International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Data Analyzing with Cloud Computing Including Related Tools and Techniques

Djabeur Mohamed Seifeddine Zekrifa[1], Abhishek Sharma[2], Satyam[3], Abhijit Janardan Patankar[4], Kalyan Devappa Bamane[5]

Submitted: 28/04/2023 **Revised**: 25/06/2023 **Accepted**: 06/07/2023

Abstract: The arrival of the digital era has led to an increase in a variety of data kinds, which continues to grow with each passing day. In point of fact, it is anticipated that by the year 2016, the cloud will store fifty percent of all data. The complexity of this data necessitates its storage, processing, and examination in order to provide information that may be put to use by organizations. The needs of big data analytics in terms of storage and computational power make cloud computing an ideal platform for carrying out the aforementioned objectives. Because of this, cloud-based analytics is now a study subject that may be pursued. However, before actual implementations of this synergistic model can be deployed in a widespread manner, a number of concerns need to be resolved, and dangers need to be reduced. This article investigates the current research in this area of study, as well as its obstacles, unanswered questions, and potential future research directions.

Keywords*: cloud-based analytics, Artificial Intelligence, Deep Learning Technology, Machine Learning,*

1. Introduction

Since the dawn of the information age, there has been a meteoric growth in the quantity of data that is being produced, stored, and distributed. Massive volumes of data may be obtained from a variety of places, data stores, online communities, blogs, and streaming media servers. This growth has resulted in the production of copious amounts of pervasive, intricate data. For these massive data sets to provide useful insights, they must be efficiently generated, stored, communicated, and analysed [1]. While this information has great potential, it is also very complicated, fraught with security risks, and mostly irrelevant. The pros and cons of acquiring access to sensitive information including medical records, social media posts, bank transactions, public documents, and DNA sequences are debatable given the nature of the study being conducted. Big data processing and analytics was developed because of the need for better analytics services, apps, programming tools, and frameworks.[2]

2. Objective

The research endeavors pursued several objectives to contribute to the field of cloud-based big data analytics. Firstly, the aim was to conduct a comprehensive study on the application of Cloud Computing in the context of Big Data Analytics. This involved exploring the various ways in which cloud technologies can be leveraged to facilitate efficient and scalable processing and analysis of large and complex datasets. The research delved into the underlying principles, frameworks, and architectures associated with Cloud Computing in the realm of Big Data Analytics, providing a solid foundation for further investigation. Furthermore, the research aimed to identify and understand the difficulties and unanswered questions that exist within the field, with a particular focus on delineating future research goals. By analyzing existing literature, engaging in discussions with experts, and examining practical challenges faced in real-world scenarios, the study sought to shed light on the gaps and areas that warrant further exploration. This involved identifying emerging trends, technological advancements, and unresolved issues that require attention to enhance the effectiveness and efficiency of cloud-based big data analytics.

The investigation into difficulties and questions for future research goals encompassed various aspects. These included addressing scalability concerns to accommodate the exponential growth in data volume, velocity, and variety [3]. It also involved exploring methods to improve the performance and processing capabilities of cloud-based big data analytics frameworks, ensuring timely and accurate analysis of data [4]. Additionally, the study aimed to identify potential challenges related to data privacy, security, and regulatory compliance in the context of cloud-based analytics. This included evaluating

1Higher School of Food Science and Agri-Food Industry, Ahmed Hamidouche Av, Oued Smar, Algiers 16200, Algeria
2Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore, Madhya Pradesh, India.
3Department of Electronics and Communication Engineering, Mohan Babu University, Tirupati, Andhra Pradesh-517102.
4Department of Information Technology, D y Patil College Of Engineering, Akurdi, Pune.
5Department of Information Technology, D y Patil College Of Engineering, Akurdi, Pune.
Corresponding Author: d.zekrifa@essaia.dz

privacy-preserving techniques, encryption mechanisms, and governance frameworks that can safeguard sensitive information and ensure compliance with relevant data protection regulations.

Moreover, the research sought to uncover potential research directions to enhance the integration of Cloud Computing and Big Data Analytics, considering advancements in technologies such as machine learning, artificial intelligence, and distributed computing [5]. This involved investigating ways to leverage these technologies to optimize resource utilization, improve data processing efficiency, and enable real-time analytics on cloud platforms. By studying Cloud Computing for Big Data Analytics and addressing the difficulties and questions for future research goals, the research aimed to contribute to the advancement of knowledge in the field [6]. The outcomes of the study provided insights, recommendations, and potential directions for researchers, practitioners, and organizations interested in harnessing the potential of cloud-based big data analytics to derive valuable insights and make data-driven decisions in various domains.

3. Methodology

Many firms and groups have found success via the analysis of massive data sets. Medical and scientific research, transportation and logistics, international security, environmental and social problem forecasting and management are just a few of the many possible areas where such technologies may be put to use. One of the most significant uses of big data nowadays is in the realm of scientific research. The many branches of medicine stand to benefit greatly from the use of both big data analytics and cloud computing. The field of science is one of the most significant users of big data. This is in addition to its already widespread use in monetary, industrial, and governmental spheres. Future-relevant fields include, but are not limited to, systems biology, the prediction of protein structure in addition to function, individualised medicine, and metagenomics. Improving the efficiency and consumer satisfaction of business models is another major use of big data analytics. This is an excellent example of how big data analytics may be put to good use. This is a perfect example of a crucial use for big data analytics.

4. Cloud Computing for Big Data Analytic

In recent years, the problem of insufficient interaction has come to be seen as a serious issue, and several initiatives have been undertaken to address this concern. Optimizing the HBase and HDFS implementation is necessary to improve its responsiveness. Purpose is to promote free and broad usage of online analytical tools by investigating the feasibility of OLAP web services in cloud-based infrastructures [7]. Recent academic efforts have aimed to create a cloud-based big data management architecture. For managing large data sets in the cloud, we provide a data paradigm and associated schema. In addition, work is done to simplify data access for the customer. Exploration of performance and operational velocity has also been critical. Investigate and implement the suggested combined Hadoop and MPI/OpenMP system to maximize speed and performance [8].

When it comes to doing an analysis of the effectiveness of the system, the amount of power that is used has become an extremely important criterion. This is because data has to be transported between data centers that are often situated at great distances from one another. During the processing and storage of large amounts of data, a network-based routing method known as Greed may be utilized to determine the route that will consume the least amount of energy and will go to the cloud data center. Several different analytics systems exist, and many of them make use of simulation. One such application models and predicts the spread of the dengue outbreak in Singapore by using a Direct Acrylic Graph (DAG The rise of online risk analytics necessitated an infrastructure that could supply users with the programming resources and infrastructure for performing the same tasks, prompting the emergence of the concept of Continuous Analytics as A Service (CAaaS), which stands for Continuous Analytics and is used to predict the behaviours of a service or a user. Predicting how a service or user will act is the primary function of CAAAS, an acronym for Continuous Analytics as a Service. Continuous Analytics as a Service (CAAAS) is an abbreviation for this concept. Predictions of future user or service behaviours may be made with the use of this technology. A user or service's behaviours may be analysed and predicted with the help of CAAAS. There are several methods for doing this. CAAAS is an abbreviation for Continuous Analytics as a Service, which describes this method. The academic community is just now beginning to show interest in real-time big data analysis, the last subfield of Big Data Analysis. The figure 1 shows the pictorial representation of big data analysis. Real-time analytic services are offered by many of the same organizations that provide cloud computing for a price. AWS Kinesis is one of the AWS-based technologies that can manage streaming data in real time. Storm, Apache S4, and IBM Infosphere Streams are just a few of the many frameworks and software systems designed for this same purpose.

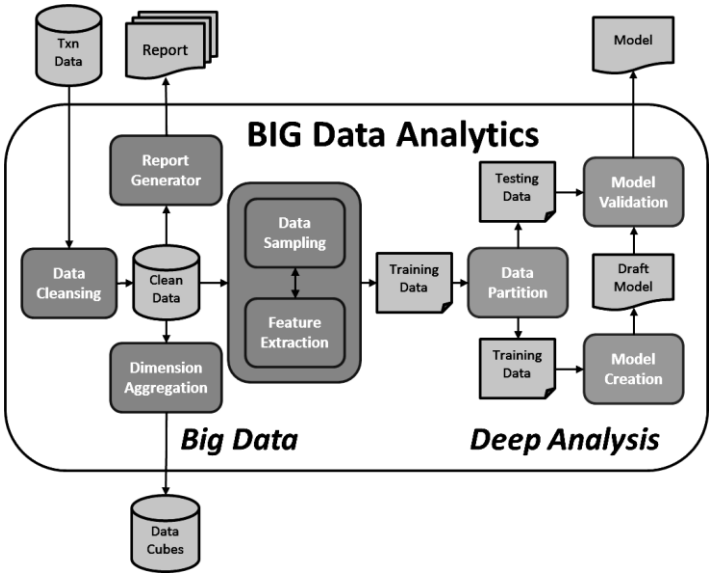

Fig 1. Big Data Analytics

5. The Difficulties and Questions for Future Research Goals

- ***CHALLENGES AND QUESTIONS TO CONSIDER***

For machine learning and data analytics to go beyond their current methods and approaches, some obstacles must first be surmounted. Only then will it be possible to move on to more advanced topics. considers the following conditions to be absolutely necessary.

1. Establishing a strong scientific basis is the first step that has to be taken before selecting an appropriate approach or design.
2. It is necessary to build new algorithms that are both efficient and scalable.
3. Appropriate development skills and technology platforms need to be identified and developed in order to ensure the correct execution of the solutions that have been designed.
4. In conclusion, it is necessary to investigate the economic value of the solutions just as thoroughly as the data format and the usability of the data.

In the context of big data analytics performed on the cloud, there are additional issues that need to be addressed, such as the acceptance and deployment of efficient big data solutions based on cloud architecture, as well as the mitigation of security and privacy threats. Data security emerges as a paramount concern when combining big data analytics with cloud computing in a unified architecture. Consequently, significant attention has been devoted to exploring this aspect of cloud-based big data analytics and devising practical implementation strategies.

A critical area of focus involves the analysis and comparison of various approaches aimed at ensuring the security of cloud and Internet of Things (IoT) data through the use of authenticators. Authenticators play a vital role in verifying the identity and integrity of users and devices accessing cloud-based systems. Researchers and practitioners have been actively investigating different techniques and methodologies to strengthen data security in the cloud environment. The goal is to develop robust frameworks that prioritize efficiency, security, and scalability/elasticity.

One approach gaining attention involves utilizing SSL (Secure Sockets Layer) and public key cryptography to ensure the safety of massive volumes of data stored in off-site cloud data centers. By implementing SSL protocols, secure and encrypted communication channels are established, safeguarding data during transit between users and cloud servers. Public key cryptography strengthens security by enabling the secure exchange of encryption keys and ensuring confidentiality and integrity of the data.

Moreover, researchers have developed frameworks that aim to streamline the procedures of submitting jobs and authenticating users in cloud-based big data analytics. These frameworks integrate multiple security methods, such as multi-factor authentication, role-based access control, and data encryption, to create a comprehensive security infrastructure. By simplifying user authentication processes and providing a range of security measures, these frameworks enhance the overall security posture of cloud-based big data analytics systems. The figure 2 shows the analysis of big data process.

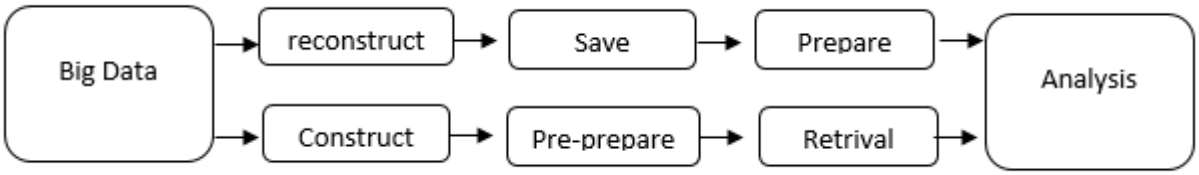

Fig 2. Analyzing Big Data [9]

As the field progresses, the future of cloud-based big data analytics must prioritize efficiency, security, and scalability/elasticity. This entails continuous advancements in security technologies and practices to adapt to evolving threats and vulnerabilities. Additionally, exploring techniques like SSL and public key cryptography allows for the safe handling and storage of vast amounts of data in remote cloud data centers. Moreover, addressing security concerns is crucial when integrating big data analytics with cloud architecture shown in figure 3. The research community is actively engaged in analyzing and comparing approaches to ensure the security of cloud and IoT data through the use of authenticators. The adoption of SSL, public key cryptography, and comprehensive security frameworks helps protect data during transmission, authenticate users, and streamline job submission procedures. By prioritizing efficiency, security, and scalability/elasticity, the field of cloud-based big data analytics can continue to evolve and provide safe and reliable solutions for analyzing massive datasets.

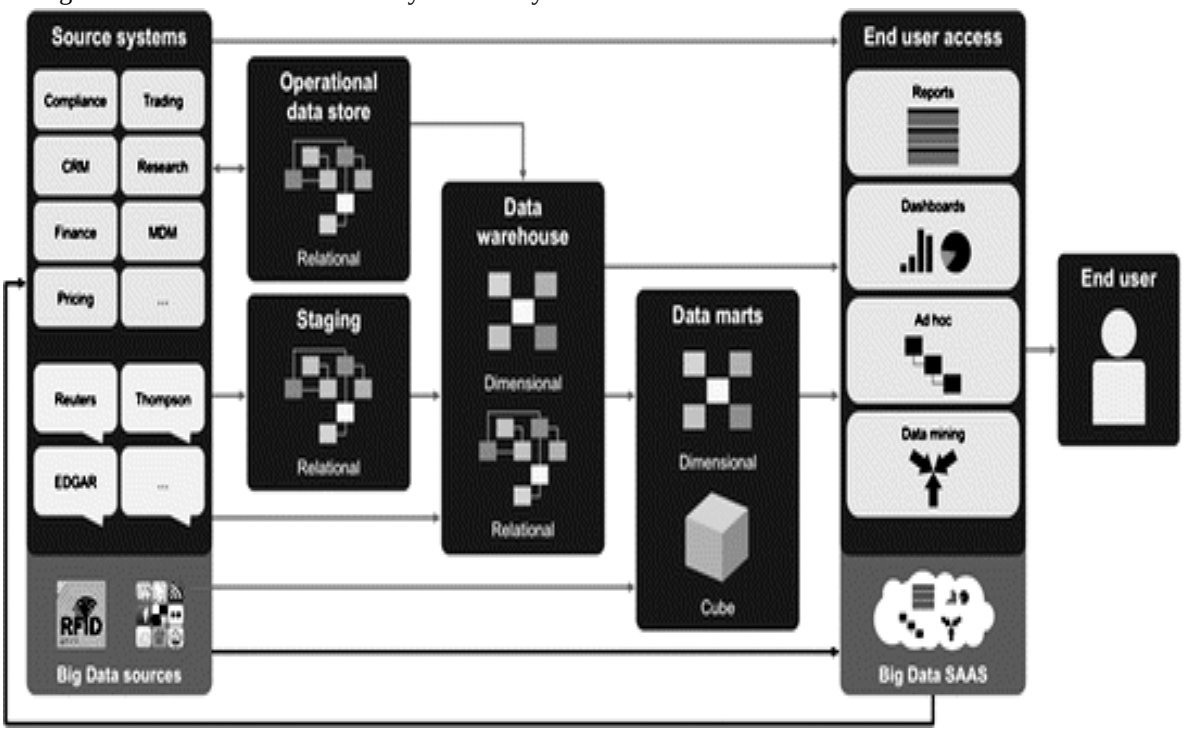

Fig 3. Big Data and Cloud Computing

- ***FUTURE RESEARCH DIRECTIONS***

The availability of open-source data mining strategies, resources, and technologies provides a diverse range of options for analysts and researchers. Notable examples include R, Gate, Rapid-Miner, and Weka, among many others. These tools, available under open-source licenses, enable cost-effective and efficient data mining solutions. When combined with the capabilities of cloud-based analytics, organizations can easily access affordable and scalable data analytics services. The simplicity of accessibility, low cost, and ease of installation and testing make cloud-based analytics increasingly popular in various domains. The figure 4 shows the importance of big data and data stored year wise.

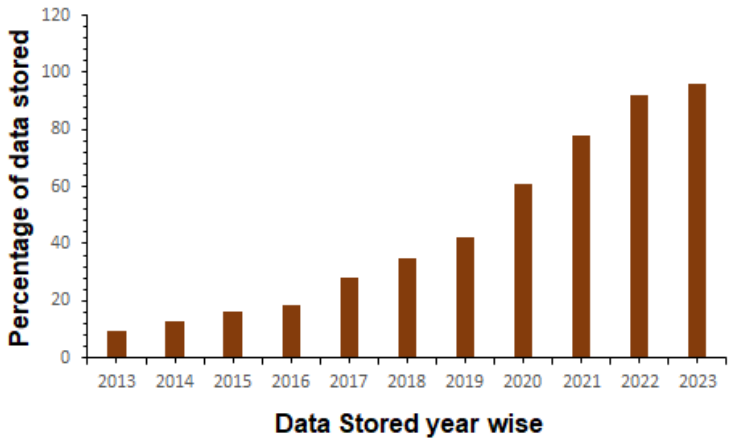

Fig 4. Data Stored Year Wise

In light of these advancements, several key study directions have been identified to further enhance the field of cloud-based big data analytics. Firstly, there is a focus on developing analytics and information management techniques specific to cloud-based environments. This involves designing novel methodologies and frameworks that leverage the unique features and capabilities offered by cloud platforms, enabling efficient and effective data analysis.

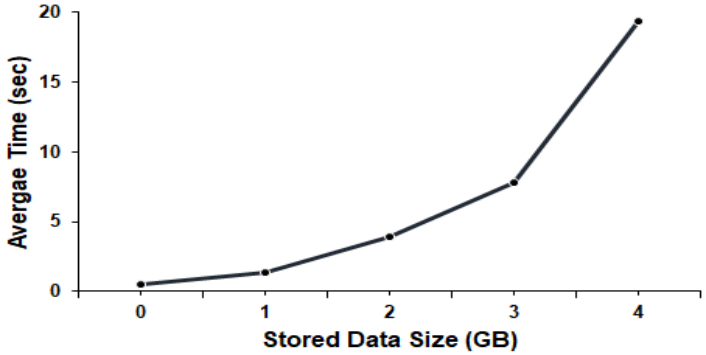

Fig 5. Dataset storage size and time

Secondly, researchers are working towards modifying and improving existing methods and approaches to enhance productivity and minimize risks associated with cloud-based analytics. This includes exploring techniques for optimizing resource allocation, reducing computational overhead, and mitigating potential security vulnerabilities. The figure 5 depicts the importance of data set stored and time taken for computing with time in seconds. Another critical area of study is centered around developing comprehensive plans and procedures to address privacy and security concerns of users in cloud-based analytics. As data privacy becomes a growing concern, efforts are being made to design robust encryption mechanisms, access control frameworks, and data anonymization techniques that safeguard sensitive information while facilitating analysis.

Furthermore, researchers are investigating the legal and ethical implications of technological advancements in cloud-based big data analytics. This involves adapting existing legal frameworks and ethical practices to align with shifting perspectives and societal expectations. Addressing issues such as data ownership, consent, and fairness in algorithmic decision-making is crucial for fostering trust and responsible use of cloud-based analytics. It is important to note that the study directions discussed above are not exhaustive, and the field of cloud-based big data analytics is continuously evolving. The primary objective remains the transition of the cloud from a mere data management and infrastructure platform to a scalable and powerful environment capable of facilitating advanced data analytics. By actively exploring these avenues of research, the aim is to unlock the full potential of cloud-based analytics, empowering organizations to extract valuable insights from their big data while ensuring privacy, security, and ethical considerations are upheld.

6. Conclusion

Professionals and academics are currently witnessing a period of great significance in the study of big data, as it continues to gain prominence in various fields. The exponential growth in digital data production in recent years has led to an increased focus on big data analytics and analysis. Notably, a substantial portion of this vast amount of data has already been uploaded to the cloud, making it a feasible option to shift big data analytics to a

cloud-based architecture. By leveraging cloud infrastructure, it becomes possible to accommodate the storage and processing requirements of data analytics algorithms effectively.

However, despite the numerous advantages, certain challenges remain unresolved. One major concern is the issue of ownership and control over the data residing in the cloud. As organizations rely more on cloud-based solutions, questions arise regarding who has ultimate ownership and authority over the data and its analysis. Additionally, safety and privacy concerns loom large, as the cloud environment raises questions about data security, confidentiality, and compliance with regulatory frameworks.

To tackle these challenges, ongoing research investigations are underway in the realm of cloud-based big data analytics. The primary objective is to develop a robust and efficient system that effectively addresses the identified risks and issues. These research endeavors aim to strike a balance between maintaining the advantages of cloud infrastructure while ensuring data ownership, control, safety, and privacy. Solutions are being sought to implement appropriate security measures, encryption protocols, and access controls that safeguard data integrity and confidentiality.

Furthermore, efforts are being made to establish transparent and accountable governance frameworks to mitigate concerns related to data ownership and control. By doing so, organizations can confidently embrace cloud-based big data analytics without compromising on privacy or exposing themselves to potential risks. The ultimate goal is to create an effective and efficient system that harnesses the power of the cloud, enabling seamless and secure big data analytics while addressing the ethical, legal, and technical challenges inherent in this domain.

In summary, the growing significance of big data analytics is driving the exploration of cloud-based architectures as a viable solution. While the cloud infrastructure offers scalability, storage, and processing capabilities, challenges related to ownership, control, safety, and privacy persist. Through ongoing research, experts are working towards developing comprehensive systems that ensure effective data governance and address the identified dangers and issues, enabling organizations to fully leverage the potential of cloud-based big data analytics.

References

[1] D. Das and M. Nayak, “Big Data Analytics: An overview,” Applications of Machine Learning in Big-Data Analytics and Cloud Computing, pp. 271–287, 2022. doi:10.1201/9781003337218-13

[2] Y. Demchenko, “Big Data Platforms and tools for data analytics in the Data Science Engineering Curriculum,” Proceedings of the 2019 3rd International Conference on Cloud and Big Data Computing, 2019. doi:10.1145/3358505.3358512

[3] P. Dharanyadevi, J. Therese M, B. Senthilnayaki, A. Devi, and K. Venkatalakshmi, “Cram on data recovery and backup cloud computing techniques,” Intelligent Network Design Driven by Big Data Analytics, IoT, AI and Cloud Computing, pp. 115–134, 2022. doi:10.1049/pbpc054e_ch6

[4] Y. Demchenko, “Big Data Platforms and tools for data analytics in the Data Science Engineering Curriculum,” Proceedings of the 2019 3rd International Conference on Cloud and Big Data Computing, 2019. doi:10.1145/3358505.3358512

[5] C. Komalavalli and C. Laroiya, “Challenges in big data analytics techniques: A survey,” 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2019. doi:10.1109/confluence.2019.8776932

[6] “2022 7th International Conference on Cloud Computing and big data analytics [front cover],” 2022 7th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), 2022. doi:10.1109/icccbda55098.2022.9778865

[7] S. Kumar, G. Mapp, and K. Cengiz, “Introduction to intelligent network design driven by Big Data Analytics, IOT, AI and cloud computing,” Intelligent Network Design Driven by Big Data Analytics, IoT, AI and Cloud Computing, pp. 1–12, 2022. doi:10.1049/pbpc054e_ch1

[8] O. Akerele, “Software process simulation modelling for Agile Cloud Software Development Projects: Techniques and Applications,” Strategic Engineering for Cloud Computing and Big Data Analytics, pp. 119–139, 2017. doi:10.1007/978-3-319-52491-7_7.

[9] Muniswamaiah, Manoj, Tilak Agerwala, and Charles Tappert. "Big data in cloud computing review and opportunities." arXiv preprint arXiv:1912.10821 (2019).

[10] Leo, L. M. ., Simla, A. J. ., Kumaran, J. C. ., Julalha, A. N. ., & Bhavani, R. . (2023). Blockchain based Automated Construction Model Accuracy Prediction using DeepQ Decision Tree. International Journal on Recent and Innovation Trends in Computing and Communication, 11(1), 133–138. https://doi.org/10.17762/ijritcc.v11i1.6060

[11] Dwarkanath Pande, S. ., & Hasane Ahammad, D. S. . (2022). Cognitive Computing-Based Network Access Control System in Secure Physical Layer. Research Journal of Computer Systems and Engineering, 3(1), 14–20. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/36