IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Smart And Secure Storage System Based On Nosql Databases

Arun Sunder .S
Information Technology
National Engineering College
Kovilpatti ,India

Dr. V.Jackins ,Asst. Professor
Information Technology
National Engineering College
Kovilpatti, India

***Abstract*— Cloud Computing which is emerging as one of the trends in computer world. Nowadays each and every organization is migrating to cloud at a principle "Data Any Where ,At Any Time". However when data grows large in size it needs to stored in an optimized way .When Optimization is considered the data should be converted into table form for easy access and retrieval .While Storing on table form it can be done in Relational Database System. When the data grows exponentially the RDBMS is not suitable. To resolve this problem the NoSQL arises with Document Object Model where we can store any amount of huge data in the form of objects .However some of the cloud functionalities involves the usage of cloud storage without secure due to cost effective. Inorder to Protect the data we have proposed an idea to implement a storage of encrypted data from client side to cloud storage by using Android application as an interface.**

Keywords- NoSql , Google Firebase,Conventional RDBMS

I. INTRODUCTION

Cloud computing has gotten a lot of attention in this new era and has become a new computation mode. Oracle, Microsoft Azure, AWS, and other rapidly growing IT industries They made a difference by establishing a cloud computing server for themselves as well as the rest of the world. When compared to data stored in the cloud, the cost of storing data locally is high. After that, the payment for the service's rent will be made. Previously, they had to purchase and maintain IT infrastructure, which was quite costly. Since then, it has evolved into a cutting-edge technology for storing data in the cloud.

Big Data is a term that refers to a large volume and variety of data that's being collected at a rapid-fire pace from colorful sources. Every time, the volume of data is anticipated to increase by 40, with a44-fold increase between 2009 and 2020 (1). Because so important of this data is unshaped and textual. Since its commencement, the use of Big Data has grown. NoSQL technology is snappily getting popular among Internet organisations. The simplicity of the business design is one of its advantages. vertical scalability and enhanced vacuity control NoSQL databases are getting more popular as a feasible volition to relational databases. As further companies realise the value of relational databases, a data model with smaller schemas is a better way to deal with them.

A. Rdbms Databases

An RDBMS, or relational database operation system, is software that allows druggies to modernize, query, and manage relational databases. Structured Query Language (SQL) is the

most common programming language used to pierce a database. The SQL standard has been modified to allow for the storehouse, reclamation, and publication of JSON data within a relational database, furnishing lesser inflexibility. The need for integrating, managing, and analysing data from multiple sources across on- demesne and pall surroundings is addressed by ultramodern database operation. For illustration, the addition of the object relational model, which is analogous to a relational database, has allowed merchandisers to offer extensions that support data types not included in the SQL standard, similar as time-series data.

B. No-sql and Big Data in Cloud

When people say "NoSQL database," they usually mean anything that isn't a relational database.Some people use the term "NoSQL" to mean "non-SQL," while others use the term "not only SQL." In any case, most people agree that NoSQL databases hold data in a way which is not in relational tables. As storage costs fell dramatically in the late 2000s, NoSQL databases arose. To avoid data duplication, there was no longer any need to build a complicated, difficult-to-manage data model. Because developers have become the primary cost of software development, NoSQL databases were designed with developer productivity in mind, rather than storage.

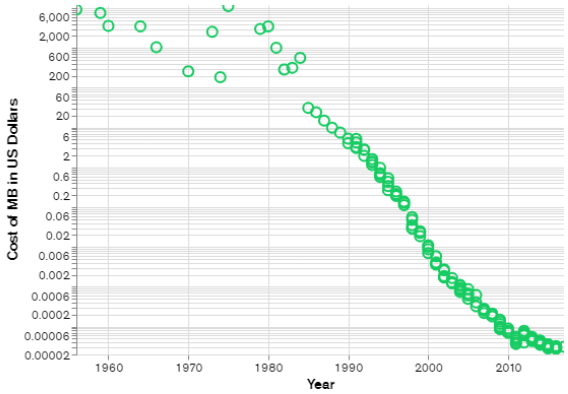

Fig. 1 Cost Analysis of NoSQL storage Based on Mongodb Reference

Big data is no longer just a buzzword in the industry. “Fig 1” clearly shows the cost estimates of NoSQL versus RDBMS storage hypothesis for better data storage utilisation. When the amount of data that needs to be stored grows large, organisations switch to NoSQL databases to achieve greater flexibility in data retrieval. Data is valuable to organisations of all sizes, because it is used to identify challenges which enhances the growth of new opportunities and used to measure performance. It is also becoming increasingly important in Ml for training complex and unstructured data models and facilitating Artificial Intelligence.

II. Related Work

A major concern with RDBs in the age of big data is that modern organizations (using big data systems)are increasingly using NoSQL databases instead. RDB to NoSQL database conversion is significantly challenging because of big data complexity must be handled.Because they have a mathematical foundation, RDB to RDB conversions, such as MSSQL to SQL Server, are simple.As opposed to normalization and general relations,NoSQL databases allow for customised calling process and requests. For RDB-trained workers it is hard to convert from RDB to NoSQL databases. Another challenge for big data companies is analyzing data. RDBs can only carry large datasets, which is post before it is saved. As a consequence, data may be obtained and analysed with ease.

Formless data, on the other hand, is first saved and subsequently processed for analysis. To increase data quality in large data systems, data conventional methods must be applied. Our groundwork will developed an adaptive marketing approach for converting current RDB records to NoSQL systems, as well as data screening activities to support data quality. Data screening is essential to ensure appropriate statistical accuracy.

Innovators must be well-versed in both RDBs and NoSQL systems in order to change them. Rapid population growth from RDB to MongoDB that uses RDB's information and manipulate the data properties.The method is created using descriptive and behavioral characteristics, and a programmed for upgrading from RDB to NoSQL is established. It is suggested to connect Database Server to NoSQL systems.

This strategy lowers the ambiguity of the translation process by squeezing a big and complicated relation system into a solitary HBase table. Another transformation plan is categorized into four categories. We began by converting 1-to-1 and 1-to-many connections into a single HBase data set. Following that, neighboring tables are joined using a recursive technique. In the third phase, the line key is designed, and in the fourth step, perspectives collected through qualitative indexes for needed access patterns are produced.

This type of data system are a brand new variety which is mostly not based on the semantic data model, that does not simulate the information. An outline regarding the storage of NoSQL data and then go through the convention of RDB to NoSQL system. It either have acid or base characteristics, but it possesses its behavior. The major points in the data system are data performs replication, RAM is used for storing enlarge amount of information, contains allotted indexes and also have an additional feature such as add columns to our data system.

The various types of Nosql databases were proposed by Ork de Rooij et al. [3].The Nosql databases can be classified into four types like graph-oriented,wide-colum,key-value and document-oriented.The available databases comparison is performed based on the features and characterstics which are most considered.This study involves about NoSql Database implementation with recent trends and technologies and useful to gain about current tools. The Traditional RDBMS database is compared against NoSql Databases using the key Concepts.It is considered to be has an ability that allow data to be stored and distribute across multiple computing nodes.Due to this abilities ,FAANG companies are using Nosql Databases for their Support on Big volume of data.

On the side , To handle complex data and providing a toolkit for analysing complex data efficiently ,introducing an Architectural support , ,[4] Han, J et al., proposed that Big Data is a solution for complex and heterogenous data . The complex data can consists of structured , unstructured and semi-structured.Still many Organizations which involves the usage of Big Data are having trouble in handling complex data which should migrate to gain multiple usage like handling large amounts of data, realtime interaction and scalability.It presents about an Novel approach for transforming the conventional RDBMS into Big Data Solution.It involves the usage of bi-fold transformation to reduce data complexity and improve analysis of data.

Under the name XDEHD, Bicevska, Z., and Oditis, I. [5] have introduced a method for coupling Heterogenous XML schemas.The coupled schems will produce data without duplication.The technique is compromised of three steps.converting all the schemas to subschema ,which in turn produce an subschema.The final schema produced is called candidate subschema which is used to match the parent data of the heterogenous XML.

According to Chevalier M. [6], et al. the lack of design and conceptual model in various NoSql databases is one of the challenges to maintain the consistency .The scope of this publication is about to propose an uniform conceptual model based on ontologies in NoSql Databases. The concept can be used in No-Sql based DW solutions to implement cube concepts for multi-dimensional data Visualisation. Data Ware Houses(DW) employs this technique because it supports scalable and dynamic data modelling capabilities. While the world is watching the rapid growth of databases in all aspects, the world is also witnessing an information revolution. Databases are interconnected by their content and schema, but they express the same concepts and relations using different elements and structures, which can lead to semantic and structural conflicts.

According to [7]Scabora L.C., et al,there are several advantages such as flexibility and scalability, while adopting an Data WareHousing Environment.For the usage of NoSql Databases three physical data warehouses design is involved using Star Schema Benchmark. . The main focus of their research is on using the MapReduce framework to process OLAP queries over column-oriented databases. Usage of OLAP Queries along with customised columnar databases makes easy to generate data based on user queries. OLAP and DW are the recent technologies today and they've piqued researchers' interest over the last decade. These technologies, on the other hand, were primarily designed for relational database systems in centralised environments. To put it another way, these technologies were not created to be used in scalable systems like NoSQL databases.

Max C. and colleagues [8]have proposed that On the one hand, it examines a variety of issues such as modelling, querying, data loading, and OLAP cuboids.It compares relational data database models with document Object model.It focuses more on OLAP cuboids which has advantage over arrays and nesting .It is proved to be efficient on complex level queries and high workloads.Due to the rise in BigData the NoSql databases are growing faster than the conventional Databases .Moreover it focusses about Document-Object Model whcich is a special class NoSql System.

As a general overview of an architecture for handling Big Data, Cuzzocrea et al. [9] have published a survey that includes about reasoning and resolution in comparison with BigData. Various issues on integrating big data architecture with existing one has been discussed in this survey. First, a definition and characteristics of Big Data are presented. Additionally, the various steps for Big Data data processing are described, along with the most common problems in managing big data.

The use of Online Analytical Processing (OLAP) architecture has been implemented by [10]Thusoo, A., et al.This study includes about the usage of both unstructured and structured data for decision support.Nosql Databases includes more flexibility than RDBMS because it involves the usage key and object pair .It has become an one of the persistent of data storage involving Big Data .To ensure about the recent trends many data models and relational models are proposed to meet the demand for Nosql Databases.

According to Bog, A.,[11] this makes investigating and standardizing such systems extremely difficult. As a result, the usual method of defining a conventional norm with which was before workloads, which has been utilized for years in negotiation and systematic is difficult to apply to Big Data systems. This document summarizes existing benchmarks as well as those in development, compares their characteristics side by side, and discusses their benefits and drawbacks. The main purpose is to gather a clear knowledge of the present status of the Big Data and gave a lecture on the use cases in our daily life. Now a days these techniques and those areas are getting developed in both professional as well as business. These will lead or this will creates a various aspects in the Big Data areas as well as carries a various use cases.

According to Vohra D [12], they used to optimize these parameters based on the large number of tests. To assess the approaches for analyses, choose two major key points: Word Count and Tera Sort. These criteria are used to calculate performance metrics: impalement time, throughput, and quickness. The amount of the given information and the specific criterion have a great impact on the performance is the outcome of our experiments.Is the new specific criterion improve the performance? For this queries become unanswered. For business sector holding enlarge amount of data is the major part in Big Data .Hadoop and the spark make it more interesting in distributed system.Due to the API and its pursuance apark has become very trendy.

Jingqi Zhang et al.,[13] present a hybrid AES-ECC cryptosystem which joins both the asset of the AES and ECC.Plaintext is encrypted with AES in this cryptosystem for faster encryption.Neverthless utilize ECC to encode AES keys ,that will improve the security purpose across many insecure route and it simplifies the key management. Latency is reduced in ECC.

III. Proposed System

The proposed model consists of Firebase password Hashing,AES cipher encryption with hashing technique to encrypt the data from the user side and transferred securely through the encrypted https firewall connections. The different components involved here are Firestore,Realtime Database and Standard encryption of AES which is implemented through Android architecture . Figure 2 shows the model of the system

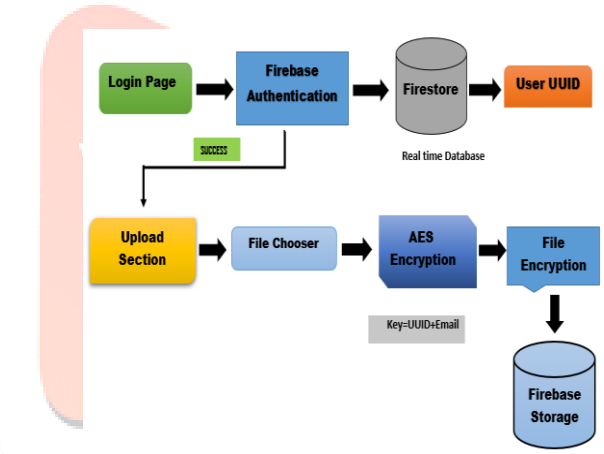

Fig. 2 Block Diagram

A. Aes Encryption

Advanced Encryption Standard (AES) is a block cipher.Most probably every cryptography system follows with key either as public or private.The key size for AES varies as 128,192,256 bits.Every Key has ceartian round of operation based on the number of bits (i.e) Key size.The encryption and decryption of AES is done by shuffling replacing the data using permutation policies.Here we have implemented our AES encryption/decryption using 256 bits which is the highest level of keying.As the 256 bit is large it is not possible to generate a key using brute force method which takes several years to rely on the key.Hence AES 256 is best way to use with combination while encrypting/decrypting.

B. Firebase Password Hashing

Firebase Password hashing is an method of storing the passwords in secure manner which is not visible to the user itself.Each and every password stored for authentication is encrypted using an scrypt technique by employing base signer,number of rounds and hashing parameters.These values are not fixed one and it can be overrided based on the needs.An

modified version of scrypt salt signer is useful to build an secure authentication in an application. The general configuration can be similar to

```
hash_config {
  algorithm: SCRYPT,
  base64_signer_key:
      jxspr8Ki0RYycVU8zykbdLGjFQ3McFUH0uiiTvC8pVMX
      An210wjLNmdZJzxUECKbm0QsEmYUSDzZvpjeJ9WmX
      A==,
  base64_salt_separator: Bw==,
  rounds: 8,
  mem_cost: 14,
}
```

{key} - The signer key parameter for the password hash. The key must be decoded from Base64 to be used.

{salt} - Concatenation of the password salt from the exported account and the salt separator from the application password hash parameters. Every part must be concatenated from Decoded Base64

{rounds} - The rounds parameter is to specify the number of rounds to be applied to the application password hash parameters.

{memcost} - The mem_cost parameter from the application password hash parameters.

[-P] - An optional parameter to read the raw text supplied to application before gets converted to hash.

C. Firestore -Realtime Database

Cloud Firestore is a cloud-hosted, NoSQL database that the apps can access directly via native SDKs.The firestore is used to store NoSql data in pile of document object model. Each and every document consists of key-value pair, which holds the required data. Accessing a record can be done by getting the document object then corresponding value using the key specified.The cloud firestore supports multiple data types unless of RDBMS. Hence firestore is relative realtime NoSql databases to store large collection of data.

The application consists of a login UI which is routed through Firebase by the application's access. Credentials can be generated in a variety of ways, including email authentication, mobile number with OTP, Microsoft account, Github, and so on. In this case, we used Gmail authentication. There are two methods for authenticating user information.

1)The first method is to create login credentials with a common password for all users from the admin panel, where individual email addresses can be loaded via a csv file.

2) The second method is for the user to generate their own credentials. However, it violates application domain security, allowing anyone who is not a member of the organization to access the application data.

When a user is created, a UUID is generated that is unique to that user. The UUID is 28 characters long and is generated to distinguish each user. Even if a user attempts to create credentials using previously registered user details, the attempt will be rejected. The UUID plays such a significant role in authentication validation. Firebase password hashing is a method of storing user passwords in the form of hash values which are not visible to the creator. Scrypt and a Base64 salt are used to compute the hash which is given by,

```
hash_config {
    algorithm: SCRYPT,
    base64_signer_key:
        BxjwgueyroiAGHdowkbd==#$%cdwjkge!@xyzFZX
        CVtyuop[dwbckdjsnwo,
    base64_salt_separator: XA=,
    rounds: 12,
    mem_cost: 14,
  }
```

The credentials are validated using hash encryption and a match with the UUID. A document (key,value pair) is created for each user in the firestore realtime database, with the key being the email address and the value being the UUID in encrypted form. The encrypted UUID's storage purpose is to map the corresponding user with Firebase storage, where all of the users' files are stored. Every user has access to a storage device with a limited capacity and a secure data transfer model. However, the files stored in the storage are visible to administrator if the console is logged in.

To avoid this data breach, data should be encrypted on the user's end before being uploaded to storage. Encryption on Android can be accomplished using the standard AES encryption technique, which employs the encryption AES 256 as it is time efficient to encrypt and decrypt . When the user selects a file to upload, it is encrypted using standard AES encryption with the key generated by the combination of the UUID and the user emailId that is currently signed in and converted into a salt hash. The salt hash is now used as the encryption key, which is stored in the firestore.

When decrypting from the user's perspective, the salted hash stored in the firestore is retrieved using the current user as the decryption key. When decrypting the file, the process is reversed. To reduce network resurrection, decryption is performed at the user's end after retrieving the file from storage, which consumes device resources. This is the model system for creating a secure storage system that has been proposed.

IV. RESULTS

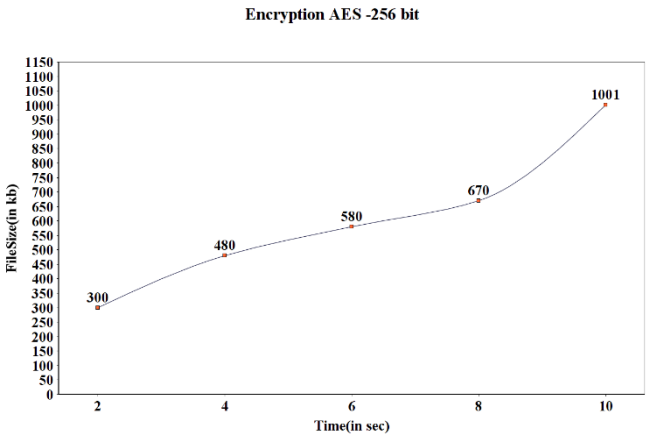

Fig. 3 AES Encryption

The "Fig 3" clearly shows that ,as the size of file increases the time taken to encrypt increases correspondingly.This is due to spectre of the file as the combination of AES and Scrypt involves the 8 rounds of keying which results in increasing the time for encryption.

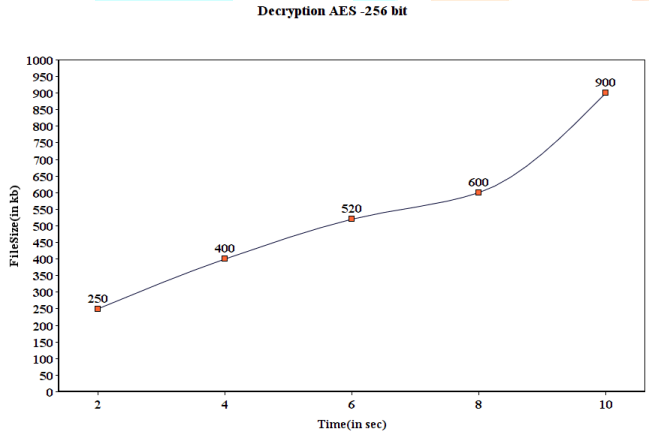

Fig. 4 AES Decryption

Similarly the "Fig 4" shows that the time increases as the file size increases.This is due to the scrypt hashing and AES combination as it involves the multiple rounding technique to decrypt. As the both encryption and decryption are done at the client side which reduces the network bandwidth usage.

V. CONCLUSION

From the above proposed system ,it is an new technique of encrypting an file with an AES and Scrypt Hashing which makes an higher level of security of storing files which is an complete end-to-end encryption from user to server side.As the firebase cloud storage provides secure encrypted data transmission it does not provide an inbuilt security feature to protect user files.In order to minimise the problem our idea has been proposed based on the sample model.Based on this future enhancement done in the area of encrypted cloud storage with higher security functions.

ACKNOWLEDGMENT

First and foremost, I want to convey my heartfelt gratitude. To everyone who assisted us in completing this project. I Want to express my gratitude to Dr.V.Jackins sir for his Patience and advice. His involvement enabled myself to gain a thorough understanding of all facets of project work.

REFERENCES

[1] Mckinsey Global Institute, "Big Data." [Report, 2011] New York: McKinsey Global Institute. IEEE Transactions on Information Forensics and Security, 10(6), 1167– 1179. doi:10.1109/tifs.2015.2400425

[2] Z. Parker, S. Poe and S.V. Vrbsky, "Comparing NoSQL MongoDB to an SQL DB", Proceedings of the 51st ACM Southeast Conference, pp. 5, 2013

[3] Ork de Rooij, Andrei Vishneuski, Maarten de Rijke, xTAS "Text analysis in a timely manner", in: Proceedings of the 12th DutchBelgian Information Retrieval Workshop (DIR 2012), Gent, Belgium,2018.

[4] Han, J.; Haihong, E.; Le, G.; Du, J. "Survey on NoSQL database." In Proceedings of the 2011 Journal on Pervasive Computing and Applications (ICPCA), Port Elizabeth, South Africa, 2018; pp. 363–366.

[5] Bicevska, Z., and Oditis, I. (2018) "Towards NoSQL-based Data Warehouse Solutions". In Research Challenges in Information Science (RCIS), IEEE pp, 104, 104-111.

[6] Chevalier M., Malki M. El, Kopliku A., Teste O. and Tournier R. (2018) "Benchmark for OLAP on NoSQL technologies comparing NoSQL multidimensional data warehousing solution." In Research Challenges in Information Science (RCIS), IEEE, pp. 480-485, Athens.

[7] Scabora L. C., Brito J. J., Ciferri R. R. and Ciferri C. D. A. (2018). Physical Data Warehouse Design on NoSQL Databases. In on Enterprise Information Systems, Lecture Notes in Business Information Processing, Springer, pp. 111-118, Poland.

[8] Max C., El Malki M., Kopliku A., Teste O. and Tournier R.(2016). Document-oriented data warehouses: Models and extended cuboids, extended cuboids in oriented document. In 10th Int. Conf. on Research Challenges in Information Science (RCIS), IEEE, pp. 1-11, Grenoble

[9] A. Cuzzocrea, I.-Y. Song, and K. C. Davis."Analytics over large-scale multidimensional data: The big data revolution"ACM 14th International Publication on Data Warehousing and OLAP, DOLAP '11, pages 101{104, New York, NY, USA, 2018. ACM.

[10] Thusoo, A., Sarma, J. S., Jain, N., Shao, Z., Chakka, P.,Zhang, N., Anthony, S., Liu, H., and Murthy, R. Hive - a petabyte scale data warehouse using hadoop. In 26th ICDE, pages 996–1005,2018.

[11] Bog, A. (2019). Benchmarking Transaction and Analytical Processing Systems: The Creation of a Mixed Workload Benchmark and Its Application. Springer Publishing Company, Incorporated, 1st edition.

[12] Vohra D. Practical Hadoop ecosystem: a defnitive guide to hadoop-related frameworks and tools. California: Apress;2018.

[13] Jingqi Zhang , Wei Gao , Jiakun Li , Xiaoyang Tian , Hua Dang1 High-Speed and High-Security Hybrid AES-ECC

Cryptosystem Based on FPGA. IEEE International Conference on Signal, Information and Data Processing (ICSIDP) - High-Speed and High-Security Hybrid AES-ECC Cryptosystem Based on FPGA. , (), 1–6,2019.