A COMPARISON OF SQL AND NoSQL DATABASES

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

OF

ATILIM UNIVERSITY

BY

SOUAD RASHD ALI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

THE DEPARTMENT OF SOFTWARE ENGINEERING

JANUARY 2018

Approval of theGraduate School of Natural and AppliedSciences, Atılım University.

Prof Dr. Ali KARA

Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Ali YAZICI

Head of Department

This is to certify that we have read the thesis A COMPARATIVE STUDY OF SQL AND NoSQL DATABASES Submitted by Souad Rashd and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Prof. Dr. Ali YAZICI

Supervisor

Examining Committee Members

Prof. Dr. Ali YAZICI

Asst. Prof. Dr. A. Murat ÖZBAYOĞLU

Asst. Prof. Dr. Çiğdem TURHAN

Date: 9 Ocak 2018

I declare and guarantee that all data, knowledge and information in this document has been obtained, processed and presented in accordance with academic rules and ethical conduct. Based on these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name: Souad Rashd Ali

Signature:

ABSTRACT

A COMPARISON OF SQL AND NoSQL DATABASES

SOUAD RASHD ALİ

M.S., Software Engineering Department

Supervisor: Prof Dr. Ali YAZICI

JANUARY 2018, 32 pages

This study comprises a research about both SQL and NoSQL databases and a comparison in terms of the advantages and disadvantages of each. The comparison is implemented depending on scalability, (ACID and CAP) theory, flexibility, performance, schema, query language, cost, speed and the stored data. We described briefly several examples of SQL databases including My SQL , MS-SQL Server Express Edition , Oracle 11g Express Edition Database. Also, the study included a description for many types of NoSQL database including Key-Value Store Database, Column-Oriented Database, Document Store Database and Graph Databases. The result of the study detected that SQL database is a table based database, supports flexibility, is vertical scalable, and utilizes Structured Language Query. On the other hand, the study discovered that SQL database is not a suitable choice for hierarchical data storage, etc. Whereas in terms of NoSQL, the study discovered that NoSQL databases are less flexible, however, horizontally scalable, have a non-structured querying language and is a good choice for hierarchical databases.

Keywords: SQL Databases, NoSQL Databases, Key-Value Stores, Column-Oriented Databases, Graph Databases, Document Store Databases

ÖZET

SQL ve NoSQL VERİTABANLARININ KARŞILAŞTIRILMASI

Souad Rashd

Yüksek Lisans, Yazılım Mühendisliği Bölümü

Tez Yöneticisi: Prof. Dr. Ali YAZICI

OCAK 2018, 32 sayfa

Bu çalışma SQL ve NoSQL veritabanlarının bir karşılaştırmasını ve avantajları ile dezavantajlarının belirlenmesine yönelik bir araştırmayı içerir. Karşılaştırmalar her iki veritabanı türünü ölçekleme, (ACID ve CAP) kuramı, esneklik, başarım, şema, sorgulama dili, maliyet, hız ve very açısından değerlendirmektedir. Çalışmada, MySQL, MS-SQL Server Express Edition, Oracle 11g Express Edition gibi SQL veritabanlarına örnekler verilmiştir. Ayrıca, çalışmada NoSQL veritabnalarına örnek olarak, Anahtar-Değer Depo veritabanları, Kolon-Yönelimli Veritabanları, Doküman Depol Veritabanları ve Dizge Veritabanları kısaca örneklerle anlatılmıştır. Çalışmanın sonuçları ve karşılaştırmalar SQL vertiabanlarının tablo tabanlı sistemler olduğunu, esnekliği, düşey ölçeklemeyi ve yapısal sorgulama dili SQL'i desteklediğini ortaya koymuştur. Diğer taraftan, bu çalışma, SQL sıra düzensel veri yapıları için uygun olmadığını, ve NoSQL veritabanlarının daha az esnek olmasına ragmen, yatay ölçeklenebilir, yapısal olmayan sorgulama dili özellikleri ile, sıra düzensel veritabanı yapıları için daha uygun olduğu ortaya çıkartmıştır.

Anahtar Kelimeler: SQL Veritabanları, NoSQL Veritabanları, Anahtar-Değer Depolama, Kolon-Yönelimli Veritabanları, Dizge Veritabanları, Doküman Depo Veritabanları

To My Family and Friends…

ACKNOWLEDGMENTS

I express sincere appreciation to my supervisor Prof. Dr. Ali YAZICI for his guidance and insight throughout the research. Thanks also to my husband, my father and my mother for their continuous support and patience.

CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ACID	-	Atomicity, Consistency, Isolation and Durability
BASE	-	Basic Availability, Soft State and Eventual Consistency
CAP	-	Consistency, Availability and Partition tolerance
DB	-	Database
NoSQL	-	Not Only Structured Query Language
RDBMS	-	Relational Database Management Systems

CHAPTER 1

INTRODUCTION

1.1 Basic Concepts and Definitions

Recently, everything became connected by the internet. This rapid development with the computer and internet technology led to an effective store and retrieval of information. The experimental results and huge amounts of online transactions led to obtain great amounts of information which require appropriate storage solutions. In order to get a successful storage method of information and data retrieved correctly, the role of the database comes into use. The most successful databases which are designed for this purpose are SQL and NoSQL databases.

The SQL databases completely depend on the relational tables to store the data and that's the reason behind its name as the relational database. Each table in the relational database can be divided into number of rows and columns depending on their information. The row in the relational database refers to the record while each column refers to the field. The relational table can use common columns and foreign keys in order to link the tables with each other. The simplest entities which can be represented in the relational tables are the departments, employees and so on. This simple scheme is useful when designing the database scheme in the real world and mapping them to the database, in addition to mapping the relations between them. Nevertheless, because of the continuous growth of the stored and analyzed data, the relational databases include variety of limits such as the storage, scalability restrictions and particularly losing of query as the volume of data is very huge, and the storage and management of greater databases become challenging. The first companies which discovered the serious weaknesses of SQL database to support the great user requirements and big data are LinkedIn, Amazon, Google and Facebook. Therefore, the NoSQL databases are configured in order to overcome the limitations and weaknesses which exist in SQL databases. The NoSQL databases include many features such as that they do not have a stable schematic relative structure, the requirements determine the different fields of each record dynamically.

NoSQL databases are considered non-relational where they access the data by the use of the key-value pairs rather than the process of storing the data in rows and columns as in the case of conventional databases. Each item in NoSQL database is stored separately with a unique key value. Moreover, it does not need to a structural scheme which defines each table and the associated columns. Thus, NoSQL databases delivers more flexible methods in order to store the data than the relational databases.

As shown in Figure1.1, the SQL database is designed to work in one machine. Thus, SQL needs a large machine to work properly. In order to solve this problem, we need larger number of machines to store the data. Hence, NoSQL databases are considered as cheaper alternatives to SQL databases and in addition they are horizontally scalable. If one of the devices stop working, the entire reliability of the cluster will stay reliable.

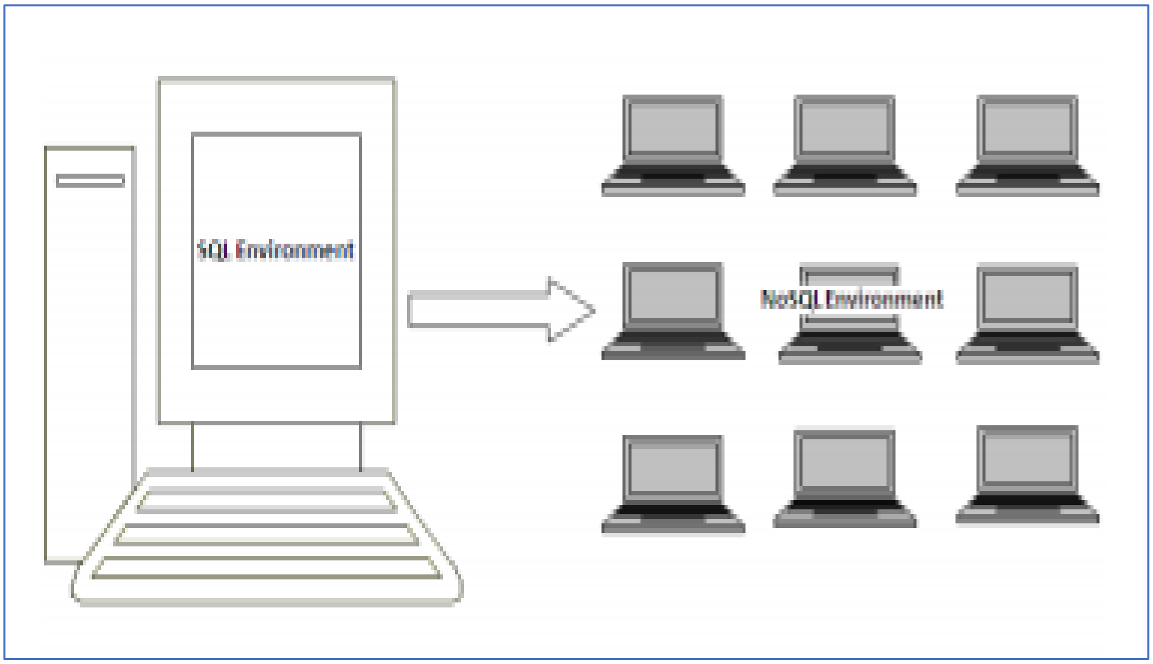


__Figure 1.1__ SQL and NoSQL environment adapted from (Sharma and Dave, 2012)

1.2 SQL Database

In general, the SQL databases are called Relational Databases (RDBMS). The SQL databases depend on the relation (table) to store the data and that is why they are known as the relational databases. The relational (table) is separated into a number of columns and rows for the purpose of storing the data. In the relation (table), the rows represent the record and the columns represent the fields. The foreign key or a common column is used to link the tables to each other. The relationship which exists between the field and tables is known as the scheme. The database scheme in SQL must be defined obviously before adding any information. The data which must be stored in the relational databases must be organized in order to make the relational database effective.

1.3 NoSQL Databases

NoSQL database is considered as a class of database management systems (DBMS) which does not follow the whole relational DBMS rules and they do not use the conventional SQL database to query data. Also, NoSQL is not just a SQL database since it is not a relational database and it is produced for the purpose of working with large datasets. In addition, it is considered as a high performance database and it is open source. NoSQL is currently used by the industries to administrate the unstructured databases. It is without a scheme and it does not work based on the rows and columns as the case of SQL. Moreover, it does not organize its data in associated tables.

1.4 Types of SQL and NoSQL Databases

1.4.1 SQL Databases

MySQL Community

MySQL is currently owned by the Oracle Company and it is created by Swedish company (Yank, 2004). MySQL is used frequently in the web applications where it is the most common database which deals with the web applications. This database is a common open source database which means that one can enhance this database and it is free and used by many sites such as Facebook.

MS-SQL Server Express Edition

MS-SQL is produced by Microsoft Company and fulfils stability, reliability and scalability requirements. It is a trusted, powerful and can easy work with database (Yank, 2004). Microsoft SQL Server Express is free version of Microsoft SQL and it is a relational database management system where it can be downloaded, distributed and used freely. It is designed and implemented to work with embedded and small size applications.

Oracle 11g Express Edition Database

This is a distributed database with advanced properties. It can be downloaded quickly and managed simply (Karunakaran et al., 2016; Kothuri et al., 2008). It is an entry-level, database based on the Oracle Database 11g Release 2 code base (Buelow and Kasi, 2009).

1.4.2 NoSQL Databases Types

Key-Value (KV) Store Databases

This type of databases has a user-friendly Application Programming Interface (API) and it is easy to be used. The key-value databases do not include a fixed scheme (schema less). KV stores records in only two columns. The key is always stored as a string and it is contained in one of the columns while the actual data is stored in the other column. When the value of the key is known, the user can access the value of the data. It is not possible to determine the value of the key when determining the data value. The value data always includes of primitive data type which can be integer, or string. The typical key value is shown in Figure 1.2.

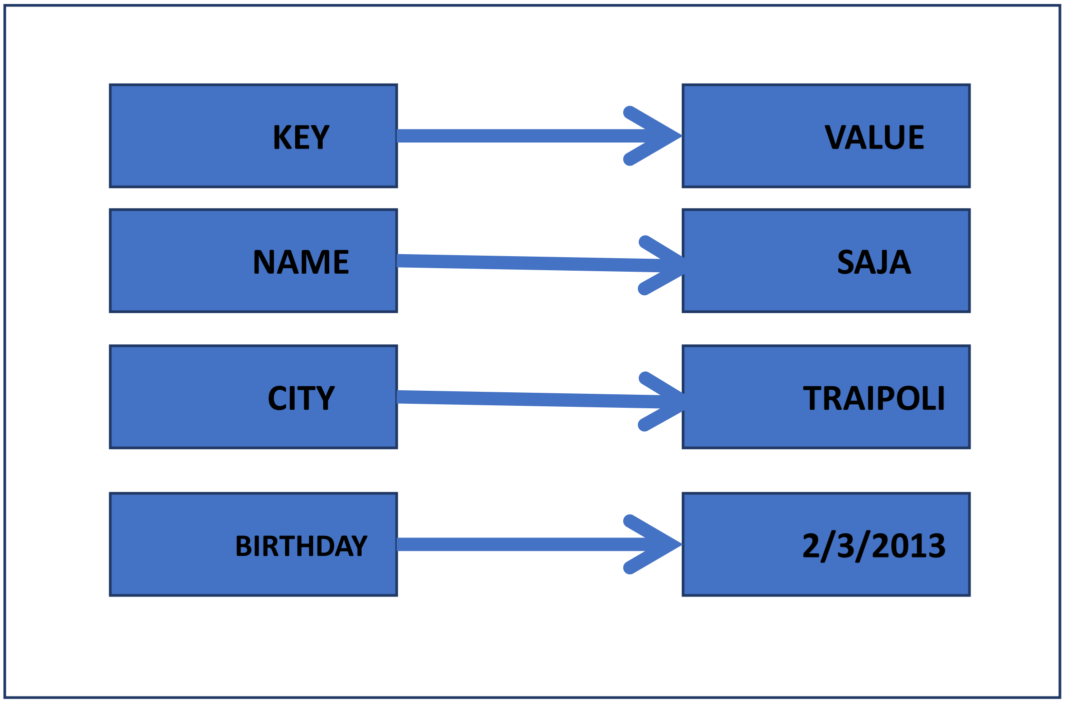
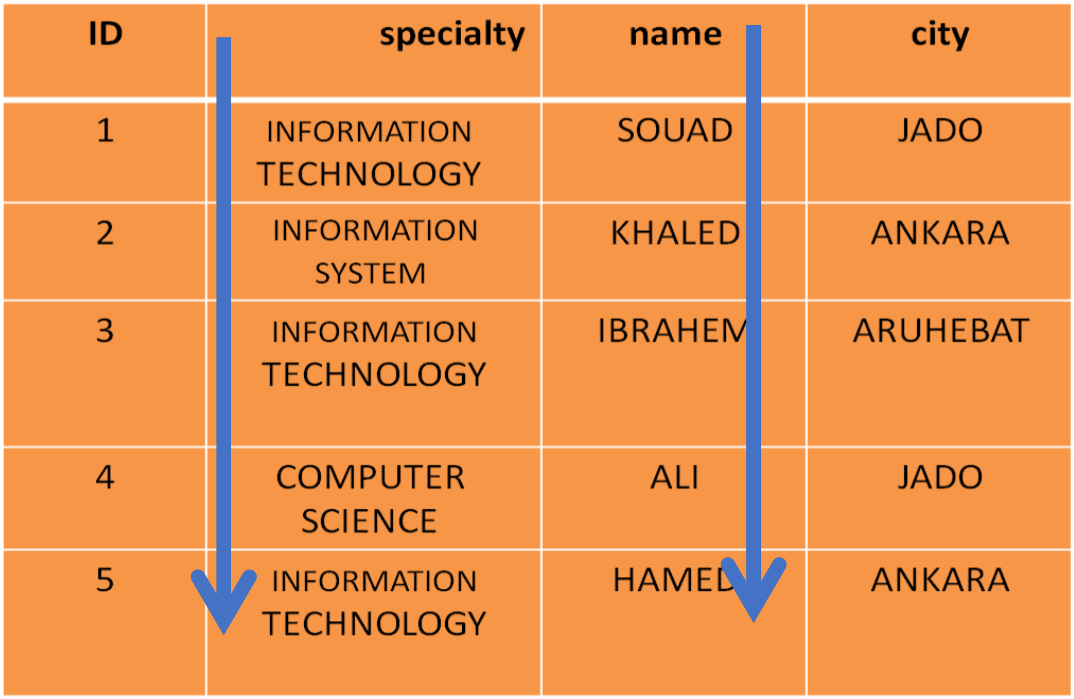


Figure 1.2 *An example of a Key-Value database*

Column-Oriented Databases

The data which is stored in this type of databases are stored in the form of columns instead of rows (Li and Manoharan, 2013; Han et al., 2011). This type of database is a DBMS which stores the data in columns rather than rows (Stonebraker et al., 2005). The purpose of it is to effectively read and write the data from the hard disk in order to have a faster data access; (Abadi et al., 2006). An example of a column-oriented database is shown in Figure 1.3.



ID	specialty	name	city
1	INFORMATION TECHNOLOGY	SOUAD	JADO
2	INFORMATION SYSTEM	KHALED	ANKARA
3	INFORMATION TECHNOLOGY	IBRAHEM	ARUHEBAT
4	COMPUTER SCIENCE	ALI	JADO
5	INFORMATION TECHNOLOGY	HAMED	ANKARA

Figure 1.3 *An example of a Column-Oriented database*

As shown in Figure 1.3, if one wants to search about a specific student in the information technology specialty by the use of SQL databases, one has to search each row and column starting from the ID of the student in the ID columns and ending by the name of CITY in the CITY column. Thus, it needs a long time if the table includes excessive amount of records. While in case of the column-oriented database, the search will be done in the *specialty* and *name* columns only from the top to bottom. This process will save time if the table is large and include hundreds of columns and millions of rows.

Document Store Databases

This type of database is yet another type of NoSQL where a table is used to store all the given data in an entity within a single document. This document includes all the related stored data. Any two documents in the document store database may include different data structures data. Also, it can scale in a horizontal way and the data can be stored in thousands of nodes (computers) with good performance of the system. The documents store database can be queried with the use of many languages including Java and Java Script. An example of a document store database is shown in Figure 1.4.

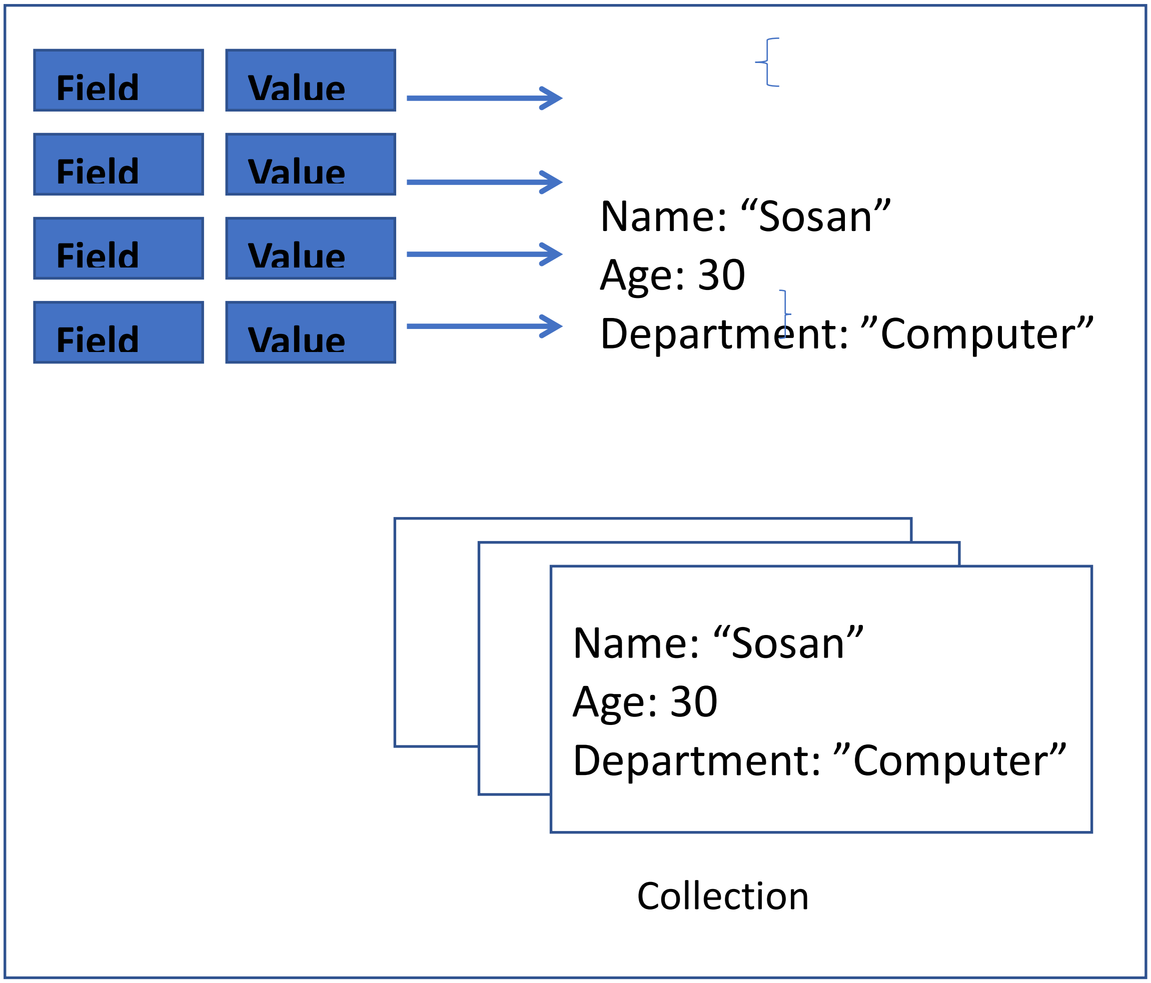


Figure 1.4 *An example of a Document Store database*

As shown in Figure 1.4, the document store database includes many collections where all of the collections include a set of documents and each document includes a set of fields and values. The field differs from one document to another.

Graph Databases

The data with the graph database is stored in the form of graphs. The graph includes nodes that act as objects and edges which denote the relationship between the objects. In addition, the graph comprises features which associate to the nodes. In terms of speed, the graph databases are quicker than the relational databases and they are scalable easily where the queries are conveyed as traversals. There are many applications which use this type of databases including content management and social networking sites. An example of graph database is shown in Figure 1.5.

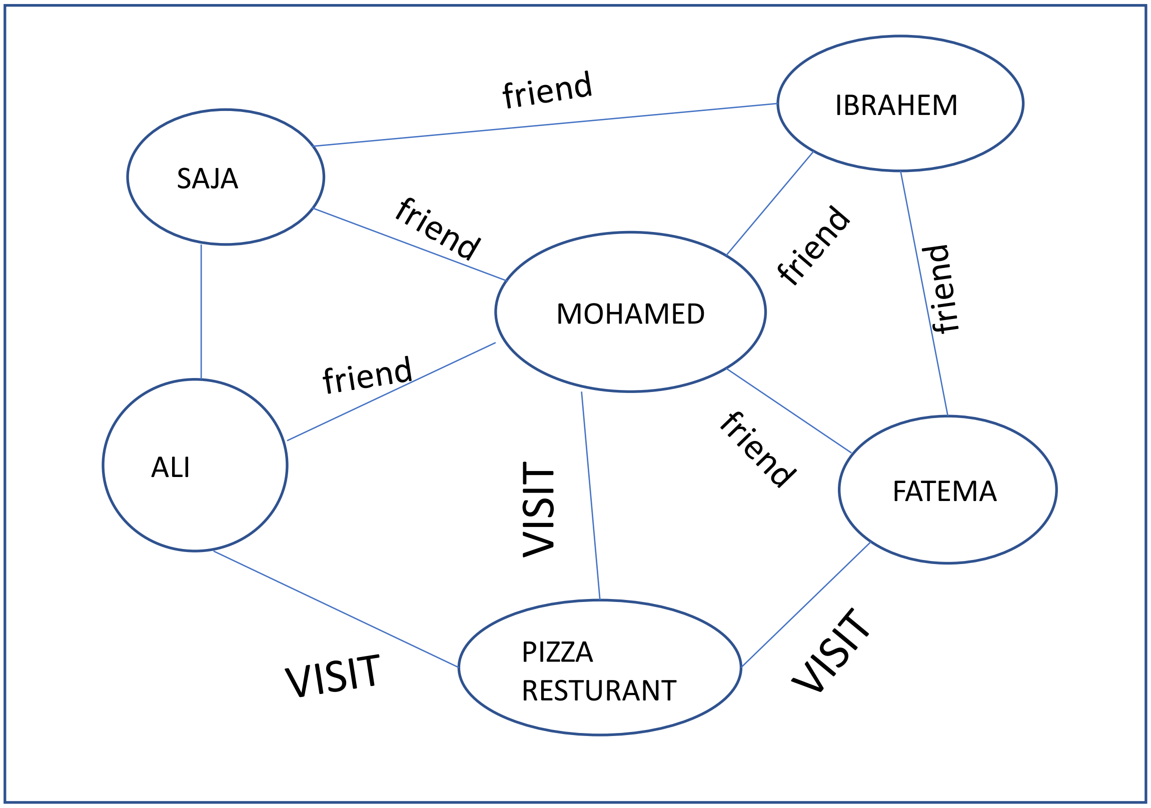


Figure 1.5 *An example of a Graph database*

Figure 1.5 illustrates an example of the graph database where Fatema is a *friend* with Mohamed, and Mohamed is a *friend* with Ali. Fatema, Mohamed and Ali *visit* a Pizza restaurant. We conclude that both Ali and Fatema may have met before and thus, we can suggest to Mohammed to add Fatema as a friend and vice versa.

CHAPTER 2

LITERATURE REVIEW

In this chapter, we will briefly discuss the main differences between the SQL and NoSQL databases by reviewing the publication about the subject from the literature. In addition, some of the advantages and disadvantages of the relational and non-relational databases according to the opinions of the authors in many scientific papers will be given.

2.1. Literature review

According to (Pore and Pawar, 2015) SQL is the standard inquiry language for relational database management structure. The relational database management system includes many typical categories including SQL Server, MS Access, MySQL, and Oracle. NoSQL is the Not Only SQL which is addition of the non-relational information storing frameworks (Pore and Pawar, 2015). The entire SQL offers the ACID characteristics (Sharma and Dave, 2012). Currently, each institution includes some portion of SQL in their IT framework (Leavitt, 2010). Furthermore, it was announced that MySQL is a basic part of SQL which is used widely in implementing the websites including Facebook and other important sites. Moreover, it was stated that the important center of NoSQL databases is more on BASE (Basically available, Soft state eventually Steady) than relational database management system ACID (Atomicity, Consistency, Isolation, and Durability) properties. NoSQL databases is considered to be a non-relational database where they access the data by the use of the key-value rather than the process of storing the data in rows and columns as the case of the conventional databases (Chow, 2014); (Greenspan and Bulger, 2001).

(Sharma and Dave, 2012) have stated that the NoSQL database is used in order to store large quantity of information. NoSQL is carried, non-relational, open source and it is on a level plane scalable. It differs from SQL databases in many forms including that it does not include the ACID feature which exists in SQL database.

NoSQL database is considered a class of database management systems (DBMS) which does not follow the whole relational DBMS rules and they do not use the conventional SQL database to query data (Li and Manoharan, 2013). Some of them are simple to be instantiated, quick to be examined, composed and erased. Whereas the others are fast to be instantiated yet reasonable on alternative processes. In addition, there is a little relationship in each database between the implantation and displaying of information (Li and Manoharan, 2013).

(George, 2013) provided the domain of NoSQL, the comparative of working tools and management framework for the great storage of information (George, 2013). Succeeding the NoSQL solutions analysis on the market, the MongoDB system has been selected as the goal application for this stage of progression, for explanations that are associated to verified presentation, pliability, make time to market and ease of use (George, 2013). Also, he stated that MongoDB is a fast and flexible database. The information can be put away as the document that allows the representation of the changeable influences in the private article documents created from individual fields with the sort of key information including records or archives vectors (George, 2013). This ability will enable the engineer to present a wide diversity of subjects in a flexible way without decreasing back on extrication information into many tables.

NoSQL databases are most widely used in online applications (Karunakaran et al., 2016). It is still not much used in standalone database applications. There is an organized way that needs to be implemented for NoSQL databases so that it can completely be ready to be replaced RDBMS.

NoSQL database is a good way for big data application but queries in NoSQL is not properly implemented and not standardized. So queries must be properly implemented using some compiler (Oliveira, 2014).

(Al Hinai, 2016) stated that the framework of non-relational database denoted as NoSQL have advanced option phases to store, stack and dissect Big Data. Most popular frameworks of NoSQL, for instance, MongoDB, Redis, HBase, and Cassandra hand over reliability for adaptability which suggest that clients will be incompetent to improve the most new deviations of the information. On the other hand, he stated that these frameworks provide what is known as reliability of possible information.

NoSQL DB represents another technique in order to put away and recover information and data (Consistent with Pretorius, 2013). Also here in this study, it has been stated that the NoSQL database is designed in order to scale out whereas the relational databases are designed to scale up. The vertical scaling which belongs to the relational databases include the addition of more hardware power (Pretorius, 2013). This may assess the institutions in order to choose if NoSQL can be a practical solution and regulate business and data invention purposes. This can help institutions to promise that the dangers recognized when using the NoSQL databases are suitably tended (Pretorius, 2013). Furthermore, he proposed that the NoSQL database type is the most suitable selection for the organizations which depend heavily on internet.

The NoSQL reliability can be more beneficial to give superior fitting to the basics of definite web-based organizations than the use of the traditional databases. The NoSQL database can be a great choice for the organizations with varying size (Pretorius, 2013). NoSQL database does not have the ability to displace each business' relational database. The NoSQL can be used as an additional tool and take into consideration the final goal of the business (Pretorius, 2013).

In (Pore and Pawar, 2015) it was mentioned that NoSQL databases are used when one has to work on unstructured data sets such as images. In the SQL database one is not able to write, read and update unstructured data such as images, videos where NoSQL databases can be used to read, write and update this sort of data.

CHAPTER 3

COMPARING BETWEEN SQL AND NoSQL DATABASES

This chapter will include a comparison between the SQL and NoSQL database according to specific characters including properties and type of both SQL and NoSQL databases.

3.1. The comparative phases between SQL and NoSQL databases

3.1.1. SQL databases type's advantages and limitations

This part includes the types of SQL database which are explained in Table A-1. Table A-1 describes the types of SQL databases according to each author and what type of SQL databases is pronounced by the authors and which system pragmatic to use SQL databases to guarantee the quality promise to use databases. Also, in this section the limitation of SQL database and the advantages of using this type of databases are given. Additionally, this section includes the description of the companies and even the countries which use or start using this kind of databases. In case of MySQL, it includes many advantages such as repetition, sharing, reduced cost, scalability and obtainability. MySQL databases can be presented by the use of Apache and PHP, server side Java scripting method using NodeJs consistent (Pore and Pawar, 2015; Greenspan and Bulger, 2001; Yank, 2004 ; Harris et al., 2009; Tuya et al., 2006; Shute et al., 2013). Also, the study displays the main differences between MySQL and MS-SQL Server in terms of consistency, scalability and constancy requirements and combined development environment. The profitability of the designers can be increased by the use of Microsoft visual studio, SQL Server Management Studio, and Visual Developer tools, disaster Recovery mechanism (Pore and Pawar, 2015), (Greenspan and Bulger, 2001), (Yank, 2004), (Harris et al., 2009).

Another type of SQL database is Oracle 11g Express Edition Database that is characterized by its fast downloading, easily upgrading, easily manageable, productivity, and reliability. It can be integrated to a new version (Pore and Pawar, 2015).

3.1.2. NoSQL Database types, advantages and limitations

In this section, we will present the types of NoSQL database to ensure the quality advertisement of the databases which are used in the available scientific papers (conferences, journals, books) during the last ten years. Table A-2 presents the types of NoSQL and the practices associating to each researcher.

The purpose of this explanation is to make full description of the NoSQL databases which is explained by many researchers and to guide the researchers, companies and institutions who aim to start using this type of databases and expand their knowledge about them. As shown in Table A-2, NoSQL database include various types of databases including Key-Value Store Databases that includes many advantages in order to motivate most of the researchers who interest to study this type of NoSQL databases where its advantages are quick and store the information and data on solid state drives and rise the performance. Moreover, NoSQL databases can lodge easily in any datatype. Nevertheless, this type of NoSQL databases include one disadvantage which has not fixed scheme as explained by (Pore and Pawar, 2015; Li and Manoharan, 2013; George, 2013; Pretorius, 2013; DeCandia et al., 2007).

The Key-value database looks like the hash table where its key is used as indexes. Key-value is used in many large companies including Amazon which is implemented by the use of Amazon's Dynamo Model, Basho technologies developed by the use of C, Java script, and Erlang, Riak clients, Riak search, Web machine, Riak KV, Riak core, Riak SNMP/JMX, Riak Pipe, Riak can be used to collect and check point of sales i.e. POS data, preserving user's personal information on social networking sites etc. if compared with the Column-Oriented Databases, it includes various advantages including reducing the disk access associated to relational table, which includes of column and rows with constant sized fields for each record, and high level of obtainability. However, Cassandra includes one disadvantage is that reading the operations are reasonably slower than the write operations. Also, according to (Pore and Pawar, 2015; George, 2013; Al Hinai, 2016; Pretorius, 2013) this kind of NoSQL databases can be used in Column-Oriented Databases, HBase developed after Google's Big Table. Also, it is open source, distributed and non-relational databases. And explained another type of NOSQL DB the graph database, which comprises features which associate to the nodes. In terms of speed, the graph databases are quicker than the relational databases and they are scalable easily where the queries are conveyed as traversals. There are many applications which use this type of databases including content management and social networking sites (Pore and Pawar, 2015; George, 2013; Chow, 2014; Pretorius, 2013).

3.1.3. SQL and NoSQL properties and authors

This part includes a comparison between the SQL and NoSQL databases according to the most significant characteristics which have been given bellow:

- ACID and CAP
- BASE
- Scalability
- Language Query
- Flexibility
- Schemas
- Stored data
- Reliability
- Cost

The comparison details are shown in Table A-3.

(ACID AND CAP) Properties:

The ACID properties is focused by SQL databases (**A**tomicity, **C**onsistency, **I**solation and **D**urability) while NoSQL databases follow the Brewers CAP theorem (**C**onsistency, **A**vailability and **P**artition tolerance). The ACID model is considered one of the oldest and significant models of databases. In order to guarantee the reliability, there are four goals which must be achieved as follow:

1. Atomicity states that the transaction must be either implemented completely or not at all implemented. If only part of the transaction is failed, the whole transaction must be failed.
2. Consistency State that just valid data must be written to the database. The entire transaction will be returned back in case of implementing a transaction which

violates the rules of database consistency. If the transaction is been implemented successfully, the database will be taken from one reliable status to another.

3. Isolation states that the transaction will be implemented independent from each other. For instance, if Anwar disputes a transaction on database ‘A’ and Lars disputes other transaction on the same database at the same time, both of the transactions must work on the database in a separated method.
4. The durability will ensure that the transaction which has been dedicated to the database will not be missed. The database backup and transaction logs must be employed in order to guarantee the durability.

CAP (Consistency-Availability-Partition Tolerance)

Consistency: the user will get the same data regardless of the type of server which has sent answer to the user request.

Availability: the server will answer the user request even if by sending a message which has not been requested by the user or just by sending a message which says that the server does not work at the moment.

Partition Tolerance: the system will continue its work effectively even if the singular servers are failed or has not been reached.

BASE

Basically available: this features promising to be available most of the time. Thus, there is no guarantee with hundred percent that the request which is directed to the database will have answer that the transaction is succeeded or failed.

Soft State: the status of the system may change without any inputs from the user and this is caused by the ultimate stability.

Ultimate stability: this refers that the system may not be stable to any certain point during time. However, the system tries to be stable ultimately when the data have not to be exactly signified through the entire system.

In general, SQL inclines to be particularly ACID, but NoSQL inclines to track BASE more. However, it is likely that NoSQL is entirely ACID and only the model of data differs. Thus, it is clear that the NoSQL database is completely different from SQL database where they existed in much larger sets.

Scalability

The NoSQL database is designed in order to scale out whereas the relational databases are designed to scale up. The vertical scaling which belongs to the relational databases include the addition of more hardware power. This involves the purchase of expensive proprietary servers with high power. These servers have the ability to deal with high traffic and heavy capacity to get bogged down when allocating with possibly millions of concurrent users for example with large names of websites such as Facebook or Amazon. This workload requires a large number of servers which have the ability to work simultaneously and preserve the storing of data in an organized manner which is difficult and may be impossible on a large scale. This matter is possible with NoSQL which works on horizontal scale and the data storage and workload are maintained on a large number of typical or virtual machines. The horizontal scale which works with NoSQL permits the workload to be distributed on a large number of nodes which make the large web servers like those existed in Facebook able to handle possibly hundred-millions or more simultaneous users.

Query Language

Structured query language is used by SQL database in order to define and manipulate the data where it is very powerful while the queries in NoSQL are focused on collecting the documents. NoSQL is sometimes called UnQL (Unstructured Query Language). The syntax which is used by NoSQL differs from one database to another.

Flexibility

There is huge difference between NoSQL and SQL databases. The data in the SQL database are split into several interrelated tables which include rows and columns. The SQL databases include the foreign key stored in columns in the same database and it will be the reference between one table and another in the same database. When the user runs a specific query on a set of data, he needs to collect the required information from more than one table and currently the data may be collected from hundreds of tables in the applications of the large enterprises before it can be delivered to run a specific application.

Likewise, when the data is written to specific database, it will be distributed on several tables. When the data is delivered to the database with small size, the SQL database will be able to collect and store them with high speed. However, the applications in today's world are constructed on the prospect that huge volume of data must be read and written in speed close to the real time. NoSQL database involves many models which differ from one to another. Specifically, NoSQL database are really "NoREL," or non-relational, significance where they do not depend on the tables and the relations between tables for the purpose of store and retrieving the information from the database.

Schema

The scheme in SQL database are the tables and fields which must be defined before the user is able to add any data to them. The scheme of the database must be designed and executed before developing any logics to deploy data. In SQL databases, the updates can be implemented later with the ability of expecting large and complicated changes. Whereas in terms of NoSQL, the data can be added in anywhere when there is no convincing aim to regulate a document design or even collecting in progress.

Stored data

In terms of the hierarchical data storage, the SQL database is not considered a suitable selection. So, the NoSQL is considered better selection for the hierarchical data storage where the NoSQL follow the key-value pair method to store data content which is related to JSON data. NoSQL databases is the most preferred selection for the large set of data.

Reliability

The reliability in NOSQL is less than the reliability in SQL. So, there is some specific mechanism which emphasizes on all columns must contain data for each row.

Cost

NOSQL database is cheaper than SQL database. SQL database tends to be expensive proprietary servers while NOSQL database use a set of cheap commodity servers and quite expensive licenses for SQL database are required while many NOSQL database are open source and free.

CHAPTER 4

CONCLUSIONS, LIMITATIONS AND FUTURE WORK

4.1. Conclusions

The main goal of this study was to assess the basic differences between SQL and NoSQL databases and the relative analysis of these two types of databases. In addition, the study describes some of the types of NoSQL and SQL databases. In case of SQL, it is considered as a vertically scalable database while the NoSQL database is considered horizontally scalable. There are many differences between SQL and NoSQL databases and the type of application and query which is used determine the use of each one of them. The volume and the type of query which will be implemented in specific application determines the performance of the database. Chapter 3 of this study included the presentation of the whole relative types of SQL and NoSQL databases and features.

As a result, SQL and NoSQL perform the same performance in different methods. It is probable to use one of them and then move to use another one later but the little planning can save the money and time. Each one of these two databases have advantages and disadvantages and characteristic which differentiate them from each other. The use of any type of the databases is determined by the requirements of the project or the system where most of the companies depend on more than one type of databases and in some cases more than one database in the same application or system.

4.2. Limitations

The study is based on the literature and related resources in order to perceive the main reason to use SQL and NoSQL types of database. In additions, this comparative study is performed by following the SQL and NoSQL types of databases and exploring the benefits and limitations to use SQL and NoSQL databases. The focus of the study is implemented by making a comprehensive comparison between the SQL and NoSQL databases. This study may provide benefits to the researchers and organizations which aim to use SQL and NoSQL types of databases in order to ensure the quality guarantee of using such applications at any field.

4.3. Future study

NoSQL databases are still developing and require some time to get mature in order to handle the unstructured data emerging form IoT applications. A systematic comparison between SQL and NoSQL databases would make more sense if certain case studies are considered in real situations where database users switch from one to another. A questionnaire needs to be designed to measure the effectiveness of each in the aforementioned case studies.

REFERENCES

Abadi, D., Madden, S., & Ferreira, M. (2006, June). Integrating compression and execution in column-oriented database systems. In Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data, pp.671-682.

Atikoglu, B., Xu, Y., Frachtenberg, E., Jiang, S., & Paleczny, M. (2012, June). Workload analysis of a large-scale key-value store. In ACM SIGMETRICS Performance Evaluation Review, 40(1), pp. 53-64.

Al Hinai, A. H. (2016). A Performance Comparison of SQL and NoSQL Databases for Large Scale Analysis of Persistent Logs.

Abramova, V., & Bernardino, J. (2013, July). NoSQL databases: MongoDB vs Cassandra. In Proceedings of the international C* conference on computer science and software engineering (pp. 14-22). ACM.

Angles, R. (2012, April). A comparison of current graph database models. In Data Engineering Workshops (ICDEW), pp. 171-177, IEEE.

Buelow, H., & Kasi, J. (2009). Getting Started with Oracle SOA Suite 11g R1, a Hands-on Tutorial: Fast Track Your SOA Adoption, Build a Service-oriented Composite Application in Just Hours! Packt Publishing Ltd.

Chow, M. (2014). Abusing NoSQL Databases. lOLl https:/Iwww. co. tt/files/defcon21 ISpeaker% 20Material s/M ing% 20Cho wIDEFCON-21-Chow-Abusing-NoSQL-Databases.pdf.

Cattell, R. (2011). Scalable SQL and NoSQL data stores. ACM Sigmod Record, 39(4), pp.12-27.

DeCandia, G., Hastorun, D., Jampani, M., Kakulapati, G., Lakshman, A., Pilchin, A., & Vogels, W. (2007). Dynamo: Amazon's highly available key-value store. ACM SIGOPS operating systems review, 41(6), pp.205-220.

Dominguez-Sal, D., Urbón-Bayes, P., Giménez-Vanó, A., Gómez-Villamor, S., Martínez-Bazan, N., & Larriba-Pey, J. (2010). Survey of graph database performance on the HPC scalable graph analysis benchmark. Web-Age Information Management, pp.37-48.

Greenspan, J., & Bulger, B. (2001). MySQL/PHP database applications. John Wiley & Sons, Inc. George, T. B. (2013). A New Application for the Management of the MongoDB Servers. International Journal of Sustainable Economies Management (IJSEM), 2(3), pp.58-71.

Han, J., Haihong, E., Le, G., & Du, J. (2011, October). Survey on NoSQL database. IEEE Proceedings of the 6th International Conference on Pervasive computing and Applications (ICPCA), pp.363-366.

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. Journal of biomedical informatics, 42(2), pp.377-381.

Han, J., Haihong, E., Le, G., & Du, J. (2011). Survey on NoSQL database. Proceedings of the 6th International Conference on Pervasive computing and Applications (ICPCA), pp.363-366.

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. Journal of Biomedical Informatics, 42(2), pp.377-381.

Idreos, S., Groffen, F., Nes, N., Manegold, S., Mullender, S., & Kersten, M. (2012). MonetDB: Two decades of research in column-oriented database architectures. A Quarterly Bulletin of the IEEE Computer Society Technical Committee on Database Engineering, 35(1), pp.40-45.

Kothuri, R., Godfrind, A., & Beinat, E. (2008). Pro Oracle spatial for Oracle Database 11g. Dreamtech Press.

Karunakaran, P., Seshagiri, V., Lakshmana Vadaga, M. & Shah, J. (2016) Data Migration Methodology from SQL to Column Oriented Databases (HBase)", IJARCET, 5(11), pp. 2631-2635.

Li, Y., & Manoharan, S. (2013). A performance comparison of SQL and NoSQL databases. In Communications, computers and signal processing (PACRIM), 2013 IEEE Pacific RIM Conference, pp.15-19.

Lee, K. K. Y., Tang, W. C., & 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.

Lim, H., Fan, B., Andersen, D. G., & Kaminsky, M. (2011, October). SILT: A memory-efficient, high-performance key-value store. In Proceedings of the Twenty-Third ACM Symposium on Operating Systems Principles, pp.1-13.

Leavitt, N. (2010). Will NoSQL databases live up to their promise? IEEE Computer, 43(2), pp.12-14.

MySQL, A. B. (2011). MySQL community server.

Oliveira, P. M. M. (2014). High throughput storage for health devices real-time information.

Pretorius, D. J. (2013). NoSQL database considerations and implications for businesses (Doctoral dissertation, Stellenbosch: Stellenbosch University).

Pore, S. S., & Pawar, S. B. (2015). Comparative Study of SQL & NoSQL Databases. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 4(5), pp.1747-1753.

Riesen, K., & Bunke, H. (2008). IAM graph database repository for graph based pattern recognition and machine learning. Structural, Syntactic, and Statistical Pattern Recognition, pp.287-297.

Sattar, A., Lorenzen, T., & Nallamaddi, K. (2013). Incorporating NoSQL into a database course. ACM Inroads, 4(2), pp.50-53.

Shetty,D, J.Chidimar. (2016). Comparative Study of SQL and NoSQL Databases to evaluate their suitability for Big Data Application, International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online) Vol. 4, Issue 2, pp.314-318, Month: April - June 2016, Available at: www.researchpublish.com

Sharma, V., & Dave, M. (2012). SQL and NoSQL databases. International Journal of Advanced Research in Computer Science and Software Engineering, 2(8), pp.598-601.

Stonebraker, M., Abadi, D. J., Batkin, A., Chen, X., Cherniack, M., Ferreira, M., & O'Neil, P. (2005, August). C-store: a column-oriented DBMS. In Proceedings of the 31st international conference on Very large data bases, pp. 553-564.

Strauch, C., Sites, U. L. S., & Kriha, W. (2011). NoSQL databases. Lecture Notes, Stuttgart Media University, 20.

Sumathi, S., & Esakkirajan, S. (2007). Fundamentals of relational database management systems, Springer, Berlin.

Shute, J., Vingralek, R., Samwel, B., Handy, B., Whipkey, C., Rollins, E., & Cieslewicz, J. (2013). F1: A distributed SQL database that scales. Proceedings of the VLDB Endowment, 6(11), pp.1068-1079.

Tuya, J., Suarez-Cabal, M. J., & De La Riva, C. (2006, November). SQLMutation: A tool to generate mutants of SQL database queries. IEEE Second Workshop on Mutation Analysis, pp. 1-5.

Yank, K. (2004). Build your own database driven website using PHP & MySQL. SitePoint Pty Ltd.

Zou, L., Chen, L., & Özsu, M. T. (2009). Distance-join: Pattern match query in a large graph database. Proceedings of the VLDB Endowment, 2(1), pp.886-897.

APPENDIX - A

Table A-1. **SQL databases types and advantages and limitations**

SQL Databases types	Advantages	Disadvantges	When can be used	Authors / years
MySQL	1.Replication 2.Sharding 3. Free download 4.Maturity 5. Available for all major operating systems 6.Reduce cost 7.İncrease scalability and availability		used with apache and PHP, server side java scripting technique using Node js.	• (Pore and Pawar, 2015) • (Greenspan and Bulger, 2001) • (Yank, 2004) • (Harris et al., 2009) • (Tuya et al., 2006) • (Shute et al., 2013)
MS-SQL Server	1. Reliability requirements 2. Scalability requirements 3. Stability requirements 4. Integrated Development Environment For proficient advancement increment designer's profitability by utilizing Microsoft visual studio, SQL Server Management Studio, and Visual Developer tools 5. Disaster Recovery mechanism by providing database mirroring mechanism 6. Cloud back-up support		1.Can be integrated to Microsoft visual studio, SQL Server Management Studio and Visual Developer tools. 2. cloud storage	• (Pore and Pawar, 2015) • (Greenspan and Bulger, 2001) • (Yank, 2004) • (Harris et al., 2009)

Oracle 11g Express Edition Database	1. Fast downloading 2. Easily Upgrade 3.Vast platform support 4. Scalability 5. Easily manageable 6. Productivity 7. Reliability		1.Can be integrated to a new version 2. In adminstartion database	• (Pore and Pawar, 2015) • (Karunakaran et al., 2016) • (Kothuri et al., 2008)

Table A-2. NoSQL Databases types and advantages and limitations

NoSQL Databases Types	Advantages	Disadvantages	When can be used	Authors / years
Key-Value Store Databases	1. Rapied and cost-effective service of NoSQL database 2.Stores data on solid state drives 3. Increase the performance 4. NoSQL databases can also much more easily accommodate any new type of data	No fixed schema	1. KV databases are like hash tables where the keys are used as indexes. 2. KV database used in amazon and implemented using Amazon's Dynamo Model. 3. Basho technologies developed RIAK by using C, Java script, and Erlang 4. Riak clients, Riak search, Web machine, Riak KV, Riak core, Riak SNMP/JMX, Riak Pipe 5. Riak can be used for collecting and checking Point of Sales i.e. POS data, maintaining user's personal information on social networking sites etc.	• (Pore and Pawar, 2015) • (Li and Manoharan, 2013) • (George, 2013) • (Pretorius, 2013) • (DeCandia et al., 2007) • (Atikoglu et al., 2012) • (Han et al., 2011)

				- (Lee et al., 2013)
Column-Oriented Databases	1.Decreases the disk access compared to relational table, which consists of column and rows with uniform sized fields for every record. 2.High level of availability	1.Cassandra is that read operations are comparatively slower than write operations.	1.Column-Oriented Databases 2. HBase is developed after Google's Big Table as well as is open source, distributed and non-relational databases 3.Cassandra dynamic schema 4.Applications that use Cassandra are banking and finance, social networking websites and real-time data analytics 5. Cassandra used by Adobe, Twitter, eBay etc.	- (Pore and Pawar, 2015) - (George, 2013) - (Al Hinai, 2016) - (Pretorius, 2013)
Document Store Databases	1. MongoDB stores the documents in BSON format which is binary form of JSON. BSON 2. Supports different data types such as integer, float, string, Boolean, date	1. Schema less 2. Collections across multiple nodes 3. A good solution to store small and large files together	1. Format of documents are PDF, XML, JSON etc. Each document in document store databases is addressed by unique key for representation of that document. 2. Applications where data need to be store as documents that having some special characteristics. 3. MongoDB is proposed by 10gen company to handle growing data storage needs 4. CouchDB storage of data, it uses JSON documents as well as to create and update database document 5.CouchDB is used for CMS system, Customer Relationship Management, Facebook apps like Horoscope 6.	- (Pore and Pawar, 2015) - (Sharma and Dave, 2012) - (George, 2013) - (Chow, 2014) - (Pretorius, 2013) - (Han et al., 2011)

Graph Databases	1.Graph Databases, data is store in the form of a graph. 2. Records can be traversed 3. Graph databases are faster than relational databases 4. Easily scalable as queries are expressed as traversals 5. Provides flexible network structure	1. Graph databases don’t have pre-defined schema and focus on connection between the data.	1.Used in number of applications like content management, social networking sites, bioinformatics, cloud management etc. 2. Neo4j is object oriented, high performance graph database. Neo4j uses property graph data model which composed of nodes and relationship with their properties.	• (Pore and Pawar, 2015) • (George, 2013) • (Chow, 2014) • (Pretorius, 2013)

Table A-3. SQL and NoSQL databases properties and authors

Properties	SQL databases	NoSQL databases	Authors name and years
History	Developed in 1970s to deal with first wave of data storage applications	Developed in 2000s to deal with limitations of SQL databases, particularly concerning scale, replication and unstructured data storage and does not use SQL language	• (Pore and Pawar, 2015) • (George, 2013) • (Chow, 2014) • (Strauch and Kriha, 2011)
Types	Table based databases	key-value databases, document store databases, wide-columnar stores, and graph databases	• (Pore and Pawar, 2015) • (Sharma and Dave, 2012) • (Li and Manoharan, 2013)
Flexibility	1.Each record conforms to fixed schematic representation 2. Each row must contain data for each column	1. Dynamic. Information can be added on the fly 2.Each 'row' doesn't have to contain data for each 'column'	• (Pore and Pawar, 2015) • (Strauch and Kriha, 2011)
Scalability	1.RDBMS can be implemented across multiple servers 2.scalabl 3. bigger server 4.increasing the horse-power of the hardware.	1. horizontally scalable 2. A cheap multiple servers 3. cost-effective 4. distribute data across servers automatically	• (Pore and Pawar, 2015) • (Strauch and Kriha, 2011)
ACID Compliancy	Varies between technologies ACID compliant	1.Solutions sacrifice ACID compliancy for performance and scalability 2.Does not use concept of ACID properties	• (Pore and Pawar, 2015)

			• (Sharma and Dave, 2012)
Language Query (LQ)	1.Structured Language Query 2.powerful	1.Unstructured Query Language 2.Not powerful	• (Pore and Pawar, 2015) • (George, 2013) • (Chow, 2014)
Reliability	good fit	not good fit	• (Pore and Pawar, 2015)
Stored data	not much good fit for hierarchical data storage	1.Better for the hierarchical data storage as it follows the key-value pair way of storing data content which is like JSON data 2.preferred for large data set	• (Pore and Pawar, 2015) • (Li and Manoharan, 2013)
Transactional and application	useful for heavy transactional type applications, as it is more stable and provides the atomicity as well as integrity of the data	1.transactions purpose but it is still not comparable and usage effective	• (Pore and Pawar, 2015)
Supportability	1.Provide to the projects 2.large scale deployments	dependent support few outside experts are available to setup and deploy your large scale	• (Pore and Pawar, 2015) • (George, 2013) • (Chow, 2014)
Data Manipulation	Done by using Selection, Insertion, and Updation statements	done by using different object-oriented APIs.	• (Pore and Pawar, 2015)
Consistency	Inconsistent state, once the transaction gets completed in the database	all machines must be present the information same in all updating to be made on all machines frequently i.e. consistent information	• (Pore and Pawar, 2015) • (George, 2013) • (Chow, 2014) • (Strauch and Kriha, 2011)
Fundamentality	Relational Database	distributed databases	• (Pore and Pawar, 2015) • (Sumathi and Esakkirajan, 2007)

limitations	Google, Amazon, Facebook	1. no fixed schematic 2. records can have different fields	• (Pore and Pawar, 2015) • (George, 2013) • (Chow, 2014) • (Strauch and Kriha, 2011)
provide	Latest feature called as sharding	lack of RDBMS support	• (Pore and Pawar, 2015) • (Strauch and Kriha, 2011)
Cost	1.Reduce cost 2.Open source database	cache data in system memory	• (Pore and Pawar, 2015) • (George, 2013) • (Chow, 2014)
Schema	record conforms to fixed schematic representation, meaning the columns must be decided and locked before data entry	dynamic. Information can be added on the fly, and each ‘row’ doesn’t have to contain data for each ‘column’.	• (Pore and Pawar, 2015)
Relationships	Relational Databases	Non-relational or distributed database	• (Pore and Pawar, 2015) • (George, 2013) • (Chow, 2014) • (Strauch and Kriha, 2011)
Functionality	Relational algebra	Graph databases=graph theory Document stores = variable (low) Column stores= minimal Key-Value Stores = variable (none)	• (Strauch and Kriha, 2011)