International Journal for Modern Trends in Science and Technology
Volume 9, Issue 08, pages 41-45
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol9issue09.html
DOI: https://doi.org/10.46501/IJMTST0909008

A Study on Data Protection and Privacy in Cloud-Based Big Data Environments

Aggala Chiranjeevi[1] | Dr. Prasadu Peddi[2] | Dr. Suneel pappala[3]

[1]Research Scholar, Dept of Computer science & Engineering, Shri JJT University-Rajasthan.
[2]Dept of Computer science & Engineering, Shri JJT University- Rajasthan
[3]Associate Professor. Dept of Computer science & Engineering, Lords Institute of Engineering And Technology, Hyderabad.

To Cite this Article
Aggala Chiranjeevi, Dr. Prasadu Peddi and Dr. Suneel pappala. A Study on Data Protection and Privacy in Cloud-Based Big Data Environments, 2023, 9(09), pages. 41-45. https://doi.org/10.46501/IJMTST0909008

Article Info
Received: 11 August 2023; Accepted: 28 August 2023; Published: 10 September 2023.

ABSTRACT

In the era of digital transformation and the explosive growth of data, organizations are increasingly relying on cloud-based big data environments to store, process, and analyze vast datasets. While these environments offer unparalleled scalability and performance, they also raise significant concerns regarding data protection and privacy. This study presents a comprehensive examination of the challenges and solutions related to safeguarding data in cloud-based big data environments, with a specific focus on privacy preservation. The research begins by analyzing the unique characteristics of cloud-based big data environments, highlighting their inherent vulnerabilities and potential threats to data privacy. We explore the risks associated with data leakage, unauthorized access, and the potential for breaches in the cloud. Additionally, we discuss the regulatory landscape surrounding data protection and privacy, including GDPR, CCPA, and other global data protection laws, and their implications for organizations operating in cloud-based big data ecosystems.

Keywords: cloud-based big data environments, GDPR, CCPA.

1. INTRODUCTION

During recent years, data production rate has been growing exponentially. Many organizations demand efficient solutions to store and analyze these big amount data that are preliminary generated from various sources such as high throughput instruments, sensors or connected devices. For this purpose, big data technologies can utilize cloud computing to provide significant benefits, such as the availability of automated tools to assemble, connect, configure and reconfigure virtualized resources on demand. These make it much easier to meet organizational goals as organizations can easily deploy cloud services. This shift in paradigm that accompanies the adoption of cloud computing is increasingly giving rise to security and privacy considerations relating to facets of cloud computing such as multi-tenancy, trust, loss of control and accountability. Consequently, cloud platforms that handle big data that contain sensitive information are required to deploy technical measures and organizational safeguards to avoid data protection

breakdowns that might result in enormous and costly damages.

2. LITERATURE REVIEW

Renu yadav (2022) Big data is the generated data from different sources sensors, digitizers, scanners, mobile phones, the internet, video, e-mail, and social media. Big data involves appropriate processing capability and analytical capabilities, but conventional methods are unlikely to meet essential expectations. Cloud service providers collaborate on a network that includes computing resources, servers, storage, applications, and a variety of other services. Cloud is a platform for cloud providers to provide services to users and manage data in the data center. With all of these advantages of cloud computing, the protection of huge data remains a serious concern, which also has an adverse influence on cloud migration. Businesses seek robust cyber security methods to protect big data against threats and intrusions while shifting data to the cloud. However, several researchers in the previous study used a variety of security approaches to assure optimum data protection and found that they were effective.

Jahoon Koo et al (2020) The use of big data in various fields has led to a rapid increase in a wide variety of data resources, and various data analysis technologies such as standardized data mining and statistical analysis techniques are accelerating the continuous expansion of the big data market. An important characteristic of big data is that data from various sources have life cycles from collection to destruction, and new information can be derived through analysis, combination, and utilization. However, each phase of the life cycle presents data security and reliability issues, making the protection of personally identifiable information a critical objective. In particular, user tendencies can be analyzed using various big data analytics, and this information leads to the invasion of personal privacy.

Cloud computing

Cloud computing can be considered as a new computing arche type that can provide services on demand at a minimal cost. The three well-known and commonly used service models in the cloud paradigm are software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS). In SaaS, software with the related data is deployed by a cloud service provider, and users can use it through the web browsers. In PaaS, a service provider facilitates services to the users with a set of software programs that can solve the specific tasks. In IaaS, the cloud service provider facilitates services to the users with virtual machines and storage to improve their business capabilities.

Cloud computing is closely related to but not the same as grid computing. Grid computing integrates diverse resources together and controls the resources with the unified operating systems to provide high performance computing services, while cloud computing combines the computing and storage resources controlled by different operating systems to provide services such as large-scaled data storage and high-performance computing to users.

Risk assessment of Energy Big Data in Cloud Environment

In the process of risk assessment, the probability of risk occurrence, loss range, and other factors need to be considered comprehensively to get the possibility and degree of system risk occurrence, determine the risk level, and then decide whether to take corresponding control measures and to what extent. Therefore, the construction of risk assessment index system should follow the principles of comprehensiveness, scientificity, representativeness, and practicability, select the representative risk elements from a scientific perspective, quantify the risk based on the practical principle, and strive to show the risk management level comprehensively and accurately.

Identification of Risk Factors

Data security management is the most prominent risk faced by big data application. Although the massive data is stored centrally, it is convenient for data analysis and processing, but the loss and damage of big data caused by improper security management will cause devastating disaster. Due to the development of new technology and new business, the infringement of privacy right is not limited to physical and compulsory invasion, but is derived in a subtler way through various data, and the data security and privacy risks caused by this will be more serious. Compared with the previous Internet and computer technology, the application advantage of big data in the cloud environment is more obvious. Big data platform has strong sharing ability, which can manage the security of information use and improve the efficiency of resource utilization. The construction of cloud platform and system application

have strict standards. Cloud computing technology provides more comprehensive technical support and makes privacy management more reasonable, which is consistent with the level of technology development in the new era. But from another point of view, it is under the influence of cloud platform sharing features that part of the data information is easy to be exposed, which provides opportunities for some illegal intrusion. Therefore, we must pay full attention to its risks.

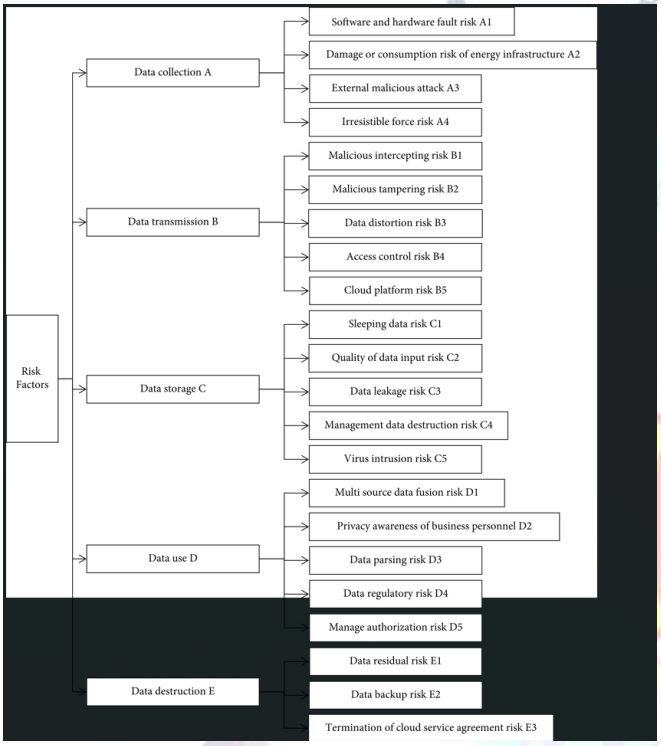


Figure: Security and privacy risk assessment index system based on the whole life cycle of energy big data

3. PRIVACY PRESERVING OF CLOUD APPLICATIONS

Due to the low cost and autonomy of source sharing, it has the potential to occur on a huge scale. In the cloud, customers rely heavily on the resources provided by the company or its third-party suppliers. Both downloadable programmers and web-based services are considered tools for the IT industry. Providers in the cloud may access client data stored in remote data centers. If a user has access to the internet and is utilizing the cloud, they may access their data and do calculations from anywhere and on any device. It provides these advantages. Cloud computing offers a compromise between data isolation and privacy concerns, but the concept of storing data on remote servers is not new. On this study, we provide a safe method of storing data on the cloud. In this study, we examine a specific sort of utility computing known as the "programmer as a tool" (or "inch") paradigm. If you're worried about security and have questions such, "a company must trust businesses and client users, network administrators?" this setup may put your mind at ease. Both personal computers and the company's Intranet are heavily used by staff. Programmers may need to make last-minute requests for resources and use data. Putting private information on the cloud opens the door to security risks.

This storage provider went out of business after an equipment administrator made a mistake that exposed the private information of almost 45 clients. Despite these benefits, many companies and people are hesitant to move their data to the cloud due to privacy and security issues with cloud computing resources such as databases, networks, operating systems, source tracing, load balancing, transactions, and memory management. Some common-sense precautions include the following: Setting a user's permissions in an access control system.

SECURING BIG DATA

The data outsourcing process involves a large number of people. Uncommon knowledge, nevertheless, has been shown to not exist at your demonstrated develop and lessen endemic levels. For that, it's necessary to consistently draw on other sources. We have to shorten a vital piece of advice since it doesn't need it anymore because it has all the knowledge it needs. Everything you send will be stored. There are a lot of people involved, so it might get struck many times if it doesn't arrive at its destination right away. Given the critical nature of the information to the operation of businesses, an insider assault or predator strike might have severe repercussions. That's why it's so important to have a reliable cloud storage platform that doesn't need any unique skills from data owners.

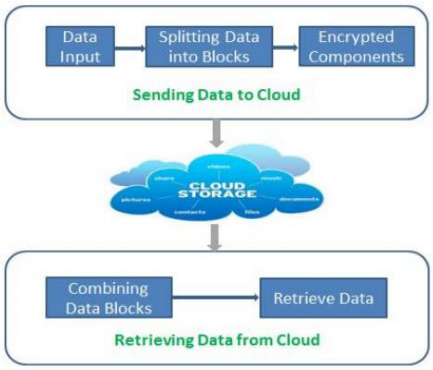


Figure: An outline for protected remote storage

4. PROPOSED FRAMEWORK

The planned structure is described in further detail in this section so that those in need of human recovery support may make the most of it. However, this even timing of the facts may be used to verify the client's operation. The data operator, who is often a company that generates massive volumes of data, generates and uploads the data to the consumers' cloud storage.

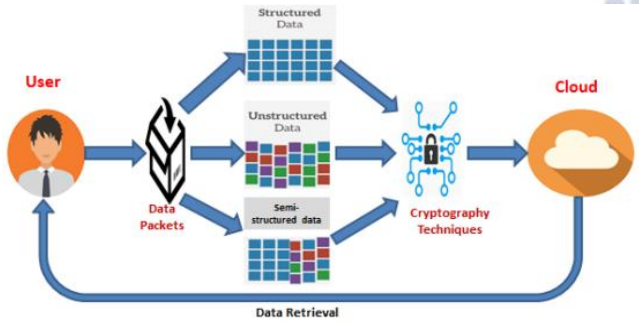
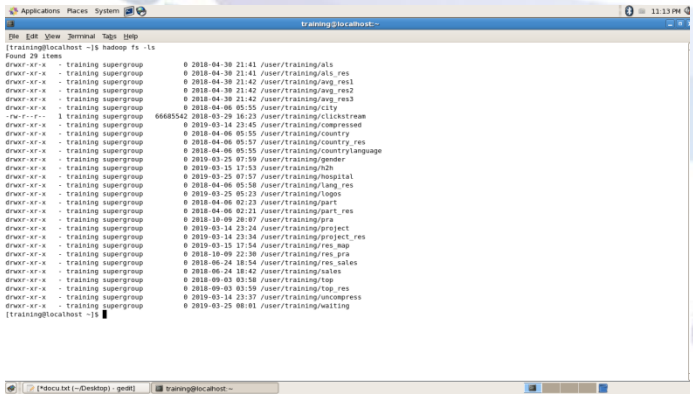


Figure: Secure blur storage and recovery using future structure

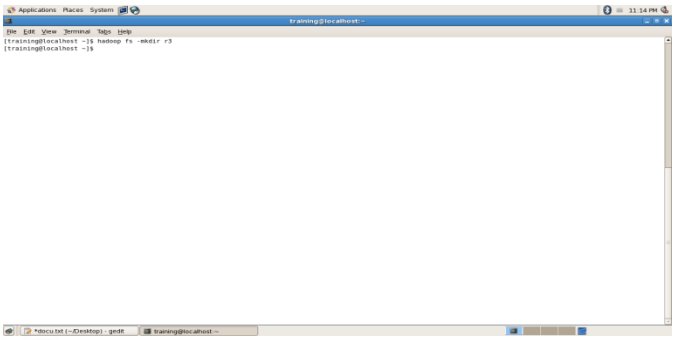
Finding an innate strategy that will allow you to successfully finish a real even version made to address the use problem is your current mission. The usefulness of the frame has been much enhanced in this updated version. Consequently, the suggested design has a means of connecting to a preexisting public cloud.

5. EXPERIMENTAL RESULTS

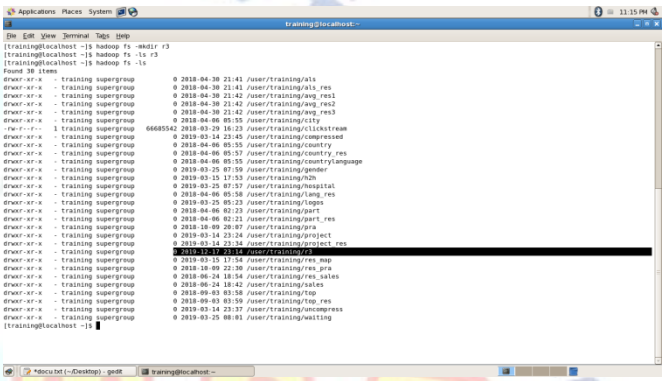
Automated sign-showing findings demonstrated the viability of the framework designed for installing, storing, downloading, and disclosing data from existing systems. The tests were developed using Amazon's EC2 and S3 cloud computing and storage services.



The above window provides a summary of the directories that may be inspected on the hadoop platform



A directory has been twisted and is been created in hadoop platform for the work to carry out and examine the data.



The directories are catalogued, and the task at hand is to determine which ones exist inside the hadoop infrastructure.

Cloud services are the way to go when dealing with massive volumes of data. Sure, there are a few clouds in the sky. Using these techniques might cut down on the time spent on these cycles of labor, saving money in the process. Connecting to a communal cloud from far away adds to the data loss that ultimately leads to an overload.

Security Monitoring and Incident Response:

For maintaining a secure environment continuous monitoring of cloud infrastructure and timely response to security incidents are critical. The proposed system incorporates robust security monitoring tools, such as intrusion detection systems (IDS) and security information and event management (SIEM) solutions. These tools analyze system logs, network traffic, and behavior patterns to identify potential security threats and anomalies. These system triggers automated alerts and initiates incident response procedures to mitigate risks and ensure prompt remediation by detecting suspicious activities.

Compliance and Regulatory Considerations: In the proposed system the regulatory consideration and industry-specific compliance requirements is very critical for cloud security. This system incorporates

features to facilitate compliance, such as data anonymization techniques, data retention policies, and secure data erasure methods. The ensures in protection of sensitive data while meeting legal obligations is well adopted with aligning general data protection regulation (GDPR). The proposed system presents a comprehensive approach to enhance data privacy and security in cloud computing environments. By integrating robust encryption, access controls, auditing capabilities, and secure data transfer mechanisms, the system addresses the challenges associated with data privacy and security in the cloud. This system enables organizations and individuals to leverage cloud computing services with confidence, knowing that their sensitive data is protected against unauthorized access and data breaches. Embracing such a system will foster trust, accelerate cloud adoption, and pave the way for a secure and privacy-aware digital future.

6. CONCLUSION

Cloud-based big data environments introduce a host of complexities related to data protection and privacy. The scale, diversity, and rapid growth of data make it challenging to maintain control and visibility over sensitive information. Organizations must recognize these complexities and invest in robust strategies and technologies. Compliance with data protection regulations is non-negotiable. Organizations must stay informed about evolving laws and regulations, such as GDPR, CCPA, and other global data protection laws, and adapt their practices accordingly. Failure to comply can result in severe penalties and reputational damage. Effective data governance is a cornerstone of data protection and privacy in the cloud. Establishing clear data ownership, classification, and access control policies is essential. Employing encryption for data at rest and in transit is a fundamental security measure. Additionally, robust access control mechanisms should be implemented to ensure that only authorized personnel can access sensitive data. Regular audits of access logs are crucial for monitoring and enforcement.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Jahoon Koo et al (2020) Security and Privacy in Big Data Life Cycle: A Survey and Open Challenges, Sustainability, 12(24), 10571; https://doi.org/10.3390/su122410571.

[2] Renu yadav (2022) A Comparative Study On Different Techniques Used To Secure Big Data In A Cloud Environment, International Journal of Creative Research Thoughts, Volume 10, Issue 9, ISSN: 2320-2882.

[3] Ankit Anand, J Lakshmi and S K Nandy (2013), "Virtual Machine Placement Optimization Supporting performance SLAs",

[4] Aleksandar Donevski, SaskoRistov, and Marjan Gusev (2013), "Security Assessment of Virtual Machines in Open-Source Clouds".

[5] B.Yamini and D VetriSelvi (2011), "Cloud Virtualization: A Potential way to reduce Global Warming", ISSN: 978-1-4244-9182.

[6] Buddhika Lakmal Warusawithana and Mithila Mendis (2013), "Next Generation Multi-Tenant Virtualization Cloud Computing Platform".

[7] Boyang Wang, Hui Li, Liu, Fenghua Li & Xiaoqing Li, (2014) "Efficient Public Verification on the Integrity of Multi-Owner Data in the Cloud", vol. 16, no. 6, pp. 592-599.

[8] Cheng-Kang, C, Sherman, SM, Wen-Guey, T, Jianying, Z, Robert, H & Deng (2014), "Key-Aggregate cryptosystem for scalable data sharing in cloud storage", vol. 25, no. 2, pp. 468-477.