

COURSE TITLE: DISTRIBUTED SYSTEM

PROJECT TITLE:-CASE STUDY ON DISTRIBUTED LIBRARY SYSTEM FOR UOG

BY: -

NAME ID

AlehegnAdane 7655/07

Lema Misgan

Lakew Tarekegn

Submitted to: Mesifin(PHD).

Gondar, Ethiopia

May, 2015

**List of Acronyms**

UOG-University of Gondar

CMHS-Lib:-College of Medical and Health Science Library

ATewd-Lib:-Atse Tewdros Campus Library

Maraki-Lib:-Maraki Campus Library

AFasil-Lib:- Atse Fasil Campus Library

Tseda-Lib:-Tseda Campus Libra

ii. Table of content

Contents

I. List of Acronomy [ii](#_Toc417782763)

II. Table of Content [iii](#_Toc417782763)

[1. Introduction 1](#_Toc417782763)

[2. Problem Statement and Justification 2](#_Toc417782764)

[3. Objective of the project 3](#_Toc417782766)

[3.1 General Objective 3](#_Toc417782767)

[3.2 Specific Objectives 3](#_Toc417782768)

[4. Scope of project 3](#_Toc417782769)

[5. Need assessment 3](#_Toc417782770)

[6. Distribution Goals 4](#_Toc417782771)

[7. Design decisions 4](#_Toc417782772)

[8. Architecture 8](#_Toc417782773)

[9. Conculusion 10](#_Toc417782774)

[9. Reference 11](#_Toc417782774)

[9. Appendex 12](#_Toc417782774)

# Introduction

In today’s world of emerging technologies, enterprises are moving towards the Internet for

businesses. People are rushing towards the e-commerce applications for their day-today needs, which in turn are making the Internet very popular. Library management system has given both an opportunity and a challenge to traditional library in academic world. Libraries play an important role in the academic world by providing access to world-class information resources and services. Advances in information technology have changed the entire gamut of Library and Information service today. It is time to develop new means of information access and resource sharing with the help of fast growing computer and telecommunication technologies. Virtually all large computer-based systems are now distributed systems. Information processing is distributed over several computers rather than confined to a single machine. Distributed system is therefore very important for enterprise computing systems.

A distributed system consists of a collection of autonomous computers linked by a computer network and equipped with distributed systemsoftware. This software enables computers to coordinate their activities and to share the resources of the system hardware, software, anddata. Users of a distributed system should perceive a single, integrated computing facility even though it may be implemented by manycomputers in different locations. Replication is having multiple copies of data and services in a distributed system. Reasons of replication are: Reliability of the system, Better protection against corrupted data, Improved Performance and faster response time, Facilitates scaling in numbers and geographical area.

The basic motivations for selecting UOG library system, when we visit the library system, librarians suffer by the following main problems 1) frequently losing data, 2) work load in one branch since all data are feed from only one location/branch 3) difficult to extend the system to new branch are some problems. Our new system will focused on improved performance, increased availability, share ability, expandability and access facility because distributed system has an opportunity to those problem. The basic activities in library distributed system will includes; registering books, updating books ,deleting books , borrow books, renew books, and reserve books being any branch or multiple location.

# Problem Statement and Justification

## Currently, UOG Library use Electronic Library Management (ELM) system. All ATewd-lib, Maraki-lib and CMHS-lib Science students who comes to the library for the first time is registered via the system by the data encoder. All books are also registered in the system. As a result of single system, students are expected to clear their status in branch library whether they borrow or not. It is impossible to check status of students and books in all branches being any branch because of undistributed system. Managing those records is highly monotonous process and there are different challenges as well. Like: - difficult to check status of every student being very where, check status of books being everywhere, workload in one branch because of once branch book should enter in that branch only, inability to assure reliability of the system( frequently data is lost). Generally, the existing system is highly tolerated by the following main problems:-

* *Transparent problems*:-because of single system, it is impossible to assure transparency of system. All branch libraries use their own database server so that branch library resource put only in that server.
* *Fault tolerance problem*: - difficult to provide fault tolerance because of lack of backup server. In case of failure of one server, totally all data will be losing.
* *Share ability problem*: it does not allow systems to use each other’s resources.
* *Expandability problem:* it does not permit new systems to be added as members of the overall system. Example:-in UOG there are 2 new branch libraries are established before 3 years, but the system couldn’t be functional in that new branch library.
* *Performance problem*: Resource replication, combined processing power of multiple computers provides much more processing power than a centralised system with multiple CPUs. However, the existing systems do not have any replication. This limits the system improvement performance. Each branch libraries uses only their own computing resource.
* *Reliability and availability problem***:** Disruption would not stop the whole system from providing its services as resources spread across multiple computers. But the disruption of College of Medical and Health Science Library’s server disrupts the other in 2003 like; load problem, report problem, redundancy problem.

# Objective of the project

## 3.1 General Objective

The general objective of this project is to design distributed system for University of Gondar library system.

## 3.2 Specific Objectives

In order to achieve the general objective, the following specific objectives are specifying:

* To analysis UOG library system,
* To identify criteria which helps us determination replica placement in UOG library,

# Scope of project

The scope of the project is case study project for University Gondar library. The project includes the following activities;

* Analysis UOG library,
* Determine placement of replica,
* Identify student attribute,
* Identify book attribute,

# Need assessment

The project team conducted need assessment on selected area. The assessment data were collected from key informants of the student (student library club members), instructors and staffs through unstructured interview and focused group discussion. As the respondents explain, starting from 2003, UOG use Electronic Library Management System (ELM) which is commercial and desktop application software. As its inception, it was implemented in main library (College of Medicine and Health Science Library) only. After some time, the system is implemented in different branch library. As the library staffs describe, the system has no any replica even they don’t back up properly. Because of lack of replica and inability to take back up properly, more than 500,000 data are lost in CMHS Library last year. Because of this the reliability and the availability of the system are very low and poor. As a result, the project teams understand there is a need of developing distributed system which able to curb the listed problems.

# Distribution Goals

Distributed System, Data, Process, and Interface components of an information system are distributed to multiple locations in a computer network. Accordingly, the processing workload is distributed across the network. We have proposed three replicas in our system.

The goals of the distribution are to get the following benefits;

* *Share ability***:** Allows systems to use each other’s computing resource resources,
* *Expandability***:** Permits new systems to be added as members of the overall system. We have planned to extend the system to all new branch library also.
* *Improved performance***:** Resource replication. Combined processing power of multiple computers provides much more processing power than a centralised system with multiple CPUs.
* *Improved reliability and availability***:** Disruption would not stop the whole system from providing its services as resources spread across multiple computers.
* *Potential cost reductions*:-this system can reduce the need of new computing resource since they can share the resource.

# Design decisions

Effective data management in today’s competitive enterprise environment is an important issue. Fast and effective access to data is very important. The actual gains in availability, reliability and performance which are achieved by object replication are a complex function of many factors including the number of replicas, the placement of those replicas, the nature of the transactions performed on the replicated object, the choice of replication protocols and the availability and performance characteristics of the machines and networks composing the system. One of the effective measures to access data effectively in a geographically distributed environment is replication. Replication is one such widely accepted phenomenon in distributed environment, where data is stored at more than one site for performance and reliability reasons. Replication is a strategy in which multiple copies of some data are stored at multiple sites (Bernstein 1987). The reason for such a widespread interest is due to following facts: (i) Increased availability (ii) Increased performance and (iii) Enhanced reliability. By storing the data at more than one site, if a data site fails, a system can operate using replicated data, thus increasing availability and fault tolerance. At the same time, as the data is stored at multiple sites, the request can find the data close to the site where the request originated, thus increasing the performance of the system. But the most important thing is deciding where the replica site should be? Replica-server Placement, Finding best location or placed where a server can be placed and Content Placement, Finding out which server is best for storing a particular content. Based on Distance between clients and locations as starting point (Latency, bandwidth). Best K out of N locations (K<N) are selected. By applying Clustering, Group nodes accessing the same content and with low inter-nodal latencies into groups or clusters and place a replica on the k largest clusters.

*Replication Criteria to be considered for a Replica Management system:*

*Openness:* **-** The replicas should be useful to many requesters, not only a single user.

*Locality***: -** Obtaining a “nearby” replica is preferable. The actual distance (or cost) metric used may include dynamic parameters such as network and server load.

*Addressability***:-**For management, control, and updates, support should be provided for enumeration and individual or group-wise addressing of replicas and Freshness*:***-**The replicas should be the most up to date version of the document.

*Adaptively***:-**The number of replicas for a resource should be adaptable to demand, as a tradeoff between storage requirement and server load.

*Flexibility:-*The number of replicas for one resource should not depend on the number of replicas for another resource. And Variability**: -** The locations of replicas should be selectable.

*State size:-*The amount of additional state required for maintaining and using the replicas should be minimum. This applies to both distributed and centralized state.

*Independence: -*The introduction of a new replica (respectively, the removal of an existing replica) on a node should depend on as few other nodes as possible.

*Performance: -*Locating a replica should not cause excessive traffic or delays.

From what we have explained so far, we have proposed three replica sites and each replica will be placed near to the client. The placement of replicas in the system plays an important role. Secured place will be the main criteria for replica placement.

The other important thing in the design decision is determine process communication protocol. Protocols are the standard rules for communications between processes. Network communications requires protocols to cover high-level application communication all the way down to wire communication. To allow for programming and distributing objects across distributed networks, TCP is used with a middleware system. At the TCP level, all the connections between objects on networked computers appear identical because the details of routing across hubs, routers, etc. are all handled at lower levels. Object communication is through a middleware system. So our process communication protocol is TCP.

Our project design shown below;

File x

File x

File x

File x

File x

CMHS-Lib

Site-5

Maraki-Lib

Site-2

AFasil-Lib

Site-4

Tseda-Lib

File x

ATewd-Lib

Site-1

User 2

User 1

User n

Figure 1: project design

As shown in Figure 1. File X’ is being replicated at all sites. Site 1, Site 2, Site 3, site 4 and site 5 are distributed site locations and connected through a Middleware infrastructure. Data stored in a file, File X, is stored in Site 1 and is replicated at all other sites. Suppose user 1 tries to access File X in Figure 1. For simplicity, let the distance shown in the figure be proportional to the access cost of the file. The above mentioned benefits of replication are clear in this scenario (as Sites 1 and 2 are close to the user in comparison to the other Sites). The files can be accessed at a cheaper cost (thus increasing the performance) and the file can still be accessed even if one of the sites is down (thus increasing availability).

# Architecture

Distributed System, Data, Process, and Interface components of an information system are distributed to multiple locations in a computer network. Accordingly, the processing workload is distributed across the network.

* Set of separate computers that are capable of autonomous operation, link by a computer network.
* Enable individual computers (different location) to share resources in the network
* Server implementation for the same interface located in different servers.

Figure 2: distributed Library system architecture

From what we have explained so far, we have proposed five replicas at each branch library. the replica placement is selected based on the criteria listed in design decision.

**Rules for Distributed system**

* Local autonomy - No site should depend on another site to perform its functions.
* Replication transparency - Users should not be aware of any data replication.
* Hardware independence - A Data must be able to run on different types of hardware.
* Operating system independence - A process must be able to run on different operating systems.
* Network independence - A process must be able to run on different types of networks
* No reliance on a central site - A process should not need to rely on one site more than any other site
* Continuous operation - Performing any function should not shut down the entire distributed database.
* Location transparency - Users should feel as if the entire database is stored at their location.
* Distributed query processing - A process must process queries as rapidly as possible even though the data is distributed.
* Distributed transaction management - A process must effectively manage transaction updates at multiple sites.

.

1. **Conclusion**

Libraries play an important role in the academic world by providing access to world-class information resources and services. It is a time to develop new means of information access and resource sharing with the help of fast growing computer and telecommunication technologies. A distributed system consists of a collection of autonomous computers linked by a computer network and equipped with distributed system software. Frequently losing data, work load in one branch, difficulty of extending the system to new branch are the basic motivations for selecting UOG library system. The general objective of this project is to design distributed system for University of Gondar library system. The project team conducted need assessment on selected area. The assessment data were collected from key informants of the student (student library club members), instructors and staffs through unstructured interview and focused group discussion. The goals of the distribution are to get the following benefits; Share ability, Expandability, Improved performance, improved reliability and availabilityand Potential cost reductions.distance and security are the main factors for deciding placement of replica and number of clients and operation are the factors to determine the number of replica. The other important thing in the design decision is determine process communication protocol. Our process communication protocol is TCP since TCP is used with a middleware system and Object communication is through a middleware system.

# Reference

1. Andrew S. Tanenbaum and Maarten Van Steen,”Distributed System: Principles and Paradigm” Vrije Universiteit, Amsterdam, Netherlands,2006.
2. Daniel McCue and Mark Little, “Computing Replica Placement in Distributed Systems”: A Position Paper for the Second Workshop on the Management of Replicated ,2008
3. J. D. Noe and A. Andressian, "Effectiveness of Replication in Distributed Computer Systems", Technical Report, Department of Computer Science, University of Washington, 2002.
4. N Geetha, ”Distributed Database Management Systems for Information Management and Access”, New Delhi,2004
5. **Appendix**

Herewith we have attached sample interview questions which were conducted between the libraries system administrator and project team members.

1. How many branch libraries do have?
2. What types of service do you providing?
3. Is your service computerized? If so, what type of system do you use?
4. Do you back up your system properly?
5. Before this time, was your system fail? Is so, what was the solution?