Scientific Journal of Impact Factor(SJIF): 4.14

e-ISSN(O): 2348-4470
p-ISSN(P): 2348-6406

International Journal of Advance Engineering and Research Development

Special Issue for ICPCDECT 2016,Volume 3 Issue 1

Secure Unique Identification using Encrypted Storage in NoSQL Database

Omkar Gurav, Swapnil Shirode, Anand Shende, Piyush Govekar

[1] *Computer, AISSMS IOIT, Pune, Maharashtra, India*

[2] *Computer, AISSMS IOIT, Pune, Maharashtra, India*

[3] *Computer, AISSMS IOIT, Pune, Maharashtra, India*

[4] *Computer, AISSMS IOIT, Pune, Maharashtra, India*

Abstract *- Database helps us to collect, retrieve, organize and manage the data. The database used in the project is used to associate user's personal documents together and keep it available for the user anytime he requires. We need every user to have a unique identity and for which we have used an Aadhar Card. This is used because there is only one card per person. The database is going to record all sensitive information of the user, so a need to provide security to it. This is done using encryption. There are various algorithms discussed, the one chosen is ETSFS (Extended Transpose Substitute Folding Shifting). A database is required which can be chosen to create tables or documents dynamically. Therefore NoSQL is chosen because it can create documents randomly. There has been a study as to where all an encryption process should take place. The encryption here is provided on the Client Application end. This project should provide a unique identity for the database which is secured using encryption. This will help the user to keep the personal information online and not carry it around everywhere physically.*

Keywords- NoSQL database, ETSFS algorithm, Application side encryption.

I. INTRODUCTION

Another module of this project is to make payments using this UID. Whenever we need to make a payment we just have to give in the UID and the rest of the transactions can be managed by itself. We can add cards to the tuples of a corresponding UID. This makes it handy and unique feature of the project.

We wanted to store all sensitive and personal information in the database which would include not only textual data but also images. This won't be easy with structured language. So we required a NoSQL which could save all the above data in it. As a NoSQL language we wanted a language that could be Consistent and Partition able always. So we decided to use MongoDB.

Whenever we talk about providing security to databases we usually use Encryption and Decryption. There are many Encryption techniques, most of the common ones are DES (Data Encryption Scheme), AES (Advanced Encryption Scheme) and their variants. Among many algorithms like these we chose ETSFS (Extended Transpose Substitution Folding Shifting).

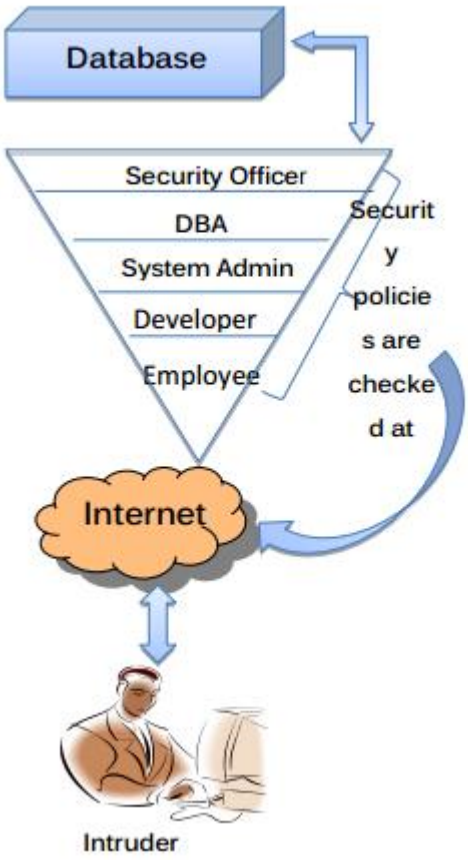
To identify every individual uniquely would make it easy to maintain data. This is done by using a unique identifier provided by the Indian Government on the Aadhar Card. Using this UID (Unique Identifier) we can save all the personal information and associate it with this. For example, the medical records of each person can be associated with this UID. Also that we can have more than one medical record for the same UID.

Another module of this project is to make payments using this UID. Whenever we need to make a payment we just have to give in the UID and the rest of the transactions can be managed by itself. We can add cards to the tuples of a corresponding UID. This makes it handy and unique feature of the project

II. LITERATURE SURVEY

The three main properties of database security are - Confidentiality - this means that we need to protect personal data so that nobody else can interfere with one's life (this is an exception in the case of doctors). Integrity this means that nobody without authority can change or modify the information in the database. Availability this means that the information should be available always. There are four types of flows according to [1], which are to be taken care of to provide database security - Access control this give access to only some of the people that have the rights to get access to the database and the information within. If the access control rights are given in the wrong hands, the effects can be hazardous. Information flow control If the flow of information is not proper then this gives the intruder a chance to latch up on the information and make wrong use of it. Cryptographic flow control this flow is the flow of encrypted data in the

network. Inference control in this type of flow it is taken care that the information reaches only the required person and with given rights. A method used to encrypt data is using a hierarchy-



As we can see that the employee has the least rights because it is not needed for him to know anything other than his job. The Developer has more rights because he has to develop the software used by employee. So he has rights of employee as well the Developer. The System Admin has right required to manage the system. The Database Admin has rights to maintain the database. The security officer has the highest rights to maintain the security and defend against attacks by intruders. The paper A NEW APPROACH FOR COMPLEX ENCRYPTING AND DECRYPTING DATA published by Obaida Mohammad Awad Al-Hazaimeh in the journal International Journal of Computer Networks & Communications (IJCNC) Vol.5, No.2, March 2013. Studies that the AES algorithm takes much more time to encrypt and decrypt than it should. So they proposed an algorithm that could make this process faster. This was successfully shown by them.

Table 2. Speed for Encryption in different key lengths

Algorithm	Key Length			
	128 - Bits	192 - Bits	256 - Bits	512- Bits
Proposed Technique	0.1814575	0.2014440	0.2377533	0.2972729
AES	0.5500000	0.6043321	0.7132599	0.8918188

Table 3. Speed for Decryption in different key lengths

Algorithm	Key Length			
	128 - Bits	192 - Bits	256 - Bits	512- Bits
Proposed Technique	0.2701433	0.2968131	0.2978936	0.3187594
AES	0.5402866	0.5936262	0.5957872	0.6375188

The advantages of this algorithm were that it took less time for encryption and decryption. But this made the algorithm complex. The paper A Research paper: An ASCII value based data encryption algorithm and its comparison with other symmetric data encryption algorithms published by A. Mathur in the journal International Journal on Computer Science and Engineering (IJCSE), Vol. 4, pp. 1650-1657, Sep 2012 ISSN: 09753397. Says that we can use the ASCII values of each character and encrypt it and then use it as a cipher. This algorithm employed a key that was used to encrypt another key called as a secret key. The encrypted secret key was used to encrypt the text. This gave a cipher text. This algorithm successfully implemented the encryption using ASCII values for characters. But the disadvantages of this were that the key had to be same as the length of the text and this key had to be entered by the user. So this made the algorithm inflexible. The paper An Encryption Algorithm Based on ASCII Value of Data published by Satyajeet R. Shinde, Rahul Patil in the journal International Journal of Computer Science and Information Technologies, Vol. 5 (6) , 2014, 7232-7234. Transforms the previous algorithm and makes it automatic and flexible by changing the key from user entered to automatically generate able. This gives us the flexibility to keep the text of any size.

COMPARISON BETWEEN EXECUTION TIME OF EXISTING ALGORITHM AND PROPOSED ALGORITHM

Size of plaintext	Execution time for existing algorithm in ms	Execution time for proposed algorithm in ms
2	322	15
4	3679	15
6	3861	16
8	4748	16
10	5543	30

The paper Database Security and Encryption: A Survey Study published by Iqra Basharat, Farooque Azam in the journal International Journal of Computer Applications (0975 888) Volume 47 No.12, June 2012 presented a survey on papers based on database encryption. The main papers in focus were:

1. Kadhem, H.; Amagasa, T.; Kitagawa, H.; A Novel Framework for Database Security based on Mixed Cryptography; Internet and Web Applications and Services, 2009. ICIW '09. Fourth International Conference on; Publication Year: 2009, Page(s):163170

2. Dr. Anwar Pasha Abdul GafoorDeshmukh; Dr. Anwar Pasha Abdul GafoorDeshmukh;TransparentDataEncryption-SolutionforSecurityofDatabase Contents; International Journal of Advanced Computer Science and Applications, Vol. 2, March 2011

3. Luc Bouganim; Yanli GUO; Database Encryption; Encyclopedia of Cryptography and Security, S. Jajodia and H. van Tilborg (Ed.) 2009.

4. TingjianGe, Stan Zdonik; Fast, Secure Encryption for Indexing in a ColumnOrientedDBMS;2007 IEEE 23rd International Conference on Data Engineering (2007) Publisher: IEEE, Page(s): 676-685.

Thefirstpaperdiscussesadifferentstrategyinwhichthedatabaseisencrypted end to end. Usually in database encryption the data residing in the databases is encrypted, but here the database scheme is itself encrypted. This is termed as Mixed Cryptography by the authors. It is seen that the security level has increased while the complexity and time required has also increased. This acts as a downfall of the strategy. The algorithm used for this type of encryption is any symmetric algorithm. This project considers the server to be a multi-party server. There are three places where encryption is implemented- 1. Server 2. Trusted Party 3. Client

This technique provides three tier encryption and is hence very secure. This also encrypts the queries at the client end which makes transmission secure as well. But this makes the performance of the queries low. Also here access control methods are not defined.

The second paper studies the encryption technique used in Microsoft SQL Server 2008. It provides security to tables, table space and columns. Before this we couldn't provide security to the databases that resided on removable media. This technique has the capability to provide security not only to the databases that reside in the internal devices but also on the external devices. This meant that databases on floppy, CD or Hard Disks can be protected in the same way. It uses a Master Key that was used to encrypt the key used by the client, this encryption gave rise to a certificate. A certificate is a digital mark or a signature that is used by the receiver to recognise who the sender. So the sender can make his identity be known. This will help the server recognise who the sender is and can make the decision of whether the databases should be shown or no. The Transparent Data Encryption consists of:-

1. Authentication- Every user is known and registered to the Server. Only recognised people can use this functionality.

2. Validation- The server must validate the request of each client. This makes the process synchronised.

3. Data Protection- The data of each client is protected.

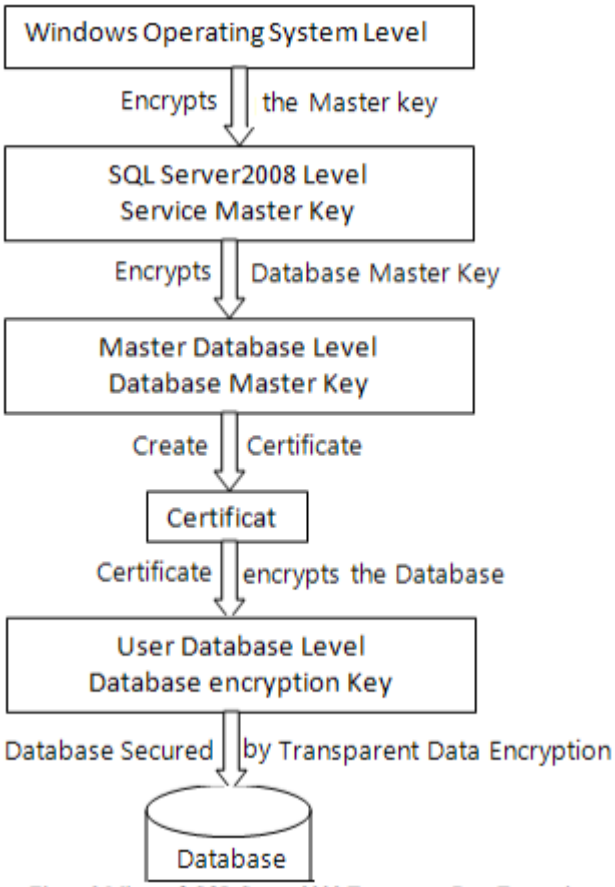


Figure 2 Microsoft SQL Server 2008 Transparent Data Encryption

This technique eliminates the problem of illegal access. Nobody can access and lay hands on your data without your permission. The cost of user management is efficiently handled. This reduces its cost. The privacy management is also maintained. This does not provide encryption across communication channels as the previous paper provided. This can highly effect affect the security across the network as the data is completely vulnerable. Not optimal for sensitive data. The database can never be opened without the certificate from the client. This increases the delay and not only that it makes the cost high because we need to send not only information we need (i.e. the request) but also the certificate. Extra information has to be sent. And also that the certificate can be easy modified and altered. This is very difficult to maintain then.

The third paper explains that we can use encryption no only on the database or storage but also on the application level. The encryption in this paper is applied on- 1. Storage level 2. Database level 3. Application level.

The advantage is that the security is maintained and it cannot be tampered with. The encryption keys can also be kept hidden and never be exposed to the outside world. The only disadvantage is that the algorithm is very complex.

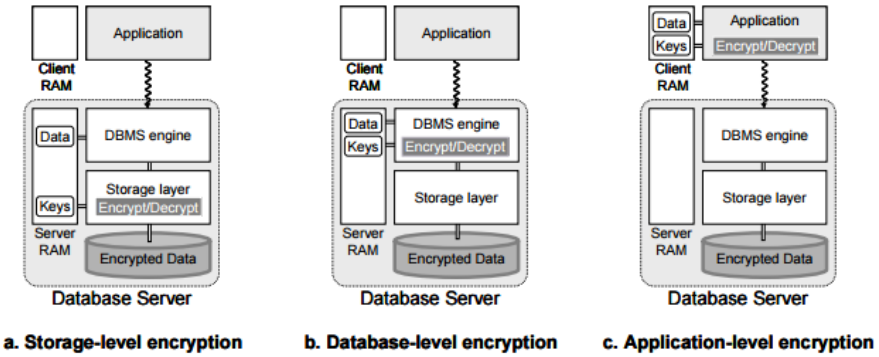


Figure 1. Three options for database encryption level

The goal of the fourth paper is to design a database system that can be encrypted as well as its performance is high. The advantages were

1. Fast indexing operation
2. Low decryption overhead

The disadvantages were

1. Complex algorithm
2. Costly

The method used is Fast Comparison Encryption. The encryption is performed at the Data Ware House level.

A few Algorithms that were studied-

1. DES:

DES is an acronym Data Encryption Standard, operating on 64 bits using a secret key which is of 56-bit long. A key is chosen randomly. The same key is used to encrypt the message and the same is used to decrypt it.

Six bits are mapped to groups of four bits. It is done in 16 rounds. Process of DES encryption:-

1. The input key is used to obtain sixteen 48-bit keys. These are used as sub keys. Each sub key is used in each round.
2. It is expanded from 32 bits to 48 bits using another fixed table.
3. The result is combined with the sub key for that round.
4. The 48 resulting bits are then transformed again to 32 bits. In the next round, combination is used as the left part.

2. TRIPLE DES:

Triple DES is the advancement of DES and covers the problems of DES. It uses three 56-bit DES keys, which creates a key with the length of 168 bits. The key is then split into three same length keys. The process is-

1. First key is for encryption.
2. Second key is for decryption.
3. Third key is for another encryption.

3. RSA

Stands after the name of developers: Ron Rivest, Adi Shamir, and Leonard Adleman. RSA is asymmetric, which means that the same key for encryption cannot be used for decryption. It uses a combination of public and private keys. The process of creating keys in RSA is as follows:

Two prime numbers are taken, *p* and *q*. 'n' is given by p multiplied by q. considering '*e'* should be greater than 1. *Gcd (e, (p-1), (q-1)) = 1.*

Then find the multiplicative inverse'd' of *e* modulo *(p-1) (q-1).* '(*N*, *e*)' is the public key. 'd' is the private key.

4. AES – Advanced Encryption Standard

AES varies between 128, 192 and 256 bits. AES is symmetric. The AES encryption process: (1) Key Schedule and (2) Encryption. The key schedule generates several sub keys. Steps taken to encrypt are:

1. ***Sub Bytes.***
2. ***Shift Rows***.
3. ***Mix Columns***.
4. ***Add Round Key.***

5. ETSFS – Extended Transposition-Substitution-Folding-Shifting

This covers the disadvantages of all the above algorithms. The steps are –

1. Transpose the data to be encrypted.
2. Substitution is then done on the transposed data.
3. Folding is then done.
4. The array is made which is used to change a letter for another.

III. PROPOSED SYSTEM

The proposed system enables the service registration to be made quickly and securely. To realize this system, several steps have been followed and are described in this section. It begins with the architectural diagram of the proposed system.

The goal of our project is to make use of an identifier, which is unique and which can be used by an individual in Medical and Banking domains or personal use. The objective of the proposed system is as follows:

The main objective of our project is to make use of the unique identifier provided by the Indian Government. The objective is to make this identifier applicable in day to day life. The application we thought of is to make use of this identifier in the field of medical health. This identifier clubs multiple medical reports and can be accessed from anywhere and at any time. We can also make transactions using this UID.

A. Architectural diagram

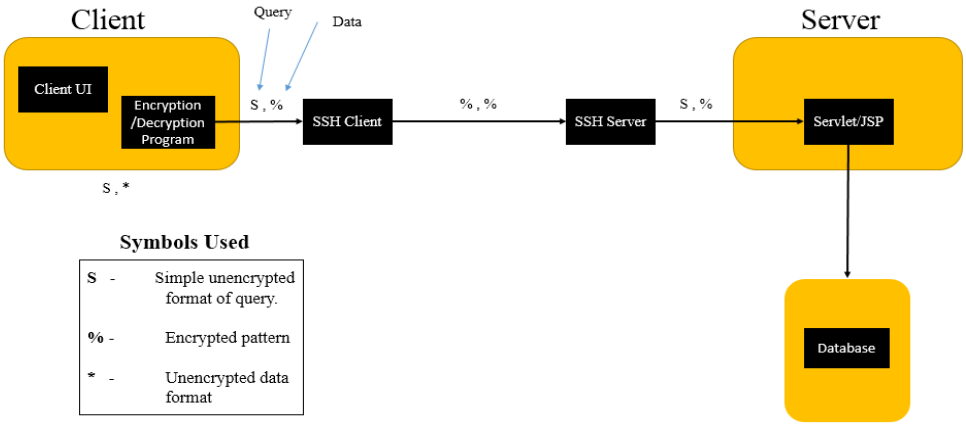


Fig-1 Architecture diagram

The given fig-1 is architectural diagram of our proposed system. As shown in given figure our system is divided into two parts

1. Server Side
2. Client Side

From the above figure, it is cogent that the client will only encrypt and decrypt the data. We use SSL protocol to transfer data. The server uses JSP and is connected to the database via a JDBC (Java Database Connectivity).
Whenever we take the data it is encrypted and is represented by '%'. The unencrypted or plain text is represented by '*'.
The S stands for the query.
Across the SSL we have the same combination of query and data as S and %.
But in the SSL, the query and data are % and %.
So data remains consistent at the start and at the end of SSL.
The data is stored in % form in database. Whenever we retrieve from the database, then only it is decrypted at the Client end.

B. System Description

The user first has to fill the registration using his Aadhar Card. We'll first check whether an account has been created for this Aadhar Card. If yes then an account is created and then personal information can be coupled with it. There are two types of users-

1. User
2. Administrator

The user has the rights to –

1. Insert
2. Update
3. Retrieve

The admin have tasks like-

1. Upgrade
2. Maintain

IV. METHODOLOGY

1) Client:

The client is an end from where he can interact with the Server. For doing so, he will have to first register with the Server. After creating an account he can couple documents with the account. After creating the account he can update, retrieve, insert and delete the documents coupled with the account.

2) Server:

The server side uses a Servlet. It takes the data from the client and stores it as it is in the database without decrypting. The Server understands the query from the client and works accordingly to reply back with the answer to the query or to store the data on insert option. It can also delete upon the delete query initiated by the client.

3) SSL (Secure Socket Layer Protocol).

SSL Protocol provides security over the Web. It is used for secured communication. It builds a secure tunnel between the client and the server.

The SSL stages are:

i. Establish key of safety communication.
ii. Authenticate the server.
iii. Authentication of client.
iv. Last stage.

4) Encryption/Decryption

The encryption is done at the client end and sent to the server end.
The decryption is also provided by the client end only.

5) Query and Flow

The information when entered, it will first encrypt the data and then sent to the server. The server then takes the data and stores it as it is without decrypting. This saves the time required for decryption.

When the client tries to retrieve data, it sends the query. The query is then encrypted by SSL Client and is decrypted by SSL Server. The decrypted query is sent to the Server. The Server then returns the documents requested by the Client.

The update query also works on the same terms. The data is updated by the server on the client's request.

V. ALGORITHM

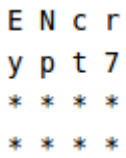
The algorithm which the paper proposes is an ETSFS algorithm along with some modifications over the ETSFS algorithm discussed in [25].

The process of Encryption-

In this modified algorithm, we have kept the Transpose phase as it is. The formula used in the Substitution phase is changed. Then next, the Folding phase has also been changed. Lastly the Shifting phase, shifts the rows upwards by one. The phase also sees a change.

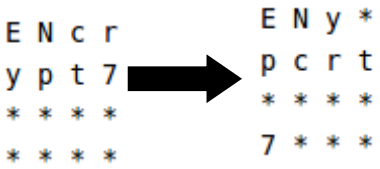
Algorithm for Encryption –

1. The string entered is converted into an array of integers. This is a 2-dimensional array of 4 rows and 4 columns.



The above is the plotting of the string 'ENcrypt7' in a 2D array form.

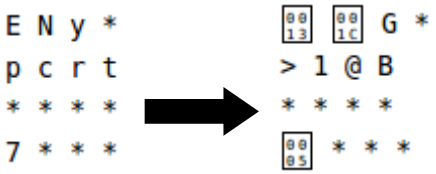
2. Transpose.



In the transpose phase we don't change anything. We just move around the elements of the array in the way shown by the arrows.
This gives a different look to our matrix than a normal one, which is also difficult to predict for the intruder.

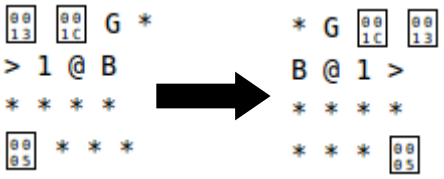
3. Substitution.
In the substitution phase, the values of the characters are substituted by other values. This adds to the security.
The formula used here is-
arr(i,j) = (((arr(i,j) + 62)%123)+11)

In case of ASCIIs less than 62,
arr(i,j) = (((arr(i,j) + 62-123)%123)+11)



The above diagram indicates the level of security the intruder will have to overcome.

4. Folding.
In the folding phase, the matrix or the array is mirrored about the center.



5. Shifting.
This shifts the rows upwards by 1. The upper row is added down at the end.

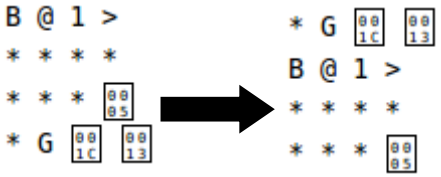
After this the Encryption shows the array in such a way that the intruder cannot identify what the sequence of characters is or which character is replaced by which character. Therefore, this algorithm gives high security.

The process for Decryption-

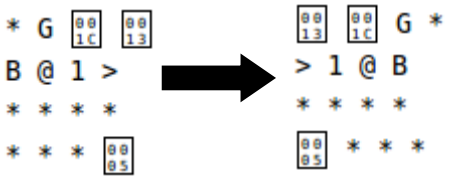
The process for Decryption is all the way opposite of the encryption. Here, first we shift, then we fold, then we substitute and then lastly we transpose.

Algorithm for Decryption-

1. Shift-
   To get back to the original array, we have to first shift the rows again. Now we shift it down by 1 unit. The lower row is added at the top.

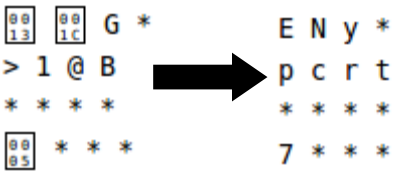


2. Folding-
   This mirrors the image again about the center.



3. Substitute-
   Here we calculate the original values instead of the substituted characters.
   The formula used here is-

   arr(i,j) = (((arr(i,j) - 62)+123)-11)



4. Transpose-
   These are hard coded to get the original form as below.

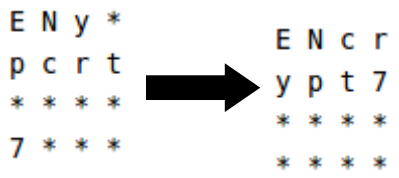
```
int [][]mat2=new int [4][4];
mat2[0][0] = arr[0][0];
mat2[0][1] = arr[0][1];
mat2[1][0] = arr[0][2];
mat2[2][0] = arr[0][3];

mat2[1][1] = arr[1][0];
mat2[0][2] = arr[1][1];
mat2[0][3] = arr[1][2];
mat2[1][2] = arr[1][3];

mat2[2][1] = arr[2][0];
mat2[3][0] = arr[2][1];
mat2[3][1] = arr[2][2];
mat2[2][2] = arr[2][3];

mat2[1][3] = arr[3][0];
mat2[2][3] = arr[3][1];
mat2[3][2] = arr[3][2];
mat2[3][3] = arr[3][3];
```


The conversion is shown below-



5. Now since we have got the original array back again, we can convert it into a string or a character array and display it.

VI. CONCLUSION

The project started with the team finding a suitable database first. The team came across the need to add documents dynamically which can be done using MongoDB. The team also realized that MongoDB does not provide security measures. This was a task that the team worked upon and decided to use encryption for these purposes. The ETSFS algorithm was chosen after studying a number of algorithms. The project was visualized from end to end and then requirements and input-outputs were decided. The project plan was then decided in a team meeting.

VII. ACKNOWLEDGEMENT

We would like to thank all those who somehow have given a contribution to this project and taking this project on the path of success.

VIII. REFERENCES

[1] Yahaya Abd Rahim, *A Secure and Fast Auto Filling Form System*, 2013 IEEE Symposium on Industrial Electronics & Applications (ISIEA2013), September 22-25, 2013, Kuching, Malaysia.

[2] Rishabh Jain., Rahul Jejurkar, *AES Algorithm Using 512 Bit Key Implementation for Secure Communication*. International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 2, Issue 3, March 2014

[3] Ning Oiu,Guoyong Dai ,"*Design and Implement of Online Intelligent Form Filling*" 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2012).

[4] Rakesh Chawla, Puneet Dhiman ,Sudhanshu Prakash Tiwari "*A SSL Based Approach to Enhance Security and QOS in Grid Networks"* 2012 Second International Conference on Advanced Computing & Communication Technologies.

[5] Quist-Aphesti kester *"A Novel Cryptographic Encryption technique for securing Digital Images in the cloud using AES and RGB pixel displacement"* 2013 European Modelling Symposium.

[6] Dongliang Lei, Ke Zhou, Hao Jin, Junping Liu, and Ronglei Wei *SFDS: A Security and Flexible Data Sharing Scheme in Cloud Environment".* 2014 International Conference on Cloud Computing and Big Data.

[7] Boyang Wang, Sherman S.M. Chow, Ming Li, and Hui Li *Storing* Shared Data on the Cloud via Security-Mediator.

[8] Obaida Mohammad Awad Al-Hazaimeh, A NEW APPROACH FOR COMPLEX ENCRYPTING AND DECRYPTING DATA. International Journal of Computer Networks & Communications (IJCNC) Vol.5, No.2, March 2013.

[9] A. Mathur, An ASCII value based data encryption algorithm and its comparison with other symmetric data encryption algorithms. International Journal on Computer Science and Engineering (IJCSE), Vol. 4, pp. 1650-1657, Sep 2012 ISSN: 09753397.

[10] Satyajeet R. Shinde, Rahul Patil ,An Encryption Algorithm Based on ASCII Value of Data, International Journal of Computer Science and Information Technologies, Vol. 5 (6) , 2014, 7232-7234.

[11] Iqra Basharat, Farooque Azam Database Security and Encryption: A Survey Study. International Journal of Computer Applications (0975 888) Volume 47 No.12, June 2012

[12] Kadhem, H.; Amagasa, A Novel Framework for Database Security based on Mixed Cryptography. Internet and Web Applications and Services, 2009. ICIW '09. Fourth International Conference on; Publication Year: 2009,Page(s):163170.

[13] Dr. Anwar Pasha Abdul Gafoor Deshmukh; Transparent Data Encryption-Solution for Security of Database Contents. International Journal of Computer Science and Applications, Vol. 2, March 2011.

[14] Luc Bouganim; Yanli GUO, Database Encryption; Encyclopedia of Cryptography and Security. S. Jajodia and H. van Tilborg (Ed.) 2009.

[15] Tingjian Ge, Stan Zdonik Fast, Secure Encryption for Indexing in a Column Oriented DBMS. 2007 IEEE 23rd International Conference on Data Engineering (2007) Publisher: IEEE, Page(s): 676-685.

[16] Kale Sarika Prakash,P.M.J Prathap,"A Survey on Iceberg Query Evaluation Stratigies",International Journal of Modern Trends in Engineering and Research ,e-ISSN NO.2349-9745,July 2015

[16] https://en.wikipedia.org/wiki/Encryption

[17] http://www.cryptographyworld.com/des.htm

[18] https://en.wikipedia.org/wiki/Advanced_Encryption_Standard

[19] https://www.mongodb.org/

[20] http://www.tutorialspoint.com/mongodb/

[21] https://en.wikipedia.org/wiki/Data_Encryption_Standard

[22] http://www.asciitable.com/

[23] https://www.digicert.com/ssl.htm

[24] https://en.wikipedia.org/wiki/Payment_gateway

[25] Hanan A. Al-Souly, Abeer S. Al-Sheddi, Heba A. Kurdi;Fast, Lightweight Symmetric Encryption Algorithm for Secure Database. *(IJACSA) International Journal of Advanced Computer Science and Applications, Special Issue on Extended Papers from Science and Information Conference 2013.*

[26] https://www.digicert.com/ssl.htm

[27] https://en.wikipedia.org/wiki/SSL

[28] http://searchsecurity.techtarget.com/definition/Transport-Layer-Security-TLS

[29] https://en.wikipedia.org/wiki/Transport_Layer_Security

