International Journal of Pure and Applied Mathematics
Volume 117 No. 21 2017, 177-197
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue

ijpam.eu

An Analysis and Performance Evaluation of NOSQL Databases for Efficient Data Management in E-Health Clouds

[1]M.P.Gopinath,[2]G.S. Tamilzharasi, [3]S.L.Aarthy and[4]R.Mohanasundram

[1]School of Computer Science and Engineering,

VIT University, Vellore.

mpgopinath@vit.ac.in

[2]School of Computer Science and Engineering,

VIT University, Vellore.

[3]School of Information Technology and Engineering,

VIT University, Vellore.

[2]School of Computer Science and Engineering,

VIT University, Vellore.

Abstract

E-health cloud offers electronic health care services across the internet. In such type of systems the patients' health data is collected from the Body Area Networks (BAN), then it is stored, processed and analysed under cloud computing infrastructures. The data generated from the BAN networks are highly dynamic and vast in nature as it continuously monitors the patients' health conditions. At present, there exist several database systems to deal with the e-health applications but the one that better suits the scaling demands of E-health clouds still remains to be undetermined. In order to solve this issue, in this paper, a clear analysis and performance evaluation of NoSQL databases over E-health clouds is presented. The major contribution of the project is listed as follows, Find and analyse the advantages and disadvantages of the NoSQL databases with respect to the E-health clouds. Derive metrics to evaluate the performance of various NoSQL databases that deploys e-health applications. Benchmarking various NoSQL databases like MongoDB, Cassandra, and Hbase.Evaluating the better among NoSQL databases that suit the needs of E-health clouds.

KeyWords:E-health clouds, NoSQL databases, relational databases, distributed systems.

1. Introduction

E-health is an emerging technique offers medical informatics and healthcare services through the internet in an electronic format [1]. It adopts modern communication and data analytics approaches to meet the on-demand requirements of patients, healthcare professionals, and all other E-health service consumers. The term E-health has acquired greater importance in today's scenario as it adopts information and communication techniques to the favor of health care services in a cost-effective and efficient manner. In e-health clouds, the patient's health data is monitored, tracked and recorded in the form Electronic Health Records (EHR) over the cloud computing environment. EHR is the digital form of the patient's health records makes the patient's data instantly available to the authorized users [2, 3]. The body sensors are projected in and around the patient's body. BAN(Body Area Networks) monitors and collects the patient's health information through the body sensors and stores into the cloud server. The health service provider or the CSP further process the data and shares it among authorized data users. The data collected from the BAN networks are highly dynamic and vast in nature as it monitors patient's health information's in a continuous manner [4,5].

Data gathered from the BAN networks are categorized into three types such as structured, semi-structured and unstructured data. The information's such as patient's name, age, identification number, blood pressure level, glucose level and diagnosis codes comes under the category of structured data as it can be easily stored in traditional databases. The structured data contains a standard format, and it can be directly stored in the relational database systems. The data without any clear structure is called the unstructured data. Documents, image, video and audio files are some of the examples of the unstructured data. The semi-structured data cannot be directly stored in the relational database systems, but it has some organizational properties using which it can be easily analysed. Sensor data is the best example of the semi-structured data as it continuously monitors the patient's health diagnostics then collects and stores it for data analysis processes. Though the data generated from BAN networks is a combination of structured, semi-structured and unstructured data, the majority of the data from BAN networks are in semi-structured format. Further, 80\% of the EHR's are semi-structured as it is collected from the sensor networks. Due to the enormous growth of EHR an efficient database system to manage, the EHR has become the essential requirement [6,7].

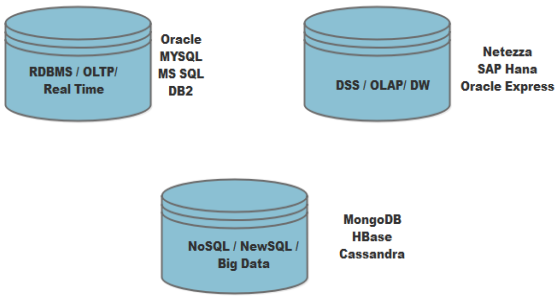

Figure 1: Evolution of NoSQL Databases

In traditional E-health management systems, the health data is stored on the relational or an object-relational database system. Relational Database Management System (RDBMS) is the basis for all the modern database systems and works by E.F.Codd relational model. It stores and retrieves every individual patient records in a separate manner. EHR encompasses a broad range of the data generated from distinct sources such as mobile devices, sensor networks, health repositories, and applications. It may even vary from the terabyte to the petabytes of the data. The traditional database systems such as Oracle, IBM, MYSQL and Microsoft Access adopts structured query language to perform create, insert, delete and update operations across database systems. However, the increased volume and velocity of the E-health data makes its application obsolete across E-health clouds. The major drawback of traditional RDBMS systems is it fails to handle the huge volume of the EHR as it ranges gigabytes to petabytes in size. This lead to the addition of more memory, and central processing unit (CPU) to the database system results in decreased scale-up and speed up measures with increased cost. Further, the majority of the health data are in the semi-structured form it cannot be maintained adequately by the traditional database systems. Also the RDBMS is designed to handle steady data retention it cannot withstand the rapid growth of the EHR. The inability of the traditional database systems to deal with the rapid growth of EHR led to the emergence of NoSQL database systems [8,9]. The evolution of the database systems are clearly illustrated in figure 1.

To deal with the semi-structured and unstructured queries at a faster rate, a newer database system called distributed database system has aroused. In a distributed database system the larger databases are divided and disseminated across multiple servers. This technique of distributed database systems led to the arrival of advanced database systems such as NoSQL databases. The term NoSQL is Not Only SQL designed to deal with semi-structured data contents [10, 11]. It acts as efficient database systems for cloud dependent applications. It does not adopt any query languages and works by JSON. The use of the NOSQL databases across E-health clouds has become prevalent as it scales well with cloud computing platform. Further, it provides increased speed and a higher degree of fault tolerance measures. There exists a variety of

NoSQLdatabase such as MongoDB, Cassandra, couch base, Hbase, etc. Its, utility varies from one application to another. Thus in this paper, a detailed analysis of NoSQL databases concerning the E-health cloud is given. First, the applicability of the NOSQL databases to the E-health clouds are discussed on its pros and cons. Next, the metrics are derived to evaluate the performance of the NoSQL databases across E-health clouds. Further, the various NoSQL databases are benchmarked under a set of standard constraints and the NoSQL databases that better suits the need of E-health clouds are identified and suggested for its application across E-health clouds.

2. Related Works

The research works related to the proposed system is discussed in this section. E-health clouds monitor patient's health status through the use of sensor networks. A set of body sensors is implanted in and around the patient's body that monitors the patient health status. The data collected from the sensor networks are highly dynamic and continuous in nature. This sensor data is stored in the form of Electronic Health Records(EHR) across the cloud server. The EHR's are then processed and analysed for medical and research purposes [12,13]. The cost effective and scalable nature of the cloud computing services enables the deployment of E-health services across the cloud computing environment [15,16]. Body Area Networks (BAN) [17,18,19] is a collection of advanced nano and micro technology components that improves the accuracy and speed of the recorded data. It possesses sensors and actuators that monitor and logs the patient's health data. However, the BAN generates an enormous amount of EHR's thus the process of storage and management of the EHR across the cloud computing environment is found to be the challenging factor [22,23].

The termNoSQL databases [24] also known as non-relational databases manages larger datasets with no single point of failure. NoSQL databases often depend on horizontal scalability that enables the performance measures of the system as it increases the computing power of the single node rather than increasing the number of nodes. There exist around 150 different types of NoSQL databases and they are grouped into four major categories such as Key-value store, document store, and column family and graph databases [28]. In Key-value store databases, all the data is stored in key-value pairs. Dynamo DB and Azure Table Storage are some of the examples of Key-value store databases. Document store databases are used to store and process the data contents in a document format. The well-known example for the document store is the MongoDB and couch base. Column family databases store the data across the columns, and it possesses an infinite number of the columns, which is also organized as a group. Apache Casandra and Hbase are the examples of column family databases. The graph databases stores and manages graph kind of data such as social network relations. Neo 4j is the best example of the graph database. The tremendous increase of big data across cloud computing enables

the adoption of NoSQL databases for storage and analytical purposes. Further, the adoption of NoSQL databases across big data applications provides improved performance measures [27,28]. An application of NoSQL databases to the e-health systems was given in [25]. This approach improves the performance and scalability measures through the combination of RDBMS and NoSQL databases for distributed storage systems. It is specifically designed for the applications of Radio Information System (RIS) and E-health Information Systems(HIS) as it comprises of a large amount of structured, semi-structured and unstructured data. This system model is deployed across the hybrid cloud environment. Followed by this work another application of NoSQL databases for E-Health system was given in [26]. It presents a review of EHRs across NoSQL databases. It further evaluates the suitability of NoSQL databases across e-health care systems. Even though there exist several NoSQL databases for big data analytics its adoption strategy widely differs from one application to another.

Scalability, availability, consistency, request and response time are some of the important metrics and benchmarks to evaluate the performance of NoSQL database systems [29]. An experimental evaluation of NoSQL databases was given in [30]. This work compares the performance of the various NoSQL databases by the storage and retrieval time metrics. It further evaluates the performance measures of the NoSQL databases concerning the cloud computing systems. Some of the benchmarks to assess the performance measures of the NoSQL databases to the relational database is given in [31]. This work examines the performance of NoSQL databases such as MongoDB, PostgreSQL, and SQLLite3 by the workload benchmarks such as a number of messages inserted, the size of the messages and number of topics inserted. It categorizes the performance of the three databases with respect to robotics logging applications. However, the metrics and benchmarks defined by this work form the basis for the other big data applications. In [32] an analysis of various NoSQL databases and its performance measure are described in detail. The performance measures of widely adopted database such as HBase, Cassandra, MongoDB, and Redis was comparatively evaluated. It applies YCSB benchmark tool for the process of performance evaluation. This work evaluates the performance of the database using transaction processing time metrics. Thus through the adoption of standard benchmarks and metrics performance measure of the NoSQL databases are quickly evaluated.

It is clearly revealed from the literature that there exist several NoSQL databases for big data analytics processes. However, the adoption of the database system varies from one application to another as each system has its own requirements. Further, the database system is found to be efficient only when it satisfies certain performance metrics and benchmarks. At present, there exist no standard database systems that better suits the needs of E-health cloud application. Further, there is also no proper benchmarks and metrics to evaluate the performance measures of the E-health database systems. Thus, in this work,

various metrics and benchmarks for the e-health systems are framed, and the performance measures of the various NoSQL databases is evaluated. Through this, the database that well suits the needs of E-health applications is analysed and identified.

3. Proposed Method

NoSQL Databases and its Application to E-Health Clouds

The challenges behind the traditional RDBS systems gave rise to the evolution of NoSQL databases across the E-health clouds. The NoSQL databases are schema-less data model supports scalable replication and distribution. The shared nothing architecture of the NoSQL databases enables it to run across a large number of nodes and provides higher performance per node in comparison to the traditional database systems. Further, the NoSQL databases possess non-locking concurrency control mechanisms such that the there is no conflicts between the real-time read and write operations. Thus, NoSQL database systems preserve consistency property. Whereas, in the traditional systems consistency become the bottleneck when it deals with the property of scalability. A clear description to the e-health architecture interms of the NoSQL database perspective is given in figure 2.

As there exists a variety of NoSQL databases, the selection of most appropriate one to the E-health clouds remains to be a difficult process. Use of right NoSQL database for right applications provides improved performance. Selection of right data model, analysis of pros and cons of consistency and identification of compromising features of the RDBMS forms the three criteria's assists in appropriate NoSQL database selection for the E-health clouds.In general the NoSQL databases falls into four major categories such as document store, key value, graph store and column family stores.

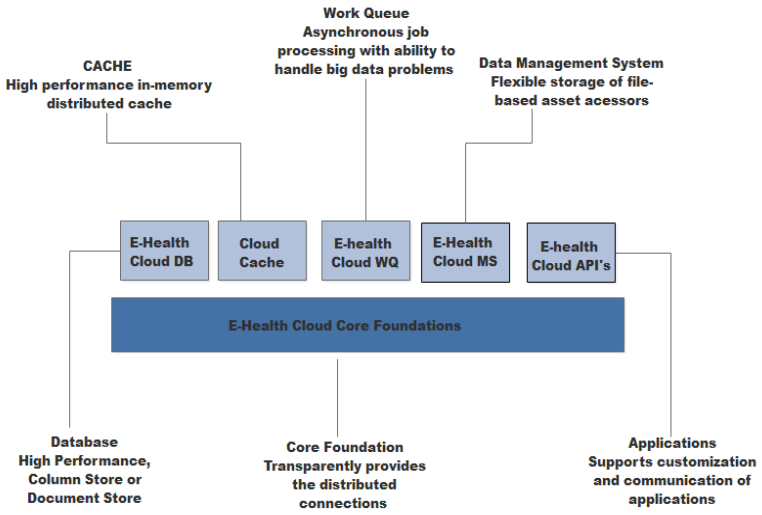

Figure 2: E-Health Cloud Architecture in a Database Perspective

Key-Value Data Stores

It is a simple form of NoSQL database systems makes use of dictionary-like data structures. It provides data access through keys, which acts as a unique identifier to the data contents. It maps every attribute with a separate key, and each key represents a value. The Value, in turn, represents the set of data. It does not adopt any structured query languages and offers greater flexibility. In a key-value, data store client can get, add or delete a value to the key. In general, it stores and retrieves data using key\-value pairs. For example, to represent the treatments undergone by a particular patient a key k1 can be mapped to the value of the treatment T1. Similarly, the values k2 and k3 are be mapped to the treatment T2 and T3 of the patient P1. Some of the well-known examples of the Key-value data stores include Redis, Riak, Couch base, Memcached, BerkeleyDB, and upscaledb. The key-value data stores are not popular among the E-health applications because of the following reasons:

- It does not support foreign key references. Since health data is mostly used for data analytics purposes references from one relation to another has become mandatory for patient health diagnosis purposes.
- Implementation of SQL datatypes to validate database entries is highly impossible in key-value data stores. Since EHR are highly sensitive in nature, it requires validation of the data types and the data elements.
- In E-health applications the data contents are highly interrelated needs to be fetched from multiple rows of the different relations or tables. Fetching this sort of results are complex and expensive in key-value data stores as it requires join operations.

Because of this reasons, key-value stores are not much widespread among E-health applications.

Document Databases

Document databases stores multiple attributes in a single document rather than storing every attribute with a key. Upon the addition of the documents, it builds the required data structures to support the document. These database systems are highly flexible and vary from the traditional RDBMS systems. In general, the document databases make use of JSON (JavaScript Object Notation) and XML (Extensible Markup Language) for querying purposes. In document database systems, a document may contain multiple documents embedded and a list of multiple values within it. The major advantage of these databases are it supports querying with various attributes. For example, to cluster the list of patients with low blood pressure. First, the identifier to list the patients with low blood pressure are stored in a document, and within the document, all the patient documents with low blood pressure are embedded. The identifiers assist in easy reference to the patient's attributes. Thus the document databases enable easier classification and clustering of the EHR's. Further it stores and combines

data of any structures without affecting data access time and indexing functions. Also, it is most suitable for the in-depth analytical processes, which is the most important requirement of the E-health cloud systems. These features enable the application of document databases across E-health cloud systems. MongoDB, terastore, CouchDB, orientDB and Raven DB are some of the examples of the document databases.

Column Family Stores

In a column family data stores, data is stored in the form of columns, and a set of columns forms the row. A row can contain n number of columns corresponding to it. Column family represent the group of related data contents, and it can be accessed together. In column family stores a key identifies a row and a row can have multiple columns. In a column family stores, it is not necessary for all the rows to have the same columns and a column can be added to a row at any time without affecting other values. It is designed for rows with many columns and can even handle millions of columns. This property makes it application most suitable across E-health clouds. Because in E-health clouds frequently accessed data can be grouped together as a single row and it is not necessary for all the rows to possess the same column. Further, the EHR's are dynamic in nature, and it can be added as columns into the column family stores without affecting other rows of the data store. Also, the vast amount of EHR's can be stored as columns in a column family data stores in a most efficient manner. For example, Patient name and ID can frequently be used together thus these two columns forms a row. Similarly, patient disease and medication details are often used together hence these two columns form a row, and all these are grouped into collection called Patient Diagnosis (column family store).The column family stores form the most suitable database for E-health clouds as it possesses the ability to handle a huge volume of EHR's in an efficient manner. Further, it is simple and effective to use in a real time scenario. Cassandra, HBase, Hypertable and Amazon DynamoDB are some of the examples of the column family stores.

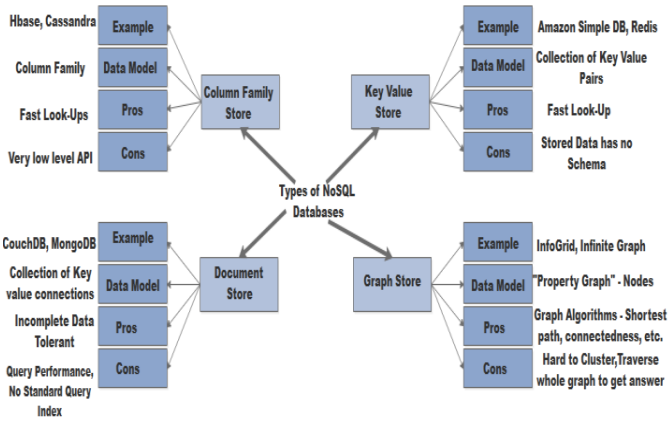

Figure 3: Summary of NoSQL Database Systems

Graph Stores

The graph data stores work by nodes and relationship. Nodes represent an object with the set of identifiers and relationship defines the link between two nodes. Its applications are prevalent across social media systems such as Facebook and Twitter. Nodes and relationship may sometimes form a complex structure. Whereas, E-health application requires simple structures for data analytics and medication diagnosis processes. Further, querying is complex across graph stores as it requires multiple SQL statements or a recursive statement to find paths. Also, the identification of the relationship between nodes is a time-consuming process. The graph stores are not much prevalent across E-health clouds as it takes more computational time and may sometimes result in reduced system performance when applied to E-health cloud scenario. The best example for the graph data stores is Neo4j, infinite graph and Orient DB.A summary on different types of NoSQL databases with its advantages and disadvantages are clearly illustrated in the figure 3.

Significance of NoSQL Databases with Respect to its Applications

In general NoSQL databases are used across the organizations under the following constraints,

- To enhance developer's productivity through the utilization of an application that better suits the requirement of the application.
- To provide improved data access performance measures. Thus the system can handle a large volume of data with improved throughput and latency.

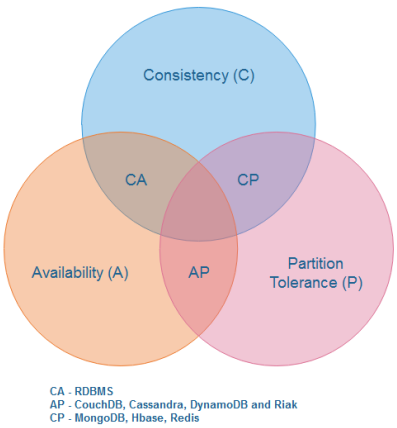

Figure 4: NoSQL Databases with Respect to CAP Properties

The adoption of NoSQL databases is highly influenced by the organizational or the application requirements. From the discussions at the above subsections, we can easily conclude that document data stores and column family stores are the most suitable NoSQL databases for the E\-health clouds. It is well-known that consistency, availability and partition-tolerance form the three basic requirements of the NoSQL database systems. Practically it is impossible for the

database systems to fulfil all the three requirements. At present, the NoSQL databases follow the combination of the CAP requirements. Some databases may be consistent and available, but it does not support the property of partition tolerance. Similarly, few databases can support consistency and partition tolerance but do not support availability. Though the document and column family data stores are identified to be the suitable No-SQL databases for the E-health clouds. There exists a variety of document and column family data stores. To derive the performance metrics and to benchmark the NoSQL databases the widely adopted and popular databases such as MongoDB, Cassandra and Hbase are taken into consideration. Figure 4 illustrates the fulfilment of CAP properties by some of the well-known database systems.Since there exists a variety of NoSQL databases a summary on NoSQL databases and its applications are given as follows:

Key-Value Stores: The key-value data stores are widely used across web applications such as online shopping management systems, user profile maintenance and session management systems. It is not suitable for the applications where the data contents are highly related to each other and require querying by data. Further, it does not operate on multiple keys at a time. Since the EHR's are highly related to each other Key-value data stores are not suitable for E-health cloud environment.

Document Data Stores: It is generally used for real-time data analytics processes and content management systems. It is simple and easy to use, but it does not support complex transaction systems. The simple and efficient use of document data stores makes its application feasible across e-health cloud systems.

Column Family Stores: This data stores are most suitable for applications that have a large number of write operations and aggregation. However, these data stores are not appropriate when the system requires earlier development and changing query pattern. However, the ability of the column family stores to handle an enormous volume of information in an efficient manner makes its application prevalent among E-health cloud systems.

Graph Data Stores: It is most appropriate for connected networks, social networking sites and for the applications that require routing information and recommendation engines.

Evaluation of NoSQL Databases in a E-Health Cloud Perspective

There exist around 150 different types of NoSQL databases, but its use cases highly differ from one application to another. In this section, we evaluate three major databases such as MongoDB, Cassandra, and Hbase on the E-health cloud environment. The performance measure of the NoSQL databases can be made in two different ways such qualitative and quantitative methods. In a qualitative method some of the quality attributes are taken into considerations, and on the quantitative approach, the performance is analyzed regarding

increasing workloads and operations. Some of the quality attributes to evaluate NoSQL databases in a qualitative way are listed as follows:

- Scalability, how efficiently the database system deals with the increasing demands of the E-Health clouds.
- Simplicity, the database should be simple and easier to deal with EHR's.
- Availability, the database should be available all times as the EHR's are continuous and dynamic in nature.
- Flexibility, the database should be flexible enough to meet with the emerging features of the E-health applications.
- Durability, the database should be durable such that the changes made to the EHR's must be reflected in the database in a permanent manner.
- Query Expressiveness, the expressive of the queries in different programming languages.
- Most suitable data models and data types. The data model and data types should be appropriate to deal with EHR's.
- Support to indexing. The database must support possess features to support efficient EHR management.
- Cost Effectiveness, the database should be affordable and cost effective to deal with E-Health cloud environment.

Concerning qualitative analysis, Cassandra supports availability, eventual consistency and scalability measures from figure 4. The consistent nature of the Cassandra improves the latency measures. Further, it requires less expense for establishment and management purposes. Cassandra supports decentralized master to master architecture this prevents a single point of failure and maintains EHR in an efficient manner. Further, it supports faster read and write operations with improved latency measures. Whereas MongoDB supports high availability and scalability as it supports sharding and master-slave replication. Sharding assists in efficient EHR management as their comparatively larger in nature. Further, it supports rich query language thus complex, and ad-hoc EHR's are managed in an efficient manner. A summary of qualitative analysis of the three major databases are given in the table 1.

To benchmark various NoSQL databases (quantitive measure) for E-health clouds, we require a set of metrics to evaluate the performance measures. Since the process of data access provision is the major task across E-health clouds the metrics are derived with the read, write and update operations. Since we consider read, write and update operations to evaluate the system performance, operational latency forms the most important metric. Hence the tests should be designed to demonstrate how the latency varies at different scenarios. Some of the important performance metrics with respect to the latency measures are listed as follows:

- Data import performance, Operational latency measures for the different workload (operations per second) and throughput operations per second.
- Read performance, latency and throughput measures achieved during read operations.

- Write performance, latency and throughput measures concerned with write and update operations.
- Operation latency measures for a different mix of operations and workloads.
- Operational latency measures for varying key clusters.

The above mentioned are the most important metrics to evaluate the database performance. Also, there exists some additional metrics given as follows:

- Elastic speedup, the extent to which the addition of servers affects the operational latency measures.
- Scalability, the extent to which the existence of more or lesser nodes affects operational latency.
- Fault tolerance, the extent to which the random failure of the system affects the operational latency measures.
- Load balancing, How efficiently the database system balance the load across with various servers and workloads.
- The level of an extent to which the choice of cloud infrastructure instance type affects the system performance. Example: Amazon EC2.
- Storage consistency (number of threads and operations per seconds).
- Eventual consistency, availability, and durability.

These metrics forms the basic requirement to benchmark E-health NoSQL database systems. In addition, Latency and throughput are the most frequently used metrics to evaluate the system performance.Throughout this paper these two performance metrics are adopted to benchmark various NoSQL databases such as MongoDB, Cassandra and Hbase.

4. Results and Discussions

To evaluate the performance measures the databases are connected to the benchmarking tools and tested using different scenarios. The test scenario follows the performance metrics defined at the previous section.

Evaluation Setup

MongoDB version 3.3, Cassandra 3.0 and Habase 1.0 are the three databases we tested. The tests are implemented in an open source cloud platform Amazon Web Services(AWS). The system configurations include the 16 GB RAM, Intel Xenon processor 2.20GHz with 4 virtual processors and 15 cores in a high-performance network. Ubuntu 16.0 operating system is used. The load tests are performed at two database server configurations. The databases are deployed on a single node to evaluate the performance of the single server and at three nodes to measure multi-node performance. HL7 Fast Healthcare Interoperability Resources (FHIR) (http://www.hl7.org/implement/standards/fhir/) is used for system prototyping. The data model contains patient's information's such as patient name, body weight, blood pressure, etc. A synthetic dataset is used for the testing purpose, and it contains one million records with 2.5 million patient

diagnostic result records. The Yahoo Cloud Serving Benchmark (YCSB) tool is used to evaluate performance and benchmark the databases. YCSB has default data models, and workloads for test execution and it is modified in accordance to the E\-Health use case scenario. The workloads are described in terms of the operations performed across the records (read, write and update).

Benchmarking the NoSQL Databases for E-Health Clouds

The first stage of the test requires the import of the dataset into the data stores. During this state around 100,000,000 records with each 1kb size are imported into the data store. Through the use of the YCSB, the throughput (threads per node) and operational latency in the millisecond are compared. This includes scenarios, where the data collected from the BAN networks are incorporated into the data stores. From the observation, it is clearly identified that during the data import phase Cassandra provides the highest performance, Hbase with lowest performance measures and MongoDB remains nearer to the Cassandra. In an average, Cassandra provides a latency measure of 0.5 seconds to insert records across 12 threads,MongoDB takes around 0.6 seconds and Hbase with 0.8 seconds.

Next, the throughput and latency measures of the three databases are measured on the read operations. This kind of workload is given to the E\-Health clouds when the data stored across the E-health clouds are accessed by various data users. The read operations are distributed across 1 to 16 nodes (threads). As a result of the observation, Cassandra and Hbaseprovides improved read latency measures, but the performance degrades with the increased number of operations per thread. Whereas, MongoDB provides consistent measures with higher latency measure.

Next, a workload with 50% read and 50% update operations are equally distributed across the databases. In this case, Hbase produces the consistent performance measures. MongoDB and Cassandra performance measures degrade with increasing write operations per second. The results are inconsistent because the read and the write operations are distributed in a random manner. The difference between latencies varies around the average of 20 to 30 milliseconds. Next, a workload of 5% update and 95% read operations are given across the data stores. In this scenario, the Cassandra provides the lowest performance with the latency of 90ms. Hbase provides the highest performance measures with the latency of 45ms, and the MongoDB provides consistent performance measure with the latency of 60ms.

A workload with 5% insert and 95% read are given across the data stores. In this scenario, Cassandra provides higher performance with a latency of around 10ms. But its performance degrades with the execution of the operations across 10 nodes. Hbase provides a maximum throughput around 7 nodes with a latency of 40ms and MongoDB provides consistent performance measures with higher latency.

Next, a complex read, write and update operations are given. In this scenario, Cassandra provides higher performance with the lesser throughout around 8 nodes. Hbase achieves standard performance measure with increased latency of 60ms. MongoDB provides lesser performance with higher latency measure of 70ms. It performance degrades with complex read, write and update operations.

Next, a workload with 90% insert and 10% read operations are given to the data stores. This includes real-time scenarios such as a large amount of the EHR's are inserted into the cloud systems. As the result of the operation, Hbase and Cassandra provides lower latency and higher throughput measures. The performance of the MongoDB degrades with the increased number of insert operations.

Thus from the experiment it is observed that MongoDB provides consistent performance measures with standard workloads. However, the performance degrades with the increased workloads. Among all the three Cassandra provides highest performance measure in all the scenarios. Hbase provides improved performance when there exist complex operations. In this manner, the data stores are benchmarked across various scenarios. The experimental results are clearly illustrated from figure 5 to 11 for better understanding purposes.

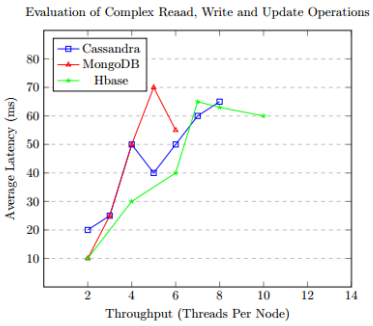

Figure 5: Evaluation of Complex read, write and update operations

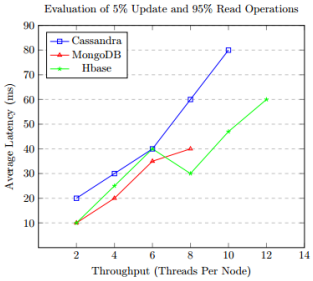

Figure 6: Evaluation of 5% update and 95% read operations

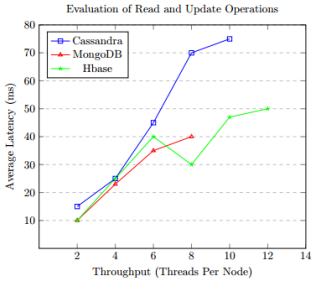

Figure 7: Evaluation of read and Update operations

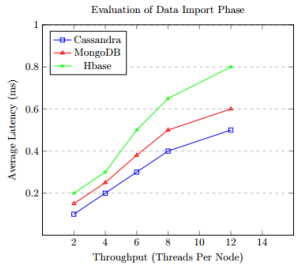

Figure 8: Evaluation of Data Import Phase

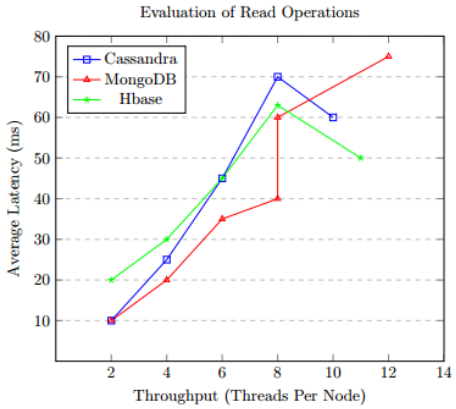

Figure 9: Evaluation Read Operations

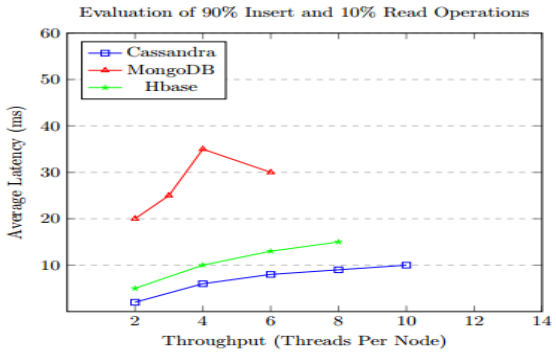

Figure 10: Evaluation of 90% Insert and 10% Read Operations

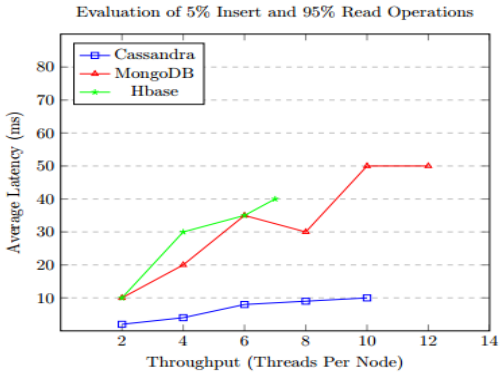

Figure 11: Evaluation of 5% Insert and 95% Read Operations

Discussions

Thus, from the experiment, it is concluded that all the three databases MongoDB, Cassandra and Hbase form the suitable data stores for the E-Health clouds. But its performance measures varies from one scenario to another. Even though the E-Health cloud possesses the same architecture pattern, its utility varies from one system model to another. Thus a discussion on the utility of these databases across different E\-Health scenarios is given in this section. In Cassandra, the process of scale up, scale down, remove or add nodes can be made quickly in an automated manner. It forms the most suitable solution when the E-health cloud scenario requires simple setup and maintenance processes. It is most efficient, when there is a high velocity of random read and write operations. It does not require multiple secondary indexes and flexible to wide or sparse column requests. In certain E-health applications such as prediction analysis, the property of strict consistency is needed. During this situations, Hbase forms the most suitable solution. Hbase is used when there is a need for optimized read operations and range-based query scan of EHR's. Also, it forms the most suitable solution when the E-health cloud requires faster read and write

operations with improved scalability. However, it does not offer much support to real-time data analytics and aggregation operations. MongoDB is widely used when the EHR's are in the form of semi\-structured data. It highly supports real-time data analytics and scalability. However, it does not form the most suitable database system when there is a need for foreign key constraints. Thus, depending upon the constraints and E-health cloud requirements, these data stores are used at a real time.

5. Conclusion

The paper provides an analysis and performance evaluation of NoSQL databases for E-Health clouds. Benchmarking the NoSQL data stores in the perspective of the E-Health cloud is an important requirement as there exists a variety of NoSQL databases and its utility differs from one application to another. Further, system performance remains to be an important factor when dealing with huge volume of EHR around E-Health clouds. A brief analysis is made to identify the most appropriate NoSQL data stores for E-Health clouds. Document datastores and column family stores are found to be the most suitable solution. Because it possesses all the capabilities to store and manage EHR in an efficient manner with improved performance. To benchmark these data stores, we derived suitable performance metrics. Scalability, availability, flexibility, durability and query expressiveness are some of the metric to benchmark the databases. Among them, latency and throughput are found to be the most important factors. The experimental result states that all the three databases Cassandra, Hbase, and MongoDB form the suitable solution to the E-Health clouds. Among the three databases, Cassandra is identified to be the most suitable one for E-Health clouds. It provides higher performance measures, but it degrades across complex write operations. MongoDB provides standard performance measures at all the scenarios. Hence it forms the most suitable solution when we require a standard and simple data store. HBase is utilized when there is complex read and write operations. In future, this work can be extended to evaluate the E-Health clouds performance measure at various situations and data model.

References

[1] Eysenbach G., What is e-health?,Journal of medical Internet research 3(2) (2001).

[2] Hoerbst A., Ammenwerth E., Electronic health records, Methods Inf Med 49(4) (2010), 320-336.

[3] Häyrinen K., Saranto K., Nykänen P., Definition, structure, content, use and impacts of electronic health records: a review of the research literature. International journal of medical informatics77(5) (2008), 291-304.

[4] AbuKhousa E., Mohamed N., Al-Jaroodi J., e-Health cloud: opportunities and challenges, Future Internet 4(3) (2012), 621-645.

[5] Lounis A., Hadjidj A., Bouabdallah A., Challal Y., Secure and scalable cloud-based architecture for e-health wireless sensor networks, 21st international conference on Computer communications and networks (2012), 1-7.

[6] Tamizharasi G.S., Manjula R., Monisha K., Balamurugan B., A Secure and Efficient Framework for Health Data Management in E-Health Clouds, International Journal of Computer Science and Information Security14(9) (2016).

[7] Bricon-Souf N., Conchon E., A 2015 Medical Informatics Perspective on Health and Clinical Management: Will Cloud and Prioritization Solutions Be the Future of Health Data Management?., Yearbook of medical informatics 10(1) (2015).

[8] Brown G.D., Patrick T.B., Pasupathy K.S. eds., Health informatics: a systems perspective. Health Administration Press (2013).

[9] Madden S., From databases to big data, IEEE Internet Computing 16(3) (2012), 4-6.

[10] Moniruzzaman A.B.M., Syed AkhterHossain, Nosql database: New era of databases for big data analytics-classification, characteristics and comparison. arXiv preprint arXiv:1307.0191 (2013).

[11] Levin, Nadine, Reza M. Salek, Christoph Steinbeck, From Databases to Big Data, Metabolic Phenotyping in Personalized and Public Healthcare (2016).

[12] Vilaplana J., Solsona F., Abella F., Filgueira R., Rius J., The cloud paradigm applied to e-Health. BMC medical informatics and decision making13(1) (2013).

[13] Benharref, Abdelghani, Mohamed Adel Serhani, Novel cloud and SOA-based framework for E-Health monitoring using wireless biosensors, IEEE journal of biomedical and health informatics 18(1) (2014), 46-55.

[14] Yüksel B., Küpçü A., Özkasap Ö., Research issues for privacy and security of electronic health services, Future Generation Computer Systems, 68, 1-13.

[15] Avram M.G., Advantages and challenges of adopting cloud computing from an enterprise perspective, Procedia Technology 12 (2014), 529-534.

[16] Dinh H.T., Lee C., Niyato D., Wang, P., A survey of mobile cloud computing: architecture, applications, and approaches. Wireless communications and mobile computing13(18) (2013), 1587-1611.

[17] Ullah S., Higgins H., Braem B., Latre B., Blondia C., Moerman I., Saleem S., Rahman Z., Kwak K.S., A comprehensive survey of wireless body area networks. Journal of medical systems36(3) (2012), 1065-1094.

[18] He D., Zeadally S., Kumar N., Lee J.H., Anonymous authentication for wireless body area networks with provable security, IEEE Systems Journal (2016).

[19] Cavallari R., Martelli F., Rosini R., Buratti C., Verdone R., A survey on wireless body area networks: Technologies and design challenges, IEEE Communications Surveys & Tutorials, 16(3), pp.1635-1657.

[20] Surendar, A., Rani, N.U."High speed data searching algorithms for DNA searching",(2016) International Journal of Pharma and Bio Sciences, 2016 (SpecialIsssue), pp. 73-77.

[21] He D., Zeadally S., Wu L., Certificateless public auditing scheme for cloud-assisted wireless body area networks, IEEE Systems Journal (2015).

[22] Zhang Y., Qiu M., Tsai C.W., Hassan M.M., Alamri A., Health-CPS: Healthcare cyber-physical system assisted by cloud and big data, IEEE Systems Journal11(1) (2017), 88-95.

[23] Tong Y., Sun J., Chow S.S., Li P., Cloud-assisted mobile-access of health data with privacy and auditability, IEEE Journal of biomedical and health Informatics18(2) (2014), 419-429.

[24] NoSQL databases: a step to database scalability in web environment

[25] Weider D.Y., Kollipara M., Penmetsa R., Elliadka S., A distributed storage solution for cloud based e-Healthcare Information System, IEEE 15th International Conference on In e-Health Networking, Applications & Services (2013), 476-480.

[26] Ercan M.Z., Lane M., An evaluation of NoSQL databases for EHR systems, Proceedings of the 25th Australasian Conference on Information Systems. Auckland University of Technology, School of Business Information Systems (2014).

[27] The rise of "big data" on cloud computing: Review and open research issues

[28] SitalakshmiVenkatraman K.F., Kaspi S., Venkatraman R., SQL VersusNoSQL Movement with Big Data Analytics (2016).

[29] Mohamed M.A., Altrafi O.G., Ismail M.O., Relational vs. nosql databases: A survey, International Journal of Computer and Information Technology3(03) (2014), 598-601.

[30] Surendar, A."Evolution of gait biometric system and algorithms-A review" (2017) Biomedical and Pharmacology Journal, 10 (1), pp. 467-472.

[31] Vimalkumar, M.N., Helenprabha, K., Surendar, A."Classification of mammographic image abnormalities based on emo and LS-SVM techniques",(2017) Research Journal of Biotechnology, 12 (1), pp. 35-40.

[32] Fiannaca A.J., Justin Huang, Benchmarking of Relational and NoSQL Databases to Determine Constraints for Querying Robot Execution Logs, Computer Science & Engineering, University of Washington, USA (2015), 1-8.

[33] Park H.J., A Study about Performance Evaluation of Various NoSQL Databases, The Journal of Korea Institute of Information, Electronics, and Communication Technology 9(3) (2016), 298-305.

[34] Mohanasundaram R., Periasamy P.S., Clustering Based Optimal Data Storage Strategy Using Hybrid Swarm Intelligence In WSN, Wireless Personal Communications (2015).

[35] Mohanasundaram R., Periasamy P.S., Hybrid Swarm Intelligence Optimization Approach for Optimal Data Storage Position Identification in Wireless Sensor Networks, The Scientific World Journal (2015).

[36] Mohanasundaram R., Periasamy P.S., Swarm Based Optimal Data Storage Position Using Enhanced Bat Algorithm In Wireless Sensor Networks, International Journal of Applied Engineering Research 10(2) (2015), 4311-4328.

[37] Mohanasundaram R., Periasamy P.S., A Meta heuristic Algorithm for Optimal Data Storage Position in Wireless Sensor Networks, Pakistan Journal of Biotechnology (2016), 463-468.

[38] Aarthy S.L., PrabuS. A computerized approach on breast cancer detection and classification, Iioab journal 7(5) (2016), 157-169.

[39] Aarthy, S.L., Prabu S., An approach for detecting breast cancer using wavelet transforms, Indian Journal of Science and Technology 8(26) (2015).

[40] Gopinath, M.P., PrabuS. Classification of thyroid abnormalities on thermal image: a study and approach, Iioab journal 7(5) (2016), 41-57.

[41] Gopinath, M.P., PrabuS. A Comparative study of Techniques Involved in Thermal Image Diagnostic System, International Journal of Applied Engineering Research, 9(24) (2014), 26393-26416.

[42] Manju, K., Sabeenian, R.S., Surendar, A."A review on optic disc and cup segmentation",(2017) Biomedical and Pharmacology Journal, 10 (1), pp. 373-379.