International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 07 | July - 2021 ISSN: 2582-3930

# Data Security and Privacy Protection in Cloud Data Warehouse

**Chikkakrishnappa[1], Harshitha S[2], Poojitha S[3], Priyanka K[4], Rajeshwari G D[5],**

[1]*Assistant Professor, Dept. of Computer Science and Engineering, Bangalore*
[2,3,4,5] *Students, Dept. of Computer Science and Engineering, Bangalore, Karnataka, India*

---

**Abstract -**In any of cloud deployment models, security and privacy of data stored in cloud database has become challenging issue. Data theft in cloud server by malicious insiders or any of intruders has become one of the most common issues. Intruders or malicious insiders might use authorized access users credentials to access the data. Data or documents that are uploaded to the server will be encrypted which makes data secured but it might not be possible to encrypt all the data in cloud database. In this paper, the proposed scheme will ensures the safety of the data which are important to the organizations development where it prevents any outsiders from accessing the data. According to the proposed scheme, users can only access the data only if they provide specific keyword in accordance with data. Otherwise they won't be able to access the file and also they can decrypt the file only if admin or data owner accepts their request for accessing the data. A proposed scheme ensures the data security and makes sure that each and every users have authorized access.

***Key Words*: Cloud Computing, AES Algorithm, Privacy and Data Security in cloud.**

## 1. INTRODUCTION

Cloud Computing is now evolving as one of the on-demanding technology where there will be the availability of computer system resources, data storage, and computing power, without active management by the end-user. Cloud Computing promotes the concept of leasing remote resources rather than using the hardware's which frees customers from maintenance costs. Since the services provided via the internet, Cloud customers can access their data anytime and anywhere in the world. Services offered to the customers/users are relatively low in cost. Therefore, many organizations, enterprises are moving from physical data centers to virtual data centers because of the advantages of cloud computing.

In Cloud computing, there are different types of deployment models such as public cloud, private cloud, and hybrid cloud. Many enterprises use the private cloud to store their information because it is more secure compared to the public cloud and also there will not be any outside threats. Private Clouds can also be referred to as a corporate cloud and it also provides high computing services to selected cloud customers. It also provides a high level of security and privacy to data through firewalls and internal hosting and also ensures that operational and sensitive data are not accessible to third-party providers.

Despite many advantages, there is some challenging issue which limits the cloud customers from using cloud services that is security and privacy of data that are stored in the cloud data warehouse. Data can either be stored in the public, private, community, or hybrid cloud. Also, many organizations use private clouds to store their information because of security reasons. In private clouds, organizations will be responsible for the security of the data. Even though Cloud service providers provide secure data storage and data transfer from one system to another but there might be unauthorized users or malicious insiders who have ill intentions will try to access sensitive data of an organization.

In order to avoid those unauthorized users from accessing the data that is stored in a cloud database those stored data are encrypted when the file is uploaded to the cloud data storage. Advanced Encryption Standard is one of the most popular algorithms used for encrypting data. AES algorithm is based on a design principle which is known as a substitution-permutation network and is quite efficient in both software and hardware. Advanced Encryption Standard algorithm is chosen because of their High speed and low RAM requirements.

## 2. LITERATURE SURVEY

Cloud Computing is the technology that is now emerging has new evolution and also promises to have far-reaching effects on the networks of federal agencies and also other organizations. Cloud customers must understand the policies, procedures, and technical controls used by the cloud service providers because it is a prerequisite to assessing the security and privacy risks involved [1]. Database-as-a-service is a model which provides user services such as data creation, storage, modification, and retrieval from anywhere in the world as long as they have access to the Net. There are two privacy issues they are First, the data owner must be assured that the data stored on the cloud server is protected against data thefts from outsiders. Second, the data owners must be assured that whether data is protected from cloud services providers if the data owners cannot trust the cloud service providers [2]. Data

stored in the cloud servers are often encrypted by using any cryptographic tools. There are several benefits of using cryptographic storage like customers will usually have control over their own data and the security properties are derived from cryptography are all legal mechanisms, physical security, and also access control. Data that are stored in cloud servers are encrypted by using symmetric encryption scheme (for example, AES algorithm) [3]. A data theft threat is mostly because of a malicious insider. A malicious insider might be a current or former employee or contractor who has or had authorized access to the data that is stored in a cloud server and also they might intentionally exceed or misuse that access credentials in a manner that negatively affect the confidentiality, integrity, or any of the information related to the company [4]. Although cloud customers prefer using cloud services rather than maintaining physical data centers. There is also a challenging issue where each and every cloud customer are supposed to face i.e. information privacy. There might be malicious insiders who try to steal sensitive data which affects an organization's growth. To prevent intruders from accessing data stored in the server, the data owner must ensure that data is encrypted and must also ensure whether the user is authorized or not authorized. The proposed scheme ensures authorized users and also secures data decryption with permission given by the data owner.

## 3. SYSTEM ARCHITECTURE

The system architecture is a conceptual model that defines the structure and behavior of the system. It comprises the system components and the relationships describing how they work together to implement the overall system. Fig.1 below shows the system's architecture and the various functionalities of our system.

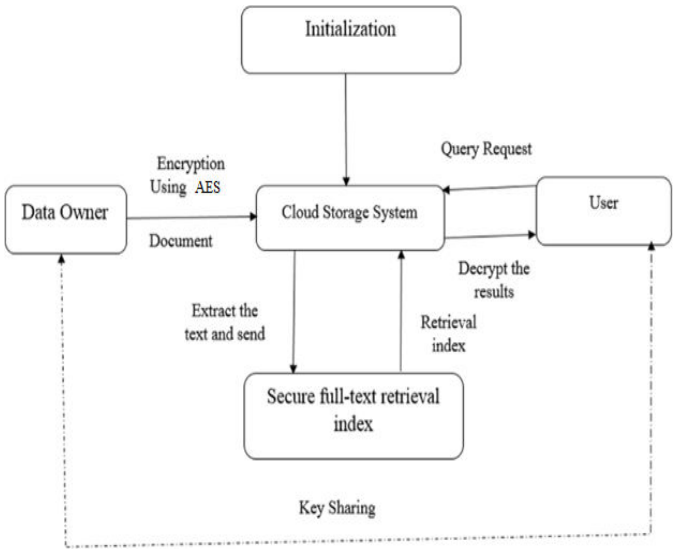


**Fig. 1** System Architecture

System architecture is as shown in Fig. 1. It represents how exactly operation is carried out. It includes operations such as login to the cloud server, uploading data or documents, encryption, and decryption, also the cloud database. According to the proposed scheme, the data owner will upload the data whereas the user will only have access to the documents by providing keywords and secret keys sent by the data owner. It consists of:

- **Initialization**: Data owner and/or user will log in to the cloud database for accessing the documents that are outsourced to cloud servers.
- **Data Owner**: Data owner is the one who will outsource their data to a cloud server. Security and Privacy of the data will be done by the data owner itself. Data encryption is done once the document is uploaded to the server data owner.
- **Cloud Storage System**: It is the storage or database system provided in a cloud server for data storage. The data owner will use the storage space for data storage.
- **User**: The user is the entity that will access the data that are stored in a cloud database. User will send the query request such as requesting for accessing documents etc. Also, file decryption is done by users themselves when they need to access the document.
- **Secure full-text retrieved index**: It is a full-text version of the original document after decrypting the file by providing the secret key.

## 4. IMPLEMENTATION

System Implementation is the structure that is created during architectural design and also the results of system analysis are used to construct system elements that are needed to meet the requirements of stakeholders and systems which is developed in the early phases of the life cycle. After constructing these system elements, then they are integrated to form intermediate aggregates, and finally, they complete the system-of-interest (SoI). Implementation is the process that actually yields the lowest-level system elements in the system hierarchy (system breakdown structure). System elements are either made, bought, or reused in which production involves the hardware fabrication processes of joining, forming, removing, and finishing, the software realization processes of coding and testing, or the operational procedures development processes for operators' roles.

Modular design is a design in which a system will further be divided into smaller parts called modules. We

can create the modules independently and it's used in different systems. A modular system can be characterized into reusable modules and discrete scalable; well-defined modular interfaces are used very carefully; and also industry standards are used for interfaces.

**Modules:**

- Data Owner
- User
- Security

**Modules Description:**

**A. Data Owner:**

The data owner is the super user in our project he is having certain rights with our application. Admin can able to perform the following tasks in our application:

- Log in
- Data owner can upload the files to the cloud.
- Data owner can view, edit and update multiple cloud datacenter details
- Data owner can view the user details.

**B. User:**

User is the regular user to this application user can perform following task with our application.

- If new user they should register with our application.
- To log in the application they need to register at once.
- If users log in they will search files using keywords, and to access the file user can send a request to the owner.

**C. Security:**

The AES algorithm is used to encrypt the data. Linear search and dept. first Search algorithms are used to search keywords i.e. to search the documents that are stored in a cloud database.

**1. AES Algorithm:**

AES is iterative and is based on a 'substitution–permutation network'. It consists of some series of linked operations, which involve replacing some of the inputs with specific outputs (which is known as substitutions) and others involve shuffling of bits around (which is known as permutations).

AES algorithm performs its computations on bytes rather than bits. Hence, AES considers the 128 bits of a plaintext block as 16 bytes, and these 16 bytes are arranged in the form of four columns and four rows for processing as a matrix. the number of rounds in AES is variable and depends on the length of the key when compared to DES. AES performs 10 rounds for 128-bit keys, 12 rounds for 192-bit keys, and 14 rounds for 256-bit keys. Every round performed by AES uses a different 128-bit round key, which is calculated from the original AES key.

The schematic representation of the AES structure is given in the following illustration.

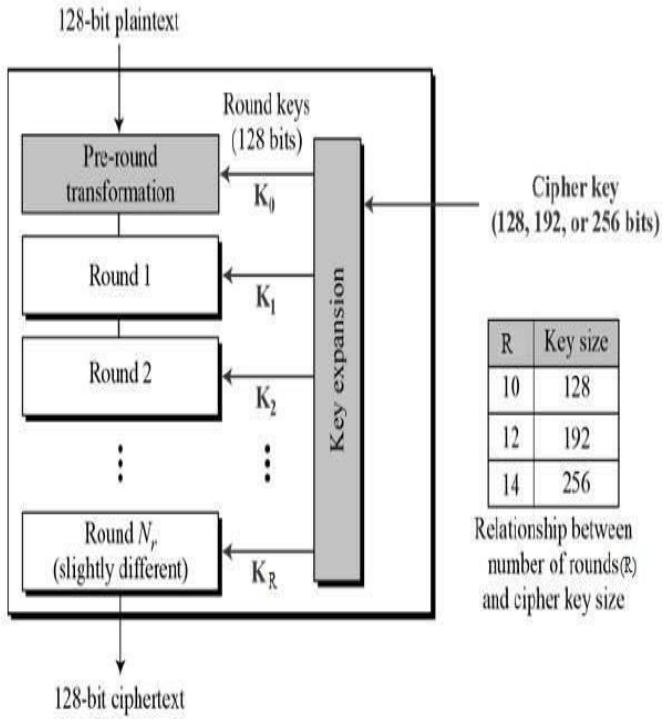


**Fig. 2** AES Algorithm

**2. Linear Search Algorithm**

A linear search is also called a sequential search. It is a method for finding an element within a list. This algorithm starts to checks each element present in the list sequentially until it finds its match or the whole list has been searched. A linear search runs in its worst linear time and makes almost n comparisons, where n is the length of the list. According to this algorithm, every element in the list will be searched so the average cases of the linear search will be (n+1)/2 comparisons, but the average cases can be affected if the search probabilities for each element vary. If the algorithm reaches the end of the list then the search terminates unsuccessfully.

### 3. Depth-first search Algorithm

Depth-first search is an algorithm for traversing or searching tree or graph data structures. The operation of the DFS algorithm starts at the root node and explores each branch in the decision tree before backtracking the same. So it basically starts from the root or any arbitrary node and marks the node and moves to the adjacent unmarked node. This loop will continue until there is no unmarked adjacent node. After that, it will backtrack and check for other unmarked nodes and traverse them. Finally, nodes will be printed.

### 5. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed a scheme in which the data uploaded by the data owner to the cloud server will be encrypted. Whenever the user needs to access the data then the user needs a security key generated to decrypt the required field. According to the proposed scheme

Whenever the user needs to access the data, the user needs to request access permission from the data owner/admin. Once the request accepted, the user will receive a confirmation mail from the admin along with the secret key. If the user enters any wrong credentials then the admin will be notified regarding the same. In this way, the proposed scheme is efficient in terms of data security because the permission will be provided directly from the data owner for accessing sensitive document.

The proposed scheme can also be further developed with features like sending a direct text message instead of the email message when there are any malicious insiders transferring document that is stored in cloud database from system to any of external devices such as USB drives etc.

### REFERENCES

[1] Guidelines on Security and Privacy in Public Cloud Computing. (2011). *Journal of E-Governance, 34(3), 149–151*. doi:10.3233/gov-2011-0269.

[2]“Executing SQL over Encrypted Data in the Database-Service-Provider”.

[3] Kamara S., Lauter K, (2010) Cryptographic Cloud Storage, ln: Sion R. et al. (eds) Financial Cryptography and Data Security. FC2010. Lecture notes in Computer Science, vol 6054, Springer, Berlin, and Heidelberg.

[4] Cloud Security Alliance “The Treacherous 12: Cloud Computing Top Threats.”

[5] S. Pearson and A. Benameur. “Privacy, security, and trust issues arising from cloud computing.” Proc. Cloud Computing and Science, pp. 693–702, 2010.

[6] NIST. “Challenging Security Requirements for US Government Cloud Computing Adoption “ https://ws680.nist.gov/publication,2012.