Handling Big Data using NoSQL Database

Mr. M. Praveenkumar[#1],
[#1]Assistant Professor,
Department of Information Technology,
Rathinam Technical Campus, Coimbatore, India.

Mr. S.P.Santhoshkumar[#2],
[#2]Assistant Professor,
Department of Computer Science and Engineering,
Rathinam Technical Campus, Coimbatore, India.

Mr. T. Gowdhaman[#3]
[#3]Assistant Professor,
Department of Computer Science and Engineering,
Rathinam Technical Campus, Coimbatore, India,

Mr. A. Wasim Raja[#4]
[#4]Assistant Professor,
Department of Computer Science and Engineering,
Rathinam Technical Campus, Coimbatore, India,

Abstract-**The traditional relational database is widely used in information management system for structured data is not effective. To handle the unstructured Data, a kind of new technology emerged "NoSQL", which is non-relational database management system for unstructured data. The biggest motivation behind NoSQL is scalability. Healthcare is one of the applications of Big Data. The data collected from every individual provides a high volume of data at high velocity. Analysis of stored healthcare data in terms of querying leads to early detection, diagnosis and also effective drug can be suggested. In order to provide security, the application must explicitly encrypt any sensitive information before writing it to the database. And also, while querying from MongoDB-NoSQL Database-, join operation will be avoided which leads to efficient data retrieval.**

Keywords: Big Data, NoSQL, MongoDB, Security, Healthcare Record

I. INTRODUCTION

Big Data is a term that refers to data sets whose size (volume), complexity (variability), and rate of growth (velocity) make them difficult to be captured, managed, processed or analyzed by conventional technologies and tools, such as relational databases and data processing applications. The size of the data ranges from 30-50 terabytes (10^{12}) to multiple petabytes (10^{15}) is termed as big data.

The complex nature of big data is primarily driven by the unstructured nature of much of the data that is generated by modern technologies, such as that from web logs, Internet transaction, Healthcare, sensors embedded in devices, Internet searches, social networks such as Facebook, portable computers, smart phones and other cell phones, GPS devices, and call center records.

When big data is effectively and efficiently captured, processed, and analyzed, it produce greater insight like finding pattern, deriving meaning, making decision. From this, companies are able to gain a more complete understanding of their business, customers, products, competitors, etc.

A. NOSQL

NoSQL is a non-relational database management system, different from traditional relational database management systems in some significant ways. It is designed for data stores where very large scale of data storing needs (for example Google or Facebook which collects terabits of data every day for their users). This type of data storing may not require well- fixed schema, avoid join operations and generally scale horizontally. NoSQL databases are not built mainly on tables, and generally do not uses SQL for data handling.NoSQL includes a wide variety of different database technologies. They are classified as

- Key-Value Database
- Document Databases
- Wide-Column Database
- Graph Databases.

Key-value store are the simplest NoSQL databases. Every item in the database is stored as an attribute name (or "key"), together with its value. User can request data using key. Examples of key-value stores are Riak and Voldemort. Document database pair each key with a complex data structure known as a document. Documents can contain many different key-value pairs or even nested documents. Wide-column stores such as Cassandra and HBase are optimized for queries over large datasets, and store columns of data together, instead of storing in rows. Graph store database is designed for data whose relations are well represented as a graph.The kind of data could be social relations, public transport links, and network topologies. Graph stores include Neo4J and HyperGraphDB.

The main contribution of the paper is organized as follows

- To Handle Big Data by using NoSQL Database – MongoDB
- To Provide Confidentiality to Sensitive Data
- Analyzing Large Amount of Data

II. RELATED WORKS

Comparison between the relational and non-relational database is shown in [1].The two most extensively used relational databases are MySQL and Oracle. Relational databases do not support high scalability and the data has to fit in predefined tables or Structure. Non-Relational database differs from the relational systems in many significant ways. It doesn't use relations (tables) as its storage structure, it doesn't use SQL as its query language, join operations cannot be performed and the data can be inserted anytime without defining a schema.

The most popular NoSQL databases (MongoDB and Cassandra).The large amounts of user-related sensitive information stored in these databases raise the concern for the confidentiality and privacy of the data [2]. Mongo data-files are unencrypted. Any attacker with access to the file system can directly extract the information from the files. In order to mitigate this, the application must explicitly encrypt any sensitive information before writing it to the database.

The implementation of textbooks management system based on MongoDB [3]. During multi-table query, lot of time is required. To solve this problem, in this paper we attempt to use NoSQL to solve it. Scheme free is one of the characteristics of NoSQL. Two basic information, students' and teachers' information will be created in MongoDB. Subscription textbooks, used textbooks, textbooks storage, entry textbook, delivery textbook, can be set into Separate Operating Collection.

III. PROPOSED SYSTEM

The traditional data models are incapable of handling complex data in the context of Big Data. NoSQL Database is one of tool to handle Big Data is used.

MongoDB, one of the NoSQL Database is a schema-free, document-oriented database and an open source database. It stores data as documents. So it is a document oriented database. A MongoDB database holds a set of collections. A collection is an equivalent term for a table. A collection holds a set of documents. A document is a set of fields. It can be thought of as a row in a collection. Every document has an id. There is no problem in storing documents that do not have the same structure.

The Healthcare Record (HR) is one of the applications of Big Data contains valuable information entered by clinicians. Knowledge Discovery from such HR is a challenging task. HR includes the data generated through the body vitals such as blood pressure, body temperature, lab Reports, prescription are stored in MongoDB. User should explicitly encrypt any the sensitive information before inserting into database by using RSA Algorithms. Analysis of healthcare data in terms of querying often leads to better diagnosis, early detection and also effective drug can be suggested. Analysis often leads to modeling of disease based on clinical documentation. For Example, Physician can Query the number of overweight patient in a database for providing a nutritional counseling services.

IV. MODULES CONSIDRED

The various modules considered in this paper are organized as follows.

B. Identifying Dataset Applications

Big data is generated by number of sources that include social network and media, mobile devices, Internet transaction, Healthcare etc.Healthcare include massive amounts of patient data. It is increasingly challenging to store the variety of structured and unstructured data required, from basic patient information and medical histories to lab results and MRI images. MongoDB is used to store data that provides a view of the patient, doctor,

procedure and other types of information in a single data store.

C. System Implementation.

Using Java program connect the MongoDB database to perform various operation like insertion, encryption of data, deletion, updating and also retrieving data through analysis.

D. Inserting Data into MongoDB

User must be authenticated to database System. MongoDB database use object-oriented thinking to implement system. MongoDB is a document-oriented database developed by 10gen.It manages collections of JSON (JavaScript Object Notation) like documents.NoSQL database have their own terminology that is different from typical relational databases such as MSSQL and MySQL.

Document = Row or Record

Collection = Table or View

Field = column

MongoDB, one of the NoSQL database collection (table) holds one or more documents (records).Each document can have one or more fields (solumn). Create a Database and then create Collection and then insert data in the form of document.

E. Encrypting the Sensitive Data

MongoDB data files are unencrypted, and do not provide a method to automatically encrypt these files. Any attacker with access to the system can directly extract the information from the files. In order to relieve this, the application must explicitly encrypt any sensitive information before writing it to the database. The Sensitive information in healthcare includes patient biodata, problem status. These information should not be disclosed. The Authorized user can enter the data into HR application. The entered data can be stored in encrypted form using RSA Algorithm.

F. Analysis of Data

Analysis of data includes querying and reporting function. The data generated through body vitals, lab reports, prescription should be stored for analysis.

Real time healthcare data refers to data like body temperature, blood pressure, pulse/heart rate and respiratory rate that can be generated every 2-3 seconds. These data are collected from every individual provides a volume of data at high velocity. These data will be stored in MongoDB.

The analysis of these data in term of querying will leads to better diagnosis, early detection of disease and effective drug can be suggested. Analysis may include querying a list of all diabetic patients or a list of all patients within a certain age range.

G. Performance Comparison Between MongoDB and MySQL

In the demonstration of inserting data, MongoDB spends fewer times than MySQL, which also improves query efficiency.
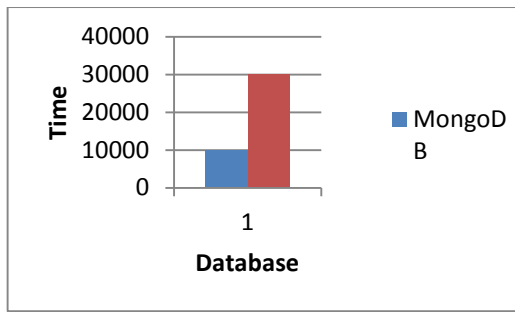
Special Issue - 2017

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCICCNDA - 2017 Conference Proceedings

Figure 1: MongoDB and MySQL Data insert time.

V. ARCHITECTURAL DESIGN

The Architecture diagram shows the how Health care Records are stored in MongoDB (NoSQL Database).

User needs to login to HR application to insert patient records into MongoDB database.To Store Sensitive data user can encrypt data using RSA algorithm and then insert data. Update details of individual patient are maintained to perform analysis for early detection of disease or to discover knowledge based on query retrieval.
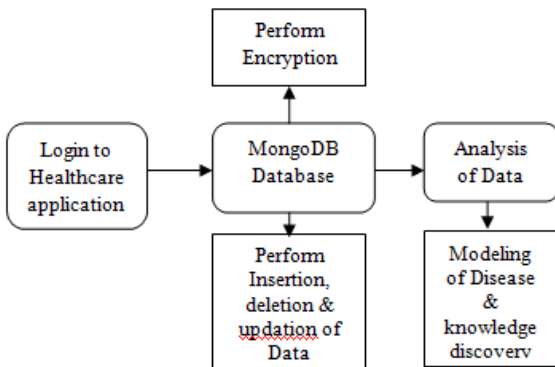
Figure 2: Overall Process in NoSQL Database (HR)

VI. CONCLUSION

Big Data needs a new generation of technologies designed to query a large volumes of a wide variety of data. The Healthcare Record (HR) contains valuable information entered by clinicians. The large volume of Healthcare records is stored in NoSQL - MongoDB Database in the form of document. Analyzing healthcare records stored in MongoDB in terms of query leads to better diagnosis and early detection of disease. MongoDB is more efficient than MySQL in terms of inserting and querying data and also join operation can be avoided. MongoDB is built for scalability, availability and high Performance. MongoDB also provide security for sensitive data by externally encrypting data by using RSA algorithm.

REFERENCES

[1] Carolyn McGregor,University of Ontario Institute of Technology,Canada,"Big Data in Neonatal Intensive Care" , Published by the IEEE Computer Society,2013.

[2] Francisco Cruz , Pedro Gomes and Rui Oliveira, "Assessing NoSQL Databases for Telecom Applications", IEEE Conference on Commerce and Enterprise Computing 2011.

[3] Lior Okman, Nurit Gal-Oz, Yaron Gonen and Ehud Gudes, "Security Issues in NoSQL Databases", International Joint Conference of IEEE TrustCom-11, 2011.

[4] Nishtha Jatana, Sahil Puri, Mehak Ahuja and Ishita Kathuria, "A Survey and Comparison of Relational and Non-Relational Database ", International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 6, August 2012.

[5] Shidong Huang ,Zhenyu Liu,Yun Hu,"Non-structure Data Storage Technology-An Discussion", IEEE/ACIS 11th International Conference on Computer and Information Science 2011.

[6] Zhu Wei-ping and Li Ming-xin," Using MongoDB to Implement Textbook Management System instead of MySQL", IEEE 3rd International Conference on Communication Software and Networks (ICCSN), May 2011.

[7] A Navint Partners White Paper for Big Data, May 2012. www.navint.com

[8] NoSQL – White Paper by couchbase. www.couchbase.com

[9] K. Chitra et al., International Journal of Advanced Research in Computer Science and Software Engineering 3(7), July - 2013, pp. 1356-1360.

[10] www.MongoDB.com Provide Overview of MongoDB.