Evaluation of NoSQL databases for EHR systems

Mehmet Zahid Ercan
Faculty of Business Education Law and Arts
University of Southern Queensland
Toowoomba, Queensland
Email: MehmetZahid.Ercan@usq.edu.au

Michael Lane
School of Management and Enterprise
Faculty of Business Education Law and Arts
University of Southern Queensland
Toowoomba, Queensland
Email: Michael.Lane@usq.edu.au

Abstract

Electronic Health Record (EHR) systems offer significant benefits for healthcare. The improved availability of healthcare information from multiple locations contributes to the accuracy and timeliness of care, and should lead to overall improved quality of healthcare delivery. Practical experience and relevant research demonstrate that there are many technological issues that need to be addressed for modern healthcare systems to be effective in sharing EHRs as the structure and size of the healthcare data have changed considerably over time. Recent literature shows that the emerging NoSQL databases have significant advantages such as easy and automatic scaling, better performance and high availability which address the limitations of relational databases in distributed healthcare systems. In this paper we reviewed EHRs and the key features of NoSQL databases. We then evaluated the suitability of NoSQL databases in meeting the requirements of national EHR systems in sharing EHRs in a distributed system environment.

Keywords

NoSQL databases, Electronic Health Record (EHR), healthcare systems, relational databases, distributed systems, ACID, CAP theorem, BASE.

INTRODUCTION

The implementation of Electronic Health Record (EHR) systems and in particular healthcare data sharing between healthcare providers remain a significant challenge for many countries despite the developments in database technology and network infrastructure (Bacelar-Silva et al. 2011; Hoerbst et al. 2010; Pearce and Haikerwal 2010). Many countries such as Australia, Finland, Germany and Turkey are still trying to establish their nationwide e-health platforms that will facilitate data sharing. However the issues about data standards, scalability, high volumes of data storage and data processing and the cost of implementation are particularly challenging for governments and healthcare providers (Bacelar-Silva et al. 2011; Drejhammar 2010; Grimson 2001; Hoerbst et al. 2010; Jin et al. 2011; Schmitt and Majchrzak 2012; Vest 2012).

Data intensive information systems require solid database management systems in order to function properly. The size and heterogeneity of data in modern distributed systems is increasing rapidly. Thus, the need to scale databases beyond the capabilities of traditional relational databases running on a single large computer system has led to the development of new scalable database systems (Borkar et al. 2012; Helland 2011; Konstantinou et al. 2011). These new systems are referred to as “NoSQL” databases. While the name is not entirely agreed upon, NoSQL stands for “Not Only SQL” (Cattell 2011). Traditional relational database management systems have limitations due to scalability and infrastructure cost issues. NoSQL database systems which have emerged in response to these limitations are mostly open-source and can run on commodity hardware architectures. NoSQL database systems can scale horizontally with no single point of failure or bottlenecks because of a shared-nothing architecture (Borkar et al. 2012; Konishetty et al. 2012).

NoSQL databases offer low-cost solutions that provide high availability and addressing scalability issues. NoSQL database systems have been heavily influenced by Google’s Bigtable and Amazon’s Dynamo systems and can easily scale up to accommodate large datasets (Borkar et al. 2012; Schram and Anderson 2012). Some NoSQL databases have already been developed and used commercially by companies such as Google and Amazon. However, there are also many open source NoSQL database systems that have been developed based on similar approaches, such as HBase, MongoDB, CouchDB, Cassandra, Memcached, etc (Schram and Anderson 2012). Moreover, NoSQL database systems are already used in large commercial applications by Google, Amazon, LinkedIn and Twitter because they offer high scalability and availability that the traditional relational database systems cannot provide. Furthermore, open source NoSQL database systems have a significant advantage in terms of implementation and software licence costs over traditional relational database systems. This is another reason to use NoSQL database systems to address the shortcomings of commercial relational database management systems (Escriva et al. 2012).

There are numerous white papers, blog entries and comments mentioning the advantages of NoSQL database systems (Parker et al. 2013). While there are significant advantages of using NoSQL database systems, there is limited research which has evaluated the use of NoSQL databases in the healthcare domain. In this paper we identify the issues and requirements of modern healthcare applications, particularly EHR systems, and demonstrate how NoSQL databases can provide viable solutions for sharing EHRs drawing on the relevant literature.

This paper is organised as follows. In the next section a general overview of Electronic Health Records (EHR), EHR systems, importance of EHR sharing, technological issues affecting EHR systems is provided. Then background information about NoSQL databases and theoretical basis of NoSQL databases is provided and related work about NoSQL databases is presented. Next, the potential benefits of NoSQL databases for EHR systems are discussed. Finally conclusions and future research on the role of NoSQL databases in facilitating EHR sharing in distributed healthcare systems are presented.

BACKGROUND

Electronic Health Record (EHR)

Storing healthcare information electronically was introduced several decades ago because paper-based records can no longer meet the requirements of an advanced health care system (van Ginneken 2002). Electronically stored healthcare information has been identified by a number of different names such as Electronic Patient Records (EPR), Computerised Patient Records, Electronic Medical Records and Electronic Health Records (EHR) (ISO 2004; Narayan et al. 2010).

While these terms might sometimes be used interchangeably, National Health Service (NHS) suggests that EPR is “the record of the periodic care provided mainly by one institution”. On the other hand, EHR is defined as the collection of patient’s health and healthcare information, from cradle to grave. According to these definitions, EHR is described as a collection of EPRs for a single individual (Executive 1998). The International Standard Organization (ISO) defined EHR as “a repository of information regarding the status of a subject of care in a computer processable form and, transmitted securely, and accessible by multiple authorized users” (ISO 2004). There is no one single definition for a EHR. Therefore synthesising well accepted industry definitions, we define an EHR as : an electronic record that holds a patient’s lifetime health-related information. Furthermore, a system that handles operations on EHRs will be referred to as an EHR system or an EHR application.

Implementing a national EHR system is one of the top priority targets for many nations including Australia, Canada, England and the United States. Many countries have developed their own national EHR system architecture. For example, Turkey has a national system called “Saglik-NET” which collects and centrally stores a wide range of medical data. In the Netherlands the data is kept locally and a central system called a “National Switch Point (NSP)” handles the links to the data and allows access to information in various services in the health network. Austria and Germany are establishing their nationwide EHR systems (Hoerbst et al. 2010). In England The National Care Record Service (NCRS) enables access to patients’ EHRs in a national system called “Spine” (Bacelar-Silva et al. 2011). Authorised professionals can access summary records of patients which include basic information such as date of birth, name, contact information, allergies, etc. In Australia, on the other hand, there is a significant effort on establishing a Personally Controlled Electronic Health Record (PCEHR) system (Vest 2012). The National E-Health Transition Authority (NEHTA) is working on establishing governing standards that will enable PCEHR. The role of NEHTA is to “develop better ways of electronically collecting and securely exchanging health information” (NEHTA 2013).

Benefits of Electronic Health Record (EHR) sharing

The literature on EHR emphasises the importance of information sharing function of EHR in improving healthcare outcomes. Iakovidis (1998) suggests that the purpose of EHR is to support continuity of care and van der Linden et al. (2009) mention the support of continuing, efficient and quality integrated health care as the primary purpose of EHR. Narayan et al. (2010) suggest that a life-time health record system is established to keep track of all healthcare-related information of individuals from birth to death to allow efficient, consistent and universal sharing of health information. Previous studies also suggest that additional purposes of EHR are providing support for the development of health policies, medical education and advanced research. (Heard 2006; Iakovidis 1998; van der Linden et al. 2009). According to the US Institute of Medicine, an EHR improves patient safety, supports efficient patient care delivery and improves the efficiency of healthcare services (Englebardt and Nelson 2002; Kohn et al. 2000). Schiff et al (2003) points out that patient safety and quality of healthcare can be increased through her sharing amongst healthcare facilities. Halamka et al (2005) demonstrated that the uncoordinated approach for medical records leads to wastage of time and medical errors. Previous literature suggests that implementing a fully functioning EHR system with the participation of all healthcare organisations could lead to a USD77.8 billion benefit for the United States (Walker et al. 2005; Yasnoff et al. 2004).

Technological issues affecting EHR systems

Establishing a nation-wide EHR system requires a significant investment as well as extensive system design and project management (Hoerbst et al. 2010; Pearce and Haikerwal 2010; Vest 2012). Poorly designed architecture not only poses a substantial failure risk for the implementation of EHR systems, but also can cause significant losses of financial and human resources (Pearce and Haikerwal 2010). There are a number of obstacles and challenges in relation to EHR systems mentioned in the literature, such as standardisation of vocabulary, security, privacy and data quality. In addition to these matters which are extensively covered in the existing literature, Orfanidis et al (2004) claims that the expanding size of healthcare data also creates an obstacle for EHR systems. Blobel (2006) suggests that an EHR system which allows the exchange of health information should be scalable, flexible and portable with Internet access. The increasing diffusion of information systems in healthcare delivery and increasing size and heterogeneity of the healthcare data has resulted in a bottleneck for storage, retrieval, high availability and analysis aspects of traditional relational databases. NoSQL database systems might be the solution to this bottleneck in addition to providing many other advantages (Jin et al. 2011; Schmitt and Majchrzak 2012).

NoSQL databases

NoSQL is a term used to describe a new category of non-relational databases. A NoSQL database is also known as a distributed data store is capable of scaling to large datasets with no single point of failure. Data may span server nodes, racks, and even multiple data centres. Recently NoSQL database systems have attracted attention of industry and researchers due to the demand for high performance access to large volumes of data without requiring significant effort for scaling and tuning (Ferreira et al. 2013). NoSQL databases gained speed in development and market acceptance after Google's BigTable whitepaper and Amazon's Dynamo which contributed to the evolution of NoSQL databases. NoSQL database technology depends on horizontal scalability which enables increasing performance and capacity by increasing the number of nodes, rather than increasing the computer power of single node (Abramova and Bernardino 2013). Figure 1 shows how NoSQL is trending in terms of attention in contrast to relational databases over last 10 years based on Google searches.

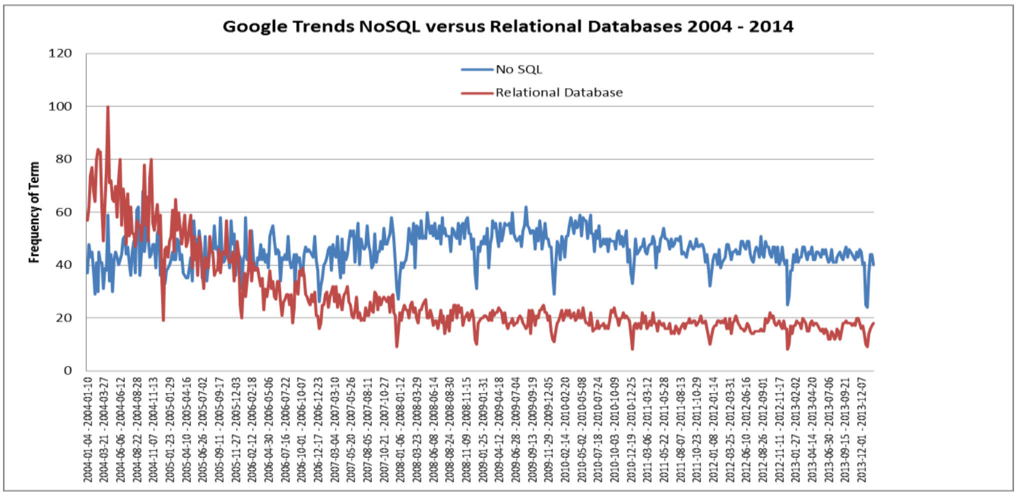

Figure 1: Google search trends NoSQL databases versus Relational Databases

Types of NoSQL databases

Open source availability of many NoSQL databases as well as vast amount of know-how that is made publicly available by large vendors of NoSQL databases has led an increasing number of IT practitioners being involved in developing NoSQL databases. There are over 150 different NoSQL databases available which can be grouped into four categories for modelling of the data requirements of a specific application: (1) Key-value store, (2) Document store, (3) Column-family, and (4) Graph database (Abramova and Bernardino 2013).

In key-value stores, all data is stored as key-value pairs, in which, the keys are unique values that are used to access the information stored in values. Redis, Azure Table Storage and DynamoDB are examples of key-value stores. Document stores are essentially similar to key-value stores. However, the values are usually documents in known formats such as XML or JSON. Well-known examples of document stores are MongoDB and Couchbase (Abramova and Bernardino 2013; Dede et al. 2013). On the other hand, in column family type of NoSQL databases, data is stored in columns, however the columns are not required to be defined at the beginning and there may be countless numbers of columns which may also be organized in groups called supercolumns. Cassandra and HBase are examples of known column family types of NoSQL databases which have been implemented in practice. Finally, Graph databases are the examples of data stores that can store and handle graph type of data such as social network relations. Neo4j and InfoGrid are examples of graph databases (Abramova and Bernardino 2013).

Relational Database Theory

The relational model and database theory has its origins with E.F. Codd, in the 1970s. The relational model has been adopted widely in industry and many of the current modern day commercial database systems are influenced by the work of Codd (Suciu 2001). In theory, Codd (1970) suggests that the data stored in large shared data banks can be defined and organised based on interrelationships of data, and redundancy and consistency problems can be eliminated by normalisation of data. This is a procedure for organising data into relational views by eliminating the copies of the same data and establishing a link between data groups using primary keys (Abiteboul et al. 1995; Codd 1970). Fundamentals of relational databases have remained unchanged for decades. However, the link between theory and practice in relation to database systems has weakened over the time (Abelló et al. 2011; Badia and Lemire 2011; Suciu 2001; Vianu 2001). The emergence of high-speed networks, fast commodity hardware and the increasing amount of unstructured or semi-structured data has raised the necessity for traditional database theory and designs to be adapted in order to meet the needs of today's business environment. Recent studies such as Valduriez (2011), Jin et al. (2011) and Konishetty et al. (2012) have explored the principles underpinning distributed database management systems and the practical implementations of NoSQL databases to help in establishing a better link between theory and practice in information management.

CAP Theorem

NoSQL database systems have received much attention from research community (Cattell 2011; Escriva et al. 2012; Floratou et al. 2012; Lee et al. 2013; Schram and Anderson 2012). The literature suggests that current research focuses on the scalability, fault-tolerance and performance advantages of the NoSQL/distributed database systems, while criticising the weak consistency approach. The issue of consistency is explained in the context of CAP (consistency, availability, and partition-tolerance) theorem (Agrawal et al. 2011; Bermbach and Tai 2011). CAP theorem, introduced by Eric Brewer in 2000, suggests that there is always a trade-off between consistency, availability and partition-tolerance. In CAP theorem, consistency means that each server returns the right response to each request, availability means that each request will eventually receive a response and while partition-tolerance means that the service can continue operating normally even when communication between some nodes are lost. The underlying idea of CAP theorem is that the communication between servers is prone to network errors and failures, thus it is not possible to have all three features (consistency, availability, partition tolerance) working together perfectly (Gilbert and Lynch 2012).

ACID (Atomicity, Consistency, Isolation, Durability) Properties

Gray (1981) suggested a number of properties for database systems to achieve reliable transaction processing, which are known as Atomicity, Consistency, Isolation and Durability (ACID). Atomicity means a transaction is either completed entirely or failed, i.e. there is no partial completion in any transaction. Consistency is the property that guarantees that every transaction changes a database into a valid new state, incorporating all rules, constraints and triggers, etc. Isolation means that each transaction happens totally independent of each other and transactions do not affect each other while being executed. Durability is the property that means if a transaction has been completed, the new state of database is guaranteed to be durable regardless of any potential failures such as power loss, network errors, etc. afterwards (Gray 1981; Sattar et al. 2013) .

BASE (Basically Available, Soft State, Eventually Consistent)

Due to the distributed nature of NoSQL databases without a coordinator or master node and based on the CAP theorem NoSQL databases cannot offer strong consistency models like traditional relational databases can do. Therefore, while NoSQL databases have many advantages such as high availability and easy scalability, NoSQL databases cannot have all strong ACID properties. NoSQL databases focus on the BASE principal instead, which stands for **B**asically **A**vailable, **S**oft state and **E**ventually consistent. BASE principal implies that the system can continue working as usual in case of a failure due to the distributed nature of NoSQL databases and even though there is no guarantee of consistency at any given point of time, data will eventually be consistent at some point in time (Bailis and Ghodsi 2013). Google has recently published a paper on "Spanner", Google's globally distributed database system, which mentions the possibility of transaction control, consistency and replication without sacrificing high-availability, and there are papers suggesting that it might be possible to achieve consistency and high availability together to the extent that distributed databases can match the properties of current relational databases (Bailis et al. 2013; Corbett et al. 2013) . Figure 2 summarises three data models which categorise NoSQL databases and comparative strengths of NoSQL databases and relational databases in terms of CAP theorem.

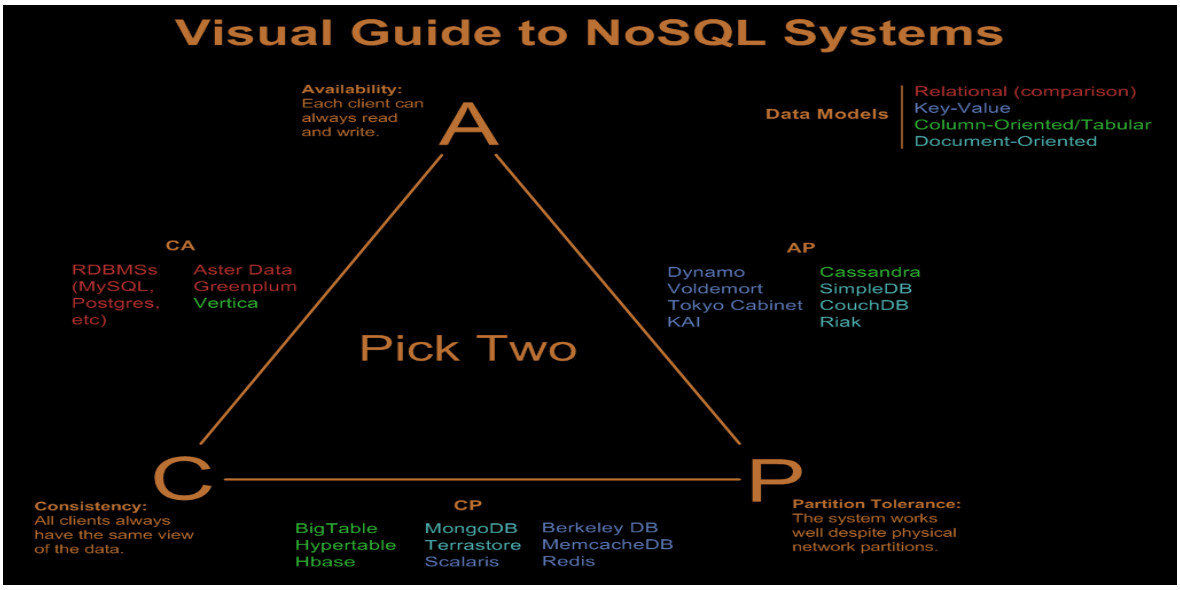

Figure 2: Comparison of three main data model types used in NoSQL databases with relational databases in terms of CAP Theorem

Figure 2 shows that the strengths of relational data models are in being able to deliver consistency and a lesser extent availability whereas strengths of No SQL key value, column oriented, tabular and document oriented data models are in being able to deliver consistency and partition tolerance or availability and partition tolerance.

POTENTIAL BENEFITS OF NOSQL DATABASES FOR EHR SYSTEMS

Table 1 presents the main requirements of EHRs in a distributed healthcare system and how NoSQL database system features can address these requirements.

Table 1. Comparison of EHR requirements and NoSQL database features that address these requirements

EHR requirement	**NoSQL database feature**
Size of healthcare data increased over time, data size became a bottleneck for EHR systems	NoSQL databases are based on horizontal scalability which allows easy and automatic scaling
Healthcare data includes free-text notes, images and other complex data. Heterogeneity of healthcare data leads to a requirement of new solutions	Flexible data models offered by NoSQL databases allow unstructured or semi-structured data to be stored easily
Healthcare data should always be accessible for continuity of healthcare services	NoSQL databases provide high availability due to the distributed nature and replication of data
Healthcare data is normally added, not updated	Eventual consistency suggested by NoSQL database architecture considered acceptable for EHR use cases
Healthcare data sharing requires access to EHRs from multiple locations which requires a high-performance system to respond data access request in a timely manner	NoSQL databases offer higher performance compared to relational databases in many use cases

The main benefits that NoSQL databases provide for EHR systems is now discussed in detail, in turn.

Scalability and Performance

The increasing size of healthcare data and the requirement of scalability for EHR systems are considered bottlenecks for EHR system implementations as most of the current systems are based on relational databases which limits scalability (Dolin et al. 2006; Guo et al. 2005; Guo et al. 2004; Jin et al. 2011; Schmitt and Majchrzak 2012; Takeda et al. 2000). The fundamental advantage of NoSQL databases is that they allow scaling up to large datasets without any changes in the overall structure of data model or architecture. Hardware requirements and costs can grow linearly as storage requirements grow. Therefore cost-effective scaling up is made possible and high initial investment in hardware requirements are avoided (Lakshman and Malik 2010).

Traditional relational database systems rely on purchasing more expensive and powerful servers to increase capacity. However, distributed data storage systems such as NoSQL database systems are based on shared-nothing approach. In a shared-nothing architecture, servers have their own resources, thus they do not share RAM, processor or storage. This enables horizontal scaling, the distribution of data and processing operations over many servers to achieve large numbers of read/write operations per second (Cattell 2011). Capacity can be increased by adding more commodity servers dynamically and redistribution of the data occurs on the fly and seamlessly without reconfiguration or a decrease in performance. This is one of the most important advantages of NoSQL database systems over traditional relational database systems (Pokorny 2011).

High Availability

Schmitt and Majchrzak (2012) suggest that the nature of healthcare data requires high availability and distributed data management to enable access to healthcare information whenever needed, even in an event of crisis when data centres fail. Achieving high availability by maintaining a number of replications, enabling high performance on transactions using distributed algorithms are also advantages of NoSQL database systems over traditional relational database systems (Featherston 2010; Mengchen 2011). In order to achieve this, NoSQL database systems trade consistency for availability and introduce eventual consistency concept (Dede et al. 2013).

Consistency Considerations

In an eventually consistent NoSQL database, data read by clients immediately after being updated may be an out-dated version as all nodes have not been updated at once. However, some NoSQL databases such as Cassandra offers different levels of consistency and users can select the level of consistency they require for each transaction. In addition to that, other studies have shown that the inconsistency windows for many NoSQL databases are less than a second. Therefore, eventual consistency model suggested by NoSQL databases is claimed to be sufficient in most use cases (Bailis and Ghodsi 2013) .

In an update-intensive database application where consistency is very important, such as a stock exchange system that handles financial transactions from all over the world where milliseconds in processing time matter, databases with strong consistency features are more suitable. However in the case of healthcare, the healthcare record of a person is usually appended not updated, and the updates to the same data such as penicillin allergy status of a patient, are not made from many different locations at the same time. Therefore weaker consistency models can be applied using NoSQL databases without major drawbacks in the healthcare domain (Frank et al. 2014).

Flexible Data Model

Research and industry projects focusing on storing healthcare information in NoSQL databases are being driven by practical experience demonstrating that the traditional relational approach for storing healthcare records has become the bottleneck for healthcare systems as the structure and size of the healthcare data have changed considerably over time. Medical databases can contain heterogeneous data including text, images, free-text physician notes, logs from medical devices, prescriptions etc. This heterogeneous medical data because of its size and structure is difficult to handle and manage using traditional relational databases (Jin et al. 2011; Schmitt and Majchrzak 2012).

NoSQL databases have different types of data structures which allow different ways for modelling data. For instance document store type - NoSQL databases can handle documents having unstructured or semi-structured data easily. Data does not need to be defined beforehand and thus any content can be stored as documents without any schema constraints. This aspect of NoSQL databases addresses some of shortcomings of relational databases in healthcare.

Open Source Availability

It is also important to note that there are many open source NoSQL database alternatives which may help in reducing the overall cost of implementation and customisation of a database system by enabling access to the source code.

CONCLUSIONS AND FUTURE RESEARCH

Data storage systems are crucial for all sorts of data intensive applications which store and manage huge amounts of data. Modern applications such as high-traffic web sites or large enterprise systems require new approaches to data storage in order to achieve higher performance and higher availability than is possible with traditional relational database management systems (RDBMS). This is particularly the case when it involves a high concurrent numbers of transactions and large amounts of data.

There is no unanimous agreement in the literature on the overall superiority of NoSQL databases over traditional relational databases as well as their generic suitability for data-intensive applications. However, previous studies suggest that NoSQL databases have many technical and financial advantages for large scale data intensive applications.

Past empirical research demonstrates that the type and the requirements of the application dramatically determine the suitability of NoSQL databases.

Considering the importance of EHR implementation for continuity of care and overall health systems, using NoSQL databases have significant potential to lead to better EHR applications in terms of scaling, flexibility and high availability.

However, there is limited research on this topic which either focuses on exploring possibility of establishing a healthcare data model using a NoSQL database or merely tries to evaluate basic database performance by comparing the performance of NoSQL databases with relational databases. Moreover inadequate attention has been given to establishing a healthcare data model and testing the performance with realistic healthcare data sets to validate the comparison between NoSQL databases and relational databases. Clearly this may lead to the results which deviate from what can be obtained in a real-world scenario.

Future work in this research project will involve designing and populating a NoSQL EHR data model based on Australian Healthcare data standards and specifications. The second phase will involve the evaluation of the performance, scalability and availability of a NoSQL database in comparison to a relational database in various configurations and scenarios using randomly generated data that is realistic of healthcare data. We believe that in the area of healthcare that NoSQL databases are worthy of empirical research. The contribution of NoSQL databases to addressing the increasing need for scalability and availability in a distributed systems architecture to facilitate EHR sharing in national healthcare systems is significant.

REFERENCES

Abelló, A., Ferrarons, J., and Romero, O. 2011. "Building Cubes with Mapreduce," *Proceedings of the ACM 14th international workshop on Data Warehousing and OLAP*: ACM, pp. 17-24.

Abiteboul, S., Hull, R., and Vianu, V. 1995. *Foundations of Databases*. Addison-Wesley.

Abramova, V., and Bernardino, J. 2013. "Nosql Databases: Mongodb Vs Cassandra," in: *Proceedings of the International C* Conference on Computer Science and Software Engineering*. Porto, Portugal: ACM, pp. 14-22.

Agrawal, D., Das, S., and El Abbadi, A. 2011. "Big Data and Cloud Computing: Current State and Future Opportunities," *14th International Conference on Extending Database Technology*: ACM, pp. 530-533.

Bacelar-Silva, G.M., Vicente, C.M.O., David, M., and Antunes, L. 2011. "Comparing Security and Privacy Issues of Ehr: Portugal, the Netherlands and the United Kingdom," in: *4th International Symposium on Applied Sciences in Biomedical and Communication Technologies*. Barcelona, Spain: ACM, pp. 1-4.

Badia, A., and Lemire, D. 2011. "A Call to Arms: Revisiting Database Design," *ACM SIGMOD Record* (40:3), pp. 61-69.

Bailis, P., Fekete, A., Ghodsi, A., Hellerstein, J.M., and Stoica, I. 2013. "Hat, Not Cap: Towards Highly Available Transactions," in: *Proceedings of the 14th USENIX conference on Hot Topics in Operating Systems*. Santa Ana Pueblo, New Mexcio: USENIX Association, pp. 24-24.

Bailis, P., and Ghodsi, A. 2013. "Eventual Consistency Today: Limitations, Extensions, and Beyond," *Commun. ACM* (56:5), pp. 55-63.

Bermbach, D., and Tai, S. 2011. "Eventual Consistency: How Soon Is Eventual? An Evaluation of Amazon S3's Consistency Behavior," *Proceedings of the 6th Workshop on Middleware for Service Oriented Computing*: ACM, p. 1.

Blobel, B. 2006. "Advanced and Secure Architectural Ehr Approaches," *International journal of medical informatics* (75:3), pp. 185-190.

Borkar, V.R., Carey, M.J., and Li, C. 2012. "Big Data Platforms: What's Next?," *XRDS* (19:1), pp. 44-49.

Cattell, R. 2011. "Scalable Sql and Nosql Data Stores," *SIGMOD Rec.* (39:4), pp. 12-27.

Codd, E.F. 1970. "A Relational Model of Data for Large Shared Data Banks," *Communications of the ACM* (13:6), pp. 377-387.

Corbett, J.C., Dean, J., Epstein, M., Fikes, A., Frost, C., Furman, J.J., Ghemawat, S., Gubarev, A., Heiser, C., Hochschild, P., Hsieh, W., Kanthak, S., Kogan, E., Li, H., Lloyd, A., Melnik, S., Mwaura, D., Nagle, D., Quinlan, S., Rao, R., Rolig, L., Saito, Y., Szymaniak, M., Taylor, C., Wang, R., and Woodford, D. 2013. "Spanner: Google&Rsquo;S Globally Distributed Database," *ACM Trans. Comput. Syst.* (31:3), pp. 1-22.

Dede, E., Govindaraju, M., Gunter, D., Canon, R.S., and Ramakrishnan, L. 2013. "Performance Evaluation of a Mongodb and Hadoop Platform for Scientific Data Analysis," in: *Proceedings of the 4th ACM workshop on Scientific cloud computing*. New York, New York, USA: ACM, pp. 13-20.

Dolin, R.H., Alschuler, L., Boyer, S., Beebe, C., Behlen, F.M., Biron, P.V., and Shvo, A.S. 2006. "Hl7 Clinical Document Architecture, Release 2," *Journal of the American Medical Informatics Association* (13:1), pp. 30-39.

Drejhammar, F. 2010. "Designing a Trusted Distributed Long-Term Archive for Health Records," in: *ERCIM News*. Sweden.

Englebardt, S.P., and Nelson, R. 2002. *Health Care Informatics: An Interdisciplinary Approach*. Mosby Incorporated.

Escriva, R., Wong, B., and Sirer, E.G. 2012. "Hyperdex: A Distributed, Searchable Key-Value Store," in: *ACM SIGCOMM 2012 conference on Applications, technologies, architectures, and protocols for computer communication*. Helsinki, Finland: ACM, pp. 25-36.

Executive, N. 1998. *Information for Health: An Information Strategy for the Modern Nhs 1998-2005: A National Strategy for Local Implementation*. NHS Executive.

Featherston, D. 2010. "Cassandra: Principles and Application." Retrieved 27.08.2011, from http://disi.unitn.it/~montreso/ds/papers/Cassandra.pdf

Ferreira, G.d.S., Calil, A., and Mello, R.d.S. 2013. "On Providing Ddl Support for a Relational Layer over a Document Nosql Database," in: *Proceedings of International Conference on Information Integration and Web-based Applications & Services*. Vienna, Austria: ACM, pp. 125-132.

Floratou, A., Teletia, N., DeWitt, D.J., Patel, J.M., and Zhang, D. 2012. "Can the Elephants Handle the Nosql Onslaught?," *Proceedings of the VLDB Endowment* (5:12), pp. 1712-1723.

Frank, L., Pedersen, R.U., Havn, C., #248, Frank, and Larsson, N.J. 2014. "The Cap Theorem Versus Databases with Relaxed Acid Properties," in: *Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication*. Siem Reap, Cambodia: ACM, pp. 1-7.

Gilbert, S., and Lynch, N. 2012. "Perspectives on the Cap Theorem," *Computer* (45:2), pp. 30-36.

Gray, J. 1981. "The Transaction Concept: Virtues and Limitations (Invited Paper)," in: *Proceedings of the seventh international conference on Very Large Data Bases - Volume 7*. Cannes, France: VLDB Endowment, pp. 144-154.

Grimson, J. 2001. "Delivering the Electronic Healthcare Record for the 21st Century," *International journal of medical informatics* (64:2), pp. 111-127.

Guo, J., Takada, A., Niu, T., He, M., Tanaka, K., Sato, J., Suzuki, M., Suzuki, T., Nakashima, Y., and Araki, K. 2005. "Enhancement of Mml Medical Data Exchange Standard for a Localized Chinese Version," *Journal of medical systems* (29:5), pp. 555-567.

Guo, J., Takada, A., Tanaka, K., Sato, J., Suzuki, M., Suzuki, T., Nakashima, Y., Araki, K., and Yoshihara, H. 2004. "The Development of Mml (Medical Markup Language) Version 3.0 as a Medical Document Exchange Format for Hl7 Messages," *Journal of medical systems* (28:6), pp. 523-533.

Halamka, J., Aranow, M., Ascenzo, C., Bates, D., Debor, G., Glaser, J., Goroll, A., Stowe, J., Tripathi, M., and Vineyard, G. 2005. "Health Care It Collaboration in Massachusetts: The Experience of Creating Regional Connectivity," *Journal of the American Medical Informatics Association* (12:6), pp. 596-601.

Heard, S. 2006. "Electronic Heatlh Records," in *Health Informatics: Transforming Healthcare with Technology*, M. Conrick (ed.). Melbourne: pp. 222-332.

Helland, P. 2011. "If You Have Too Much Data, Then 'Good Enough' Is Good Enough," *Commun. ACM* (54:6), pp. 40-47.

Hoerbst, A., Kohl, C.D., Knaup, P., and Ammenwerth, E. 2010. "Attitudes and Behaviors Related to the Introduction of Electronic Health Records among Austrian and German Citizens," *International journal of medical informatics* (79:2), p. 81.

Iakovidis, I. 1998. "Towards Personal Health Record: Current Situation, Obstacles and Trends in Implementation of Electronic Healthcare Record in Europe," *International journal of medical informatics* (52:1), pp. 105-115.

ISO. 2004. "Ts 18308 Health Informatics-Requirements for an Electronic Health Record Architecture."

Jin, Y., Deyu, T., and Xianrong, Z. 2011. "Research on the Distributed Electronic Medical Records Storage Model," *IT in Medicine and Education (ITME), 2011 International Symposium on*: IEEE, pp. 288-292.

Kohn, L.T., Corrigan, J.M., and Donaldson, M.S. 2000. *To Err Is Human: Building a Safer Health System*. National Academies Press.

Konishetty, V.K., Kumar, K.A., Voruganti, K., and Rao, G.V.P. 2012. "Implementation and Evaluation of Scalable Data Structure over Hbase," in: *Proceedings of the International Conference on Advances in Computing, Communications and Informatics*. Chennai, India: ACM, pp. 1010-1018.

Konstantinou, I., Angelou, E., Boumpouka, C., Tsoumakos, D., and Koziris, N. 2011. "On the Elasticity of Nosql Databases over Cloud Management Platforms," in: *Proceedings of the 20th ACM international conference on Information and knowledge management*. Glasgow, Scotland, UK: ACM, pp. 2385-2388.

Lakshman, A., and Malik, P. 2010. "Cassandra: A Decentralized Structured Storage System," *SIGOPS Oper. Syst. Rev.* (44:2), pp. 35-40.

Lee, K.K.Y., Tang, W.C., and Choi, K.S. 2013. "Alternatives to Relational Database: Comparison of Nosql and Xml Approaches for Clinical Data Storage," *Computer Methods and Programs in Biomedicine* (110:1), pp. 99-109.

Mengchen, Y. 2011. "Cassandra to Back Applications." Retrieved 27.08.2012, from http://salsahpc.indiana.edu/b534projects/sites/default/files/public/1_cassandra%20to%20back%20applications_Yu,%20Mengchen.pdf

Narayan, S., Gagne, M., and Safavi-Naini, R. 2010. "Privacy Preserving Ehr System Using Attribute-Based Infrastructure," in: *2010 ACM workshop on Cloud computing security workshop*. Chicago, Illinois, USA: ACM, pp. 47-52.

NEHTA. 2013. "Nehta - National E-Health Transition Authority." Retrieved 25.02.2013, 2013, from http://www.nehta.gov.au/

Orfanidis, L., Bamidis, P.D., and Eaglestone, B. 2004. "Data Quality Issues in Electronic Health Records: An Adaptation Framework for the Greek Health System," *Health informatics journal* (10:1), pp. 23-36.

Parker, Z., Poe, S., and Vrbsky, S.V. 2013. "Comparing Nosql Mongodb to an Sql Db," in: *Proceedings of the 51st ACM Southeast Conference*. Savannah, Georgia: ACM, pp. 1-6.

Pearce, C., and Haikerwal, M.C. 2010. "E-Health in Australia: Time to Plunge into the 21st Century," *Med J Aust* (193:7), pp. 397-398.

Pokorny, J. 2011. "Nosql Databases: A Step to Database Scalability in Web Environment," in: *Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services*. Ho Chi Minh City, Vietnam: ACM, pp. 278-283.

Sattar, A., Lorenzen, T., and Nallamaddi, K. 2013. "Incorporating Nosql into a Database Course," *ACM Inroads* (4:2), pp. 50-53.

Schiff, G.D., Klass, D., Peterson, J., Shah, G., and Bates, D.W. 2003. "Linking Laboratory and Pharmacy: Opportunities for Reducing Errors and Improving Care," *Archives of Internal Medicine* (163:8), pp. 893-900.

Schmitt, O., and Majchrzak, T.A. 2012. "Using Document-Based Databases for Medical in-Formation Systems in Unreliable Environments," *9th International ISCRAM Conference*, Vancouver, Canada.

Schram, A., and Anderson, K.M. 2012. "Mysql to Nosql: Data Modeling Challenges in Supporting Scalability," in: *Proceedings of the 3rd annual conference on Systems, programming, and applications: software for humanity*. Tucson, Arizona, USA: ACM, pp. 191-202.

Suciu, D. 2001. "On Database Theory and Xml," *SIGMOD Rec.* (30:3), pp. 39-45.

Takeda, H., Matsumura, Y., Kuwata, S., Nakano, H., Sakamoto, N., and Yamamoto, R. 2000. "Architecture for Networked Electronic Patient Record Systems," *International journal of medical informatics* (60:2), pp. 161-167.

Valduriez, P. 2011. "Principles of Distributed Data Management in 2020?," *Database and Expert Systems Applications*: Springer, pp. 1-11.

van der Linden, H., Kalra, D., Hasman, A., and Talmon, J. 2009. "Inter-Organizational Future Proof Ehr Systems: A Review of the Security and Privacy Related Issues," *International journal of medical informatics* (78:3), pp. 141-160.

van Ginneken, A.M. 2002. "The Computerized Patient Record: Balancing Effort and Benefit," *International journal of medical informatics* (65:2), pp. 97-119.

Vest, J.R. 2012. "Health Information Exchange: National and International Approaches," *Advances in health care management* (12), p. 3.

Vianu, V. 2001. "A Web Odyssey: From Codd to Xml," *Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*: ACM, pp. 1-15.

Walker, J., Pan, E., Johnston, D., Adler-Milstein, J., Bates, D.W., and Middleton, B. 2005. "The Value of Health Care Information Exchange and Interoperability," *HEALTH AFFAIRS-MILLWOOD VA THEN BETHESDA MA-* (24), p. W5.

Yasnoff, W.A., Humphreys, B.L., Overhage, J.M., Detmer, D.E., Brennan, P.F., Morris, R.W., Middleton, B., Bates, D.W., and Fanning, J.P. 2004. "A Consensus Action Agenda for Achieving the National Health Information Infrastructure," *Journal of the American Medical Informatics Association* (11:4), pp. 332-338.

COPYRIGHT