (2018). Modern Business Intelligence The Path to Big Data Analytics. <https://www2.deloitte.com/content/dam/Deloitte/tr/Documents/deloitte-analytics/Modern%20Business%20Intelligence.pdf>
- Douglas T., Todd T., and Miron L. (2004). 'Distributed computing in practice: The condor experience' in Concurrency and Computation Conference.
- Dublin, (Aug. 24, 2020). 'The Global Big Data Analytics Market Fueling Artificial Intelligence, 2020.' Report, (Globe Newswire), ResearchAndMarkets.com.
- Elgendy, N. and Elragal, A., (2014). Big data analytics: a literature review paper. s.l., Springer, cham, pp. 214-227.
- Grover, V., Chiang, R.H., Liang, T.P. and Zhang, D., (2018). Creating Strategic Business Value from Big Data Analytics: A Research Framework. *Journal of Management Information Systems*, pp. 388-423.
- Günther, W.A., Mehrizi, M.H.R., Huysman, M. and Feldberg, F., (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems*, Vol. 26, pp. 191-209.

- Herodotos H., Harold L. and Liang D. et al. (2011). 'Starfish: a self-tuning system for Big Data Analytics' in Biennial Conference on Innovative Data System Research.
- Huston L., Sukthankar R. and Ailamaki A. (2004). 'Diamond: A storage architecture for early discard in interactive search' in USENIX file and storage technology Conference April.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Byers, A.H., (2011). Big data: The next frontier for innovation, competition, and productivity.
- Maritz J., Eybers S., Hattingh M. (2020) 'Implementation Considerations for Big Data Analytics (BDA): A Benefit Dependency Network Approach.' In: Hattingh M., Matthee M., Smuts H., Pappas I., Dwivedi Y., Mäntymäki M. (eds) Responsible Design, Implementation and Use of Information and Communication Technology. I3E 2020. Lecture Notes in Computer Science, vol 12066. Springer, Cham.
- Mohn Rehan G., Durgaprasad G., (2015) 'Hadoop, MapReduce and HDFS: A Developers Perspective' in Procedia Computer Science, vol. 48 pp: 45 – 50.
- Paramita (Guha) Ghosh (February 18, 2020). 'Business Intelligence and Analytics Trends in 2020.' Dataversity <https://www.dataversity.net/business-intelligence-and-analytics-trends-in-2020/>
- Pavlo A., Paulsone and Abadi D. J. (2009). 'A comparison of approaches to large scale data analysis' ACM SIGMOD International Conference, New York.
- Russom, P., (2011). Big data analytics, s.l.: TDWI best practices report, fourth quarter. Sagioglu, S. and Sinanc, D., 2013. Big data: A review. s.l., IEEE, pp. 42-47.
- Sanjay G., Howord G. and Shuntak L. (2003). 'The Google file system' in SOSP Conference, ACM J., New York.
- Sebastian Stan (7th Jan 2020) 'Data Analytics Market in 2020: Trends, Forecasts & Challenges.' Cognetik, <https://www.cognetik.com/blog/data-analytics-market-in-2020-trends-forecasts-challenges/>
- Sijie G., Jin Xiong and Rubao L. (2012). 'Mastiff: A MapReduce based system for Time based Big Data Analytics' in IEEE Conference on Cluster Computing.
- Singh, D.; Reddy, C.K. (2015). A survey of Platforms for Big Data Analytics. J. Big Data vol. 2, pp. 8.
- Singh, J. and Singla, V., (2015). Big data: tools and technologies in big data. International Journal of Computer Applications., Vol. 112.
- Soila K., Jiaqi T., Rajeev G. and Priya N. (2009). 'An Analysis of Traces from a Production MapReduce Cluster' CMU-PDL-09-107
- Thomas H., Martin P.-Z. and Alexander S. (2012). 'Giant Data: MapReduce and Hadoop' University of Freiburg computer science department - Research Article

- Wamba, S.F., Gunasekaran, A., Akter, S., Ren, S.J.F., Dubey, R. and Childe, S.J., (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, pp. 356-365.
- Watson, H., 2019. Update Tutorial: Big Data Analytics: Concepts, Technology, and Applications. *Communications of the Association for Information Systems*, Vol 44, pp. 21.