Encrypted Data storage in Cassandra with ECC

Mr. Abdul Haseeb
Department of Computer Science & Information Technology
Maulana Azad National Urdu University- Hyderabad
Mrs. GeetaPattun
Assistant Professor
Department of Computer Science & Information Technology
Maulana Azad National Urdu University Hyderabad

Abstract-*When the storage system, servers are used formanagement of large amounts of data, wehave gained increased interest for the advantages such as availability and scalability. A major disadvantage of storage providers being the lack of security features. In this article we have analyzed confidentiality of outsourced data by encrypting all data records before sending them to thestorage system.While traditional databases usually rely onthe SQL model, a lot of alternative approaches, commonly referred to as NoSQL (short for "Not only SQL") databases, also other approaches were in use to meet the new requirements so called "Web 2.0". Cassandra is one among the several existing NoSQL databases whose primary concern is scalability, high availability and no single point of failure for retrieving the stored data. As mentioned earlier, for Cassandra also the security is not a primary concern. Here we proposed a method for storing the data into Cassandra securely using explicit data encryption scheme. This paper gives a detailed overview of Cassandra database and security issues and finally the proposed architecture for storing the data into Cassandra securely using explicit data encryption method.*

Key words- Cloud storage, NoSQL, Cassandra and Cryptography Algorithms.

1.0 Introduction- In today's world e-commerce and social media occupy large portion of web usage, as a result there is a growing need for technology to handle large amount of data related to users. This particular need has been so far meet by using technologies like cloud computing and distributed web applications. But there is a huge draw back in terms of scalability and reliability as it involved very large amount of data. To overcome these many companies have adopted various types of non-relational databases, normally called as NOSQL databases. Different NOSQL databases take different approaches. One such NOSQL database is Cassandra which is developed by Prashant Malik and AvinashLaxman for Facebook. The main purpose of Cassandra is to have high availability with no single point of failure. Cassandra is based on the combination of concepts of Google's Big Table [3].NoSQL database declared, it will not replace relational database. It just provides more options for different scenarios [18].NoSQL databases running in distributed cloud environments were designed to meet those requirements. They provide ease of use and flexibility at low costs, with least consumption of resources for storing and sharing data. Furthermore cloud service providers often provide such storage space which can be booked flexibly on demand.Encryption is always a handy countermeasure in such untrustworthy environments. It can ensure confidentiality of the externally stored data records against any illegitimate read accesses, but it is usually connected to some limitations concerning the interacting possibilities with the encrypted data with the help of some algorithms.Thus in this paper we make the following contributions [2].Cassandra does an excellent job of storing our tweets in a reliable and robust fashion, it is not particularly suited for the task of supporting the random access and display of sequences of tweets [21].

- We quantify the performance of all schemes in a small distributed environment consisting of two nodes.
- We compare the performance of data stores in Cassandra in both the encrypted and the non-encrypted case in a single-server environment.
- The tests executed analyze the two systems when (1) increasing the number of nodes (2) increasing the no of bit size and (3) increasing the total number of operations

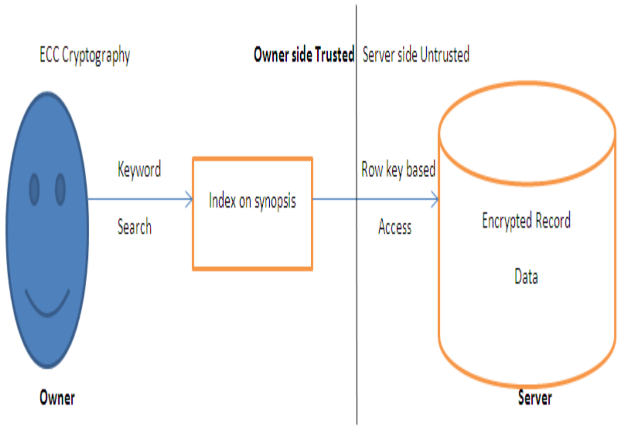

Figure-1 Data outsourcing with encryption and confidential index

If we want to implement security in Cassandra we should aware about some basic terminology like as

2.0Cloud Storage- is a service that allows saving data on offsite storage system managed by third-party and is made accessible by a web services API. There are of two types Block Storage Devices and File Storage Devices now the block storage devices offer raw storage to the clients. This raw storage is partitioned to create volumes. And the file Storage Devices offer storage to clients in the form of files, maintaining its own file system. This storage is in the form of Network Attached Storage (NAS).Cloud computing has been devised as an alternative to intra-company data processing by outsourcing administration tasks to cloud service providers. Individual users take advantage of external processing resources, too, to store and share data [1].

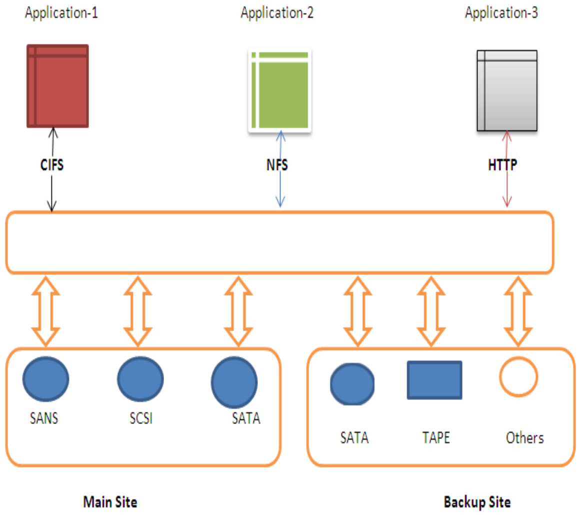

Figure-2 Cloud storage system

3.0NoSQL- database means NOT ONLY SQL where Many companies are now using NoSQL databases to store bulk of their data. These are non-relational databases without full SQL functionality. Usually NoSQL databases have simple and flexible data model and are highly scalable and reliable.Most of the NoSQL databases have been developed for bulk storage and high availability of data and so they often lag behind in incorporating proper security features to protect the data stored in them. But it doesn't mean that we cannot enable security in them. We can make use of third party tools to ensure security of data. One such solution is IBM's InfoSphereGuardium Data Encryption. It sits between the OS file system and the database. An important feature of this tool is that it is completely transparent to the database and the applications and also it can be used in heterogeneous environments to encrypt structured and unstructured data [4].NoSQLs employ asynchronous replication, allowing writes to complete faster and smoother as they are independent of network traffic [15].Big Data demands by means of highly parallel processing on a large number of commodity nodes. NoSQL and NewSQL data stores have emerged as alternatives to Big Data storage [16].

4.0Cassandra- Cassandra is often called a columnar data base [11]. Cassandra is a distributed storage system for managing very large amounts of structured data spread out across many commodity servers, while providing highly available service with no single point of failure. Cassandra aims to run on top of an infrastructure of hundreds of nodes. At this scale, components fail often and Cassandra is designed to survive these failures.Cassandra was designed to support the Inbox search feature of Facebook. As such it can support over 100 million users which use the system continuously [5]. But it has some challenges; one of the most important challenges is how to secure data with encryption technique when we are storing the data and ithas no data encryption and their authentication mechanisms are highly vulnerable in Cassandra data base [6] and it is currently being developed in a top-level project of the Apache Software Foundation [14].Apache Cassandra is a NoSQL-system exhibiting the highest performance at the present time [17].Apache Cassandra NoSQL system aiming to strengthen its consistency while preserving its key distribution mechanism [20]. Cassandra is a distributed system.it loads the network to handle its read/write requests and replication of data across nodes. Therefore, the network for communication between the clusters was constructed independently [22].

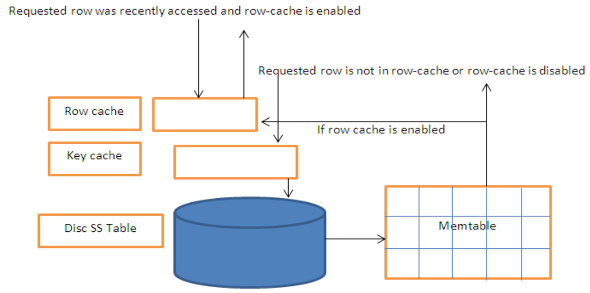
Requested row was recently accessed and row-cache is enabled
Requested row is not in row-cache or row-cache is disabled
Row cache
If row cache is enabled
Key cache
Disc SS Table
Memtable

Figure-3 Cassandra caching

i. with settled length series of bits. This settle length of a string in bits is called Block measure. This square size relies on calculation.

5.0 RSA Algorithm-

RSA stands for Rivest, Shamir and Adelman, who discovered the scheme in 1977. Clifford Cocks had independently discovered this earlier in 1973, but his work was classified and remained unknown for many years.The security level which is given by RSA can be provided even by smaller keys of ECC. For example, the 1024 bit security strength of RSA could be offered by 163 bit security strength of ECC. Other than this, ECC is particularly well suited for wireless communications, like mobile phones, PDAs, smart cards and sensor networks.EC point of multiplication operation is found to be computationally more efficient than RSA exponentiation [10].

6.0 ECC Algorithm-

Elliptic Curve Cryptography (ECC), which was initially proposed by Victor Miller and Neal Koblitz in 1985, is becoming widely known and accepted. The way that the elliptic curve operations are defined is what gives ECC its higher security at smaller key sizes.The elliptic curve is used to define the members of the set over which the group is calculated, as well as the operations between them which define how math works in the group [10].The elliptical curve can be used to obtain the same level of security as RSA-based systems [13].

i. MATHEMATICAL FORMS

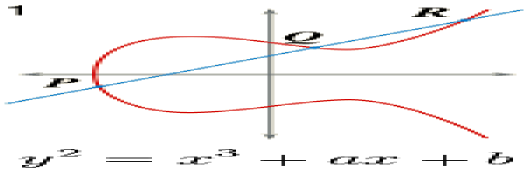

Figure-4 Simple Elliptic curve

Q = d * P
d = The random number that we have selected within the range of (1 to n-1).
P is the point on the curve.
Q is the public key and ‘d’ is the private key.

Encryption- Let ‘m’ is the message that we are sending. We have to represent this message on the curve. This has in-depth implementation details. All the advance research on ECC is done by a company called certicom.

Consider ‘m’ has the point ‘M’ on the curve ‘E’. Randomly select ‘k’ from [1 – (n-1)].

Two cipher texts will be generated let it be C1 and C2.

$C1 = k*P$

$C2 = M + k*Q$

C1 and C2 will be sending.

Decryption- We have to get back the message ‘m’ that was send to us,

$M = C2 – d * C1$

M is the original message that we have send.

How does we get back the message,

$M = C2 – d * C1$

‘M’ can be represented as ‘C2 – d * C1’

$C2 – d * C1 = (M + k * Q) – d * (k * P)$ where $C2 = M + k * Q$ and $C1 = k * P$

$= M + k * d * P – d * k *P$ (canceling out $k * d * P$ and $Q=d*p$)

$= M$ (Original Message)

So finally we will implement Data can be Stored in Encrypted form using this algorithm and also improve performance and secure data to unauthorized users.

a. Point Addition

If P(X1,Yl) and Q(X2,Y2) are points on the elliptic curve and if –X1 ≠ X2 (equally P ≠-Q),then, R(X3,Y3)=P+Q can be defined geometrically, in the case of P≠Q, a line intersecting the curve at points P and Q must also intersect the curve at the third point -R, and R(X3, Y3) is the answer, if P=Q,the tangent line is used[10].

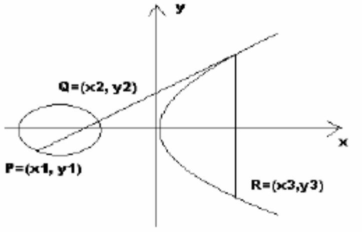

Figure-5 Point of Addition

b. Scalar Multiplication (Point Multiplication)

Point Multiplication (also called scalar multiplication) isdefined by repeated addition. Q=kP=P+P+....+P.(k timesaddition) Elliptic curve discrete logarithm problem (ECDLP),is based on ECC's security and is described as follows. Givenan elliptic and a point on it, to determine k from Q=kP, whereQ and P are points on the curve and kP means P added itself ktimes. It is easy to get Q from k and P, especially for the bignumbers. [10].

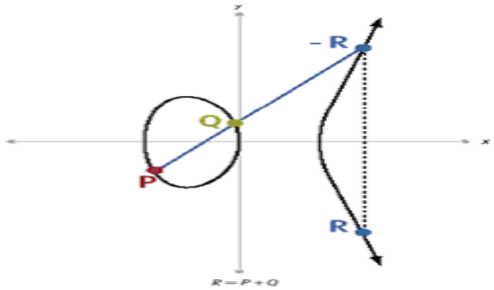

Figure-6 Point of Multiplication

Group law for E/F : $y2 = x3 + ax + b$, char(F) ≠ 2, 3

Point Addition

Let P = (x1, y1) ϵ E(F) and Q = (x2, y2) ϵ E(F), where P ≠ ±Q. Then P + Q = (x3, y3), where $x3 =((y2−y1)/(x2−x1))2− x1−x2$ and $y3 = ((y2 − y1)/(x2 − x1)) (x1−x3) − y1$.

Point Doubling

Let P = (x1, y1) ϵ E(F), where P≠ −P. Then 2P= (x3, y3), where $x3 = ((3x12+ a)/2y1)− 2x1$ and $y3 =((3x12+ a)/2y1))(x1 − x3) − y1$.

Group law for E/F: $y2 + xy = x3 +ax2 +b$, char(F) = 2

ii. **Performance of ECC**

When we use ECC cipher suites can offer significant performance benefits to SSL clients and servers especially as security needs increase [12].The elliptic curve infrastructure will be usable and such infra-structure will provide lower key size, faster access, and high performance to accomplish identity theft [13].

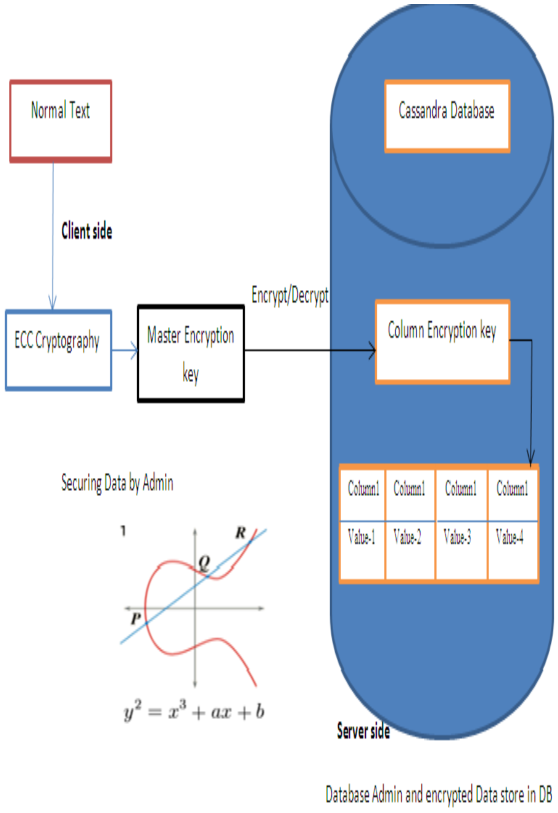

Figure-8 Block Diagram of securing Cassandra DB using ECC Cryptography

iii. COMPARISON

ECC is a kind of public key cryptosystem like RSA. But it differs from RSA in its quicker evolving capacity and by providing attractive and alternative way to researchers of cryptographic algorithm.Comparison between the two asymmetric cryptographic algorithms such as RSA and ECC, same level of security data sizes, encrypted message sizes and computational power. ButECC have smaller keys than other cryptographic algorithms (RSA). ECC offers equal security for a far smaller key size, thereby reducing processing overhead the best known algorithm for solving hard the elliptic curve discrete logarithm problem (ECDLP).It takes full exponential time[10].

ECC Key size bits	RSA key size bits
161	1024
225	2048
268	3072
400	7680
560	15360

Table-1key size for ECC and RSA

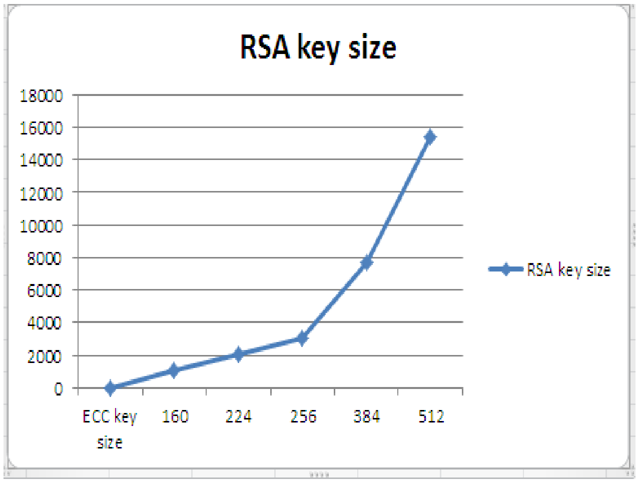

Figure-7 Generation of key graph and its comparision

7.0 Experimental Results

Elliptic curve crypto systems allow to significantly reducing size of the encryption keys. The small key size enables faster execution of various cryptographic operations. According to the literature, it is concluded that RSA key generation takes place substantially slower than elliptic curve based crypto systems of comparable level of security. The results are listed on the table.

Key size bits		Generation Time(seconds)	
ECC	RSA	ECC	RSA
161	1024	0.07	0.16
225	2220	0.17	7.47
268	3072	0.28	9.89
400	7680	0.63	133.90
560	15360	1.42	678.05

Table-2 comparison between ECC and RSA

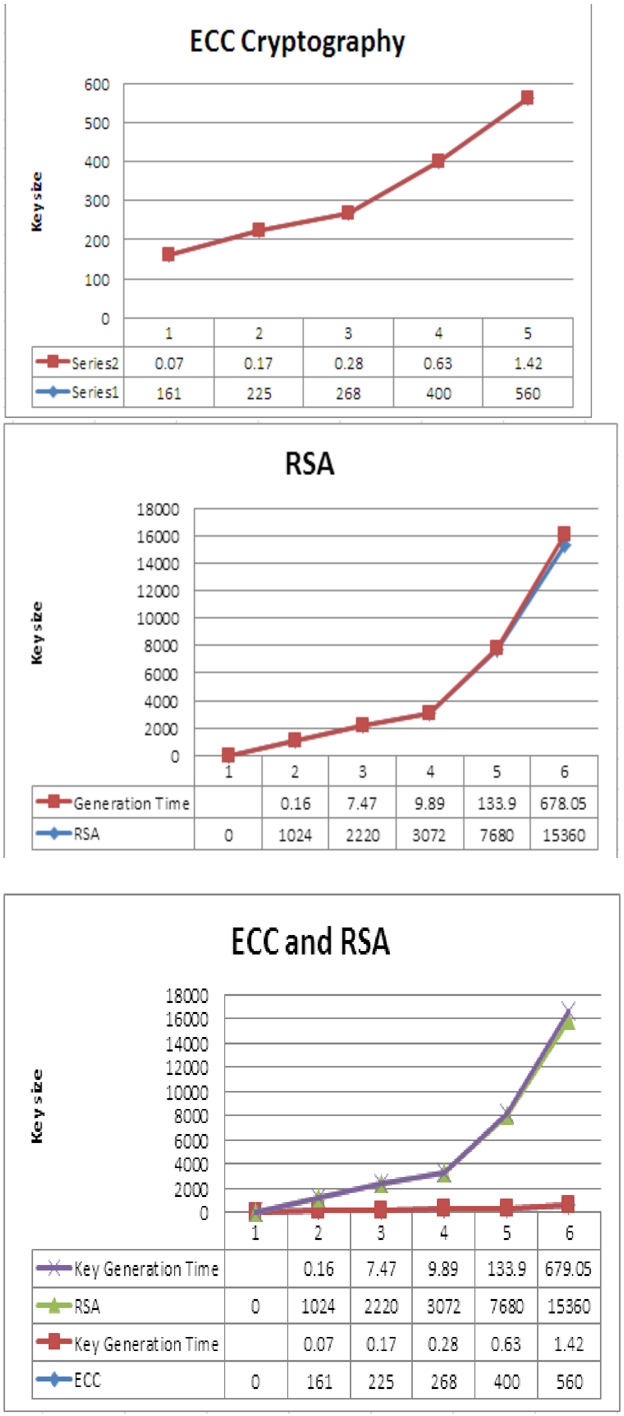

Figure-8 Comparison graph

8.0 Conclusion and future work

This paper focused on ECC, and its mathematical functions and it applications and also we compare between ECC and RSA. Here above table shows that if we use ECC cryptography for encryption Decryption then you have to save lots of time and also more secure in comparison to other cryptography.So we can say that ECC a shorter key length have an advantage in compare to RSA. The performance of the projected is good when using ECC and security and future enhancement for future purpose.

9.0 References

[1] Waage, T., & Wiese, L. (2014, November). Benchmarking Encrypted Data Storage in HBase and Cassandra with YCSB. In International Symposium on Foundations and Practice of Security (pp. 311-325). Springer International Publishing.

[2] Waage, T., Jhajj, R. S., & Wiese, L. (2015, October). Searchable encryption in apache cassandra. In International Symposium on Foundations and Practice of Security (pp. 286-293). Springer International Publishing.

[3] Swamy, R., &Srinivasa, K. (2013). Explicit Data Encryption Architecture for Cassandra. International Journal of Engineering Innovations and Research, 2(4), 376.

[4] Sathyadevan, S., Muraleedharan, N., &Rajan, S. P. (2015). Enhancement of Data Level Security in MongoDB. In Intelligent Distributed Computing (pp. 199-212). Springer International Publishing.

[5] Okman, L., Gal-Oz, N., Gonen, Y., Gudes, E., &Abramov, J. (2011, November). Security issues in nosql databases. In Trust, Security and Privacy in Computing and Communications (TrustCom), 2011 IEEE 10th International Conference on (pp. 541-547). IEEE.

[6] Abramova, V., Bernardino, J., & Furtado, P. (2014, June). Testing cloud benchmark scalability with cassandra. In Services (SERVICES), 2014 IEEE World Congress on (pp. 434-441). IEEE.

[7] Schlesinger, R. (2004, October). A cryptography course for non-mathematicians. In Proceedings of the 1st annual conference on Information security curriculum development (pp. 94-98). ACM.

[8] Magons, K. (2016). Applications and Benefits of Elliptic Curve Cryptography. In SOFSEM (Student Research Forum Papers/Posters) (pp. 32-42).

[9] Vegh, L., &Miclea, L. (2016, June). Secure and efficient communication in cyber-physical systems through cryptography and complex event processing. In Communications (COMM), 2016 International Conference on (pp. 273-276). IEEE.

[10] Prabu, M., &Shanmugalakshmi, R. (2010, February). A study of elliptic curve cryptography and its application. In Proceedings of the International Conference and Workshop on Emerging Trends in Technology (pp. 425-427). ACM.

[11] Butgereit, L. (2016, September). Four NoSQLs in Four Fun Fortnights: Exploring NoSQLs in a Corporate IT Environment. In Proceedings of the Annual Conference of the South African Institute of Computer Scientists and Information Technologists (p. 8). ACM.

[12] Gupta, V., Gupta, S., Chang, S., &Stebila, D. (2002, September). Performance analysis of elliptic curve cryptography for SSL. In Proceedings of the 1st ACM workshop on Wireless security (pp. 87-94). ACM.

[13] Al-Hamdani, W. A. (2011, September). Elliptic curve for data protection. In Proceedings of the 2011

Information Security Curriculum Development Conference (pp. 1-14). ACM.

[14] Miyokawa, S., Tokuda, T., & Yamaguchi, S. (2016, January). Elasticity Improvement of Cassandra. In Proceedings of the 10th International Conference on Ubiquitous Information Management and Communication (p. 37). ACM.

[15] Chandra, D. G. (2015). BASE analysis of NoSQL database. Future Generation Computer Systems, 52, 13-21.

[16] Grolinger, K., Hayes, M., Higashino, W. A., L'Heureux, A., Allison, D. S., &Capretz, M. A. (2014, June). Challenges for mapreduce in big data. In Services (SERVICES), 2014 IEEE World Congress on (pp. 182-189). IEEE.

[17] Kozlov, A. A., Aleshina, A. A., Kamenskikh, I. S., Rovnyagin, M. M., Sinelnikov, D. M., &Shulga, D. A. (2016, February). Increasing the functionality of the modern NoSQL-systems with GPGPU-technology. In NW Russia Young Researchers in Electrical and Electronic Engineering Conference (EIConRusNW), 2016 IEEE (pp. 242-246). IEEE.

[18] Wang, G., & Tang, J. (2012, August). The nosql principles and basic application of cassandra model. In Computer Science & Service System (CSSS), 2012 International Conference on (pp. 1332-1335). IEEE.

[19] Ciriani, V., Vimercati, S. D. C. D., Foresti, S., Jajodia, S., Paraboschi, S., &Samarati, P. (2010). Combining fragmentation and encryption to protect privacy in data storage. ACM Transactions on Information and System Security (TISSEC), 13(3), 22.

[20] Garefalakis, P., Papadopoulos, P., Manousakis, I., &Magoutis, K. (2013, June). Strengthening consistency in the cassandra distributed key-value store. In IFIP International Conference on Distributed Applications and Interoperable Systems (pp. 193-198). Springer Berlin Heidelberg.

[21] Anderson, K. M., Aydin, A. A., Barrenechea, M., Cardenas, A., Hakeem, M., & Jambi, S. (2015, January). Design challenges/solutions for environments supporting the analysis of social media data in crisis informatics research. In System Sciences (HICSS), 2015 48th Hawaii International Conference on (pp. 163-172). IEEE.

[22] Kago, M., & Yamashita, A. (2013, October). Development of a scalable and flexible data logging system using NOSQL databases. In Proceedings of ICALEPCS (p. 532).