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

Big Data Architecture Framework for Data Analysis and Processing in Ecosystem

K. Chanthirasekaran[1], S. Yogeeswaran[2], Rafikullah Deraman[3], T. Jackulin[4], S. V. Manikanthan[5]

Submitted: 16/07/2023 **Revised**: 08/09/2023 **Accepted**: 25/09/2023

Abstract: Big Data is a well-known buzzword in industry and academics, but there is still much conceptual uncertainty surrounding it, making it difficult to understand what it really means. The phrase is used to denote a wide range of concepts, from technology's ability to receive, repair, and use information to the cultural transformation that is extensively influencing both industries and societies, both of which are drowning in information. Research has gone in a lot of different areas because there is no official definition. This article explores the nature of big data, which can come from various fields of science, business, and social activity. It also suggests an improved big data definition that contains the following elements: Aspects of large data include facilities, safety, data infrastructure, data concepts, and data architectures. This article talks about the paradigm change in big data operations from traditional host- or service-oriented design to data-centric architecture. The Big Data Architecture Framework (BDAF), which is proposed to handle every area of the Big Data Ecosystem, is made up of the following components. As a result, the analysis of big data is currently the subject of study and development. Examining the possible effects of big data challenges, unanswered research problems, and related tools is the primary objective of this work. Thus, this paper provides a paradigm for investigating big data at different stages. Academics now have novel possibilities to develop the solution as a result of the difficulties and open research problems.

Keywords: *Unstructured data, Big Data Architecture Framework (BDAF), Ecosystem, bid data technology.*

1. Introduction

Big data analysis is at the heart of contemporary research and business. These data are generated by chats on social networking sites, scientific data, sensors, communications, visit waterways, records, announcements, pictures, speech, photographs, and other transactions via the internet as well as by handheld devices and apps, click rivers, records, and searches. They are stored in databases, which grow dramatically over time and become more and harder to collect, arrange, oversee, share, analyze, and visualize using conventional database software technologies [1]. Because it is heterogeneous, unstructured, and inconsistent and comes from various sources, this data necessitates extensive storage and effective processing capabilities. Big data's most crucial component is value, which is the process of extracting enormously valuable hidden information from massive datasets of diverse types that are produced quickly. Due to the information that big data contains—information that is not present in small-sized data—we are able to make informed decisions from both structured and unstructured data. There are numerous applications for big data skills in figure 1.1.

The Emotional Face of Big Data, a global initiative that centers on the real-time collection, visualization, and analysis of massive amounts of data, was completed in 2012. Numerous statistics are derived following this media initiative. Facebook boasts 955 million monthly active users who use 70 different languages, 140 billion photographs that have been posted, 125 billion friend relationships, 30 billion pieces of content that have been submitted every day, and 2.7 billion remarks and likes. 48 hours of videos are posted to YouTube every minute, and 4 billion views are made there each day. Google offers a wide range of services, monitoring 7.2 billion pages daily, processing 20 petabytes (1015 bytes) of data each day, and translating into 66 different languages. More than 140 million active Twitter users send 1 billion Tweets every seven days. There are 571 brand-new websites launched each minute of the day. The amount of information will grow by 50 times over the next ten years, but the number of technologists who can keep up with it will only grow by 1.5 times.

[1]*Professor, Department of Electronics and Communication Engineering, Saveetha Engineering College, Chennai.*
chanthirasekarank@saveetha.ac.in,
ORCID: https://orcid.org/0000-0002-1856-028X
[2]*Assistant professor, Department of ECE, P.T.Lee Chengalvaraya Naicker College of Engineering and Technology, Oovery, Kanchipuram-631 502, Tamilnadu, India*
yogash28s@gmail.com
[3]*Project and Facilities Management Research Group, Faculty of Civil Engineering and Built Environment, Universiti Tun Hussein Onn Malaysia, Batu Pahat, Malaysia.*
Email: rafikullah@uthm.edu.my
[4]*Associate Professor, Panimalar Engineering College, Chennai, India.*
Mail id – karthijackulin@gmail.com,
Orcid: 0000-0003-4015-7718
[5]*Director, Melange Academic Research Associates, Puducherry, India*
prof.manikanthan@gmail.com

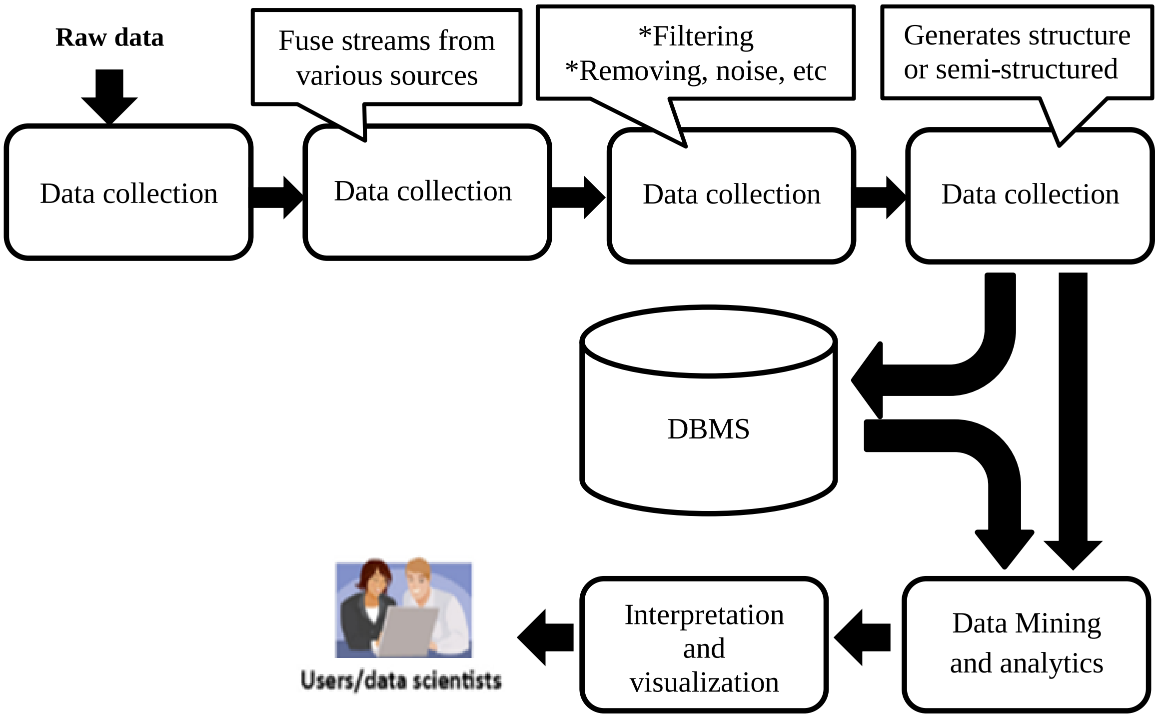

Fig 1.1. Basic Big Data components that make up an infrastructure

Big data is a crucial component for businesses looking to gain a competitive edge,4 as it can aid in the creation of new goods and services, the quicker automation of strategic and operational decisions, the identification of past events and the prediction of future ones, the identification of consumer behaviour [3], the direction of targeted advertising, the production of greater returns on investments, the identification of sales and market opportunities, the planning and forecasting of future sales, and the increase of manufacturing.

- o Project managers and system builders are assisted in implementing big data ecosystems in commercial organizations by this novel solution to the problem of developing frameworks.
- o Several of the current frameworks employed for this goal are covered in the linked literature review on big data for business management.
- o The big data project life cycle is covered by the methodology component of the suggested framework, which also specifies when and how to employ the other components.

The text of the article is as follows: Section II discusses important problems, benefits, difficulties, survey findings, samples, techniques, and knowledge discovery from huge data. The significant security-related issues are reviewed in Section III. Big data's advantages, potential drawbacks, difficulties, and hurdles are discussed in Section IV. The task is concluded in Section V.

2. Literature Review

Cui, Y., Kara, S., et.al [4] In closing, understanding the data needs of manufacturing applications, being aware of the strengths and limitations of big data technologies, and identifying the gaps can help define future research areas and inspire fresh concepts for creative applications. Big data in manufacturing is presented holistically in this systematic literature review to investigate the potential applications for manufacturing. It is conceptual. This systematic literature review's framework is composed of three levels: information source, big data ecosystem, and information users.

Khalifa, S., et.al [5] The shortage of analytics talent can be filled with the help of consumable analytics. By offering in-tool support to make analytics easier to design, manage, and execute, it is hoped that the talents currently present in organizations will have a greater impact. Assistance may be static, displaying the same information regardless of the dataset being analysed. Users can benefit from static support when configuring operations, but not when choosing them. Because there are so many analytics approaches available, inexperienced analysts frequently find them unable to confidently choose the best one and must instead resort to time-consuming experimentation.

Anwar, M. J., et.al [6] With the need for digital transformation rising, many companies are turning to smart devices to profit from advancements in technology. These devices generate enormous amounts of data, which calls for real-time processing. Massive data is a word for enormous data collections with massive and varied formats that is used to describe them. Big data presents a challenge for conventional techniques of storing, analysing, and visualization for advanced processing due to its inherent nature. Big data has numerous definitions.

Allam, S. et.al [7] These records must be digitalized for the current digital world. In order to respond to new issues, a great amount of information must be accurately analyzed. This will improve healthcare efficiency while lowering costs. The government continues to generate petabytes of data each day. It requires tools that make it possible to review the massive amounts of data being collected in real time. This would help the government provide people with amenities that enhance value. By using machine learning techniques to understand information structures and linkages, big data analytics aids in the discovery of advantageous solutions.

Benhlima, L. et.al [8] Technologies based on big data have drawn a lot of attention because they handle large volumes of data more successfully than conventional methods. Structured, semi-structured, and unstructured information are all supported by big data frameworks, which also offer several functionalities. Designing predictive models and using large data mining tools are examples of those features that improve decision-making by choosing pertinent data. Stream manufacturing, on the other hand, is appropriate for applications that need real-time feedback. Through the collection and storage of batches, batch processing tries to process a large amount of information to provide results.

Qin, P., Li, W. et.al [9] presents the MetaCloudDataStorage strategy, a cloud computing security framework. In the Meta cloud centre, data is separated and stored at many cloud tiers. Data security is the responsibility of the cloud. Data segmentation and the real storage's physical details are stored in the Meta cloud center. The approach lacks data access security monitoring and authentication, and the cloud information is not encrypted either.

Farley, S. S., et.al [10] Since ecology is a diverse field, comprehending and incorporating the range of ecological data is a first-order big-data challenge. Data systems, which here are defined as a collection of data gathered by scientists and the accompanying quantitative and theoretical structures for understanding these measures, can be used to organize ecological information; a data system typically consists of several different data kinds. Each research community uses a diverse set of data systems, has varying degrees of competence, and has made varying amounts of past investments in ecoinformatics. The scope and nature of big data challenges and their associated solutions differ between societies and data platforms.

Kumar, S., et.al [11] In the past two years, forecasting has been regarded as one of the major business intelligence approaches, but its applications in the real-world reach beyond the corporate setting. Among the many methods employed in big data analysis, text mining and multimedia insights are just two. One of the most significant subcategories is predictive analytics, which covers statistical methods like data extraction and artificial intelligence that use past and present data to create forecasts.

3. Methods and Materials

Massive Data Analysis

Recently, the term "Big Data" has been employed to encompass datasets that get large enough that they're challenging to organize through traditional database management platforms [12]. These are collections of information that are excessively big for frequently employed software tools and storage devices to gather, archive, handle, and deal with in a feasible span of time.

The quantity of data in a single massive data collection currently ranges anywhere from a few dozen terabytes (TB) to many petabytes (PB). So, a few of the difficulties associated with mass data include collecting, preserving, seeking, sharing, analyzing, and displaying. Enterprises nowadays are looking at immense quantities of meticulous data to unearth previously unexplored realities.

Thus, big data analytics implies the integration of advanced analytical techniques into massive data sets. Big data sample analytics uncovers and takes advantage of organizational transformation. Anyway, controlling a greater collection of data becomes more challenging.

Our commitment is to begin this part by exploring big data's characteristics and importance. Naturally, massive and more intricate collections of information can frequently be examined for business purposes, necessitating actual time or near-real-time capabilities; subsequently, that ultimately results in a demand for fresh information designs, analytical techniques, and instruments. As a consequence, the subsequent part will dive into more details on big data analytics instruments and procedures, beginning with the big preservation and management of data and ending with big data analytic computation. Ultimately, several of the multiple big data analytics that have become more popular as a result of big data is show in below figure 3.2.

Big Data features

Massive data is information that is so massive, scattered, broad, or sensitive to time that it mandates the adoption of novel technology structures, analysis, and equipment for the purpose to generate breakthroughs that open up fresh sources of economic value. Big data can be summarized as the aforementioned three V's, or volume, variety, and velocity. The quantity and scope of the information provided have been measured by its volume. Velocity is the speed at which this information changes or how often it is generated. The varied formats and types of info, as well as numerous applications and their techniques of analysing information will round off the diversity. Big data's most significant attribute is data volume. The magnitude of massive information in TBs or PBs, as well as the total amount of documentation, the transactions, tables, and records, may all be evaluated.

Furthermore, a component that makes massive information so important is the fact that it emanates from numerous sources than ever prior to, such as clickstreams, logs, and social media. Due to the application of these avenues for data analytics, common organized information is currently mixing semi-structured information with unstructured information, such as text and spoken language. This section lists AI challenges and links each challenge to a particular Big Data component. The dimensions of Big Data are depicted in Fig. 3.1, together with the difficulties they entail, for further discussion in the following sections.

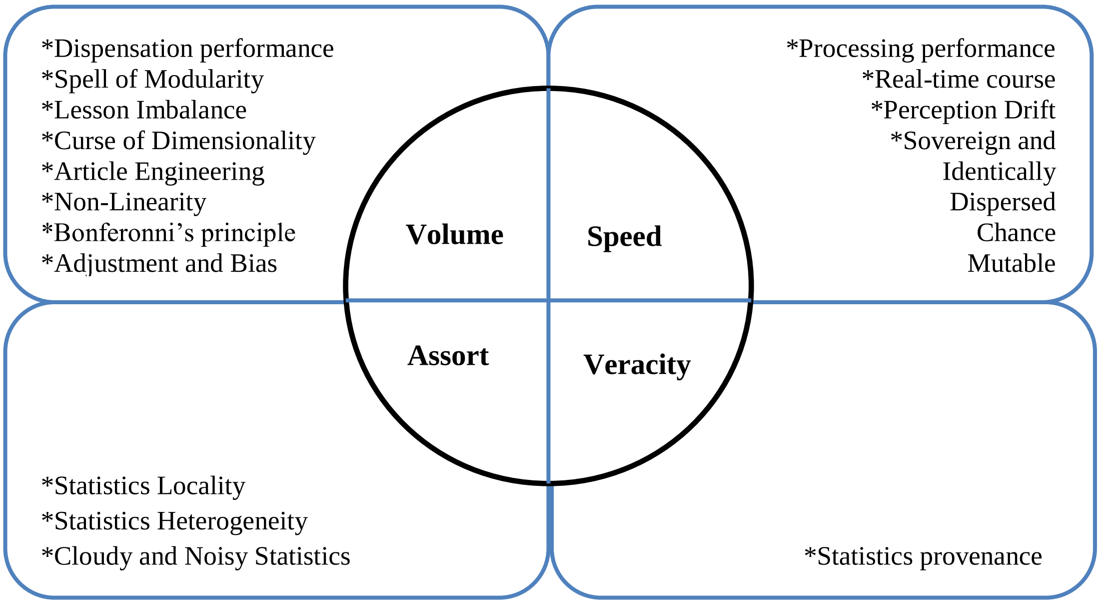

Fig 3.1. Big Data characteristics and related difficulties

In addition, to give substantial historical context, multidimensional information can be retrieved from a warehouse of data. So, in the context of large volumes of data, variation is exactly as vital as volume. Big information may additionally be defined through its velocity or speed. In simple terms, this describes the pace of data production or transfer. Streaming data, especially gets collected in the moment from web pages, is at the cutting edge of big data.

The inclusion of a fourth V, or veracity, was recently debated by specific scholars and institutions. The level of accuracy of the data becomes its primary point of veracity. As a direct consequence of data inconsistency, unfinished business, unclear information, latency, deception, and estimations, large-scale information quality has been defined as excellent, awful events, or undefined.

Large-Scale Ecosystem

Although the reality is those databases and Hadoop are the major technology and tools for massive information statistical analyses and big data processing. That's a broad variety of devices to keep, process, observe, and convey solutions to the intended applications. Data mining is "an all data-related operations, spanning source, goal, and the ultimate result demand "fuel," according to Wikipedia.

The Big Data Ecosystem, which interacts with the fluctuating information, theories, and underlying equipment across the entire Big Data lifetime, is sometimes alluded to as a complicated set of interconnected components. We will get into more information concerning our goal for the BDE in the sections that follow.

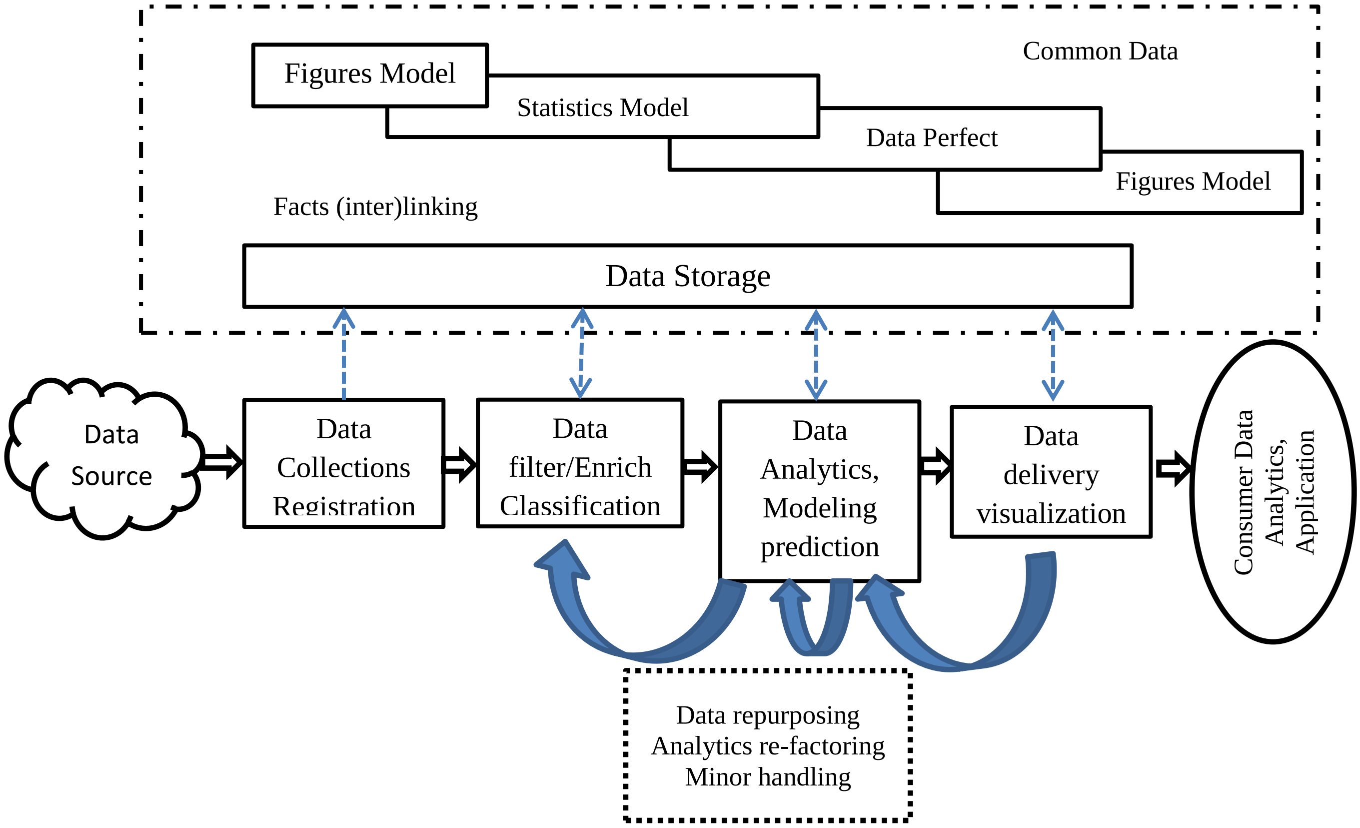

Fig 3.2. Large Data Lifecycle in the Large Data Ecosystem

The Big Data Lifecycle Management framework in Figure 3.2, it was put forth as a result of an analysis of current procedures in various academic and commercial technology areas [12], reflects the new approach to data management and processing that is necessary in the Big Data business.

Big data and data in-depth knowledge and technologies model change

The change in outlook in present e-Science and business practices has been rendered achievable by recent breakthroughs in ICT overall, cloud computing, and technologies for large-scale data. The evolution is illustrated by several key features.

- Digital transformation of present-day artifacts alongside additional material, as well as the conversion of all procedures, occurrences, and finished goods into electronic format using multidimensional, multifaceted tests, surveillance, and management.
- Automation of any information generation, utilization, and administration operations, including gathering information, preservation, categorization, indexing techniques, and additional aspects of wider data governance and traceability.
- Possibility to recycle and redeploy the elementary information sets based on model construction for secondary and current analysis.
- International information availability and admission throughout the system allow joint research organizations or technological professionals, to offer open access to knowledge used in manufacturing or science.
- The readily accessible nature of vital elements and administration solutions, allows rapid architecture, resource arrangement, modification, and planning as per the requests for specific academic assignments or projects.
- Cutting-edge protection and communication control technologies guarantee the safe operation of the complex industrial and research facilities and enable the creation of a trustworthy secure ecosystem for coordinated research groups and technologically certified professionals.

Screening Sensor Data via KF

(1) Start-up stage

U_l represents the state transition model (used on the earlier state g_{l-1})

P_l is the model used for observation

R_l indicates the covariance of the process noise

S_l illustrates the covariance of the observation noise

D_l is also known as control input model (as it relates to the control vector w_l)

$$x_l \sim O(0, R_l)$$

(2) Using the old form g_{l-1}, computing the new state g_l

$$g_l = U_l g_{l-1} + D_l w_l + x_l$$

$$i_l = P_l g_l + v_l \quad v_l \sim O(0, S_l)$$

(1)

(3) Evaluation of current state using earlier state

Predicted state

$\hat{g}_{l|l-1} = U_l \hat{g}_{l-1|l-1} + D_l w_l$
(2)

Predicted covariance

$H_{l|l-1} = U_l H_{l-1|l-1} U_l^U + R_l$
(3)

(4) Integrating the most recent forecast with the most recent figuring
Current observation
$\tilde{y}_l = i_l - P_l \hat{g}_{l|l-1}$
(4)

Observation covariance

$T_l = P_l H_{l|l-1} P_l^U + S_l$
(5)

Optimal gain

$L_l = H_{l|l-1} P_l^U T_l^{-1}$
(6)

Update state (forecast and observation)

$\hat{g}_{l|l} = \hat{g}_{l|l-1} + L_l \tilde{y}_l$
(7)

(5) Update Covariance (prediction and observation)

$H_{l|l} = (J - L_l P_l) H_{l|l-1}$
(8)

In the beginning, it considers that the present state, g_l, emerged from state g_{l-1}. Here $i_l . \hat{g}_{l|l-1}$ indicates the estimated value of g at time l, whereas the prediction accuracy signifies the estimation of g at time l. $H_{l|l-1}$is defined as the estimation precision. It calculates useful details from an immense gathering of hazy and indirect data. The KF processes knowledge when it enters in because it performs iteratively. As a consequence of this, it guarantees the intelligent city functions in real-time. Furthermore, it permits lightning-fast execution with minimal memory usage.

Big Data Lifecycle and handling of data

The business sector and others are stepping into an entirely new arena as a consequence of the growing acceptance of modern technologies in every facet of their enterprise tasks. To successfully take delight in the fresh chances to gather and mine information needed for important insights [13, 14], which include market forecasting, consumer behavior projections, social class behavior assumptions, etc., they must utilize scientific approaches.

The implementation of biological exploration techniques, such as periodic model construction and acquiring of enhanced data, as well as recycling of acquired information with updated models, has been proposed in several instances of blog posts in addition to industry papers.

We generate allusion to the Scientific Data Lifecycle Management paradigm that served as the centerpiece of a comprehensive study in a separate publication and was discussed in our previous paper, which portrays the intricate and iterative procedure for scientific investigation and consists of numerous subsequent stages: Conducting an inquiry or experiment, gathering information, evaluating it, distributing conclusions, getting remarks, discussing it, and protecting (or discarding) the outcomes are all involved.

Data retention and safeguarding are mandated by the new BDLM throughout its segments, which should offer data reuse, repurposing, secondary investigation, and subsequent evaluation of the processed information and reported conclusions. The scenario is only feasible, however, if BDI completely implements data authentication, cross-referencing, and connection. Throughout the whole data lifespan, reliability of data, managing access, and fiscal responsibility must be presented. A vital aspect of the characterized BDLM is data duration, which somewhat is required to be accomplished in a trustworthy and safeguarded manner.

4. Implementation and Experimental Results

The research results and responses are discussed in this portion.

Experimental Setup

Three sets of 8-bit, 64-bit, and 256-bit inputs for data were employed in the present research to evaluate the utilization of RSA and ECC to retrieve information privacy alongside randomly generated secret keys. The research projects were performed on a Windows 10 computer platform together with MATLAB and 3 GB of RAM.

Experimentation Outcomes

For the ECC and RSA methods, the evaluations use a wide range of the inputs, including data inputs with 8-bit, 64-bit, and 256-bit resolutions. Table 1 show the encryption as well as decryption times for both approaches for 8-bit input. The research results show that in comparison to the RSA method, ECC may decode data more rapidly.

Table 1. Shows the encryption, Decryption and Cumulative times for eight bits (in seconds)

Safety Bit level	Encryption Interval		Decryption Interval		Cumulative Time	
	ECC	RSA	ECC	RSA	ECC	RSA
113	3.20	1.02	1.50	1.70	3.70	1.72
129	4.87	1.03	1.80	2.94	5.67	2.97
145	5.72	1.04	2.00	3.64	6.72	3.68

Table 2. Shows the encryption, Decryption and Cumulative times for 64 bits (in seconds)

Safety Bit level	Encryption Interval		Decryption Interval		Cumulative Time	
	ECC	RSA	ECC	RSA	ECC	RSA
113	10.99	1.17	7.94	21.42	17.92	21.58
129	16.09	1.17	8.36	47.49	23.45	47.65
145	21.24	1.14	9.48	78.77	29.71	78.91

Similar to the first table, Table 2 illustrates the time required for the encryption, decryption, and total processing of 64-bit data input using the ECC and RSA methods. The research results indicate that the RSA tactics consume considerably fewer moments for encryption than do ECC solutions. The ECC algorithm, likewise, decrypts data at the 113, 129, and 145 security bit levels considerably quicker.

As an outcome, the ECC technique needs a shorter period altogether. Additionally, as the authentication bit value goes from 112 to 144, the time inequality between ECC and RSA expands [15, 16]. Whenever the security bit level goes up to 145, the gap that exists between ECC and RSA techniques, for instance, climbs from 3.66 seconds to 49.2 seconds, demonstrating RSA requires 52.2 seconds longer than ECC to process 64-bit input.

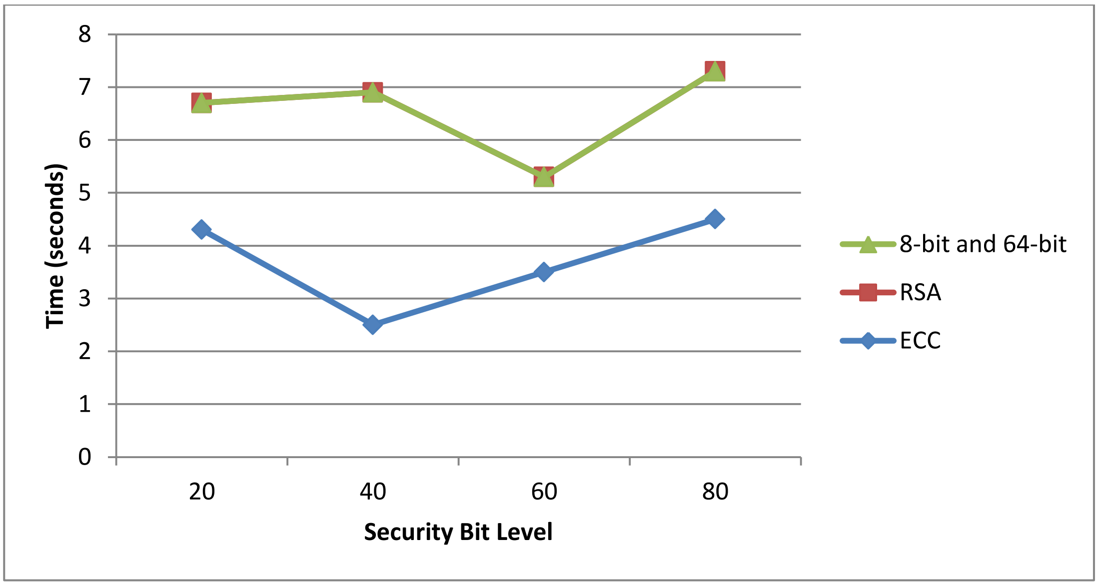

Fig 4.1. Execution times (in seconds) for RSA and ECC, as well as the total input times for 8-bit and 64-bit data

The 256-bit data input demonstrates identical patterns, as illustrated in Table 3. In comparison with RSA, ECC is a far more effective and trustworthy approach that provides powerful safeguards while lowering key lengths. The overall user experience becomes better as an outcome of a lesser amount of network and mathematical resource requirements. Although ECC is capable of handling the same amount of responses per second in just 75 milliseconds, RSA is capable of performing 460 requests per second alongside an average response time of 160 ms.

Table 3. Shows the encryption, Decryption and Cumulative times for 256 bits (in seconds)

Safety Bit level	Encryption Interval		Decryption Interval		Cumulative Time	
	ECC	RSA	ECC	RSA	ECC	RSA
113	40.71	1.59	27.3 4	103.0 4	67.04	103.6 2
129	59.44	1.57	28.4 1	210.6 1	86.85	211.17
145	78.51	1.58	33.1 6	312.6 1	110.6 6	312.6 4

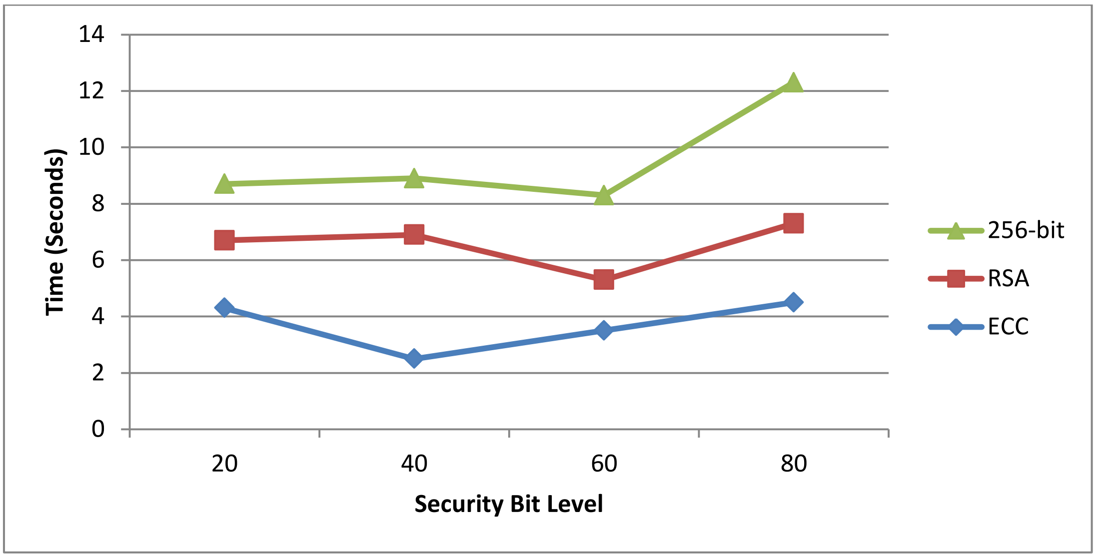

Fig 4.2. Total input time for 256 bits

Figure 4.2 offers total input time for 256 bits. The experiments revealed that RSA is lightning-fast at encrypting data however it requires a lot of time to decode it, whereas ECC is sluggish at encrypting data but speedy at decrypting. According to the research, it is often argued that ECC usually proves more efficient and safer than RSA.

5. Discussion

The experiment's goal was to figure 4.1 out the time length necessary to perform for RSA and ECC encryption and decryption on three distinct input patterns—8, 64, and 256 bits—using randomly selected keys that corresponded with NIST regulations. Conclusions indicate that the ECC exceeded, RSA in terms of reduced variable operational risk and safety. ECC is excellent for applications with limited funds. One of the primary benefits of ECC is the fact that in terms of contemporary key sizes, it is considerably more secure than RSA. A 2048-bit RSA key, which is equivalent to a 3072-bit RSA key, is 10,000 times worse than a 256-bit traditional ECC key. For a computer to exceed an adversary's processing capability, RSA keys must be lengthier. ECC is also faster for many different kinds of reasons. The initial advantages of employing smaller keys lies in the fact they require less information to be sent throughout the SSL transaction, from the web server to the user. ECC dramatically reduces CPU and memory usage, boosting performance and response speeds anytime it is active on Web servers. The encrypted suites accepted by modern Web servers and browsers that establish PFS additionally impose ECC, disregarding the fact that PFS is not an additional benefit from ECC. Website hosting providers that support ECDHE use cipher packages that reap the benefits from both PFS and ECC.

6. Conclusion

The present investigation examined a framework for directing the creation and application of big data ecosystems. With the addition of new information, we designed its basic design using the existing literature.

Readers can gain a better understanding of which solution might be used for specific non-functional criteria thanks to our poll. We have now gone on to discuss pertinent issues that may help such systems evolve properly in the future. Each problem outlined in the part titled "Lessons learned and future research goals" has been derived from the analysis that has been done. Finally, we have offered a set of exciting ideas for future research work paths relevant to both the functionality associated with Big Data administration (i.e., Big Data storage and deployment) and the Big Data lifecycle administration as a whole in order to meet the stated issues.

The definition of big data will be expanded upon in future study and development, as was done in this work. We made an effort to condense and reconsider some definitions of big data that are frequently used. However, more formal approaches and taxonomies of common big data use cases in various big data origin and target areas, as well as an examination of various stakeholder groups, will be needed for further research.

References

[1] Sagiroglu, S., & Sinanc, D. (2013, May). Big data: A review. In *2013 international conference on collaboration technologies and systems (CTS)* (pp. 42-47). IEEE.

[2] Arfat, Y., Usman, S., Mehmood, R., & Katib, I. (2020). Big data for smart infrastructure design: Opportunities and challenges. *Smart Infrastructure*

and Applications: Foundations for Smarter Cities and Societies, 491-518.

[3] Orenga-Roglá, S., & Chalmeta, R. (2018). Framework for implementing a big data ecosystem in organizations. *Communications of the ACM*, *62*(1), 58-65.

[4] Cui, Y., Kara, S., & Chan, K. C. (2020). Manufacturing big data ecosystem: A systematic literature review. *Robotics and computer-integrated Manufacturing*, *62*, 101861.

[5] Khalifa, S., Elshater, Y., Sundaravarathan, K., Bhat, A., Martin, P., Imam, F., ... & Statchuk, C. (2016). The six pillars for building big data analytics ecosystems. *ACM Computing Surveys (CSUR)*, *49*(2), 1-36.

[6] Anwar, M. J., Gill, A. Q., Hussain, F. K., & Imran, M. (2021). Secure big data ecosystem architecture: challenges and solutions. *EURASIP Journal on Wireless Communications and Networking*, *2021*(1), 130.

[7] Allam, S. Research on the Secure Medical Big Data Ecosystem Based on Hadoop. *International Journal of Creative Research Thoughts (IJCRT), ISSN*, *2320*(2882), 815-819.

[8] Benhlima, L. (2018). Big data management for healthcare systems: architecture, requirements, and implementation. *Advances in bioinformatics*, *2018*.

[9] Qin, P., Li, W., & Ding, K. (2022). A big data security architecture based on blockchain and trusted data cloud center. *Wireless Communications and Mobile Computing*, *2022*.

[10] Farley, S. S., Dawson, A., Goring, S. J., & Williams, J. W. (2018). Situating ecology as a big-data science: Current advances, challenges, and solutions. *BioScience*, *68*(8), 563-576.

[11] Kumar, S., & Singh, M. (2018). Big data analytics for healthcare industry: impact, applications, and tools. *Big data mining and analytics*, *2*(1), 48-57.

[12] Demchenko, Y., De Laat, C., & Membrey, P. (2014, May). Defining architecture components of the Big Data Ecosystem. In *2014 International conference on collaboration technologies and systems (CTS)* (pp. 104-112). IEEE.

[13] Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of business research*, *70*, 263-286.

[14] L'heureux, A., Grolinger, K., Elyamany, H. F., & Capretz, M. A. (2017). Machine learning with big data: Challenges and approaches. *Ieee Access*, *5*, 7776-7797.

[15] Begoli, E., & Horey, J. (2012, August). Design principles for effective knowledge discovery from big data. In *2012 Joint Working IEEE/IFIP Conference on Software Architecture and European Conference on Software Architecture* (pp. 215-218). IEEE.

[16] Begoli, Edmon, and James Horey. "Design principles for effective knowledge discovery from big data." In *2012 Joint Working IEEE/IFIP Conference on Software Architecture and European Conference on Software Architecture*, pp. 215-218. IEEE, 2012.

[17] Khare, S. ., & Badholia, A. . (2023). BLA2C2: Design of a Novel Blockchain-based Light-Weight Authentication & Access Control Layer for Cloud Deployments. International Journal on Recent and Innovation Trends in Computing and Communication, 11(3), 283–294. https://doi.org/10.17762/ijritcc.v11i3.6359

[18] Kamau, J., Ben-David, Y., Santos, M., Joo-young, L., & Tanaka, A. Predictive Analytics for Customer Churn in the Telecom Industry. Kuwait Journal of Machine Learning, 1(3). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/130

[19] Chaudhury, S., Dhabliya, D., Madan, S., Chakrabarti, S. Blockchain technology: A global provider of digital technology and services (2023) Building Secure Business Models Through Blockchain Technology: Tactics, Methods, Limitations, and Performance, pp. 168-193.