ISSN:2320-0790

An International Journal of Advanced Computer Technology

Privacy preserving in data-mining: A survey on security of outsourced transaction databases

Jinal Parmar[1],Vinit Gupta[2],IndraJeet Rajput[3]

[1, 2, 3]Department of Computer Engineering, HasmukhGoswami College of Engineering, Vehlal, Gujarat, India

Abstract*:* Database Outsourcing is a promising data management system in which Data-owner stores the confidential information at the third party service provider's site. The service provider manages and administers the database and avails the readymade services to the data owner and their clients to create, update, delete and access the database. Although database security is required because more service providers are not trustworthiness.The major requirements for getting security in outsourced databases are confidentiality, privacy, integrity, freshness in case of dynamic updates, access control in multi-user environment, availability and query authentication and assurance. To achieve these all requirements various security mechanisms like access control based approach, order preserving encryption based approach, hardware based encryption approach, fake tuple based approach, secret sharing approach, authenticated data structure approach, attributes based approach, combined fragmentation and encryption based approach, have been put forth till date. In this paper various security mechanisms analyzed and their significance given in this survey paper.

Keywords*:* Access Control, Confidentiality, Freshness, Integrity, Outsourced Databases, Query Authentication, Security mechanisms

I. INTRODUCTION

"Privacy Preservation" in data-mining means the Confidential or important data must be preserves or secure by the unauthorized person or attacker. The problem of privacy-preserving data mining has become more important in recent years because of the increasing ability to store personal data about users, and corporate data of private institute for the purpose of outsourcing and many different various purposes. Recently, the privacy of outsourced databases is a popular research topic. The third party provides a mechanism to allow their customers to create, store and access their databases at provider end. Using outsourced database can help organization reduce hardware equipment cost, system building, but also reduce cost of the personnel department. However, when the all of data be placed in outsourced database service provider, the provider is not trusted, sensitive data may have leaked crisis. Hence, the preserving privacy of database becomes very important issues [6].

The term "Database as a Service" (DBaaS) appeared in [7]. DBaaS is the breakaway technology of the recent era. The data owner of the organization stores their data at the third party service provider's site and delegates the responsibility of administering and managing the data to the service provider. This paradigm alleviates the need of installing data management software and hardware, hiring administrative and data management crew (personnel) at the organization's site. Due to this, the organization can concentrate on their core business logic rather than on the tedious job of data management leading to the saving in data management cost. Cloudant, Amazon DynamoDB, Hosted MongoDB are some examples of database service providers.

Preserving the security of the outsourced databases is a great challenge in the current scenario.As the data is stored at the service provider's site, it may be the case that service provider is distrustful in terms of disclosing and misusing the data. In this case, security of the database can be hampered in a dramatic way. If proper security is not enforced, then there are chances of data breaches and hacking the data in an unauthorized manner. Data breaching means disclosing the sensitive data intentionally or unintentionally. According to the survey taken by Trust wave Global Security [1], out of 450 data breach samples, 63% of investigations were related to the administration of third party service providers. According to the data breach investigation done by Trust wave in 2012, 76% of security deficiencies were caused by the third party service provider [2]. Therefore, it is very essential for the companies to be aware about security carrying out in their outsourced databases to keep the data confidential and thereby complying with the government rules and regulations. *Confidentiality*, *integrity* in context of *completeness* and *correctness*, *authenticity*, *accountability*, etc are considered as the foundation of security services. Therefore, implementing them in an systematic way is very important from the security point of view. Various techniques are used for realizing the security in database outsourcing. These techniques include encryption, authenticated data structures, order preserving encryption, signature schemes, etc. In this paper, we have given the complete analysis of security techniques along with their benefits and drawbacks.

The objective of this paper is to focus mainly on various security techniques for outsourced transaction datasets. The remaining portion of the paper is organized like this Section II presents the theoretical background of this paper. Section III presents comparative study/analysis of different security

techniques and section IV concludes the paper with summary and future direction.

II. DEFINATION AND THEORITICAL BACKGROUND

This section describes the concept of Database as a Service and benefits, architecture of database outsourcing model, challenges, and security services associated with the same.

A. The term Database as a Service

Database as a Service (DBaaS) is an architectural and operational approach enabling IT providers to deliver database functionality as a service to one or more consumers. DBaaS affords organizations an opportunity to standardize and optimize on a platform that eliminates the need to deploy, manage and support dedicated database hardware and software for each project's multiple development, testing, production, and failover environments [3].

The DBaaS ameliorates the need to purchase and install the data management hardware and software at the data owner's site. The data owner and clients use the readymade database service availed to them by service provider. The Organizations pay for the database service they are getting from the service provider. For the companies with less amount of resources limited hardware and time-bound projects, DBaaS best suits the scenario. Due to its inherent scalable property, DBaaScan scale up well in case of increasing user demands and also scale down when the demand subsides. The deployment of infrastructure for industries gets easier with the help of DBaaS. It offers flexible and on-demand services, optimizes performance tuning of the system, lowers the operating cost and complexity, accelerates the provisioning i.e. allows to clone the old database with a new schema, shortens the sales cycle, provides failover environment for project execution, enables the centralized administration and management of all kinds of databases[4].

Considering the adoption of DBaaS in industries [5], the research states that in 2016, the revenue generated by DBaaS providers will be $1.8 billion which is almost twice of the revenue generated in 2012 which is $150 million as shown in figure 1.

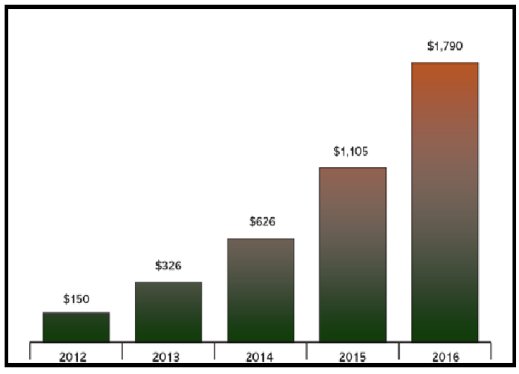

Fig. 1 DBaaS Market Revenue and Forecast ($ Million) [5]

B. Architecture of Outsourced Transaction Database Model

There are mainly three entities involved in the Outsourced Transaction Database Model. Three entities are:

1. **Data Owner**
2. **Service Provider**
3. **Clients**

The architecture is look like given below of Outsourced Transaction Database Model:-

Generally, data owner and clients are considered as trustful entity while service provider is distrustful in context of disclosing data in an unauthorized manner.

The **Data-owner** is responsible for update, insert, delete, modify, access databases. The data owner has the authority to permit or deny the clients for accessing the database. The **Service provider** performs all the data maintenance tasks. Data management hardware and software tools are deployed and maintained at the provider's site. The responsibilities of Service-provider are given below:-

- Provide Database as a Service
- Maintenance & administration of database
- Transaction Management
- Backup/Recovery System
- Fault Tolerance
- Scalability
- Database Availability
- Disaster Protection
- Efficient Query Processing

The query is processed efficiently and results are sent back to the queried. The **Clients** are given permission to access the data according to their privilege level.

There are three types of outsourced database model which are categorized on the basis of number of data owners and clients involved.

1. unified client model
2. multiple client model
3. Multiple data owner model

The first model is "unified client model" in which the database is used by single entity i.e. here functionality of client and data owner is same. The data owner does all the operations on the database. The communication link between data owner and client has high bandwidth. This model is adopted in [8], [9], [10].

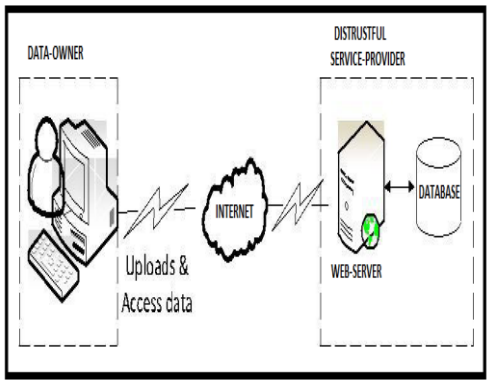

Fig.2 Single Data-owner and Service-Provider (Unified client model)

The second type of outsourced database model is "multiple client model" where multiple clients are given the authority of read only access. Here, the database can be accessed through mobile devices, laptops, PCs with limited bandwidth communication link. This model is adopted by [8], [9], [11].
"Multiple data owner model" is the third type of model which is adopted in [12]. In this model, each data owner uploads data at service provider's site. So, for every group of data owner and client, the separate access control and security policies are needed to be applied. This model can also be called as multi-authority outsourced database model.

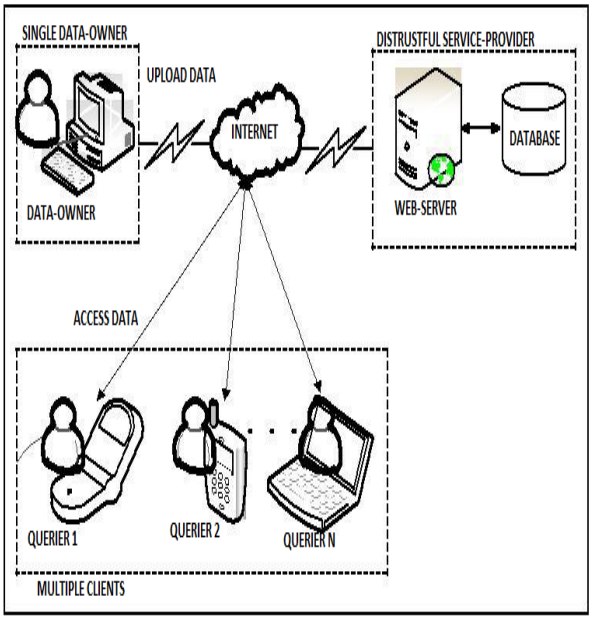

Fig. 3 Single Data-owner Multiple clients and Service-provider(Multiple Client Model)

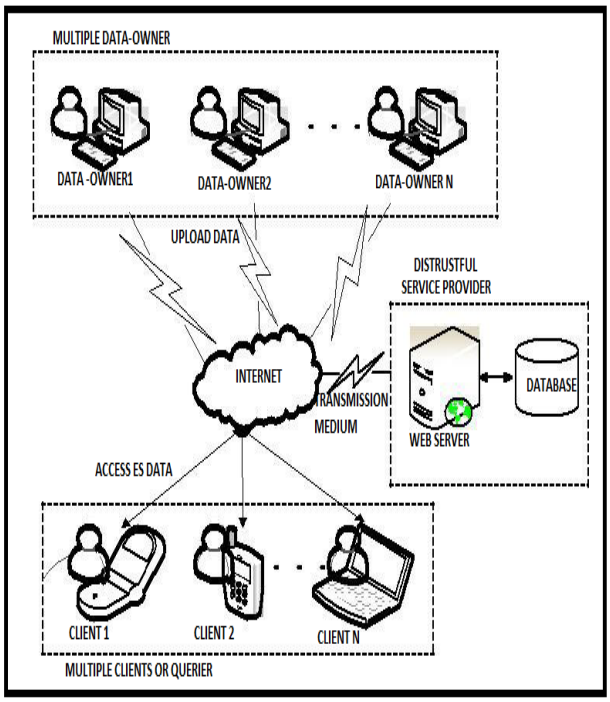

Fig.4 Multiple Data-owner Multiple Clients and Service Provider (Multiple Data Owner Model)

C. *Security Requirements*

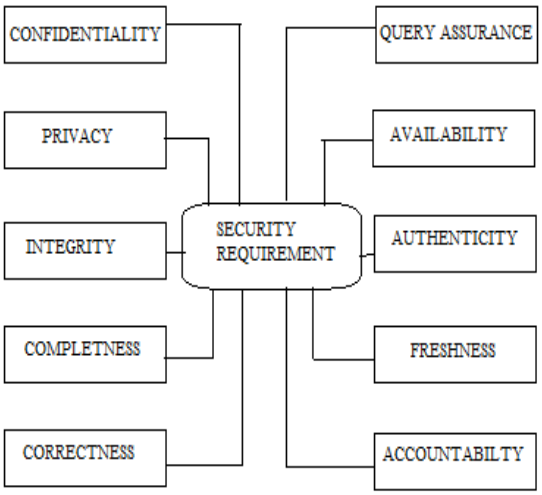

Fig. 5 General Security Requirements for achieving Security in Database Outsourcing.

The figure 5 describes the general security requirements for implementing them in database outsourcing.

1. The *confidentiality* is one of the important aspects in security. Making the data unintelligible when it is in the transit state or stored in data canter is referred to as preserving the *data confidentiality*.
2. *Privacy* is also considered while maintaining the *confidentiality*. Generally, the *privacy* comprises of *user privacy* and *access privacy*. For hiding the identity of the user, the *user privacy* is considered. *Access privacy* conceals the database access pattern for a particular user.
3. *Integrity* assures that the data being stored in the database or being transmitted in the network is tamperproof or unaltered. *Integrity* can be considered as the combination of two dimensions as *completeness* and *correctness*.
4. The *completeness* guarantees that the query results are retrieved by executing the query over all the database records which contain the predicate (tuple) expressed in the query.
5. *Correctness* promises that the results gained by executing the query against the database are unaltered, correct and are produced by the genuine database servers or genuine processes accessing the database.
6. *Query assurance* lets the client believe that query is executed over the genuine database server only.
7. *Availability* is important aspect in the security triad of *CIA (Confidentiality, Integrity, Availability)*. *Availability* is defined as the degree to which the database system is up and working in an operable state. It is very much crucial for the service provider to make the database service available all the time.
8. *Authenticity* refers to the trustworthiness and genuineness of databases, communication via transmission links, transactions, clients, data owners and the service provider. All the entities must be validated for ensuring the authenticity. Digital signature provides the better way to achieve the authenticity.

9. *Freshness* is the new aspect considered in database outsourcing. *Freshness* of database is assured only when the query is executed on the most recent edition (version) of the database uploaded by the data owner. Maintaining the *Freshness* has a great significance when the database is continuously or periodically updated and upgraded by the data owner. By sending the timestamp to the clients showing the validity of database is a good approach for ensuring the *Freshness* of the database.
10. The tasks performed by each entity are accountable for that entity only. This is called as *accountability*. *Access control* is referred to as allowing only the authorized users to access the protected data they are permitted to. *Access control* can be realized by following the three steps viz. *Identification*, *Authentication* and *Authorization.*

Identification is the act of finding which entity is querying the system. Once the identification is completed, authentication comes into picture. It refers to verify the claim of an entity to be genuine. For implementing the robust security, the multifactor authentication mechanism can be implemented. The multi-factor authentication can be the combination of username, passwords, biometric authentication and the unique assets like swipe cards. Once the specific entity is identified and authenticated, which data is permitted to access and which kinds of operations on data (Read, Write, Execute, Update) are allowed to be performed is found out. This is called as authorization.

III. DETAIL DISCUSION OF VARIOUS SECURITY TECHNIQUES

A. Access Control Based Approach[13]:-

Data *confidentiality*, *integrity*, and *privacy* of the clients' information protected by this approach. Among various services of cloud computing, enabling secure access to outsourced data lays a solid foundation for information management and other operations. However, more research efforts are needed to achieve flexible access control to large-scale dynamic data. In this environment, the data can be updated only by the original owner. At the same time, end users with different access rights need to read the information in an efficient and secure way. Both data and user dynamics must be properly handled to preserve the performance and safety of the outsourced storage system.

In[13]"**Secure and Efficient Access to Outsourced Data**",Weichao Wang, Zhiwei Li,Rodney Owens, BharatBhargava proposed their techniques that include:-

(1)The proposed approach provides fine grained access control to outsourced data with flexible and efficient management. The data owner needs to maintain only a few secrets for key derivation.

(2)It does not need to access the storage server except for data updates. They propose comprehensive mechanisms to handle dynamics in user access rights and updates to outsourced data.

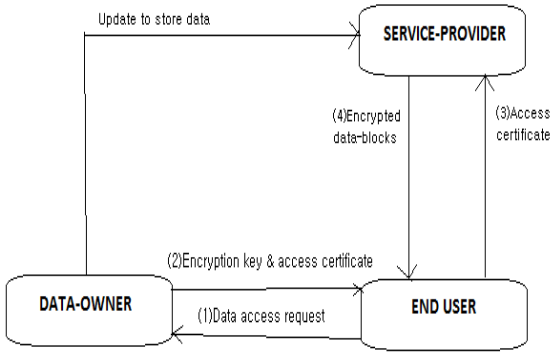

Figure.6 Illustration of the application scenario [13]

Therefore, the proposed approach is robust against collusive attacks if the hash function is considered safe. Analysis shows that the key derivation procedure based on hash functions will introduce very limited overhead. They propose to use over-encryption and/or lazy revocation to prevent revoked users from getting access to updated data blocks.

The main benefit of this approach is very limited overhead, avoid collusive attacks. The verification scheme of PKI is used for maintaining the integrity data access and the communication done for resource sharing. The accountability is also supported in this approach by tracing the user request for data using the timestamp. The drawback of this system lacks the robustness in terms of agent recovery. The approach does not support the scalability for acquiring large number of clients.

B. Attribute Based Access Control Approach[14]:-

To achieved *Confidentiality*, *Accountability*, *Access Control*Attribute based access control approach is used in which the access structure is related to the set of attributes of the user. In[14]*"***Achieving Secure, Scalable, and Fine-grained Data Access Control in Cloud Computing***",*Shucheng Yu, Cong Wang, KuiRen, and Wenjing Louaddress the open issue and propose a secure and scalable fine-grained data access control scheme for cloud computing. They proposed scheme in which each data file can be associated with a set of attributes which are meaningful in the context of interest. The access structure of each user can thus be defined as a unique logical expression over these attributes to reflect the scope of data files that the user is allowed to access. As the logical expression can represent any desired data file set, fine-grainedness of data access control is achieved. To enforce these access structures, they define a public key component for each attribute. Data files are encrypted using public key components corresponding to their attributes. User secret keys are defined to reflect their access structures so that a user is able to decrypt a cipher-text if and only if the data file attributes satisfy his access structure.
Here achieved these all Security requirement:-

1. Fine-grainedness of Access Control
2. User Access Privilege Confidentiality
3. User Secret Key Accountability
4. Data Confidentiality

The benefit of this method is that computation and communication cost incurred for revocation is less. It suffers from one weakness. The attributes associated with the users are placed in Attribute Authority. The revoked user can corrupt this authority by updating their own secret key also the secret key of non-revoked users.

*C. **Fake Tuple Insertion Based Approach[15][16][17][18]:-***

Fake tuple based approach is mainly used in outsourcing transaction database for the main purpose is to confuse the service-provider which may be attacker and also the security services like to *integrity* and *privacy*. Because of the fake tuple service-provider can't find the original support of the items in the dataset. The insertion of fake tuple based approach is adopted in [15],[16] and [17] to provide the *integrity* services. It mainly includes two approaches as probabilistic approach and deterministic approach.

In probabilistic method [15]"**Integrity auditing of outsourced data**", M. Xie, H. Wang, J. Yin, and X. Meng proposed the fake tuples are created and inserted into the database. For verifying the query *integrity*, the query is fired against the database server which contains both the real and fake tuples as the predicates. The server returns the query results. These results are verified by the client who knows all the fake tuples in the database. The client evaluates the fake tuples returned by server through result and the tuples determined by him. If tuples from server and from client are found out to be different, then the server is considered as dishonest and it is declared that the data has been tampered; else if tuples from both client and server are same, then it can be claimed that *completeness* is achieved i.e. *integrity* of the data is maintained. As already mentioned, the client should be aware of the fake tuples. The client has to maintain the copy of recent tuples. In case of large databases, a local database of fake tuples has to be maintained which causes extra storage overhead on client and it is against the concept of outsourcing. Freshness is guaranteed by using the fake update operation. The client deletes and inserts the fake tuples and analyse the results obtained by the server and evaluates the *freshness.*

In [16] "**Providing *freshness* guarantees for outsourced databases**,"M. Xie, H. Wang, J. Yin, and Meng proposedthe deterministic approach alleviates the need to save the fake tuples. The deterministic functions are used to recreate the fake tuples. These created fake tuples have prominent pattern and it can be easily noted by the hacker. Therefore, an encryption is applied on the real and fake tuples. Due to this, computation overhead increases.

In[17]"**Data Integrity Evaluation in Cloud Database-as-a-Service**",P. Ghazizadeh, R. Mukkamala S. Olariucreates the tuples with no distinguishable pattern using uniform distribution and hence removes the need of encrypting the tuples.

In [18]" **Privacy-Preserving Mining of Association Rules From Outsourced Transaction Databases**", FoscaGiannotti, Laks V. S. Lakshmanan, Anna Monreale, Dino Pedreschi, and Hui (Wendy),Wang's goal is to devise an encryption scheme with fake tuple which enables formal privacy guarantees to be proved, and to validate this model over large-scale real-life transaction databases (TDB). The architecture behind our model is illustrated in Fig. 7.

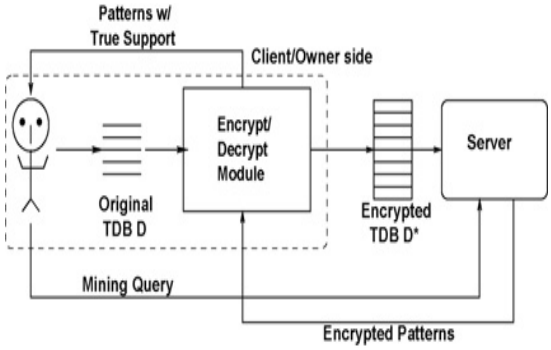

Fig.7 Architecture of mining-as-service paradigm [18].

The client/owner encrypts its data using an encrypt/decrypt (E/D) module, which can be essentially treated as a black box from its perspective. Here, Given a plain database *D*, construct a *k*-private cipher database *D**by using substitution ciphers and adding fake transactions such that from the set of frequent cipher patterns and their support in *D**sent to the owner by the server, the owner can reconstruct the true frequent patterns of *D* and their exact support. So here we can see that D* is the database with fake tuple which make complex structure for attacker or made confused to find original support of the item. Because of complex structure attacker may become confused to find real support of item so this is the benefit of this approach. The drawback of this approach does not provide *correctness* guarantees to the user.

*D. **Hardware-level Encryption based approach[19][20]:-***

For the security service the *data privacy*, special encryption hardware for IBM DB2 has been used in [7]. Here all the rows of the database are encrypted as a whole using DES (Data Encryption Standard) algorithm. The query execution time in hardware-level encryption is much less as compared to the software-level encryption.

In[19]"**SHAROES: A Data Sharing Platform for Outsourced Enterprise Storage Environments**" , Aameek Singh,Ling Liu propose a platform called SHAROES thatprovides data sharing capability over such outsourced storage environments. SHAROES provide rich *nix-like data sharing semantics over SSP stored data, without trusting the SSP for data *confidentiality* or *access control*. SHAROES is unique in its ability in reducinguser involvement during setup and operation through the use of in-band key management and allows a near-seamless transition of existing storage environments to the new model. It isalso superior in performance by minimizing the use of expensive public-key cryptography in metadata management.

In[20]"**TrustedDB: A Trusted Hardware based Database with Privacy and Data Confidentiality**",Sumeet Bajaj, RaduSion proposeda server side hosted and robust prototyping hardware. It provides privacyand data *confidentiality* performing the query optimization and supports any type of query fired against database.

The benefit of this approach is provides privacy, confidentiality and access control. But, due to in-built hardware processing of query.it suffers from cost overhead and it has some performance limitations. It is useful for small databases. In caseof large databases, encryption and decryption causes extraoverhead on the system leading to degradation in systemperformance and efficiency. Encryption based techniquessuffer from key management overheads.

E. ***Authenticated data structure based approach[21]:-***

Authenticated data structure based approach is used for mainly *Authentication* and *Integrity* services in the outsourced transaction database model. For Authenticated data structure approach some techniques are used in is one-way hash function, cryptographic signature approach, merkle hash tree, Bloom filters, Elliptic curve cryptography.

In [21]"**Scalable Verification for Outsourced Dynamic Databases**", HweeHwa Pang Jilian Zhang KyriakosMouratidis study the problem of verifying the *authenticity*, *completeness* and *freshness* of query answers from frequently updated databases that are hosted on untrusted servers. They introduce a protocol, built upon signature aggregation, for checking the *correctness* of query answers. Their approach has the important property of allowing new data to be disseminated immediately, while ensuring that outdated values beyond a pre-set age can be detected. They also construct authentication mechanisms for the B+-tree and standard relational operators that are suitable for dynamic databases.

The benefit is achieving considerably higher transaction throughput. The drawback is communication cost for transferring the page-level data is also more. For implementing the digital signature scheme, large storage and bandwidth is required.

F. ***Secret Share Distribution based approach[22]:-***

Though the encryption makes the data confidential forsafety, it creates extra overhead of encryption anddecryption on the system and degrades the performance of database. So to protect the data, secret share distribution based approach best suits in the system where encryption is not applied. Rather than performing encryption on data, data is distributed on multiple servers, called as shares.

To achieved *confidentiality*, *integrity*, *correctness* in [22]"**Privacy-Preserving Computation and Verification of Aggregate Queries on Outsourced Databases**",Brian Thompson, Stuart Haber, William G. Horne, Tomas Sander, Danfeng Yao present a solution in which service providers can collaboratively compute aggregate queries without gaining knowledge of intermediate results, and users can verify the results oftheir queries, relying only on their trust of the data owner. Our protocols are secure under reasonable cryptographic assumptions, and are robust to collusion among k dishonest service providers. They focused on computing aggregate queries including SUM and AVERAGE with SELECT clauses. Themain goal of PDAS is to prevent micro data (i.e., individual data entries) frombeing accessed by users or any of the third-party service providers who are delegated by the data owner to answer queries. They introduced two main techniques:

-A distributed architecture is introduced for outsourcing databases using multiple service providers. They extended threshold secret sharing schemes to support sophisticated aggregation operations by leveraging the additive property of polynomials over a field.

- A verification protocol is developed for the user to verify that the outsourced computation is indeed computed correctly, without leaking any microdata. They provided security analysis that our protocol achieves*secrecy*, *integrity*, *correctness*, and collusion-resistance properties. They also discussed possible variants.

The benefit of this approach is Encryption is not required so overhead is not occurs. Thedrawback of this approach is that it only supports thenumeric data. It does not support aggregate queries.

G. ***Order Preserving Encryption based approach[23]:-***

To gained better *privacy* in [23]"**Order Preserving Encryption for Numeric Data**",RakeshAgrawal , Jerry Kiernan, RamakrishnanSrikant, YirongXu present an order-preserving encryption scheme for numeric data that allows any comparison operationto be directly applied on encrypted data. Query results produced are sound (no false hits) and complete (no false drops). It allows standard database indexes to be built over encrypted tables and can easily be integrated with existing database systems. The proposed scheme has been designed to be deployed in application environments in which the intruder can get access to the encrypted database, but does not have prior domain information such as the distribution of values and cannot encrypt or decrypt arbitrary values of his choice. The encryption is robust against estimation of the true value in such environments.

When encrypting a given database P, OPES makes use of all the plaintext values currently present P, and also uses a database of sampled values from the target distribution. Only the encrypted database C is stored on disk. At the same time, OPES also creates some auxiliary information K, which the database system uses to decrypt encoded values or encrypt new values. Thus K serves the function of the encryption key. This auxiliary information is kept encrypted using conventional encryption techniques.

OPES works in three stages:

1. *Model:* The input and target distributions are modelled as piece-wise linear splines.
2. *Flatten:* The plaintext database P is transformed into a "flat" databaseF such that the values in F are uniformly distributed.
3. *Transform:* The flat database F is transformed into the cipher database C such that the values in C are distributed according to the target distribution.

The main advantage of encryption is that it makes the data unintelligible and scheme handles updates gracefully and new

values can be added without requiring changes in the encryption of other values.

It is useful for small databases. The drawback is that this approach supports only the range queries and suffers from plain-text chosen attacks and the size of encryption key is twice as large as the number of unique values in the database.

H. ***Fragmentation based approach[24]:-***

To gained *confidentiality* of the constraints and *correctness* and *completeness* in [24]"**Horizontal Fragmentation for Data Outsourcing with Formula-Based Confidentiality Constraints**", Lena Wiese introduces horizontal fragmentation in which rows of tables are separated (instead of columns for vertical fragmentation). They give a formula-based definition of confidentiality constraints and an implication-based definition of horizontal fragmentation correctness. Then they apply the chase procedure to decide this correctness property and present an algorithm that computes a correct horizontal fragmentation. In their approach for vertical fragmentation, only projection onto columns is supported and thus the so called "confidentiality constraints" are merely defined as sets of attributes of the database schema.

To extend the "vertical fragmentation only" approach they make the following contributions:-

– They propose to use not only vertical but also horizontal fragmentation. In particular, their aim to filter out confidential rows to be securely stored at the owner site. The remaining rows can safely be outsourced to the server.

-They extend expressiveness of the "confidentiality constraints" by using first order formulas instead of sets of attribute names. This implies that vertical fragmentation can be data-dependent in the sense that only some cells of a column have to be protected.

– They explicitly allow a full database schema with several relations symbols and a set of database dependencies. With these dependencies they introduce the possibility of inferences to the fragmentation topic and provide an algorithm to avoid such inferences. In their horizontal fragmentation approach fragments are sets ofrows instead of sets of columns. The fragments (the rows in the server and theowner fragment) have to be combined again by simply taking the union U ofthe fragments.

The benefit of this approach is encryption not require here so extra overhead not occur here. The drawback is that how to adaptively update the data and structure is become complex.

I. ***Combined fragmentation and encryption based approach[25]:-***

Toachieve data *confidentiality* and *privacy* in [25]" **Adaptive, Secure, and Scalable Distributed Data Outsourcing: A Vision Paper**" Li Xiong, SlawomirGoryczka, VaidySunderam made a framework in which they combine data partitioning, encryption, and data reduction such as compressed or statistical data outsourcing to ensure data confidentiality and privacy while minimizing the cost for data shipping and computation. Each resource provider may store parts of the data in original, encrypted, or reduced form. Algorithms can be developed to allow users to pre-process their data for secure outsourcing on distributed resource providers that systematically balance the requirements on confidentiality and privacy, scalability, and analytical utility of the data for a given workload. Adaptive outsourcing design that allow users to dynamically provision their outsourcing needs with data updates and changing query workload. Control-theory based mechanisms can be developed to effectively model and estimate the changing query workload and changing data for dynamically adjusting the outsourcing design.

An important building block of their framework is encryption and partitioning (or fragmentation) techniques. Encryption consists in encrypting all the values of an attribute, thus making them unintelligible to unauthorized users. Fragmentation consists in partitioning data records (horizontal partitioning) or attributes (vertical partitioning) in subsets such that only records or attributes in the same fragment are visible together.

The benefit of this systems techniques that make different types of cloud and local platforms compatible, host practical manifestations of remote databases, and perform at optimal levels in order to make the technology eminently usable.
The drawback is that how to adaptively update the data in the cloud while balancing the computational overhead and accuracy of the synopsis is a challenge. However, updating the deployed data too often increases the amount of noise that need to be added to the synopsis. Careful privacy budget management needs to be performed.

IV. CONCLUSION AND FUTURE WORK

The Database as a Service is a recent database management solution which is growing popular day by day due to its usefulness.In this paper, we have discussed the concept of DBaaS, its architecture and its benefits.The thorough analysis of general security requirements for the outsourced databases is done in this paper.We have mainly focused on how the security applied in outsourced databases and analyzed the techniques with their usefulness for the same. The detailed discussion of achieving the *confidentiality*, *integrity*, *completeness*,*Correctness*, *accesscontrol* and *accountability* in single and multi-user environment is given.The generalized securityframework can be developed such that it supports all types of databases and all the types of queries. Here summarized all the different security techniques with their benefits and drawbacks in table. The future enhancement can also focused on providing security for outsourced transaction database along with reducing the communication, computation cost and optimization of query processing time.

TABLE I: COMPARISION OF ALL SECURITY TECHNIQUES

Sr.No	Security Techniques	Achieved Security Services	Benefits	Drawbacks
A.	Access Control Based Approach[13]	-Confidentiality -Integrity -Privacy	-Very limited overhead, avoid collusive attacks. -The accountability is also supported in this approach by tracing the user request for data using the timestamp.	- This system lacks the robustness in terms of agent recovery. - The approach does not support the scalability for acquiring large number of clients.
B.	Attribute Based Access Control Approach[14]	-Confidentiality -Accountability -Access Control	-Computation and communication cost incurred for revocation is less.	-The attributes associated with the users are placed in Attribute Authority. The revoked user can corrupt this authority by updating their own secret key also the secret key of non-revoked users.
C.	Fake Tuple Insertion Based Approach[15][16] [17][18]	-Integrity -Privacy	-Because of complex structure attacker may become confused to find real support of item.	-This approach does not provide *correctness* guarantees to the user.
D.	Hardware-level Encryption based approach[19][20]	-Privacy -Confidentiality -Access Control	-It is useful for small databases.	-It suffers from cost overhead and it has some performance limitations. -It suffer from key management overheads.
E.	Authenticated data structure based approach[21]	-Authenticity -Integrity -Completeness -Freshness	-Achieving considerably higher transaction throughput.	-Communication cost for transferring the page-level data is also more. -For implementing the digital signature scheme, large storage and bandwidth is required.
F.	Secret Share Distribution based approach[22]	-Integrity -confidentiality -Completeness -Correctness	-Encryption is not required so overhead is not occurs.	- It only supports the numeric data.
G.	Order Preserving Encryption based approach[23]	-Privacy -Confidentiality	-The main advantage of encryption is that it makes the data unintelligible. - It is useful for small databases.	-These approaches suffer from plain-text chosen attacks. -This approach supports only the range queries. -The size of encryption key is twice as large as the number of -Unique values in the database.
H.	Fragmentation based approach[24]	-Confidentiality -Completeness -Correctness	-The benefit of this approach is encryption not require here so extra overhead not occur here.	-The drawback is that how to adaptively update the data and structure is become complex. -Chances to occur inference attack by the service provider.
I.	Combined fragmentation and encryption based approach[25]	-Confidentiality -Privacy	-This systems techniques that make different types of cloud and local platforms compatible, host practical manifestations of remote databases, and perform at optimal levels in order to make the technology eminently usable.	-The drawback is that how to adaptively update the data in the cloud while balancing the computational overhead and accuracy of the synopsis is a challenge. -However, updating the deployed data too often increases the amount of noise that need to be added to the synopsis. -Careful privacy budget management needs to be performed.

ACKNOWLEDGMENT

I acknowledge here my debt to those who have contributed significantly in this survey paper. I indebted to my internal guide Mr. Vinit Gupta,Department of Computer Engineering,HasmukhGoswami College of Engineering, Vehlal,Gujarat Technological Universityfor helping me and his experience is very helpful to me.

REFERENCES

[1] http://www.computerweekly.com/news/2240178104/Bad outsourcingdecisions-cause-63-of-data-breaches.

[2]http://www.networkworld.com/news/2012/020712-data- breach-255782.html

[3]www.oracle.com/technetwork/topics/.../oes-refarch-dbaas508111.pdf

[4]http://dbaas.wordpress.com/2008/05/14/whatexactly-is-database-as-a-service/

[5]https://451research.com/reportshort?entityId=78105&referrer=marketing

[6]Yung-Wang Lin, Li-Cheng Yang, Luon-Chang Lin, and Yeong-Chin Chen, *Preserving Privacy in Outsourced Database*, International Journal of Computer and Communication Engineering, Vol. 3, No. 5, September 2014.

[7] H. Hacigumus, B. Iyer and S. Mehrotra, *Providing database as a service*, in Proc. of IEEE 18th ICDE, 2002, pp. 29-38.

[8] E. Mykletun, M. Narasimha, and G. Tsudik,*Authentication and integrity in outsourced databases*, In Proc. of ACM Trans. On Storage, vol. 2, 2006, pp. 107-138.

[9] M. Xie, H. Wang, J. Yin, and X. Meng, *Integrity auditing of outsourced data***,**"VLDB 2007, pp. 782-793.

[10] Zheng-Fei Wang, Ai-Guo Tang,*Implementation of Encrypted Data for Outsourced Database*, In Proc. of Second International Conference on Computational Intelligence and Natural Computing (CINC), IEEE, 2010, pp. 150-153.

[11] Li Feifei, Marios H, George K, *Dynamic Authenticated Index Structures for Outsourced Database***,** In Proc. of ACM SIGMOD'06. Chicago, Illinois, 2006, pp. 121-132.

[12]SomchartFugkeaw, *Achieving Privacy and Security in Multi- Owner Data Outsourcing*, In Proc. of IEEE Transactions 2012, pp.239-244.

[13] Weichao Wang, ZhiweiLi,Rodney Owens, Bharat Bhargava,*Secure and Efficient Access to Outsourced Data***,** *CCSW'09*, November 13, 2009, Chicago, Illinois, USA ACM 978-1-60558-784-4/09/11.

[14] Shucheng Yu, Cong Wang, KuiRen, and WenjingLou,*Achieving Secure, Scalable, and Fine-grained Data Access Control in Cloud Computing*, IEEE INFOCOM 2010.

[15] M. Xie, H. Wang, J. Yin, and X. Meng, *Integrity auditing of outsourced data*, VLDB 2007, pp. 782-793.

[16] M. Xie, H. Wang, J. Yin, and Meng, *Providing freshness guarantees for outsourced databases*, in Proceedings of the 11th international conference on Extending database technology: Advances in database technology, ser. EDBT '08. New York, NY, USA: ACM, 2008, pp.323–332.

[17] P. Ghazizadeh, R. Mukkamala S. Olariu, *Data Integrity Evaluation in Cloud Database-as-a-Service*", In Proceedings of IEEE Ninth World Congress on Services, 2013, pp. 280-285.

[18] FoscaGiannotti, Laks V. S. Lakshmanan, Anna Monreale, Dino Pedreschi, and Hui (Wendy) Wang, *Privacy-Preserving Mining of Association Rules From Outsourced Transaction. Databases*, IEEE SYSTEMS JOURNAL, VOL. 7, NO. 3, SEPTEMBER 2012.

[19] AameekSingh,Ling Liu**,** *SHAROES: A Data Sharing Platform for Outsourced Enterprise Storage Environments***,**Data engineering, IEEE 2008.

[20] Sumeet Bajaj, RaduSion, *TrustedDB: A Trusted Hardware based Database with Privacy and Data Confidentiality*, In Proc. of IEEE Transactions on Knowledge and Data Engineering, 2013.

[21] HweeHwa Pang Jilian Zhang KyriakosMouratidis,*Scalable Verification for Outsourced Dynamic Databases*, ACM.VLDB '09, August 2428, 2009, Lyon, France Copyright 2009.

[22]Brian Thompson, Stuart Haber, William G. Horne, Tomas Sander, Danfeng Yao, *Privacy-Preserving Computation and Verification of Aggregate Queries on Outsourced Databases*, HP Laboratories HPL-2009-119, published by Springer Aug-2009.

[23] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, *Order pre-serving encryption for numeric data*, In Proceedings of the 2004 ACM SIGMOD international conference on Manage-ment of data,SIGMOD '04, pages 563–574, 2004.

[24]Lena Wiese,*Horizontal Fragmentation for Data Outsourcing with Formula-Based Confidentiality Constraints*, Advance in information and computer security, Springer 2010.

[25] Li Xiong, SlawomirGoryczka, VaidySunderam, *Adaptive, Secure, and Scalable Distributed Data Outsourcing: A Vision Paper*,3DAPAS'11, June 8, 2011, San Jose, California, USA. Copyright 2011 ACM 978-1-4503-0705-5/11/06.