PERFORMANCE INDICATORS TO CONSIDER WHILE SELECTING CLOUD DATABASE SERVICE PROVIDER

ABBAS HUSSEIN ALI

DECEMBER 2014

PERFORMANCE INDICATORS TO CONSIDER WHILE SELECTING CLOUD DATABASE SERVICE PROVIDER

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES OF
ÇANKAYA UNIVERSITY

BY
ABBAS HUSSEIN ALI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF
MATHEMATICS AND COMPUTER SCIENCE / INFORMATION TECHNOLOGY PROGRAM

DECEMBER 2014

Title of the Thesis: **Performance Indicators to Consider While Selecting Cloud Database Service Provider.**

Submitted by **Abbas Hussein ALI**

Approval of the Graduate School of Natural and Applied Sciences, Çankaya University.

Prof. Dr. Taner ALTUNOK
Director

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

Prof. Dr. Billur KAYMAK ÇALAN
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Dr. Sinan Can AÇAN
Co-Supervisor

Assist. Prof. Dr. Özgür Tolga PUSATLI
Supervisor

Examination Date: 19.12.2014

Examining Committee Members

Assist. Prof. Dr. Özgür Tolga PUSATLI	(Çankaya University)
Assist. Prof. Dr. Abdül Kadir GÖRÜR	(Çankaya University)
Tuncay KÜÇÜKPEHLİVAN	(Başarsoft)

STATEMENT OF NON-PLAGIARISM PAGE

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name	:	Abbas Hussein, ALI
Signature	:	
Date	:	19.12.2014

ABSTRACT

PERFORMANCE INDICATORS TO CONSIDER WHILE SELECTING CLOUD DATABASE SERVICE PROVIDER

ALI, Abbas Hussein
M.Sc., Department of Mathematics and Computer Science / Information Technology Program
Supervisor: Assist. Prof. Dr. Özgür Tolga PUSATLI
Co-Supervisor: Dr. Sinan Can AÇAN

December 2014, 40 pages

Cloud computing is a new technology that provides software and hardware resources depending on customer specifications. A cloud database is one of the services that a cloud provider may offer. Many service providers are available on the market; however, selecting a qualified provider is an issue for customers. Performance indicators can be used to represent, measure, and compare the quality of service providers, which leads to creating a broad understanding among cloud service providers. This thesis proposes performance indicators to evaluate cloud database service providers. The first general attributes mostly related to software as service depending on the Service Measurement Index (SMI), then cloud database attributes are introduced. As a limitation of the scope of this thesis, the performance indicators have not been tested in an organization using cloud database service. The performance indicators are suggested to help customers while selecting suitable service provider to their requirements.

Keywords: Cloud Computing, Service Provider, Performance Indicators.

ÖZ

BULUTTA VERİTABANI HİZMET SAĞLAYICI SEÇİMİNDE DÜŞÜNÜLMESİ GEREKEN PERFORMANS GÖSTERGELERİ

ALI, Abbas Hussein
Yüksek Lisans, Matematik-Bilgisayar Anabilim Dalı / Bilgi Teknolojileri Bölümü
Tez Danışman: Yrd. Doç Dr. Özgür Tolga PUSATLI
Eş Danışman: Dr. Sinan Can Açan

Aralık 2014, 40 sayfa

Bulut bilişim, müşterilerin ihtiyaçlarına göre yazılım ve donanım kaynaklarını sağlayan yeni bir iş modelidir. Bulut veritabanı bulut servis sağlayıcısı tarafından sunabilir hizmetlerden biridir. Var olan birçok bulut servis sağlayıcı arasından nitelikli sağlayıcı seçmek müşteriler için bir sorun teşkil eder.

Bulut servis sağlayıcılarının kalitesini ölçmek, değerlendirmek ve karşılaştırmak için performans göstergeleri kullanılabilir. Bu tezde, bulut veritabanı sağlayıcısı değerlendirmek için kullanılabilecek performans göstergeleri önermektedir. Öncelikle Servis Ölçüm Endeksindeki (SÖE) Servis-Olarak-Yazılımla ilişkili özellikler tarif edilmiş, sonra da bulut veritabanı özellikleri eklenmiştir. Bu tezde, performans göstergeleri, bulut veritabanı servisini kullanan bir kuruluşta test edilmemiştir. Performans göstergelerinin organizasyonların ihtiyaçlarına uygun bulut servis sağlayıcıları seçilmesinde yardımcı olacağı düşünülmektedir.

Anahtar Sözcükler: Bulut Bilişim, Servis Sağlayıcı, Performans Göstergeleri

ACKNOWLEDGMENT

I would like to thank my supervisors and co-supervisor, assistant professor. Dr. Özgür Tolga PUSATLI and Dr. Sinan Can AÇAN, for their encouragement. I also thank my family for their support, especially my mother during this tough time, and I appreciate their patience, understanding and helping hand.

Thank you all.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES

LIST OF TABLES

TABLES

LIST OF ABBREVIATIONS

3G	Third Generation
4G	Fourth Generation
ACID	Atomicity, Consistency, Isolation, Durability
API	Application Program Interface
App	Application
AWS	Amazon Web Service
BPaaS	Business Process as a Service
CDBMS	Cloud Database Management System
CPU	Central Processing Unit
CRDBMS	Cloud Relational Database Management System
CSMIC	Cloud Service Measurement Index Consortium
DbaaS	Database as Service
DBMS	Database Management System
DOS	Denial of Service
FISAM	Federal Information Security Management Act of 2002
GB	gigabyte/gigabytes
GLBA	Gramm-Leach-Bliley Act
GUI	Graphical User Interface
HIPAA	Health Insurance Portability and Accountability Act
HTTP	Hypertext Transfer Protocol
I/O	Input / Output
IaaS	Infrastructure as Service
IBM	International Business Machines
IEC	International Electro technical Commission
IEEE	Institute of Electrical and Electronics Engineers
ISO	International Organization for Standardization
IT	Information Technology

JDBC	Java Database Connectivity
KPI	Key Performance Indicator
NIST	National Institute of Standard Technology
NoSQL	No Structured Query Language
ODBC	Open Database Connectivity
PaaS	Platform as Service
PC	Personal Computer
RDBMS	Relational Database Management System
RDS	Relational Database Service
REST	Representational State Transfer
S3	Simple Storage Service
SaaS	Software as Service
SAS	Statistical Analysis System
SLA	Service Level Agreement
SMI	Service Measurement Index
SOA	Service Oriented Architecture
SOAP	Simple Object Access protocol
SPI	Software, Platform and Infrastructure
SQL	Structured Query Language
TB	terabyte
TPC	Transaction Processing Performance Council
VM	Virtual Machine
Wi-Fi	Wireless Fidelity

CHAPTER 1

INTRODUCTION

1.1 Background

Cloud computing is an emerging service delivery method in Information Technology. It enables the on-demand request of IT requirements, including greater computation power and more software services.

A cloud computing model means centralizing all software and data hosted on a virtual server or a pool of virtual servers, and accessing the servers through the Internet without the need for numerous items of local hardware, such as memory or processor power. This allows the use of light-weight client computers in the end user side to access resources. In some cases, the user just has a device with a minimal OS and a web browser that enable to access the computing resources from anywhere with an Internet connection. This is what makes a cloud service flexible. Furthermore, the resources may be allocated according to the needs of an organization, such as the needs of a small company in the first month of business differing from their needs one year later. This is what gives cloud computing cost efficiency. Because of this efficiency, many more organizations are opting to use cloud computing services. Nevertheless, many questions arise from such choices. One of these questions is: *Which service provider will be most suitable for the organization?*

1.2 Objective

The primary aim of this study is to address performance indicators in order to assist small and medium-sized organizations to select a suitable cloud database service provider. The proposed performance indicator takes the Software Measurement Index (SMI) as a basis.

1.3 Research Problem

Selecting cloud computing service providers is a real problem for organizations. The selection process must include all requirements for the organizations and cover the most important characteristics for the required services. The increase in the number of cloud computing service providers presents the user with the daunting problem of selecting service provider services.

In Chapter Two 2.16.1, the literature review shows that there are many performance indicators for evaluating and comparing cloud service providers in software services in general. However, they are not necessarily showing those for the cloud database service known as Database-as-Service (DbaaS), which provides Database Management System (DBMS) functionalities to consumers.

The research question for this thesis is:

How can an organization select a suitable cloud database service provider?

1.4 Thesis Organization

This thesis contains four chapters. Fundamental information on cloud computing and cloud databases is found in the background chapter.

Chapter One provides the objective and the research problem. Chapter Two includes an introduction to cloud computing technology, its roots, the service and deployment models, benefits, issues, cloud database structure and advantages, Service Level Agreement (SLA) important points, encryptions and scheme consideration and related work. In Chapter Three, performance indicators are proposed and discussed. In Chapter Four, findings, limitations, future work, recommendations and conclusion are given.

CHAPTER 2

LITERATURE REVIEW

2.1 Cloud Computing Technology

There existed in the 1960s centralized computing services that provided computing services over a network by using mainframe time-sharing technology.

This mainframe time-sharing technology provided acceptable performance to users by utilizing computing resources; however, because of high hardware costs, mainframes had become difficult to scale and develop. Furthermore, the performance of applications that run on the mainframe was out of the user's control because it depended on the number of users utilizing the mainframe at any given moment. After the technology jump in the semiconductor industry, personal computers became reasonably priced. With the broad spread of the personal computer, business was liberated from using the mainframe technology. However, at the same time a new problem arose: how to share the data?

For the above question, the client-server model was the answer that provided centralized data-sharing which gave management and processing challenges. The initially simple client-server model became more complex because of the needs of business for growth. Therefore, the cost of complexity, maintenance and management of IT infrastructure was high. For many organizations, the dream was to focus full time on the essential business idea instead of wasting time with information technology issues. Although the impact of the cloud computing model is not touchable worldwide, many companies believed that cloud computing may present a model that may promote operational efficiency while minimizing costs and complexity [1].

2.2 Cloud Computing Technologies

We can follow the emerge of cloud computing by studying the progress of number of technologies, mainly in hardware, such as virtualization technique, multi-core processor, Internet technologies such as Web 2.0 services, service oriented architectures, and distributed computing such as clusters, grids and systems management technologies. Figure 1 shows the correlation of technologies that safely developed and interacted in the existing cloud computing model. Some of these technologies did not receive enough interest in their start up steps of development; however, the technologies received serious attention later from academia social and by industry companies' communities. Then, a properties and standardization procedure followed, leading to touch the advantages [2].

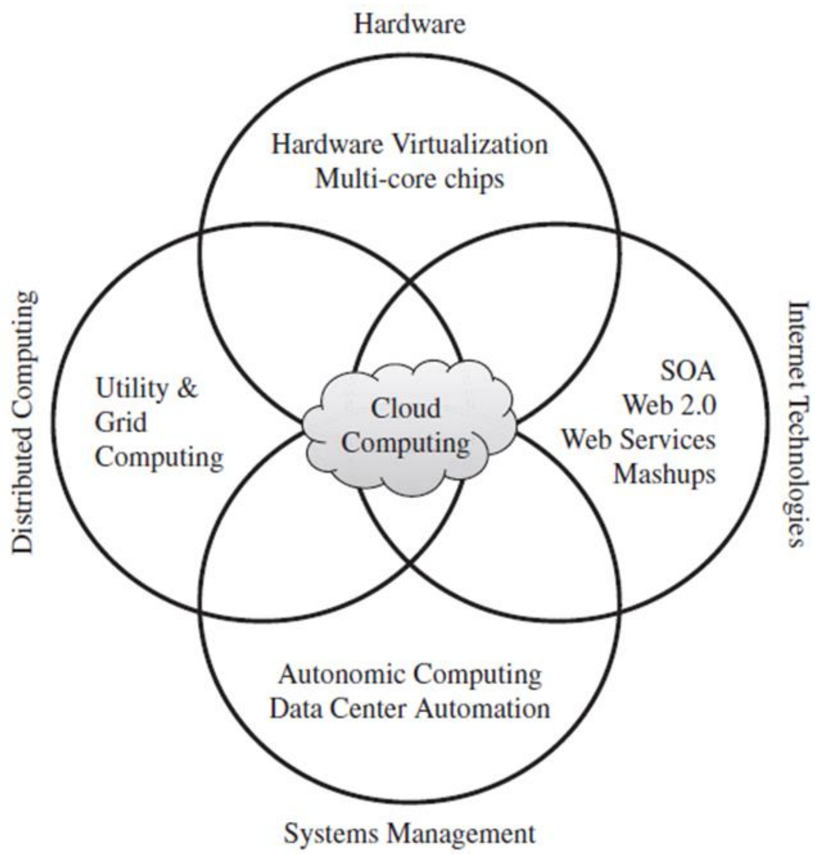
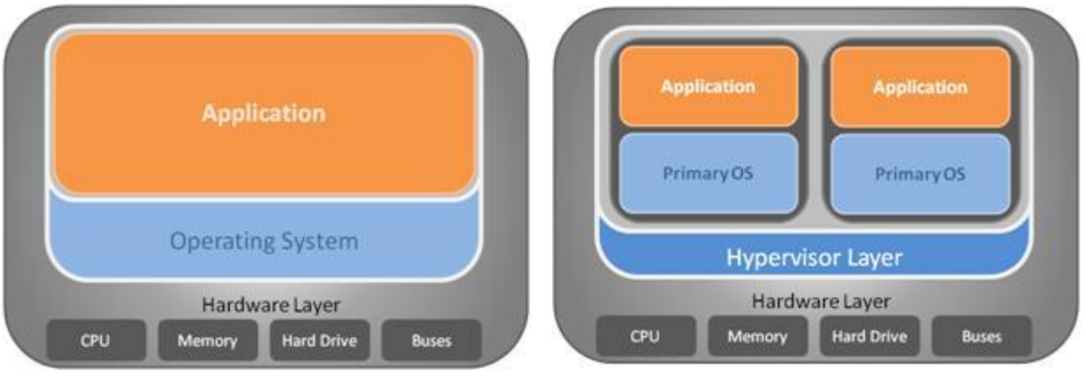


Figure 1 Convergence of various advances leading to the advent of cloud computing **[5]**.

2.2.1 Virtualization

Virtualization is a technology that allows sharing of a single of an application or resource among multiple tenants. It does so by assigning a intelligible name to a physical resource, and on demand will provide a pointer to that physical resource in order to satisfy [3]. The basic concept was centralizing physical computing resources and run them one part. Then the separate requests can be serve as required from the pool resources. For example, generating dynamic platform for an application is possible when needed. The virtualization techniques have a number of advantages for providers, such as resource usage in virtualization in way that every single load order can be satisfy from the resource pool. In cases of demands increasing, supplying large data centers with electric power becomes increasingly difficult. Centralizing reduces the number of physical components, and as a result, reduces the expenses for power supply. The virtualization also has benefits for the customer, such as scalability. Any scale up or scale down can be done quickly without delay. In cases of bottlenecks, the resource pool using virtualization technique can provide more computing resources (for example storage, Input/output capabilities). Another benefit for the customer is availability; where services will be available for daily use. Moreover, it is possible to migrate applications to another up-to-date system in the event of technology upgrades [4].



Traditional Architecture Virtualization Architecture

Figure 2 Virtualization architecture **[45]**

2.2.2 Grid computing

Both the grid and cloud computing are technologies that supply consumer with operational computing resources as state to their specific demands [5]. Grid computing is often confused with cloud computing, but they are quite different. Grid computing sets up a number of computers as resources in a network to work simultaneously on a single problem. This is usually done to address a scientific or technical problem. The Grid computing was developed to sharing hardware/software between many users. Grid computing enables users to gain access to physical resources by deploying interoperable services. The resources in the grid paradigm are administrated by their owners. By using remote machines without paying and without service level guarantees, authorized users can call grid services [6]. In contrast, cloud computing was developed using a top down approach. The concept of cloud computing is to provide users computing resources with high performance. In the cloud computing paradigm the users receive virtual resources. The hardware and software that are used in the underlying infrastructure are not exposed to public. The only important piece information that a user needs to know, it is the quality of services (QoS) for which they will pay. Cloud consumers call for operational software, platform and infrastructure with specific performance. Cloud computing uses virtualization to create insulated environments that can configure to specify users' requirements. The cloud gives users a virtual machine with exact properties that they ask for. Furthermore, it is easy to use and scalable [4].

2.2.3 Big data

Big data points to the collection and the following analysis of any large set of data that may include hidden information such as use, sensor data. When analyzing big data in right way, it can open new business fields, creating new markets and competitive benefits [7]. As compared to the structured data, According to IBM big data is classified to three major classes Variety, Volume and Velocity as show in Figure 3**.** [8].

Figure 3 IBM characterizes big data by its volume, velocity, and variety **[8]**.

2.3 Cloud Computing Definition

In order to analyze and specify performance indicators for selecting cloud database service provide, it is important to defined cloud computing, what are the services it provides?, Moreover, what are the advantage, disadvantage of the cloud computing. Cloud computing can consider as the fifth generation architecture in the information technology world. We can say that cloud computing emerged by collaboration between the following technologies: Main frames, Client Servers, Web, SOA were developed in years 1970,1980,1990,2000 respectively, and the cloud in 2010 The National Institute of Standards and Technology (NIST) define cloud computing as follows [9]

> *"Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."*

For example, we can describe cloud computing as following example: when plugging a device into an electric point in the wall, we do interest how the electric power is generated or how it reaches the wall socket. It is easily available from the socket that hiding the power generation process and the distribution lines. When we applied this idea to information technologies, this is referring to delivering useful functions while hiding the specifics of internal working. Computing can be considered completely virtualized, by allow computers to be built from distributed hardware software components, for example CPU, storage space, and software resources [9].

2.4 Cloud Computing Service Models:

Back to the definition of the National Institute of Standards and Technology (NIST), cloud computing paradigm is divided to three main service models, namely the SPI models: Software as Service (SaaS), Platform as Service (PaaS) and Infrastructure as Service (IaaS), as shown in Figure 4. Software as a Service (SaaS) represents the ability provided to the user to use applications runs on the provider's infrastructure. The applications can be access form many user devices by thin client interface or a program interface, for example a web browser for accessing web based email such as Hotmail. In cloud computing environment the consumer does not have control over the infrastructure such as operating system, network bandwidth, storage space, with possible configuration limitations to user application settings. The second service model, Platform as a Service (PaaS), represents the ability provided to the user for deploying into the cloud infrastructure consumer applications that are created by one of programming languages, libraries, and tools that were supported by the service provider. The user does not have authority to control the cloud infrastructure, such as network bandwidth, storage space or operating systems. Moreover, the consumer does not have control over the underlying applications and configuration settings for the application-hosting environment. The third service model, Infrastructure as a Service (IaaS), represents the ability provided to the user to have processing power, storage space, network bandwidth and other essential computing resources. In this service model the user will have the ability to run and manage qualitative software, such as operating system and applications. The consumer cannot manage or can controlling the deployed infrastructure, the user can have full control over the storage space, underline applications and operating system, and with control limitations over select network components [9].

2.5 Cloud Computing Deployment Models:

Cloud computing paradigm is deployed for using by four deployment models, the private cloud, public cloud, community cloud and the hybrid cloud, as shown in Figure 4. The **Private cloud** is the cloud computing infrastructure that is limited for especial use by a one organization and includes multi consumers (e.g., business units). It is possible to owned, administer managed by the organization, also can be mange by a third party, or by sharing the management by the organization and third party. It can exist in or out of the consumer building. The **Community cloud** is the cloud computing infrastructure which is restricted for using by a specific community of users or even organizations that have the same concerns such as tasks, requirements and policy. It is possible to be own, operat and managed by one or more of the community organizations , also can be mange by a third party, or even by combination of organization and third party. It can exist on or off of consumer building. The **Public Cloud** infrastructure is open to public use by the general users. It can be manage, own, and run by a company, academic or government organization, or by some combination of them. It exists in the cloud provider building. The **Hybrid cloud** is consists of cloud computing infrastructure that collects two or more of cloud deployment model (private, public or community) by keeping component models as separate units but connected to each other by technology which enables data and application portability [9].

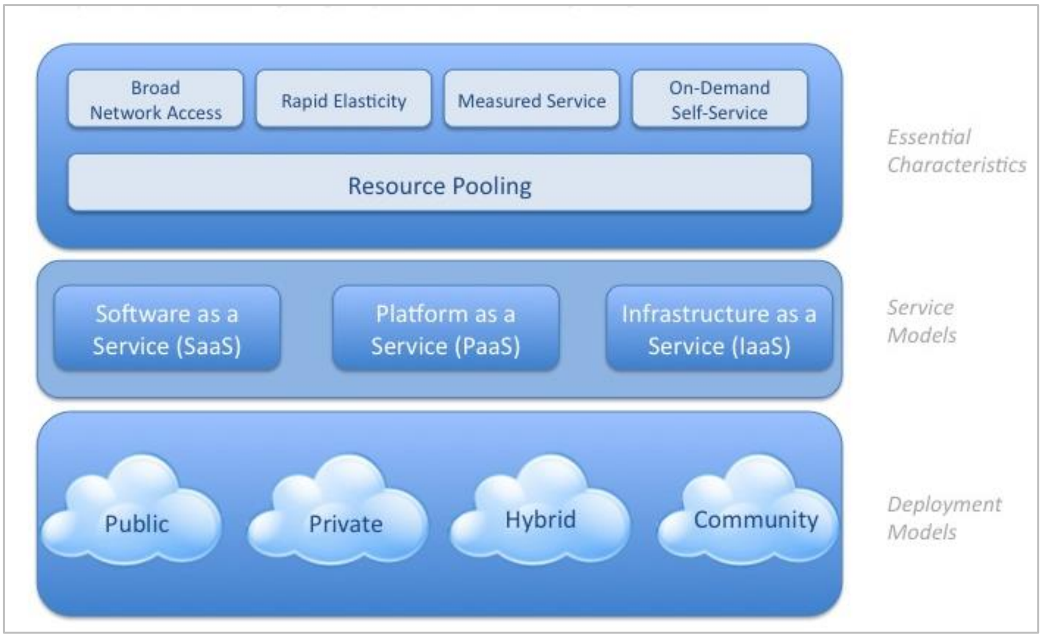


Figure 4 Visual of NIST working definition of cloud computing

2.6 Benefits of Cloud Computing:

Cloud computing model is Internet based computing of public participation. Its computing resources (computing power, storage capacity, and integrating capacity) are dynamic, scalable and virtualized. The cloud model has many benefits. For example, the low cost of all resources and network equipment is shared via the Internet, resulting in reduced costs for small-sized applications. Moreover, dynamic scalability in the case of the server-client model and most IT infrastructures uses 20 percent of their equipment processing power with 80 percent remaining unutilized. However, in the cloud model, they have the real need of storage resources, computing resources, by using a pay-per-use model. Furthermore, for simplified maintenance in the server-client model, the user (organization) must keep the software up-to-date. In the cloud model, however, the service provider relieves users from routine software updates and backups [10]. Furthermore, agile supply-on-demand enables faster setup of resources on an as-needed basis; for example, when a project is started, a manager can simply ask cloud services from any provider. Then if the project is finished or fails, one simply terminates the Contract with the service provider [11]. Another benefit for the cloud is that it is green, meaning it is more environmentally friendly than in-house computing because the service provider can run their infrastructure more efficiently. Cloud computing service providers with hundreds of servers will rarely need to have many machines powered up while standing idle. In contrast, in most data centers in smaller organizations, many servers and desktop computers are kept running without a single process. This means wasting electrical power for cooling and running hardware. There are many services providers that proved that cloud computing is green by reducing power consumption. Netsuit, for example, made an announcement in 2008 and its customers saved $61 million in energy bills by using its web-based customer relationship management system [2]. Organizations have increasingly started to move their database systems into cloud database services because of the advantages that they provide. One advantage of a cloud database is affordability. Because the cloud database paradigm reduces maintenance and operational costs and simultaneously provides users services on a per usage basis only, many organizations have adopted cloud database services. Flexibility and scalability, according to increases in video and image resolution, databases can grow quickly in size. New scalable solutions are required in that case. By using cloud databases, users can scale their services up/down to meet the changes needs in their database automatically without any interaction, as result increases efficiency. With databases reside in the cloud, database administrators and users will have the ability to access the database from anywhere by using a computer, tablet or mobile device, equipped with a browser and Internet connection. This is the increase in mobility that a cloud database provides. At the same time, connecting an application to the database can be achieved without any configuration to the cloud databases [12].

2.7 Obstacles to Cloud Computing Adoption

Organizations see the cloud as being risky for many reasons that accompany any new technology. New technologies usually take some time before becoming widespread and accepted. For cloud computing, there were two reasons for lack of adoption. The first is technical issues, such as Data management, security attacks and privacy. For organizations, there are the unstable operations which cause outages in this new cloud computing model. This notably affects some of the largest providers, such as Amazon, whose outages are shown in Table 1. Furthermore, there are non-technical issues, such as legislation and government policies. Organizations faced with privacy law, regulatory requirements, and data security regulation, and for government agencies, they even face more [13]. Another concern is the management of data by third party. Another legal issue which is complex, and not well understood; is holding the data and application by a third party, Moreover, there is more chance to lose the control over data when becomes in third party position. It is not well defined whether it is the responsibility of the customer or the provider to maintain the privacy of any personal data. Part of the experts' advice is to build a private cloud computing network in order to gain the advantages of cloud computing [14]. Regulations affect many areas of the lives of companies, such as how can a company run and secure their digital data. Data security laws in the United States often address specific industries, such as the financial and healthcare industries. However, these data security laws do not specifically address the cloud computing model. Some government agencies have begun to take action to address security concerns after understanding the associated risks with these security concerns in the cloud computing model. Compliance is another issue must considered, it is obstacle for cloud service providers and consumers how involve in using services. For example, in the USA financial sector data security is governed by Gramm-Leach-Bliley Act (GLBA), which is the rules covering many aspects of cloud networking, including information security and privacy, and sets standards to ensure the confidentiality of user records and protects them against threats of unauthorized access. The financial sector is expected to comply with GLBA regulations, but the GLBA does not explicitly refer to the cloud computing model. Another example is the healthcare sector regulation Health Insurance Portability and Accountability Act (HIPAA). These regulations have many data security requirements. HIPAA stipulates that the healthcare sector ensure the confidentiality, integrity, and availability of personal data as well as the protection of data against threats and unauthorized access. Moreover, HIPAA regulations do not explicitly refer to the cloud computing model [15]. According to a survey organize by the consulting company Capgemini, the greatest obstacle to the adoption of cloud computing services in general is the fear of security breaches (41%), followed by issues with data sovereignty (35%), as illustrated in Figure 5 [16].

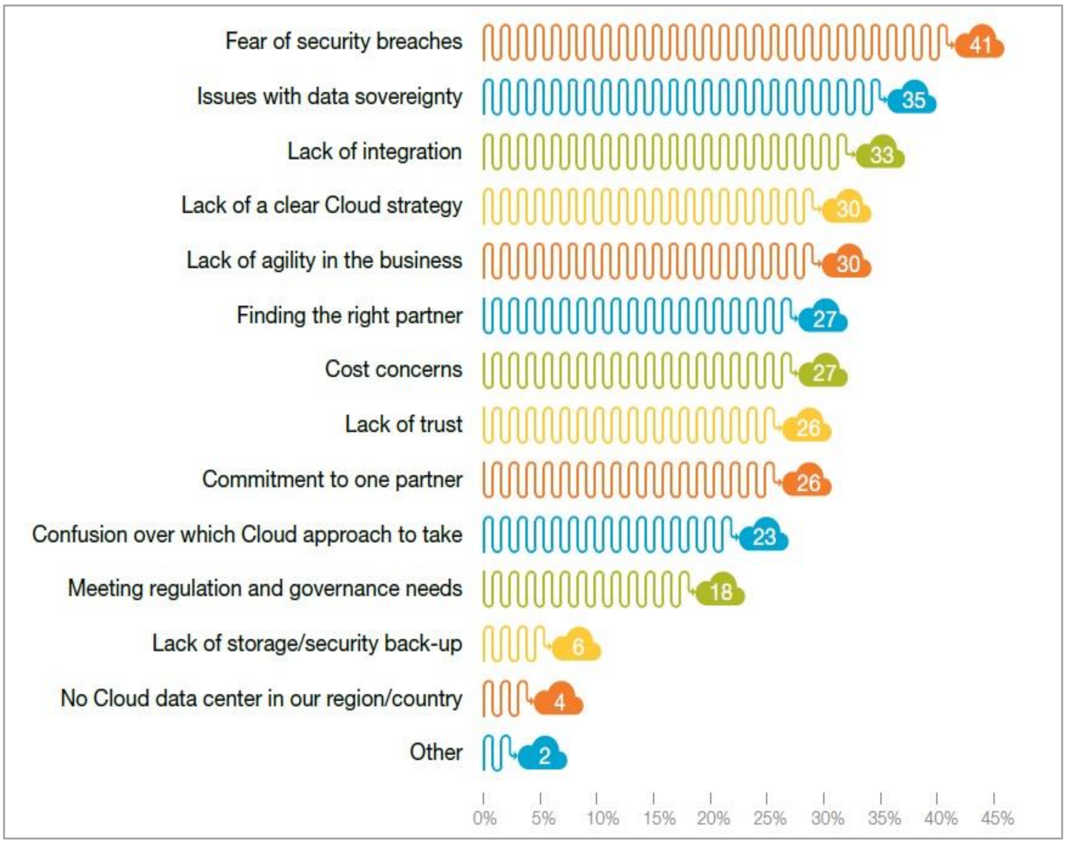


Figure 5 Obstacles to Cloud Computing Adoption [16]

The adoption of cloud services is spreading and still growing with a few exceptions in the EU in the private and public sectors. In organizations that are considered to be strong adopters, the cloud represents a small part of an IT department. However, adoption varies according to factors such as availability, price, and speed of Internet connection. The attractiveness of cloud adoption varies; for example, for organizations that have used cloud computing before, do so because of the efficiencies and flexibility of cloud computing; and for those have not started using cloud computing, the attraction will be the lower cost and ease of use. In order to increase the adoption of cloud computing across the EU, action must occur, such as ensuring accountability, which means providing a clear definition about cloud service providers' accountability, liability and portability. This ensures that the user is able to move data and applications between cloud service providers [17]. However, this may mean an additional 45€ billion of direct spending on cloud computing by 2020, at the same time overall cumulative impact on gross domestic product by 957€ billion, and with the availability of 3.8 million jobs in 2020 [18].

2.8 Cloud Database Services

Cloud computing is fast rising in IT department in the world. Many companies have begun moves to cloud computing by accessing their Database Management System (DBMS) from cloud databases providers. In 2011, over 36 percent of US companies were already running applications in the cloud infrastructure. Cloud databases are accessible by the user through the Internet in the cloud database service provider infrastructure when a user demands. The cloud database is carried out using the cloud provider software and hardware resources [12]. Cloud databases engage companies for two main reasons in terms of financial costs. Firstly, economies of scale; in the case of the client-server, the company must purchase all hardware (server and networking) that is required to setup the database management system; in other words, operating everything by themselves. However, in a cloud database, the cost is falling because the user pays for a shared resource. Secondly, expertise is required to extract good performance from DBMS in cases of the client-server, which increases operational costs for companies. In cloud computing, by centralizing the expertise to run hundreds of DBMSs, this leads to a reduction of operational costs and an increase in performance [19]. In general, the available CDBMS in cloud computing has two structures. Number of providers offer multi-instance architecture, while another number of providers support a multi-tenant architecture. Both structures install the CDBMS on a Hypervisor. (A Hypervisor is computer software that runs virtual machines, as shown in Figure 2). In the multi-instance architecture, every user is provided with an exclusive Database Management System (DBMS) running on a virtual machine for each user. This architecture allows the user to have good administrative control over security tasks. In contrast, the multi-tenant architecture provides the database area that is used by many tenants by using a labeling method.

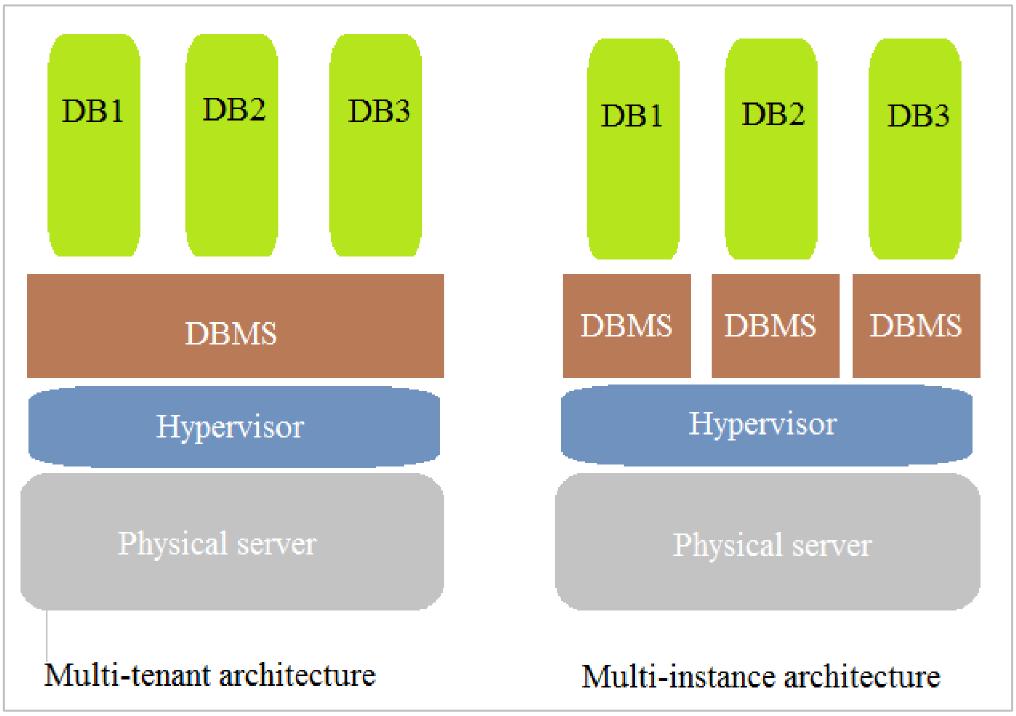


Figure 6 Multi-tenant vs. multi-instance architecture

In Figure 2, the data for each tenant is labeled by identifiers. The cloud provider in the multi-tenant model is responsible for the maintenance and establishment of a secure database environment. Furthermore, the multi-instance structure is more easily configured than the multi-tenant structure, such as in security features and data encryption [20].

2.9 Structure of Cloud Database

In 2011, over 36 percent of United States companies already ran their applications in the cloud. According to the trend in the Information Technology world, in the coming years a large number of IT departments will run their applications using the cloud. We can describe the cloud database as a distributed database that offers a query service over multiple distributed database nodes existent in multiple geographical locations in virtual environments. An example would be an organization that has a database application running in the cloud database. The service provider would have more than one data center on that server. In such an organization in this case, the user query will split between these data centers to answer a user query, as illustrated in Figure 3. The user query can be generated from anywhere, such as from a PC or from a laptop via a Wi-Fi connection, or from a smart phone

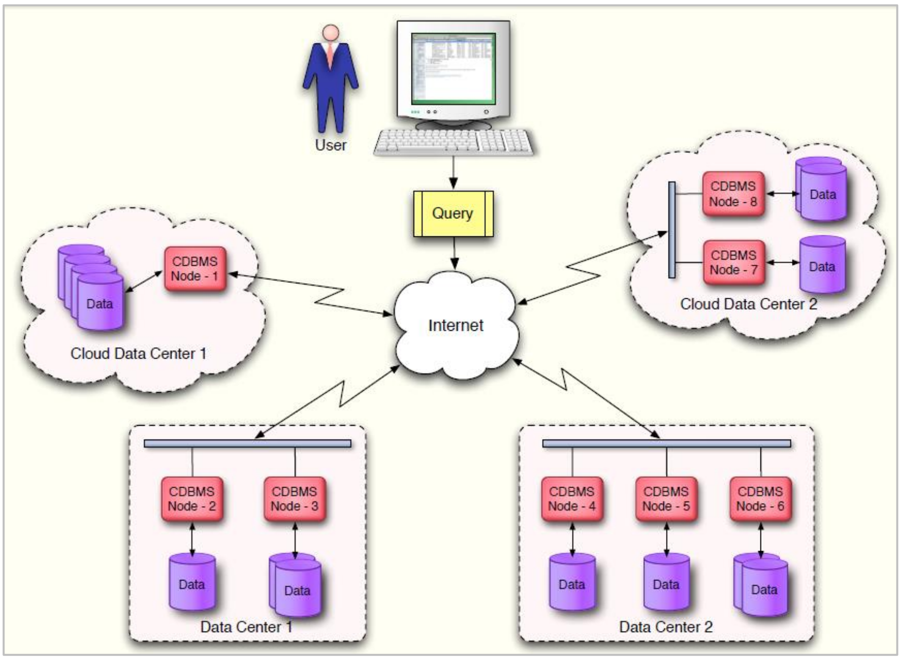


Figure 7 Example of the structure of a cloud database

Through 3G and 4G connection; for that reason, in Figure 3 the query is represented as coming "through the Internet;" at the same time, it implies that the response will return back to the user through the Internet [21].

2.10 Challenges of Cloud Computing

In spite of the fact that cloud computing has many advantages and benefits and large number of vendors and tools are available, In parallel serious risks and security challenges are associated with the cloud model. All participants in the cloud model (providers, developers, users) must take into consideration these security challenges. One of the important issues in the cloud paradigm is that users lose physical control over storage and computing resources. Privacy can be a major issue with regard to security in cloud computing. The level of privacy can be attributed to the ability to control data. However, the architecture of cloud computing saves data in distributed locations in plaintext format in storage resources. The firewall can offer a reasonable degree of protection. However, this architecture will put some important data into a position of leakage [5].

Data availability is another concern in cloud computing in cases of natural disasters or in cases of hardware or software error. How can a service provider recover the user (organization) data? The migration procedure for data or applications in case a consumer decides to move to another service provider does not guarantee data erasure, which is another concern in cloud computing [22].

The virtual database environment that used in cloud computing can differ in providing service. For example, some service provider environment support multi-instance architecture, however others providers support multi-tenant architecture [23]; therefore, in both architectures data are stored in a shared environment with other consumers data. Organizations outsourcing critical and important data to the cloud, therefore, must take into consideration the access, modification and use of these. The data are controlled and secured. There are many challenges that cloud database services is facing such as availability, physical access control, Data Isolation, connection speed and security attacks on cloud computing. [5]. These are explained below.

2.10.1 Availability

Availability refers to the probability of working the Cloud Database Management System (CDBMS) continuously during a particular time interval. It is a primary concern for companies providing data outsourcing. It can be affected temporarily or permanently. Many factors may contribute to the non-availability of a cloud database, including Denial of service attacks (DOS), equipment failure and natural disasters [24]. Furthermore, outages have been known to occur, examples of which include outages experienced by Gmail, Amazon S3 (Simple Storage Service) and

AppEngine (as shown in Table 1) and a notable outage experienced by FlexiScale, which lasted for 18 hours on 31 October, 2008 [14].

Service Providers and Outages	Duration/ Hours	Date
Amazon simple storage service: authentication system excessive load led to unavailability	2	02.15.2008
Amazon simple storage service: Single bit error leading to gossip protocol blowup	6-8	07.20.2008
AppEngine: coding and programming error	5	06.17.2008
Gmail: error in contacts system led to unavailability	1.5	08.11.2008

Table 1 Outages in Amazon web Service, AppEngine, and Gmail [25]

2.10.2 Physical access control

To establish physical security that can be defined as the procedures taken in order to prevent any illegal actions (theft, espionage, or harm) against IT and network equipment in case of the client-server, it is done simply by adding a locked door or assigning security guards. When organizations and government agencies outsource their data into a cloud database, they lose physical control over the data. When data are outsourced to a cloud provider’s database environment, organizations will lose the ability of implementing the standard level of traditional security and physical access limitations. Moreover, to guarantee operations and high rate of availability for the system to all consumers that use services, the provider’s staffs often authorized with unbounded access to the service provider infrastructure facility. Therefore, users must review cloud service providers control procedure that followed in provider physical infrastructure [26].

2.10.3 Data isolation

Outsourcing data into the cloud model means sharing computing and storage resources with other consumers because all consumers' data will be stored in only one instance of a software system. In order to have an acceptable level of data isolation, the service provider must use an isolation mechanism to ensure the confidentiality of data [27].

2.10.4 Connection speed

The Internet speed that is used to access the cloud service in organizations is slow if it is compared to the data transfer speed in the cloud data centers. Therefore, this will affect the performance and efficiency of the cloud database. The speed of sending the query is very fast in data centers, but data retrieval is slow in organizations. The solution for this problem is to use an expensive fast speed line [12].

2.10.5 Security attacks on cloud computing

The security attacks in the cloud environments include computer software and network security attacks. By moving to cloud it will be likely to happened or easier. There are many types of cloud computing attacks, including the Denial-of-Service attack (DoS). Because the nature of cloud computing depends on sharing computing resources with many users in a virtual environment, more DoS attacks will occur. Another attack is the malware injection security attack. It is considered the first attack that attempt into the cloud. The purpose of malware injection security attacks might be anything interests attacker, which include modifying of data, change or reverse the functionalities or even blocking the service. In this security attack, the first step done by the attacker is to build his/her own malicious service application module according to targeted service such as SaaS, PaaS and IaaS, and second step is to add the malicious service application to the cloud system. After that the attacker acts to the cloud system like an instance for a new implementation. If such behavior occurs, the cloud in automated way redirects the requests that done by a valid user to the malicious service, after that attacker code start to execute. Another security attack is the authentication attacks. This is a weak point for services and applications that hosted in virtual environment and it is frequently occur. Based upon what user knows many authentication methods are created. Mostly the attacker targets the authentication securing mechanism and methods [28]. On the other hand, the cloud service provider has solutions to these security attacks; for example, for DoS attacks, activating rate-limiting and access control list capability on switches and using firewalls to allow or deny access through ports or IP addresses. Also, for malware injection, one solution is done by saving the OS type for every user in the first stage when a user create the account for later checking in case of any request. For authentication, attacks, service providers recommend using more complex user name and password companied by a secret question [29].

2.11 Encryption Considerations

Encryption is not a new technology for databases. Many encryption methods have been used for encrypted databases. The main variance between the encryption methods is the level of detail present for each method. For example, some methods encrypt a relation and others the tuple and while some others encrypt the entire database. With the amount of important data sitting in one pool, encryption appears to be a logical solution for security. However, this level of security will cost in terms of computation power. Encryption affects the performance of the database because the decryption operation for huge amount of data takes time, and the decryption occur in every single query statement, Therefore, the decryption process becomes redundant [30].

2.12 RDBMS and NoSQL Schemes

Databases are collections of arranged data. In general, the expression database at most points to the whole database system; however in fact, it points only to a set of data. The system that manages the data, transactions, modification (Adding, Updating, and Deletion) is the Database Management System (DBMS).

Relational databases use the concept of databases that isolate data into tables where each column refers to a field and each row represents a tuple. Tables can be connected to each other by using foreign keys or mutual columns. The collection of fields and columns creates a table which represents the entities, including users, customers or suppliers [31]. The RDBMS, which provides ACID (Atomicity, Consistency, Isolation, Durability), is a set of properties that secures the database transactions operation completely and dependably [32].

The main characteristic of NoSQL databases is that they have a dynamic schema; records can have different fields as needed. In another words, NoSQL databases have no predefined schema. Furthermore, NoSQL databases do not provide an ACID transaction such as RDBMS. The growing popularity of NoSQL databases occurred in the 2000s when Amazon and Google began to conduct research deeply into distribution databases. Therefore, the classes of NoSQL develop and contain many other subtypes, each of which is the best for a fixed data group. There are three main public NoSQL database classes. The first is **Document stores.** The concept of documents is the main idea here; by replacement of the documents instead of the tuples in relational databases and the set of documents represents the tables. The second is **Key-value stores.** Data is stored as values with a key assigned to each value similarly to hash-tables. In addition, by relying on the database, a collection of values can have one key. The third is **Graph databases.** The concept of graph structure is the basic idea of a graph database, in which nodes and edges are the major idea. Edges represent the relations between the entities (nodes). Nodes represent entities, such as a user or a music recording. NoSQL provides index-free adjacency. This means every item has a direct pointer to its next element; therefore, there is no need for indexing of every element [33].

2.13 E-government and C-government

The e-government is a management system, in which the agencies and departments of government employs information technologies in order to deliver public services to society. E-government integrates several organizations and departments with different tasks and responsibilities. E-government in general depends on the in-house IT infrastructure to apply information technologies. E-government depends on a huge budget for IT infrastructures in order to deliver services to citizens. Many e-government applications have been built for years, and as is certain to happen, different problems will arise, including data explosions, cost of operation and maintenance that makes it difficult to provide high quality performance services to business and government sectors. As a result, e-government initiative will fail to meet the requirements of public service and will face difficulties with business and governmental department collaboration [34]. In cases of in-house, the real utilization of such hardware is low. The cloud model offers solution to reduce the expenditure and cost of owning in-house IT infrastructures. E-government implements similar applications in every department. Cloud computing services can reduce such redundancies and make the collaboration between different departments more flexible. Moreover, an IT service is needed for any task that can be added with less effort and time as in the case of in-house IT infrastructure. This will increase e-government scalability and flexibility. Table 1 shows a summary for deferent governmental application and importance of the public requirements to them [35].

Performance Indicator	**Governmental Applications**	**Financial Services**	**Healthcare Applications**
Archiving	H	H	H
Audit	H	H	H
Availability	M /H	H	H
Flexibility	M	H	M
Privacy	H	H	H
Security	H	H	H
Quality of Service	M	M	M
Scalability	M	M	M
H = High, M= Medium.			

Table 2 General Requirement for E-government Applications [35]

2.14 SLA Agreement Important Points

The Service Level Agreement (SLA) is the agreement part of the service that clearly and specifically defines the limitation of the service, such as scope, quality and responsibilities. SLA is considered to be an individual valid agreement between the service provider and the consumer. As a result, SLA becomes the key to understand the nature of the service. SLA is a method to reduce and manage the risks as well as become aware of and deal with lack of security. The SLA varies from one service vendor to another as well as between various cloud services. In these days, the SLA focuses more on financial side and exceptions, instead of reviewing service level expectations. The SLA is the only official document between the consumer and service provider. Therefore, the consumer has to use the SLA to achieve better services as well as tangible outcomes from the service and service provider. The customer must be clear as to what the service provider will do in cases of recovery from disaster, physical and logical security, restriction on data location, support for investigation, data destruction, and encryption key management and network security. The SLA must contain a clear framework for the responsibilities of both parties; including understandable steps to handling and solving any issues that may arise. Furthermore, other business issues to be reviewed include guarantees and conditions for termination of service. The consumer is required to overview the nature of the service provided and any associated risk, including the most important aspects, such as the strength of the security solutions offered by the provider, which in general, are not necessarily perfect solutions. Nevertheless, by realizing what the service provider is able to cover, the consumer can make informed decisions and judgments [30].

2.15 Selecting a Suitable Cloud Database

Carnegie Mellon University Silicon Valley has built a Service Measurement Index [36] which is a set of business-relevant Key Performance Indicators (KPIs) that provider users and organizations with a standard procedure to measure and compare the quality of services in spite of the fact that a service is internally sourced from a service provider. It has been built to be a method to assist users and organizations to evaluate cloud service providers depending on organization specific business features and the technology requirement. The development of the Service Measurement Index presently is done by the Cloud Services Measurement Initiative Consortium (CSMIC). The Service Measurement Index (SMI) has a top-down framework. The top main levels contain 7 categories. Each category is subdivided into three or more attributes. Every attribute is defined by a set of Key Performance Indicators. The

SMI can be used for BPaaS, IaaS, PaaS and SaaS. Depending on Service Measurement Index (SMI), we propose a performance indicator to evaluate the service providers for a Cloud Database (DbaaS), which is a subtype of Software as Service (SaaS).

2.16 Literature Review

In the cloud computing paradigm, hardware and software resources are provided as services over the Internet. One of the advantages of the cloud computing model is reducing the cost, where user expenditures include only service using payment and not for software installation, updating and maintenance. Database as service (DbaaS) is part of Software as service (SaaS) and the typical function is to provide users with the Database Management System (DBMS) in cloud computing. The market for this solution has grown and many companies, including IBM, Microsoft and LongJump, provide DbaaS. According to the cloud paradigm, DBMS moves from traditional client-servers, where the organization is responsible for managing data and answering user queries to outsourcing, and where the third part is responsible for data management. In spite of cloud computing's great benefits in cost reduction and better service, the widespread use of DbaaS has not been adopted because of the security issue [37].

The cloud computing market is growing. Cloud computing consumers in a few easy steps sign up tens or hundreds of virtual machines (VMs). They payment is for what only they really use, without any operation costs. This sort of flexibility has encouraged more organizations, including governments, enterprises and businesses, to migrate their applications to cloud computing. Many companies are running cloud computing services, such as Google's AppEngine, Amazon Web Services (AWS), and Microsoft's Windows Azure. From the perspective of competition, the variety of cloud computing service providers is healthy; however, it makes selecting a suitable cloud service provider for an application difficult [38].

2.16.1 Related Work

Jae Yoo Lee *et al*. define a quality model to evaluate Software as Service (SaaS), by defining the key feature of SaaS, and dependent on them, they define the metrics for the quality model attributes. Additionally, they are based on IEEE 1061. The quality model attributes were reusability, efficiency, reliability, scalability and availability. The metric for the attributes were functional, Non-Functional Utilization and coverage of variability for reusability, utilization of resource and time behavior for efficiency, service stability and service Accuracy for reliability, coverage of scalability for scalability and robustness of service for availability. The quality model can help the service providers to evaluate their service to anticipate any development trends in the service. Moreover, the consumer can view these evaluations to select the best service providers [30].

Waleed Al Shehri starts the paper with an interview into the cloud database and how cloud computing growth until 36% of companies in United State of America has running their applications in the cloud. Additionally, the author provides an overview about the structure of cloud databases which are built depending on the idea of distributed database management systems and how users can access the database through any device with Internet access, such as a computer, tablet or mobile phone. Furthermore, the author mentions distributed queries and how a main query can be divided into sub queries if the required data are in another location in order to answer the user. The researcher also listed the main companies that provide cloud database management system services, for example Amazon Relational Database Service (RDS), Microsoft SQL Azure, Google AppEngine Data store and Amazon SimpleDB. He further mentioned that these companies provide cloud database services of different quality and type of services.

The author attempted to come out with parameters or performance indicators that can be consider in selecting service provider that meets with user (organization) requirements. According to the researcher the first parameter is Data Sizing. The author considered the data size that is provided by the service provider to be important. For example, Amazon RDS provides users with total store capacity up to 1 TB of data; while, SQL Azure provides users with 50 GB of data for every database instance. The second is Portability. The author considered the portability of a database to be important in cases of the service provider going out of business. In that case, the user (organization) is required to transfer the database to another service provider. The third is Transaction Capabilities. Many companies generally use money transactions in their database systems, so they require a complete read and write operation in order to guarantee that transactions are made. The fourth is Configurability. All database setups are carried out by the service provider and a few options to be configured are left to the database administrator, which is considered to be the best option according to the writer. The fifth is Certification and Accreditation. The writer recommended selecting a service provider with certification and accreditation, such as FISAM certification, in order to reduce the risks of the services to the company. The final parameter is Storage Location. The author recommended choosing the service provider whose storage location is declared geographically [37].

Shalini Ramanathan, *et al*. start with an introduction to the cloud and database service in cloud model. Furthermore, the author overviews Amazon’s SimpleDB and Google’s Bigtable database as a service by analyzing the characteristics, architectures, advantages and points to some limitation in the mentioned providers’ DbaaS. Moreover, the author puts some criteria in this paper as a standard to make a comparison between Amazon’s and Google’s Database service; for example, programming language, scalability, data storage costs and usage. All in all, from the researcher’s perspective, Google’s Bigtable is deemed to be the better choice for storage of huge data. Furthermore, the paper also comes out with detailed guidance for users to choose DbaaS [39].

Jialin Li, *et al*. state that in order to select a good service provider for cloud databases, the user faces multiple service providers with their respective performance or price of service. Therefore, they make comparisons between Amazon Redshift and Google BigQuery (cloud Database service providers) and compare the results with PostgreSQL 9.1 by benchmarking the quality of performance and scalability by evaluating the time with increasing data size and number of nodes. They depend on the Transaction Processing Performance Council (TPC) and measured the query response time for the first different type of query and the second specific query by changing the dataset size, and the third specific query by changing the number of server nodes. They produced these results: for BigQuery, the user has no control over how many nodes are being used to process the query. Furthermore, query run time increases significantly when the data size increases. Query run time increases when the SQL query joins the operation. In the redshift, the user has control over the number of nodes being used to process the query. With constant data size and an increase in the number of nodes, the runtime for the query decreases. By increasing the data size and number of nodes, the result of the runtime is decreased. Moreover, when running the same query more than one time, the run time decreases because of the caching feature in Redshift. After comparing the run time for both BigQuery and Redshift with PostgreSQL 9.1, both were found to be faster than PostgreSQL 9.1. Moreover, BigQuery has better performance than Redshift in parallelizable queries processing even as Redshift has the caching effort. For small data sizes, BigQuery is faster, but in multiple joins and nested sub queries, Redshift is faster. Redshift is much better at handling complex queries than BigQuery [26].

Pang Xiong Wen *et al*. start their paper with a definition of cloud computing and continue with some statistic on SaaS usage and the financial profit income for companies. After that, they attempt to frame a problem statement in two questions, the first of which is *How to measure the quality of SaaS including platform and software*. Second question is *How to explore an evaluation model that takes into account the key features of SaaS.* After exploring the key features of SaaS, for example, Data Security, Quality of Service, Multi-tenant, Configurability, Scalability, Pay-for-Use licenses, the authors put forward a quality model which consists of Security, Quality of Service, and Software quality. After these three dimensions of the elements of quality model, the authors present some metrics for security, namely the five metrics of Customer Security, Application Security, Network Security, Data Security and Management Security. For Quality of Service, they specified three metrics: Quality of Platform, Quality of Application and Quality of Experience. For software quality metrics, they depended on the same product quality model as the ISO/IEC 25010:2011 product quality model. After completing the quality model, the authors created an evaluation model for the SaaS Service by depending on the above quality model to classify the SaaS service into four levels, namely Basic SaaS, Standard SaaS, Optimized SaaS, Integrated SaaS, by which SaaS Services meet the quality models and its element metrics [40].

Xianrong Zheng, *et al*. proposed a quality model to evaluate cloud service providers. The model has six quality dimensions: usability, availability, reliability, responsiveness, security and elasticity. Moreover, the writers make a comparison between three cloud storage service providers: namely, *Aliyun*, *Amazon S3* (Simple Storage Service) and *Azure Blob* using Usability metrics. For example, Aliyun has API and GUI; Amazon has API and GUI; and Azure Blob has only API. Furthermore, the researchers compared service providers using Responsiveness time (by calculating cloud services performing a request during a time interval where the closer value of responsiveness to 1 is the better responsiveness) for three operations: Uploading, Downloading and Deleting, where the Azure service provider has the best Responsiveness value (of 0.993) [36].

Eileen Marie Hanna, *et al*. they tried to define eight requirements for cloud services which they are security, privacy, availability, auditing, flexibility, archiving, quality of service and scalability. By studding governmental cases for large-scale applications financial services, healthcare applications and online entertainment, they explained the importance of these eight requirements in the context of e-government, such as archiving, which is of high importance and is required for governmental applications, financial services, healthcare applications and of low or medium importance for large-scale computations and online entertainment. Auditing is highly important and is required for governmental applications, financial services and healthcare applications but is of low importance for large-scale computations and online entertainment. Availability is of high importance for financial services, healthcare applications and online entertainment, and of low and medium importance for governmental applications and large-scale computations. Flexibility is of low importance for large-scale computations and online entertainment and of high importance for financial services and of medium importance for governmental applications and healthcare applications. Privacy is of high importance for governmental applications, financial services and healthcare applications and of low importance for large-scale computations and online entertainment. Security is of high importance for governmental applications, financial services and healthcare applications but less so for online entertainment and of low importance for large-scale computations. Quality of Service is of high importance for online entertainment but less for governmental applications, financial services, large-scale computations, and healthcare applications. Scalability is high importance for large-scale computations and online entertainment but less for governmental applications, financial services and healthcare applications. The authors have endeavored to create a quality model to evaluate service providers [35].

Saurabh Kumar Garg, *et al*. started their paper with an introduction to cloud computing. According to these researchers, the increasing numbers of service providers in cloud computing present important challenges to the consumer, most important of which is how to determine and select a suitable cloud service provider that can meet their needs. Depending on the Cloud Service Measurement Index Consortium (CSMIC) that is used to evaluate cloud service providers, the researchers proposed a framework to compare different cloud providers based on user requirements. The proposed framework consists of three parts. The first part is the SMI Cloud Broker, which receives customer requests for an application. The SMI calculator computes the different key performance indicators, after which the ranking system sorts the service providers according to customer requirements. The second part is Monitoring, which, self-explanatorily, is responsible for monitoring the performance of the selected service providers in the previous step. The third part is the Service Catalogue, which stores the service providers and their characteristics with different cloud service providers [31].

2.16.2 Summary

In the previous section (2.16.1), we reviewed related works, most of which were quality models to evaluate cloud computing software service providers. They depend on a collection of attributes for their models. The objective of this thesis is to propose performance indicators to evaluate cloud database service providers. In the next chapter, we present our proposed quality model attributes.

CHAPTER 3

PERFORMANCE INDICATORS

3.1 Performance Indicator Elements

Because of the advantages of the cloud services that are realized by the organization, the number of organizations that adopt cloud service is increasing. Simultaneously, Service vendors are increasing. Therefore, the difficulty of choosing the service provider increased and become complex [38].

The Service Measurement Index (SMI) helps consumers to find the most suitable service provider. In general, the SMI can assist users to select a SaaS. In order to evaluate the cloud Database as services (DbaaS) providers, we depend on the SMI as a basis. Additionally, we defined other performance indicators which are mostly related to database characteristics, such as Transaction Capabilities and Data size [41].

3.1.1 Security

Traditional data security, in cases of in-house hosting, refers to solving three main issues. The first is data confidentiality, which means preventing unauthorized users from using information. The second is data integrity, which is, protecting data from unauthorized modifications (editing or adding) or deletions [42]. The third is availability that means recovery and protection from software and hardware failure and also from malicious actions that make sensitive data or databases in cloud or a part of it unavailable. The Database Management System (DBMS) has its own security service, the most important is the reference monitor that is a software module interrupts every single access request to the system, and depending on the specified authorizations that are pointed to according to the access control policies, determines whether access can occur partially or totally. In the cloud database, data confidentiality consists of two different confidentiality requirements. The first is that of user confidentiality, that is, protecting data from unauthorized modifications and operations by end users. This is also carried out in cases of outsourcing data to service providers that may be untrusted. The second confidentiality issue related to database service providers is known as publisher confidentiality, that is, the threat of saving the organization data from modifications by publishers [37]. The level of security differs from one deployment model to another in the cloud. Different deployment models mean different control capacities and security levels. The e-government decision making sector can select various deployment models according to data sensitivity, control capacity, business requirements and budgets, as Table 3 shows[2].

Deployment Model	Security	Cost	E-government Department
Private Cloud	High	High	Department storing sensitive data and calling high security
Community Cloud	Medium	Medium	Department sharing same concerns and calling for collaboration
Public Cloud	Low	Low	Department directly providing public services
Hybrid Cloud	That Depend	That Depend	Large government organizations

Table 3 Comparison of Deployment Models [34]

3.1.2 Efficiency

Efficiency refers to the number and amount of resources that are used in order to deliver the require service functionalities to the user. In other words, it can refer to how well SaaS utilizes resources. These resources include network bandwidth, storage capacity and CPU [20].

3.1.3 Reliability

Reliability refers to the average number of failure occurs over a given period of time. It is a description of how a service can work without shortages over a given time and situation [2]. Moreover, reliability measures the ability of the SaaS to continue to run with a given standard of performance over time. The reliability in a cloud database has many aspects, one of which is the reliability of the services themselves. The second aspect is the messages reliability that interchanges among the service provider the consumer. The logical reasons for defining this quality attribute is that cloud databases are reusable and are used by many consumers. Therefore, a sudden failure occurs in the cloud database. That is, it transmits the failure to the service users who hired the cloud database service. Furthermore, in cloud databases, most of the data are managed by the service provider (server side); therefore, if any failure occurs in the cloud database, the reliability will affect the services that are produced by the consumers themselves [43].

3.1.4 Scalability

This performance indicator measures the capability of scaling IT hardware resources, such as compute nodes, CPU power, and network bandwidth, storage units while maintaining an acceptable level of performance. Scalability is important for e-government applications, especially for financial services, as Table 1 shows. The logical reason for defining this performance indicator is; cloud computing is a type of virtual computing environment in which dynamic hardware and software resources are provided and deployed as a service through the Internet [5]. Cloud database must support the growing amount of resources whenever the consumer requests it, since a cloud database is a subtype of SaaS providing a set of functionalities [22]. There are many different types of growth in the amount of IT resources among various organizations. Some organizations have firms growing in their computational and storage resources. In additional, some organizations require high growth at times and low growth at other times. Moreover, many organizations start with a small amount of resources and there is bound to be a request for a growth of resources. However, the stable increase in computation and storage resources can easily be managed [27].

3.1.5 Availability

This performance indicator measures the rate of the total unavailable service times into the total service time for a consumer who can reach the service [2]. It is highly necessary for e-government applications, such as governmental and healthcare applications, as Table 1 shows. The logical reason for defining this performance indicator is that; the availability of a cloud database is a basic condition for the effective running of a cloud database since cloud databases are deployed and run on the service provider server side. The consumer has no control of the service when high availability of service is not guaranteed and as a result, the service consumer cannot use the service on demand [43]. Many organizations ask for high availability levels for their computing and storage resources due to their importance to their organizational operations. One example can be found in financial and banking organizations. Cloud service providers have recently experienced technical issues due to issues such as natural disasters, security attacks and infrastructure failures [35]. The general idea of a cloud service is to share computing and storage resources with other consumers. Additionally, it is [30]. In March 2009, the law obligated a particular service provider to shut down in order to review the potential criminal activity of one of its consumers. As a result of the investigation, the businesses of other consumers of this service provider had failed [13].

3.1.6 Transaction capabilities

Transaction capabilities are a major characteristic for Databases Management System that require to provide warranty read/write operations. For example, financial and banking organizations that transfer money need to supply the user with complete feedback as to whether or not a transaction has succeeded. This type of transaction is known as ACID. The ACID transaction requires full processing and saving in order to ensure that all data are either written or deleted as one part. In cases of providing ACID, extra costs for the service will be incurred. If this step of insurance is not required, the cost of the service will be lower and faster by offering a non-ACID transaction [35].

3.1.7 Pricing

Most organizations would be willing to compare the pricing issue of cloud-based services from different vendors in order to determine acceptable pricing. Service providers usually rate their services price according to data size that can be store in individual database, and the amount of data that is transfer into and out of the database. Therefore, applications that transfer high amounts of data in/out to the database will incur a higher cost. As shown in Table 2 [23], cloud database service providers use different pricing plans; for example, Google Big Query charges per unit of data processed, whereas Amazon Redshift charges per hour per node [37]. Purchasing a virtual server with specific properties (CPU, storage and memory) will cause the expenditure on IT to become operational rather than capital. Additionally, the physical room required for racks of servers in case of in-house system is no needed. And an organization no to pay high power bills to run and cool their servers room [44].

3.1.8 Database accessibility

Database service providers offer a set of connection mechanisms for storage and data retrieval. In general, there are three methods for information retrieval. First, RDBMS offers general database connection drivers, for example, Java Database Connectivity (JDBC), Open Database Connectivity (ODBC), which allows the outer application to the service to connect to the database for storage and data retrieval. Second, NoSQL usually offers interfaces, which use standard based, Service Oriented Architecture (SOA) protocols, for example REST or SOAP, with Hypertext Transfer Protocol (HTTP). Third, some NoSQL databases may be bounded by connecting through to software running in the provider system, which restricts the portability but increases security [35].

Pricing Scheme	Amazon RDS (MySQL)	Microsoft SQL Azure	Google Datastore	Amazon SimpleDB
Example Pricing of Processing	$0.11 for small instant $2.60 per hour for largest instant	Range from $9.99 per Database with up to 1 GB of storage to $499.95 per database up to 50GB of storage per month	$0.10 per App Engine CPU hour	$0.14 per SimpleDB unit hour
Example of pricing for data transfer	Inbound $0.10 per GB and outbound range from $0.15 per GB to $0.08 per GB depending on volume	In bound $0.10 per GB outbound $0.15 per GB	Inbound $0.10 per GB Outbound $0.12 per GB	Inbound $0.10 per GB and outbound from $0.15 to $0.08 per GB depending on volume
Example Monthly pricing for data storage	$0.10 per GB plus $0.10 per million I/O request	Included in processing pricing	$0.15 per GB	$0.25 per GB

Table 4 Pricing Scheme's and Examples [26]

3.1.9 Data sizing

E-government and organization applications need to be certain about the data size in the database that will store the data. Furthermore, when outsourcing their data, the data sizing becomes a very important standard for selecting the service provider. Cloud database providers offer different sizes of storing data. For example, Amazon RDS offers the user with storage up to 1 TB of data for one database where Microsoft Azure SQL offers only 50 GB of data for a database, as shown in Table 3. Archiving is important key feature for e-governmental department and agencies applications, financial services and healthcare sectors, when using cloud computing environment as shown in Table 1. All these three sectors are required to keep record of their procedure and operations for later work. By result, all information and records must be archived and keep even this information is not presently require or used [35].

Attribute	Amazon RDS MySQL	Microsoft SQL Azure	Google Data store	Amazon Simple DB
Scheme Type	RDBMS	RDBMS	NoSQL	NoSQL
Amount of data for one database	1 terabyte per database	50 gigabyte per database	Not published for entire database, but 1 MB limit on a subset of data (called an entity) limits to the number of indexes	10 gigabytes per database domain (roughly equivalent to an RDBMS table)

Table 5 Cloud Database Provider – Data Sizing for One Database [23]

CHAPTER 4

FINDINGS AND CONCLUSION

4.1 Findings

Cloud computing, which is a new model for computing, is growing fast at the same time gaining attention not only by large organizations but also by academic, government and small and medium business organizations. It offers resources on demand with its three types of services models Software as Service (SaaS), Platform as Service (PaaS), and Infrastructure as Service (IaaS), and by its deployment models: the public cloud, private cloud and hybrid cloud. The findings of this thesis are

1- Adoption of cloud computing by organizations is low.

   Fears of the cloud largely arise from the perspective of losing the control of sensitive data. As a result, the adoption of cloud computing by organizations is low. Present control procedures do not perfectly pointing data management and storing by third party in cloud computing environment.

2- Traditional security concerns.

   This includes computer software and network security attacks. By migration the application and data to the cloud these security attacks will be more possible or easier for attackers. Cloud providers react to this issue by showing that their security procedures and actions are widely tested and grown than traditional client-server security procedures.

3- Data sizing is important.

   Data sizing is a critical condition for e-governmental applications and financial applications for archiving and safe records for later use. Even these records are not used currently.

4- Database Accessibility importance

In e-government, many departments are required to share information with each other. From this point of view, the type of database accessibility that a service provider offers is an important condition for e-government application collaboration.

5- Transaction capabilities and the importance of the Database Scheme

In cases of financial organizations, ACID capabilities which are provided by the CRDBMS scheme is a vital condition for supplying transaction process completion. While in the case of dealing with huge amounts of unstructured data, such as media and audio files, there is no need for ACID capabilities The NoSQL database scheme is preferable.

4.2 Limitations

1. Testing

The performance indicators were tested neither as simulation nor with an organization using the cloud database model.

2. Immature indicators

The performance indicators do not address all characteristics of cloud databases, such as portability, and they do not consider the infrastructure of the service provider, such as a network's hardware.

4.3 Future Work

For future work, we plan to do the following:

1. Extend the performance indicators

   By extending, we mean include more SaaS attribute such as usability, accountability into the performance indicators in order to evaluate better for service providers

2. Ranking.

   By ranking, we refer to assigning specific value for all performance indictors in order to determine the priority of performance indicators for the consumer.

In order to rank the performance indicators, we will use a ranking algorithm, such as the ranked voting system.

3. Develop tool.

   After Steps 1 and 2 are completed, a tool to collect the performance indictor's priority feedback from the consumer is built and programmed in order to recommend the type of service that the user should rent.

4.4 Recommendations

We need to define the authorizations, which basically declare who can access which resource under any mode. In order to control access requests, we recommend setting up the ***access control mechanism,*** which is a security service that interrupts each access request to the system and determines whether it has partial or total access depending on predefined specific authorizations. Additionally, we recommend using a digital signature to protect data from unauthorized auditing. Furthermore, for the evaluation process to select a cloud database service provider, we recommend including certificated service providers. Different service providers have various certifications; for example, Amazon has **SAS 70 TYPE II**, **ISO 27001,** whereas Azure has certifications of **Safe Harbor policy**, **ISO 27001**.

4.5 Conclusion

The cloud database is a new service delivery model that both the consumer and service provider benefit. As more service providers are established, the quality of evaluating services becomes a critical issue for consumers.

In this thesis, we presented performance indicators to evaluate cloud database service providers. We defined the performance indicators which are, for the most part, related to software as service depending on the Service Measurement Index (SMI), which is based on the International Organization for Standardization (ISO). Additionally, we add more performance indicators mostly related to the characteristics of cloud databases, such as data sizing and transactional capabilities.

The performance indicators cover most important features that are used to evaluate cloud database service providers in a way that enables the consumer to select a suitable service provider that meets the requirements.

Coming back to the research question *"How can an organization select a suitable cloud database service provider?"* we can say that our performance indicators can help consumers in evaluating services while selecting suitable cloud database service providers.

REFERENCES

1. **Sosinsky B.**, **(2011)**," *Cloud Computing Bible",* Wiley Publishing, Incorporation, Indiana, pp.25-26.
2. **Barnatt C., (2010),** "*A Brief Guide to Cloud Computing",* Robinson, London, pp. 62-90.
3. **(2014),** http://www.tutorialspoint.com/cloud_computing/cloud_computing_tutorial.pdf (Data Download Date: 05.07.2014).
4. **Baun C., Marcel K., Jens N., and Stefan T., (2011)**, " *Cloud Computing Web-Based Dynamic IT Services*", 2th Edition, , Germany, Spring Publishing, pp.5-7.
5. **Buyya R., Broberg J., and Goscinski A., (2011)**,"*Cloud Computing Principles and Paradigms"*, Wiley Publishing Inc., New Jersey, pp. 113-120
6. **Jacob B., Brown M., Fukui K., and Trivedi N., (2005),** "Introduction to Grid", IBM, pp. 3-12
7. **(2012),**"*Introduction to Big Data Infrastructure and Networking Considerations*", California, Unitaed State of America, pp.4-5.
8. **Zikopoulos P., and Eaton C., (2012),** "*Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data: Analytics for Enterprise Class Hadoop and Streaming Data",* Mc Graw Hill, pp. 3-8.

9. **Grance T., and Jansen W., (2011)***” The NIST Definition of Cloud Computing”,* National Institute of Standards and Technology. http://www.nist.gov/customcf/get_pdf.cfm?pub_id=909494 (Data Download Date: 20.07.2014).

10. **Ganesh Chanadra D., and Singh Bhadroia R., (2012),** *"Cloud Computing Model for National e-Governance Plan(NeGP),"* , IEEE, India, pp. 520 – 524.

11. **Hogan M., (2008),** *"Cloud Computing and Database: How databases can meet the demands of cloud computing”*, ScaleDB Incorporation, pp.6-12.

12. **Bloor R., (2011),** *"What is a Cloud Database ?"*, The Bloor Group, pp.4-7.

13. **Weis J,. and Alves Foss J., (2011),** *"Securing Database as a Service: Issues and Compromises,"*, IEEE, vol. 9, no. 6, pp. 49-55.

14. **Chow R., (2009),** *"Controlling data in the cloud: outsourcing computation without outsourcing control,"*,16th ACM Conference on Computer and Communications Security (CCS), US, , pp. 85-90.

15. **Nicholas Hoover J., (2013)**, *"Compliance in the Ether: Cloud Computing, Data Security and Business Regulation,"*, Journal of Business and Technology Law, vol. 8, no. 1, pp. 255-273.

16. **Armstrong S., (2014),***” A Close-up View of Microsoft Azure Adoption”,* Capgemini**.**, http://www.capgemini.com/resource-file-access/resource/pdf/azure_report_2014-06-05.pdf. (Data Download Date: 24.12.2014).

17. **Cattaneo G., Kolding M., Bradshaw D., and Folco G., (2011),** *"Quantitative Estimates of the Demand for Cloud Computing in Europe and the Likely Barriers to Take,"* International Data Corporation, Brussels, Belgium, pp.132-134.

18. **European Commission, (2012)** , *"Unleashing the Potential of Cloud Computing in Europe "* , European Commission Brussels, Belgium.

19. **Tiwari S., (2011),** *"Professional NoSQL,"*, John Wiley and Sons, pp.3-6.

20. **Ferrari E., (2009),** *"Database as a Service: Challenges and Solutions for Privacy and Security,"*, IEEE , Services Computing Conference, Singapore, , pp. 46 - 51.

21. **Li A., Yang X., Kandula S., and Zhang M., (2011),** *"Comparing Public Cloud Providers,"* , IEEE, vol. 15, pp. 50-53.

22. **Hurwitz J., Bloor R., Kaufman M., and Halper F., (2009)** " *Cloud Computing For Dummies*", Wiley Publishing Incorporation, pp.15-16.

23. **Pizette L. and Cabot T.,** (2012), *" Database as a Service: A Marketplace Assessment."*,http://www.mitre.org/sites/default/files/pdf/cloud_database_service_dbaas.pdf. (Data Download Date: 12.08.2014).

24. **Grance T. and Jansen W., (2011),** "*Guidelines on Security and Privacy in Public Cloud Computing"*, National Institute of Standards and Technology., www.nist.gov (Data Download Date: 25.06.2014).

25. **Armbrust M., (2009),** *"Above the Clouds: A Berkeley View of Cloud"*, Electrical Engineering and Computer Sciences University of California at Berkeley, California , Technical Report, pp.14.

26 **Li J., Sharma N. Kr., and Szekeres A., (2013),** *"Performance Analysis of Cloud Relational Database Services,"*, Washington Univesrity, pp 1-14.

27. **Xiao-hui L. and Xin-fang S., (2013),** *"Analysis on Cloud Computing and its Security "*, IEEE, 8th International Conference on Computer Science and Education, Srilanka, pp. 126-128.

28. **Batra S. and Chibber A., (2013)**, *"Preliminary Analysis of Cloud Computing Vulnerabilities,"* Journal of Engineering, Computers and Applied Sciences, vol. 2, no. 2319-5606, pp. 49-51.

29. **Singh S., (2014),** *"Cloud Computing Attacks: A Discussion With Solutions,"* Open Journel of Mobil Computing and Cloud Computing, vol. 1, pp. 1-8.

30. **Yoo Lee J., Woo Lee J., Wan Cheun D., and Dong Kim S., (2009)** *"A Quality Model for Evaluating Software as a Service in Cloud Computing,"*, IEEE, 7th

International Conference on Software Engineering Research, Management and Applications, China, pp. 312-311

31. **Kumar Garg S., Versteeg S., and Buyya R., (2011),** *"SMICloud: A Framework for Comparing and Ranking Cloud Services,",* IEEE, 4th International Conference on Utility and Cloud Computing, Melbourne, Australia, vol 29, pp.1012-1023.

32. **Elmasri R. and Navathe S. B., (2011),"** *Fundemental of Database systmes"* 6th edition. United States of America, Addison-Wesley, pp. 23-24.

33. *"Cloud Commons: Cloud Service Measurement Initiative Consortium",* http://www.cloudcommons.com/about-smi. (Data Download Date: 25.07.2014).

34. **Liang J., (2012),** *"Government Cloud: Enhancing Efficiency of E-government and Providing Better Public Service,",* IEEE, International Joint Conference on Service Science, Shanghai, pp. 261-265.

35. **Hanna E. M., Mohamed N., and Al-Jaroodi J.,(2012),** *"The Cloud: Requirements for a Better Service,",* IEEE ,12th International Symposium on Cluster, Cloud and Grid Computing, Ottawa, Canada, pp.787-792.

36. **Zheng X., Martin P., Brohman K., and Xu L. D., (2014),** *"CLOUDQUAL: A Quality Model for Cloud Services,"*,IEEE, Industrial Informatics, IEEE Transactions, vol. 10, no. 2, pp. 1527-1536.

37. **Al Shehri W., (2013),** *"Cloud Database : Database as Service,"* International Journal of Database Management Systems, vol. 5, no. 13, pp 1-12.

38. **Curino C., (2011),** *"Relational Cloud: A Database as a Service for the Cloud",* 5th Biennial Conference on Innovative Data Systems Research , Asilomar,Canada, pp. 3-4.

39. **Ramanathan S., Goel S., and Alagumalai S., (2011),** *"Comparison of Cloud Database: Amazon's SimpleDB"* , International Journal of Computer Science, Vol. 8, pp 243-246, India.

40. **Xiong Wen P. and Dong L., (2013),** *"Quality Model for Evaluating SaaS Service"*, IEEE, 4th International Conference on Emerging Intelligent Data and Web Technologies, pp 83-87, China.

41. **Beynon-Davies P., (2004) "***Database Systems*", Plagrave Macmillan, New York, United State of America, pp. 523-524.

42. "Security Guidance for critical Areas of focus in cloud computing", https://cloudsecurityalliance.org/guidance/csaguide.v3.0.pdf (Data Download Date: 20.12.2014).

43. **Sakhi I., (2012),** *"Database security in the cloud"* M.S. Thesis, Department of Computer Engineering**,** Royal Institute of Technology, Sweden, pp. 16-19.

44. **Tripathi A., and Parihar B.,(2011),** *"E-Governance challenges and cloud benefits,"*, IEEE, Computer Science and Automation Engineering Conference vol.1, pp. 351 - 354.

45. VMware http://www.vmware.com/virtualization/virtualization-basics/how-virtualization-works (Data Download Date 25.12.2014).

APPENDICES A

CURRICULUM VITAE

PERSONAL INFORMATION
Surname, Name: Abbas Hussein Ali
Date and Place of Birth: 5 August 1981 Kirkuk, Iraq
Marital Status: Single
Phone: +90 553 440 33 96, +964 770 104 6104
Email: abbasalsaka@gmail.com

EDUCATION

Degree	**Institution**	**Year of Graduation**
B.Sc	Kirkuk University	2006
D.Sc	Technical institute of Kirkuk	2002
High School	Secondary of Ata Begler	2000

WORK EXPERIENCE

Year	**Place**	**Enrollment**
2007	Kirkuk University College Science, Library	Manager
2008	Kirkuk University College Science, Department of The computer Science	Computer Lab.
2009-2012	The Presidency of the Kirkuk University, Planning and Follow-up Department	Employee

FOREIGN LANGUAGES:
Turkish, Arabic, English, Kurdish
HONOURS AND AWARDS:
1. Graduate within Honor Students 2006
   Kirkuk University
2. Graduate within Honor Students 2002
   Technical Institute of Kirkuk

HOBBIES:
Cooking, Travel, Photography.