ISSN No. 0976-5697

Volume 5, No. 4, April 2014 (Special Issue)

International Journal of Advanced Research in Computer Science

REVIEW ARTICLE

Available Online at www.ijarcs.info

An Analysis of Data Privacy Techniques in Cloud Computing

Arvind S. Kapse
Dept. of CSE
P. R. Patil College of Engg. & Tech
Amravati, Maharashtra, India

V. M. Thakare
Dept.of CSE
Sant Gadge Baba Amravati University
Amravati, Maharashtra, India

Sudarson Jena
Dept. of IT
GITAM University Hyderabad, India

Abstract **-** Cloud Computing is one of the emerging technologies in Computer Science. Cloud provides various types of services to us. Database Outsourcing is a recent data management paradigm in which the data owner stores the confidential data at the third party service provider's site. The service provider is responsible for managing and administering the database and allows the data owner and clients to create, update, delete and access the database. There are chances of hampering the security of the data due to untrustworthiness of service provider. So, to secure the data which is outsourced to third party is a great challenge. The major requirements for achieving security in outsourced databases are confidentiality, privacy, integrity, availability. To achieve these requirements various data confidentiality mechanisms like fragmentation approach, High-Performance Anonymization Engine approach etc are available. In this paper, various mechanisms for implementing Data Confidentiality in cloud computing are analyzed along with their usefulness in a great detail.

Keywords: Cloud computing, Data Confidentiality, Integrity, Outsourced Databases, Performance

I. INTRODUCTION

Data confidentiality is one of the pressing challenges in the ongoing research in Cloud computing. as soon as confidentiality becomes a concern, data are encrypted before outsourcing to a service provider. Hosting confidential business data at a Cloud Service Provider (CSP) requires the transfer of control over the data to a semi-trusted external service provider. Existing solutions to protect the data mainly rely on cryptographic techniques. However, these cryptographic techniques add computational overhead, in particular when the data is distributed among multiple CSP servers[1].Storage as a Service is generally seen as a good alternative for a small or mid-sized business that lacks the capital budget and/or technical personnel to implement and maintain their own storage. But the main issue is to maintain CIA (Confidentiality, Integrity and Authentication) (Fig.1)to the data stored in the cloud.

Trusted Cloud provides you with the ability to create a unified data protection policy across all clouds.The impact of privacy requirements in the development of modern applications is increasing very quickly. Many commercial and legal regulations are driving the need to develop reliable solutions for protecting sensitive information whenever it is stored, processed, or communicated to external parties.

To this purpose, encryption techniques are currently used in many scenarios where data protection is required since they provide a layer of protection against the disclosure of personal information, which safeguards companies from the costs that may arise from exposing their data to privacy breaches. However, dealing with encrypted data may make query processing more expensive. For that the fragmentation procedure is applied to a relational databases where the tables are treated as independent fragments. This fragmentation and distribution approach reduces the trust expectancies towards the external service providers and thus improves privacy and confidentiality.

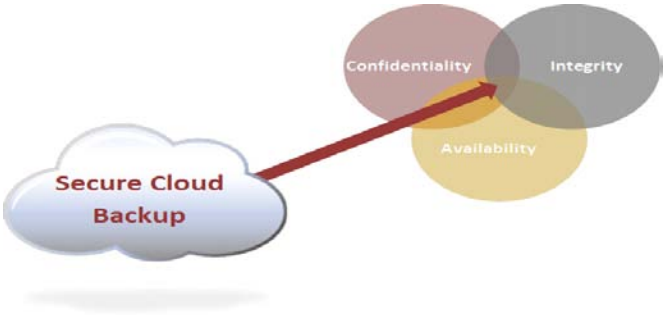

Figure. 1 cloud computing

II. BACKGROUND

Data storage is the central part of Cloud computing, which renders data confidentiality as one of the most critical issues in the Cloud[1].This section describes the concept of DBaaS and benefits, architecture of database outsourcing model challenges associated with the same.

A. Database As A service:

Database-as-a-Service (DBaaS) is a service that is managed by a cloud operator (public or private) that supports applications, without the application team assuming responsibility for traditional database administration functions. With a DBaaS, the application developers should not need to be database experts, nor should they have to hire a database administrator (DBA) to maintain the database. True DBaaS nirvana will be achieved when application developers can simply call a database service and it works without even having to consider the database. This would mean that the database would seamlessly scale and it would be maintained, upgraded, backed-up and handle server failure, all without impacting the developer in any way. From the developer's perspective, this is the definition of DBaaS.

CONFERENCE PAPER
Two day National Conference on Innovation and Advancement in Computing
Organized by: Department of IT, GITAM UNIVERSITY Hyderabad (A.P.) India
Schedule: 28-29 March 2014

The ultimate goal of a DBaaS is that the customer doesn't have to think about the database. Today, cloud users don't have to think about server instances, storage and networking, they just work. Virtualization enables clouds to provide these services to customers while automating much of the traditional pain of buying, installing, configuring and managing these capabilities. Now database virtualization is doing the same thing for the cloud database and it is being provided as Database as a Service (DBaaS).

B. Architecture of Outsourced Database Model:

a. Overview:

In the Outsourced Database Model (ODB), organizations outsource their data management needs to an external service provider. The service provider hosts client's databases and offers seamless mechanisms to create, store, update and access (query) their databases. This model introduces several research issues related to data security which we explore.

b. System Model:

(Fig.2)depicts the architecture of Outsourced Database Model. It consists of 3 main entities as (1) the data owner(s), (2) the database service provider (server) and (3) the client(s) (also referred to querier(s)). The data owner creates, modifies and deletes the contents of the database. The server hosts the owner's database, i.e., the owner outsources its database to the server. The clients issue queries about the database to the server. Some of the parameters identifying a specific ODB include the number of owners and clients and the type of trust in the server. Is the server trusted with the data contents but not with integrity? Or do we not trust the database administrators and therefore need to employ encryption to provide data privacy?

c. Database as a Service Provides Many Benefits:

As data continues to expand, and technology advances at an ever more rapid pace, Database as a Service (DBaaS) can provide enterprises with a database solution that is simple to use and easy to update. DBaaS provides developers with a cloud-based database through which scaling, load balancing, failover and backup can all be managed.

DBaaS is perfect for applications that require quick provisioning and dropping of databases, such as prototype testing, sales promotion and other short-term projects. In the past, soon after a database was created for one of these purposes, it might already be obsolete.

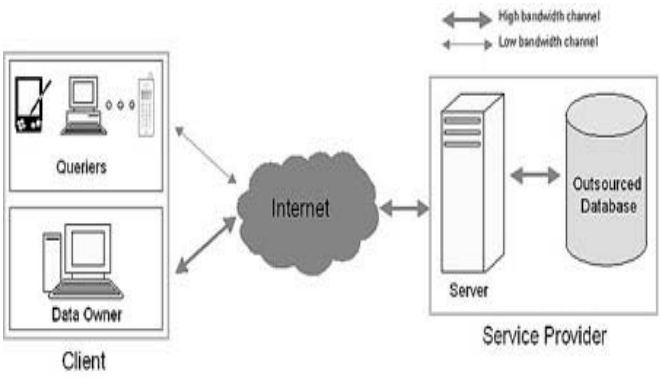

Figure. 2 Architecture of outsource database model

A web-based database works better for these types of scenarios because it is more adaptable and economical.

There are several reasons why DBaaS is gaining popularity and recognition across a variety of industries. The growth of data has complicated the management of information. Storing multiple databases on site is not only expensive and inefficient, but it can also be risky in terms of compliance or lost, misplaced, and mishandled data.

Another trend that is driving DBaaS adoption is the importance of increased agility and useable data. With applications that can help companies better understand customer behaviors and patterns, the need for organized and accessible data is more important than ever.

DBaaS works well for dynamic applications that require custom data and a tailored experience for different users. These apps need a more agile and cost-efficient platform than what a traditional, on-premise database can provide.

Many enterprises are realizing the need for a scalable database solution that will save money and manage data more effectively. With DBaaS you can stop thinking about and dealing with your database and just let it work for you.[4] Our technical support specialists will ensure that all aspects of handling your data are taken care of, from storage and backups, to tuning and security

III. OUR APPROACH

A. Difference Between Data Confidentiality And Data Security:

This paper concentrates on a particular aspect of database security, that is data confidentiality. Data confidentiality refers to the ability to share sensitive data among a community of users while respecting the privileges granted by the data owner to each member of the community. Any user external to the community is assumed to have no privilege at all. A special case of data confidentiality is data privacy. Data privacy means that the data owned by an individual will never be disclosed to anyone else. Privacy is easier to enforce than confidentiality since sharing is precluded. The simplest and most effective way to ensure data privacy is to encrypt the user's data thanks to a symmetric key algorithm (e.g., DES). The user being the unique holder of the cipher key, no one else can access the clear text form of the data. Several Storage Service Providers propose to manage encrypted backups for personal data.[11]They guarantee that data is encrypted at all times from transmission of a customer's computer to their server and back and remains safe from unauthorized access even by their staff. Data privacy solutions cover only a restricted range of applications considering that even private data is subject to sharing (e.g., patient's medical records are shared by doctors, customer's information is shared by e-commerce sites). Thus, the remainder of the paper focuses on the more general problem of data confidentiality and places much emphasis on access right management.

B. Data Confidentiality Requirements:

a. Confidential data must be managed by an auto administered DBMS to cast off the DBA privileges.
b. This DBMS must be hosted by an auto-administered computing system to cast off the system administrator Privileges.

c. This computing system must constitute a Secure Operating Environment (SOE)1 to cast off any Intruder action.

The traditional database server approach suffer from a strong handicap to meet these requirements because existing DBMSs, as well as the computing systems they rely on, are far too complex, first to be auto-administered and second to constitute a SOE. The first assumption is strengthen by the analysis done in which measures the distance separating current technologies from future self-tuning and zero-admin DBMSs2.[11]The worrying numbers regularly published by the Computer Security Institute and the FBI on database vulnerability truly confirms the second assumption

C. Data confidentiality A Top Concern:

Everybody agrees that maintaining financial application confidentiality in a public cloud is critical. It is worth mentioning since it goes against the conventional wisdom that this is essential also in a private cloud. Private clouds are not an abstract concept; they are used for some very practical needs. Often an organization will use a private cloud to serve its customers, employees or supply chain. These stakeholders have their own cloud data security concerns. From their point of view, they are using a public or community service, even if the technical implementation is called a "private cloud". This imposes many security requirements on the private cloud as well.

For example, consider a financial institution which is selling financial packages to the employees of its customers. The customers are large organizations, but the end-users are individuals: employees who need to manage their financial benefits.This institution has set up a software solution providing self-service tools to the end-users, to view and assess their financial packages. A fundamental part of the system is security, and the choice was made to base the system on a private cloud. But the end-users and – even more important – their employers, who are paying for the system, see this as a public cloud. Essentially they have outsourced their employee's data to an external financial provider. [6] They are therefore very strict about security, and ask many of the same questions they would ask in a pure "public cloud" implementation.

D. Achieving A Confidentiality In Cloud:

When moving to the cloud, all the traditional threats still exist. In addition, there are new, cloud specific threats. Cloud providers preach a "shared responsibility" mode shown in (fig.3) , claiming (for good reason), that you - the customer - should take all means to ensure application privacy and security. Trust cannot be outsourced, which is why each organization must own the responsibility to keep its data private. Some examples for new and specific cloud threats include shared infrastructure, employees of cloud providers who may be "malicious insiders," and unapproved usage of cloud infrastructure (for example a developer provisioning a new virtual server to test drive a recently developed app).

Regardless of the threat, a fundamental building block technology for achieving privacy in a public cloud is data encryption. Cloud encryption allows organizations to build "virtual walls" around their sensitive data, and therefore achieve privacy in a shared environment. But cloud encryption is only one part of the equation. Managing the encryption keys in a shared, public compute environment is the bigger obstacle. Another equally large issue is securing the most sensitive resources, such as the encryption keys themselves, when they are in memory of servers in the cloud.[6]

E. The 5 Commandments For Confidentiality In The Cloud:

To maintain client confidentiality, cloud providers need to follow the following five commandments:

a. Know Thy Requirements**.** Clients must have a firm understanding of their data and security requirements. Cloud-based platforms, like locally deployed solutions, can provide the security clients require if configured properly. Customers need to define their requirements and review the vendor's abilities to provide the appropriate level of security.

b. Thy Regulator is Everywhere so Follow all Thy Regulator's Commandments**.** SaaS-based vendors are required to be proactive with global regulatory agencies to understand the mandates and best practice recommendations. These policies will address data retention and access and must be incorporated into the organizations internal policies, covering all employees and applications.

c. Thou Must Learn and Teach. Vendors must continually educate and train their employees on procedures and best practices. Security awareness and client confidentiality should be regularly discussed. Simulations are an excellent way to test how internal staff would react when faced with unethical scenarios.

d. Thou Should Trust, but Verify**.** Trust that your employees are honest, but verify that the controls in place are working. Technology can audit access trails to ensure only those employees and applications entitled, are accessing the data. Proactively look for unauthorized access to data and regularly review the entitlements, both of employees and applications, for those granted access to sensitive client information. End-users should request on-site audits to gain confidence in the controls technology providers have implemented.

e. Only the Anointed Shall Enter the Temple Lest They Die**.** Data should not be accessed via low level (SQL) means. Data should be controlled through applications and granted via role-based entitlements. Vendors have a need to provide a high level of service to their clients, however, service and support personnel should have a different access entitlement than the sales team.

The cloud offers unique advantages for efficiency, transparency, and savings that have never been available before. [7]By working together with clients, and following the five commandments for maintaining confidentiality in the cloud, technology vendors can continue to provide this critical service in a secure environment while adding unparalleled value to their end-user.

There are some points regarding cloud computing

a) Cloud computing has significant implications for the privacy of personal information as well as for the confidentiality of business and governmental information.

b) A user's privacy and confidentiality risks vary significantly with the terms of service and privacy policy established by the cloud provider

c) For some types of information and some categories of cloud computing users, privacy and confidentiality rights, obligations, and status may change when a user discloses information to a cloud provider.
d) Disclosure and remote storage may have adverse consequences for the legal status of or protections for personal or business information.
e) The location of information in the cloud may have significant effects on the privacy and confidentiality protections of information and on the privacy obligations of those who process or store the information
f) Information in the cloud may have more than one legal location at the same time, with differing legal consequences
g) Laws could oblige a cloud provider to examine user records for evidence of criminal activity and other matters.
h) Legal uncertainties make it difficult to assess the status of information in the cloud as well as the privacy and confidentiality protections available to users.
i) Responses to the privacy and confidentiality risks of cloud computing include better policies and practices by cloud providers, changes to laws, and more vigilance by users.[8]

IV. KEY REQUIREMENTS OF CONFIDENTIALITY ASPECTS IN OUTSOURCING

We have discussed the key requirements of security aspects in database outsourcing in this section.

a. ***Confidentiality*** assures that only the authorized and intended users or systems are given consent to access the data. Data confidentiality refers to keep the data concealed from unauthorized access when it is stored and also when the data is in transit state. While ensuring the confidentiality, some additional dimensions are considered viz. *user privacy* and *access privacy*. Privacy is one of the primary requirements to achieve the security. User privacy conceals the user identity when he fetches or manipulates the data. Access *privacy* is assured when the access pattern of database and intended database records for a particular user are kept secret.

b. ***Integrity*** assures that the data stored in database or in transmission state are not modified or manipulated except by trusted persons or processes. *Completeness* and *correctness* are two important dimensions of *integrity*. *Completeness* means that query results obtained by fetching all records from the database and no any record containing the predicate in the query is excluded. It ensures that entire results are obtained when a query is fired on the database. *Correctness* means that the query results obtained from server are tamperproof and generated by original server. To verify that *integrity* is maintained, *query assurance* mechanism needs to be incorporated which ensures that query fired against the database is correctly and completely executed by service provider or process including all matched predicates in query.

c. ***Availability*** ensures that data are available to the trusted users and systems when they access database in an authorized manner. It is degree or extent to which database is in operable state which is calculated in terms of reliability. It is recommended for service providers to provide always-on availability of database to their valued and authorized users. To provide high level of availability, database system's down-time should be kept low.

d. ***Authenticity*** implies that contracts, query transactions and communication are genuine and identities of the involved entities (users and system) are verified and known. To provide the *authenticity*, the digital signatures are used.

e. ***Freshness*** of query results forms an important aspect of security when the database is periodically updated by the data owner. In such dynamically changing databases, *freshness* guarantees that the query results are obtained by executing the queries over the most updated database.

f. A person who works with data is delegated some responsibilities of data assurance. The tasks for which that person is responsible are the part of data security plan. This is called as *accountability*. It assures that all the operations performed by the users, processes or systems can be traced and identified to the respective entity. This ensures that all the valuable assets are refrained from the illegal access. This is also called as *access control*. Access control can be implemented by assigning role to the individual so that they can access the data according to their privilege level only. Access control matrix, access control list and access control capabilities list are some mechanisms for achieving the same.

g. ***Risk management*** is considered for proper implementation of security in outsourced databases. It includes the set of activities to identify and track the data security vulnerabilities and also the control measures are set to avoid the further risk to security[12].

These terms are maintained in the security policy agreed by both the data owner and service provider to meet the data owner's organization needs.These are the security requirements to be achieved in outsourced databases.

V. ANALYSIS OF DATA CONFIDENTIALTECHNIQUES APPLIED IN OUTSOURCED DATABASES

A. Fragmentation Approach:

Existing data fragmentation techniques are aimed at enhancing the data manipulation process, decreasing processing time, facilitating data manipulation, optimizing storage, increasing exibility, distributing processing costs, and facilitating data distribution and transportation, but are not specially designed with data security in mind. Current state-of-the-art approaches that focus on security aspects in data confidentiality using fragmentation 3 fragmentation rely on encryption for ensuring data confidentiality.[1]Our approach aims at minimizing the amount of encryption needed and relies on unlink ability of data entities to limit the privacy impact in case of single-point data leakages.

The aim of this research is to secure the sensitive outsourced data with minimum encryption within the cloud provider. Unfaithful solutions for providing privacy and

security along with performance issues by encryption usage of outsourced data are the main motivation points of this research. Secure and confidential storage of data in the cloud environment based on fragmentation method supports minimal encryption to minimize the computations overhead due to encryption. The proposed method uses normalization of relational databases, tables are categorized based on user requirements relating to performance, availability and serviceability, and exported to XML as fragments.[13] After defining the fragments and assigning the appropriate confidentiality levels, the lowest number of Cloud Service Providers (CSPs) is used required to store all fragments that must remain unlinkable in separate locations.

B. High-Performance Anonymization Engine Approach:

a. The development of an approach employing anonymization to protect confidential data on a public cloud. The objective is to protect customers' sensitive data, using the HPAE, from Storage As A Service (SaaS) providers when data is processed in the cloud. Moreover, as the concept of confidentiality varies among users, our solution aims to be flexible enough to allow users to configure their privacy policies.

b. The design of an architecture that efficiently handles large volumes of data offering high-throughput and fast processing. Performance is a determining factor in the adoption of a new approach, thus we aim to provide a solution that offers good performance (in terms of throughput) such that it is practical and allows users to anonymize data on-the-fly. Furthermore, the design should be modular and extensible to facilitate the accommodation of new components, such as new anonymization techniques, new input/output types, more efficient libraries/data structures etc.

c. The development of a prototype tool in Java to demonstrate our approach and measure the performance of our architecture.[10] To facilitate ease of integration of the HPAE with existing systems inside organizations ,our Java implementation will provide support for various types of input data sources as well as various types of output destinations.

VI. CONCLUSION AND FUTURE DIRECTIONS

Database outsourcing is a popular data management technology in the recent era which is accepted in the industries due to its inherent profitable features. In this paper, we have discussed the concept of DBaaS, its architecture and its benefits. Cloud computing offers attractive and cost effective solutions to customers, but it also imposes many challenges regarding confidentiality, privacy, control and laws and legislation. Most of the security measures are based on trust reliance where the customer relies strongly on the providers' trustworthiness. This paper focuses on secure and confidential data outsourcing to Cloud environments by using Fragmentation and high Performance Anonymazation Engine techniques and applying only minimal encryption to prevent data exposure. In this paper We have mainly focused on how the data confidentiality applied in outsourced databases and analyzed the techniques with their usefulness for the same.

The future work can also be focused on providing security for outsourced database along with reducing the communication, computation cost. There is much scope for improving the optimization of query processing time. The generic system can be developed which efficiently provides DBaaS and works on any database providing all the security mechanisms.

VII. REFERENCES

[1]. Aleksandar Hudic, Shareeful Islam, Peter Kieseberg, Edgar R. Weippl "Data Confidentiality using Fragmentation in Cloud Computing" Int. J. Communication Networks and Distributed Systems, Vol. 1, No. 3/4, 2012, pp. 325-329.

[2]. http://www.scaledb.com/DBaaS-Database-as-a-Service.php

[3]. Abhishek Patel, Mayank Kumar "A Proposed Model for Data Security of Cloud Storage Using Trusted Platform Module", 2010, pp. 587-593.

[4]. http://www.contegix.com/database-as-a-service-dbaas-pro HYPERLINK "http://www.contegix.com/database-as-a-service-dbaas-provides-many-benefits/"vides-many-benefits.

[5]. http://www.tomsitpro.com/articles/cloud-computing-ibm-homomorphic-encryption-cloud-security,1-1506.html.

[6]. http://www.wallstreetandtech.com/technology-risk-management/the-holy-grail-of-cloud-computing-maint/240006774.

[7]. http://tabbforum.com/opinions/the-5-command HYPERLINK "http://tabbforum.com/opinions/the-5-commandments-for-confidentiality-in-the-cloud"ments-for-confidentiality-in-the-cloud.

[8]. http://www.netop.com/solutions/report-privacy-and-confidentiality-in-cloud-computing.htm.

[9]. http://www.amazon.com/The-Next-Frontier-Confidentiality-Integrity/dp/0769549780.

[10]. Vanessa Ayala-Rivera, Dawid Nowak, Patrick McDonagh "Protecting Organizational Data Confidentiality in the Cloud using a High-Performance Anonymization Engine", 2011, pp. 2348-2352.

[11]. Luc Bouganim, Philippe Pucheral "Chip-Secured Data Access: Confidential Data on Untrusted Servers" Proceedings of the 28th VLDB Conference, Hong Kong, China, 2002, pp. 1381-1386.

[12]. Ajeet Ram Pathak, B. Padmavathi "Analysis of Security Techniques Applied in Database Outsourcing" (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (1), 2014, pp. 665-670.

[13]. http://www.emeraldinsight.com/journals.htm?articleid=17084777.

[14]. R. Datta, D. Joshi, J. Li, and J. Z.Wang, ``Image retrieval: Ideas, in_uences, and trends of the new age," ACM Comput. Surveys, vol. 40, no. 2, 2008, pp. 1-5.

[15]. J. K. Pillai, V. M. Patel, R. Chellappa, and N. K. Ratha, ``Secure and robust iris recognition using random projections and sparse representations," IEEE Trans.

Pattern Anal. Mach. Intell., vol. 33, no. 9, Sep. 2011, pp. 1877-1893.

[16]. Z. Brakerski, C. Gentry, and V. Vaikuntanathan, ``(Leveled) fully homomorphic encryption without bootstrapping," in Proc. 3rd Innov. Theoretical Comput. Sci. Conf., 2012, pp. 309-325.

[17]. Z. Brakerski, ``Fully homomorphic encryption without modulus switching from classical GapSVP," in Advances in Cryptology Crypto. New York, NY, USA: Springer-Verlag, 2012, pp. 868-886.

[18]. M. Naehrig, K. Lauter, andV.Vaikuntanathan, ``Can homomorphic encryption be practical?" in Proc. 3rd ACM Workshop Cloud Comput. Security Workshop, Ser. CCSW, New York, NY, USA, 2011, pp.113-124.

CONFERENCE PAPER
Two day National Conference on Innovation and Advancement in Computing
Organized by: Department of IT, GITAM UNIVERSITY Hyderabad (A.P.) India
Schedule: 28-29 March 2014