A NOVEL FRAMEWORK FOR BIG DATA SECURITY IN CLOUD COMPUTING: PERFORMANCE AND RELIABILITY

[1]**Renu Yadav**

(*Research Scholar*), School of Computer Application & Technology,

Career Point University, Kota, (Rajasthan)

[2]**Dr. Abid Hussain**

(Associate Professor), School of Computer Applications,

Career Point University, Kota, (Rajasthan)

ABSTRACT

The quantity and variety of data being produced by both humans and technology are increasing throughout the world. Although this exponential growth, security and data protection needs are not intended to be addressed by the tools and technology created to manage large data volumes. Organizations must ensure that the cloud environment is secure from all threats and that the cloud service providers and data users are on the same page to protect the security of large amounts of data. Due to its significant role in tackling existing issues such as security reliability, data governance, and performance, cloud computing offers a wide variety of promises to safeguard sensitive data. Cloud service providers are exposed to several security risks since they supply the full scope of the existing big data security. This study presents a revolutionary approach for big data security in cloud computing. We explain the various security threats that arise when migrating data to the cloud. With the support of an identity and access key management system, access control, encryption, and centralized key management strategies, along with the suitable enforcement of policies, governance, and security level agreement, we utilized proper mechanisms in this framework to improve performance and reliability.

Keywords: Big data security, cloud computing, identity, and access management, cloud service provider, centralized encryption key management, access control.

1. INTRODUCTION

Big data security focuses on the measurement of all the techniques used to safeguard the process as well as the data from theft, unauthorized use, and other harmful activities like cyberattacks, data breaches, and other malicious activities. Considering cloud computing's tremendous potential to offer accessible and affordable access to significant processing capacity, the benefits of implementation are indisputable. The security issue has become even more vital because outsourced data may contain sensitive information. The fundamental security issue encompasses the data movement in and out of the cloud environment as well as the processing and storage of voluminous data inside of the data center. Multi-tenancy and virtualization, which ensure improved usage of resources but make it extremely difficult to provide secure processing, are addressed in the context of the cloud. For organizations that outsource to cloud computing, there are numerous threats and vulnerabilities including service and deployment model factors that affect its security. Data is processed and stored on servers owned by the service provider for big data organizations or individuals using the cloud. In this paper, we design a novel framework for big data security that provides a detailed and constructive approach to cloud security. Organizations must identify any cloud-specific, uncertainty security controls that are implemented and stored on cloud servers to ensure security while moving massive data to the cloud. Before an organization implements a cloud, security issues and vulnerabilities need to be carefully addressed. We also highlight several security concerns and the techniques used by cloud computing to improve performance and reliability, including access controls, service level agreements, centralized encryption key management, and secure identity and access management.

BIG DATA TECHNOLOGY

Big information is usually considered to refer to a variety of data types, such as structured, semi-structured, and unstructured data, that are getting more complex using the conventional method of data analysis. The definitions depict the five big data perspectives (Volume, Variety, Velocity, Veracity, and Value). Volume suggests a huge amount of information. When processing large datasets, a series of issues come up, including the obstacle of storing or preserving the entire information in memory and on a storage device as well as the complex problems with numerous features and characteristics. Variety pertains to many distinct data formats, such as text, audio, and video, that can be categorized as structured, semi-structured, or unstructured. The problems arising with diversity include data inconsistency, data granularity, inaccurate data, and dirty data. Velocity quantifies the data's flow and speed. The increase in high-definition videos, video streaming, and video gaming is another factor contributing to the congested traffic. The aspect of data quality is called veracity. In other words, the data must be reliable, accurate, and comprehensive. Value is a significant factor in big data. Big data is usually unnecessary on its own, as a result, data must be transformed into knowledge to remain efficient [1].

2. ROLE OF CLOUD COMPUTING FOR BIG DATA

Big data necessitates the use of high-performance computation and storage facilities. Cloud computing is the preferred platform for big data as it provides computation facilities as paid services. The capability of clouds such as infrastructure, scalability, flexibility, and cost-saving ability motivates organizations to prefer cloud computing-based services for data storage and processing. When compared with traditional processing platforms, the cloud provides capabilities like virtualization, secured and distributed storage, and processing competency that makes it suitable for big data. Cloud provides infrastructure, management, cost efficiency, and remote access from anywhere in the world with an internet connection effectively. Big data all 5V's features can be addressed by cloud computing. A variety of data types are dealt with by elasticity, pooled resources, and self-service. The volume describes the enormous amount of data that is created and stored, not in terabytes but in zettabytes or yottabytes. The velocity describes the speed at which new data is generated and the speed at which data moves around. By using self-service to choose the best-suited services, a pay-as-you-go pricing model, correct data representation of value, and cloud computing's elasticity capabilities, the validity of the data is alleviated. Combining the cloud computing utility model with a variety of cloud services for computations, infrastructures, and storage also creates a very desired environment.As a result, cloud computing works as a service model in addition to offering resources for big data computation and processing. Big data use cloud storage as opposed to local storage that is physically tied to a device. Rapidly cloud-based applications created with virtualized technologies are what fuel big data analysis. As a result, cloud computing works as a service model in addition to offering resources for big data computation and processing. However, security concerns are one of the main problems with big data, especially when a company stores and processes a lot of private and sensitive data for a cloud service provider [2].

3. CHALLENGES AND ISSUES OF BIG DATA SECURITY

Because of its inherently remote operations, resource residency, decentralized administrations, and management control, the security of Big Data throughout exporting are paramount. Data Owners have no complete control over the services that are using their data because of the cloud's architecture. The concerning issues are the most important ones to consider when evaluating data protection before migrating to a cloud environment. According to CSA, the Big Data Working Group has emphasized the primary hurdles to establishing appropriate security and privacy threats, particularly infrastructure security, data protection, management, integrity, and reactive security. We also recognized and addressed the most significant threats to big data security, including data loss, inefficient management of user identity, credentials, and access, insecure user interfaces and application programming interfaces, security vulnerabilities, account takeover by insiders with a nasty behavior, persistent advanced threats, unsatisfactory due diligence, cloud service mishandling, and suspicious use, service denial, and shortcomings in resource sharing[3]. Furthermore, organizations transitioning to cloud computing confront a variety of big data security threats. Organizations must determine all cloud-specific protection systems and procedures in advance to ensure security when shifting data to the cloud. In certain instances, it could be required to demand from cloud service providers that security controls and components be fully and precisely implemented through contractual mechanisms including service-level agreements (SLA). Cloud computing big data security necessitates a variety of techniques. The organization's security, governance, and regulatory compliance strategies may be impacted by all of these considerations. Users will take measures to ensure implementation is fully protected, configured, and managed because big data security in the cloud is a shared responsibility[4].

4. RELATED WORK

The security of big data is a well-known topic in today's world, but no ground-breaking work appears to have been done to include security concerns in cloud computing software development. This paper proposed by **Lai, R H Deng, C. Guan, and J. Weng et al.** mixes the three cryptographic techniques including KP-ABE, PRE, and lazy re-encryption associates each data file with a set of attributes, and assigns each user an access structure that is defined over these attributes. To enforce the access control, KP-ABE escorts data encryption keys of data files. It enables immediate fine-grain of access control of their data files with minimal overhead in terms of computation effort. **Sudhansu Ranjan, Lenka et. al.,** presented the RSA encryption and the digital signature method were combined with all cloud computing characteristics, including PaaS, SaaS, and IaaS. The three-way security offered by this technique includes data security, authentication, and verification. They suggested the RSA encryption method for data secrecy and the MD 5 algorithm for authentication in this work. **L. Tawalbeh, et al.,** Encrypt the data by applying TLS, AES, and SHA security mechanisms based on the type of classified data. It established the idea of manual classification of data, not automatic classification. **Liu et al.,** present a privacy-preserving multi-factor authentication system in which user passwords are the first factor and collected hybrid user behavior profiles serve as a second authentication factor and it is protected from the authenticated server. They used fully homomorphic encryption (FHE) and a fuzzy hashing (FH) technique to preserve the user's personal information is not leaked to the server or third party. Secure smart card authentication and authorization framework utilizing the multimedia cloud was introduced by **Yang et al**. They suggested employing the Diffie-Hellman algorithm and a two-factor authentication system with an open ID to gain access to material stored in a multimedia cloud**.** The work of **T. Xia et al.** additionally includes two different use-case scenarios to authenticate the relevance of the meta-model. The IaaS cloud deployment architecture suggested in this paper, nonetheless, somehow doesn't consider BD-specific security issues into consideration in any of the previous findings. Big data systems' necessary privacy and security measurements, definitions, standards, and features are addressed by the recently formed NIST BD security Sub-Working Group. Based on the NIST big data compatibility architecture, **Y. Demchenko, C. De Laat, et al.** thoroughly demonstrate key elements of the big data ecosystem. Their big data architecture, big data analytics, data structures and topologies, big data lifecycle management, and big data security make up their architecture framework. They do not, look at the deployment within an IaaS cloud or the security measures, features, or requirements specific to clouds, as this study indicates. **Valentina Casola et al.** introduced automatic security-by-design for cloud applications relating to security service level agreements (SLA). To evaluate security requirements, their approach takes a strategic analytical method and a fully automated information security procedure. To emphasize the need for increased security through architecture, **D. Polverini et al.** investigated data privacy and security in the information and communication system. Meanwhile, the definition of a security-by-design building strategy for the development of multi-cloud systems was advised by **V. Casola, A. De Benedictis, et al.** Overall, neither of the research mentioned previously seriously believed in considering the large data security standards, which affect the security deployments during the construction process.

5. BIG DATA CLOUD SECURITY REFERENCE ARCHITECTURE

Classical security arrangements and techniques can be used to mitigate several security threats related to big data security in the cloud. By introducing the Big Data Security Cloud Reference Architecture, it is important to recognize modules to address security service expectations for Big data and storage systems. The framework additionally outlines a variety of implementation and integration requirements and features that can be exploited to construct a secured Big data security in the Cloud system. National Institute of Standards and Technology (NIST's) study on reference architecture concentrates on big data processing in the cloud in a generalized way by identifying the architecture for whichever is specific to the big data system.

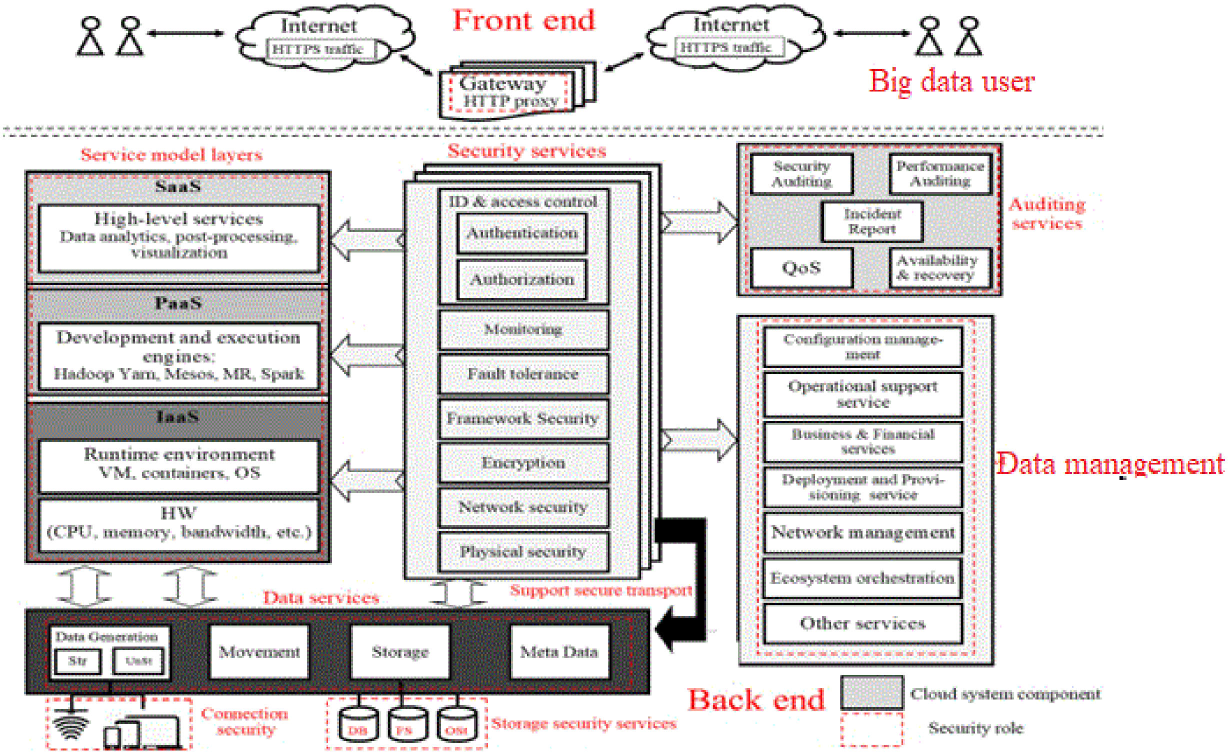

Figure 1 Big Data Security Cloud Reference Architecture

Service Model: This is an explanation of the three different cloud delivery models (IaaS, PaaS, and SaaS). Both the users and the cloud service provider would cooperate within those responsibilities. In an *IaaS model*, computing resources, such as CPU, RAM, and bandwidth, provide the model's foundation. The characteristics are then separated into hardware, network, and virtual machine. The containers are used by big data systems as virtual machines for their programs. The client-cloud provider connection is described by the *PaaS model*. The user is allowed to use the resources, applications, processes, and methods being developed by the providers to cloud applications. Increased features like refinement procedures and visualization tools are offered by the *SaaS model*. The user has no management or control over the runtime, development tools, or cloud infrastructure. As a result, the service provider is solely accountable for efficiency and security. An organization formed up of numerous users is granted privileged access to private cloud infrastructures. The organization, a recognized third party, or a blend of both can own, govern, and operate it. The infrastructure may be located on-site at the organization or off-site with the cloud service provider. Public cloud infrastructure is given access for utilization by anyone. One or more organizations, an authorized third party, or a combination of these entities must own, manage, and operate it. Off-site is really where the infrastructure is located. The deployment models that make up hybrid cloud architecture might be private, public, or a combination of both. *Multi-Cloud* is going to operate in a multi-cloud scenario where they use one or more PaaS and one or more IaaS services in addition to one or more SaaS services. To maximize their use of multi-cloud settings while maintaining tactical awareness and adequate security standards within every CSP, they function within, organizations adopting multi-cloud environments should formulate a plan[6].

Data Management: By doing this data management, the service delivery model is maintained. These are frequently connected services that the provider uses rather than the client. Data management services are an additional process that controls the interaction and process between the user and the cloud service provider. The service provider is in charge of upholding managed services with greater security in a sustainable manner.

Auditing services: The evaluation of cloud implementation's operational, reliability, and security assessment. Furthermore, it guarantees mitigation plans, quality of service, and reliability. Security measures are established and evaluated along through vulnerability analysis. Although maintaining the notification of cybersecurity threats, it evaluates treatment plans and the quality of security services.

Data services: A hypervisor and containers are being used by the data service to create virtual storage devices for use as on-demand storage, or it might have direct access to actual storage. All data processes including data in motion, data in use, and data in rest inside storage devices, datasets, and object storage will be included in the tasks to these services. Moreover, these providers offer data processing and activities, such as storing meta-data, as well as data migration from storage to virtual machines.

Security services: It describes many different technologies, regulations, and controls used to safeguard cloud computing's data, services, applications, and important infrastructure components. The reference architecture demonstrates that security is a critical concern that affects every part of the model [7].

6. PROPOSED NOVEL FRAMEWORK

All potential problems with cloud computing should be resolved by an effective paradigm for cloud data security. By protecting big data from all hazards involved and offering a cloud system that is safer and more effective, we hope to give the advantages of cloud computing without any obstacles. The security threats outlined below, as described by CSA, will be addressed by our suggested approach. Some security techniques address the issues and dangers stated above. It contains user interface security, an incident response strategy built into the application programming interface, centralized encryption and key management system (CKMS), intrusion detection system, and intrusion prevention system.

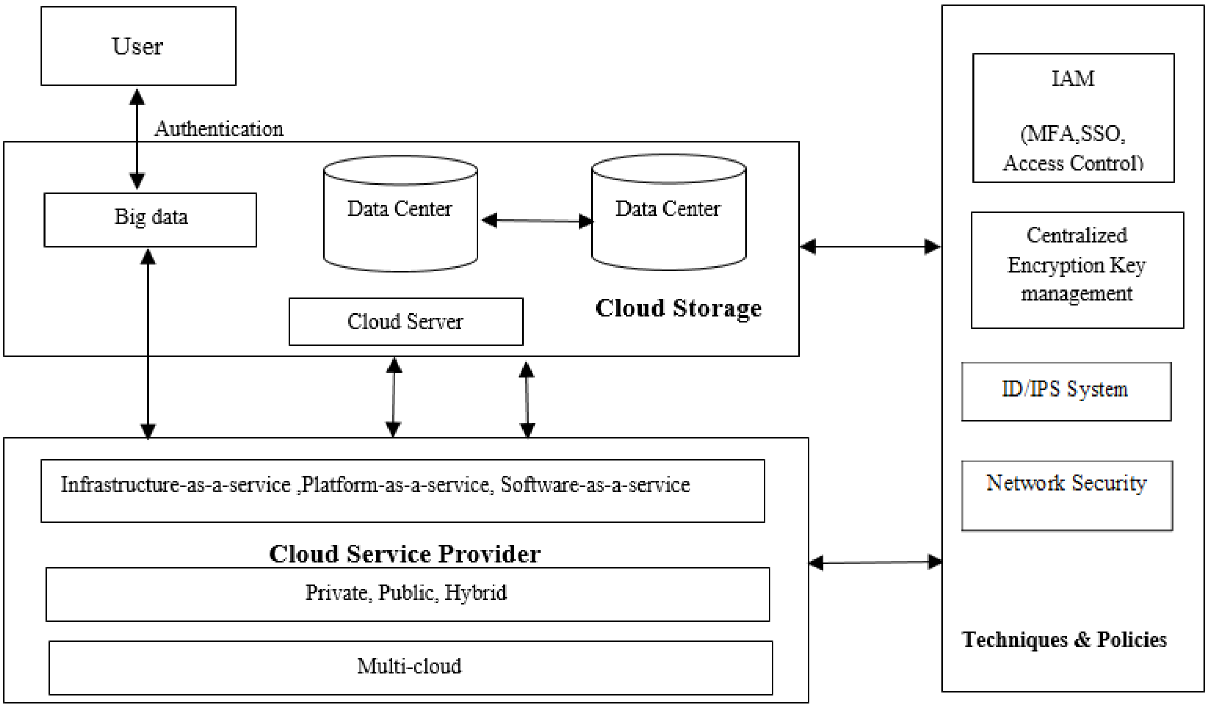

Figure 2 Proposed Framework Big Data Security in the Cloud Computing

In this framework, fig.2 (BDO), (CSP), and (CS) are the component of our framework that are works through an approach service level agreement (SLA). *Big Data Owner (BDO)* can be a user or organization or a combination of both. In the cloud system, the major security threat for Big data owners does not have control over their data once it has been stored in the cloud. Effective identity and access management policy in the cloud enables the proper access to the services. *Cloud Service Provider(CSP):* It gives the services like (IaaS, PaaS, and SaaS) and deployment models like (Public, private, and hybrid,) to the big data owners. *Cloud storage* is a virtual platform that provides organizations with scalable storage resources that can be provisioned rapidly as needed by the organization. It uses a highly virtualized, multi-tenant infrastructure. A cloud server is a virtual server that runs in a cloud computing environment rather than a physical server. The contract between the big data owner or user of the cloud services and the *service level agreement* (SLA) is known as this. A cloud server is a virtual server that runs in a cloud computing environment rather than a physical server. During the migration, users are required to authenticate through security identity and access management and access control policies to proper service access to the cloud[8].

1. *Identity and Access Management:* Data breaches can also be caused by human error or vulnerabilities in the control system. Incorporating Identity and Access Management (IAM) techniques into data handling makes ensures that only authorized users have access to the necessary resources. In the deployment phase, it first registrations and validates access privileges before being used for managing, identifying, and authenticating users who need access to applications, systems, or networks. In three separate situations at rest, in use, and in-transit data can be encrypted. At rest, encryption is used for data saved on resources. Encryption is used to protect data that is currently being created, changed, or viewed. Encryption is used to mitigate risk while it is in transit between two places. In fig.3 IAM facilitates single sign-on, authentication, authorization, and access control to centralize the login process for all business-related applications. With SSO, supervisors can quickly deploy and counter users, set permissions, and regulate which applications their users might access. A significant extra level of security exceeding conventional username and password are easily shared or affected delivered by multi-factor authentication. MFA comprises physical keys and gives access to login or one-time credentials generated by applications on systems. The ability to establish privileges at a greater increase slightly with service quality. An administrator verifies and permits access privileges in the system. Users identify and authenticate themselves during the transaction. The cloud is a crucial resource for identity management when it comes to data storage. Since the majority of small and medium-sized businesses lack the space for an on-site server rack, outsourcing that work to an identity access management provider reduces both operational expenses and security risks. Administrators can configure permissions for access to the identity management software using access control tools based on the kind and extent of access. In the configuration, an administrator registers and authorizes access rights. In the operation, users identify and authenticate themselves, and the access to applications is based on previously authorized access rights.
2. *Centralized Key Management System (CKMS):* Organizations can keep secure control of their systems and data because of the explosive growth in the use and volume of encryption keys in the digital age. The initial distributed key management systems quickly found themselves unable to meet demand. In addition to the management of various encryption methods across platforms, the effort to maintain numerous keys needed to maintain it quickly became unsustainable. The management of keys can be centralized to address each of these issues. Management is made simpler by having all required gear and software in one location. Additionally, as additional locations are added, they are incorporated into the KMS rather than having new infrastructure built there. It is possible to reduce implementation and upkeep costs [9].
3. *Intrusion detection systems (IDS) and intrusion prevention systems (IPS)*: It continuously monitors networks, detect possible threats, logs information regarding them, puts a stop to the occurrences, and notifies security administrators. IDS/IPS is used by many networks to recognize vulnerability management concerns and prohibit users from breaching security protocols. Because it may intercept attackers while they are obtaining network intelligence, it has become a vital component of the security infrastructure of the majority of organizations.

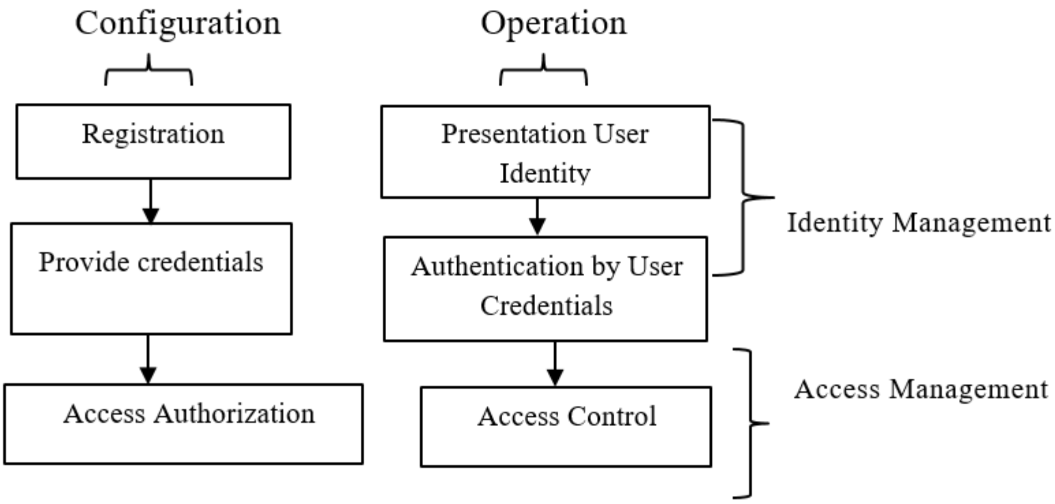

Fig.3 Identity and Access Management

4. *Reliability and performance:* The stability and availability of cloud computing are key indicators of its power. Resources must be provided continuously, without data loss, and depending on the failure rate. When a user switches to the cloud infrastructure, the cloud must offer increased performance. Applications operating on the cloud system's capabilities are typically used to gauge performance. Security addresses threat avoidance, whereas performance evaluates the speed of data implementation and interchange. Both are reliant on the user and service provider as well as the hardware and software components of the system. To improve the performance and reliability various threats and challenges of cloud computing have been triggered by the techniques in this paper[10].

5.

7. DISCUSSIONS AND RESULTS

In this research paper, we considered techniques for protecting huge datasets that enhance cloud computing services. A strong security framework has been designed to protect the data stored in the cloud. The data is protected against hostile users by using centralized encryption mechanisms relevant to the sensitivity and relevance of the data. Multifactor authentication, which is part of access control, protects sensitive data from unauthorized access. Every authorized and authorized access can be evaluated to find potential threats, and control actions can be made based on the threats that are found by the security measures to comply with the organization's changing data sensitivity needs. A personal intrusion detection and prevention technique is another name for it. It gathers data on an activity, assesses dangerous conduct, and reports that independent actions must be taken to halt or avoid it. By making it more difficult to identify the encryption mechanism, data encryption reduces the difficulty of detecting the statement's content. When considering adopting cloud computing, it's imperative to conduct a thorough investigation and have clear channels of communication with the cloud provider to establish the immediate requirements of operating a company effectively in the cloud. We eliminate numerous potential barriers to cloud computing with our framework. We will be able to give users of cloud computing the safety and security they have to make use of all the features and advantages of cloud computing when we safeguard the cloud environment from these vulnerabilities and attacks.

8. CONCLUSION AND FUTURE SCOPE

As the growth of digital services tends to grow around the world, cloud servers face significant risks of internal and external data leakage and data breach. The paper creates a secure environment for big data exchanging, storing, and transmitting using various identity and access management strategies, including multi-factor authentication, encryption for authentication, and intrusion detection and prevention. Overall, it can be determined that the security architecture enabled is flexible and able to adapt to the often cloud environment by providing a practical means to defend the data residing in the cloud servers. To keep the data safe and secure in the cloud, we proposed the framework and architectures be continuously improved. For upcoming work, we recommend incorporating cutting-edge technologies like the Internet of Things (IoT) and blockchain-based implementation into a comprehensive big data security scenario. Owing to the widespread cybersecurity assaults, a new modern method of access control in cloud computing should be provided. We also need to give more attention to the blockchain, which is extremely significant for cloud computing, and access control would perform with this technique when an event between the cloud and the user is monitored.

9. REFERENCES

[1] S. Kaisler et al. "Big Data: Issues and Challenges Moving Forward". 46th Hawaii International Conference on. IEEE, Jan. 2013, pp. 995– 1004.

[2] H. Miller, et.al., "Big-data in cloud computing: A taxonomy of risks", Information Research, (2013).

[3] Hashem, I.A.T., et al., 2014. The rise of "big data" on cloud computing: Review and open research issues. Information Systems, 47, pp. 98–115.

[4] CSA, "Top Threats to Cloud Computing, Tech. Rep. V1.0, Cloud Security Alliance 2010," https://cloudsecurityalliance.org/topthreats/csathreats.v1.0. pdf.

[5] Cloud Security Alliance, "The Treacherous 12 Cloud Computing Top Threats in 2016," Security, no. February, pp.1–34, 2016.

[6] NIST Big Data Working Group (NBD-WG) Oct. 14, 2020. http://bigdatawg.nist.gov/

[7] Yang, P.; Xiong, N.; Ren, J. Data security and privacy protection for cloud storage: A survey. IEEE Access 2020, 8, 131723–131740.

[8] Gupta, M., Awaysheh, F.M., Benson, J., Alazab, M., Patwa, F., Sandhu, R., "An Attribute-Based Access Control for Cloud-Enabled Industrial Smart Vehicles". 2020, 17, 4288–4297.

[9] Amr M. Sauber,Passent M. El-Kafrawy, Amr F. Shawish , Mohamed A. Amin,and Ismail,"New Secure Model for Data Protection over Cloud Computing", September 2021.

[10] Khalil Ahmad, Maher Ali Khemakhem, Abdullah Ahamad, et al.," A Proposed Framework for Secure Data Storage in a Big Data Environment Based on Blockchain and Mobile Agent", Published: 21 October 2021.