An Analysis of Relational Database and NoSQL Database on an Ecommerce Platform

By

Xuejiao Liu

A dissertation submitted to the

University of Dublin, Trinity College

in partial fulfilment

of the requirements

for the degree of

Master of Science in Computer Science

August 2015

Dissertation supervisor: Siobhán Clarke

Declaration

I, the undersigned, declare that this work has not previously been submitted as an exercise for a degree at this, or any other University, and that unless otherwise stated, is my own work.

I also agree that Trinity College Library may lend or copy this dissertation upon request.

Signed: .....

Xuejiao Liu

August 28, 2015

Permission to lend and/or copy

I also agree that Trinity College Library may lend or copy this dissertation upon request.

Signed: .....

Xuejiao Liu

August 28, 2015

Acknowledgments

I would like to express my thanks to Professor Siobhán Clarke for her valuable guidance and support during the course and especially during this dissertation.

Additionally, I would like to thank Launchbox and Citi Group for creating such a great environment and letting me participate of such an amazing experience.

I am also grateful to Jinwei and Tian for their cooperation to concretize our idea, and also to Jiayi, because without them this dissertation would not have been possible.

Finally, I would like to give my most sincere thanks to my parents Mrs. Xiaohong Zhang and Mr. Wei Liu who encouraged me in life and sponsored me for my master degree.

XUEJIAO LIU

An Analysis of NoSQL Database and Relational Database on an Ecommerce Platform

Xuejiao Liu

MSc. Computer Science (Networks and Distributed Systems)

University of Dublin, Trinity College

2015

Supervisor: Dr. Siobhán Clarke

Lots of businesses recognize the Internet as an effective approach to creating valuable business opportunities. Currently, there is a trend towards an online marketing model, which refers to posting business information to Internet users with the goal of increasing offline consumption. However, as this model has gained popularity it has also increased challenges. One of them is that large amounts of data are now available on online commerce platforms, which leads to database maintenance and performance issues for businesses.

In order to handle large amounts of data and processing, a Database Management System for online commerce platforms has to address demanding performance requirements, especially scalability and processing efficiency. On the other hand, from a business perspective, the usage and maintenance of their database must be cost-effective. These conditions demand that online commerce platforms must be built on a scalable and practical database.

To improve a platform's ability to process data, this dissertation investigates different kinds of databases and how they help to mitigate issues related to large volumes of data. There are two major types of database: Relational Database and NoSQL Database. In this dissertation, the response time, error rate and throughput of different Database Management Systems are compared. Database process time complexity and space complexity are used to measure the efficiency. Furthermore, the dissertation comparatively analyzes different databases' economic costs.

The research is based on a real business platform - Soosokan, which is a project conducted within the Trinity College innovation incubator, called LaunchBox. Two Database Management Systems: MySQL database and Cloudant database, respectively based on relational SQL and NoSQL technology, were built on the Soosokan platform to evaluate processing big data performance and economic benefits. The results of two database experiments show that the scalability of MySQL and Cloudant is similar, but Cloudant is more efficient than MySQL in querying and inserting. In addition, Cloudant occupies more space but MySQL is more expensive. In conclusion, Soosokan employed the cost-effective Cloudant database with better performance.

Contents

List of Tables

List of Figures

Abbreviations

SQL	Structured Query Language
IDS	Integrated Data Store
DBMS	Database Management System
RDBMS	Relational Database Management System
OS	Operating System
TCP/IP	Transmission Control Protocol/Internet Protocol
HDFS	Hadoop Distributed Files System
API	Application Programming Interface
PDW	Parallel Data Warehouse
MPP	Massively Parallel Processing
DBaaS	Database-as-a-Service
SWOT	Strengths - Weaknesses - Opportunities - Threats
MVP	Minimum Viable Product
MVC	Model-View-Controller
HTTP	Hypertext Transfer Protocol

REST	Representational State Transfer
XML-RPC	eXtensible Markup Language-remote procedure call
SOAP	Simple Object Access Protocol
JSON	JavaScript Object Notation
MD5	Message-Digest Algorithm 5
HTML	HyperText Markup Language
XPath	XML Path Language
URL	Uniform Resource Locator
ER Model	Entity-Relationship Model

Chapter 1

Introduction

1.1 Background

The development of ecommerce platforms has experienced in recent years a rapid growth due to these online ecommerce services breaks the limitations of traditional business models, such as long geographical range and limited opening time. Ecommerce platforms can make merchants be able to target users, explore business opportunities and economize resource. But on the other hand, Ecommerce platforms also lead to intensified competition, as ecommerce platforms increase consumers' ability to gather information about products and prices. The basis of an ecommerce platform is storage and processing of product information. With the development of the ecommerce platforms, data has increased in the large scale in various fields, such as inventory data, transactions and operation logs.

In recent years, the number and scale of ecommerce platform have increased dramatically that leads to the big data issues for ecommerce platforms. And with the explosive increase of data, “big data” this term is mainly used to describe large amount data with high complexity and varying structure. And the “big” dataset’s size is beyond the ability of traditional database tools to capture, manage, store and analyze. In 2001, Doug Laney [1] firstly defined big data, he summarized that big data challenges brought by increased data with a 3Vs model, Volume, Velocity, and Variety. On ecommerce platform, a great number of products and transactions create large volumes of data. And the most significant impact is that large number of data allows businesses to make more informed decisions. Because based on data insights, business owners can take the guesswork out of decisions and make the wisest decision to prevent unnecessary

spending, or even mistakes. Big Data also allows ecommerce platforms provide more personalized offers and communications [2]. The big data brings a lot of potential benefits. However, there are many problems arose on big data. For any companies or platforms, if they want to profit from big data, they must to provide the strong technical support. All in all, the big data is a significant challenge that ecommerce platform must to face.

In Kaisler Stephen et al. research, they suggested that in big data age there should be three fundamental issues to address: storage issues, management issues and processing issues [3]. In essence, these issues are related to databases' basic features, and these challenges also be suitable to describe ecommerce platforms' problems. A database of platforms need to handle data storage, management and processing. In other world, to some degree, addressing three fundamental issues of big data is equivalent to improving databases' performance. Thus, for an ecommerce platform the primary consideration is how to build a strong database infrastructure to handle large amount data and processing operations.

1.2 Motivation

In order to seize the large market opportunities, we aim to implement an ecommerce platform - Soosokan[1], which is an eco-system which based real-time and geographic location. Soosokan provides mobile channels to buyers and sellers. Buyers can search the product detailed information and view the promotion information. Sellers upload products' information into Soosokan database and publish advertisements within the valid distance. Building a platform to share information between buyers and sellers is benefit on both sides. Sellers can explore new users without the limitation of geography and increase opportunities that own shop can expand influence and improve

[1] http://www.soosokan.com

competitiveness. For buyers, it provides an access to detailed product information and a platform to compare similar products attributes such as price, distance and promotion.

As an information retrieving platform, the search function is core business of Soosokan, and there is massive data will be stored and searched. In addition, with the development of amount of users and products, we have to consider the scalability of database and search efficiency in Soosokan system.

In order to address these issues arose from large amount of data, the scalability and data processing ability of databases need to be improved. In recent years, there are diverse types of database, such as traditional Relational databases, rapidly widely accepted NoSQL databases and new technology NewSQL database. Thus, today is an era of competition in a variety of databases, and each database has own advantages and disadvantages. The physical world is a relational world, Relational database is more suitable for physical objects to present their relationship, which is more intuitive and straightforward. The new generation of database such NoSQL and NewSQL weaken the relationship of objects and emphasize the advantages of Key-Value Stores and document database, rather than simply oppose Relational databases. Thus, developers need to compare and evaluate these database in their projects.

In this dissertation, I purpose to understand the advantages and disadvantages of different kinds of databases, and chose the best data infrastructure for e-commerce platform - Soosokan. Different database scalability and effectiveness performance will be tested and compared. In addition, as a business application, the economic costs also are the important part of databases. Therefore, the costs of databases will be compared as well.

1.3 Research Questions

The research question that this dissertation aims to answer is as follows:

What kind of database can be employed to reduce the issues that ecommerce platforms are currently facing when handling large number of data?

1.4 Research Goals

The research goal of the dissertation is:

> *Compare the NoSQL technology with Relational database, analyze the advantages and disadvantages of different approaches to certify the availability of NoSQL technology on ecommerce platform.*

A set of features of NoSQL technology and Structured Query Language (SQL) are analyzed and compared in order to be used as guidelines to develop information retrieving platform in the big data. These features include:

- Scalability

Analyze Relational database's and NoSQL database's ability to handle a growing amount of data and its ability to be enlarged to accommodate that growth.

- Time complexity

Searching is the core business in an information retrieving platform. In order to evaluate the time complexity of the system, the search efficiency will be compared in the different database infrastructure.

- Space complexity

Non-Relational database and Relational database employ different database frame, thus the same data item occupy different storage space. Analyze the utilization rate of storage media by Relational database and NoSQL database technology.

- Economic cost and benefits

From an economic point of view, compare the cost and benefits of Relational database and NoSQL database technologies. The cost depends on the space complexity in storage media with different technology. It also involves pricing policy from different database or storage vendors.

1.5 Structure of Dissertation

This dissertation is organized as follows:

Chapter 1 introduces the topic and sets the question and goals that will drive the research of this dissertation.

Chapter 2 explores the state of the art in different database to handle large number of data.

Chapter 3 presents the development of the idea that drives this dissertation.

Chapter 4 describes the implementation and deployment of Soosokan platform

Chapter 5 introduces the design and setup of experiments that test different database performance.

Chapter 6 details the methodology followed to conduct the studies with Relational database and NoSQL database and outlines the outcomes of those studies.

Chapter 7 concludes this dissertation and discusses the future work and improvements.

Chapter 2

State of the Art

2.1 Database Development

The history of Database can go back 50 years ago. At that time, database management was very simple. The data was processed by a lot of punched cards, which analyze and compare and build data tables. And the result of database management printed on the paper or created the new punched card. Thus, physical database management is that a system to manage, process and store the physical punched cards. And, in 1951, Remington Rand Inc. developed Univac I computer, which supports a tape drive that can enter hundreds records per second [4]. And this technology triggered a revolution in data management.

Database management system appeared in 1960s in the true sense. In 1961General Electric Co.[2] developed the first network database, which also is the first database management system - Integrated Data Store (IDS). It laid the foundation for the network database, and at the time IDS has been largely used by industry, known for its high performance.

Hierarchical Database Management System (DBMS) is followed by the emergence of network-based database. Also in the 1960s hierarchical database was firstly purposed, which is a data model which stores data into tree-like structure. The records or the database objects are found primarily by following references from other objects. The common type of this database is that data was stored on magnetic tape. Typical product

[2] http://www.ge.com

is IMS, which is IBM hierarchical database management system model developed by IBM[3] in 1969 [5].

Hierarchical database is a good solution to the problem of pooling and sharing of data, but there is still much lack of data independence and abstraction level. When handle the two kinds of database, users need the specific storage structure of data to point out the access path. However, the subsequent Relational database can solve these problems.

Today, there have been several generations of database. Moreover, the common problems of all databases are the big data. And in 2011, a report by the McKinsey Global Institute[4] analyzed challenges and values brought by big data, and listed the some technologies to handle big data issues. Currently, two kinds of database are used to process large amount of data: Relational database and NoSQL database.

2.2 Relational Database

The concept of Relational database was proposed by Edgar F. Codd in June of 1970 [6], and Relational databases are based on the relational data model. Relational databases use a set algebra and other mathematical concepts and methods to process data. And linkages in the real world between entities are represented by the relational model.

1974, IBM built research project - IBM System R, it was the first product with the implementation of SQL, which has become the standard relational data query language and virtually RDBMS (Relational Database Management System) uses SQL as the language for basic operations and maintaining the database system. However, until mid of 1980s, computing hardware was becoming more powerful to lead relational system was widely deployed.

[3] http://www.ibm.com

[4] http://www.mckinsey.com

Relational database is widely used for over 40 years. Nowadays, although it has appeared to be inadequate to handle massive data, it is still the mainstream database infrastructure. There are several famous Relational databases, such as Oracle[5], MySQL[6] and Microsoft[7] SQL Server. These three databases are the most popular databases in the current market.

2.2.1 Oracle

Based on Codd's relational data storage model, Larry Ellison developed Oracle database, which is a typical Relational database. And currently, Oracle is the most widely accepted database and it can run on all major operating system. In addition, it also fully supports for all industry standards. Oracle products using standard SQL, and it is compatible with DB2, IBM SQL / DS, IDMS / R and INGRES databases. Oracle products can run on a wide variety of hardware and operating systems (OS) platforms, and it also can be installed on more than seventy different size devices, such as mainframe computers, midrange computers and minicomputers. Oracle products can communicate with multiple communication networks and support multiple protocols, such as Transmission Control Protocol/Internet Protocol (TCP/IP), DECnet and LU6.2 etc. Moreover, Oracle provides several development tools to help developers. Oracle has structured, high availability and high performance features, which made Oracle to become the most popular Relational database. In the early years, almost all large enterprises used Oracle as database.

With the development of the information technology, Oracle products also face the issue from big data. For Relational database, the most commonly used method to address large number of data is that introducing Hadoop[8] into data infrastructure. The Hadoop is an

[5] http://www.oracle.com

[6] http://www.mysql.com

[7] http://www.microsoft.com

[8] https://hadoop.apache.org

open-source framework, supported by the Apache Foundation. It is written in Java for storage and processing of large data sets. Hadoop improves the database fault tolerant and parallelizes data processing across many nodes. And it leverages Hadoop distributed file system (HDFS) to replicate data sets.

In Oracle 12c product, the new big data architecture has introduced into Oracle data warehouse [7]. This architecture employs Cloudera[9] Hadoop and allows the majority of business users to access the data from the data warehouse, using SQL-based environments. And in any types of data warehouse environment, Oracle can provide performance optimizations.

From an economic point of view, the Hadoop is not a cost-effective approach to address the big data issue. Hadoop is free, but the servers which Hadoop reside do indeed cost money. In addition, Hadoop as Big Data platforms are designed to scale out, rather than scale up, businesses need more servers to meet data requirements, which will increase the economic costs.

2.2.2 MySQL

MySQL initial purpose is to build an open-source and easy to use database. And its history can go back to 1985, but, the first release version of MySQL was published in 1995. Until 1998, MySQL supported more than ten operating systems. However, there were many problems on MySQL at that time, for instance that MySQL could not support transactions, subqueries, foreign keys and view function. Figure 1 shows database market share by 2006.

[9] http://www.cloudera.com

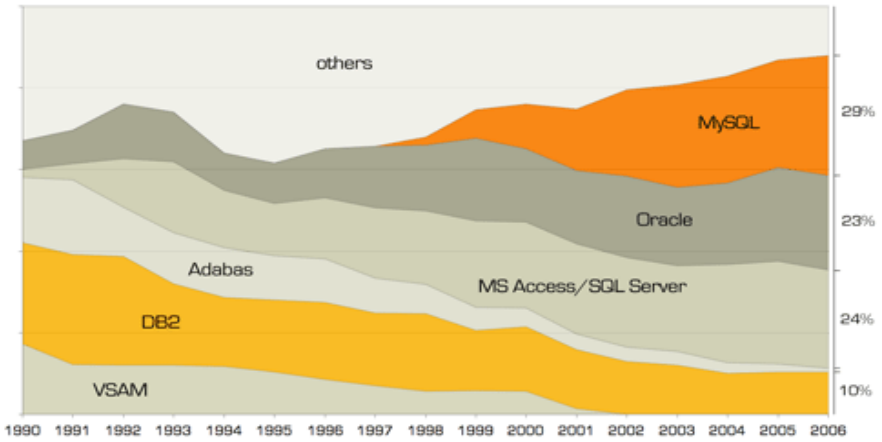

Figure 1 The development of MySQL database

Source from JoinVision E-Services GmbH, July 2006

As above figure shows, MySQL database's outbreak was actually arose around 2001 and 2002. Especially, in 2002, MySQL 4.0 Beta version was released, meanwhile MySQL selected InnoDB as the default search engine, which improved the ability of processing and caching data. In the same year, MySQL 4.1 can support subqueries. And at this point, MySQL finally transformed into a full-fledged Relational database system. MySQL 5.0 version added triggers, query optimization, and distributed transaction functions in 2005. In 2009, MySQL was acquired by Oracle.

Early market position of MySQL is focused on the development of Internet. Thus, the rapid growth of the Internet led to MySQL being popular in the world. In addition, MySQL makes a significant contribution to open-source software framework. Currently, the most widely accepted framework LAMP is based on the MySQL database. The LAMP is an acronym of Linux-Apache-MySQL-PHP, which provide open source software infrastructure with great performance, including operating system, server, database and front-end script language.

In order to improve the ability of processing large amount of data, MySQL 5.6 provides Hadoop interface [8]. Data can be organized and analyzed within Hadoop platform.

Similarly with Oracle, all Relational database including MySQL to address big data issues by Apache Hadoop framework [9]. In addition, MySQL 5.6 database also implements NoSQL technology via a Memcached daemon plug-in to process, with Memcached protocol mapped to the native InnoDB Application Programming Interface (API) [10].

2.2.3 Microsoft SQL Server

SQL Server initially was developed by Microsoft, Sybase and Ashton-Tate for IBM OS/2 operating system. However, OS/2 project failed later, and three companies disaggregated for their own business. Microsoft developed SQL server, as a part of Windows NT software solutions, for windows operating system. Sybase focused on Linux/Unix database development.

Microsoft SQL Server mainly aims to small and medium enterprises. Its greatest advantage is that integration of Microsoft other products and resource, and it can provide a powerful visual interface, highly integrated management development tools. Microsoft SQL Server is an important part of Microsoft integrated software program, also has made great contributions to the Windows operating system's popularity in enterprise applications.

Until now, most databases are Relational databases. But there is a trend that database development to the present need richer data models and more powerful data management capabilities features. In addition, new database model needs to support object-oriented, and can be employed on different platforms. In order to meet these demands, there are some distributed data or file frames be built, such as Hadoop and MapReduce [11]. Additionally, many projects are trying to bridge between these frames and Relational database management systems.

SQL Server in order to address big data issue, it also employs the Hadoop framework, it supports simplify access to unstructured data with HDInsight, an absolute Apache

Hadoop based distribution. In addition, SQL Server expanded the role of Parallel Data Warehouse (PDW), which has a massively parallel processing (MPP) architecture. As such, Microsoft has billed PDW as being well-tuned for big data processing. In SQL Server 2012, the PolyBase is introduced into Parallel Data Warehouse [12]. SQL Server seamlessly combines relational and non-relational data with PolyBase. The PolyBase is a part of PDW and it allows SQL directly query data stored in Hadoop, and even can do the "join" operations with local relational tables [13]. And the architecture is shown below.

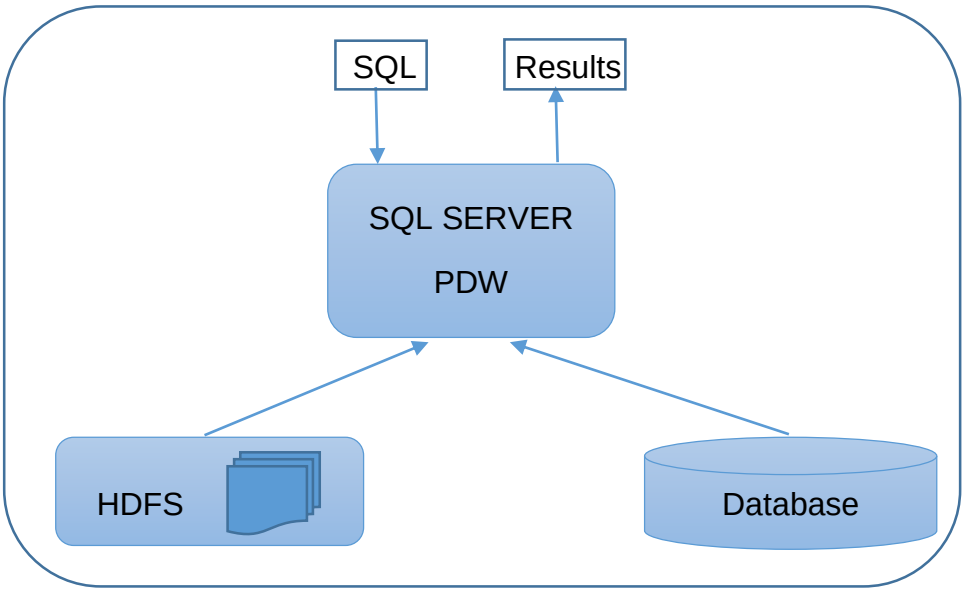

Figure 2 Microsoft SQL Server Database with Hadoop Architecture

2.3 NoSQL Database

Big data requires technologies to process large quantities of data in an efficient way within tolerable elapsed times. In order to address the big data challenges, unlike traditional database frame is introduced. NoSQL, commonly referred to as "Not Only SQL", denotes a new databases' framework that provides high performance and flexibility in the large-scale data set. In other words, it is a new database infrastructure that been very well-adapted to the heavy demands of big data issues.

With large volumes of data, there are some challenges brought by using Relational database management systems, which led to the development of non-Relational database management systems, commonly known as NoSQL databases. In Grillenberger et al. book, NoSQL is meanwhile used as a generic name for all non-Relational databases and is interpreted as an abbreviation of "not only SQL". In addition, Andreas et al. also described three features of NoSQL databases are [14]:

– a non-relational data model,

– distributed and horizontally scalable,

– schema-free or only with weak restrictions on schema.

With the development of databases' demanding requirements, NoSQL databases are widely used to address big data issues [15]. And for data structures, NoSQL databases use different data structure with Relational databases, making some better performance than Relational databases.

Generally, the non-relational data stores are categorized according to their data model:

- Column Stores: In NoSQL database, a column is an object of the lowest level in a key space. It is a key-value tuple consisting of name, value and timestamp [16].

- Document Stores: Document NoSQL database systems store documents, and all data is encapsulated or encoded with some standard formats or encodings in documents. In addition, these kind of databases can index the documents and provide a simple query mechanism. In document databases, it is allowed that nested values associated with each key.

- Key-value Stores: In this model, data is represented by key-value pairs. Similarly with key-value data structure, systems store values and its index into

databases. These indexes are based on programmer defined key. Key-value stores model is simple and easy to implement.

- Graph Stores: This model database is suitable for data whose relations are shown as a graph. And this graph relational data could be network topologies, transport links, social relations and road maps.

- Multi-model Stores: This model database can support multiple data models rather than a single, integrated backend [17]. Document, graph, relational, and key-value models, above mentioned models are examples of data models that may be supported by a multi-model database.

With the development of Web 2.0, a lot of NoSQL databases are emerged. Currently, there are about 150 NoSQL databases [18], and three NoSQL databases: Cassandra[10], MongoDB[11] and Cloudant[12] will be introduced as follows.

2.3.1 Apache Cassandra

Cassandra is an open-source distributed NoSQL database system, which developed by Facebook[13] in Java to storage simple format data. And it based on the key-value stores and integrated Google[14] BigTable data model and Amazon[15] Dynamo distributed framework. In 2008, Cassandra was independent from Facebook project. Since then, it became a popular distributed structured data storage solution.

In 2012, researchers from the University of Toronto conducted a benchmarking analysis of several different NoSQL platforms [19]. As their results shows, compared with other

[10] http://cassandra.apache.org

[11] https://www.mongodb.org

[12] http://www.cloudant.com

[13] http://www.facebook.com

[14] http://www.google.com

[15] http://www.amazon.com

databases, Apache Cassandra has the best performance of throughput on multiple nodes. In addition, In Endpoint[16] [20] white paper, they compared the Apache Cassandra, Couchbase[17], HBase[18], and MongoDB and presents that the scalability of Apache Cassandra is better than the MongoDB.

2.3.2 MangoDB

MangoDB, based on the document stores, is a database between the Relational database and non-Relational database [21]. It employs the JSON-liked data structure, called BSON, making the integration of data is easier and faster. MangoDB provides indexes on collection and supports master-slave replication to automatically recover and failover. In addition, there is no data lock [22].

MongoDB is the most popular NoSQL database [23] and ranked fourth in the all databases rankings, after three traditional Relational databases Oracle, MySQL and SQL Server. And some famous companies employ the MongoDB as their data infrastructure, such as LinkedIn[19], eBay[20] and The New York Times[21].

MongoDB provides high performance of storage and retrieval at large scale. Processing of data in MongoDB uses the MapReduce and aggregation framework, and outside, it employs the Hadoop and other external tools.

2.3.3 Cloudant

Cloudant is a non-relational, open source, and distributed database service provided by IBM. Similarly with MangoDB, Cloudant also is document-oriented database. Cloudant

[16] https://www.endpoint.com

[17] http://www.couchbase.com

[18] http://hbase.apache.org

[19] http://www.linkedin.com

[20] http://www.ebay.com

[21] http://www.nytimes.com

provides data distributed replication and Lucene querying mechanism. Cloudant with good read/write scalability, Cloudant Local scales for data size, large numbers of concurrent users, and multiple locations.

Cloudant based on the BigCoach project and supported by the Apache CouchDB[22] project. BigCouch, a fault-tolerant, horizontally scalable clustering framework, was designed to address a disadvantage of CouchDB - "it doesn't scale," which means that it does not scale horizontally across many servers. And CouchDB uses this feature to address big data issues [24]. And in January 2012, Cloudant official blog announced that they would contribute to BigCouch framework [25]. Thus, currently, the BigCoach is released and maintained by Cloudant to address big data issue.

2.4 Summary

Some researchers did similar experiments on different databases before, in [26], the author compared the performance of Oracle and MySQL. He tested databases inserting, querying, updating, deleting and pagination and showed the results in his articles. As the results show, based on 5 million data entries with different simultaneous requests (from 5 to 500), the scalability and efficiency of Oracle database are better than MySQL database.

According to three companies white paper [27, 28, and 29], among three different databases, Oracle database is the most expensive one, Microsoft SQL Server database is followed, and MySQL is the cheapest database.

The features and performance evaluation are shown as in the following table, and three "★★★" represents the most outstanding performance, and one "★" shows relatively poor performance of the database.

[22] http://couchdb.apache.org

Table 1 The comparison of Oracle, MySQL and SQL Server

	Oracle	MySQL	SQL Server
Operating System	AIX, HP-UX, Linux, OS X, Solaris, Windows ,z/OS	FreeBSD, Linux, OS X, Solaris, Windows	Windows
Current Release	12 Release 1 (12.1.0.2), July 2014	5.6.26, July 2015	SQL Server 2014, April 2014
Developer	Oracle	Oracle	Microsoft
Implementation Language	C and C++	C and C++	C++
License	Proprietary	Open Source	Proprietary
Scalability	Integrated with Hadoop	Integrated with Hadoop	Integrated with Hadoop
Efficiency	★★★	★★	No information
Price	★	★★★	★★

Similarly, some researchers evaluated performance of different NoSQL database. In Sergey Sverchkov's article [30], author compared three different database: Cassandra, HBase and MongoDB. According to his results of experiments, the scalability of Cassandra is better, but the MongoDB is more efficient.

Regarding economic costs, Cassandra and MongoDB both are free, but they are not the Database-as-a-Service (DBaaS), they need server to support them and the fee of the server should be regarded as costs of database. Cloudant is a chargeable service, but it does not need additional database server.

The features and performance evaluation are shown as in the following table, and three "★★★" represents the most outstanding performance, and one "★" shows relatively poor performance of the database.

Table 2 The comparison of Cassandra, MongoDB and Cloudant

	Cassandra	MongoDB	Cloudant
Operating System	BSD, Linux, OS X, Windows	Linux, OS X, Solaris, Windows	Cross-platform
Current Release	Cassandra 2.2.0, 20th July 2015	MongoDB 3.0.5, 28th July 2015	No information
Developer	Facebook	10gen	IBM
Implementation Language	Java	C++	Erlang
License	Proprietary	Open Source	Proprietary
Distribution	Multi-master replication	Master-Slave-Replica Replication	Multi-master replication
Scalability	★★	★★★	No information
Efficiency	★★★	★★	No information
Price	★★★	★★★	★★

Chapter3

Inception and Development of the Soosokan Idea

3.1 Initial Hypotheses

Soosokan initial idea firstly was proposed in the business innovation programme of the MSc Computer Science (Network and Distributed Systems) course at Trinity College Dublin. At beginning, our team discussed many times to target the market, and repeatedly refined Soosokan products to increase feasibility. And we got a lot of constructive comments from business innovation programme. We learned some basics of business and inspired an idea to do a new business.

As a part of this programme, our team got a lot of useful advises from Citi Group[23]. Mentors from Citi helped us to clarity users' requirements and refined our idea. In addition, they also provided some technology assistance to improve our product feasibility. More importantly, they helped us to master the skills of business presentation.

Finally, the conception of Soosokan was refined as:

> *Soosokan is aimed at developing an Eco-system which based on real-time and geographic location. Currently, we are focusing on developing two separate mobile apps for buyers and sellers.*

[23] http://www.citi.com

One of them could help users to find right products in right location right now. For example, it can help buyers to search the nearest shop which sells camera or specific milk and buyers can also look for near petrol stations by using the application. Detailed product information including price and distance would also be listed. Another app could help sellers to popularize their shops to target users with high possibilities. Sellers could use the app to manage their product's information as well as publish advertisements (also includes voucher or discount info) during a particular range (1km, 2km and 3km). These advertisements will be pushed to buyers who use Soosokan app and are currently in the target area.

As the achievement of this programme, subsequently, Soosokan business idea was presented to investors and IT industry personalities at the Citi Upstart Challenge and got the Microsoft BizSpark Start-up project support.

3.2 LaunchBox – Student Start-up Accelerator

As the hypotheses clearly stated, this initial idea needs to be tested and verified by market audiences. Because, there is a gap between academic and commercial, from a research point, the simulation of user cases or application environment is adequate for research product development. However, these simulation cannot guide the development strategy in real business market. Therefore, for the developers of new product, if they want to promote their products to the public, it was necessary not only to perform academic research, but also to interact with industry professionals and potential users to discover first-hand the state of the industry.

In order to work in a real business start-up environment and draw on the experience from successful entrepreneurs, I joined LaunchBox[24], which is the student start-up accelerator

[24] http://www.launchbox.ie

in Trinity College Dublin. In 2015, it founded 8 teams and helps them to concretize idea or hypotheses and develop their startups. LaunchBox organized a wide variety of workshops or seminars, and invited angel investors and experts in various fields of business to help teams to evaluate critically their proposals and set their objectives in the right direction.

In LaunchBox entrepreneurs and mentors mentioned some practical methods to solve startups' detailed problems, such as analyzing requirements of users, defining products' market positioning, gathering users' feedback, promoting market channels and so on. For whole start-up market strategy, they praised the lean start-up concept, which is a popular entrepreneurship methodology in Silicon Valley, and its core concept is that firstly put a minimalist prototype into market, then through continuous learning and valuable user feedback to iteratively optimize product, in order to business can adapt to the market [31]. Summed up in a word to express "Lean Startup" thinking, it should be "your vision is not important, the most important is that to find market demand rapidly with the lowest cost." And there are five principles from Lean Start-up concept:

- **Entrepreneurs are Everywhere**

  The new start-up is individual institutions that in the case of uncertainty, to develop new products and services for the purpose. This means that entrepreneurs everywhere, and Lean Startup method can be applied to all walks of life, in companies of any size and even large enterprises.

- **Entrepreneurship is Management**

  New start-ups not only represent a product, but also a body system. So it needs some new management, in particular to be able to respond to a situation of extreme instability.

- **Validated Learning**

The presence of new enterprises is not only to manufacture products and earn money, service customers. Their existence is to learn about how to build a sustainable business. Entrepreneurs can detect all aspects of their vision through a lot of experiments, and this knowledge can be proven.

- **Build-Measure-Learn**

  New Start-up basic activities are to transfer the idea to a product, a measure of customer feedback, and then realize that it should change its ways or stick unwavering. All successful startups process steps should be to accelerate the feedback loop for the purpose.

- **Innovation Accounting**

  In order to improve business results, and allow innovators to assume corresponding responsibilities, we need to focus on those tedious minutiae: How to measure progress, and how to prioritize allocation of work, and how to determine the milestones. These demands lead start-ups to design a new accounting system, which lets everyone shoulder the responsibilities.

3.2.1 Market Self-Positioning: Business Canvas

Clearing requirements of users and reasonable market position are the most important part of all product process. In order to understand users' need and clear market demand, some mentors introduced the Business Model Canvas to communicate with users and from the user point of view to fetch market's demand.

Business Model Canvas is straightforward and it is mainly used to help entrepreneurs to build, visualize or test the feasibility of their business model in order to avoid squandering funds or blindly overlaying feature [32]. Small companies use it to open up new areas, large companies can use it to explore new models to maintain industry competitiveness. The importance of Business Model Canvas for entrepreneurs is

summarized as follows: to birth of creativity, reduce speculation and ensure them find the right target audience, get a reasonable solution to the problem [33].

Business Model Canvas not only to provide flexible plans, but easier to meet the users' requirements. In addition, it also standardizes necessary elements of the business model, and emphasizes the interaction between different elements. "Business Model" is a clich é and ambiguous word. Some people understand it as profitable, and other people consider it as products, processes, technologies, or sales channels. Business Model Canvas clarity business concept, and making the entire development process more vivid.

Canvas consists of nine squares, each square represents a lot of possibilities and alternatives, business owners have to find the best one [34].And these nine parts of Business Canvas are: Key Activities, Key Activities, and Partner Network, Value proposition, Customer Segments, Channels, Customer Relationships, Cost Structure, Revenue Streams, which respectively represent the infrastructure, product, customers and finances of a business.

3.2.2 Competitiveness Analysis – SWOT Analysis

In order to understand our business's competitiveness in the current market, LaunchBox's guests introduced an analysis method – SWOT method, which is a competitiveness analysis including strengths, weaknesses, opportunities and threats [35]. Therefore, SWOT analysis actually is a method that generalize enterprises internal and external conditions, further analysis organization strengths and weaknesses, opportunities and threats faced.

Overall, SWOT can be divided into two parts: SW (Strengths & Weaknesses) and OT (Opportunities & threats), respectively used to analyze the internal conditions and external conditions. In SW analysis, competitive advantage means that a business or its product has any superior thing, it can be the any feature of products. Similarly, the competitive disadvantage of business may be reflected in these minutiae. Enterprise as a

whole with extensive source of competitiveness, so do the advantages and disadvantages analysis must be from the entire value chain. In OT analysis, environmental trends are divided into two categories: threats and opportunity. The environmental threats refer to a challenge formed by an environment unfavorable trends, if business does not take wise strategic behavior, this negative trend will weaken the company's competitiveness. The environmental opportunities for companies is that external factors, including new markets, new demand and competitors blunders etc. Refers to the challenge of environmental threats in an environment unfavorable trends formed. Environmental opportunities for company behavior are attractive areas, and the company has competitive advantages in this area.

The details are shown in the following figure.

Figure 3 SWOT Analysis
Source from Wikipedia.com

Through SWOT analysis, we can help business to gather resources and enable them focus on our products' strengths and let business strategy became clear.

In addition, SWOT model by McKinsey proposed a long time with the limitations of age. Previous businesses may be more concerned about the concrete elements such as cost,

quality. However, as the development of business becoming more mature, now the businesses may be more emphasis on organizational processes. Currently, in the SWOT analysis of the process, businesses may encounter some problems. And the SWOT analysis method must be adaptive because there are too many occasions to it. However, this can also lead to abnormal phenomenon. Problems basis SWOT analysis can be generated by a higher power SWOT analysis is resolved.

3.2.3 User Feedback Loop

In lean startup concept, it is crucial to validate product's value and growth hypotheses as soon as possible, therefore, the first step a start-up is to put a minimalist prototype into market [36]. And get users feedback to optimize their product. In order to do that, business can use a minimum viable product (MVP) to confirm (or refute) its value and growth hypotheses. Then, the next step for business is to gather users' feedback and refine products.

Ries introduced the concept of validated learning to start-up business, and in this concept, users' feedback is an important part of new product development [37]. In his book, he also suggested a build-measure-learn loop, as the following figure shows.

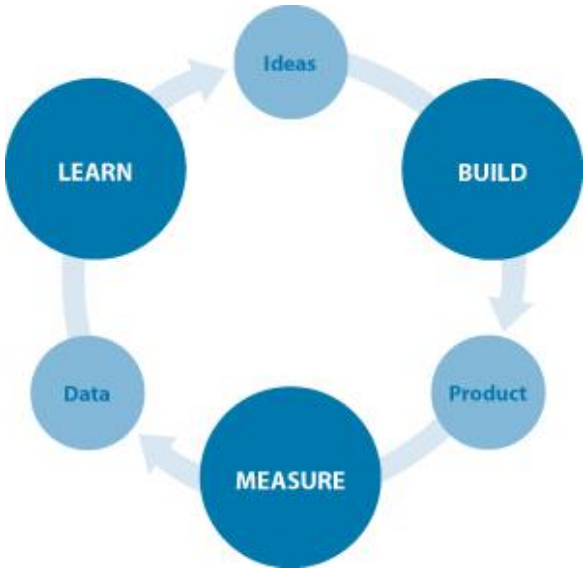

Figure 4 Build-Measure-Learn Loop
Source from 400minutes.com

All start-up companies originated from a new idea, and through building a MVP to put on market, the idea becomes a product. Then businesses measure their products by feedback of market and gather data to optimize product or market strategy. And all companies focus on minimizing the total time through the loop.

3.2.4 Failure Analysis

Everyday there are a lot of new start-up companies emerged, and also a lot of start-up companies failed. According the summary of startups were funded during 2014, "Nine out of ten startups fail" [38]. And in Jurica Dujmovic's blog, he mentioned the top 20 most common reasons by CB Insights[25], as the following figure show [39].

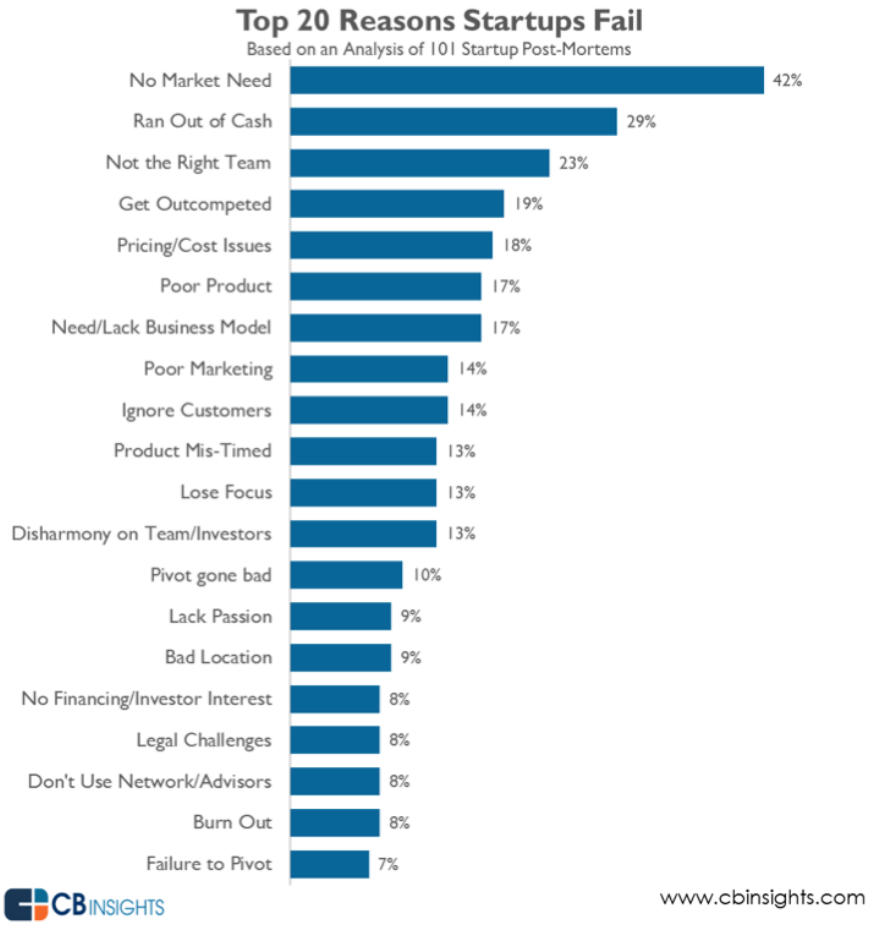

Figure 5 Top Reasons for Startups Failure in 2014
Source from cbinsights.com

[25] http://www.cbinsights.com

The reasons for failure are varied indeed. The most important one is a lack of users' requirements analysis. In the early stages of development, businesses did not analyze market's needs or analyze the requirements with an incorrect method. Therefore, 42% failure reason is that products cannot meet market demand. In addition, founders also cited a lack of adequate funding (29%), the wrong team and inappropriate collaboration for the project (23%), and fierce competition (19%) as top reasons for startups failure. These reasons for the failure should be a warning for all start-up companies.

Chapter 4

Implementation and Deployment of Soosokan

In order to complement the goal of Soosokan, this chapter will purpose the technical architecture that could be used to design, develop and deploy a platform based on this proposal. In addition, this chapter begins with a description of the architecture design and then proceeds to explore the components and technologies suggested for the implementation and deployment.

4.1 Technology Architecture

Soosokan is a location-based and real-time Android application, the front-end services are two Android applications for sellers and buyers, which support at least Android 4.0 version. Server-end services employ RESTful frameworks and Servlet to handle requests from Android applications and invoke Database API. Data infrastructure is Cloudant, which is a DBaaS to provide interfaces to developers. Soosokan uses the Cloudant to manage all text information, and File System to manage picture files. In addition, Soosokan combines with Facebook API and uses web scraper to get information from other retailer official websites. The whole project is established on the Model-View-Controller (MVC) model, and the architecture is shown as in following figure.

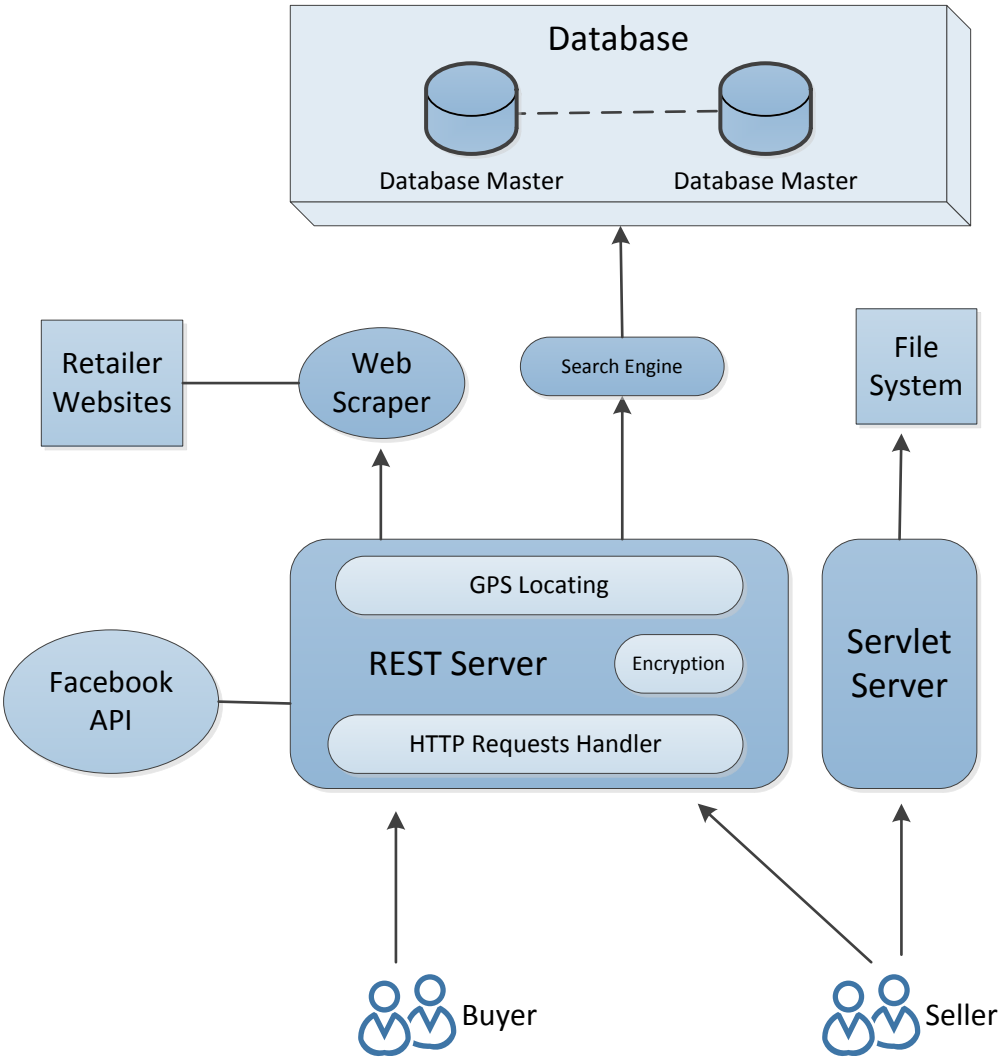

Figure 6 The Overview Architecture of Soosokan

One server employs encapsulated Hypertext Transfer Protocol (HTTP) requests in Jersey interfaces. In addition, there is another server, which is based on Servlet to process seller pictures and store these pictures into the file system. The file system is a storage space in server hard disk.

4.1.1 Mobile Channel

For Soosokan, there are two kinds of users: buyers and sellers. Common users are those who want to buy products and use Soosokan search or save information. Another users are sellers who can upload products' information and publish advertisements. Thus, we need to support two different users' requirement with two different Android applications.

4.1.1.1 Facebook Third-party API

In order to complement the goal of customization in Soosokan, in certain circumstances, the common users also be required to log in. For example, if users want to subscribe one shop they need to log in, then the subscription will be stored in database.

Soosokan is an easy-to-use application, registration will discourage users' enthusiasm. Therefore, Soosokan employs Facebook authentication API to login with a Facebook account. Facebook API allows applications to collaborate with the Graph API on behalf of Facebook users. As a result, if you want to collect data of individuals legally, you will need users' authentication. To obtain permission from Facebook users, we need to build an application on Facebook. When people join the application, they need to authorize Soosokan permission.

4.1.2 Servers

In Soosokan, in order to meet different data structure requirements, two servers are used to handle different types of requests.

The overview of servers as the figure 7 shows.

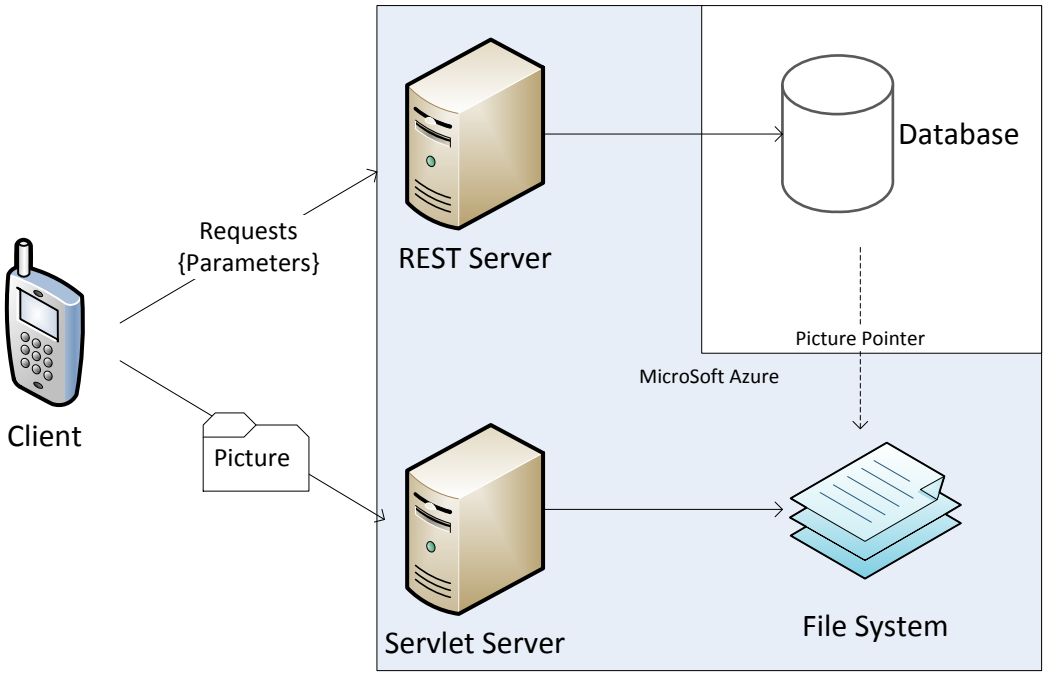

Figure 7 The Diagram of Soosokan Servers

One server handles the text message based on the Representational State Transfer (REST) framework, which is a software architecture style for building scalable web services. For web service, it has a simpler style that makes it easier to use than eXtensible Markup Language-remote procedure call (XML-RPC) and Simple Object Access Protocol (SOAP) mode. REST frame can reduce development complexity and improve system scalability.

The following figure illustrates using REST for Web Services.

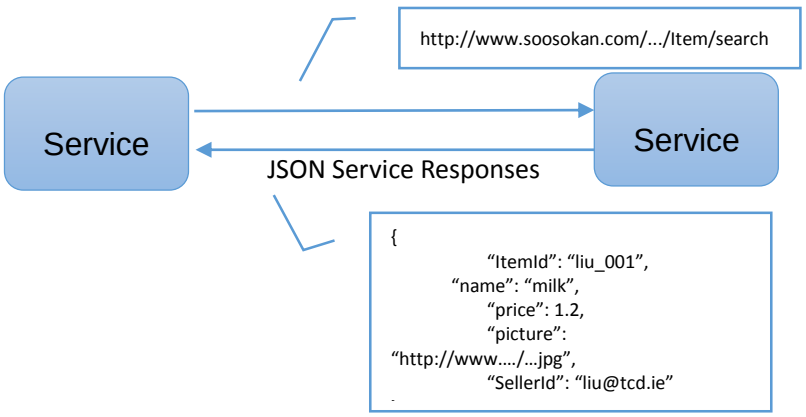

Figure 8 The Diagram of REST Framework

Typically, REST frame communicates with servers by HTTP with the same HTTP verbs (GET, POST, PUT, DELETE, etc.), which web browsers use to send data to remote servers and to retrieve web pages and. In REST Server with Jersey[26] interfaces, the presentation layer as a HTTP requests handler to execute user requests and return results to front-end. The business logic layer mainly is used to process objects logic functions and the data access layer acts as the database management system to connect database and release database operations. The last tier is database tier, which is data persistence layer to store data and support data processing. The architecture of servers as in the following figure shows.

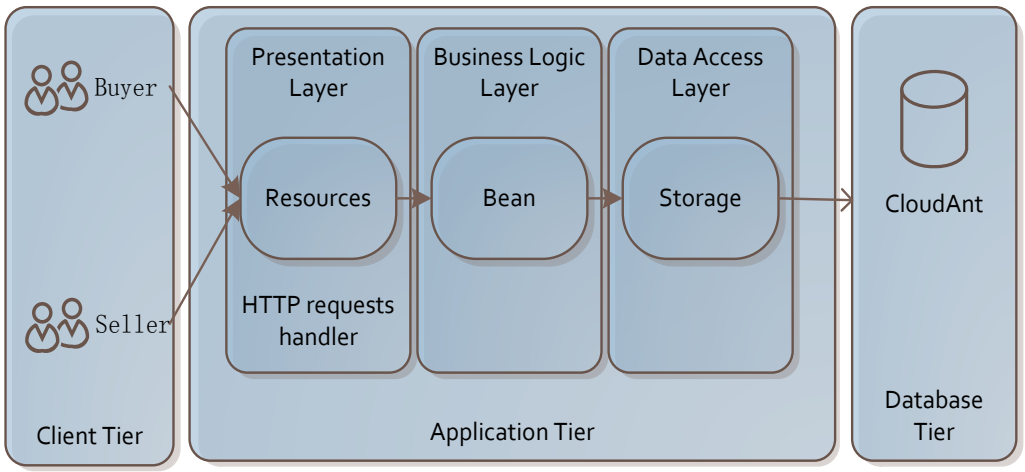

Figure 9 The Diagram of REST Server

One of the most important characteristics of the REST framework is that all messages are transferred in JavaScript Object Notation (JSON) format, which is a light-weight character string data format. This format is human-readable and easy to transmit data between server and web application. In addition, the REST frame provides JSON API to transfer objects to JSON format, which can automatically converse object class into JSON format to reduce the workload of server and client.

However, when image or multimedia files are transmitted in JSON format, all data types need to be transferred into strings that will increase the load and lack files' fidelity.

[26] https://jersey.java.net

Soosokan need to solve the problem to upload and download pictures, thus we used the double server to support Soosokan business logic. One server, based on REST framework, transmits text messages, Servlet server transmits image files.

The Servlet also called Java servlet, which is a Java program that can enhance the capabilities of a server. And servlets can response any type of requests, in Soosokan, Servlet only to handle uploading pictures and store these pictures into file system. And Servlet processes pictures by binary byte stream, which can keep pictures without distortion and easy to compress.

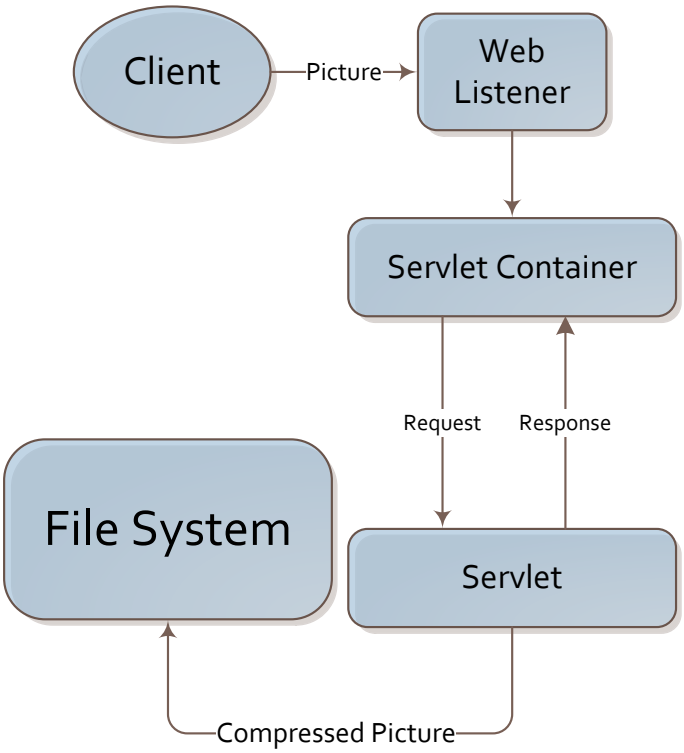

Figure 10 The Diagram of Servlet Server

Soosokan supports two servers, client mobile application upload the pictures to Servlet server. And Servlet server stores these pictures into file system. At the same time, client mobile applications post other requests and parameters to REST server, including the pointers of pictures. Then the REST server will handle the requests or manage database.

And these two servers and file system all are deployed on the cloud platform – Microsoft Azure[27], which also provides the load balancer.

4.1.2.1 Encryption

In order to protect users' sensitive information, Soosokan encrypted users' password by Message-Digest Algorithm 5 (MD5), which is a widely accepted cryptographic hash function. In addition, Soosokan stores ciphertext in database to avoid the risk of privacy leaks.

4.1.2.2 Search Engine

Essentially, Soosokan is an information retrieving platform. It supports searching products' and advertisements' information. In order to improve the search efficiency, Soosokan need to employ a suitable search engine. Because Soosokan database is document-based NoSQL database, so we choose the Lucene[28] as search engine. The Lucene is a full-text open source search library, supported by Apache Software Fundation, and originally written in Java. And the Lucene search engine can be used for any application that requires full text indexing and searching capability. Additionally, Lucene has been widely accepted by many big projects such as Linkein and Elasticsearch.

4.1.2.3 Web Scraper

Integrating other brand retailer information can improve Soosokan competitiveness, thus when user search products information we scraped some official websites to expand to our service area.

[27] https://azure.microsoft.com

[28] https://lucene.apache.org

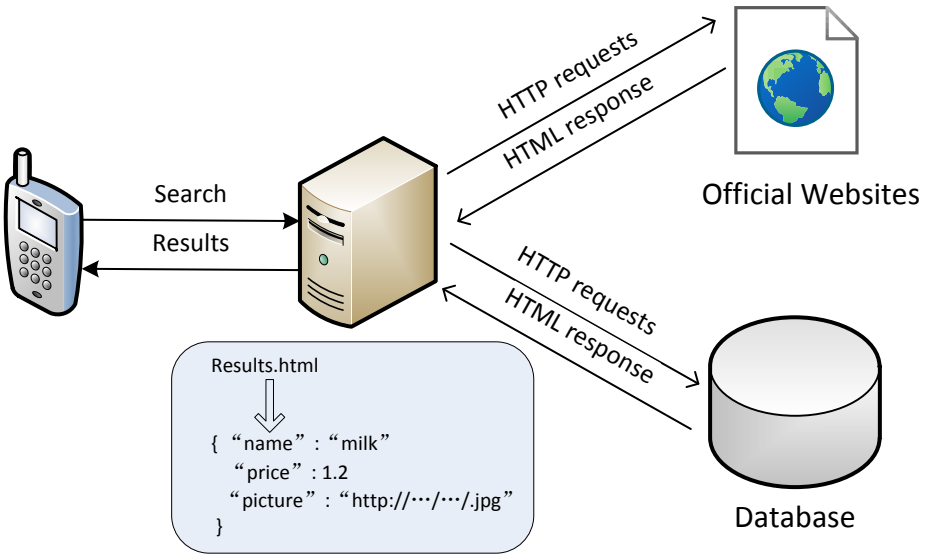

Figure 11 The Diagram of Web Scraper

As the figure 11 shows, on the server, Java program simulates HTTP requests by URLConnection class to official websites. Then server got the response from official websites within search results. Next, these HyperText Markup Language (HTML) file of search results will be analyzed by XML Path Language (XPath) to get the target information, such as product title, price and Uniform Resource Locator (URL) of product picture.

4.1.2.4 System Management Thread

In order to mitigate server and database workload, Soosokan need to manage invalid and expired information such as the overdue advertisements and update the limit of sellers' advertisements. To achieve the goal, we create a system management thread to manage the data daily.

4.1.3 File System

It is unwise to store large size files in databases, which will increase databases' workload and space, and the transmission of large size files will waste network bandwidth. Thus, a better solution is file system. Large size files can be stored in the file system, and the

pointers to the files will be kept in database. When clients retrieve files, server can find these files according to the pointers in database.

In Soosokan, there are pictures from sellers need to be stored in File System. Therefore, sellers use application to upload pictures to server, and this server saves pictures into File System. This server only processes pictures and communicates with File System.

4.1.4 Database

4.1.4.1 Database Management System

Soosokan employs the Cloudant as data infrastructure, and the Cloudant is a NoSQL Database-as-a-Service (DBaaS) and provides API for developers. Therefore, in Soosokan, we need to invoke cloud service on Cloudant platform rather than to build our local database.

Compared with Relational database, in Cloudant, there is need not to demonstrate the data structure of each attributes, because that data stored in JSON document and all attributes are stored in the string format. For developers, they just need to build the empty database. However, there also are some drawbacks of JSON data structure. When a null value or substandard data is stored in database, such decimal number to represent data, there will be not alarms for these situations.

In Cloudant all data is presented in documents and it supports the index to query.
The database index is a data structure in DBMS to sort. And it can improve the speed of data retrieval operations at the cost of additional storage space and writes to maintain the index data structure. Data entries can be located quickly by indexes without having to search every document in a database. Indexes can be built in one or more documents, providing the basis for both efficient access of ordered records and rapid random lookups. Therefore, every type of database need to create one or more indexes. For example, in advertisements database, the advertisement need to be searched by time

or ads id. We created two indexes with “ads_time” and “ads_id” for querying advertisements.

4.1.4.2 Database Replication

In order to improve Soosokan’s fault tolerance and robustness, we designed the replication of our database. And take account into our data stored in the cloud database, we employed the Cloudant database API to replicate data. The backup will be distributed in the different Cloudant data center. Moreover, these data centers built all over the world. Soosokan’s primary data stored in UK data center, and replication stored in USA.

4.2 Deployment of Soosokan

4.2.1 Google Play Store

We choose the Google Play Store[29] as the platform to release Soosokan application. Google Play Store is the biggest Android application market in the world and provides a graphic user interface to collect application information and users’ feedback. In addition, when the application updates to new version, Google Play Store will automatically update new application on users’ devices.

4.2.2 Alpha Test

The first phase of testing is Alpha test phase, application is tested in the development environment by one or more users. Alpha testing aims to refine software products by finding and fixing the bugs that were not discovered through previous tests.

At the beginning of August, we used the Google Play Store platform to release and invite Soosokan (for Buyers) Alpha test version. And there were 11 users to be testers

[29] https://play.google.com

and report some problems. And now, we released the Soosokan (For Sellers) as Alpha test version.

According to users' feedback and bug reports, we debugged some problems and released six following versions.

4.2.3 Beta Test

In software testing, Beta test is the second phase, also is a preliminary software field test carried out in the actual usage environment by beta testers who are some audiences outside of the programming team. Beta testing follows alpha testing, it also can be considered a "pre-release" testing of external user acceptance testing. The Beta versions of products can be provided to public to increase the feedback field to a maximal number of future users and to deliver value earlier.

When the Alpha testing finished, we released the Beta test version on Google Play platform. And we invited 8 users to participant in beta tests and got their feedback, then fixed and improved Soosokan application.

4.2.4 Official Version

Soosokan (For buyers) fixed all found bugs, it was released as an official version. Everyone can search and download Soosokan (For buyers) in Google Play Store.

Chapter 5

Experiments Design and Setup

5.1 Experiments Design

In current market, there are many kinds of traditional Relational databases and NoSQL databases. And in this dissertation, the performance and economic costs of different databases will be compared. From two types of database, MySQL and Cloudant are chosen to evaluate the Database features. These two databases are very representative in two types of databases. MySQL is an open-source database software and the Cloudant is a popular Database-as-a-Service, these two databases both are excellent choice for startups. Specifically the experiments procedures are:

5.1.1 Databases Design

In this dissertation, the experiments are based on MySQL database and Cloudant database. Firstly, these two databases infrastructure and operations functions should be established. And the Cloudant database details have been described in chapter 4. Thus in this chapter, only the establishment of MySQL database will be introduced.

All Relational databases employ relational table to present a logical relationship between the physical entities. And each object corresponds with a table, which includes all attributes of this object. According to Soosokan business logic, there are five entities tables: Seller, Buyer, Item, Ads and Payment. And in Relational database, the relationship of different objects is presented in the relationship tables, for example two relationship tables: Subscription and Favorite. The Entity-Relationship (ER) model of Soosokan as the figure 12 shows.

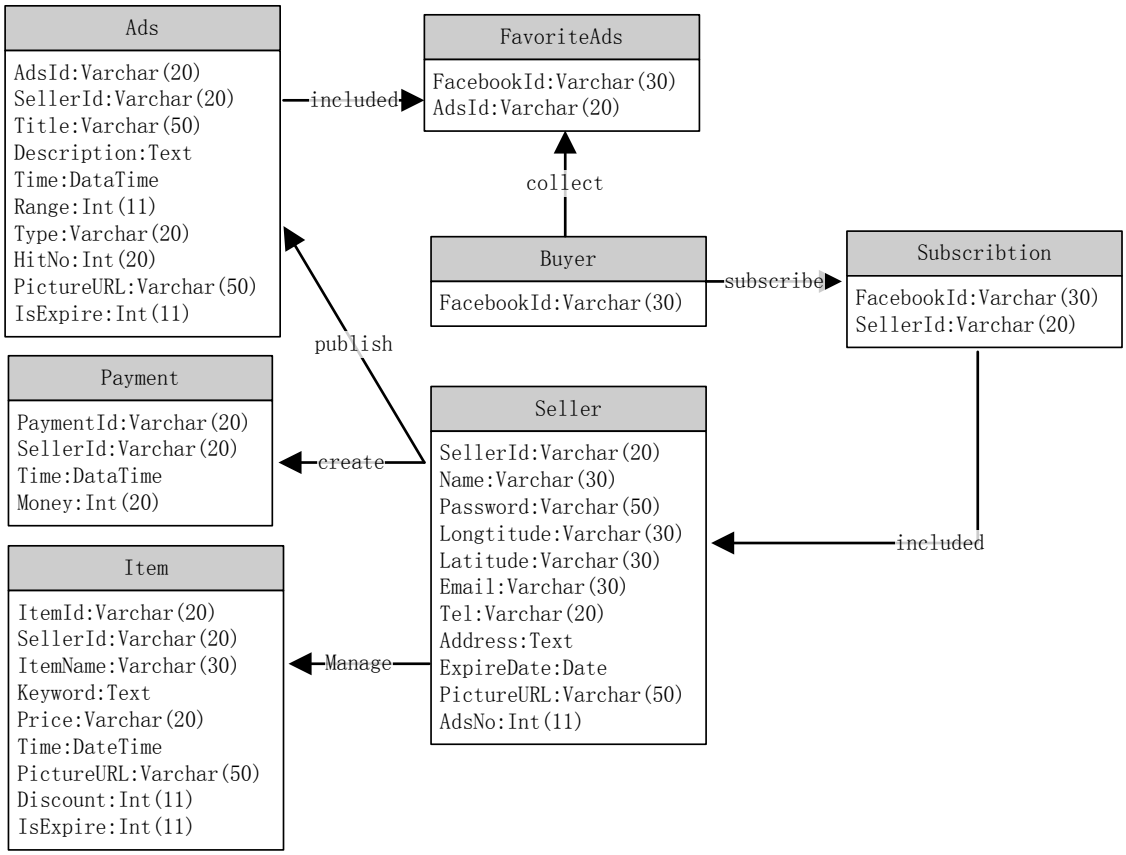

Figure 12 The ER (Entities – Relationship) Model of Soosokan

According to the ER model, the entity and relationship tables were created in MySQL database. Then, I need to build a database management system in Soosokan project, because Soosokan is an Android platform, I employed the JDBC to connect the database and Android front-end through RESTFul framework server.

5.1.2 Scalability Experiment Design

Scalability Test is a kind of performance test, which aims to understand how an application scales as it is deployed on larger systems or as more load is applied to it. The load will be tested by simulating the users or requests. The test will start at a scale factor – say 500 users and increment by the same factor every time (e.g. 500 users at a time). The goal is to get to the performance and evaluate how the system behaves at every step.

It is difficult to test system scalability in the real environment, so scalability test or load test needs some tools to simulate. There are some existed softwares to test the service

scalability, such as Apache JMeter[30] and LoadUI[31]. These tools can simulate the Database accesses and HTTP requests.

5.1.3 Time Complexity Experiment Design

Actually, database operation time depends on infrastructure. I can test the same "Query" requests and "Insert" requests run time in two databases with different amount of data. For example searching all products which keyword contains "milk" from the 500 and 10000 data entries.

5.1.4 Space Complexity Experiment Design

Relational databases store data in relational tables and columns. Information in different table can be accessed by the foreign key, which indicates the relationship of data. However, document-oriented database stores data in standalone document. Each document records one object, and does not describe the relationship between different objects.

According to different database infrastructure, the same information stored in Relational database and document database will occupy different space. I can test the same data such as 1000 seller information storage space in different databases.

5.1.5 Economic Cost Experiment Design

The economic cost depends on different database vendors' price and data amount. So I will survey different database.

In addition, the cost also depends on the storage media, including price of database hard disk, and server machine rent. Soosokan database is deployed on the cloud service, so the cloud service vendors' price will be assessed.

[30] http://jmeter.apache.org

[31] http://www.loadui.org

5.2 Experiments Setup

In this dissertation, the experiments are based on e-commerce platform - Soosokan, which is a mobile application and supports to insert and search product information. In addition, for these experiments I configured and deployed Relational database and NoSQL database - MySQL and Cloudant.

5.2.1 Soosokan E-commerce Platform Setup

The analysis of different database based on Soosokan platform, which is based on RESTFul framework and employs JSON data transmission format. In this dissertation, the experiments employed two databases. In different tests, the Soosokan server called different database operations, as the figure 13 shows.

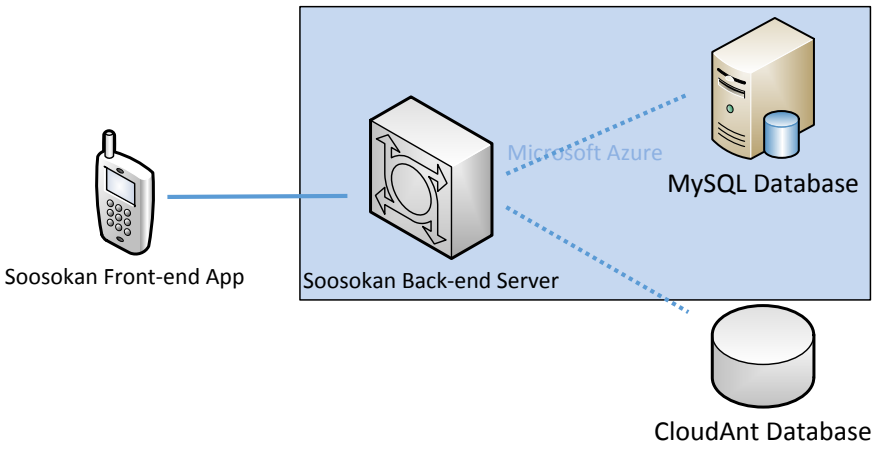

Figure 13 Overview of Experiment Setup

In addition, the Soosokan back-end server deployed on the Microsoft Azure cloud server. The Cloudant database is a DBaaS and provides integrated API, it does need to deploy on the server. However, the MySQL database needs to be deployed on the Microsoft database server.

5.2.2 MySQL Database Setup

In Relational database, data tables need to be designed to denote realistic relationship of objects. Step one is to design and create relational data tables. For each object, database management system should support basic “insert”, “delete”, “update” and “query” operations. Thus, step two is to implement these database operations. Additionally, in order to support Soosokan public network service, database need to be deployed on the database server with a public IP address. Soosokan choose the cloud service to provide database service. Thus, step three is to deploy MySQL database on Microsoft Azure cloud service. Finally, the database infrastructure was built and step four is to input data.

5.2.3 Cloudant Database Setup

Cloudant is a NoSQL full-text database, all data is stored in JSON format. Thus, the design of dataset is not required. But on the other hand, Cloudant supports index query. The first step to build Cloudant database is to create indexes. Similarly with Relational database, in Cloudant step two also is to implement basic database operations. By contrast, Cloudant is a Database-as-a-Service. There is no specific server to support database service. Developers employ Cloudant API without the deployment and installation. The last step for Cloudant database setup is to input data.

5.2.4 Test Tools and Script

The database performance test experiments are conducted on the MySQL database and Cloudant database. For scalability test, the Apache JMeter was used to simulate synchronous HTTP requests. All requests accessed the server and test tool recorded the run time, error rate, through put for each request.

Time and space complexity test experiments directly run on databases. Thus, there are database scripts or test program to record run time and storage space.

Chapter 6

Evaluation

There are several experiments explained in the last chapter have been undertaken, and this section presents all the experiments results, and detailed experiment result data is shown in Appendix.

6.1 Simulation Tool – Apache JMeter

In order to carry out these experiments mentioned in the last chapter, I used the Apache JMeter to simulate large scale of users' concurrent requests. Apache JMeter is an open source Java desktop software to test load functional behavior and measure performance [40]. Originally, it just could test Web Applications, but now, it has since expanded to test other services' performance. It can be used to simulate a heavy concurrent load and to analyze the overall performance under various load types, it also allows for comprehensive analysis and graphical analysis of the performance metrics (e.g. response time, error rate and throughput) measured [41].

In experiments, Apache JMeter configures test plans with different parameters, such as number of concurrent, request time, target URL and request parameters etc. In dissertation, the number of test concurrent requests are various from 500 to 7000, there are 14 Apache JMeter test scripts to test scalability of database with different load. And these scalability tests are undertaken on MySQL and Cloudant database with 10,000 entries.

6.2 Scalability Test

6.2.1 Response Time

According Apache JMeter tests, the scalability of MySQL database and Cloudant database is shown in the following figures. Figure 14 presents the average time of HTTP response with different concurrent requests.

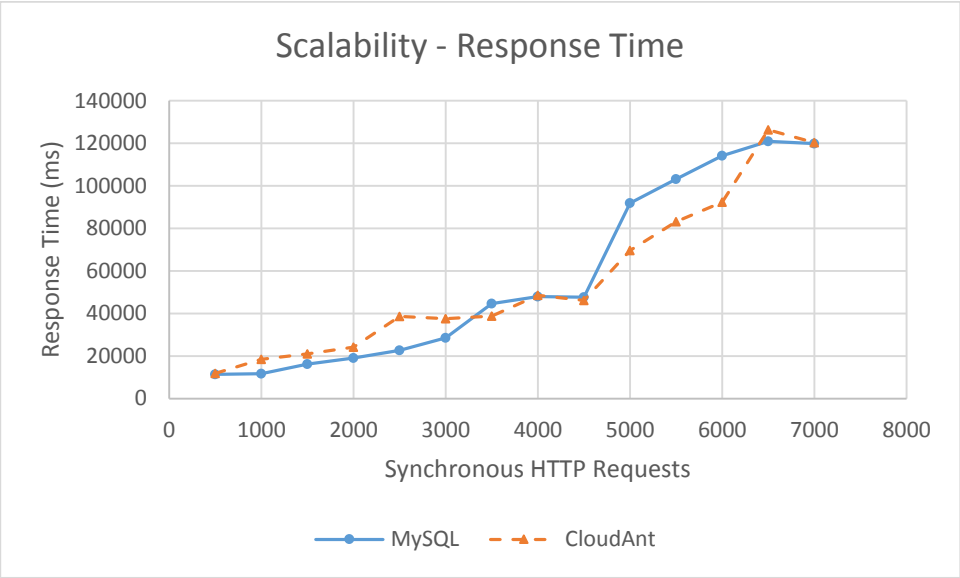

Figure 14 Response Time of Databases

Seen from figure 14, the tendency of response time about MySQL database and Cloudant database are similar. The overall trend is that with the increase of synchronous HTTP requests, the response time gradually increases, and the difference of two databases' performance is small. In this experiment, there is no a significant distinction between MySQL and NoSQL performance.

6.2.2 Error Rate

In figure 15, the error rate denotes incorrect responses' proportion of HTTP requests.

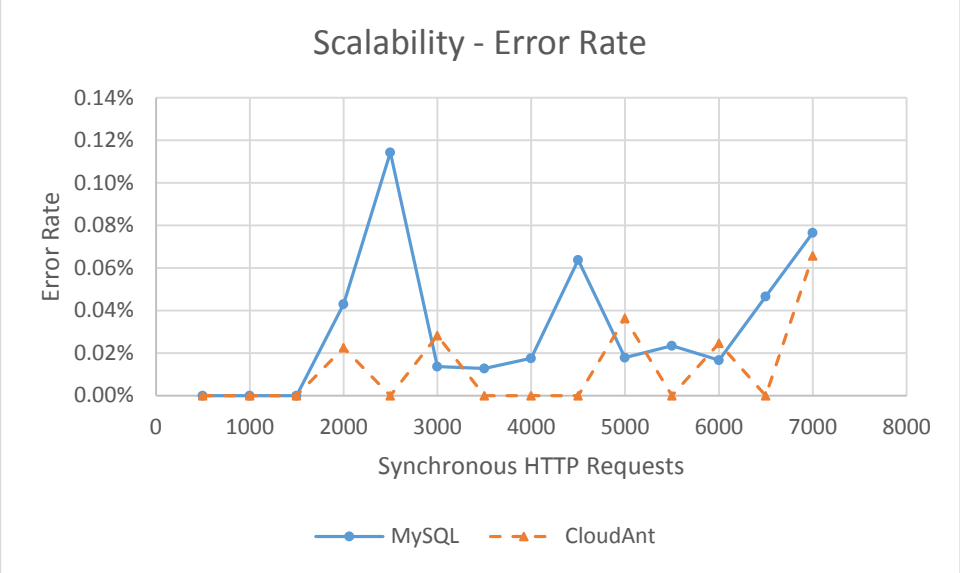

Figure 15 Error Rate of Databases

The results, Compared MySQL and Cloudant, shows that the error rate of two databases are fluctuating and irregular. At the beginning, both MySQL and Cloudant have no error when the number of concurrent requests less than 1500. Next, some errors are emerged on MySQL database and Cloudant database, and there also is a small gap between two databases' error rate.

6.2.3 Throughput

In the scalability experiments, the throughput also has been tested. The throughput in this experiment is that the number of requests per unit of time (seconds, minutes, hours) that are sent to your server during the test. Figure 6-3 shows that the average number of handled requests per second.

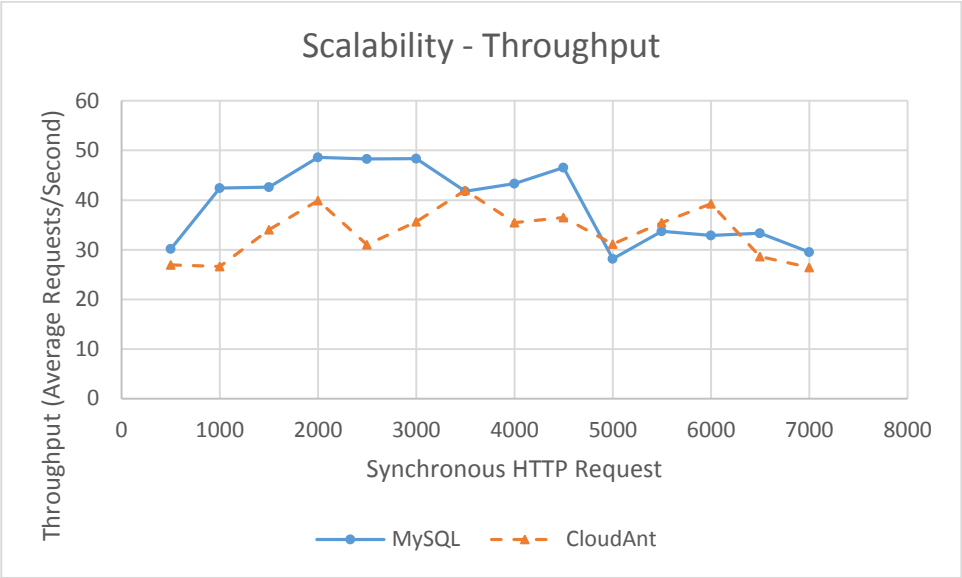

Figure 16 Throughput of Databases

Both overall tendency of figure 16 are that as synchronous requests keep increasing, both throughput increase first then start to drop. When the synchronous HTTP requests less than 5000, the MySQL database response in less time and supports more throughput. And when the synchronous HTTP requests are more than 5000, the concurrent connections weaken two databases' performance and there is a small gap between two databases.

6.3 Efficiency Test

In efficiency experiments, I tested "Query" and "Insert" two operations' run time on MySQL database and NoSQL database with different number of entries. The test script will record database operation run time and log in document, and these scripts directly run on server. In the following figures, two lines on behalf of two databases' efficiency performance. And on every node, the deviation lines show maximum and minimum values of run time.

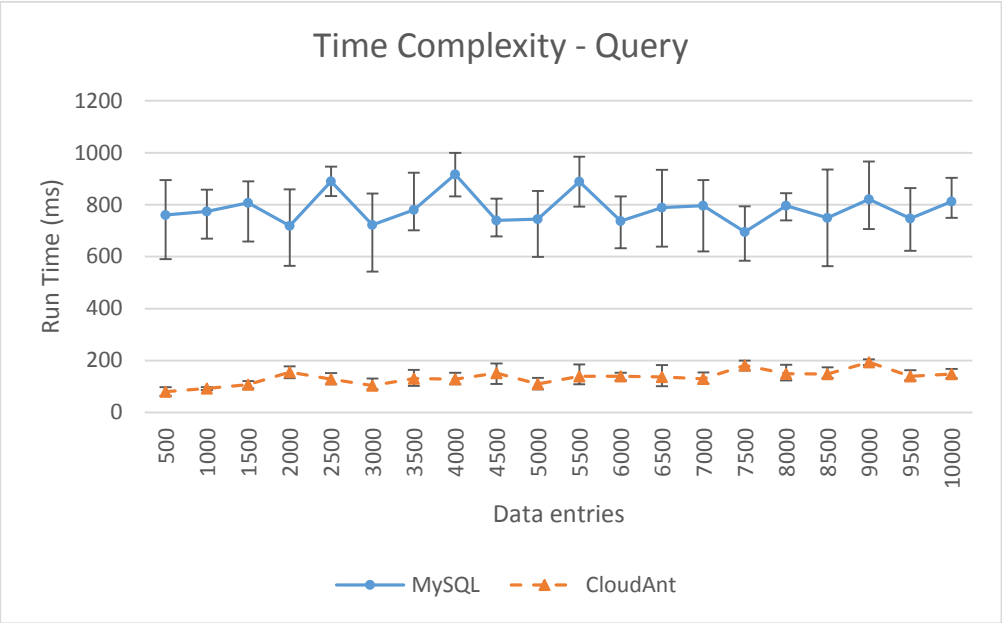

Figure 17 Query Time of Databases

Seen from the figure 17, for one database, there is no a significant changes with different data entries. But compared two databases, between MySQL database and Cloudant database, the efficiency of Cloudant is much better.

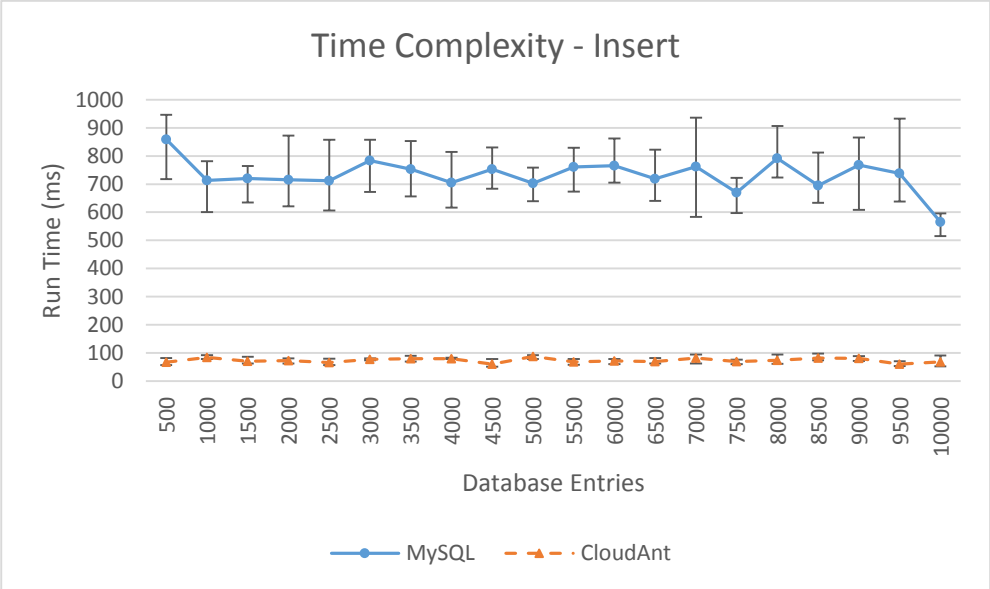

Figure 18 Insert Time of Databases

Similarly with the efficiency of "query" operation, the "insert" efficiency of Cloudant database also is much better than MySQL database.

6.4 Storage Space Test

For Cloudant, the website provides the graphical dashboard for developers. And for MySQL database, the space complexity test script directly run on database. And the SQL test script is:

```
SELECT table_schema AS "Database"
   , table_name AS "Tables"
   , ROUND(((data_length + index_length) / 1024 / 1024), 2) "MB"
FROM information_schema.TABLES
Where TABLE_SCHEMA = "soo" and table_name = "item";
```

The following figure shows the two database space complexity, which means database size in this dissertation.

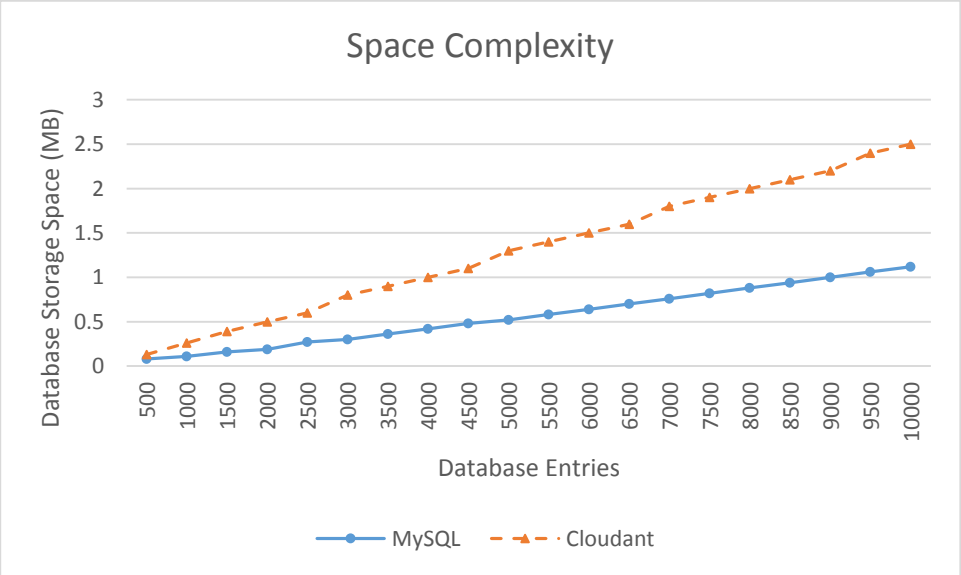

Figure 19 Storage Space of Databases

And obviously, the database storage space has an approximately proportional linear relationship with the number of database entries. The Cloudant database size is larger than MySQL database size. In Relational database, the data is stored in tables. Same type data can be represented by the same table column name, as the figure 20 shows.

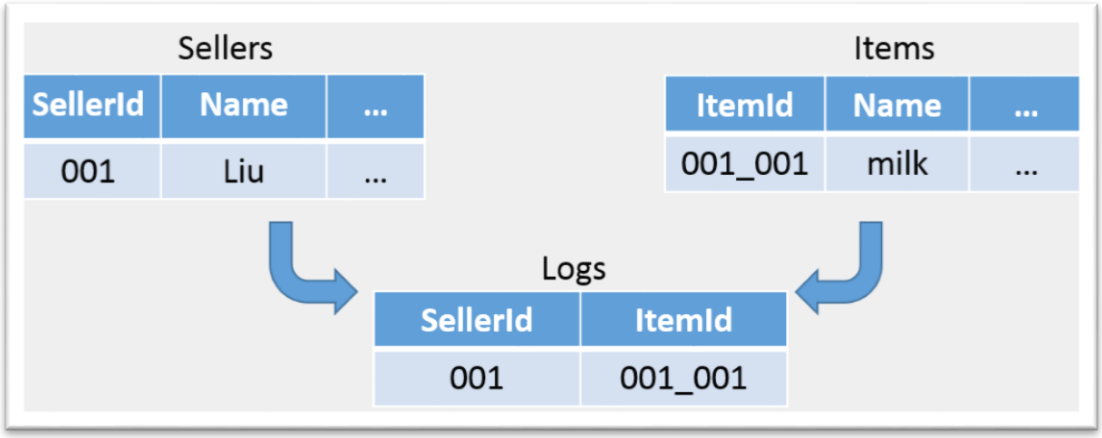

Figure 20 The Diagram of MySQL Tables

For example, if there are many sellers in database, all sellers' id can be indicated by "SellerId". However, in Cloudant database, data is stored in documents and each attribute need to be represented by the string name, for example as the following.


```
{
  "_id": "liu@tcd.ie",
  "_rev": "44-cc795c7b53833e2da2da8e138f026bdb",
  "name": "liu",
  "password": "E10ADC3949BA59ABBE56E057F20F883E",
  "longtitude": -6.2588234,
  "latitude": 53.34451,
  "email": "liu@tcd.ie",
  "expireDate": "Sep 3, 2015",
  "description": "Dublin",
  "adsNo": 32
}
```


The attributes of sellers need to be declared in every document, such as string "_id" will be repeated for each seller. This naming redundancy increases the NoSQL database size. In addition, the Cloudant database also needs additional storage space to maintain the indexes data structure.

6.5 Economic Costs Test

In economic costs experiment, I surveyed the MySQL, Cloudant and cloud server pricing strategy and summarized different database's price in the following table. As we can see, the MySQL is free and Cloudant charge according to the database size. However, the MySQL database need to be supported by the database server, thus, developers need to rent servers for MySQL database. Therefore, MySQL database itself is free and server need to pay (in following table, the fee is the price of Microsoft Azure database server). Cloudant is a Database-as-a-Service, it does not need servers and the database costs depending on the amount of data.

Table 3 Price of Databases

(Per Month)	**MySQL**		**Cloudant**	
	Database	Server	Database	Server
≤50GB	Free	20$+2.59$/GB	Free	Free
>50GB	Free	70$+1.59$/GB	1$/GB	Free

And from MySQL, Microsoft Azure and Cloudant websites, I summary this price and show details in figure 21.

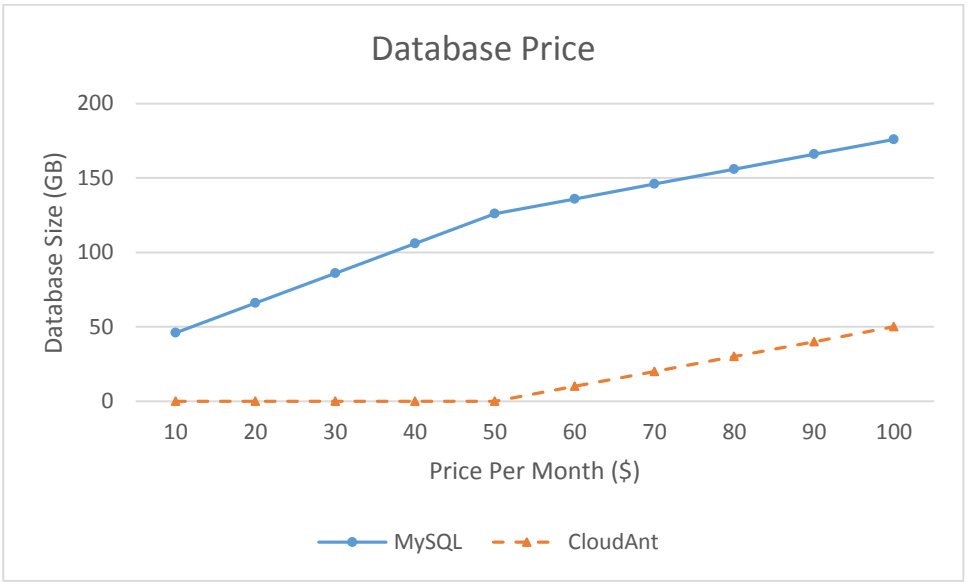

Figure 21 Price of Databases

6.6 Summary

According to these experiments mentioned above, the features and performance of MySQL database and Cloudant database are summarized in the following table. The checkmark "✓" indicates that in terms of this performance, this database shows better. And when two databases both get checkmark "✓", which means performance of two databases is similar.

Table 4 The comparison of MySQL and Cloudant

		MySQL	Cloudant
Scalability	Response Time	✓	✓
	Error Rate	✓	✓
	Through Put	✓	✓
Efficiency	Query Time		✓
	Insert Time		✓
Storage Space		✓	
Economic Costs			✓

Detailed experiments results are shown in above table. Under the current experimental conditions, the scalability of MySQL and Cloudant are similar. But Cloudant is more effective than MySQL in querying and inserting, and Cloudant database size is bigger. From a business point of view, the Cloudant database is more cost-effective.

Chapter 7

Conclusions

The dissertation proposed an analysis of Relational database and NoSQL database about different databases' features and performance on a real e-commerce platform – Soosokan. In the early stages of development, I and my team focused on developing the Soosokan product and did market research. In addition, since the real and uncertain market environment, we used an adaptation of the business development process of the lean startup methodology in the development of Soosokan.

Before put the official product into market, this dissertation aim to choose the best database between MySQL database and Cloudant database for Soosokan. MySQL and Cloudant respectively represent the tradition Relational database and NoSQL database. And each database has advantages and disadvantages, there is some difference between two kinds of database abilities. Therefore, based on features of Soosokan, several performance of database is tested in experiments, including databases' scalability, efficiency, storage space and economic costs.

Seen from the tests results, the scalability of MySQL database and Cloudant database have a small gap with the normal scale of the amount of data. And regarding efficiency, Cloudant database is much better than MySQL database. However, due to the different data fundamental, Cloudant database occupied large storage space. For economic issue for Soosokan, the economic costs of databases and servers that support database operation both need to be considered. Taken together, the Cloudant is the more cost-effective database. According to comprehensive consideration of MySQL database and Cloudant database performance, Soosokan chose the Cloudant as the data infrastructure.

7.1 Future Work

In the future, Soosokan still has a long way to go. As an application, Soosokan need to gather users' feedback and continue refining product. And as a business product, Soosokan need to promote itself to public and develop a large market.

And in this dissertation, since the experiments have undertaken in a limited environment, the results may be far from those obtained if it was carried out in a real "big data" application scenarios. In addition, only MySQL and Cloudant two databases have been researched in my dissertation, the tests of these two databases' performance have some representation. But from a commercial point of view, developers should compare more databases to decide the best one.

Reference

[1] Laney, Doug. "3-D Data Management: Controlling Data Volume." *Velocity and Variety, META Group Original Research Note* (2001).

[2] Jerry Jao. “Why Big Data Is A Must In Ecommerce”, [Online] Avaliable: http://www.bigdatalandscape.com/news/why-big-data-is-a-must-in-ecommerce/ [Cited 9th August 2015]

[3] Kaisler, Stephen, et al. "Big data: Issues and challenges moving forward."*System Sciences (HICSS), 2013 46th Hawaii International Conference on*. IEEE, 2013.

[4] “UNIVAC: UNIVersal Automatic Computer”, (2013). [Online] Avaliable: http://www.thocp.net/hardware/univac.htm#1/ [Cited 9th August 2015]

[5] Meier, Andreas, et al. "Hierarchical to Relational database migration."*Software, IEEE* 11.3 (1994): 21-27.

[6] Codd, Edgar F. "A relational model of data for large shared data banks."*Communications of the ACM* 13.6 (1970): 377-387.

[7] “Oracle Database 12c for Data Warehousing and Big Data” *Oracle White Paper*,2014

[8] “Guide to MySQL and NoSQL - Delivering the Best of Both Worlds”, *MySQL White Paper*, 2012

[9] “Unlocking New Big Data Insights with MySQL”, *MySQL White Paper,* 2015

[10] “Guide to NoSQL with MySQL - Delivering the Best of Both Worlds for Web Scalability”, *MySQL White Paper*, 2015

[11] Mistry, Ross, and Stacia Misner.”*Introducing Microsoft SQL Server 2014”*. Microsoft Press, 2014.

[12] “ Big Data Datasheet - Big Data, Small Data, Any Data”, *Microsoft White Paper*, 2015

[13] Blakeley, José A., et al. "Microsoft sql server parallel data warehouse: Architecture overview." *Enabling Real-Time Business Intelligence*. Springer Berlin Heidelberg, 2012. 53-64.

[14] Grillenberger, Andreas, and Ralf Romeike. "Big Data–Challenges for Computer Science Education." *Informatics in Schools. Teaching and Learning Perspectives*. Springer International Publishing, 2014. 29-40.

[15] Adlinger, Paul. "RDBMS dominate the database market, but NoSQL systems are catching up." *DB Engines. Np, nd Web* 19 (2014).

[16] Chaker Nakhli. "Cassandra’s data model cheat sheet: Data model elements: Column", (2012).[Online] Avaliable: http://www.sinbadsoft.com/blog/cassandra-data-model-cheat-sheet/ [Cited 12th August 2015]

[17] Matthew Aslett, “Neither fish nor fowl: the rise of multi-model databases”, (2013). [Online] Avaliable:

https://blogs.the451group.com/information_management/2013/02/08/neither-fish-nor-fowl/ [Cited 12th August 2015]

[18] "List of NoSQL databases", (2015). [Online] Avaliable: http://nosql-database.org/[Cited 12th August 2015]

[19] Rabl, Tilmann, et al. "Solving big data challenges for enterprise application performance management." *Proceedings of the VLDB Endowment* 5.12 (2012): 1724-1735.

[20] "Benchmarking Top NoSQL Databases - Apache Cassandra, Couchbase, HBase, and MongoDB", *EndPoint*,(2015)

[21] Han, Jing, et al. "Survey on NoSQL database." *Pervasive computing and applications (ICPCA), 2011 6th international conference on*. IEEE, 2011.

[22] Cattell, Rick. "Scalable SQL and NoSQL data stores." *ACM SIGMOD Record*39.4 (2011): 12-27.

[23] "DB-Engines Ranking", (2015). [Online] Avaliable: http://db-engines.com/en/ranking/ [Cited 13th August 2015]

[24] Adam Kocoloski, "ScalingOut CouchDB with BigCouch", (2010). [Online] Avaliable: http://oreillynet.com/pub/e/1760/ [Cited 13th August 2015]

[25] Adam Kocoloski , "The Future of Apache CouchDB", (2012) [Online] Avaliable: https://cloudant.com/blog/the-future-of-couchdb/ [Cited 13th August 2015]

[26] LEE J, “Oracle vs. MySQL vs. SQL Server: A Comparison of Popular RDBMS”, (2013) [Online] Avaliable: https://blog.udemy.com/oracle-vs-mysql-vs-sql-server/ [Cited 13th August 2015]

[27] “Oracle Technology Global Price List, Software Investment Guide”,*Oracle White Paper*, 2015

[28] “MySQL Global Price List December”, *MySQL White Paper*, 2014

[29] “SQL Server Pricing List” [Online] Avaliable: http://www.microsoft.com/en-us/server-cloud/products/sql-server/purchasing.aspx / [Cited 13th August 2015]

[30] Sergey Sverchkov ,“Evaluating NoSQL performance: Which database is right for your data?”, (2014) [Online] Avaliable: https://jaxenter.com/evaluating-nosql-performance-which-database-is-right-for-your-data-107481.html/ [Cited 13th August 2015]

[31] Eisenmann, Thomas R., Eric Ries, and Sarah Dillard. "Hypothesis-driven entrepreneurship: The lean startup." *Harvard Business School Entrepreneurial Management Case* 812-095 (2012).

[32] Osterwalder, Alexander, and Yves Pigneur. *Business model generation: a handbook for visionaries, game changers, and challengers*. John Wiley & Sons, 2010.

[33] Osterwalder, Alexander, and Yves Pigneur. "Business model canvas." *Self published. Last retrieval May* 5 (2010): 2011.

[34] Alex Cowan, “The 30 Minute Business Plan: Business Model Canvas Made Easy”, (2013). [Online] Avaliable: http://www.alexandercowan.com/business-model-canvas-templates/ [Cited 14th August 2015]

[35] Hill, Terry, and Roy Westbrook. "SWOT analysis: it's time for a product recall." *Long range planning* 30.1 (1997): 46-52.

[36] Moogk, Dobrila Rancic. "Minimum viable product and the importance of experimentation in technology startups." *Technology Innovation Management Review* 2.3 (2012).

[37] Ries, Eric. “The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses.” Random House LLC, 2011.

[38] “A Look Back at Startup Funding in 2014”, (2015). [Online] Avaliable: https://www.fundable.com/learn/resources/infographics/look-back-startup-funding-2014/ [Cited 14th August 2015]

[39] Jurica Dujmovic, “20 biggest reasons why startup companies fail”, (2015). [Online] Avaliable: http://www.marketwatch.com/story/20-biggest-reasons-why-startup-companies-fail-2015-06-16/ [Cited 14th August 2015]

[40] Wu, Q., & Wang, Y. (2010, March). Performance testing and optimization of J2EE-based web applications. In *Education Technology and Computer Science (ETCS), 2010 Second International Workshop on* (Vol. 2, pp. 681-683). IEEE.

[41] Ufimtsev, A., Parsons, T., Patcas, L. M., Murphy, J., & Murphy, L. (2006, October). Introducing performance engineering by means of tools and practical exercises. In *CASCON* (p. 379).

Appendix

Table 1. Results of MySQL Scalability Experiments

MySQL	Samples	Average	Median	90%Line	95%Line	99%Line
HTTP Requests	500	11333	11507	15071	15578	16035
HTTP Requests	1000	11682	12014	19389	20696	21459
HTTP Requests	1500	16235	16346	28169	29047	30438
HTTP Requests	2000	19075	18981	34897	37379	38824
HTTP Requests	2500	22628	22310	44530	46803	48912
HTTP Requests	3000	28465	30459	48818	51828	54310
HTTP Requests	3500	44635	45065	73093	76561	79755
HTTP Requests	4000	47922	48207	80796	84544	87311
HTTP Requests	4500	47577	48077	82500	87740	91067
HTTP Requests	5000	91818	92286	156450	161989	166151
HTTP Requests	5500	103213	119842	166345	173453	180544
HTTP Requests	6000	114221	126423	186454	205234	209845
HTTP Requests	6500	120873	135232	201245	214235	220345
HTTP Requests	7000	119830	136456	224124	242354	250345

MySQL	Samples	Min	Max	Error%	Throughput	KB/sec
HTTP Requests	500	6590	16439	0	30.17684	118.7918
HTTP Requests	1000	298	21720	0	42.45023	167.1063
HTTP Requests	1500	203	30670	0	42.62453	167.7925
HTTP Requests	2000	677	39220	0.043	48.62512	177.9246
HTTP Requests	2500	306	50580	0.1144	48.2728	173.64
HTTP Requests	3000	196	55034	0.013667	48.34109	181.0372
HTTP Requests	3500	7598	80605	0	41.80502	164.5664
HTTP Requests	4000	1093	88230	0.0175	43.3515	169.0006
HTTP Requests	4500	998	91875	0.063778	46.53327	176.7105
HTTP Requests	5000	1001	168445	0.0178	28.16679	109.7864
HTTP Requests	5500	892	182345	0.0231%	33.7362	112.2353
HTTP Requests	6000	984	210230	0.022%	32.849d2	153.3245
HTTP Requests	6500	1034	222234	0.0542%	33.3234	195.342
HTTP Requests	7000	998	256345	0.083%	29.5345	143.2345

Table 2. Results of Cloudant Scalability Experiments

Cloudant	Samples	Average	Median	90%Line	95%Line	99%Line
HTTP Requests	500	11817	12100	17062	17789	18293
HTTP Requests	1000	18397	19293	31919	33905	34562
HTTP Requests	1500	21039	20091	36957	39131	40534
HTTP Requests	2000	24145	24803	42399	44858	46414
HTTP Requests	2500	38549	36389	65253	70603	75247
HTTP Requests	3000	37547	35574	69463	73531	78856
HTTP Requests	3500	38752	37984	66826	70204	72701
HTTP Requests	4000	48490	46878	91790	98015	104356
HTTP Requests	4500	56103	56398	104447	110559	115074
HTTP Requests	5000	69564	70182	136478	145306	151742
HTTP Requests	5500	83212	89323	106080	157371	166521
HTTP Requests	6000	92321	98342	134323	204323	232345
HTTP Requests	6500	126324	130932	153242	219342	245235
HTTP Requests	7000	120294	132423	165345	235234	256346

Cloudant	Samples	Min	Max	Error%	Throughput	KB/sec
HTTP Requests	500	4567	18422	0	26.91066	105.9344
HTTP Requests	1000	403	35651	0	26.61202	104.7588
HTTP Requests	1500	462	41359	0	34.06381	134.0929
HTTP Requests	2000	673	47158	0.0225	39.89945	155.1083
HTTP Requests	2500	941	76093	0	31.00775	121.4609
HTTP Requests	3000	265	79832	0.048333	35.6574	134.0968
HTTP Requests	3500	376	73821	0	41.97649	165.2414
HTTP Requests	4000	159	105488	0	35.4676	134.0874
HTTP Requests	4500	359	116695	0	36.48895	141.4756
HTTP Requests	5000	163	152914	0.0564	31.15692	115.4505
HTTP Requests	5500	194	188349	0	35.4342	112.3964
HTTP Requests	6000	345	253422	0.02%	39.2312	120.4223
HTTP Requests	6500	362	274354	0	28.6453	132.2342
HTTP Requests	7000	298	286456	0.07%	26.43456	117.5234

Table 3. Results of MySQL Insert Experiments

MySQL	Insert1	Insert2	Insert3	Ave-SQL	Standard Dev	Max	Min
500	718	913	947	859.3333	100.8971	87.66667	141.3333
1000	782	600	758	713.3333	80.73551	68.66667	113.3333
1500	764	762	635	720.3333	60.3453	43.66667	85.33333
2000	873	651	621	715	112.3922	158	94
2500	858	672	606	712	106.6958	146	106
3000	672	820	858	783.3333	80.23853	74.66667	111.3333
3500	656	853	750	753	80.45288	100	97
4000	687	814	616	705.6667	81.90374	108.3333	89.66667
4500	684	743	831	752.6667	60.40052	78.33333	68.66667
5000	639	759	710	702.6667	49.26346	56.33333	63.66667
5500	780	829	673	760.6667	65.13746	68.33333	87.66667
6000	730	705	862	765.6667	68.87831	96.33333	60.66667
6500	640	822	694	718.6667	76.32096	103.3333	78.66667
7000	768	583	936	762.3333	144.1673	173.6667	179.3333
7500	597	690	722	669.6667	53.01782	52.33333	72.66667
8000	723	907	747	792.3333	81.67143	114.6667	69.33333
8500	633	812	640	695	82.78084	117	62
9000	866	827	609	767.3333	113.085	98.66667	158.3333
9500	644	638	933	738.3333	137.6719	194.6667	100.3333
10000	515	596	585	565.3333	35.87323	30.66667	50.33333

Table 4. Results of Cloudant Insert Experiments

Cloudant	Insert1	Insert2	Insert3	Average	Standard Dev	Max	Min
500	62	56	81	66.3333	10.6562	14.6666	10.3333
1000	78	92	82	84	5.88784	8	6
1500	63	62	86	70.3333	11.0855	15.6666	8.33333
2000	62	76	80	72.6666	7.71722	7.33333	10.6666
2500	63	79	56	66	9.62635	13	10
3000	78	80	74	77.3333	2.49443	2.66666	3.33333
3500	78	69	89	78.6666	8.17856	10.3333	9.66666
4000	78	83	78	79.6666	2.35702	3.33333	1.66666
4500	78	52	51	60.3333	12.4988	17.6666	9.33333
5000	78	91	92	87	6.37704	5	9
5500	78	58	68	68	8.16496	10	10
6000	78	60	75	71	7.87400	7	11
6500	62	82	64	69.3333	8.99382	12.6666	7.33333
7000	94	87	62	81	13.7356	13	19
7500	60	72	76	69.3333	6.79869	6.66666	9.33333
8000	61	94	67	74	14.3527	20	13
8500	97	75	73	81.66667	10.873	15.33333	8.66666
9000	81	71	88	80	6.97615	8	9
9500	57	53	70	60	7.25718	10	7
10000	52	61	91	68	16.67333	23	16

Table 5. Results of MySQL Query Experiments

MySQL	Query1	Query2	Query3	Ave-SQL	Standard Dev	Max	Min
500	796	590	895	760.3333	127.0442	134.6667	170.3333
1000	858	669	795	774	78.57481	84	105
1500	873	890	658	807	105.5872	83	149
2000	859	734	565	719.3333	120.4722	139.6667	154.3333
2500	889	947	833	889.6667	46.54269	57.33333	56.66667
3000	780	543	843	722	129.1588	121	179
3500	717	701	923	780.3333	101.0918	142.6667	79.33333
4000	920	832	1000	917.3333	68.61163	82.66667	85.33333
4500	718	678	823	739.6667	61.14645	83.33333	61.66667
5000	780	599	853	744	106.7739	109	145
5500	889	793	985	889	78.38367	96	96
6000	748	633	832	737.6667	81.56933	94.33333	104.6667
6500	795	639	934	789.3333	120.4999	144.6667	150.3333
7000	873	620	895	796	124.7745	99	176
7500	708	584	794	695.3333	86.19874	98.66667	111.3333
8000	804	740	845	796.3333	43.20751	48.66667	56.33333
8500	748	563	936	749	152.2783	187	186
9000	790	707	967	821.3333	108.4323	145.6667	114.3333
9500	753	623	864	746.6667	98.48971	117.3333	123.6667
10000	785	749	903	812.3333	65.77402	90.66667	63.33333

Table 6. Results of Cloudant Query Experiments

Cloudant	Query1	Query2	Query3	Average	Standard Dev	Max	Min
500	78	65	98	80.3333	13.5728	17.6666	15.3333
1000	94	87	98	93	4.54606	5	6
1500	109	93	121	107.666	11.4697	13.3333	14.6666
2000	156	132	178	155.333	18.7853	22.6666	23.3333
2500	125	109	152	128.666	17.7451	23.3333	19.6666
3000	93	89	131	104.333	18.9267	26.6666	15.3333
3500	125	103	164	130.666	25.2234	33.3333	27.6666
4000	125	109	153	129	18.1842	24	20
4500	156	110	189	151.666	32.3968	37.3333	41.6666
5000	109	89	133	110.333	17.9876	22.6666	21.3333
5500	125	109	185	139.666	32.7142	45.3333	30.6666
6000	141	123	153	139	12.3288	14	16
6500	125	102	183	136.666	34.0816	46.3333	34.6666
7000	125	111	154	130	17.9071	24	19
7500	183	162	200	181.666	15.5420	18.3333	19.6666
8000	141	123	184	149.333	25.5908	34.6666	26.3333
8500	138	134	174	148.666	17.9876	25.3333	14.6666
9000	195	183	205	194.333	8.99382	10.6666	11.3333
9500	131	123	163	139	17.2819	24	16
10000	144	132	168	148	14.9666	20	16

Table 7. Results of Storage Space Experiments

Entries	MySQL	Cloudant
500	0.08	0.13
1000	0.11	0.26
1500	0.16	0.39
2000	0.19	0.5
2500	0.27	0.6
3000	0.3	0.8
3500	0.36	0.9
4000	0.42	1
4500	0.48	1.1
5000	0.52	1.3
5500	0.58	1.4
6000	0.64	1.5
6500	0.7	1.6
7000	0.76	1.8
7500	0.82	1.9
8000	0.88	2
8500	0.94	2.1
9000	1	2.2
9500	1.06	2.4
10000	1.12	2.5

Table 8. Results of Price Experiments

Database Size(GB)	MySQL	Cloudant
10	45.9	0
20	65.9	0
30	85.9	0
40	105.9	0
50	125.9	0
60	135.9	10
70	145.9	20
80	155.9	30
90	165.9	40
100	175.9	50