Master Thesis
Electrical Engineering
December 2012

Performance Evaluation of Cloud Database and Traditional Database in terms of Response Time while Retrieving the Data

Kaushik Donkena
Subbarayudu Gannamani

School of Computing
Blekinge Institute of Technology
371 79 Karlskrona
Sweden

This thesis is submitted to the School of Computing at Blekinge Institute of Technology in partial fulfillment of the requirements for the degree of Master of Science in Electrical Engineering with emphasis on Electrical Engineering. The thesis is equivalent to 20 weeks of full time studies.

Contact Information:

Authors:

Kaushik Donkena
E-mail: kaushik2d@gmail.com
Subbarayudu Gannamani
E-mail: rayudu218@gmail.com

University advisor:
Prof. Lars Lundberg
School of Computing
E-mail: lars.lundberg@bth.se

School of Computing
Blekinge Institute of Technology
371 79 Karlskrona
Sweden

Internet : www.bth.se/com
Phone : +46 455 38 50 00
Fax : +46 455 38 50 57

ABSTRACT

Context: There has been an exponential growth in the size of the databases in the recent times and the same amount of growth is expected in the future. There has been a firm drop in the storage cost followed by a rapid increase in the storage capacity. The entry of Cloud in the recent times has changed the equations. The Performance of the Database plays a vital role in the competition. In this research, an attempt has been made to evaluate and compare the performance of the traditional database and the Cloud Database.

Objectives: This thesis investigates about the prior works on the issues that affect the performance of Cloud Database. And compares the performance of a Database in Traditional to that Cloud Environments

Methods: Two different research methods are used to carry the research. They are Systematic Literature Review (SLR) and Quantitative Methodology. Articles from Scientific Databases are chosen for SLR process.

Results: From the SLR process, 4 issues were identified. From the Experimentation results, Cloud Database is having poor performance compared to the Traditional Database.

Conclusions: Issues that affect the performance of Cloud Database are identified and a test bed is created to test the performance of a Database. Attempts are to be made to improve the performance of Cloud Database.

Keywords: Database, Cloud Computing, Performance, affects

ACKNOWLEDGMENT

Any attempt at any level cannot be satisfactorily completed without the support and guidance of our Supervisor. We express heartfelt gratitude to Prof. Lars Lundberg for his immense support to carry out this work. We are much thankful to librarian Sophia Swartz for her guidance in SLR. We are greatly thankful to our beloved parents, brothers and friends for their relentless support that they had given us to reach our goals.

Yours truly,
Kaushik Donkena,
Subbarayudu Gannamani.

CONTENTS

ABSTRACT .... **I**

ACKNOWLEDGMENT .... **II**

CONTENTS .... **III**

LIST OF FIGURES **1**

LIST OF TABLES **2**

LIST OF ABBREVIATIONS **3**

1 INTRODUCTION **4**

1.1 AIMS AND OBJECTIVES 4
1.2 RESEARCH QUESTIONS 5
1.3 THESIS OUTLINE 5

2 BACKGROUND **6**

2.1 DATABASE 8
2.1.1 Database Management System 9
2.1.2 Database Optimization 10
2.2 CLOUD COMPUTING 12
2.3 DEPLOYMENT MODELS 13
2.2.1 Private Cloud 13
2.2.2 Community Cloud: 13
2.2.3 Public Cloud 14
2.2.4 Hybrid Cloud 14
2.3 SERVICE MODELS 14
2.3.1 Software as a Service (SaaS) 14
2.3.2 Platform as a Service (PaaS) 14
2.3.3 Infrastructure as a Service (IaaS) 14

3 RESEARCH METHODOLOGY **15**

3.1 SYSTEMATIC LITERATURE REVIEW (SLR) 15
3.1.1 Planning the review 16
3.1.2 Conducting the review 17
3.1.3 Identification of Research 17
3.1.4 Study Selection Criteria 18
3.2 EXPERIMENT 19
3.2.1 On Traditional Database 19
3.2.2 Constructing a test bed 21
3.2.3 Database Normalization 22
3.3 CLOUD DATABASE 22

4 RESULTS **24**

4.1 SLR RESULTS 24
4.2 EXPERIMENTAL RESULTS 24
4.2.1 QUERY 1 25
4.2.2 QUERY 2 26
4.2.3 QUERY 3 (SELECT COMMAND USING SIMPLE JOIN) 27
4.2.4 QUERY 4 (SELECT COMMAND USING COMPLEX JOIN) 29

5 DISCUSSION **31**

5.1 VALIDITY THREATS 31

5.1.1 Construct Validity *32*
5.1.2 Internal Validity *32*
5.1.3 External Validity *32*
5.1.4 Conclusion Validity *33*

6 CONCLUSIONS 34

6.1 LINKING RESEARCH QUESTIONS 34
6.1.1 Research Question 1 .... *34*
6.1.2 Research Question 2 .... *34*
6.2 FUTURE WORK 34

REFERENCES 35

APPENDIX A 38

APPENDIX B 39

APPENDIX C 40

APPENDIX D 41

APPENDIX E 42

APPENDIX F 43

APPENDIX G 44

APPENDIX H 45

APPENDIX I 46

APPENDIX J 47

LIST OF FIGURES

Figure 2-1 Journey of Relational Database Management System 7
Figure 2-2 Cloud Database as a Service 8
Figure 2-3 Database.. 9
Figure 2-4 Database Management System 10
Figure 2-5 Database Performance Optimization Dependency levels 11
Figure 2-6 Cloud Usage.. 12
Figure 3-1 showing the entity relationship diagrams for Employee database....................... 20
Figure 3-2 Database schema of EMPLOYEE Database.. 21
Figure 4-1 Slow Down Factor between Traditional and Cloud Databases for different entries for Query 1 26
Figure 4-2 Slow Down Factor between Traditional and Cloud Databases for different entities for Query 2 27
Figure 4-3 Slow down Factor between Traditional and Cloud Databases for different entities for Query 3 28
Figure 4-4 Slow Down Factor between Traditional and Cloud Databases for different entities for Query 4 30

LIST OF TABLES

Table 2-1 Advantages and Disadvantages of using Indexes 12
Table 3-1 Research plan 15
Table 3-2: Defining Research Questions 16
Table 3-3 Quality Assessment checklist 17
Table 3-4 Selection Criteria 17
Table 3-5 SLR Process 19
Table 3-6 Entities and attributes in Employee database 21
Table 3-7 Entity relationship and keys information 22
Table 4-1 SLR Results 24
Table 4-2 Query 1 Response Time Values of different entries for Traditional and Cloud Database in milliseconds 25
Table 4-3 Data entries of the Query 1 25
Table 4-4 Query 2 Response Time Values of different entries for Traditional and Cloud Database in milliseconds 27
Table 4-5 Data entries of the Query 2 27
Table 4-6 Query 3 Response Time Values of different entries for Traditional and Cloud Database in milliseconds 28
Table 4-7 Data entries of the Query 3 28
Table 4-8 Query 4 Response Time Values of different entries for Traditional and Cloud Database in milliseconds 29
Table 4-9 Data entries of the Query 4 29

LIST OF ABBREVIATIONS

DBMS	Database Management System
DaaS	Database as a Service
DML	Data Manipulation Language
IaaS	Infrastructure as a Service
PaaS	Platform as a Service
SaaS	Software as a Service
SQL	Structured Query Language
SLR	Systematic Literature Review
RDMS	Relational Database Management System

1 INTRODUCTION

A Cloud can be defined as a parallel and distributed system which has a number of virtualized and interconnected computers. These are actively provisioned and presented as single or more united computing resources depending upon the service level agreement. Cloud has three popular computing paradigms Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). These services include distributed operating system, the distributed database and other services.

The Cloud Computing database is required apace and effectively and should reduce the burdens during routing configuration. The Cloud Database is constructed by collecting a number of sites. The sites are also called as nodes which are interlinked by a communication network. Every single node is a database class. Each database class has its own database, terminals, the central processor and their individual local database management system.

A database is an organized collection of data. A Database Management System (DBMS) is a software package with computer programs that controls the creation, maintenance, and use of a database. It allows the organizations to conveniently develop databases for various applications. A database is an integrated collection of data records, files and other objects. A DBMS allows different user application programs to concurrently access the same database. DBMSs may use a variety of database models, such as the relational model or object model to conveniently describe and support applications. The term database is correctly applied to the data and their supporting data structures, and not to the database management system. The database along with DBMS is collectively called Database System.

A Cloud Database is a database that typically runs on a Cloud Computing platform, such as Windows Azure, Amazon EC2, GoGrid and Rackspace. There are two common deployment models: users can run databases on the cloud independently, using a virtual machine image, or they can purchase access to a database service, maintained by a Cloud Database provider. .Of the databases available on the Cloud, some are SQL-based and some use a NoSQL data model.

1.1 Aims and objectives

Aim of the thesis is to evaluate the performance comparisons of traditional and normal database and open doors for research on the performance issues in Cloud Database.

- Creating and deploying data into the traditional database
- Migrating and deploying data into Cloud Database
- Test traditional database performance
- Test Cloud Database performance
- Compare the results of traditional database and Cloud Database in terms of response time

1.2 Research questions

1. What are the issues that affect the performance of a Cloud Database?
2. What is the performance in terms of response time of a Cloud Database compared to traditional database?

1.3 Thesis Outline

Introduction part describes the brief introduction to the research work. Background consists of background of Databases and the background of Cloud Computing. Research Methodology discusses the methodologies used for the research. This consists of SLR and Quantitative Methodology. Results chapter presents the SLR(Systematic Literature Review) Results and Experimentation results. Discussion gives a brief discussion on the obtained results. Conclusions chapter discusses the conclusions linking the research questions and the future directions of the research. References give the list of used citations and Appendix gives information on the experiment and its results.

2 BACKGROUND

The concept of database management system is quite interesting to look at over a particular period of time. According to [27], Database Management is developed in four phases from 1970's to late 1990's. Figure [1] clearly illustrates four phases of Database Management System. In early 1970's, organizations used IBM's information management system (IMS) which stores the data using hierarchical model. But the organizations have to maintain expensive main frames in order to relay on IBM's IMS. By early 1980's, IBM's IMS is replaced by the Relational Database Management System (RDMS) such as Oracle. In 1980's and 1990's amplification of networking DBMS technology is allowed on personal computers. After RDBMS progress to client /server environments and it's implemented on large organizations. In 1990's because of the fast growth of the technology symmetric multiprocessing system and data warehousing options are made available on the RDBMS.

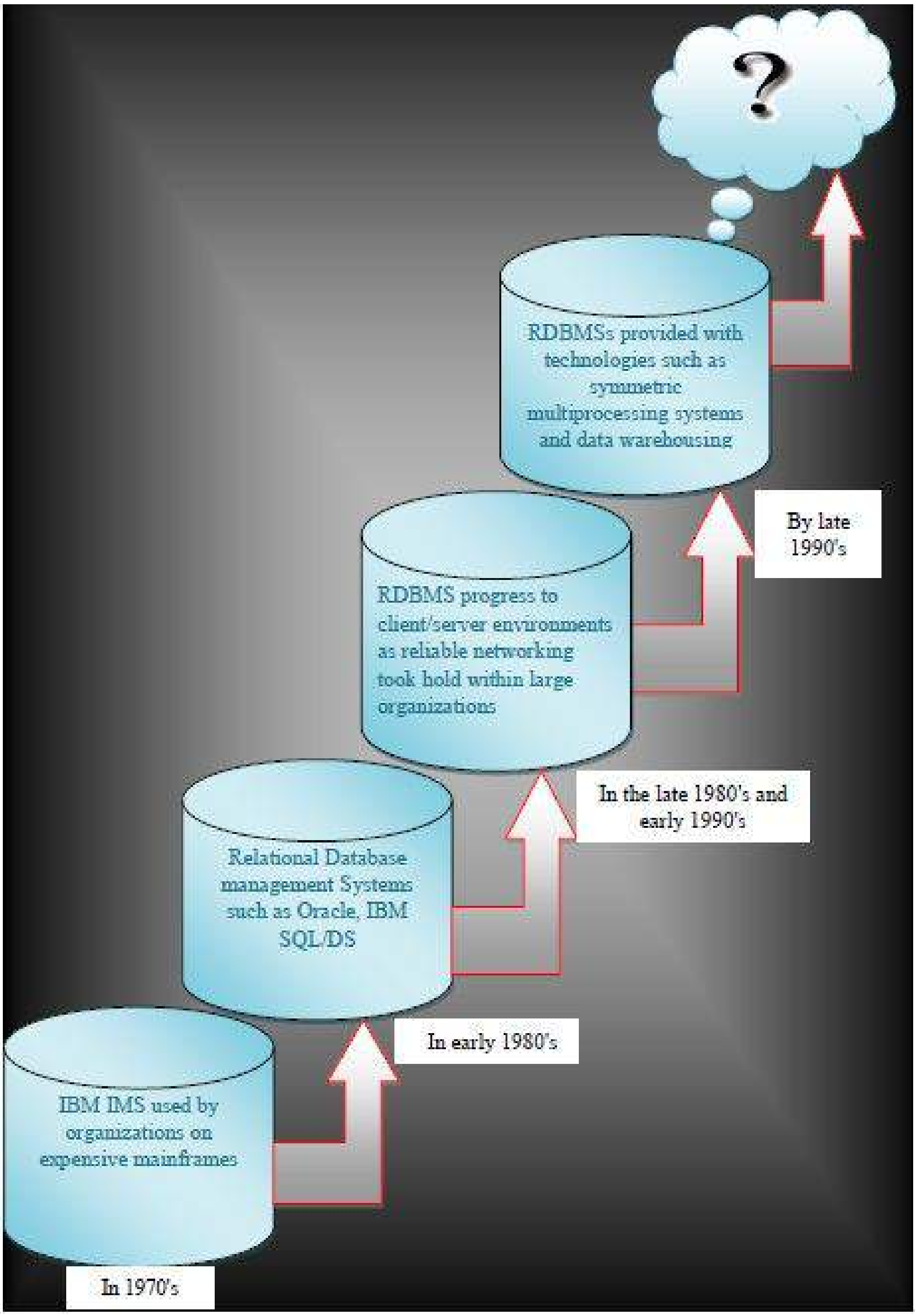

Figure 2-1 Journey of Relational Database Management System

According to [29] Figure 2-1 shows the phases of the Relational Database Management System. This has kept growing and now this time it shifted to other dimension i.e Cloud Computing. Cloud Computing has been an interesting paradigm in the recent times due to its advantages like scalability, virtualization and pay per use. As pay per use is involved, it is important to consider the resource utilization. Cloud Computing is more helpful for IT industries to improve the management of their own resources in an easy manner. Cloud Computing provides different services such as Infrastructure-as-a-Service(SaaS), Platform-as-a-Service(PaaS) and Software-as-a-Service(SaaS). According to [33] there is an addition to this list of services, called Database-as-a-Service(DaaS). In this service, organizations host their own databases in

Cloud Computing. This service provides the acess for DML(Data Manipulation Language) statement features (strore, retrieve, update and delete the data) via the internet following [29].

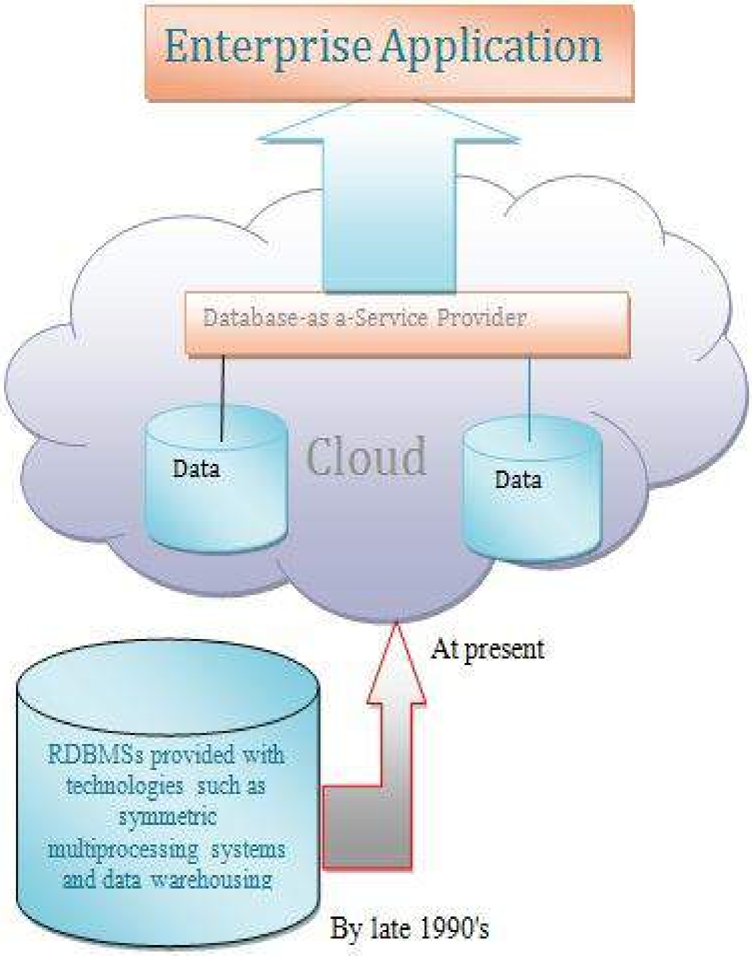

Figure 2-2 Cloud Database as a Service

According to [28], a Cloud Database is a combination of different number of nodes (or site collections) and each node has its own database, linked together in the communication network. Cloud Database system is a novel trend in the research because many organizations want to migrate their databases into Cloud to exploit the benefits Cloud Computing. Organizations look at the performance factor of the databases regardless of the paradigm, whether traditional or Cloud. In [30], authors conducted various experiments on On-premisis traditional database in terms of IBM’S DB2, Oracle database and Microsoft SQL Server. The performance of the Cloud Database is evaluated in this research and a comparison is made with that of an on-premises traditional database.

2.1 Database

Database is a collection of data or information in a well-organized manner so that data can be accessed, updated and managed easily. It can be imagined as a large data file storing the data as in the following.

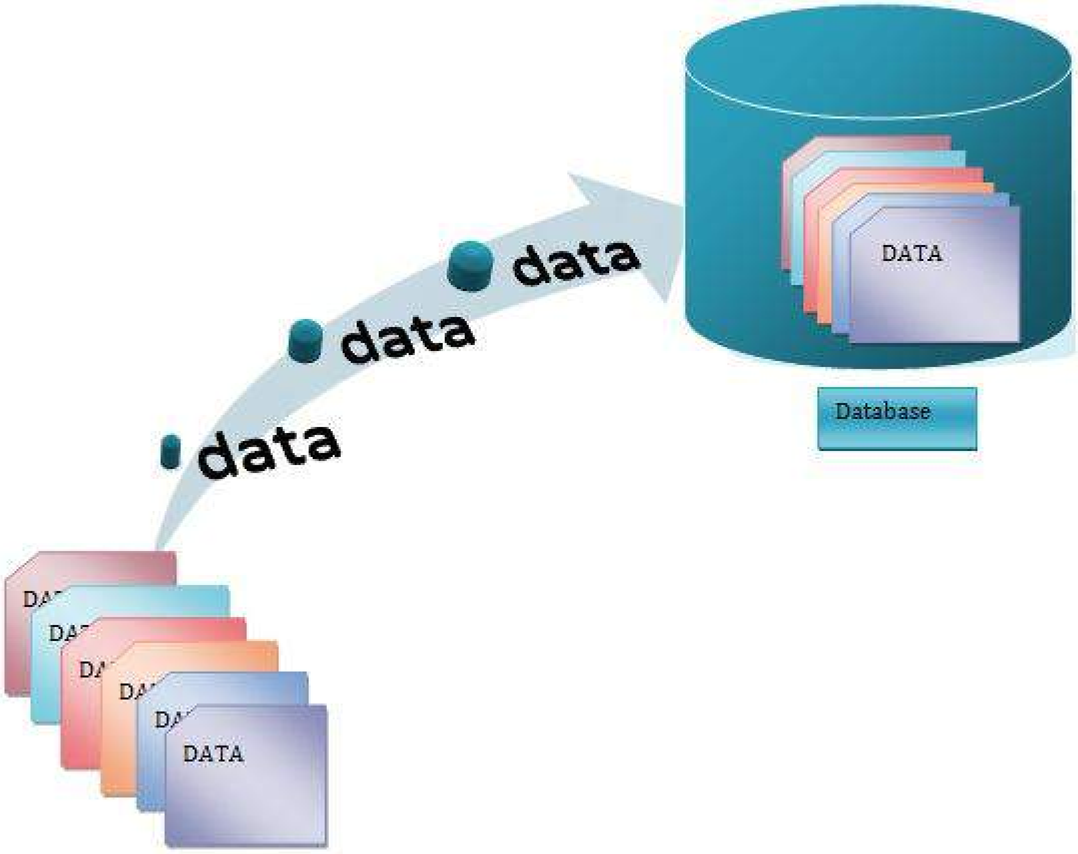

Figure 2-3 Database

As shown in the figure database is an integrated collection of data items or files. According to [31], the authors suggested the databases have to support features such as high reliability, high availability, high throughput and security. A database is rated as a high quality database if it supports aforementioned features in all operations such updating, managing, and retrieving of data. Enterprises will plan for the provision of these features while providing service to the database users.

2.1.1 Database Management System

A Database Management System is software with computer programs that lets the user control the creation, maintenance, and use of a database. According to [32] database package provides to the user a database engine, a data dictionary and a user interface. The database engine is used for the purpose of effective storage and retrieval of data. The purpose of user interface is to create a new database or update an existing database in the system. According to IBM dictionary of computing a data dictionary is a centralized repository of information about data such as meaning, relationships to other data, origin, usage, and format. It is a document which determines the structure of a database and describes a database. A DBMS can facilitate the concurrent access of multiple databases via user interface.

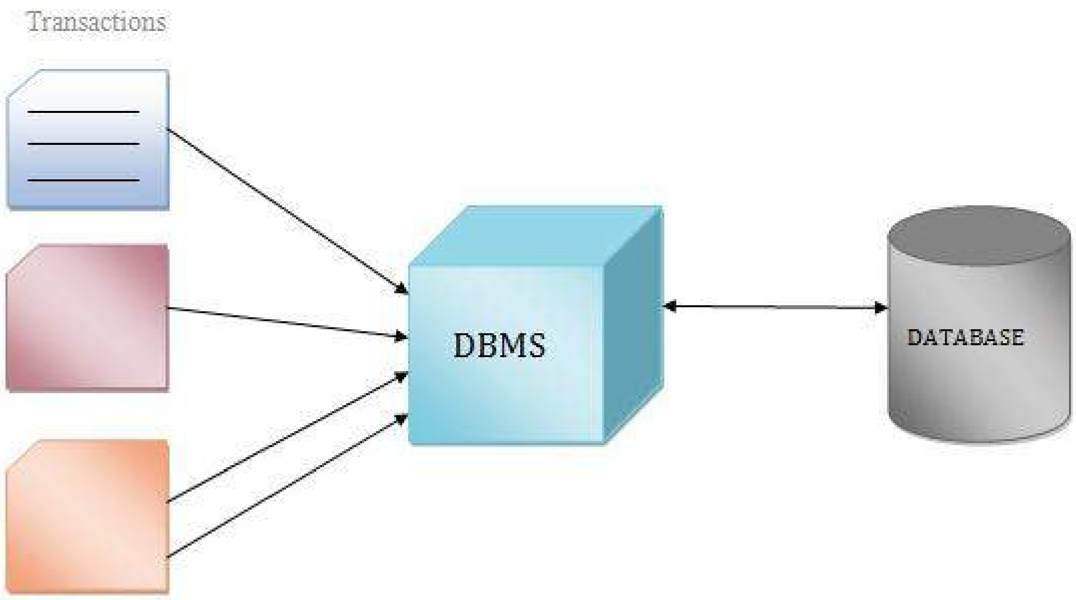

Figure 2-4 Database Management System

According to [31], a Database Management Systems acts like a platform for database administrators to manage, create and update the database. Users can run certain applications in the DBMS to access, modify and update the data. According to [32] there are different kinds of databases such as network, hierarchical and relational. Relational database was proposed by E.F.Codd in 1963. A relational database is the predominant choice in storing data, over other models like the hierarchical database model or the network model.

2.1.2 Database Optimization

According to [4] enterprises are becoming data-centric and increasingly producing humongous amounts of data in the form of sales, retail records and other commercial information. This data stored in the database needs to be effectively managed. Enterprises analyze these databases continuously and take informed decisions based on the analysis, so database performance plays a vital role in the overall functioning of the database. At the time of creation of database the scale of meta-data related to the database is small. As the size of the database increases, it encounters gradual deterioration in the performance. This performance degradation motivated the researchers to search for ways to improve the performance by database optimization. Database optimization can be performed at four different layers as shown in Figure 2-5.

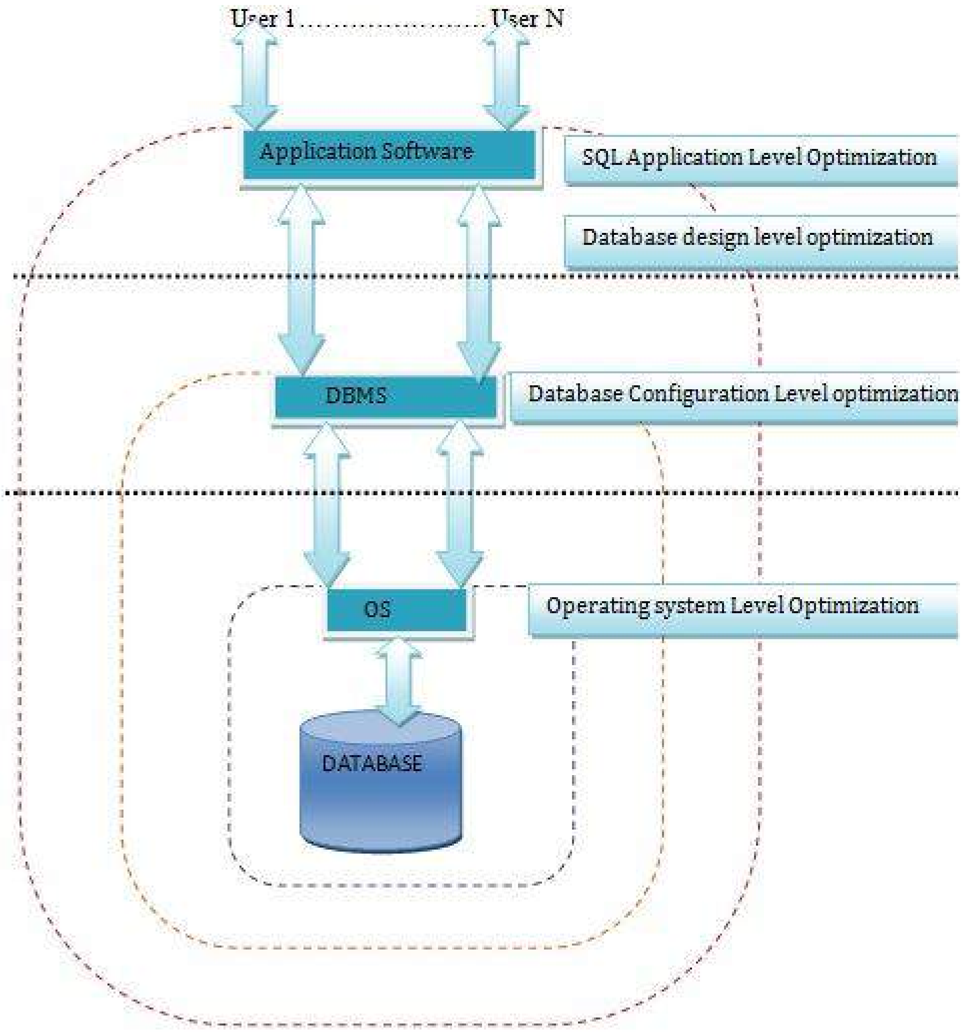

Figure 2-5 Database Performance Optimization Dependency levels

In these four levels top most level is the SQL application level optimization. In this optimization the transaction time is reduced by indexing the database thereby leading to improvement in the performance. The database performance translates to reduction in CPU costs in [35]. By indexing the database, the DBMS is enabled to maintain a separate database object storing the metadata related to database. These objects contained a sorted list of column values which contains row identifiers to the corresponding rows in that table as shown in [34].

Indexes are internally organized in a tree structure. According to [37] there are certain disadvantages of using the indexes to the database. Usage of the indexes results in speed up in the query execution, retrieval of data but every additional index added to the index table slows down the manipulation further. Since every INSERT/DELETE/UPDATE can be processed only after updating all the corresponding indexes it takes additional CPU cycles and time to keep the indexes synchronized with the tables. This also results in Database consuming additional space in database.

Table 2-1 Advantages and Disadvantages of using Indexes

	Advantages	Disadvantages
1	Optimize the database performance	Using Index slows down manipulation further
2	Using indexes we can speed up queries	Maintenance overhead
3	Reduce CPU cost for query execution	Indexes occupy the additional space in database
4	Avoids full table scan in search queries	INSERT/DELETE/UPDATE can be processed only after updating all the corresponding indexes
5	Table data can be stored in an organized way	Need to maintain index and table synchronization every time.

2.2 Cloud Computing

It is hard to define what Cloud Computing is because different authors have different definitions on Cloud Computing. But according to NIST (National Institute of standards and technology)“Cloud Computing is a model for enabling ubiquitous, convenient, on-demand network access to shared pool of configurable computing resources (e.g.., networks, servers, storage and applications) that can be rapidly provisioned and released with minimal management effort or service provider interaction”.

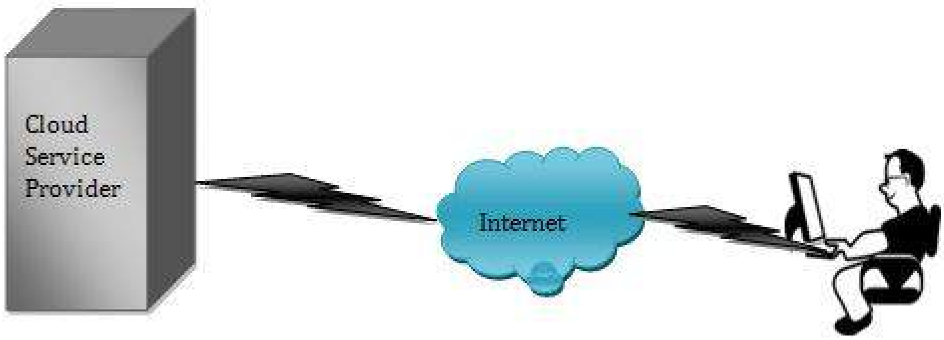

Figure 2-6 Cloud Usage

Cloud Computing has five essential characteristics (On-demand self-service, Broad network access, Resource pooling, Rapid elasticity and Measured service), three service models (Software as a service, Platform as a service and Infrastructure as a service) and four deployment models (Private Cloud, community Cloud, public Cloud and Hybrid Cloud).

2.3 Deployment Models

According to [36] four types of deployment services available in the Cloud they are Private Cloud, Public Cloud, Hybrid Cloud, and Community Cloud. Below Figure 2-7 Cloud Deployment Models clearly illustrates

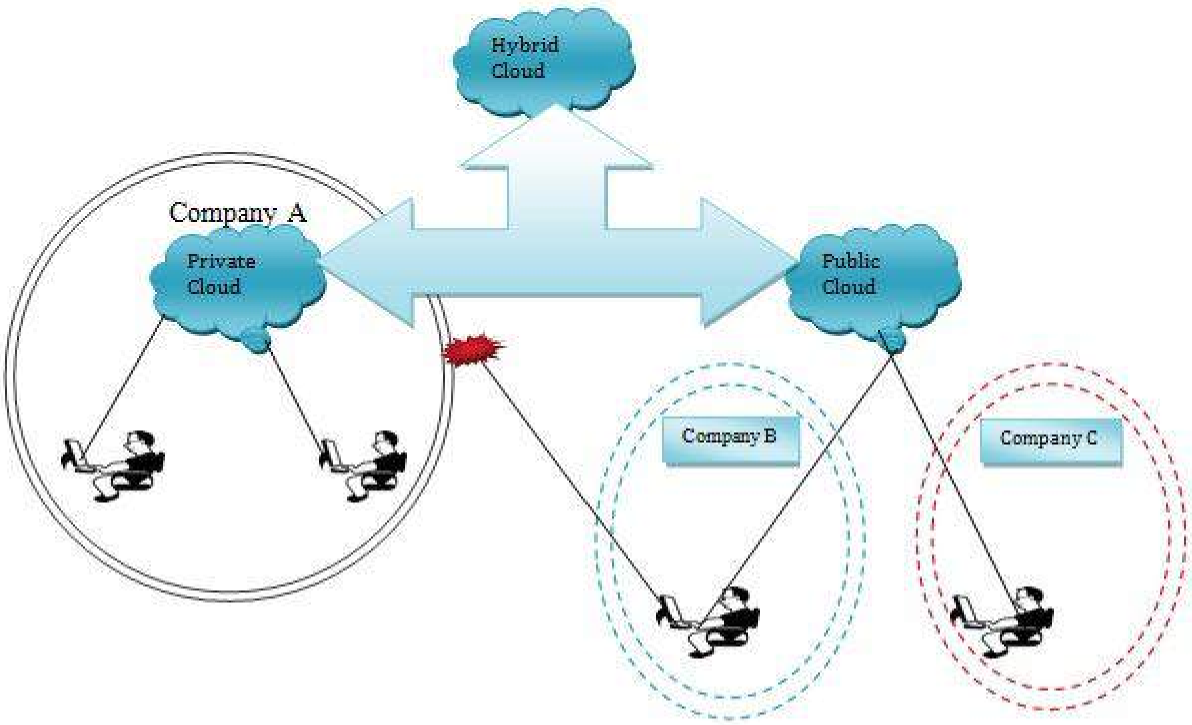

Figure 2-7 Cloud Deployment Models

The above figure clearly shows the variation between the private, Public, and Hybrid Clouds. Company 'A' owns private Cloud whereas company 'B' and company 'C' owns Public Cloud.

2.2.1 Private Cloud

Private Cloud is also called as internal Cloud or corporate Cloud. Private Cloud is providing resource, storage of data to a limited number of hosted services. This Cloud may be managed and operated by the organization behind a firewall. Private Cloud can access who are positioned within the boundaries of an organization.

2.2.2 Community Cloud:

Community Cloud is a type of infrastructure to share a resource to many organizations from a specific community with common concerns (e.g. security requirements, mission, policy, compliance considerations).

2.2.3 Public Cloud

This cloud infrastructure is employed for delivering resources to general public over the internet for open use. It may be managed and owned by academia for academic purposes or by the government or corporate for commercial purposes.

2.2.4 Hybrid Cloud

This cloud infrastructure is a combination of two or more distinct clouds. In this model an organization provides and manages some resources in-house and has others provided externally. It offers the benefits of multiple deployment models to the users.

.

2.3 Service models

2.3.1 Software as a Service (SaaS)

"This capability provided to the consumer is to use the provider's applications running on a Cloud infrastructure. The applications are accessible from various client devices through either a thin client interface, such as a web browser (e.g., web-based email), or a program interface." [26]

2.3.2 Platform as a Service (PaaS)

"This capability provided to the consumer is to deploy onto the Cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider." [26]

2.3.3 Infrastructure as a Service (IaaS)

"This capability provided to the consumer is provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications." [26]

3 RESEARCH METHODOLOGY

The two research questions follow two different methodologies. They are represented in the table.

Table 3-1 Research plan

Research Question	Research Methodology
1. What are the issues that affect the performance of a Cloud Database?	Systematic Literature Review
2. What is the performance in terms of response time of a Cloud Database compared to traditional database?	Experimentation

3.1 Systematic Literature Review (SLR)

SLR is an important research methodology in research work. SLR is a means of identifying, evaluating and interpreting all the available relevant work for a particular topic or phenomenon of interest [10]. SLR's provide a fair evaluation of research work with a trustworthy, auditable and rigorous methodology. This can be attempted by a predefined search strategy. This search strategy should be able to cover the whole related research to be assessed. The researchers should make every effort to identify the related research which is helpful as well as non-related research which is not helpful for his research work. SLR's are mainly used to summarize the existing evidence, identifying the gaps in the ongoing research and designing a frame work for a novel research.

According to research question one; there is a necessity to know the issues that affect the performance of a database. There are a very few articles which summarize the performance of a database. It has become a major cause to conduct a SLR to bridge the gap and to get a clear understanding on the issues affecting the database performance.

The three phases of SLR are:

- Planning the review
- Conducting the review
- Reporting the review

3.1.1 Planning the review

There will be a number of normal literature reviews conducted which normally lacks scientific value and contribution. In order to identify any prior SLR's, a preliminary search is done with framed search string. A selection procedure of the publication is done based on the title, abstract, introduction and conclusion if necessary. In every publication, deep scrutiny is needed for a SLR. The scientific databases used are Scopus, ScienceDirect and Inspec. As there are no hits for this search, this motivated to perform a systematic literature review.

{Cloud Database} OR {Cloud Database affects} AND {systematic review} OR {systematic literature review}

3.1.1.1 Defining the research question

Research Question for Systematic Literature Review

Research Question	Purpose
What are the issues that affect the performance of a Cloud Database?	To identify the issues that have affect on the performance of a Cloud Database.

Table 3-2: Defining Research Questions

3.1.1.2 Defining keywords

As per the guidelines provided by [10], a PICO criterion is used for defining the key words.

PICO – Population Intervention Comparison Outcomes

Population: Population refers to a specific role, kind, area or application. Here "Cloud Computing" is chosen as population for the research.

Intervention: Intervention addresses the technology or procedure or tool that deals with a specific issue. "Database" and "Performance" are chosen as intervention for this research.

Comparison: Comparison is the tool or procedure or technology with which the intervention is to be compared. No comparison is done in this research.

Outcomes: The outcomes must relate the factors that are important for a specific tool. These relevant outcomes should be presented. "affects", "problems" and "issues" are chosen as outcomes.

3.1.1.3 Study Quality Assessment

The quality assessment is required to assure that the relevant and primary studies were included during the process and must fulfill the overall aims and objectives of the research. A checklist is prepared according to the guidelines given by [10]. They are

Table 3-3 Quality Assessment checklist

Quality Assessment questions	Yes/No
Does the study clearly state aims and objectives?	-
Was it clear which research method was carried out and explained?	-
Are the findings of research clearly stated?	-
Does the author discuss the limitation constraints?	-

3.1.1.4 Selection Criteria

The guidance for the selection criteria is given in [10]. According to the guidelines, relevant articles are chosen. The inclusion and exclusion criterion helped to filter out the irrelevant articles. The selection criterion is shown in the following table.

Table 3-4 Selection Criteria

Relevance	Criteria
By Search	According to Search String Publication Year (2005-2012)
Title	Language used (English) Related to Database performance
Abstract/Introduction/Conclusion	Background in industrial or academic in related area
Full text	Performance issues on Cloud Database

3.1.2 Conducting the review

3.1.2.1 Data Extraction Strategy

Data extraction strategy is performed for this study. The aim of the extraction strategy is to extract the information concerned with Cloud Database performance and its affects. The information is collected from the popular databases and the inclusion and exclusion selection criteria are applied. The formation of the search strings becomes the first step for the search. Here Cloud Computing, Database, Performance, Issues, Problems and affects has become the components of our search strings. This search will be refined according to year from 2005 to 2012.

3.1.2.2 Identification of Research

The first step of systematic review is to create a search strategy to get the primary information related to the research question [10]. The keywords are selected as mentioned and search stings are constructed using the Boolean operators like ANDs and ORs. The papers are identified by searching them with different search strings in the standard databases like Inspec, ScienceDirect and Scopus. The relevant papers are chosen as references.

The keywords that are used for the construction of the search strings are

- Cloud Computing
- Database
- Performance
- Issues
- Problems

The following are the search strings that are constructed according to the research question for systematic review.

((("Cloud Computing") OR (Cloud)) AND (Database) AND (Performance) AND ((Issues) OR (Problems)))

3.1.3 Study Selection Criteria

The study selection criteria provide the evidence for the primary studies about the research question [10]. For this research, the intrusion and exclusion criteria are used for the filter and refine the papers.

Inclusion Criteria

- Studies which are covering the database issues in Cloud Computing
- Studies that reflect the factors that affect the performance of a database in Cloud
- Studies that include the future challenges on the performance of Cloud Databases

Exclusion Criteria

- Studies in languages other than English
- Studies which are not reflecting the database issues in Cloud Computing

Table 3-5 SLR Process

Steps	Inspec	Scopus	ScienceDirect
Articles found in initial search	106	118	2447
Refinement specified in the Appendix	91	86	79
Refinement of Cloud keyword in the title	55	55	38
Screening by topic relevant titles	27	14	0
Combined relevant titles of 3 databases	41		
Screening by duplicates and language	31		
Screening by Abstract, Introduction and Conclusion	14		
Screening by reading full text	5		

3.2 Experiment

3.2.1 On Traditional Database

For the second research question, quantitative work was done to measure the mentioned parameters in traditional and Cloud database. Performance of a database can be measured in terms of response time, throughput, cost per transaction and resource utilization (amount of system resources utilized for particular user operation) [8]. When the queries take long time to execute, it shows a negative impact on the response time. This results in the performance of a database getting diminished. So the query response time is considered as the parameter for the measurement of database performance. The CPU cycles can also be taken as a parameter but the configurations of the Cloud Database are undisclosed. So it is not chosen as a parameter to measure the performance

According to [5] the response time is defined as the time taken by the system to complete user command. The optimum response time of a system must not exceed by the limit of specified response time. A similar study has been made measuring the performance of different Cloud Databases [38].

For the experimentation, the relational database named 'Employee Database' is created in the traditional database environment. Microsoft SQL 2008 R2 is chosen as the traditional database. The data for Employee database is collected from an online data generator [20]. The data is filled into the relational database using the 'Insert' statement. The whole experiment is planned in the single table i.e. Employee Database. The experiment aims to check the performance of both the databases while increasing the data entries. First the 30,000 entries are entered and the queries are performed. Later another 30,000 entries are added to the existing entries and the database is doubled. The data is entered into the database and the number of data entries is added as 30,000 entries, 60,000 entries, 120,000 entries and 240,000 entries. Windows Azure is chosen as the Cloud Database. Windows Azure offers SQL Database. As the number of database entries increased, the performance of both the databases is tested with the queries framed and this is repeated in each case.

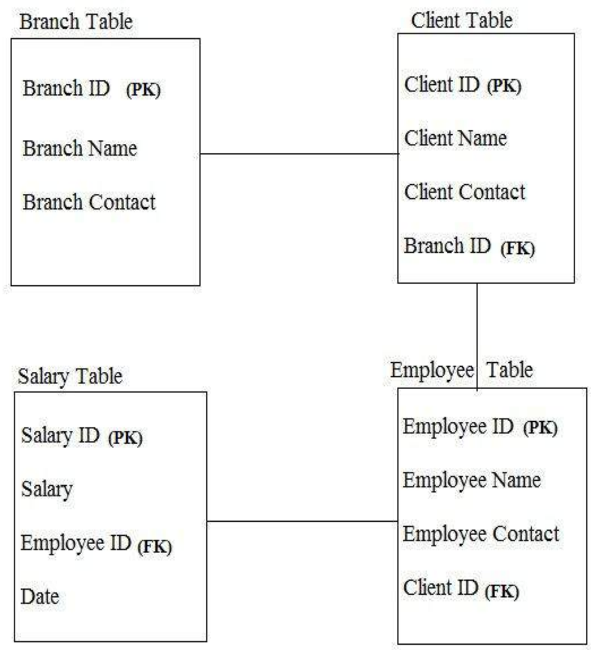

Figure 3-1 showing the entity relationship diagrams for Employee database

A better platform is build with suited relationships among the tables for testing the performance. The query elapsed time (Response time) is taken as measurement in both the databases across data manipulation language statement SELECT (to scan the data).The operations can be

1. Select few rows among many rows in the table by using simple and complex joins operations in both Cloud and traditional database
2. Repeat the above task 30 times and take the average value of the response time
3. Repeat step1 and 2 in four tables of the Cloud and traditional databases

There are other DML statements INSERT, UPDATE and DELETE. Only SELECT statement is chosen to test in the experimentation as the SELECT statement is used to retrieve the data and used in most operations in the organizations. As a first step of research, the SELECT statement is evaluated. Tabulate the above results. At the end a comparative study is done and conclusions are drawn in the user point of view. Hardware Specifications for traditional database work station:

- RAM: 4 GB
- Hard Disk: 500GB
- Processor: Intel core I5

3.2.2 Constructing a test bed

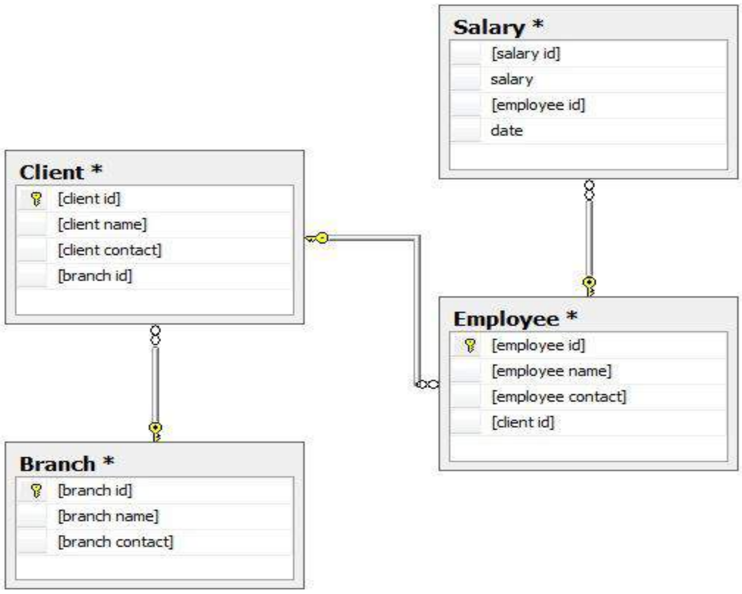

Figure 3-2 Database schema of EMPLOYEE Database

Table 3-6 Entities and attributes in Employee database

Entity	Attributes
Client table	Client ID Client name Client contact Branch ID
Employee table	Employee ID Employee name Employee contact Client ID
Salary table	Salary ID Salary amount Employee ID Date
Branch table	Branch ID Branch name Branch contact

In order to build better platform for performance testing, afore mentioned database was created with suitable relationships among the tables available in order to avoid redundant data we have also used simple and complex join queries while testing database performance.

3.2.3 Database Normalization

Database normalization is a way to produce good relationship between the fields by minimizing redundancy and dependency among data in the database. Normalization aims at isolation of data so that inserting, updating, and deleting the data can be made in just one table and then propagated through the rest of the database via predefined relationships. The goal of this technique is creation of tables with minimal amount of redundant data while preserving consistency. In normalization, each row should be unique and eliminate the duplicate columns in the same table of the database. Set the primary keys for the columns and foreign keys to the tables establishing the relationship between the tables because of the logical order in the storage of data. With this procedure, query execution and data retrieval will not take much time thereby resulting in better performance.

3.2.3.1 Relationships Among tables

Table 3-7 Entity relationship and keys information

Table	Primary Key	Foreign Key
Branch	Branch ID	----
Clients	Client ID	Branch ID
Employee	Employee ID	Client ID
Salary	Salary ID	Employee ID

3.3 Cloud Database

In order to test Cloud Database performance Windows Azure is used as platform. Following are the reasons to select Windows Azure as Cloud platform.

- Windows Azure also uses SQL similar to Microsoft SQL 2008 R2, the traditional database employed
- Windows Azure provides user friendly interface to develop database as shown in Appendix B
- Because of using SQL Server as on-premises database, database migration to Cloud is an easy process with SQL migration wizard tool. Using this tool EMPLOYEE database is migrated to Windows Azure

The Windows Azure is accessed on a webpage via work station which is connected to the Internet (BTH environment). There are no specific cache settings in SQL Server 2008 R2 and Windows Azure.

4 RESULTS

4.1 SLR Results

This section discusses the results and analysis of the papers that are extracted in the SLR process. The relevant articles are found, addressing the issues that affect the performance of Cloud Database. There are 5 papers about the topic which are relevant to meet the goals of the research.

The systematic literature review has yielded 5 results. Detailed descriptions of the list of identified issues which affect the performance of Cloud Database are given below.

Table 4-1 SLR Results

S No.	Ref. No.	Issue	Description
1.	[15]	Data Acquisition	This can be time consuming as copying data to clusters or nodes in Cloud Database can impact performance.
2.	[16]	Parallelism	With huge databases, especially Cloud Databases, the sequential processing paradigm will not cope. Thus parallelism determines the performance in huge databases.
3.	[17]	Data Management	The opportunities for parallelization and distribution of data in Clouds make storage and retrieval processes very complex, especially in facing with real-time data processing thereby affecting the performance.
4.	[18], [19]	Data mining in large databases	Data mining with many-task issue in large databases degrades the performance of a Cloud Database [18]. Growth of the size of database or the decrease of the minimum support increases the memory requirement and execution time thereby affecting the performance of database [19].

4.2 Experimental Results

In order to test the performance of on-premises and Cloud Databases, query response time was taken as a measurement across Data Manipulation Language Statements (SELECT) with different conditions. Each statement was iterated at least 30 times and for every attempt query response time was noted and finally average was calculated for all iterations. All the SQL queries were executed using EMPLOYEE database in SQL Server 2008 R2 and in Windows Azure. Running the SELECT statement results in the retrieval of data and the number of results fetched in each case is tabulated along with the response time values.

The “Slow Down” curve was drawn with the help of obtained response time values. It is obtained by dividing all the entry response times with the initial entry response time. The response time values of all the data sizes (30,000, 60,000, 120,000, 240,000 and 480,000 entries) in traditional database are divided by the initial entry response time of the traditional database i.e. 30,000 entries. Graph is obtained by the values. The same procedure is repeated for the Cloud Database response time values. Graph is drawn with the values and both the curves are plotted. These curves show the ‘Slow Down’ as a comparison between the two.

4.2.1 QUERY 1

The main aim this exercise is to find out query elapsed time for a query which retrieves small number of rows from large table, by scanning the complete table.

Command:

```
select EmployeeID, Date, Salary from Salary where Date = '01/02/2009' and EmployeeID>0 and EmployeeID<A;.
```

Above query retrieves data EmployeeID, Date and Salary columns for the date '01/02/2009' with EmployeeID range ‘0’ and A from Salary table. By executing the above query we end up retriving the data in between the EmployeeID 12000 to 30000. The table 4-2 shows the average query elapsed time for both traditional database(SQL Server 2008 R2) and Cloud Database(Windows Azure).

Table 4-2 Query 1 Response Time Values of different entries for Traditional and Cloud Database in milliseconds

Response time for	Retrieved results	Traditional Database (ms)	Cloud Database (ms)
30,000 entries	15	6	11
60,000 entries	15	7	9
120,000 entries	38	16	40
240,000 entries	89	20	74
480,000 entries	184	62	178

For convenience, the value is given as A in the query and the resemblance of A is tabulated as follows

Table 4-3 Data entries of the Query1

	30,000 entries	60,000 entries	120,000 entries	240,000 entries	480,000 entries
A	22500	45000	110000	220000	440000

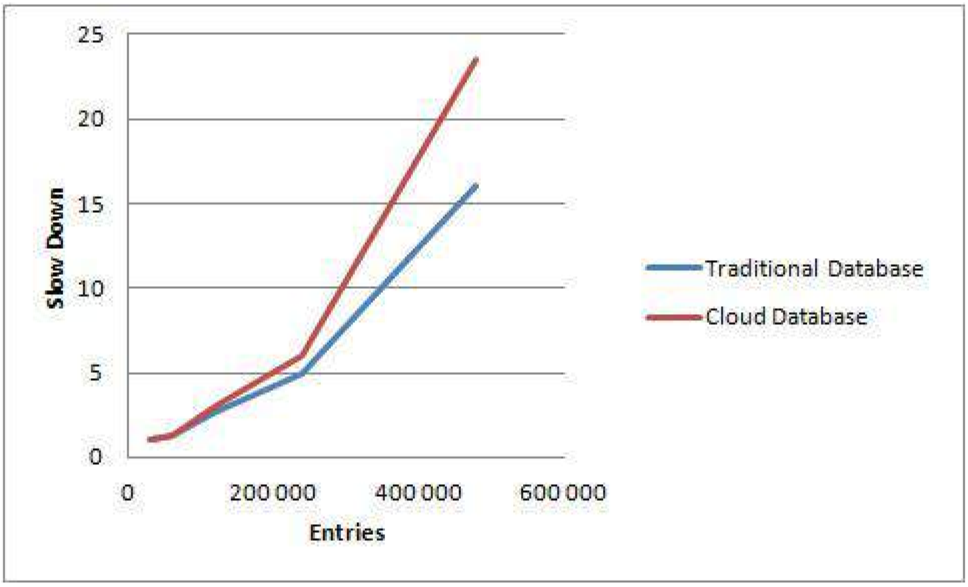

Figure 4-1 Slow Down Factor between Traditional and Cloud Databases for different entries for Query 1

From the above there is a drastic change between Cloud Database and traditional database performance while retrieving rows from tables. These results show that the Traditional Database is performing well for this query. At 30,000 the response time is almost doubled in Cloud. At 60,000 entries, both the databases have almost the same response time. At 120,000 entries, the Cloud has 2.5 times higher response time. At 240,000 entries, the Cloud has 3.5 times higher response time. At 480,000 entries, the Cloud Database is 2.9 times higher.

4.2.2 QUERY 2

In this query by using SELECT command we retrieve the data from a large table by scanning the complete table.

Command:

select EmployeeID, EmployeeName, EmployeeContact from Employee where EmployeeName > 'b%' and ClientID>0 and ClientID<A:

Above query retrieves data EmployeeID, EmployeeName and EmployeeContact columns for the EmployeeName > 'b% within ClientID range ‘0’ to A from a Employee table The task of above query is to pull out large number of rows from a single table, above query retrieves data from Client table in between ‘0’ and A. The average response time in Cloud Database and traditional database is shown in Figure 4-2.

Table 4-4 Query 2 Response Time Values of different entries for Traditional and Cloud Database in milliseconds

Response time for	Retrieved results	Traditional Database (ms)	Cloud Database (ms)
30,000 entries	25,667	310	1,546
60,000 entries	48,360	387	2,452
120,000 entries	95,981	739	4,996
240,000 entries	190,696	1421	9287
480,000 entries	380,755	2836	19056

For convenience, the value is given as A in the query and the resemblance of A is tabulated as follows.

Table 4-5 Data entries of the Query 2

	30,000 entries	60,000 entries	120,000 entries	240,000 entries	480,000 entries
A	750	1400	2800	5600	11200

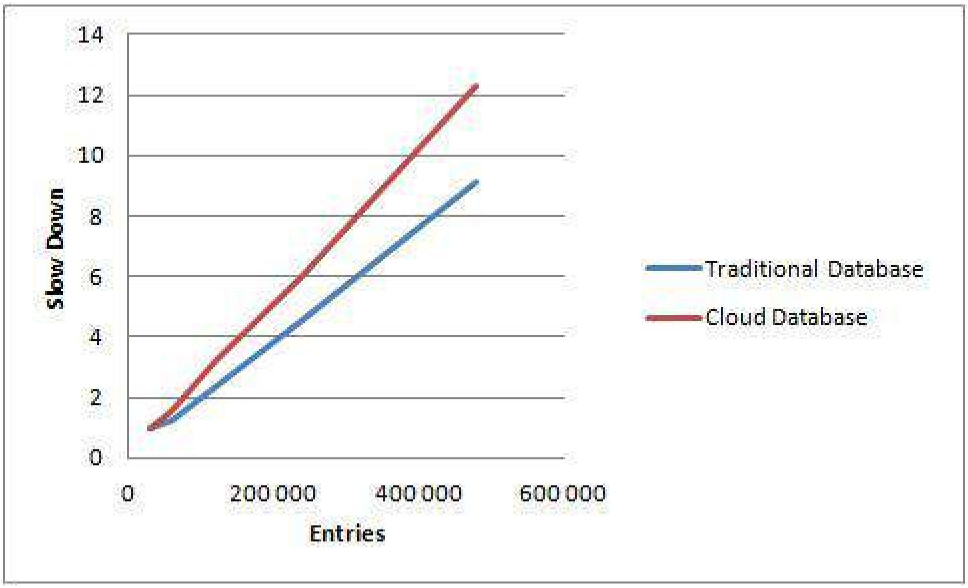

Figure 4-2 Slow Down Factor between Traditional and Cloud Databases for different entities for Query 2

These results show that the Traditional Database is performing well for this query. At 30,000 the response time is 5 times more in Cloud. At 60,000 entries, Cloud has 5.9 times higher response time. At 120,000 entries, the Cloud has 6.9 times higher response time. At 240,000 entries, the Cloud has 6.5 times higher response time. At 480,000 entries, the Cloud Database is 6.7 times higher.

4.2.3 QUERY 3 (SELECT COMMAND USING SIMPLE JOIN)

Test is carried out based on Employee and Salary table. By using simple join query we try to retrieve the data EmployeeName, EmployeeID from Employee table and Salary, Date from the Salary table.

Command:

```sql
set statistics time on
select e.EmployeeID, e.EmployeeName, s.Salary, S.Date from Employee e inner join Salary
s on e.EmployeeID = S.EmployeeID where EmployeeName > 'a%' and s.SalaryID> 0 and
S.SalaryID<A;
```

Above query retrieves the EmployeeID, EmployeeName, Salary and Date within the SalaryID range ‘0’ to A. The task of above query is to pull out large number of rows from the two tables. The average response time values in Cloud Database and traditional database are as shown in figure below.

Table 4-6 Query 3 Response Time Values of different entries for Traditional and Cloud Database in milliseconds

Response time for	Retrieved results	Traditional Database (ms)	Cloud Database (ms)
30,000 entries	24,499	324	1373
60,000 entries	48,999	465	1928
120,000 entries	97,999	885	4359
240,000 entries	195,999	1690	7777
480,000 entries	391,999	3235	15587

For convenience, the value is given as A in the query and the resemblance of A is tabulated as follows.

Table 4-7 Data entries of the query3

	30,000 entries	60,000 entries	120,000 entries	240,000 entries	480,000 entries
A	25000	50000	100000	20000	40000

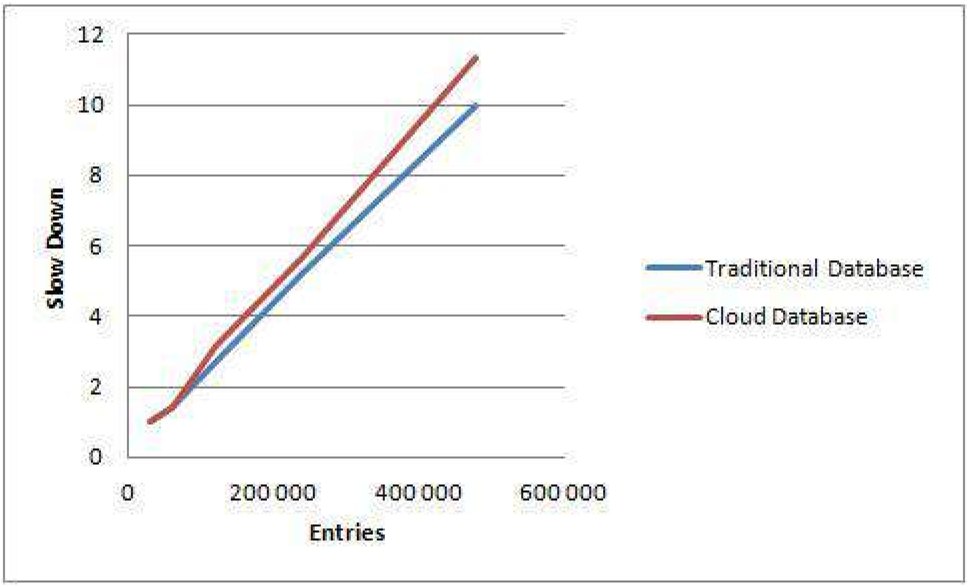

Figure 4-3 Slow down Factor between Traditional and Cloud Databases for different entities for Query 3

These results show that the Traditional Database is performing well. At 30,000 the response time is tripled in Cloud. At 60,000 entries, the response time 4 times higher in Cloud. At 120,000 entries, the Cloud has 5 times higher response time. At 240,000 entries, the Cloud has 4.6 times higher response time. At 480,000 entries, the Cloud's response time is 4.8 times higher.

4.2.4 QUERY 4 (SELECT COMMAND USING COMPLEX JOIN)

Test is carried out based on SELECT command that uses complex join to retrieve the data. Following query has been constructed to retrieve data by joining multiple tables with specific conditions.

Command:

```
select e.EmployeeName, e.EmployeeContact, c.ClientName, c.ClientContact,
b.branchName, b.branchContact, s.Salary from Employee as e join Client as c on
e.ClientID=C.ClientID join Branch as b on b.branchid=c.BranchID join Salary as s on
s.EmployeeID= e. EmployeeID where s.Salary>0 and s.Salary<A;
```

The above query retrieves the data BranchName, BranchContact from Branch table, ClientName, ClientContact from Client table, EmployeeName, EmployeeContact from Employee table, and Salary from Salary table by satisfying the range in between '0' and A.

Table 4-8 Query 4 Response Time Values of different entries for Traditional and Cloud Database in milliseconds

Response time for	Retrieved results	Traditional Database (ms)	Cloud Database (ms)
30,000 entries	8,204	204	1107
60,000 entries	32,430	1097	8921
120,000 entries	65,006	1590	31258
240,000 entries	130,479	3080	33973
480,000 entries	261,537	7083	77654

For convenience, the value is given as A in the query and the resemblance of A is tabulated as follows.

Table 4-9 Data entries of the Query 4

	30,000 entries	60,000 entries	120,000 entries	240,000 entries	480,000 entries
A	10000	20000	40000	80000	160000

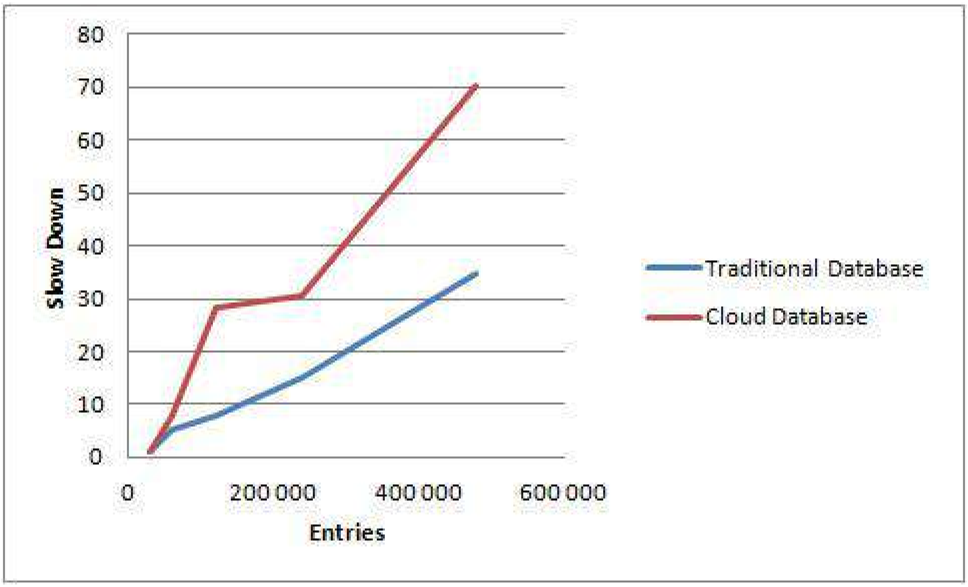

Figure 4-4 Slow Down Factor between Traditional and Cloud Databases for different entities for Query 4

These results show that the Traditional Database is performing well for this query. At 30,000 entries, the response time is 5.4 times higher in Cloud. At 60,000 entries, the response time is 8 times higher in Cloud. At 120,000 entries, the Cloud has 19 times higher response time. At 240,000 entries, the Cloud has 11 times higher response time. At 480,000 entries, the Cloud's response time is 10.9 times higher.

5 DISCUSSION

The goal of this research is to identify the previous research attempts on issues that affect the performance of a Cloud Database and compare the performance of a Cloud Database to that of a traditional database in terms of response time. Response time is considered as a metric to compare the performance of both the databases. In the research, SLR and the Quantitative methodology are followed to answer the RQ1 and RQ2 respectively. To answer the RQ1 for the SLR part, search strings are framed initially. Three databases are chosen for the extraction of the articles. Articles are selected using the search strings and the intrusion and exclusion criteria specified in Appendix A. With the obtained results specified in the Section 5.1, issues such as Data Acquisition, Parallelism, Data Management, Integrity of data storage, Data mining in large databases, Resource allocation and management, Database migration, Disaster recovery and Applications which affects the performance of Cloud Database are identified. To answer the RQ2, a quantitative methodology is followed. A relational database named Employee database is designed, normalized, optimized and deployed into the Cloud environment and traditional environment. The Employee database consists of four tables namely Branch, Client, Employee and Salary. The relational database is designed in such a way that it is normalized properly and the primary keys and foreign keys are set accordingly. The Microsoft SQL 2008 R2 and Microsoft Azure are chosen as Traditional and Cloud Databases respectively. By using the SELECT statement, queries are framed with the Simple and Complex Join techniques for the performance testing. Each query is executed in Traditional database and Cloud Database for 30,000 entries, 60,000 entries, 120,000 entries, 240,000 entries and 480,000 entries. Each query is repeated 30 times and response time values are noted. The average and standard deviation values are calculated and tabulated based on the response times. A Slow Down curve is drawn with the results.

The response time results and the curve shows that the Cloud Database performance is poor compared to that of the traditional database. As this research issue is a novel one relatively, a less amount of related work is done on the performance analysis of Cloud Database. From the results from Appendix F, Appendix G, Appendix H, Appendix I and Appendix J it is speculated that the traditional database has the better performance. Maintaining the same hardware configuration stays as a limitation for the research as the hardware configurations of the Cloud provider are undisclosed.

5.1 Validity Threats

A number of validity threats are identified in the research. These include the threats concerning the SLR and threats concerning experimentation. According to [21], any research may have four kinds of threats. They are:

- Construct Validity
- Internal Validity
- External Validity
- Conclusion Validity

5.1.1 Construct Validity

- *"Construct validity involves generalizing from your program or measures to the concept of your program"* [22]

As specified in the earlier sections, the articles are primarily extracted from Inspec, Scopus and ScienceDirect databases. From the published articles, the required articles are systematically reviewed and the issues that affect the performance of Cloud Database are identified. There is a threat that if this process could yield better results. In order to mitigate this type of threat, the guidelines provided by Kitchenham et al. [10] are used.

5.1.2 Internal Validity

- *"Internal validity is the approximate truth about inferences regarding cause-effect or causal relationships"* [23].

SLR: This threat occurs while extracting the articles related to the research. This is considered as a threat when a study is done on prior works, there is an anticipation that some of the issues may be missing during this process. In order to mitigate this type of threat, a systematic method [10] is followed. The formed search strings from the key words of the research question are verified in the discussions with the Supervisor. And a second discussion is done with the librarian. The articles are extracted from the scientific databases jointly by both the researchers. Based on the mutual understanding, with the use of inclusion and exclusion the articles are filtered. This even helped to mitigate and eliminate redundancy and inconsistency amongst the articles.

Experiment: This type of threat has a high impact on the experimentation. The data for the experimentation is collected from [20]. As the chosen Employee database is a relational database, the necessary primary and foreign keys are to be set properly in order to deploy into the Traditional Database and Cloud Database. If at all the primary keys and the foreign keys are not set properly, it would result in data insertion errors. Care is taken in order to set the keys while collecting and inserting the data. There are problems even while deploying the data into Cloud environment as both the researchers is new to the research. At every step, help is taken from the professionals and answers from stack overflow database forums helped to mitigate this threat.

5.1.3 External Validity

- *"External validity is the degree to which the conclusions in your study would hold for other persons in other places and at other times"* [24]

SLR: In SLR certain criteria is followed for the data extraction. They are 2005-2012

- Based on relevant key words
- Based on topic relevant title
- Based on abstract and introduction

This will make sure that the results are general and relevant for later research. But these might vary if the period chosen is other than 2005-2012 or if there is a change in the inclusion or exclusion criteria. To avoid this, the years prior to 2005 are verified but no results were found. The search is limited to 2012 which is tabulated in the Appendix A. In order to minimize this threat, the search is done multiple times and verified with the Supervisor at every step.

Experiment: As the Cloud is accessed via internet, this factor will have affect on the performance. In order to mitigate this threat, the experiment is repeated several number of times. This experimentation process is carried at BTH environment.

5.1.4 Conclusion Validity

- *"Conclusion validity refers to the statistically significant relationships between the treatment and outcome"* [25]

In order to reduce bias in this research, the inclusion and exclusion criteria are followed which separate out the irrelevant articles. Thus the threat is mitigated in the SLR. In the experimentation, as the research is new of this kind, discussions are conducted frequently as and when the results are obtained. Thus the threat is mitigated in the Experimentation.

6 CONCLUSIONS

6.1 Linking Research Questions

6.1.1 Research Question 1

The SLR results conclude that the Data Acquisition, Parallelism, Data Management, Integrity of data storage, Data mining in large databases, Resource allocation and management, Database migration, Disaster recovery and Applications which affect the performance of Cloud Database are identified. In total 4 issues were identified using SLR which are also having effect on the performance of the Cloud Database.

The lists with details can be found in the Table 4-1 SLR Results, while the description and analysis of the results of issues that affect the performance of a Cloud Database are discussed in results section.

6.1.2 Research Question 2

Apart from the advantages provided by the Cloud Database, it is important to consider the performance. To answer the RQ2, the relational Employee database is deployed with 5 different levels of entries into both Traditional and Cloud environments. The response time values are obtained and the slow down curves are drawn. It is observed that the performance of Cloud Database is poor compared to Traditional Database for all the four queries i.e., the slow down factor is larger in the Cloud when the size of the databases increases.

6.2 Future Work

The scope of this research is to give an introduction to the issues that are involved in the performance of a Cloud Database and a testing environment for the comparison of traditional and Cloud environments. The research is limited as the hardware configurations of the Cloud Database are undisclosed by the provider. Only SELECT operation of the DML statements is evaluated for now. In future, the other DML statement such as INSERT, UPDATE and DELETE can be evaluated. In future, the effort can be made to keep the hardware configurations same while comparing both the databases. And framework can be designed to overcome the issues identified in the SLR.

REFERENCES

[1] ”Windows Azure” [Online]. Available: http://www.microsoft.com/windowsazure/ [Accessed:12-March-2012]

[2] ”Microsoft Relational database components” [Online]. Available: http://msdn.microsoft.com/en-us/library/aa174501(v=sql.80).aspx/. [Accessed: March-2012].

[3] V. Matelan, D Cisic, and D Ogrizovic “Cloud Database-as-a-Service (DaasS-ROI)”, presented at the MIPRO, 2010 proceedings of the 33rd International Convention, May 24-28, pp. 1185-1188.

[4] Jia Liu and Lei Huang Ting” Dynamic Route Scheduling for Optimization of Cloud Database,” Presented at the Intelligent Computing and Integrated Systems (ICISS), Oct 22-24, 2010, pp.680-682.

[5] D.Petkovic “Performance Tuning, in Microsoft SQ.L Server 2008: A Beginner’s Guide”, 4 th ed. The McGraw-Hill Companies, pp. 517-525.

[6] ”Relational Database Components” [Online]. Available: http://msdn.microsoft.com/en-us/library/aa174501(v=sql.80).aspx [Accessed: February-2012].

[7] Pratt, P. J., & Adamski, J. J. (2008). CONCEPTS OF DATABASE MANAGEMENT. (pp. 29-34). GEX Publishing Services.

[8] ”IBM Informix Dynamic Server Performance Guide” Online. Available: http://publib.boulder.ibm.com/infocenter/idshelp/v10/index.jsp?topic=/com.ibm.perf.doc/perf43.htm [Accessed: February-2012]

[9] Unterkalmsteiner, M.; Gorschek, T.; Islam, A.; Cheng, C.; Permadi, R.; Feldt, R.; , "Evaluation and Measurement of Software Process Improvement—A Systematic Literature Review," *Software Engineering, IEEE Transactions on* , vol.PP, no.99, pp.1.

[10] Kitchenham, B.; Charters, S.;, "Guidelines for performing Systematic Literature Reviews in Software Engineering," Keele University and Durham University Joint Report EBSE 2007- 001, 2007.

[11] Meng-Ju Hsieh; Chao-Rui Chang; Li-Yung Ho; Jan-Jan Wu; Pangfeng Liu; , "SQLMR : A Scalable Database Management System for Cloud Computing," *Parallel Processing (ICPP), 2011 International Conference on* , vol., no., pp.315-324, 13-16 Sept. 2011.

[12] Zhang Jian-hua; Zhang Nan; , "Cloud Computing-based Data Storage and Disaster Recovery," *Future Computer Science and Education (ICFCSE), 2011 Int. Conf. on* , vol., no., pp.629-632, 20-21 Aug. 2011.

[13] Ying Hua Zhou; Qi Rong Wang; Zhi Hu Wang; Ning Wang; , "DB2MMT: A Massive Multi-tenant Database Platform for Cloud Computing," *e-Business Engineering (ICEBE), 2011 IEEE 8th International Conference on* , vol., no., pp.335-340, 19-21 Oct. 2011.

[14] W.J.Ting, D. Hui, L.M. Constance, “Accounting For The Benfits Of Database Normalization,” vol. 3, No1, June 2012.

[15] J. Baodong, “Performance Considerations of Data Acquisition in Hadoop System,” in Proc. Cloud Computing Tech. and Science, Indianapolis, Ind, pp. 545-549.

[16] D.Taniar, “High Performance Database Processing,” in Proc. Advanced Information

Networking and Application, Fukuoka, 2012, pp. 5-6.
[17] H.B. Amir, I.K. Asad, “Evolution of information retrieval in cloud computing by redesigning data management architecture from a scalable associative computing perspective,” in proc. 17th int. conf. on Neural information processing: model and applications, Berlin, 2010, pp.275-282.
[18] W.K. Lin, L.C. Yu, “Efficient strategies for many-task frequent pattern mining in Cloud Computing environments,” in proc. Systems Man and Cybernetics, Istanbul, 2012, pp.620-623.
[19] W.K. Lin, C.L. Pei, C.L. Weng ”A novel frequent pattern mining algorithm for every large databases in Cloud Computing environments,” in Proc. Granular Computing, Kaohsiung, 2011, pp.399-403.
[20] “Generate data” [Online]. Available: http://www.generatedata.com/#about [Accessed: March-2012].
[21] C. Wohlin, P. Runeson, M. Host, M. C. Ohlsson, B. Regnell, and A.Wesslén, Experimentation in software engineering: an introduction. Norwell, MA, USA: Kluwer Academic Publishers, 2000.
[22] “Construct Validity.” [Online]. Available: http://www.socialresearchmethods.net/kb/constval.php [Accessed: May-2012].
[23] “Internal Validity.” [Online]. Available: http://www.socialresearchmethods.net/kb/intval.php [Accessed: May-2012].
[24] “External Validity.” [Online]. Available: http://www.socialresearchmethods.net/kb/external.php [Accessed: May-2012].
[25] “Conclusion Validity.” [Online]. Available: http://www.socialresearchmethods.net/kb/concval.php [Accessed: May-2012].
[26] P. Mell. (2011) 'The NIST Definition of Cloud ', Reports on Computer Systems Technology, Sept., p. 7.
[27] H. Thomas, “’A veritable bucket of facts’ origins of the database management system,” in Proc. Medford, New Jersey, June 2006, pp.33-49.
[28] J.Liu,” Dynamic Route Schedule for Optimization of Cloud Database,” in Proc. ICISS Conf. 2010, pp.680-682.
[29] V.Mateljan,”Cloud Database-as-a-Service (Daas)-ROI,” in Proc. 33rd International Convection May 2010, pp.1185-1188.
[30] “Comparing Real World Database Performance” [Online]. Available: http://public.dhe.ibm.com/common/ssi/ecm/en/iml14276usen/IML14276USEN.PDF [Accessed: May-2012]
[31] K.Michael, A.Bernstein, M.L Philip “Database Systems” Pearson Education, Inc, 2006, pp.1-8.
[32] Ritchie, Colin,” Relational Database Principles,” Ashford Colour Press, 1988, pp. 6-40.
[33] A. Divyakant, D. Sudipto, A.Amr EI, “Big data and Cloud Computing: new wine or just new bottles?” Vol. 3, pp.1647-1648, Sep. 2010.
[34] D. Lex, M. Karen, G. Tim, J. Inger, F.Daniel (2009, Dec.),” Beginning Oracle SQL ,“ Apress, USA, 2009, pp.178-184.
[35] "Sqlserverbible” [Online]. Available: http://www.sqlserverbible.com/files/databasedesignroi.pdf [Accessed: 24-May-2012]
[36] Sosinsky, Barrie, “Cloud Computing Bible,” Wiley Publishing, Inc, Indianapolis, Indiana, 2011, pp.1-11.

[37] H. Lex de, F.Daniel, G. Tim, J. Inger, M. Karen,"Beginning Oracle SQL," Apress, United States of America, 2009, pp.178-184.

[38] “Amazon RDS Performance vs. Xeround Cloud Database: Benchmark Results” [Online]. Availabe: http://xeround.com/cloud-database-comparison/xeround-vs-amazon-rds-benchmark/ [Accessed : May-2012]

APPENDIX A

Table Appendix A Search Query used for RQ1

DATABASE	SEARCH QUERY USED FOR RQ1
Inspec	((((("Cloud Computing") OR (Cloud)) AND (Database) AND (Performance) AND ((Issues) OR (Problems)))) WN ALL) +(2012 OR 2011 OR 2010 OR 2009 OR 2008 OR 2007 OR 2005) WN YR
Scopus	(((("Cloud Computing") OR (Cloud)) AND (database) AND (performance) AND ((issues) OR (problems)))) AND (LIMIT-TO(PUBYEAR, 2012) OR LIMIT-TO(PUBYEAR, 2011) OR LIMIT-TO(PUBYEAR, 2010) OR LIMIT-TO(PUBYEAR, 2009) OR LIMIT-TO(PUBYEAR, 2008) OR LIMIT-TO(PUBYEAR, 2007) OR LIMIT-TO(PUBYEAR, 2006)) AND (LIMIT-TO(SUBJAREA, "COMP") OR LIMIT-TO(SUBJAREA, "MULT")) AND (LIMIT-TO(LANGUAGE, "English"))
ScienceDirect	Selected “Computer Science” and searched with the query ((("Cloud Computing") OR (Cloud)) AND (Database) AND (Performance) AND ((Issues) OR (Problems))) and limited to Computer Science, Cloud Computing, Clouds, 2005, 2006,2007, 2008,2009, 2010, 2011, 2012, English

APPENDIX B

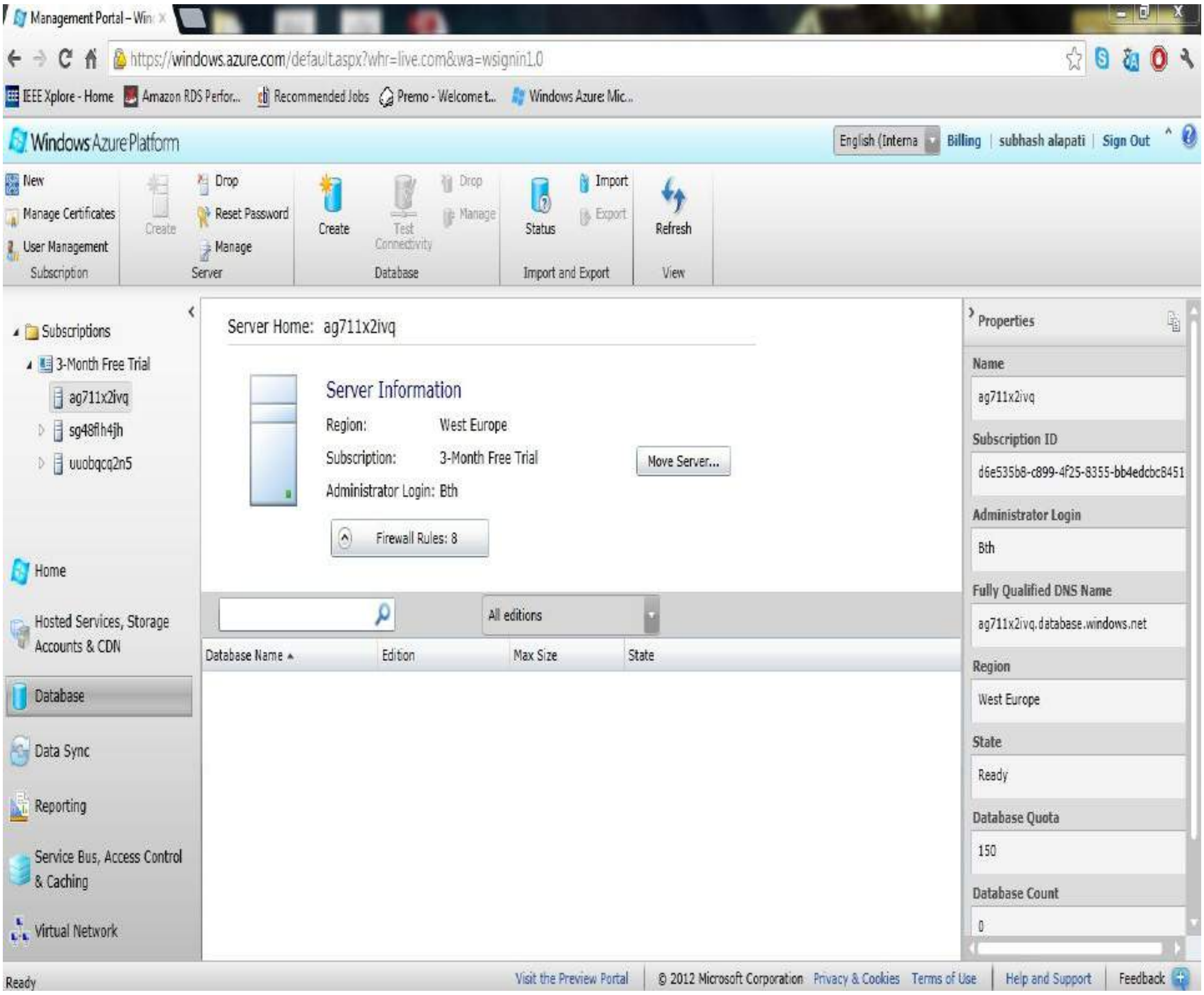

Figure Appendix B Microsoft Windows Azure Platform

APPENDIX C

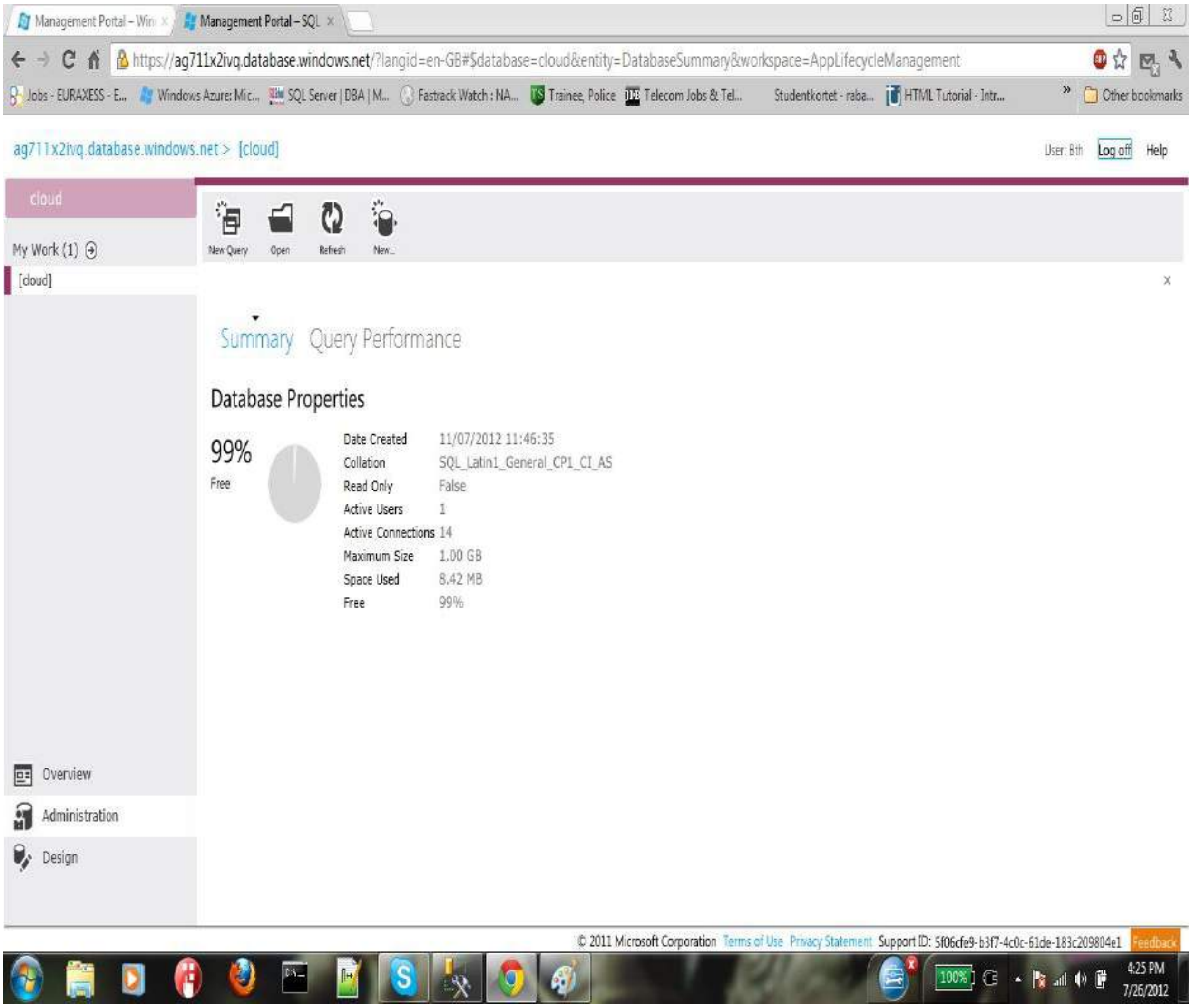

Figure Appendix C Cloud Database properties

APPENDIX D

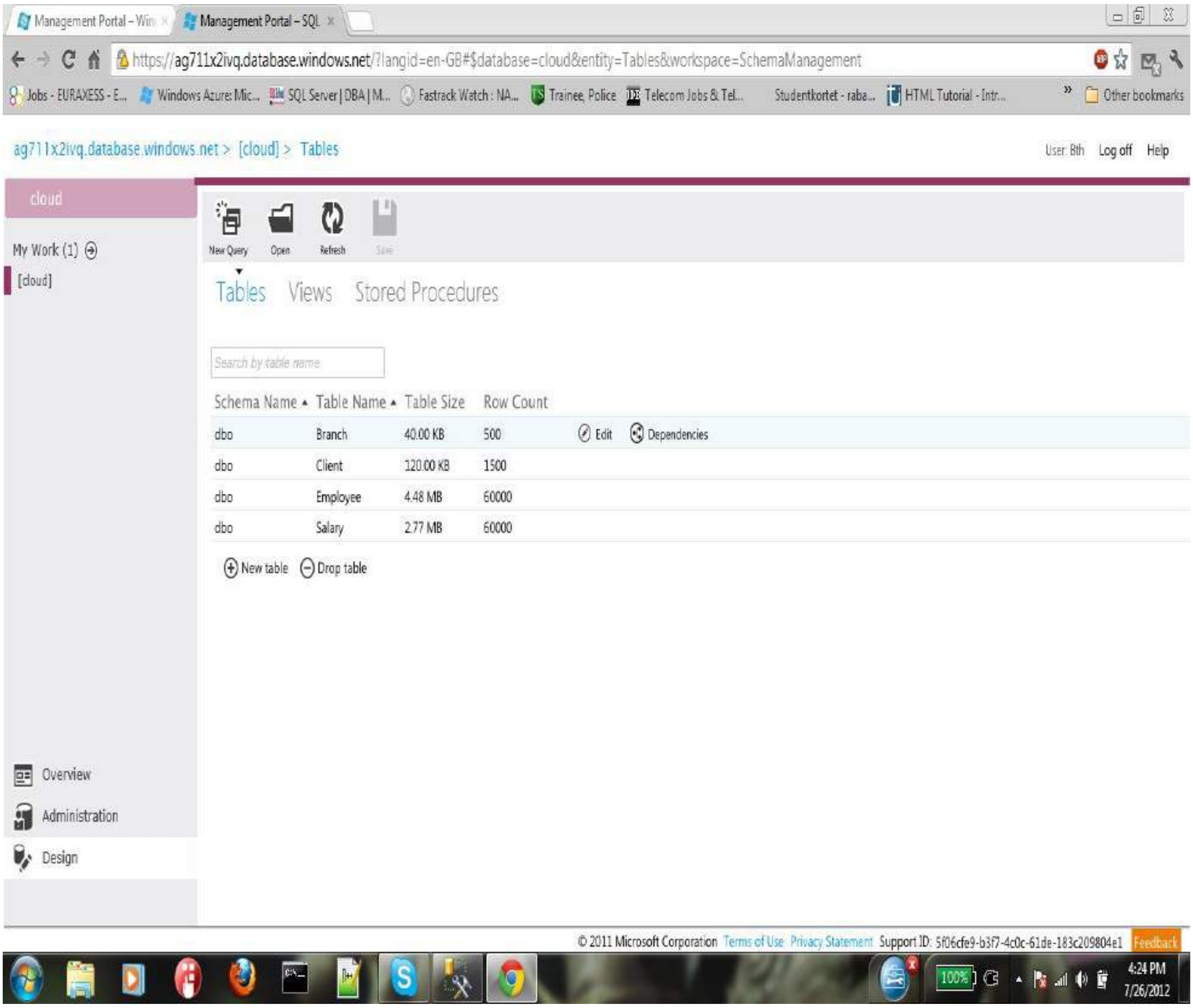

Figure Appendix D Employee database table sizes

APPENDIX E

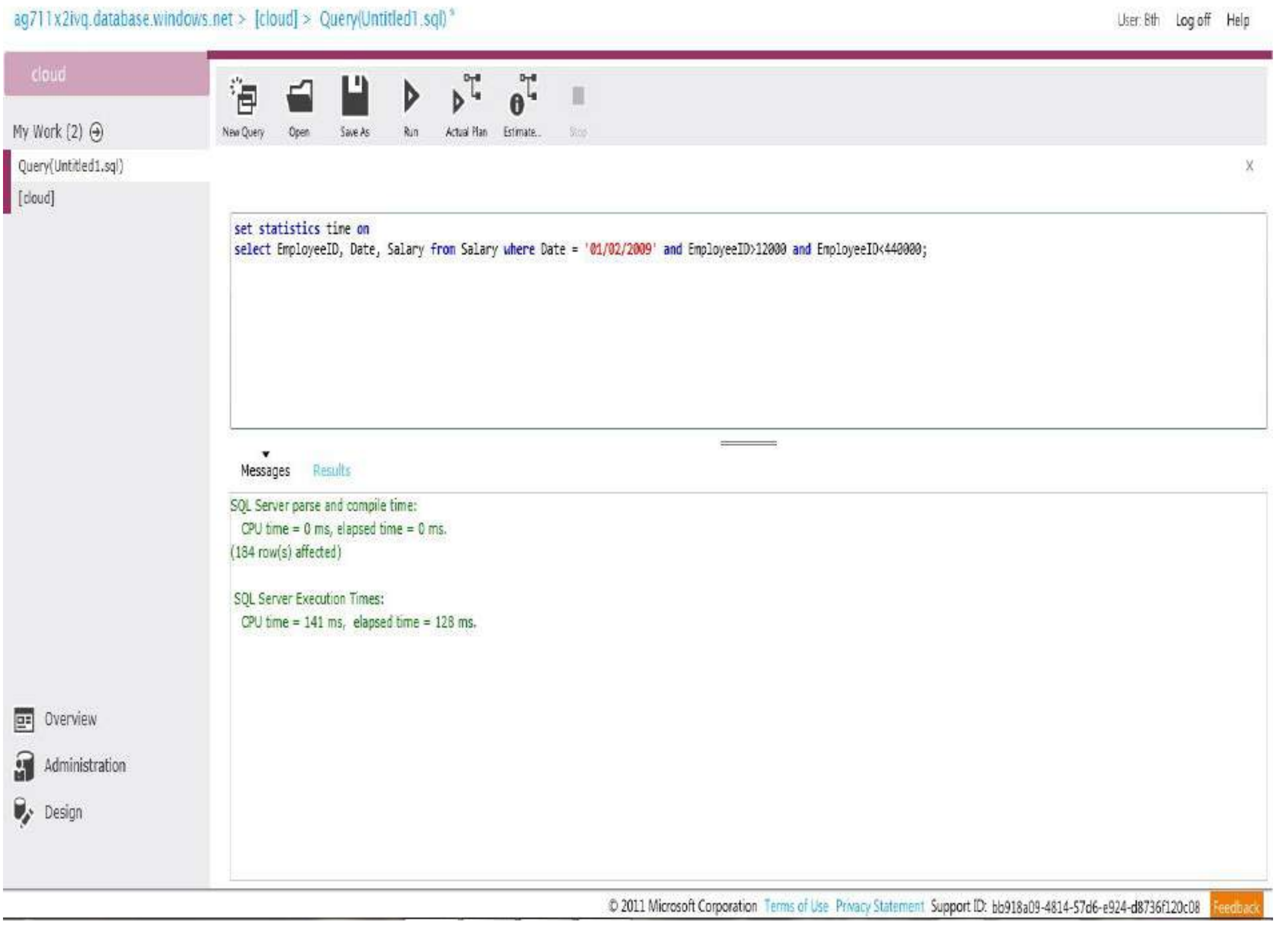

Figure Appendix E Query executions Windows Azure platform

APPENDIX F

Table Appendix F Query results for 30,000 entries

Query	Query 1		Query 2		Query 3		Query 4	
Database	Traditional	Cloud	Traditional	Cloud	Traditional	Cloud	Traditional	Cloud
	6	11	311	997	272	872	191	1529
	6	11	238	969	298	911	218	1695
	6	11	369	1041	261	867	269	1416
	6	11	317	1050	342	904	205	1394
	4	11	303	1018	269	822	167	1532
	7	11	293	1258	424	822	173	1591
	7	16	345	1368	255	965	190	1583
	7	12	257	1191	372	823	179	1366
	5	11	333	1004	260	828	205	1431
	9	11	424	1018	264	865	285	1439
	7	11	410	1052	241	813	215	1750
	6	11	247	952	265	847	179	1346
	3	11	338	1009	393	951	232	1675
	12	11	339	970	317	840	216	1658
	6	10	240	1034	310	807	182	1383
	6	11	342	1111	343	760	166	1641
	6	11	413	1021	302	944	181	1386
	6	10	348	1056	354	958	238	1471
	6	11	374	999	309	926	252	1639
	6	11	396	1080	413	852	173	1332
	5	10	335	1348	328	896	179	2705
	6	10	363	1084	352	789	230	2633
	7	11	315	980	323	805	191	1640
	6	10	454	1231	331	753	182	1785
	6	11	378	1213	294	849	196	1494
	7	10	312	1325	309	1007	203	1344
	6	10	390	1224	324	1123	224	1294
	7	10	417	1121	363	919	190	1439
	4	11	380	1011	347	796	223	1585
	4	10	362	968	333	1356	187	1460
Average	6.166667	10.9	344.7667	1090.1	318.9333	889	204.03333	1587.867
Std. Dev.	1.59921	1.09387	55.61238	121.64	47.00032	118.334	30.012047	322.7481

APPENDIX G

Table Appendix G Query results for 60,000 entries

Query	Query 1		Query 2		Query 3		Query 4	
Database	Traditional	Cloud	Traditional	Cloud	Traditional	Cloud	Traditional	Cloud
	5	8	340	1639	491	1410	1120	18411
	9	8	396	1701	498	1386	1083	16928
	6	8	362	1611	484	1416	1110	17292
	9	9	356	1667	548	1458	1085	18861
	5	8	350	1691	607	1429	1103	16401
	7	8	382	1681	509	1346	1113	16779
	5	8	372	1645	578	1390	1151	16156
	7	8	383	1629	506	2371	1001	17685
	5	8	356	2264	559	1450	1152	16791
	5	8	355	1758	454	1364	1165	16944
	9	8	403	1775	521	1411	1125	19357
	5	8	362	1703	459	1390	1119	18619
	5	8	332	1582	481	1380	994	19816
	9	8	350	1643	464	1399	1078	16171
	8	16	423	1629	447	1486	1203	18095
	4	8	375	1678	465	1392	1114	17733
	11	8	360	1653	459	1307	923	17816
	10	8	392	1779	551	1444	1312	17805
	6	16	358	1639	437	1470	1096	16112
	6	8	353	1613	591	1429	1179	17356
	4	9	354	1544	527	1442	1248	18284
	5	9	378	1638	458	1378	1019	18669
	9	8	365	1651	534	1343	1050	16144
	5	8	378	2501	452	1416	888	16846
	10	8	385	2125	459	1426	1058	16666
	11	8	346	1641	509	1461	911	16820
	5	17	357	1569	445	1398	1246	17199
	4	8	335	1705	479	1408	1129	16388
	5	8	364	1671	465	1378	1020	18397
	6	8	348	1616	535	1361	1103	17054
Average	6.6666667	8.9333	365.666667	1721.37	499.066667	1438	1096.6	17453.17
Std. Dev.	2.2180037	2.5316	20.6637002	208.314	46.9195447	180.78	96.1638688	1008.283

APPENDIX H

Table Appendix H Query results for 120,000 entries

Query	Query 1		Query 2		Query 3		Query 4	
Database	Traditional	Cloud	Traditional	Cloud	Traditional	Cloud	Traditional	Cloud
	13	31	755	3595	825	2905	1559	31300
	13	30	731	3232	841	2686	1709	28319
	13	31	696	3268	971	2975	1545	29477
	20	33	797	3203	833	2590	1523	28111
	13	31	748	3321	814	2697	1527	28568
	14	70	745	3444	845	2774	1518	31082
	18	53	816	3203	856	3639	1739	33375
	24	32	776	3182	866	3349	1562	32029
	24	33	706	3322	850	2785	1646	31140
	23	31	805	4598	854	3153	1828	30522
	13	34	810	4850	853	2666	1485	31701
	15	84	770	3192	829	2787	1713	31057
	22	39	775	3448	862	2658	1583	31233
	23	31	757	4272	823	2626	1537	33985
	13	31	744	3096	876	2667	1445	29261
	12	99	879	3629	841	2685	1546	32836
	17	76	853	4307	862	3763	1748	33154
	23	31	789	5436	951	3203	1774	32222
	19	36	770	4393	866	3660	1502	31988
	13	51	730	4429	817	3862	1581	31952
	13	34	682	3764	830	4409	1593	32362
	14	31	783	3841	858	2802	1627	31276
	13	33	753	3109	971	2800	1793	29127
	18	31	715	3310	823	2787	1545	29107
	23	31	774	3177	846	2614	1497	31904
	13	32	711	3228	818	2704	1528	33805
	18	31	831	3350	870	2676	1495	30333
	13	31	889	3227	839	3833	1482	30011
	17	31	684	2983	884	2619	1462	33263
	13	31	793	3063	921	2587	1603	33244
Average	16.66667	40.1	768.9	3615.733	859.8333	2998.7	1589.833	31258.13
Std. Dev.	4.229073	18.10001	52.59759	619.9876	42.23341	487.9902	106.1096	1681.244

APPENDIX I

Table Appendix I Query results for 240,000 entries

Query	Query 1		Query 2		Query 3		Query 4	
Database	Traditional	Cloud	Traditional	Cloud	Traditional	Cloud	Traditional	Cloud
	17	75	1286	5714	1590	5148	3150	46514
	22	74	1262	5594	1532	6426	3118	53936
	19	76	1342	7995	1607	5162	2970	55854
	21	78	1292	6042	1591	6142	3065	56148
	19	74	1395	5707	1682	5934	3142	51935
	23	75	1434	5707	1743	5112	3021	53550
	17	73	1295	7351	1699	5274	3183	52457
	21	74	1302	5611	1922	5154	3203	46735
	21	74	1347	5801	1743	5308	2988	51873
	18	78	1292	5671	1750	4943	3332	53359
	23	73	1624	5746	1930	6197	3001	55120
	18	74	1318	7954	1740	5146	3004	55826
	20	73	1297	5630	1799	5306	3249	55470
	21	73	1315	5714	1723	5062	3220	54198
	20	75	1313	5922	1645	4987	3311	55968
	20	73	1433	6407	1712	5956	2976	60122
	20	73	1299	5662	1714	5030	3275	63548
	19	73	1392	5753	1666	5152	3130	59036
	21	73	1302	7409	1579	5073	3156	59393
	20	72	1299	5696	1608	5168	2936	52944
	19	73	1327	6070	1564	5157	2987	60636
	23	73	1307	5525	1624	5349	2724	56403
	21	72	1294	5691	1618	5154	2963	63015
	21	73	1275	5885	1645	5114	3025	57967
	21	73	1332	5711	1568	6703	3095	60521
	25	73	1321	5770	1707	5137	2827	62827
	19	75	1418	5840	1622	6092	3196	49864
	19	72	1300	6327	1626	6139	3064	60686
	20	80	1295	5775	1595	5552	2993	63498
	22	74	1311	6635	1613	5145	3116	63722
Average	20.3333333	74.03	1333.96667	6077	1671.9	5440.7	3080.6667	56438
Std. Dev.	1.82574186	1.866	71.1906054	693.2	95.9663555	494.14	138.21381	4767.3

APPENDIX J

Table Appendix J Query results for 480,000 entries

Database	Traditional	Cloud	Traditional	Cloud	Traditional	Cloud	Traditional	Cloud
	53	164	3197	19375	3215	13481	7153	110874
	53	160	3206	16912	3219	12061	6892	107662
	50	159	2812	13334	3327	11494	6827	103032
	55	161	3428	15053	3129	14275	6702	103550
	71	160	2626	12789	3133	13191	6938	102679
	52	160	2614	15326	3194	13245	6607	110820
	57	163	2689	15289	3175	12762	6920	112020
	71	162	2639	14318	3127	18317	6490	103543
	59	159	2685	13132	3188	12843	6510	106587
	65	159	2705	17730	3150	11722	7382	108038
	57	164	3287	18784	3353	17191	7873	104584
	73	162	3140	14491	3197	13617	6815	105147
	74	202	2592	18246	3230	15212	7203	103887
	60	159	2750	16484	3113	12834	6744	103244
	46	164	2727	14432	3131	15730	7766	104868
	72	161	2733	13505	3159	12283	7117	101854
	60	205	2738	15970	3223	11764	8409	107734
	64	162	2593	23032	3249	12744	7366	104583
	64	161	2853	14304	3186	14071	8020	102726
	69	288	2597	14875	3231	12024	6921	111551
	66	162	3247	15625	3187	14283	6667	101576
	60	290	2659	14935	3156	11279	6687	102583
	70	179	2716	13391	3139	13146	8586	106740
	65	270	2669	13360	3276	12537	7018	110417
	74	158	2678	21594	3173	11851	7032	110866
	68	196	2766	17808	3114	14626	7063	103426
	74	164	2580	12824	3139	13807	6906	110244
	60	164	2816	14206	3181	12537	6807	106184
	56	159	3304	14789	3126	12193	6298	105417
	63	160	2643	14963	3197	13041	6794	100987
Average	62.7	177.9	2822.96667	15696	3187.23333	13338.7	7083.7667	105914.1
Std. Dev.	7.909619	37.754	256.826989	2531.3	59.2188713	1624.16	548.80545	3358.366