SELECTING A SUITABLE DATABASE SYSTEM FOR KIRKUK UNIVERSITY

YALMAZ AL-KHANJI

JANUARY 2015

SELECTING A SUITABLE DATABASE SYSTEM FOR KIRKUK UNIVERSITY

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

BY
YALMAZ AL-KHANJI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF
MATHEMATICS AND COMPUTER SCIENCE

JANUARY 2015

Title of the Thesis: **Selecting a Suitable Database System for Kirkuk University.**

Submitted by **Yalmaz AL-KHANJI**

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 KAYMAKÇALAN
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.

Prof. Dr. Taner ALTUNOK
Supervisor

Examination Date: 20.01.2015

Examining Committee Members

Prof. Dr. Taner ALTUNOK	(Çankaya Univ.)
Assoc. Prof. Dr. Fahd JARAD	(Türk Hava Kurumu Univ.)
Assoc. Prof. Dr. Ersin ELBAŞI	(Ipek Univ.)

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: Yalmaz, AL-KHANJI

Signature :

Date : 20.01.2015

ABSTRACT

SELECTING A SUITABLE DATABASE SYSTEM FOR KIRKUK UNIVERSITY

AL-KHANJI, Yalmaz

M.Sc., Department of Mathematics and Computer Science

Supervisor: Prof. Dr. Taner ALTUNOK

January 2015, 59 pages

In this thesis, the situation for Kirkuk University was especially studied in order to change the current mechanism of managing the procedures based mostly on papers and manual work to that of another mechanism based on a database system. After studying the conditions of the University, we suggested and designed two models for two database systems that are adopted to be studied and compared in this thesis, namely the centralized database system and distributed database system. By using these models, comparisons are obtained between these two systems by applying three methods. The first method is to compare the general features for these two systems, the second by using three suggested queries as a simulation of the most used queries in the University, and the third by creating a questionnaire at the Presidency of the University and College of Agriculture, considering it as one of six colleges that communicates with the University campus via the WAN. Finally, the results of the three methods are analyzed in order to select a suitable database system for the conditions at Kirkuk University.

Keywords: Kirkuk University, Centralized Database, Distributed Database.

ÖZ

UYGUN VERİ TABANI SİSTEMİ SEÇİMİ KERKÜK ÜNİVERSİTESİ İÇİN

AL-KHANJI, Yalmaz
Yüksek Lisans, Matematik-Bilgisayar Anabilim Dalı
Tez Yöneticisi: Prof. Dr. Taner ALTUNOK

Ocak 2015, 59 sayfa

Kerkük Üniversitesinin özel durumunu inceleme ve kağıt ortamına dayalı iş takip sisteminin geliştirmesi için uygun veritabanı ortamının tespit etmeyi amaçlayan bir çalışmadır. Kerkük Üniversitesinin özel durumunu inceleyip en uygun veritabanı yöntemi tespit ederek üniversitenin tüm ihtiyaçlarını verimli bir şekilde karşılamaktır . Bu çalışmada iki veritabanı sistemi incelendi. Bu sistemlerin ilki merkezi veritabanı sistemidir. İkincisi ise dağıtık veritabanı sistemidir. Bu iki sisteminin Kerkük Üniversitesi ile uyumluluğu, üç farklı yöntem ile karşılaştırıldı. Kullanılan karşılaştırma yöntemlerin ilki seçilen iki farklı sisteminin özelliklerini ve Kerkük Üniversitesinin özel şartlarıyle uygunluluğunu ölçeceğiz. İkinci yöntem ise üniversitenin en çok kullanacağı sorgulamaları her iki sistemde simile ederek üniversitenin şartlarına uyumluluğunu ölçeceğiz. Üçüncü yöntem ise üniversitenin rektörlüğü ve ziraat fakültesinde anket uygulaması yapılacaktır. Ziraat fakültesinin seçme nedeni ise üniversitenin rektörlüğü ile WAN ortamında bağlanan altı fakültenin birisi olmasıdır. Bu üç yöntemin sonuçlarını kullanarak Kerkük Üniversitesine en uygun veritabanı yöntemi seçilecektır.

Anahtar Kelimeler: Kerkük Üniversitesi, Merkezi Veritabanı, Dağıtık Veritabanı.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Prof. Dr. Taner ALTUNOK for his supervision, special guidance, suggestions and encouragement through the development of this thesis.

It is a pleasure to express my special thanks to my mother for her patience of her being far from me for such a long time and for her wishing me God speed.

To my wife for her patience and who worked to accommodate herself to my duties so that I could dedicate the needed time to prepare this thesis.

To my brother Firnas Hanci for his advice at each stage when I consulted him and for his help in times of confusion.

To my friend Kameran Ali Ameen, who helped me to complete the required number of questionnaires that were used in this thesis.

To all the friends who provided me with advice throughout the preparation of this thesis.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES

LIST OF TABLES

TABLES

LIST OF ABBREVIATIONS

DBS	Database System
DBMS	Database Management System
DDBS	Distributed Database System
DDBMS	Distributed Database Management System
LAN	Local Area Network
WAN	Wide Area Network
ALP	Attribute Locality Precedence
VPP	Vertical Partitioning Problem
HF	Horizontal Fragmentation
VF	Vertical Fragmentation
MF	Mixed Fragmentation
IT	Information Technology
SOA	Service-Oriented Architecture
ESB	Enterprise Service Bus
ALP	Attribute Locality Precedence
CRUD	Create Read Update Delete
MCRUD	Modified Create Read Update Delete
GUI	Graphical User Interface
CPU	Control Processor Unit
RSA	Rivest, Shamir, and Adleman
DSA	Digital Signature Algorithm
SQL	Structured Query Language

CHAPTER 1

INTRODUCTION

1.1 Background

Databases are important in our lives and they can be used to be warehouses for all the data that we deal with in our life, from the smallest volume of data to the greatest. In addition, they provide us quick and easy retrieval of information in forms we may require. As a result, the negative effects of the absence of databases from any field in our lives is observed. For this reason, establishing a database for each foundation is necessary due to the absence of a database, not too dissimilar from Kirkuk University's need for a database to meet all its requirements.

In fact, each database needs software known as a database management system, whose very important role is to define, create, manage, protect, etc. a database [1]. The Database with Database Management System (DBMS) forms the Database System (DBS). The world variation in requirements led to the development of several types and variations of the DBS, such as the centralized database system, the distributed database system and the parallel database system. Each type has its particular advantages and disadvantages; however, the conditions and requirements for any project will require one of these systems to be adopted to meet the project requirements in a suitable form if not in an optimized form.

The following will introduce some previews of works which utilize some of these systems. For the centralized database system, the Palestinian e-Government system will be presented, and for the distributed database system, the system of the Centralized Bank of Iraq will be presented in addition to the Basrah University system and its distributed register system. Moreover, a model compound of two systems will present centralized DBS and DDBS, which were developed by the Michigan

Alzheimer's Disease Research Center. Finally, while the DDBs have three methods of distribution the database, horizontal, vertical and mixed [2], the last two works discuss the horizontal and vertical methods of DDBS respectively.

Rebhi S. Baraka and Suhail M. Madoukh et al. start their papers by explaining the Palestinian e-Government technical framework, which was adopted by the IT department (www.mtit.gov.ps) and the Ministry of Telecom. The framework is formed from four layers, each layer with a variety of components, one of which on the Data Access Layer is the Centralized database. They describe the Centralized database as a solution for integrating and sharing data among the various ministries; whereas the applied centralized database model is approved by the Palestinian Government Data Integration Committee(www.takamul.gov.ps).They specify the applied centralized database model as a broker which integrates data from a variety of sources. As a result, there is a dependence on database replication and synchronization techniques. In the paper, the SOA is proposed as a solution to achieve limitations that appears in the applied model, such as interoperability, manageability, undermining the flexibility, scalability, and governance of the centralized database. The proposed framework for the Palestinian e-Government (SOA-based centralized database structure) is based on the SOA, Web Services and ESB. Moreover, by the Centralized database Bus, interaction between their components is obtained [3].

Mayson Mohammad Talab and Dr. Bassam Ali Mustafa et al. designed and implemented a distributed database system, which was proposed for the Centralized Bank of Iraq. In the analysis step, they used different methods to collect data, such as questionnaires, personal interviews, observations, record searching and referenda, whereas the current system is described as several departments and as a group of separate systems with different languages. In the design, they used a top-down method. The two-tier client/server model was adopted as the architecture for the proposed system. Using twelve client computers and three hubs, to which the clients were connected by a local area network (LAN), the clients were spread over all offices with one hub on each floor. To distribute the database, they used the partially replicated technique with Windows XP as the operating system for the distributed system, which was designed to be homogeneous. In order to manage, connect and implement the database, *Oracle 10g* software was used. In addition, they utilized *Oracle Forms*

Builder 6i to design the user screen and *Oracle Reports Builder 6i* to prepare the reports. By applying passwords and user roles, the system was protected from unauthorized access. In addition, they used the export and import mechanism to protect data from loss [4].

Sana J. Alyaseri et al. start their paper by introducing an interview about the DDB, followed by the main author discussing *Oracle9i* terminologies. This is followed by Sana J. Alyaseri proposing a distributed registration database system for Basrah University, using the *Oracle9i* distributed features. Two campuses were formed for Basrah University, the first of which is located in the Basrah governorate and the other in the Mesan governorate. The Basrah campus is defined as all colleges and departments that are permitted to be added to both campuses such that the Basrah campus is described as a centralized campus. In the proposed design, the author relied on distribution techniques to distribute the database between these two campuses; therefore, she selects the technique for each case in the distribution, according to the requirements of the situation. For example, the tables of instructors, colleges, courses, and departments are identical for both sites; therefore, the tables will be replicated and placed into both sites. However, student information will be stored on the site of the campus to which the students belong. As a result, the author fragmented the student information tables, such as the registration table and teaching table. Finally, she explains in detail the steps of updating the distributed registration database system, starting from installing *Oracle9i*, and finishing by placing fragments of tables in a specific side [5].

Norman L. Foster, Eszter Gombosi, Cheryl Teboe and Roderick J. A. Little et al. introduce a study about a special database design which balances the centralized database approach with the distributed database approach by applying it in a clinical research environment. They refer to the Michigan Alzheimer's Disease Research Center, which produced a database design with a centralized entry and a distributed copy of the database over the network; for example, providing copies to investigators in their offices, laboratories and clinics. The developed design avoided the disadvantages in the clinical research environments, whereas the data management in the multi-center investigations needed great effort. As a result, the efforts are scattered, and conflict increases. However, with their developed database design, the updating of information is carried out on one site and the data is easily available for others. In the developed design, the collaborative research is promoted and integrates the advantages of distributed and centralized designs. The master copy for the database file is held in one core so that the copies of the back-up are archived and stored weekly. Over the campus Ethernet network, the database is distributed weekly to personal computers. A password is given to each user as protection for database access and to specify the accessible data for each user. Users can read the database and update it, but the updating will be temporary since it is overwritten to the personal sites by the master copy that is applied weekly [6].

Shahidul Islam Khan and Dr. A. S. M. Latiful Hoque et al. proposed in their paper a new technique for horizontal fragmentation in distributed databases, considering it as a solution for horizontal fragmentation. They proposed a technique that enables the partitioning of the relations in the initial stages of the distributed database. This is dependent on the attribute locality precedence (ALP) matrix in order to fragment the relations horizontally. The ALP refers to the degree of importance of an attribute according to the sites on which the distributed database placed. According to the cost functions and modified CRUD (Create, Read, Update and Delete) matrix, the ALP value is computed for all attributes belonging to the relations, followed by the ALP table being constructed. The CRUD matrix describes the data mapping to variety sites, whereas the MCRUD Matrix is formed from rows indicating the predicates for the attributes of a relation, and columns indicating the applications on the sites for the DDBMS. Finally, the predicate set P is produced for the attribute with a highest value in the ALP table, followed by relations using the produced predicated being fragmented horizontally [7].

Shamkant Navathe, Stefano Ceri, Gio Wiederhold and Jinglie Dou et al. in their paper discuss vertical partitioning of a relation into fragments. They specify three situations for applying vertical partitioning, with the third situation being the distributed database. For the vertical partitioning problem (VPP) in the distributed database, they solved it by applying two approaches. The first approach is described by placing an object in multiple sites and without any replication in the allocation. For that, they applied the MULTI-ALLOCATE-N algorithm recursively. For example, by applying the algorithm to object Q, the object splits into two blocks *U* and *L*. Then the algorithm repeats the same process with *L* and *U* if possible, as well as with all fragments if possible. Finally, the algorithm selects a pair of fragments with the least cost. When the object Q is divided into two parts *U* and *L*, the object Q is deleted, and then *U* and *L* are inserted to the algorithm. The dividing operation is repeated while there is a partition that can be split. The second approach is to solve the VPP, described by placing the object into multiple sites and with replication in the allocation. To do so, they used two algorithms. First they used the MULTI-ALLOCATE-N algorithm, after which they applied the MULTI-ALLOCATE-R algorithm. The second algorithm was applied to each fragment resulting from the first algorithm, so as to set the new copy of the fragments on the sites which do not have a new copy of the fragments [8].

It can be observed in the cited preview works that there is work regarding central DBS, DDBS, the combination of two types of database in addition to works about distributed techniques. However, we do not have works about comparing both types of systems, especially comparisons obtained under specific conditions. Therefore, we introduce in this thesis a comparison between a central database system and a distributed database system for Kirkuk University.

1.2 Motivations

The main motivation and aim of this thesis is informed by the reality of the conditions at Kirkuk University, which include the University lacking computerization systems for storing most of its data, procedures, jobs, etc. and the need for exchanging any required data between different sites in the university.

Computers are used in the university to carry out work such as the preparation of official books within the various units and departments. However, there is no communication between the units within a single college in order to exchange information via a network. This is mainly due to there being no database system covering all colleges and departments of these colleges and no links between them to facilitate the exchange of data, as well as meeting all requirements in a suitable form for the conditions at Kirkuk University.

1.3 Purpose of the Thesis

This thesis introduces a study on a centralized database system, a distributed database system and the conditions and requirements of Kirkuk University in order to offer recommendations as to which database system would be suitable for Kirkuk University conditions.

1.4 Research Question

Which database system is suitable for Kirkuk University conditions: *the centralized database system* or *the distributed database system*?

1.5 Organization of the Thesis

This thesis contains five chapters covering all that is needed to be known regarding a suitable database system for Kirkuk University.

Chapter 1 presents an introduction to database system with preview works.

Chapter 2 presents the centralized DB and DDB systems.

Chapter 3 compares the centralized DBS and DDBS systems bearing in mind the conditions at Kirkuk University.

Chapter 4 presents the results for suggested queries and the questionnaire.

Chapter 5 presents an analysis of results, future work and a conclusion.

CHAPTER 2

CENTRALIZED AND DISTRIBUTED DATABASE SYSTEMS

2.1 Database System

In this section, we discuss databases, database management systems and concepts related to database management systems.

2.1.1 Database

A database can be defined as a collection of related data which allows users to retrieve desired information efficiently. It can be any collection of data, such as student names, phone numbers and addresses and salaries, or a simple collection of images, videos, sounds as well as complex collections. Such databases can be in a computerized form or a non-computerized form; for example, non-computerized forms include dictionaries and TV guides and so on. Examples of computerized forms include employee registers, salaries, a data bank and sales transactions [9].

Five issues that establish a set of basic principles of the database concept [10]:

1. Building an environment that can contain all the company’s data. The main feature is the ability to access these data from inside and outside the company.
2. The ability to integrate and store data without redundancy. However, this feature is available only in the centralized database approach.
3. To have the ability to store data representing an entity which participates in several relationships without becoming redundant or experiencing any structural problems.
4. Foundation of an environment which manages certain data controls, such as the concurrency control, backup and recovery, and data security.
5. Foundation of an environment which allows for a high degree of independence of data.

2.1.2 Database management systems

A database management system is a piece of software that is designed to enable users to create and maintain a database. The DBMS is a general-purpose software system that facilitates the processes of defining, constructing, manipulating and sharing databases among various users and applications [1].

2.1.3 Concepts related to the DBMS

The definition of a database is specifying the type of data, structures and constraints for data which are stored in the database. In addition, the definition of data and the description of information is stored in the database using the DBMS in the form of a dictionary or database catalogue. Construction is a process of storing data in a storage medium that is controlled by DBMS. Manipulation consists of functions such as querying data, representing a mini world or generating reports from the database. Sharing a database permits multiple users to access the database synchronously. An application program accesses the database by sending a request for data or queries to the DBMS, which results in the retrieval of data. Transaction is the reading or writing of data in the database. Protection consists of system protection from software or hardware crashes, and security protection prevents unauthorized access to the database. Any database during its life cycle is maintained, and the DBMS allows for the system to develop and change as required. The database and database management system together form the database system. Figure 1 illustrates some of these concepts [11].

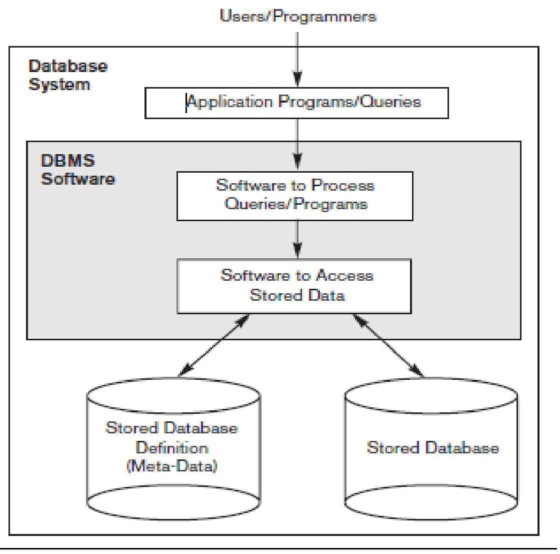

Figure 1 Simplified database system environment

2.2 Centralized Database System

In this section, we explain the definition of a centralized database system, centralized and client/server architectures and centralized database system advantages.

2.2.1 Centralized database system definition

A centralized database has all its data in one place. It is completely different from the distributed database, which has data in different places. In the centralized database, all the data reside in one place and as a result, problems such as bottlenecks can occur, and data availability is not efficient as in the case of the distributed database [12], a database with DBMS formed from a DBS [13].

2.2.2 Centralized and client/server architectures

A client and server are computer systems. The client sends a request to the server and the server receives it, processes it and returns the result or any information to the client. The server and client are almost always found in different locations. A client computer has an interface that enables end users working on the client side to use a server containing the database system. Examples of servers include web servers, application servers, database servers, etc. [9].

All client/server architecture can be implemented in one of the two approaches. The first approach is known as two-tier architecture, in which an application program and the user interface are run on the client side and the database system is found on the server machine. On the client machine, the application program invokes the DBMS running on the server machine. Figure 2 shows the two-tier architecture [14].

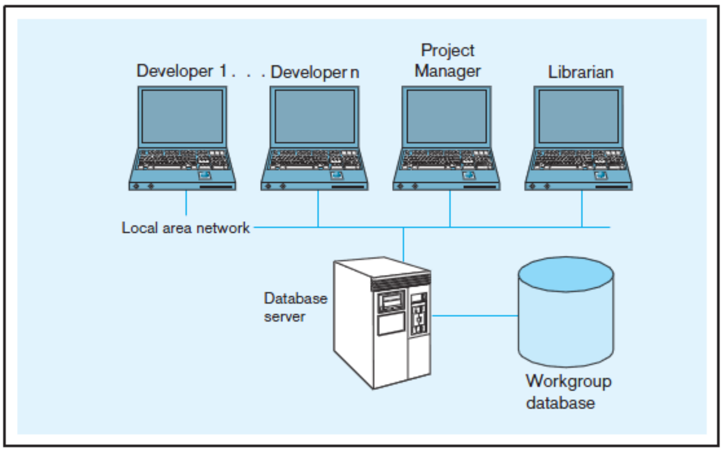

Figure 2 Two-tier database with local area network

The three-tier architecture is the second approach to implement the client/server architecture. This approach simply adds an intermediate layer known as the application server. Between the server layer and the client layer, this server connects the client machine to the database server and contains the mechanism work rules (constraints and procedures) used to access the database server. Whenever the client sends a request, the application layer checks the request and then forwards it to the database server side, thereby raising database security.

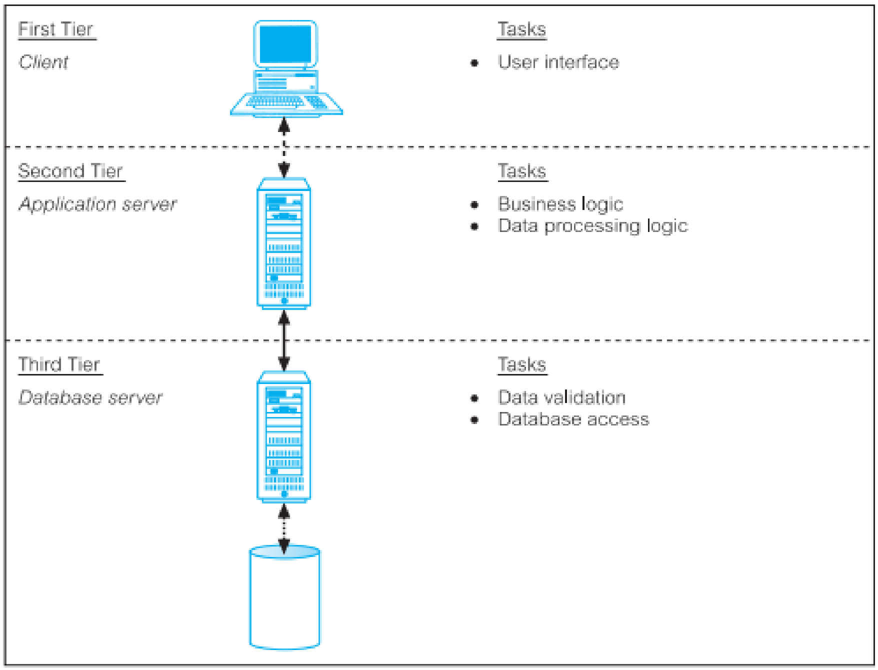

Figure 3 Three-tier architecture

Again, when the client sends a request for information to the application layer, the request is processed in the application layer, followed by the application layer preparing the appropriate database commands and sending it to the database server. Then the result is sent back to the application layer and is finally converted into a GUI format and presented to the client. Figure 3 shows the three-tier architecture [13].

2.2.3 Centralized database system advantages

The advantages of the centralized database system are explained as follows [15]:

1. Reduced Redundancy: When the design of the database depends on a good plan, it permits us to solve the issue of similar or duplicate data, where these data are used in a variety of files for a variety of applications. The solving is carried out by combining all duplicates and similar data and then storing it only once.

2. Improved Availability: Through the database management system, the information may be present for any application program.

3 Reduced Inconsistency: When several sites store identical data, updating on one site but not on another leads to inconsistency in the system.

4. Enforced data security: Because all data is stored centrally, the authentications to utilize the information are also centralized.

2.3 Distributed Database System

In this section, the distributed database system definition, distributed database management system architectures, distributed database and client/server systems, types of distributed database management and distributed database design techniques are explained.

2.3.1 Distributed database system definition

A *distributed database* is a collection of multiple, logically interrelated databases distributed over a computer network. A distributed database management system (DDBMS) is the software system that permits the management of a distributed database and makes the distribution transparent to users. The term distributed database system refers to the combination of the distributed database and the DDBMS [16]. Figure 4 illustrates a distributed database environment.

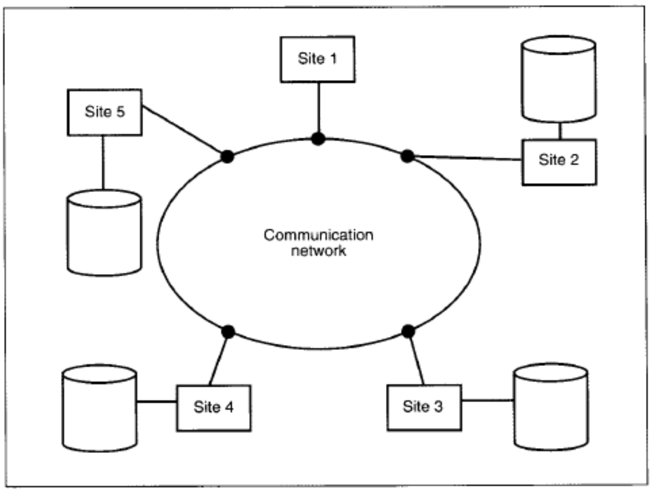

Figure 4 Distributed database environment

2.3.2 Distributed database management system architectures

The distributed database management system architecture can be classified according to three dimensions: autonomy, distribution and heterogeneous. (See Figure 5)

Autonomy indicates the distribution of control among sites, while the *distribution* refers to the physical distribution of data among sites, and the *heterogeneity* refers to the degree of dissimilar components among the DBMSs which communicate in the distributed database system [17].

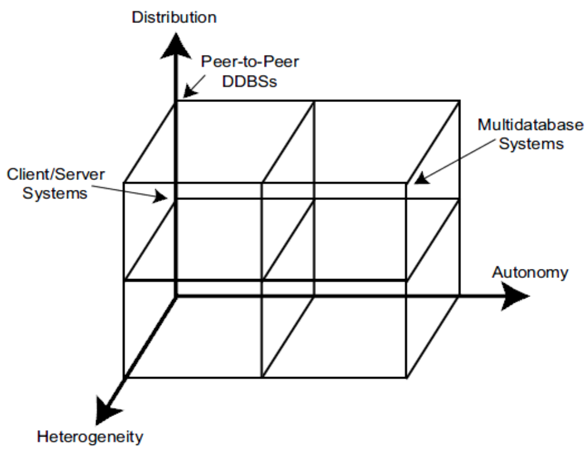

Figure 5 Implementation alternatives

2.3.3 Distributed database and client/server systems

Client/server architecture for DBMS is represented in a simple idea. By determining the DBMS functions and dividing them into two levels, we see that the first level is client functions and the second level the server functions, thereby leading us toward the idea of the "two-tier architecture." This manages the complexity of the distribution and the complexity of modern DBMS easily. In this architecture, the client requests services and the server answers the requests. The server carries out most of the data processing. Therefore, the query processing, optimization, and any other transaction is done on the servers. On the other hand, the client, despite having the user interface and the application, also holds a DBMS client module which manages the data that arrives at the client. (This architecture is depicted in Figure 6) With an uncommon case for the client, can placed in it consistency checking, which requires the replication of the system catalog and placing a copy of it on the client side.

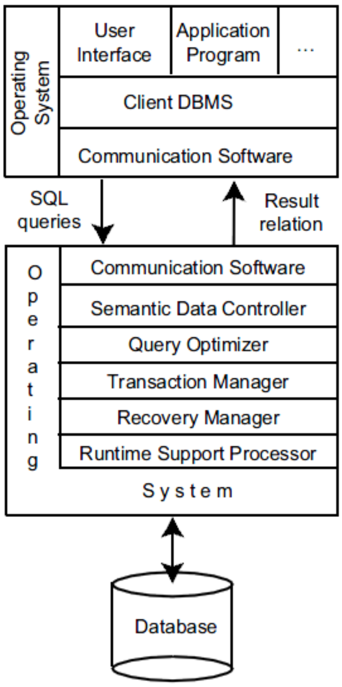

Figure 6 Client/Server reference architecture

Client/server architectures can be formed using many approaches depending on the machines which form them. When there are multiple clients and a single server, it is called a multiple client/single server approach. The latest approach, from the data management scene, is to use the approach of the centralized database system, where all data are stored on a single site. The more advanced client/server architecture is the multiple clients with multiple servers, on which the database is distributed over multiple servers. This is known as the multiple client/multiple server approach. The extension in the client/server architecture led to an improvement of the distributed function such that is is now more efficient, and on various types of servers. For example, the application server run the applications programs and the database servers run the DBMS functions, which leads to an architecture in the distributed client/server model that consists of three levels, known as the "three-tier distributed system architecture." Therefore, this architecture, with its multiple database servers and multiple application servers, leads us toward what is called the "n-tier distributed approach." In this case, the database is distributed among database servers, and the application server is used for one or more applications, as shown in Figure 7.

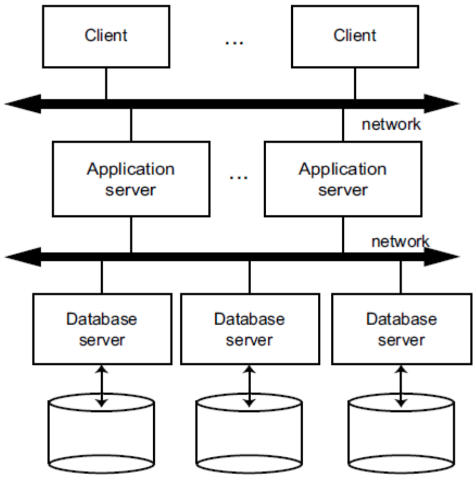

Figure 7 Distributed database servers

2.3.4 Types of distributed database management systems

The distributed database management system may be classified as either homogeneous or heterogeneous depending on the similarity of the components using it. All sites in a homogeneous database use the same DBMS product, which is easier to manage and design. However, heterogeneous databases may use different DBMS products, which is the only option when sites are required to communicate with a different DBMS [13].

2.3.5 Distributed database design techniques

Distributed database design techniques are detailed in three parts as follows:

A-Data fragmentation: Fragmentation means the division of one relation into two or more smaller relations or fragments in order to store them on several different sites. Thus, the message transmission cost is minimized. Re-combining the fragments will give us the original relation. As a result, the data will be near to those who use it but separate from other data being used by other end users. In addition, local optimization increases. Fragmentation may be horizontal, vertical or hybrid/mixed [18].

1- Horizontal Fragmentation (HF): meaning partitioning relations into separated tuples. A set of tuples forms a horizontal fragment. All sites must hold data that can be queried from it. Moreover, these data are in fragment form, thereby leading to an increase in the queries' performance. For example, the PROJ1 is divided into three fragments, as shown in Figure 8 [17].

$PROJ_1$

PNO	PNAME	BUDGET	LOC
P1	Instrumentation	150000	Montreal

$PROJ_2$

PNO	PNAME	BUDGET	LOC
P2	Database Develop.	135000	New York
P3	CAD/CAM	250000	New York

$PROJ_3$

PNO	PNAME	BUDGET	LOC
P4	Maintenance	310000	Paris

Figure 8 Horizontal fragmentation

2- Vertical Fragmentation (VF): meaning partitioning the relation into separated attributes. The set of attributes forms a vertical fragment, wherein each fragment contains the primary key attribute of the original relation. The original relation is constructed with the rejoined partitioned fragments. In the vertical fragmentation, many applications can run on one partition, as shown in Figure 9 [19][20].

Eid	Fname	Lname	Site	Eid	Pos	Salary
Fragment1				Fragment2		

Figure 9 Vertical fragmentation

3- Mixed Fragmentation (MF) can be obtained by two methods. The first method is to apply the horizontal fragmentation to the result of vertical fragmentation, followed by the horizontal fragmentation being applied to only one fragment prepared by vertical fragmentation. (See Figure 10) The second method is to apply the vertical fragmentation to the result of the horizontal fragmentation, followed by the vertical fragmentation being applied to only one fragment prepared by the horizontal fragmentation, as shown in Figure 11 [21].

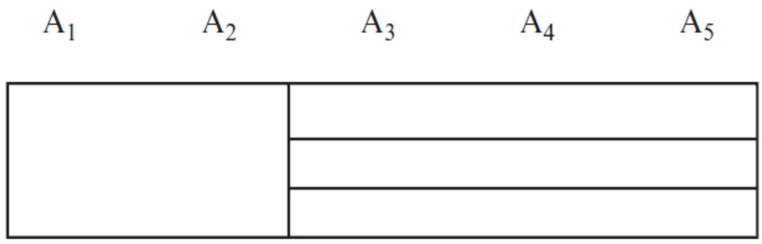

Figure 10 Vertical fragmentation followed by horizontal fragmentation

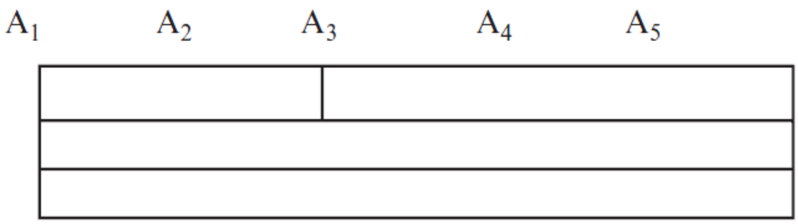

Figure 11 Horizontal fragmentation followed by vertical fragmentation

B- Data Replication and Allocation

There are three types of data replication [13]:

1- Fragmented (partitioned): With the first strategy of the replication, also known as no replication strategy, the database is partitioned into disjoint parts (fragments) with each fragment being placed into one site. The locality of the references are high with each data item placed into a site close to where they are used very frequently. The cost of storage is low, with no replication., Additionally, the availability and reliability is low, in spite of the fact that it is still higher than the centralized situation. This is due to the fact that if a site is down, data loss will only occur one site. Furthermore, with the appropriate design for a distributed database, the communication cost will be low and performance will be good.

2- Complete replication: With the second strategy, the availability and reliability is maximized in addition to the performance and the locality of reference remaining the same due to there being a complete copy of the database at every site. However, there are also disadvantages, such as the cost of communication and storage being high.

3- Selective replication: In this strategy, the two previews strategies, fragmentation and replication, are combined with centralization. Therefore, will be replication for data items that's require to use in many sites but not frequently updated, and fragmentation for some data items to get the high locality of reference, and centralization for the others. Thus, the aim of this strategy is to combine the advantages for all the preview strategies, but without any of their disadvantages. Because of the flexible features in this approach, it is the most commonly used.

Synchronous versus asynchronous replication: When the source data is updated, the replicated data is updated immediately. This type of replication is known as synchronous replication. While this mechanism may be suitable for environments that, by necessity, must keep all replicas fully synchronized (such as for financial applications), it also has several disadvantages. For example, a transaction will be unable to be fully completed if one or more of the sites that hold replicas are unavailable. Furthermore, the number of messages required to coordinate the synchronization of data places a significant burden on corporate networks.

Asynchronous replication is an alternative mechanism to synchronous replication. With the asynchronous mechanism, after modifying the source database, the target database will update. The delaying in regaining consistency may occur from a few seconds to several hours, or even days. However, the data eventually synchronizes to the same value at all sites [13].

2.3.6 Advantages of the distributed database system

Many organizations prefer the distributed database management system for many reasons. Below is a list with some advantages for distributed database system [11].

1- Improved the ease and flexibility of application development: When we can easily and flexibly develop and maintain any application distributed geographically, we have transparency of control and distribution.

2- Increased reliability and availability: Increasing these two features (availability and reliability) can be understandable when isolating a failed site from other sites without affecting those other sites. In a distributed database system, the DDBMS and the data are distributed over several sites and when a failure appears in a site, only that particular site will be unusable while the others continue to function. In a centralized database, when a failure occurs at the centralized site, the entire system will be unavailable. On the other hand, when a failure appears in a site in a distributed database, only information at that site becomes inaccessible. In addition, the user is not affected by the failed site if the information at the failed site has been replicated previously.

3- Improved performance: For a database fragmented by the distributed database management system, a data item is placed close to where it will be used, which leads to data localization, thereby reducing the competition for the input/output and CPU services simultaneously reducing the access delay in wide area networks. When distributing a larger database into a smaller database so as to be present on different sites, the local transactions and queries will have high performance because the existing database is smaller.

4- Easier expansion: The expansion in the distributed database is easier when increasing the database size, adding more data or appending more processors.

CHAPTER 3

COMPARISON BETWEEN A CENTRALIZED DATABASE AND DISTRIBUTED DATABASE IN LIGHT OF KIRKUK UNIVERSITY

3.1 Introduction

A database system provides the ability to store great amounts of data in addition to retrieval and indexing methods for these data, all of which is considered important today. When work is carried out without database programs, the required effort, cost and time is similar to how any organization manages its data without using computers, as is the case at Kirkuk University, which has a limited range of use of computers without using special programs, such as database programs for the integration of data. Therefore, in such cases, all work is carried out manually by people. Additionally, work will proceed slowly in addition to the natural factors affecting paper medium storage.

In fact, several types of database system and in order to select a suitable system, we need to study and analyze the specific conditions at Kirkuk University.
Therefore, in this chapter we present a study on Kirkuk University and its conditions which will provide the necessary information we need to select and adopt a specific database system. Next we show a current dealing mechanism with some type of documents. In addition, the suggested models for the two database systems for Kirkuk University are introduced. Lastly, we compare these two systems using the suggested models for the conditions at Kirkuk University. This comparison is carried out using three methods.

The first method is a general comparison between the features of these systems. The second is the use of suggested queries that are amply used. The third method is carried out by creating a questionnaire in order to the adoption of a suitable database system from these systems.

3.2 Analysis of the Current Situation at Kirkuk University

Before making a decision to adopt a more practical database system for the conditions at Kirkuk University, it is necessary to analyze the current general situation of the university in order to acquire a comprehensive idea about the matter. The gathering data operation is conducted by searching in order to acquire static and general information. Personal interviews with financial officers and administrators at the university are conducted to determine the procedures and data which uses repetition in addition to administering a questionnaire in the Prsidency of the University and the College of Agriculture to obtain information that directly or indirectly affects to the communication between the Presidency and the colleges via the Internet.

We can say clearly that Kirkuk University consists of a Presidency of the University and ten colleges. The Presidency and four of the colleges are located in one place, namely the University campus, found in the southernmost part of the city. There are two colleges, the College of Agriculture and College of Medicine, which are located in the northernmost part of the city. The distance between the College of Medicine and the Kirkuk University campus is 10.60 kilometers. The distance between the College of Agriculture and the University campus of is 15.30 kilometers. The distance between the College of Medicine and the College of Agriculture is 7.30 kilometers. The remaining four colleges, the College of Nursing and the College of Pure Science, are close to one another in the city center. The College of Management and Economics is also close to the College of Veterinary Medicine in the city center, but far from the other two colleges.

The Presidency of the University consists of administration departments which contain the Administrative Affairs Department, the Financial Affairs Department and the Legal Affairs Department. Additionally, the scientific departments contain the Scientific Affairs Department, the Computer and Internet Center and the Education Methods Center. Moreover, there are other departments such as the General Relational Department, the Central library, etc. The colleges contain a similar composition of departments.

The following sections present the designing models for database systems which focus on the Presidency of the University as a place on the main Campus of Kirkuk University in the south of the city and on the colleges that are located in the middle and north of the city.

In this thesis, we take into consideration the more frequently used data and disregard other less used data. Therefore, we utilize some of the procedures of the administrative and financial departments, which are considered to be important departments in both the Presidency of the University and the colleges.

3.2.1 Mechanism work at Kirkuk University

The current mechanism for dealing with documents such as official books and the sending official books from a college to the Presidency of the University can briefly be carried out in five steps, as shown in Figure 12:

1- Initial Acceptance: The department manager manually provides a request in writing to the Administrative or Scientific Assistant, who transfers it to the General Manager in order to obtain initial acceptance.

2- Final Acceptance: After printing two copies of the document, the manager of the department transfers it to the Administrative or Scientific Assistant, who transfers it to the General Manager to obtain final acceptance.

3- Issue: The two copies of the document are taken to the Administrative Department to have a date and number issued.

4- Save a copy: One of the two original copies is saved in the Administrative Department and the other is prepared for sending.

5- Sending an original copy: An original copy is given to a trusted employee to transfer it to the target side.

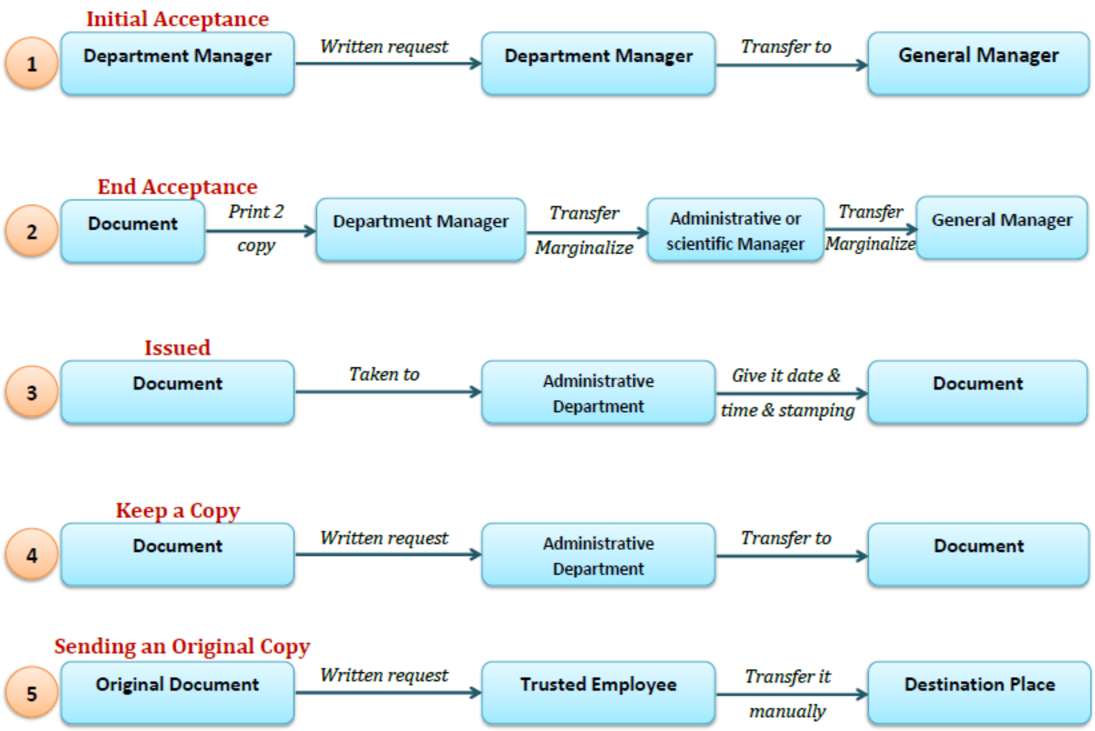

Figure 12 Steps of sending an official book from a college to the Presidency of the University.

As stated previously, one can see that all steps take at least one day. However, if we use database systems and digital signature techniques, it will take at most one hour. The Iraqi government has adopted many standards, and particularly among these standards are standards for digital signatures, such as RSA and DSA [22]. This implies that Kirkuk University can computerize all its work, including the preparation and transfer of official books. However, the efficiency of computerization for Kirkuk University requires the selection of a suitable database system to store and manage all jobs.

3.2.2 Using programs at the university

With regard to the programs used at Kirkuk University to carry out general work, a variety of Microsoft Office programs are used, including *Microsoft Office Word*, so that most of the departments can prepare and print official books. *Microsoft Office Excel* is used by the Financial Department to calculate and prepare salaries. *Microsoft Office Access*, which is used to create small databases, can be found in the Student

Affairs Department. The main point is that despite the fact that there are various programs used in various departments, each department functions without available interdepartmental communication or the sharing of data between them in order to ease the workload, unless it is done so manually, as is the case also between the Presidency of the University and colleges.

3.2.3 Frequent jobs and data

We can say that after searching and studying in the Presidency of Kirkuk University and in some of the colleges, specifically Administration and Financial Departments. The frequency of jobs in the Administrative Department in the Presidency of the University and colleges is represented by the issuing of official books and administrative orders in addition to the administrative orders for employees' bonuses, the formation of committees, saving copies of these documents to the department archive, sending other copies to targeted places, managing employee and lecture files, preparing seasonal and yearly reports for the General Manager and answering official books. Other jobs include the preparation of the employees' list for those deserving their sixth-monthly upgrading. This list is sent from colleges to the Presidency of the University for inspection and confirmation by central committees. The jobs of the Financial Department include the preparation of monthly salaries, preparing monthly and yearly balances and sending them from colleges to the Presidency of the University.

Therefore, after checking and analyzing all that has been mentioned previously, it is clear that most of the jobs for any college are conducted within the same college without being sent to the Presidency of the University. The same situation applies for the Presidency, with little data being transferred between the colleges and the Presidency of the University. The main part of the transferred data is represented in official books which are issued to colleges or the Presidency of the University. Currently, this is carried out manually.

After searching and investigating in order to determine the data for the main jobs that are processed frequently at Kirkuk University, the results from personal interviews with the Financial and Administrative Departments show that most of the data during one month revolves around the employees and lectures.

3.3 Designing database models

In designing our database model, we cover the data which are the most updated and dynamic. In addition, we take into consideration the locations of the colleges and distribute them on a geographic map so as to select suitable data for suitable places. This led to the building of a model that reflected the reality at Kirkuk University. When considering the colleges located on the Campus of the University, it can be seen that they communicate with the Presidency of the University and with each other via LAN. However, the more distant colleges communicate with the University campus via WAN. The main point here is that the WAN is not established in Kirkuk governance except as an Internet service, especially after abandoning landline phones.

The current manner of managing jobs at Kirkuk University prioritizes one of the two database systems which are under study. These two systems are the centralized database system and the distributed database system . To clarify, we designed a model for each of these two systems to simulate the Kirkuk University case in two systems.

3.3.1 Centralized database system model

For the design of our database model, we adopted the client/server method with a multiple clients/one server topology and a three-tier model for a centralized database. The main server is located in the Presidency of Kirkuk University on the University campus and the computers that are available in the Presidency of the University or in the colleges are identified as clients. These computers send requests as query data, ask about reports, updating, etc. The server provides the appropriate answers to these requests. an application server functioning as a mediator is placed between the server and the clients [11][23]. Figure 13 shows the client/server system indicating the location of the database.

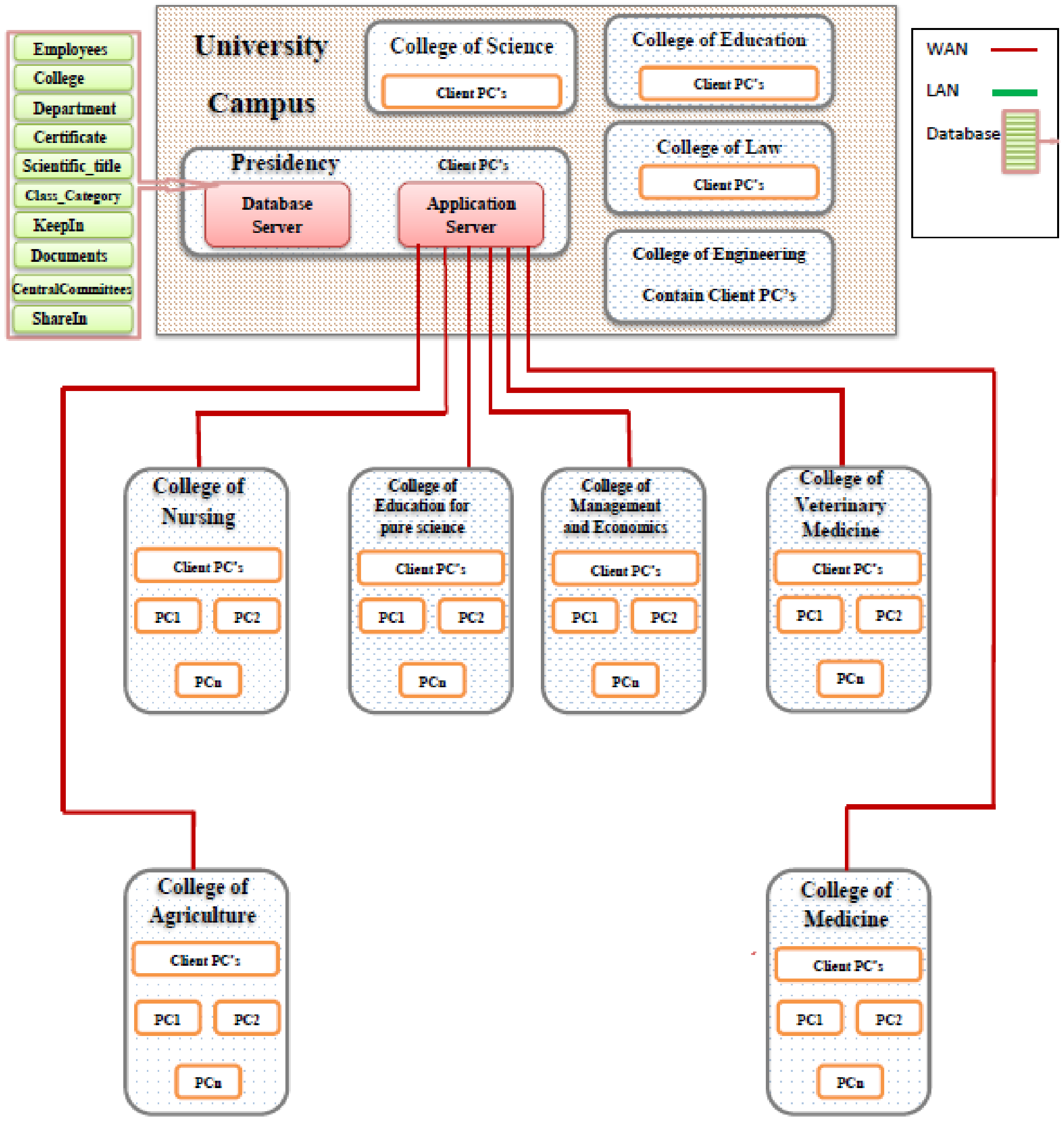

Figure 13 Centralized database model for Kirkuk University.

The designing database operation passes through three stages [11]. These three stages are detailed as follows:

1- Conceptual design: The conceptual design for a database in the centralized method for Kirkuk University is based on subjects that have been determined previously and taken into consideration during the design of any model. These subjects are limited to several entities, as follows: (Employees, Departments, Colleges, Certificates, Scientific title, Degree and Categorize for employee, Central Committee, Documents). This is followed by identified relations in order to communicate these entities to simulate the reality of communication among these entities for the Kirkuk University case. (See Figure 14)

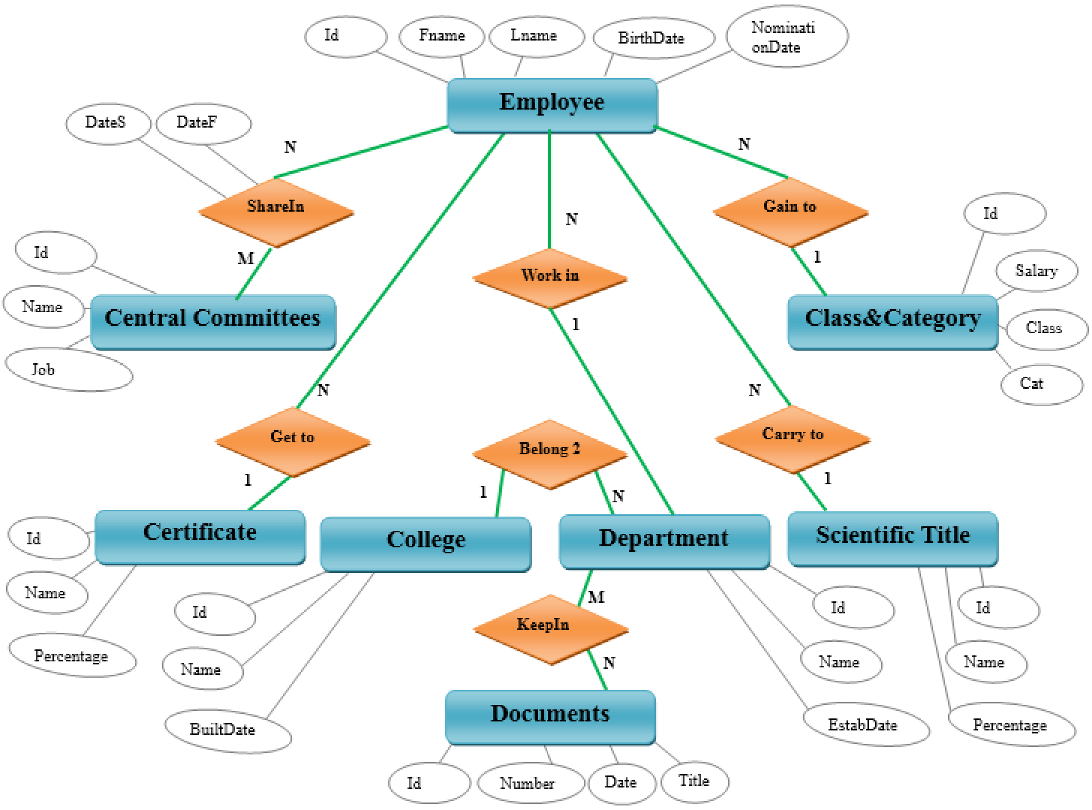

Figure 14 Conceptual design for the centralized database model for Kirkuk University

2- Logical design: The logical design is designed by converting the conceptual design into tables representing all entities and relations that are available among them [11], as shown in Figure 15.

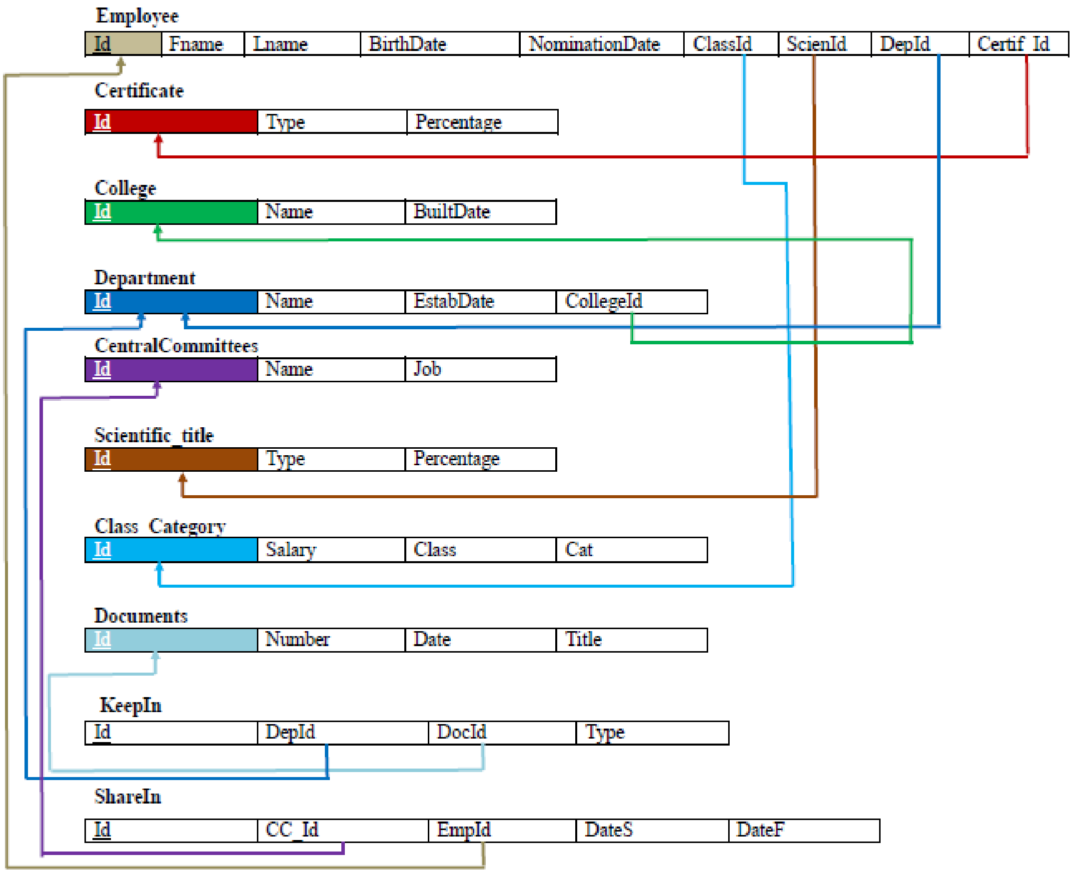

Figure 15 Logical design for the centralized database model

3- Physical design: This thesis does not introduce a physical design for our models. However, it does indicate what occurs in the physical design. These tables are created using the SQL language followed by data being entered into these tables. The main feature in the centralized database is that all tables are located on one server, and in the case of Kirkuk University, this server is found in the Presidency of the University. Therefore, the colleges placed on the Kirkuk University campus will communicate with the server via the LAN while the other colleges communicate via the WAN (i.e., the Internet).

3.3.2 Distributed database model

Before designing the database model in the distributed method, we specify the places that data will be available and which data will be available so as to meet university requirements while considering the geographical distribution for the college and the amount of transferred data among these sites. Therefore, the client/server system, the multiple client/multiple server topology and the three-tier model were adopted in the distributed database model design. To meet all the requirements, the efficiency of the five database servers will be specified.

In order to serve all clients that are available in the Presidency of the University and colleges, a server is placed on the Campus of the University inside the Presidency of the University. The second and third servers are placed in the College of Agriculture and the College of Medicine; the fourth and fifth servers are placed in the middle of the city in two different locations. For each of the two colleges in the city center, there is a server nearby. With each database server, an application server is also installed to function as the mediator between the servers and clients. The purpose of all of this is to have the user in any place feel that all of his needed data are easily and readily available without feeling the reality of the distribution of the data through the several servers located in different places [11][23]. Figure 17 shows the servers, clients and locations of the available tables depending on the database servers. Moreover, there are three stages for the design of a distributed database model as follows:

1- Conceptual design: From the definition of a distributed database, the database is joined logically and distributed physically. Therefore, the conceptual design used in the centralized database is identical to the distributed database but is known here as a global conceptual design. This represents the global logical joining for distribution data in different places as found in the suggested distributed model for Kirkuk University. However, in the local places containing a server, there is another conceptual design known as the local conceptual design which indicates to the database that which is available in it and the relations among the tables [17]. Therefore, the difference can be seen between the global conceptual design and the local conceptual design for the College of Agriculture, as seen in Figure 14 and Figure 16 respectively.

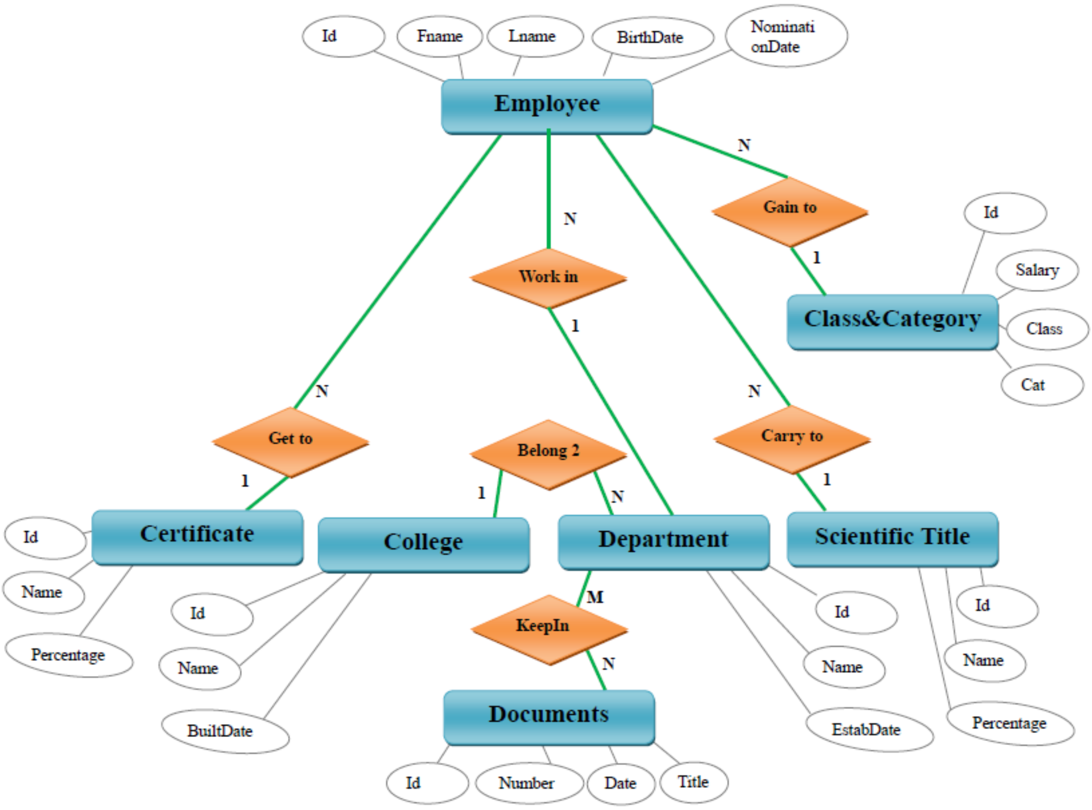

Figure 16 Local conceptual design in the distributed model for the College of Agriculture

2- Logical design: The logical design is a simulation for each local conceptual design in the distributed model by using some distributed database techniques in designing the operation, such as partitioning and replication. For example, all tables that contain constant data that seldom change are needed frequently by each college are replicated, such as the Certificates, Science_title, Class&Category and Colleges tables.

Another technique is the partitioning technique, which divides tables, into groups of rows that are required frequently in one college and required rarely from other colleges. An example includes employees' tables such that partitioning is carried out horizontally by putting rows belonging to a college on a server near to that particular college. The same approach is used for Department, Documents and KeepIn tables. The results of applying these techniques are seen in Table 1. By using an asynchronous replication method in order to avoid power or Internet outages on a site while updating data on another site, the update is generalized to all sites that share in the replication group at a later time [13].

It is clear that the logical design in the two suggested models (centralized and distributed) will not change in the structure of tables unless the number of tables is a result of replication or partitioning without change in the number and type of fields since it is partitioned horizontally.

The partitioning and replication techniques are obtained based on the requirements of the Presidency of the University and colleges. The tables needed frequently by the Presidency of the University and nearby colleges are placed on the server located in the Presidency of the University. The tables needed by the distant colleges are placed on servers near to them. After this distribution of the database, each college as well as the Presidency of the University acquires their most needed daily data via LAN.

T	Tables	Presidency's Server	Far Six Colleges' Servers	Type of Technique
1	Employees	∃	∃	Horizontal Fragmentation
2	Class_Category	∀	∀	Replication
3	Scientific_title	∀	∀	Replication
4	Department	∃	∃	Horizontal Fragmentation
5	College	∀	∀	Replication
6	Documents	∃	∃	Horizontal Fragmentation
7	Certificate	∀	∀	Replication
8	CentralCommittees	∀	∄	------------
9	ShareIn	∀	∄	-----------
10	KeepIn	∃	∃	Horizontal Fragmentation

Table 1 Distribution Tables among Used Servers in the DDB Model

3- Physical design: In the physical design, the tables are created and the data are inserted into each one. Figure 17 shows the distributed database model.

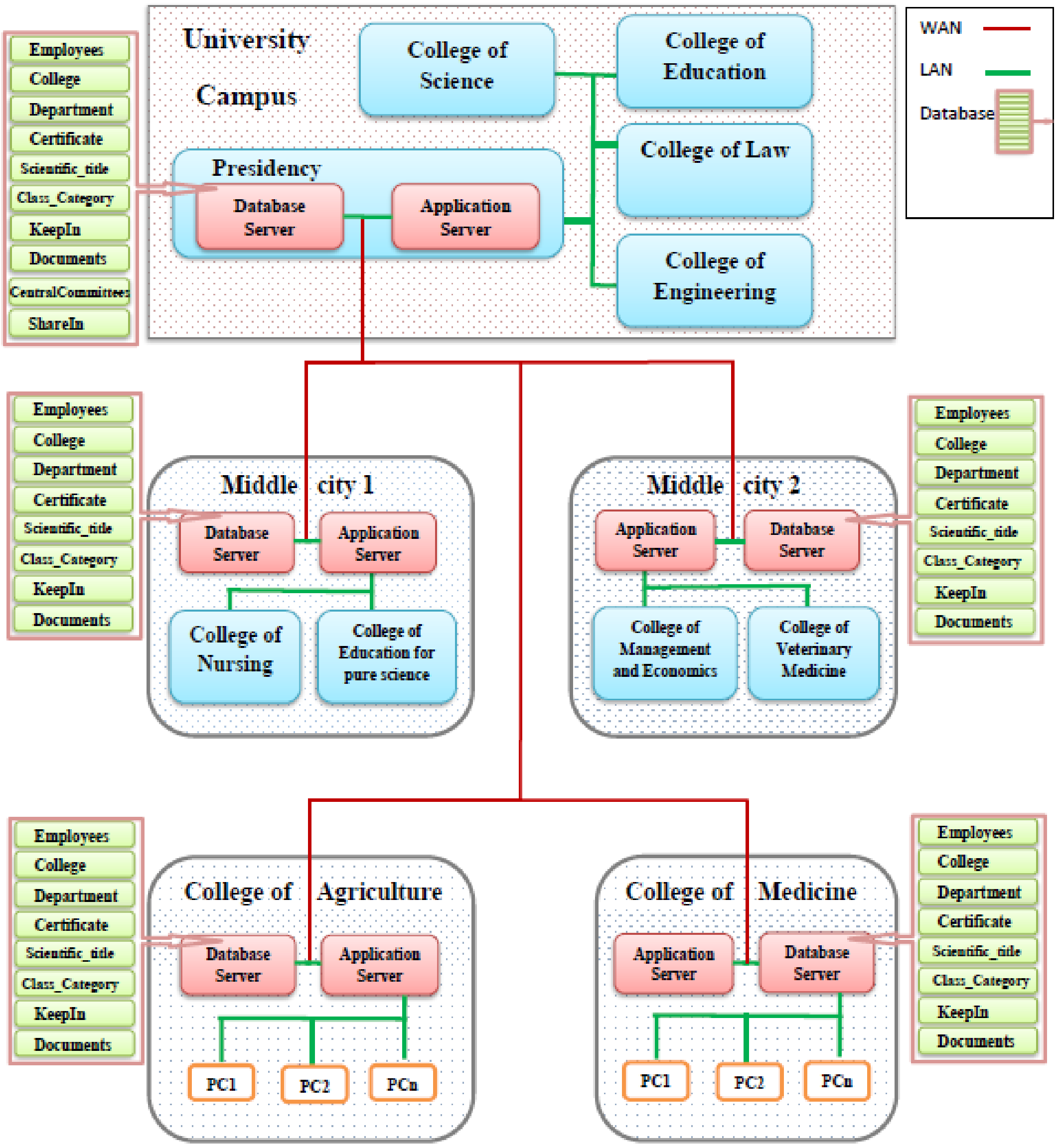

Figure 17 Distributed database model for Kirkuk University

The speed of the internally used network at the Presidency of the University and the colleges featured as good relative to the speed of the WAN that is represented on the Internet. The Internet at the Presidency of the University operates at 10 Mbps and at 2 Mbps at the College of Agriculture. Therefore, the communication between the Presidency and College of Agriculture is limited to 2 Mbps.

3.4 Comparing Between Two Systems Using the Models

The comparison operation between these two systems was obtained in light of the suggested models for these systems that were designed with Kirkuk University conditions in mind. Additionally, the comparison operation was obtained through three methods as detailed below.

3.4.1 General features

There are many differences between the centralized database and distributed database and for each method there are advantages and disadvantages. At times, one database will be preferred to the other when we have special conditions imposing themselves during the design of the database.

For Kirkuk University, we need to take into consideration the following conditions:

- A geographical distribution of the colleges to the five main places throughout the city.
- The WAN strength representing the speed of the internet between these colleges.
- Stoppages of the internet service at random times resulting from either power outages or an outage of the Internet itself.
- When we change the current applied transfer mechanism with the transfer database mechanism, the amount of transferred data between the Presidency of the University and the colleges will be taken into consideration.

We will be dealing in this thesis with some of these features as follows:

1- Availability: In the centralized database model, all the data, especially for the university, are placed on one server which is located on the campus of the University, where we find four colleges as well as the Presidency of the University communicating with campus's server via LAN and the remaining six colleges via the Internet. When all data are on the campus, the large amount of data will be transferred daily among the server and colleges, including the daily official books and special updating for employees' files and so on. The transfer of these data for the colleges that communicate with the server via LAN will be efficient; however, the colleges that communicate with the server via the WAN will operate in accordance with Internet strength and availability. When the Internet is stopped at the server, all communication channels between the colleges that communicate with the server via the Internet are also cut, thereby rendering it impossible to send or receive any data from the server until the Internet returns.

However, in the distributed database model, we can see when the WAN is cut between the campus of the University and the distant colleges. Only the data available on the campus of the University is unable to reach the distant colleges which communicate with the campus of the University via the WAN unless each distant college is able to obtain most of their daily required data from the nearest server via LAN.

2- Performance: When a query is issued from a distant college to the University campus, such as College of Agriculture asking the following query: (display details for all employees at the College of Agriculture aged 25 to 40).

In the centralized database model, the result of this query will come from the Presidency of the University (which has the main and only server in the system) and will be transferred to the College of Agriculture affected by the available WAN strength. In addition, the size of the database will be huge because all university data are on the same server. There is a fact that the relationship between the required time to complete the execution of a query and the size of database is a direct correlation; that is, when the size of a database increases, more time is required to complete the execution of a query. For example, if thirty departments from different colleges (or the same colleges) were to send queries asking similar or different questions to the server simultaneously, there will be competition to utilize the power of the server and each query will be affected by the other.

In the distributed database model, the answer for the same query above will come from a server that is available inside the college containing most of the college's daily required data. The answer will be affected by the internal network without passing through the WAN. In addition, the size of the database will be small on the server because it is a part of a larger database [7][11]. Lastly, there is no competition for this server from other colleges, thereby leading to an increase in its performance.

3- Costs: When considering the calculation of costs of the number of servers utilized in each suggested model, it can be observed that in the centralized case, there are only two servers with LAN and WAN equipment, such as wires, routers, switches, etc. In addition, maintenance will be carried out in one place at most. However, in the distributed case, it can be seen that there are ten servers in different places with LAN and WAN equipment, such as wires, routers, switches, etc. Lastly, the maintenance that is required for ten servers in five places leads to the total cost of the distributed database model exceeding the cost of the centralized database model.

4- Recurrence: Usually in traditional databases, we observe a reduction of redundancy. In the two suggested database models for Kirkuk University, the rate of recurrence can easily be observed.

In the centralized model, all data are in the same location. Therefore, the rate of recurrence is minimal where there is one copy from each table. On the other hand, we can observe the distributed model for the university, with each replication of the tables leading us toward increasing the rate recurrence. Then it is easy to determine that the rate of recurrence in the distributed database model exceeds the centralized database model.

5- Safety: It is worth mentioning the particular conditions at Kirkuk University, at which the security situation has been troubling. From time to time, we can see explosions occurring. Therefore, it ought to be taken into consideration when deciding to select a suitable database system for a university for the conditions in which a university such as Kirkuk University finds itself. Explosions may occur near to any college thereby leading to Internet cuts for many days. In the centralized database model, the whole database is located in one place; therefore, any incident that stops the Internet leads to a cutting of communication from the only server. However, in the distributed database model, when explosions occur nearby any college, only that college is disconnected from service.

Lastly, it should be noted there are many differences between the centralized database system and distributed database system. We focused on only the features of the centralized system being preferred to the distributed system or vice versa, in light of the conditions at Kirkuk University, and avoided the features of the two systems introducing the same results, such as the centralized control feature or data integrity feature.

3.4.2 Suggested queries and execution

The cost of execution for any query from any database server consists of five elements: Access cost to secondary storage, Disk storage cost, Computation cost, Memory usage cost and Communication cost. The difference is found in the last element.
In the centralized case it concludes the return for the amount of data as a result from the server to client. However, in the distributed case, it concludes transferring the result of the query as well as transferring the tables among the servers and clients [11].

The communication cost will have less effect on the calculation of the cost function when we use a network with high efficiency among the sites [11]. However, when the sites communicate with each other via the WAN, the communication cost will have a more significant effect than all the elements shared in calculating the cost function [8][11]. Particularly when using a WAN, it can be described as being between good and weak, as it occurs in the reality of Kirkuk University.

We can observe there are six colleges from ten that communicate with the University campus via WAN. Therefore, the result of calculating the communication costs for some suggested queries for the two suggested models will be useful in order to determine which one of the two models is faster in processing queries, such as the suggested queries in this thesis. We consider this element to be the most effective element in the cost calculation function for queries.
Computing the communication costs for a specific query requires calculating the amount of transferred data between the server and client during execution of the query.

The Join and Semijoin are the most important operations used in databases for the extraction of required information from two or more tables as a result of a query [24]. To show the computation of communication costs in the centralized database and distributed database models, we will assume the following queries:

1- Query X: Display the first name and surname for College of Agriculture's employees on the central committees, in addition to the committees' name and type. The query is issued from the College of Agriculture.

In SQL:

%Code Start

Select e.Fname, e.Lname, c.Name, c.Job

From Employees e, ShareIn s, CentralCommittees c, Department d, College l

Where e.Id=s.EmpId and s.CC_Id=C.Id and e.depId=d.Id and d.collegeId=l.Id and l. Name="Agriculture";

%Code End

The number of employees who share in the central committees from the College of Agriculture is 16 from a total of 120 employees available in the College of Agriculture. The Employees, CentralCommittees, and ShareIn tables are shown in Figures 18, 19 and 20 respectively.

Employees

Id	Fname	Lname	BirthDate	NominationDate	ClassId	SciensId	DepId	Certif_Id

1250 records
Each record 80 bytes
Id field 4 bytes

Fname field 15 bytes
Lname field 15 bytes

Figure 18 Employees' table

CentralCommittees

Id	Name	Job

16 records
Each record 39 bytes
Id field 4 bytes

Name field 15 bytes
Job field 20 bytes

Figure 19 CentralCommitees table

ShareIn

Id	CC_Id	EmpId	DateS	DateF

62 records — EmpId field 4 bytes

Each record 32 bytes — DateS field 10 bytes

CC_Id field 4 bytes — DateF field 10 bytes

Figure 20 ShareIn table

The execution time for this query is calculated with two models:

A- In the centralized database model: The answer for this query is obtained from the the Presidency of the University, which contains the main and sole database server in the university. All tables on this server making the join operation easy use the joining field in the ShareIn table to join it with the Employees and CentralizedCommittees tables. Then the required fields, namely (Employees.Fname, Employees.Lname, CentralCommittees.Name, CentralCommittees.Type) will be extracted and transferred to the College of Agriculture.

B- In the distributed database model: For Query X, we use two approaches for the distributed database model as follows:

1- Join approach: Using the Join approach, the query can be executed by following two strategies:

First: The answer for Query X is by transferring the ShareIn and CentralCommittees tables from the Presidency of the University to the College of Agriculture so as to carry out the joining operation and extract the required fields.

Second: The answer for Query X is obtained by transferring the employees' table and department table that are available in the College of Agriculture to the Presidency of the University in order to execute the joining operation and extract the required fields and return it to the College of Agriculture, the point of origin of the query.

2- Semijoin approach: By using the semijoin approach, the number of rows is reduced before sending it to the other side [24]. The joining field in the employees' table in the College of Agriculture will be transferred to the server, which is located in the Presidency of the University to execute the joining with the ShareIn and CentralCommittee tables and return the required fields (SharIn.EmpId, CentralCommittees.Name, CentralCommittees.Type) to the College of Agriculture to execute the final joining operation at the College of Agriculture with the employees' table followed by the extraction of the required fields.

2- Query Y: Display first name, surname and date of nomination for College of Agricultures' employees? The query is issued from the Presidency of the University,

In SQL:

```
%Code Start
Select e.Fname, e.Lname, e.NominationDate
From Employees e, Departments d, Colleges e
Where e.DepId=d.Id and d.CollegeId=c.Id
And c.Name="Agriculture";
%Code End
```

The execution time for this query is calculated with two models:

A- In the centralized database model: Because all the university tables in one server which placed on the University campus, the same place where Query Y was issued, the joined operation among the tables and the extraction for the required fields is obtained in the same place without transferring any data via the WAN.

B- In the distributed database model: We can observe that the tables of Employees and Departments are partitioned. The rows containing data, especially for the College of Agriculture, were disjoined from the main Employees and Departments tables and were set on a server in the College of Agriculture as well as at all the other colleges. Therefore, the employees and the department, especially parts of the College of Agriculture, will exist in the College of Agriculture. In addition, as a result of the replications, the table of Colleges also exists in the Presidency and the College of Agriculture. Therefore, the required tables in the joining operation are present in the College of Agriculture. Therefore, the joining operation is obtained in the College of Agriculture and the required fields will be extracted and transferred to the Presidency of the University.

3- Query Z: Display first name, surname and date of nomination for the College of Agriculture's employees? Repeating for the query Y but now issued from College of Agriculture. This query is calculated with two models:

A- In the centralized database model: Because all tables present on the University campus that communicate via the WAN with the College of Agriculture, the joining operation for the tables that share in providing the required fields, are carried out in the Presidency, followed by the required fields being extracted and transferred to the College of Agriculture.

B- In the distributed database model: Because all the tables present on a server are placed in the College of Agriculture, the joining operation for the tables that are shared in providing the required fields are carried out in the College of Agriculture, followed by the required fields being extracted.

3.4.3 Questionnaire

The questionnaire operation included about fifty employees from the university and covered the Presidency of the University considered as the Campus of the University, and the College of Agriculture, considered as the farthest college from the Campus of Kirkuk University. The questionnaire contained seven questions which were classified in three directions. The first direction pertained to computerizing the current applied mechanism at the university. The second pertained to knowing about the rate of quality and availability of the Internet at the Presidency of the University and colleges. The third direction pertained to knowing about the availability of the electric power at the Presidency of the University and colleges. It is noticeable that some of these directions divided into two questions in order to obtain information in an accurate form. For example, instead of asking about the number of the absences of electric power, they are divided into questions about the number of electricity cuts and a question about the time required to operate the generator. The same questions apply to the Internet. All are based on the idea that the persons answering these questions can easily guess the correct answer in case of divided questions instead of complex questions. These questions can be seen in Figure 21.

Computing the probability of the existing number of hours for the WAN between the University campus and the far colleges can be carried out by calculating the probability of the unavailability the Internet in the Presidency of the University and colleges. This probability depends on computing the unavailability of the Internet and electric power that affects the unavailability of the Internet.

The probability of unavailability of electric power during a month in a college or Presidency of the University can be computed with equation (3.1) and equation (3.2) based on the results that are gained from the questionnaire.

$$\mathrm{C} = \sum_{i=1}^{n} N_i \times N_{ip} \times D \qquad (3.1)$$

N_i : Number of times outages occur in one official work day.

N_{ip} : Percentage for N_i

D : Constant number denoting the official work days in a month equal to 22.

C : Probability the number of outages in a month.

Subject of Questionnaire: Computerized the Administrative Work in the University By Adopting a Database System Depending to Availability the Internet and electricity

Questionnaire

1- Do you support the idea of transforming the paperwork to the computer in the university and colleges such as prepare and printing administrative orders, sending, receiving it among the presidency of the University and colleges, archiving and retrieval it in specific speed and time?
   - Yes
   - No

2- What do you think about the quality of the Internet in presidential / college?
   - Good
   - Medium
   - Weak

3- Does the Internet service is stopped at any time in the Presidential / college?
   - Once in a month
   - Twice in a month
   - Once in a week
   - Twice in a week or more

4- How much time will require even the Internet service go back after interruption?
   - Less than a quarter of an hour
   - One hour
   - Two hour
   - Three hour or more

5- What do you think about the power outages during the official working hours? Is it:
   - Once a day
   - Twice
   - More

6- What do you think about the electric generating Presidential / college when power outages?
   - Works directly and automatic without feeling the power outages
   - Works after a power outage a quarter of an hour or less
   - Needs one hour or more

7- How many times it happened that the power outages and the generator not work because technical fault or absence Diesel or any other reason?
   - Does not happen at all
   - Once or twice in month
   - More

Figure 21 The used question list in the questionnaire operation

$$T = \sum_{i=1}^{n} S_i \times S_{ip} \times C \qquad (3.2)$$

S_i : Required time to operate the generator per outage.

S_{ip} : Percentage for S_i

T : Number of hours of unavailability of electric power per month.

On the other hand, the computation of the probability of the hours of unavailability of the Internet in the Presidency of the University and colleges by equation (3.3) and (3.4).

$$\mathrm{Y} = \sum_{i=1}^{n} X_i \times X_{ip} \qquad (3.3)$$

$$H = \sum_{i=1}^{n} M_i \times R_i \times Y \qquad (3.4)$$

Xi : Number of Internet stopping in a month (as one from the suggested answers).

X_{ip} : Percentage for the X_i (as the percentage of choosing the answer Xi).

Y : Probability the number of times of Internet stopping in a month.

Mi : Numbers hours of delaying the Internet until coming back in each cut.

R_i : Percentage for M (as the percentage of choosing the answer Mi).

H : Number hours of unavailability of the Internet per month with existing electric power.

As an outcome for all that is introduced, the percentage of the availability of the Internet in the Presidency of the University or in any far college per month can be calculated by multiplying (the percentage for unavailability of electric power per monthe subtracted from one) with (the percentage for the unavailability of the Internet per monthe with existing electric power subtracted from one), using equation (3.5).

$$U = (1 - T\%) \times (1 - H\%) \tag{3.5}$$

U : Percentage for the availability of the Internet

The rate of availability the WAN between the Campus of the University and any far college at any day can be computed by multiplying the result of U for the two sites that communicate together through the WAN.

CHAPTER 4

RESULTS

This chapter will present the results for the suggested queries in addition the results of the questionnaire that were obtained from the Presidency of the University and the College of Agriculture to find the availability time for the WAN between them.

4.1 Results of Query X

The results of query X will showed in two models, the centralized database model and the distributed database model.

4.1.1 In the Centralized database model

The communication cost for query X can be computed by calculating the transferring requirement fields (e.Fname, e.Lname, c.Name, c.Type) from the Presidency of the University to the College of Agriculture. The details are shown in Table 2.

T	A	B	C	D =B*C
1	e.Fname	15	16	240
2	e.Lname	15	16	240
3	c.Name	15	16	240
4	c.Type	20	16	320
S =				1040

Table 2: Results of Query X in the Centralized Database Model

A : Field name.

B : Field size at each record.

C : Number of required records depending to the joining condition in query.

D : The transfer size for the required field.

S : Total transfer size for the result of query.

4.1.2 In the distributed database model

The query for this model is executed by two approaches as follows:

1- Join approach: which is executed using two strategies, as in the following:

A- Join approach using strategy one: In this strategy the transferred data passes in one direction through the WAN, only by sending data from the Presidency of the University to the College of Agriculture. The details for the computation and results are shown in Table 3.

T	A	B	C	D =B*C
1	c.Id	4	16	64
2	c.Name	15	16	240
3	c.Type	20	16	320
4	s.Id	4	62	248
5	s.CC_Id	4	62	248
6	s.EmpId	4	62	248
7	s.DateS	10	62	620
8	s.DateF	10	62	620
S =				2608

Table 3 Results of Query X in the DDB Model by Join Approach in Strategy One

B- Join approach by using strategy two: In this strategy, the transferred data passes through the WAN in two directions, first by transferring data from the College of Agriculture to the Presidency of the University, and second by transferring data from the Presidency of the University to the College of Agriculture. Therefore, the total transferred data through the WAN is calculated by summing the results for two directions of transfer. The details for the computation and results are shown in Table 4.

T	A	B	C	D =B*C
First direction: from College of Agriculture to Presidency				
1	e.Id	4	120	480
2	e.Fname	15	120	1800
3	e.Lname	15	120	1800
4	e.BirthDate	15	120	1800
5	e.NominationDate	15	120	1800
6	e.ClassId	4	120	480
7	e.SciensId	4	120	480
8	e.DepId	4	120	480
9	e.Certif_Id	4	120	480
	d.Id	4	18	72
	d.Name	15	18	270
	d.stabdate	15	18	270
	d.CollegeId	4	18	72
S1 =				10284
Second direction: from Presidency to College of Agriculture				
1	e.Fname	15	16	240
2	e.Lname	15	16	240
3	c.Name	15	16	240
4	c.Type	20	16	320
S2 =				1040
S =				11324

Table 4 Results of Query X in the DDB Model by Join Approach in Strategy Two

S1: Transfer size of data from the College of Agriculture to the Presidency of the University as the first direction.

S2: Transfer size of data from the Presidency of the University to the College of Agriculture as the second direction.

1- **Semijoin approach:** In this approach, the transferred data also passes in both directions through the WAN first by transferring data from the College of Agriculture to the Presidency of the University, and second by transferring data from the Presidency of the University to the College of Agriculture. Therefore, the total transferred data through the WAN is calculated by summing the results for the two directions of transfer. The details for the computation and results are shown in Table 5.

T	A	B	C	D =B*C
[First direction] from College of Agriculture to Presidency				
1	e.Id	4	120	480
S1 =				480
[Second direction] from Presidency to College of Agriculture				
1	s.EmpId	4	16	64
2	c.Name	15	16	240
3	c.Type	20	16	320
S2 =				624
S =				1104

Table 5 Results of Query X in the DDB Model by Semijoin Approach

4.2 Results of Query Y

The results of Query Y are shown in two models: the centralized database model and the distributed database model.

4.2.1 In the centralized database model

While this query is issued from the Presidency of the University, the result of the total size for the required data transferred through the WAN is equal to zero.

4.2.2 In the distributed database model

In this model, the required fields are sent from the College of Agriculture to the Presidency of the University. The details for the computation and results are shown in Table 6.

T	A	B	C	D =B*C
2	e.Fname	15	120	1800
3	e.Lname	15	120	1800
4	e.NominationDate	15	120	1800
S =				5400

Table 6 Results of Query Y in the Distributed Database Model

4.3 Results of Query Z

The results of Query Z are shown in two models: the centralized database model and the distributed database model.

4.3.1 In the centralized database model

While this query is issued from the College of Agriculture, all tables in the Presidency of the University lead to the transfer of the total size for the required data through the WAN, as shown in Table 7.

T	A	B	C	D =B*C
2	e.Fname	15	120	1800
3	e.Lname	15	120	1800
4	e.NominationDate	15	120	1800
S =				5400

Table 7 Results of Query Z in the Centralized Database Model

4.3.2 In the distributed database model

While this query is issued from the College of Agriculture, the data for the College of Agriculture resides on the server inside the college, which results in the total size for the required data being transferred through the WAN equaling zero.

The summary results for all suggested queries is in Table 8

T	Query name	Centralized DB	DDB
1	Query X	1040	Join st1 = 2608
			Join st2 = 11384
			Semijoin = 1104
2	Query Y	0	5400
3	Query Z	5400	0

Table 8 Summary Results for All Suggested Queries

4.4 Results of Questionnaires

The results of questionnaire are presented for the Presidency of the University and the College of Agriculture.

4.4.1 Questionnaire for the Presidency of the University

Table 9 showed the results for questionnaire that applied for the Presidency of the University

Questions	Answers				Percentages of the answers			
	A1	A2	A3	A4	A1%	A2%	A3%	A4%
1	22	3	0	0	85	12	0	0
2	13	9	3	0	50	35	12	0
3	4	9	4	8	15	35	15	31
4	7	7	5	6	27	27	19	23
5	6	9	10	0	23	35	38	0
6	9	12	4	0	35	46	15	0
7	2	14	9	0	8	54	35	0

Table 9 Results of the Presidency of the University's Questionnaire

From the results of Question 5 in Table 9 by using equation (3.1), the value *C* can be calculated for the Presidency of the University. The detail of the operation is shown in Table 10.

t	N_i	N_{ip}	$C_i=(N_i*N_{ip}*22)$
1	1	0.23	5.06
2	2	0.35	15.4
3	3	0.38	25.08
C=			45.54

Table 10 Calculating the Value *C* for the Presidency of the University

From the results of Question 6 in Table 9 by using equation (3.2), the value *T* can be calculated for the Presidency of the University. The detail of the operation is shown in Table 11.

t	S_i	S_{ip}	$T_i=(S_i*S_{ip}*C)$
1	0	0.35	0
2	0.25	0.46	5.2371
3	1	0.15	6.831
T=			12.0681

Table 11 Calculating the Value *T* for the Presidency of the University

From the results of Question 3 in Table 9 by using equation (3.3), the value *Y* is calculated for the Presidency of the University. The detail of the operation is shown in Table 12.

t	X_i	X_{ip}	$Y_i=(X_i*X_{ip})$
1	1	0.15	0.15
2	2	0.35	0.7
3	3	0.15	0.45
4	4	0.31	1.24
Y=			2.54

Table 12 Calculating the Value *Y* for the Presidency of the University

From the results of Question 4 in Table 9 by using equation (3.4), can be calculated the value *H* for the Presidency of the University. The detail of the operation is shown in Table 13.

t	M_i	R_i	$H_i=(M_i*R_i*Y)$
1	0.25	0.27	0.17145
2	1	0.27	0.6858
3	2	0.19	0.9652
4	3	0.23	1.7526
H=			3.57505

Table 13 Calculating the Value *H* for the Presidency of the University

By using equation (3.5) and the value *T* and *H* for the Presidency of the University can be computed the probability of the number hours of availability the Internet for the Presidency of the University at any day.

$$U = (1-0.12)*(1-0.035)$$
$$= 0.847$$
$$= 85\%$$

4.4.2 Questionnaire for the College of Agriculture

Table 14 shows the results for the questionnaire that were applied to the College of Agriculture.

Questions	Answers				Percentages of the answers			
	A1	A2	A3	A4	A1%	A2%	A3%	A4%
1	23	2	0	0	92	8	0	0
2	7	9	9	0	28	36	36	0
3	4	6	7	8	16	24	28	32
4	9	6	4	6	36	24	16	24
5	1	5	19	0	4	20	76	0
6	5	19	1	0	20	76	4	0
7	3	17	5	0	12	68	20	0

Table 14 Results of the College of Agriculture's Questionnaire

From the results of Question 5 in Table 14 by using equation (3.1), the value C can be calculated for the College of Agriculture. The detail of the operation is shown in Table 15.

t	N_i	N_{ip}	$C_i=(N_i*N_{ip}*22)$
1	1	0.04	0.88
2	2	0.2	8.8
3	3	0.76	50.16
C=			59.84

Table 15 Calculating the Value C for the College of Agriculture

From the results of Question 6 in Table 14 by using equation (3.2), the value T can be calculated for the College of Agriculture. The detail of the operation is shown in Table 16.

t	S_i	S_{ip}	$T_i=(S_i*S_{ip}*C)$
1	0	0.2	0
2	0.25	0.76	11.3696
3	1	0.04	2.3936
T =			13.7632

Table 16 Calculating the Value *T* for the College of Agriculture

From the results of Question 3 in Table 14 by using equation (3.3), the value *Y* can be calculated for the College of Agriculture. The detail of the operation is shown in Table 17.

t	X_i	X_{ip}	$Y_i=(X_i*X_{ip})$
1	1	0.16	0.16
2	2	0.24	0.48
3	3	0.28	0.84
4	4	0.32	1.28
Y=			2.76

Table 17 Calculating the Value *Y* for the College of Agriculture

From the results of Question 4 in Table 14 by using equation (3.4), the value *H* can be calculated for the College of Agriculture. The detail of the operation is shown in Table 18.

t	M_i	R_i	$H_i=(M_i*R_i*Y)$
1	0.25	0.36	0.2484
2	1	0.24	0.6624
3	2	0.16	0.8832
4	3	0.24	1.9872
H=			3.7812

Table 18 Calculating the Value *H* for the College of Agriculture

By using equation (3.5), the values *T* and *H* for the College of Agriculture and the probability of the number of hours of availability of the Internet for the College of Agriculture on any day are computed.

$$U = (1-0.137)*(1-0.037)$$
$$= 0.829$$
$$= 83\%$$

The probability of availability of the WAN between the Presidency of the University and the College of Agriculture is computed by multiplying the value of *U* for the Presidency of the University with the value of *U* for the College of Agriculture as follows:

AvWAN: Availability of WAN between the two sites at any day

$$AvWAN = U1*U2$$
$$= 0.847 *0.829$$
$$= 0.70$$
$$= 70\%$$

CHAPTER 5

CONCLUSION

5.1 Analysis of Results

When we look at the results and analyze and study them, a suitable database system for the conditions at Kirkuk University will be selected.

5.1.1 General features

With regard to the features, the five features were studied, three of which, namely Availability, Performance and the Safety, were preferred as the distributed database models. However, the other two features, Cost and the Recurrence rate, were preferred as the centralized database model. Furthermore, when studying these results carefully, the features of Availability, Performance and Safety features (that were preferred the distributed database model) can be classified as priority features which cannot be ignored during the designing phase of the database system for Kirkuk University.
However, when analyzing the Cost feature, which is preferred for the centralized database model, the question arises as to how the system benefits by being less expensive but not efficiently meeting most university requirements. Therefore, if the distributed database model is chosen as a suitable database model for Kirkuk University, it will be expensive, more than the centralized database model. The same applies for the Recurrence feature.

5.1.2 Suggested queries

Queries that used in this thesis suggested depending on two suggested database models and the mostly subjects and procedures that used at the University.

From the results of Query X, we can observe clearly that the centralized database model will be better suited than the distributed database model when the adopted approach in the distributed model is the Join approach. However, when using the Semijoin approach, the difference will be minimal or negligible, and when checking the results of Query Y, we can clearly say that the centralized database model is also better.

The bottom line here is using queries such as Query X, which can be described as being seldom used. The same applies for Query Y; however, Query Y is used more than Query X. Nevertheless, it is still seldom when compared with using queries such as Query Z, considering queries such as Query Y to meet the general requirements for the Presidency of the University, which deals with the colleges' matters unlike the Presidency's specific daily requirements. Thus, the previously mentioned can be generalized for all similar queries.

In addition, the high efficiency for Query Z can be observed when working with the distributed database model. There are great differences when comparing it with the results of Query Z with the centralized database model. Query Z is issued from the College of Agriculture, whereas, the required data are available in the college of Agriculture in a distributed database model, and these data are available at the Presidency of the University when working with the centralized database model. Therefore, when we analyze these required data, it can be said that these are the daily requirements for data by the College of Agriculture which can therefore be measured for the most daily requirements for all the distant colleges that communicate with the Campus of the University by WAN. What is said about Query Z is naturally said about all the queries.

Based on all that is introduced in the suggested queries and all that is similar to these queries can be used in the suggested database models, we can say that the distributed database model is preferred due to the fact that it meets most daily requirements with lower communications costs when compared with the centralized database model.

5.1.3 The questionnaire

For the questionnaire, it can be observed that the most sharing persons in it support the adoption of the idea of the database system in order to computerize the current usage mechanism at Kirkuk University. Most of the sharing persons in the questionnaire from the Presidency of the University evaluated the strength of the Internet between medium and good, while most of the sharing persons in the questionnaire from the College of Agriculture evaluated the Internet strength between medium and weak. This means that the strength of the Internet is classified as a weakness for 25 percent of all the employees who participated in the questionnaire. For the remaining employees, it is ranged between medium and good strength.

In addition, one can see the rate of availability of the Internet between the Presidency of the University and College of Agriculture, which is 70 percent, thus implying that there is unavailability of the Internet between the two sides at a rate of approximately 30 percent, meaning one-third of a normal work day without WAN communication. If we assume the official books and data that are small in size can be transferred at this rate, and if we adopted the transferal of all daily requirements between the far colleges and the Campus of the University as it is in the centralized database model, at that time it will be considered an undesirable method. Therefore, the distributed database model is a preferable model at this rate of availability for a WAN between the Presidency of the University and College of Agriculture, as can an identical situation for all far colleges from the University campus.

5.2 Recommendations and Future Work

Future work may include making a comparison between the distributed database system and parallel database system in order to know which is suitable for the conditions at Kirkuk University in addition to using practical programs in order to apply the suggested queries in reality to the known speed of execution for the same queries in the two different systems.

Furthermore, another line of future work can include making a studying for the possibility of generalization of distributed database systems between Kirkuk University and the High Education and Scientific Research Ministry.

5.3 Conclusion

From all mentioned, it is clear for us that the comparison between the two suggested models be done by the three methods. After the analysis of the comparisons results for these methods, we obtained many facts. Firstly, in what is related to the results of the first method that are represented in comparison the general features for the two systems in light of Kirkuk University using the two suggested models, we ensure the most-used features in the comparison operation preferred the distributed database model. For the second method that represents the suggested queries to be simulated in the most university queries, we ensured the queries that are used daily confirm to a priority of adoption of the Distributed Database model considering it meets most daily requirements with lower communication costs. For the third method which is represented in the questionnaire, the results obtained from the questionnaire say that the rate of unavailability of the WAN between two sides around one-third of the normal work day. In addition, the strength of availability of the network is good for only half of the employees who share in the questionnaire and medium or weak for others “depending on the questionnaire results in Question 2.”

As a final result, after analyzing the results of the three methods, we are certain that the distributed database system is a suitable database system for Kirkuk University.

REFERENCES

1. **Ramakrishnan R., Gehrke J, (2003),** "*Database Management System*", McGraw-Hill, New York, vol. 3, pp. 8-15.

2. **Gadicha A. B., Alvi A. S., Gadicha V. B., Zaki S. M., (2012),** "*Top-Down Approach Process Built on Conceptual Design to Physical Design Using Lis, Gcs Schema*", International Journal of Engineering Sciences & Emerging Technologies, vol. 3, pp. 90-96.

3. **Baraka R. S., Madoukh S. M., (2012),** "*A Conceptual SOA-Based Framework for e-Government Central Database*", IEEE International Conference on Computer Information and Telecommunication Systems, Amman.

4. **Talab M. M., Mustafa B. A., (2013),** " *Design and Implementation of a Distributed Database System for the Central Bank of Iraq Using Oracle*", International Journal of Computational Engineering Research, vol. 7, pp. 38-45.

5. **Alyaseri S. J., (2010),** "*Distributed University Registration Database System Using Oracle 9i*", Computer and Information Science, vol. 3, pp. 59-67.

6. **Foster N. L., Gombosi E., Teboe C., Little R. J. A., (2000),** "*Balanced Centralized and Distributed Database Design in a Clinical Research Environment*", John Wiley & Sons, vol. 19, pp. 1531-1544.

7. **Khan S. I., Latiful Hoque A. S. M, (2010),** "*A New Technique for Database Fragmentation in Distributed Systems*", International Journal of Computer Applications, vol. 5, pp. 20-24.

8. **Navathe S., Ceri S., Wiederhold G., Dou J., (1984),** "*Vertical Partitioning Algorithms for Database Design*", ACM Transactions on Database Systems, vol. 9, pp. 680-710.

9. **Limited I., (2010),** "*Introduction to Database Systems*", Pearson Education, pp. 2-18.

10. **Gillenson M. L., (2012),** "*Fundamentals of Database Management Systems*", John Wiley & 1ons, pp. 48.

11. **Elmasri R., Navathe S. B., (2011),** "*Fundamentals of Database Systems*", Addison-Wesley, pp. 5-6-47-201-880-882-901-902.

12. **http://www.answers.com/Q/What_is_centralized_database,** (Data Download, Date : 22.08.2014).

13. **CONNOLLY T. M., BEGG C. E., (2005),** "*Database Systems A Practical Approach to Design, Implementation, and Management*", Addison Wesley, pp. 4-60-697-709-723.

14. **Hoffer J. A., Ramesh V., Topi H., (2011),** "*Modern Database Management*", Prentice Hall, New Jersey, pp. 18-19

15. **http://www1.gantep.edu.tr/~aoztas/ce332-book/XIV/07.html,** (Data Download, Date : 27.10.2014).

16. **Özsu M. T., Valduriez P., (1991),** "*Distributed Database Systems: Where Are We Now*?", IEEE Computer Society, vol. 24, pp. 68-78.

17. **Özsu M. T., Valduriez P., (2011),** "*Principles of Distributed Database Systems*", Pearson Education, Inc., London, pp. 25-27-32-76.

18. **Tomar P., Megha, (2014),** "*An Overview of Distributed Databases*", International Research Publications House, vol. 4, pp. 207-214.

19. **Bhuyar P. R., Gawande A. D., Deshmukh A. B., (2012),** "*Horizontal Fragmentation Technique in Distributed Database*", International Journal of Scientific and Research Publications, vol. 2, pp. 1-7.

20. **Ma H., Schewe K., Kirchberg M., (2006),** "*A Heuristic Approach to Vertical Fragmentation Incorporating Query Information*", IEEE International Baltic Conference on Databases and Information Systems, Vilnius, pp. 69-76.

21. **Sumathi S., Esakkirajan S., (2007),** "*Fundamentals of Relational Database Management Systems*", Springer, Berlin, vol. 47, pp. 570-580.

22. **www.egove.gov.iq,** (Data Download, Date : 15.04.2014).

23. **Yadav S. C., Singh S. K., (2009),** "*An Introduction to Client/Server Computing*", New Age International Publishers, New Delhi, pp. 7-8.

24. **Sharma M., Singh G., (2012),** "*Analysis of Joins and Semi-joins in Centralized and Distributed Database Queries*", International Conference on Computing Sciences, pp. 15-16.

APPENDICES A

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Al-Khanji, Yalmaz

Date and Place of Birth: 20 August 1982, Kirkuk-Iraq

Marital Status: Married

Phone: +964 770 400 53 13

Email: yelhan1982@gmail.com

EDUCATION

Degree	**Institution**	**Year of Graduation**
M.Sc.	Çankaya University, Mathematics and Computer Science	2015
B.Sc.	Kirkuk University, Computer Science	2006
Diploma	Kirkuk Technical Institute, Computer System	2002
High School	Al-Tameem High School	2000

WORK EXPERIENCE

Year	**Place**	**Enrollment**
2008- 2012	Kirkuk University, Financial Department	Salaries Accountant
2007 April	Kirkuk University, Computer Science Department.	Teaching Assistant

FOREIN LANGUAGES
Advanced English, Beginner Turkish

HONOURS AND AWARDS

1. Graduate Honor Student 2002
   Kirkuk Technical Institute
2. Graduate Honor Student 2006
   Kirkuk University

HOBBIES
Coding, Football, Reading.