

## **Enabling Data Storage Security with Blockchain Technology**

Thiruvankatasamy S<sup>1</sup>, Subhashri B<sup>2</sup>, Vikasini D<sup>3</sup>, Madhumitha A<sup>4</sup>

<sup>1</sup>Assistant Professor, <sup>2,3,4</sup>UG Students - Final Year, Department of Computer Science and Engineering, Nandha College of Technology, Perundurai – 638052, Tamilnadu, India

### **Abstract**

There are multiple ways of keeping data by using a single or Distributed Database on the cloud, yet a single database client can keep data only on a single server, and if multiple clients try to access those at the same time with data consistency, the data may be altered at the time of concurrent data access. Handling data on a single database could be easy, but problems associated with data consistency, confidentiality, availability, and bottleneck/single point of failure are always there. To eliminate these problems, today the author uses the Distributed Database approach, in which data is stored on multiple servers, and makes clients access the data concurrently. Here, data is mirrored in multiple places and made available anytime/anywhere. Moreover, various replicas of the data are kept on various servers so if the data is lost during concurrent access, the replica is available and made easily available to users. Data security is administered using various encryption algorithms like DES (Data Encryption Standard), 3DES (Triple Data Encryption Standard), AES (Advanced Encryption Standard), and the like. Different kinds of data are encrypted in different cryptographic schemes and thereby level of security is enhanced. Moreover, an error recovery process is also executed in the system resulting in auto-correction of a failed cloud instance if the data is found to be modified there. The main advantage of this approach is data security as well as a single point of failure elimination. In this architecture, the client is unaware of what kind of encryption is made to required data and what cloud server instance is providing those data. Hence data security, availability, and consistency are achieved better than single cloud database architecture.

**Keywords:** Database security, cryptographic algorithms, database replica, error recovery process.

### **1. Introduction**

Cloud Computing is a content delivery service that utilises and expands computer technology via the Internet. Data centres are being transformed into massive computing service pools as cheaper and more powerful processors are combined with the "software as a service" (SaaS) computing technique. Meanwhile, increasing network bandwidth and dependable, supple network connections enable clients to subscribe to high-quality services originating from data/software located solely in remote data centres. Although intended as an Internet service platform, the Cloud's innovative data storage paradigm poses a number of

design challenges that have a significant impact on the entire system's performance and security.

In the cloud context, the secured and critical information is stored in the cloud. Cloud Security Alliance's recent report lists data loss/leakage as one of the top confidential concerns in the cloud. Recent laws, regulations /compliance frameworks measure the risks; offended companies are held responsible for the loss of sensitive data and face heavy penalties over data breaches. To misplace data security practices harm on a personal level. Stolen medical records, and credit card numbers/bank information cause emotional as well as financial ruin. One of the biggest problems with cloud data storage is data integrity verification at untrusted servers. For example, the service storage provider may experience Byzantine failures intermittently and may decide to bury the data errors from clients for their benefit. This is more serious for saving amount and storage space the service provider might ignore to maintain or deliberately erase rarely accessed data files that belong to ordinary clients. By considering the large size of outsourced electronic data as well as the client's inhibited resource potential, the problem is generalized as to how could the client find an enhanced way to perform periodical integrity verifications without data files' local copy.

Confidential data stored within the cloud must be protected. Ensuring the confidentiality of information necessitates the best data management preferences like only the privileged/authorized party only can access the original content. The author proposes a solution for these choices that eliminate third-party involvement. The earlier designed architecture to achieve this is Proxy Less architecture (PLAC). There, the use of the proxy server between the client and database is eliminated. The PLAC architecture contains two problems. 1) Single point failure (means that data will be lost if the system failure occurs due to some problems) and 2) Bottleneck (too many requests coming for the same operation). Due to these drawbacks, the performance of the cloud is decreased.

The DD-PLA (Distributed Database-Proxy Less Architecture) was created to address these difficulties. There is no intermediary proxy server between the client and the cloud database in this architecture. Data is dispersed over the cloud in this distributed cloud. The advantage of storing data on a distributed database is that it reduces the load that a single database bears. Each approved user is given access permission based on their role. The vertical fragmentation of data is used to implement the DD-PLA architecture. The problem of a bottleneck is solved via fragmentation. The AES technique is employed in this design, together with a hash function, to store data in encrypted format using a symmetric key. The concurrent execution of operations is supported by the DD-PLA architecture.

## **2. Literature Review**

The major problems in previous designs are that of supporting dynamic data operations in cloud data storage applications. In Cloud Computing, the remotely kept electronic data might

not only be retrieved but also modified and updated by the clients, e.g., through insertion, deletion, modification, etc. Regrettably, the state of the art in the remote data storage context focuses mainly on static data files and the importance of dynamic data updates has received narrow attention so far.

In the cloud environment, clients are unreliable or might not be able to afford the perform frequent integrity checks overhead. Thus, in a practical scenario, it seems more rational to provide the verification protocol for error checking, which plays a more important role in achieving economies of scale and integrity for Cloud Computing. Moreover, for efficiency contemplation, the outsourced data themselves should not be modified in any one of the replicas. This study deals with the above considerations and provides ways to enhance the distributed database storage with various cryptographic security levels.

Since users cannot retain local copies of outsourced data, there exist many incentives for CSPs (Cloud Service Providers) to perform unfaithfully for the cloud users regarding their outsourced data status. For example, for increasing profit margin using cost reduction, they can dispose of infrequently accessed data without being detected promptly. Similarly, CSP attempts to secrete data loss incidents to maintain a reputation. Therefore, although outsourcing data into the cloud is economically good-looking for the cost and difficulty of long-term large-scale data storage, its lacking of offering a strong promise of data integrity and availability may hamper its wide adoption by both individual and enterprise cloud users. Existing systems regarding database security in cloud environments are discussed below. In *Distributed, Concurrent-Independent Access to Encrypted Cloud Databases*, the authors Luca Ferretti et al., stated that placing critical data to a cloud provider should guarantee the security/availability of data in use, in motion, and at rest.

Storage services provide several options, however data confidentiality solutions for the database as a service paradigm are still in their infancy. This research presented a revolutionary architecture that combines cloud database services with data security and the ability to run several operations on encrypted data at the same time. This is the first solution that allows geographically dispersed customers to connect directly to an encrypted cloud database and perform concurrent and independent activities, including database structure modification. The suggested design also has the benefit of eliminating intermediate proxies, which limit the inherent elasticity, availability, and scalability of cloud-based systems. The suggested architecture's efficacy is assessed using theoretical studies and substantial experimental data based on a prototype implementation using the TPC-C standard benchmark.

Data confidentiality is crucial in the cloud, since critical information is stored in the infrastructures of untrustworthy third parties. This imposes obvious data management choices: original plain data must be accessible only by trusted parties, which exclude cloud providers, intermediaries, and the Internet; data must be encrypted in any untrusted setting. Depending on the type of cloud service, achieving these objectives is more difficult. There are various ways for ensuring confidentiality in the storage as a service paradigm, but ensuring secrecy in the database as a service (DBaaS) paradigm is still a work in progress.

In this context, SecureDBaaS is proposed as the first solution that allows cloud tenants to fully utilise DBaaS characteristics such as availability, stability, and elastic scalability without exposing unencrypted data to the cloud provider. The architecture was created with three goals in mind: allowing multiple, independent, and geographically distributed clients to run concurrent operations on encrypted data, including SQL statements that alter the database structure; maintaining data confidentiality and consistency at the client and cloud level; and eliminating any intermediate server between the cloud client and the cloud provider.

A prototype of SecureDBaaS, which supports the execution of concurrent and independent operations to the remote encrypted database from many geographically distributed clients as in any unencrypted DBaaS setup, demonstrates the possibility of combining the availability, elasticity, and scalability of a typical cloud DBaaS with data confidentiality. SecureDBaaS combines available cryptographic algorithms, isolation techniques, and unique strategies for the management of encrypted metadata on untrusted cloud databases to fulfil the following aims.

Theoretical analysis of data consistency issues induced by concurrent and independent client access to encrypted data is presented in this paper. Due to their great computational complexity, we cannot use totally homomorphic encryption techniques in this case. There is no intermediary proxy or broker server between the client and the cloud provider in the SecureDBaaS architecture, which is built exclusively for cloud platforms. By removing any trusted intermediate servers, SecureDBaaS may achieve the same levels of availability, reliability, and elasticity as a cloud DBaaS. Other concepts based on intermediate server(s), such as , were deemed unsuitable for a cloud-based solution since any proxy creates a single point of failure and a system bottleneck, limiting the key benefits of a cloud-based solution (e.g., scalability, availability, and flexibility).

It protects data by allowing a cloud database server to perform concurrent SQL operations (not only read/write operations, but also database structure changes) on encrypted data. It provides the same availability, elasticity, and scalability as the original cloud DBaaS because it does not require an intermediate server. For most SQL operations, network latencies obscure cryptographic overheads, which affect response times. Multiple clients, who may be geographically distant, can use a cloud database service concurrently and independently. The cloud database does not require a trusted broker or a trusted proxy because tenant data and metadata are always encrypted. It is compatible with the most popular relational database servers and can be used with a wide range of DBMS implementations.

A significant portion of the research focuses on solutions for supporting concurrent SQL operations (including statements that change the database structure) on encrypted data issued by heterogeneous and possibly geographically distant clients. The proposed approach does not necessitate any changes to the cloud database and can be implemented immediately with

existing cloud DBaaS such as PostgreSQL Plus Cloud Database, Windows Azure, and Xeround. There are no theoretical or practical constraints to extending our methodology to other systems and using other encryption methods. According to the authors Amjad Alsarhani et al. in cloud computing technology allows for multiple configurable resources to be controlled in a decentralised manner. However, because data is not under the control of the content owner, data security risks arise.

To improve database confidentiality, they proposed a combination of encryption algorithms and a distribution scheme. The database was dispersed among the clouds based on the amount of protection given by the encryption techniques used. They evaluated their approach by creating and conducting trials as well as comparing it to existing solutions. The results showed that their strategy provided a highly secure approach that ensured user data security while also delivering acceptable overhead performance. For improving database secrecy, they presented a new encryption algorithm combination and a distribution scheme. The database was dispersed among the clouds based on the amount of protection given by the encryption techniques used. They evaluated their approach by creating and executing experiments and compared it to existing solutions. The outcomes proved that their plan worked.

The authors of claimed that the deployed cloud computing systems reflect a significant shift in the way things are done. They are able to popularise the Internet on a massive scale and establish some significant service companies. It's a pay-as-you-go, scalable, ubiquitous computing technology that could fulfil long-held ambitions. With cloud computing, anyone can start small and scale up quickly. That is the foundation on which the technology is based. Even though cloud computing is still novel, technology is evolving. The resources are limitless virtual/physical systems on which the programme runs out of user details that contradict that view. They discussed data migration issues in cloud computing as well as various cloud computing reviews.

## ***2.1 Their parameters of Data Migration are:***

### ***2.1.1 Migration***

Virtual machines migration is moving a running virtual machine or an application among various physical machines without disconnecting the client or network of a virtual machine that is transferred from the original host machine to the destination. The main advantage of this is getting almost zero downtime of (hardly) any milliseconds

### ***2.1.2 Response Time***

It represents the amount of interval a specific load balancing algorithm takes to reply in a system that is facing a situation like overloading. It should be minimized to raise the system performance.

### *2.1.3 Fault Tolerance*

This is the capability of the load balancing algorithm to carry out uniform load balancing either in a situation of some arbitrary node failure or network link failure.

### *2.1.4 Migration Time*

This time is when one node gets overloaded then the jobs /resources are transferred from one node to another node. It should be reduced to enhance the performance of the system.

### *2.1.5 Performance Goals in Migration*

Migrating Virtual Machines helps in reducing downtime. Load balancing and consolidation are possible. Reducing the network activity.

### *2.1.6 Load balancing*

The objective of load balancing is to make better performance among network links, central processing units, and disk drives for attaining optimum resource utilization, maximum throughput, maximum response time, and avoiding overload/underload situations by balancing the load between several resources. They concluded that cloud computing technology is a new buzzword in the IT industry and comes to the world with a new horizon of hope. It is provided as a service on the Internet that is running a) dynamically scalable and b) over virtualized resources, which is a style of computing. In their paper, they have presented various aspects and pros and cons of cloud computing. They have also presented here various reviews by different researchers and focused on the data migration problems in Cloud computing. These reviews further lead to the idea to further research in cloud computing. In the future, they would present an agent-based service to resolve the data migration problem in cloud computing, and also the data security will be ensured by using the AES concept.

In the author John Harauz et al. discussed that the cloud environment is reliable, dynamic, and customizable with a guaranteed QOS (quality of service). Within this cloud system, users have a myriad of implicit resources for their computing requirements, and they don't need a full understanding of infrastructure. Cloud computing's advent made the declaration by Scott McNealy, (Sun Microsystems' founder), that "The network is the computer" a reality and gave the old Sun marketing motto a new life. In this new computing world, users are required to accept the underlying trust premise. To advance cloud computing, the community should take proactive steps to make sure of security. A movement exists to apply universal standards (e.g., open-source) to ensure interoperability between service providers. Attempts to be made to develop security standards to make sure data's

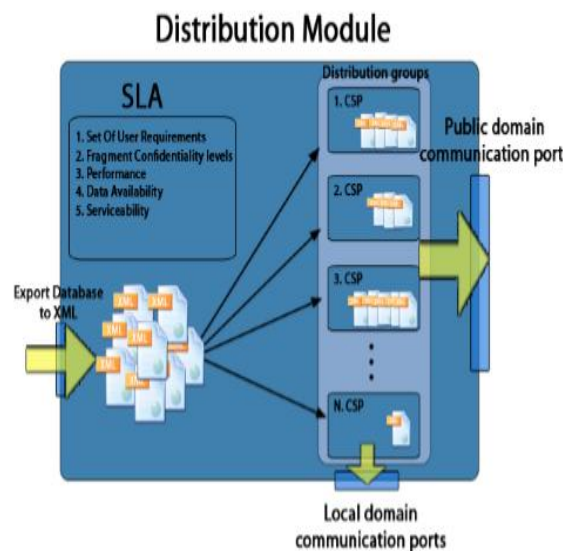
CIA. Even though the community is aware of the need for security and is attempting to initiate effective measures, a realm of security concerns exceeds these efforts.

As with most technological developments, regulators are typically in a “catch-up” mode to recognize policy, law, and governance. Cloud computing shows an extension of problems heretofore experienced on the Internet. Legal decisions will determine who “owns” the responsibility to secure information shared inside clouds. To make sure that the decisions are informed and suitable for the cloud computing environment, the industry should establish coherent/effective policy and governance to find and apply proper security methods. Within the cloud computing world, the virtual environment lets users access computing power that exceeds that contained within their physical worlds. To enter this virtual environment requires them to transfer the data throughout the cloud. Consequently, several data storage concerns can arise.

Consumers are typically ignorant of where their data is stored, as well as other sources of data that are stored alongside theirs. Storage providers must provide capabilities such as a tested encryption schema to ensure that the shared storage environment safeguards all data, stringent access controls to prevent unauthorised access to the data, and scheduled data backup and safe storage of backup media to ensure data confidentiality, integrity, and availability (CIA).

Legal difficulties such as e-discovery, regulatory compliance (including privacy), and audits also occur. The diversity of these legal considerations mirrors the diverse variety of interests that use or may employ cloud computing. These concerns, as well as their yet-to-be-determined solutions, provide light on how security is critical to cloud computing's continued growth and development. We must build a security model that promotes the CIA to address these and other challenges. This model could allow any cloud to provide a gauge of its current and prospective CIA, but the apparent challenge is gathering security data, which is difficult, if not impossible. Due to financial, business, and national security considerations, this dilemma has existed since the dawn of computing. It could be made worse by cloud computing.

In [5] the authors Aleksandar Hudec et al. stated that data confidentiality is one of the major challenges in ongoing cloud computing research. Hosting confidential business data at a CSP (Cloud Service Provider) needs the control transfer over the data to the partially-trusted external service providers. Current solutions to safeguard the data are mainly relying on cryptographic techniques. But, these cryptographic techniques create computational overhead, when data is spread among multiple cloud service provider servers in particular. They proposed a segmentation technique that efficiently saves the data on cloud servers using minimum possible encryption volume. The segmentation procedure is applied to relational databases where tables are treated as independent segments. This segmentation/distribution approach reduces trust expectancies among the external service providers and thereby improves confidentiality and privacy.



**Figure 1 The distribution module**

The segmentation process, which is illustrated in as distribution, Figure 1, is applied to each segment (table) regarding the level of confidentiality, ER model, and user requirements. The ER model helps to illustrate the relational viewpoint between tables and thereby facilitates the segmentation process. Each table is analyzed/ designated to corresponding spread groups, (step five of the segmentation algorithm). Each group denotes one CSP, where data is going to be distributed. The distribution process, (step six of the segmentation algorithm), gives secure/confidential data transport through VPN connections between customers and providers.

Existing data fragmentation techniques mainly aim to reinforce the method of knowledge manipulation using time interval reduction, data manipulation facilitation, storage optimization, increasing flexibility, processing costs distribution, and data distribution and transportation facilitation, but aren't specifically deliberated with data security in mind. Current state-of-the-art approaches only specialize in security aspects in data fragmentation



which believe encryption for ensuring data security. The approach aims to minimization of the quantity of encryption needed and relies on data entity unlinkability to limit privacy impact just in case of single-point data leakages.

Unlinkability is accomplished by distributing parts of the first data to varied storage providers. This study focuses on data fragmentation via relational databases which conform to the Normalization paradigm. The normalized data tables are considered standalone fragments and are then distributed to varied Cloud storage providers. These storage providers must be non-colluding, which might be ensured by e.g. Service Level Agreements (SLAs)/legal regulations. Note that these regulations usually specify confidential quality requirements only, but no countermeasures. Although costs play a big role in Cloud computing we don't ask them together of the prior goals of this paper to be ready to take additional confidentiality constraints counting on the data's domain under consideration, the info sets got to be analyzed first, to ultimately create SLAs conforming to well-defined user-specific confidentiality requirements. They considered the approach of applying the encryption on whole columns, even including the entity name, to cover the domain of stored data and to feature some additional complexity regarding correlation attacks: For example, a user could store differing types of knowledge marked as highly confidential and encrypt them. during this case attacker cannot know if the info set holds Mastercard numbers or simply some different, less sensitive, information. When using encryption, the encryption keys must be stored on a local domain and be accessible to appropriate users only. The problem with the encryption approach is, that it increases computational cost because each time data has to be decrypted before query processing.

**Table 1 Encrypted data of highly confidential tables prepared for distribution**

Emp ID	536480aa02d0014998bba86ba20d5855
EN1-33	acb195087d0d5424c17724a425c07ba1
EN1-26	d78c3b972ad9cf53ddf8ac7144aea80c
EN1-33	ddd97a3e5dbf2908fc1f58307e63f894
EN1-35	c8e562a275af91a90acffd7180999c6
EN1-35	2d9abf9143f741b41dab9c9a38b6c318

Medium Confidentiality Tables, (cf. Table 3), contain tables that have to be treated carefully in case of dependencies with other tables and distributed accordingly.

**Table 2 High Confidentiality table**

Emp ID	Credit Card Number
EN1-33	4111 1111 1111 1111
EN1-26	5500 6469 0000 0004
EN1-33	3400 6546 0000 0029
EN1-35	3000 9183 0000 0234
EN1-35	7010 9828 0000 0124

**Table 3 Medium Confidentiality tables**

Project Num	Project Budget
30-452-T3	10.000.000.000,00
30-457-T3	25.896.500.000,00
30-482-TC	3.200.000,00
31-124-T3	9.870.000.000,00
31-238-TC	150.000,00

Cloud storage providers possess high computational and storage resources. generally, the distribution model distinguishes two domains, the trusted local domain where the info originates from and therefore the semi-trusted property right where the info is distributed the very fact that the Local domain is taken into account trustworthy is employed to unravel the difficulty with tables that contain sensitive data without applying encryption the general public domain is taken into account non-confidential and thus the distributed data must be protected appropriately during the distribution process.

To ensure secure and reliable transport between the purchaser's local and public domains, we establish Virtual Private Network (VPN) sessions for the insecure property right. Insurance of appropriate levels of services and confidentiality from the Cloud storage service providers is established by using SLAs. The SLAs encapsulate all three essential requirements alongside performance, availability, and serviceability requirements, which the user demands data outsourcing. The local domain constitutes the start line of the info distribution process.

The distribution process is illustrated with the fragmentation algorithm stated below. The first step of the fragmentation process deploys each table (t) of a database schema S, as a private fragment, (step one among the Fragmentation algorithm) to make sure efficient transport and compatibility with different types of relational databases, the info is exported to XML files. counting on the utilization case scenario, the user defines a group of essential user requirements (data availability, serviceability, performance), which depend upon the info that the fragments contain, (step two of the Fragmentation algorithm). Afterward, requirements are assigned alongside confidentiality levels, for every fragment regarding the info that they contain.

### **3. Proposed Methodology**

To secure the validity of users' data in the cloud, this thesis suggested an effective and versatile distributed method with explicit dynamic data support. In cloud storage, some of the data is encrypted using symmetric key encryption. For example, the day-to-day transaction database records. Their aggregated values are encrypted in data owner storage which is less in size. The database in cloud storage may be redundant and can be accessed by a greater number of users. To check the data in cloud storage is safe, the sample data can be fetched from cloud storage and decrypted. The aggregated data is also decrypted so that the data from the cloud produce the same aggregated data. This ensures the data in the cloud storage is unaffected by users. The storage correctness verification is made in the above manner. The various modules used to implement the windows application project concept are

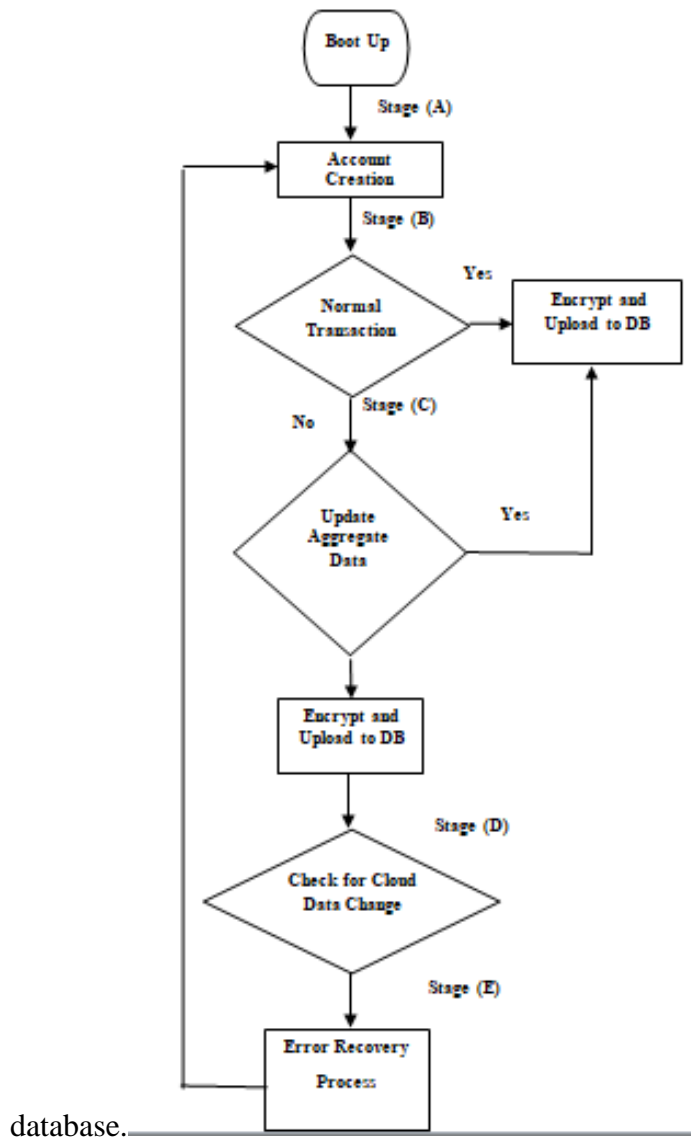
- Accounts Details
- Account Data Updation
- Aggregated Data Updation
- Error Recovery
- Accounts Details

Here, the accounting details, Reference Id holders, and contact details of the customers (account holders) can be keyed in. The account number is the primary key and is used to enter the transaction details such as credit or debit the account.

#### ***3.1 Account Data Updation***

Here, the data such as account number, date of transaction, amount, and type of transaction of accounts can be given. The type of transaction is selected as 'Debit' or Credit.

All the values are encrypted and stored in the



**Figure 2 Overall System Flow**

### **3.2 Aggregated Data Updation**

Here, the date of the transaction is given, so that the aggregated data is calculated for the given data for both 'Debit' and 'Credit' entries and stored in the database.

### **3.3 Error Recovery**

Here, the account numbers are populated in a combo box. An account number is selected. Four data grid controls are provided to fetch the records from four databases when the 'show data' button is clicked. 'Check For Errors' button is provided to check and display the error records. Figure 1 shows all the processes involved.

## **4. Conclusions**

This survey's goal is to provide information to the companies entering the cloud for taking steps to ensure they can trust the companies providing them with services, as well as the entities they are transacting with inside the cloud. Enterprises must have the ability to safeguard proprietary information on virtual servers and storage while giving cloud administrators the access and privileges needed to do their jobs. All cloud issues relate to establishing trust relationships, which form the conceptual foundations for cloud security. Many of the time-tested practices and technologies for managing trust relationships in traditional enterprise IT environments can be extended to work effectively in both private and public clouds. Those practices include data encryption, strong authentication and fraud detection, etc. Through this study, the data management process becomes easy. The interface is made according to our suggestion, then bank officials and cloud providers along with customers for access the data will be in the safest way. All the day-to-day activities can be assigned to them through the browser interface. The new system will eliminate the difficulties in the existing system. It will be developed in a user-friendly manner. The system will be very fast and any transaction can be viewed or retaken at any level.

## **References**

1. D. Dave, S. Parikh, R. Patel, and N. Doshi, "A survey on blockchaintechnology and its proposed solutioninProcedia Computer Science,vol. 160, 2019, pp. 740–745.
2. Cocco, L., Pinna, A. and Marchesi, M. (2017) Banking on Blockchain: Costs Savings Thanks to the Blockchain Technology. Future Internet, 9, 25.<https://doi.org/10.3390/fi9030025>
3. Harris, W.L. and Wonglimpiyarat, J. (2019) Blockchain Platform and Future Bank Competition. Foresight, 21, 625-639. <https://doi.org/10.1108/FS-12-2018-0113>
4. Dozier, P.D. and Montgomery, T.A. (2019) Banking on Blockchain: An Evaluation of Innovation Decision Making. IEEE Transactions on Engineering Management, 67, 1129-1141.<https://doi.org/10.1109/TEM.2019.2948142>

5. Guo, Y. and Liang, C. (2016) Blockchain Application and Outlook in the Banking Industry. Financial Innovation, 2, Article No. 24. <https://doi.org/10.1186/s40854-016-0034-9>
6. Sharma, A. (2018) Blockchain to Boost Regional Banks' Efficiency and Cut Costs. <https://www.thenationalnews.com/business/technology/blockchain-to-boost-regional-banks-efficiency-and-cut-costs-1.765312>.
7. Blockchain technology catches Axis Kotak Mahindras fancy, 2017. [Online]. Available:  
<http://www.livemint.com/Industry/loztj0R98Ea6m58Ng8jUzM/Blockchain-technology-catches-Axis-Kotak-Mahindras-fancy.html>
8. D. Dave, S. Parikh, R. Patel, and N. Doshi, "A survey on blockchain technology and its proposed solutions," in Procedia Computer Science, Vol. 160, 2019, pp. 740–745.
9. M. Wu, K. Wang, X. Cai, S. Guo, M. Guo, and C. Rong, "A Comprehensive Survey of Blockchain: From Theory to IoT Applications and beyond," IEEE Internet of Things Journal, vol. 6, no. 5, pp. 8114–8154, 2019.
10. Xin Wang, Xiaomin Xu, Lance Feagan, Sheng Huang, Limei Jiao, Wei Zhao, "Inter-Bank Payment System on Enterprise Blockchain Platform", IEEE 11th International Conference on Cloud Computing, 2018.