10-2001

Analyzing Key Database And Network Issues Of An Institution: A Case Study Of The Proposed MSIS Lab In Dakota State University

Anup Pokhrel

DAKOTA STATE UNIVERSITY

MSIS INTEGRATIVE EXPERIENCE
INFORMATION SYSTEMS PROJECT

Project Approval Form (PAF)

Student Name: Anup pokhrel

SSN [redacted]

Expected Graduation Date: Sept 30

Master's Project Title: Analyzing key D.B & N/W issues of an Institution

Approvals/Signatures:

Student: Anup pokhrel Date: 9·24·01

Faculty supervisor: Wang Date: 10/10/01

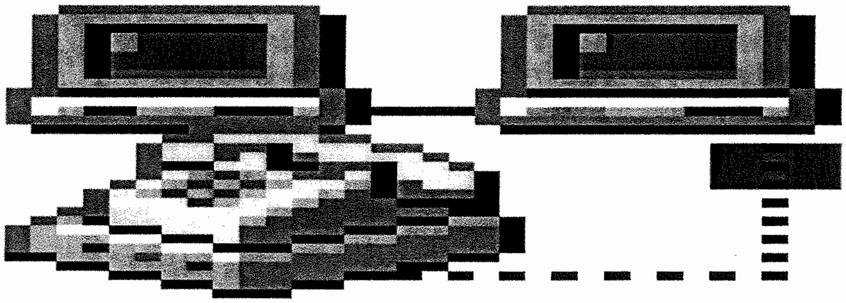
Committee member: Date: 10/10/01

Committee member: Shan Date: 10/10/01

Dean: Richard Christoph Date: ______

Copies to:
Original to Graduate Office; copies to: Advisor, Dean's Office, and Student

Analyzing Key Database And Network Issues Of An Institution: A Case Study Of The Proposed MSIS Lab In Dakota State University



A project report submitted as per the requirements of the MSIS program of Dakota State University, Madison , SD

Submitted by:
ANUP POKHREL
MSIS Student
October 2001

Acknowledgement

I would like to take this opportunity to thank my project committee faculties who during my entire project work have helped me to finish the project work. My project would not have been completed without their inspiring guidance and efforts. I would also like to thank Dr. Terry and Dr. Rick for their help and guidance in enabling me to finish my project work.

My sincere thanks to Prof. Wang, supervisor of my committee for his guidance and supervision in the networking part of my project. I would also like to thank Dr.Shan and Prof. Krebsbach for their guidance in the database part helping me to design and implement the database.

At last I would like to thank my fellow colleagues of the MSIS program who have helped me in this project work testing the prototype model of the project.

CONTENTS

PAGE

Abstract

The project deals with the implementation of a MSIS lab for Dakota State University. I have divided this project into 3 chapters. The MSIS lab concentrates in various networking and database issues as well as some project management skills and concepts.

Chapter 1 of the project report is an outline of the project management skills and the overall concepts of the project.

Chapter 2 is based upon the networking concepts in subnetting, Intranet design and performance analysis of NT and web server using IIS 4.0. The key concepts in this report will be in the design aspects of a subnetting of a network using class C network and finding the corresponding network and host addresses using a JavaScript program. This part of the report also tries to list the design considerations on building an intranet with web-enabled features. The main issue will be the creation of a www based web site with a corresponding URL. A DNS name was created for the translation of the host IP to the URL addresses. A similar FTP site was also created for the Intranet. A prototype of this proposed site has been generated by considering our networking course as a pilot project. The other part in this section was in the performance monitoring of an IIS 4.0 web server and some guidelines in improving the performance of the web server. This part also implemented some of the guidelines to improve the performance for an IIS server to host multiple web pages and FTP sites. Analyzing log files has always been a challenge for web administrators as it can generate a lot of details regarding the performance and security features of the web site. Log files

traditionally are text based files and can be read easily but as the web sites grow in number, the log files starts to grow in size and in a busy site generating reports and storing them in a sequential manner becomes an arduous task, so if this log file was mapped to a relational database then the storing of the log files will be more ordered and structural. This is helpful in generating reports through queries. The parameters that can be generated by these log files include the client's IP address, login time, page, bytes sent and received. By using efficient SQL statements various reports can be generated.

Chapter 3 of the project deals with the database design. This part discusses the need for an efficient backup and recovery system for an ORACLE 8i server. This part also covers the need of a well-tuned ORACLE server. I have also analyzed some tuning parameters for a DSS and OLAP applications using ORACLE.

Chapter 1 Outline

1.1 Work Breakdown Structure[*]

The work breakdown structure (WBS) is a powerful tool for expressing the scope or extent of a project in simple graphic terms. It represents the project in terms of the hierarchy of deliverables and services it will produce. The project is therefore described just as a manufacturer would document the bill of materials breakdown for a washing machine or automobile. The WBS starts with a single box at the top, which represents the whole project. The project is then partitioned into its components with lower level boxes.

This tool enabled me to structure the project into simpler breakdowns and helped me deal with the scope of the project more effectively. The details of the WBS can be found on the Appendix A.

1.2 Pert Chart[**]

A PERT chart is a project management tool used to schedule, organize, and coordinate tasks within a project. PERT stands for Program Evaluation Review Technique. A PERT chart presents a graphic illustration of a project as a network diagram consisting of numbered nodes (either circles or rectangles) representing events, or milestones in the project linked by labeled vectors (directional lines) representing tasks in the project. The direction of the arrows on

[*] APPENDIX A, http://appl.nasa.gov/tools/tools_wbs.htm

[**] APPPENDIX B, http://criticaltools.com/pertmain.htm

the lines indicates the sequence of tasks. A pert diagram of this project can be found at Appendix B.

1.3 System Development Life Cycle

A project normally comprises of mainly five stages:

- Planning
- Analysis
- Design
- Implementation
- Support & maintenance

Planning

This stage dealt with the idea of the project in a nutshell. The planning phase was more of gathering details and vision of the proposed project outline. The first step on my planning part was to gather all the materials and resources from various sources. I then documented my ideas and proposals for the approval from my project supervisor and the faculty members. I briefly described the things I wanted to perform as part of my project work and discussed the necessary measures and resources available on hand.

The next step was to modify the plan according to the suggestions of my project committee members and to review of the paper accordingly.

Analysis

The analysis of the project was divided into two parts:

- Network analysis
- Database analysis

The network analysis was based on the concepts of communication engineering. The first step was to study the network design for future perspectives on the expansion of the network. The network could be expanded so a virtual LAN concept known as subnetting was considered. The process of subnetting involves a great deal of calculation. During this phase of the analysis I felt a need to calculate these addresses, which would be instrumental for system administrators and network engineers for designing the network and assigning a valid IP address.

After the subnetting design I also analyzed a need for a separate Intranet for this proposed lab. The prime objective of my project work was to isolate the college network from the proposed lab network. The intranet analysis phase gave me a clear understanding in setting up a different domain name and security measures for a proper functioning of the network.

The next step in the network analysis was to set up a web server to host web sites for different courses within the Intranet for the students and professors to find or post information. Since there were a number of graduate courses offered for the MSIS program there was a need to setup multiple FTP and WWW sites for each course. In any system as the number of resources and usage increase then there needs to be a scheme to monitor the resources of such system. Therefore considering the volume of the usage of the resources I decided to monitor the performance of the web server increasing some tuning parameters that would increase the performance of the web server. Special consideration was in the analysis of the log files generated. The analysis of the log files is important as we can get a lot of security and web usage information.

The next part of the analysis was in the analysis of the database part. Any intranet system requires a stable and robust database structure. The analysis phase of this part was to come up in creating a database, configuring the database, designing backup and recovery schemes and performance tuning of the database.

Design

The design phase was more of using the ideas generated in the analysis phase and trying to document the details. In the network part the design was concerned in formulating the security schemes and Intranet design. In this phase I formulated the domain name of the server, the course domain names, ftp names of each individual courses. The web server design phase was in the consideration of the performance parameters of the web server, analyzing log files and writing SQL queries for the retrieval of the information from the RDBMS used. The other design aspect was in the security aspects of the network forming groups and subgroup and determining the level of access. I also designed the physical topology of the network and determined the type of cable used and speed and other characteristics of the cable to be used.

The database design dealt in the structure design and the space requirements. This part also dealt with a creation of various objects within a database. In this phase I also tried to analyze the difference between OLTP and OLAP systems and formulate guidelines in implementing these technologies in the server used.

I also designed some parameters on the server that might be useful with a large volume of load and also tried to generalize the solutions that may be needed.

The last part of the design phase was in the formulating the backup and recovery schemes of the server. As an institution grows it needs to protect the data and every organization needs to formulate an effective backup and recovery scheme. This was perhaps the most difficult part of the design phase as there was no clear understanding of the nature of the work to be performed. I did some research in some universities and other institutions and came up with a backup scheme depending upon the nature of the application used. I also did a research in dealing with some unexpected errors that may encounter during the backup process and have tried to recommend some solutions.

Implementation

This phase was in putting the things to work as designed in the previous phase. The network implementation phase was in the installation and configuration of a Windows NT server 4.0. The other part of this implementation phase was in the connection of the cables, configuration of network cards and switches for remote connectivity with other client computers. The web server I used was IIS 4.0 to host the web sites and SQL server was used to map the log files to this database.

In the database part I used ORACLE 8i as my application database server. I successfully installed the server in NT4.0 without using the GUI interface or wizard.

Support and Maintenance

This phase requires a constant monitoring of the resources and maintenance of the resources. A complete guideline can be found at the back of this report under the project implementation report.

Chapter2 Networking

2.1 Subnetting*

Abstract from 3COM Corporation:

In the mid-1990s, the Internet was a dramatically different network than when it was first established in the early 1980s. Today, the Internet has entered the public consciousness as the world's largest public data network, doubling in size every nine months. This is reflected in the tremendous popularity of the World Wide Web (WWW), the opportunities that businesses see in reaching customers from virtual storefronts, and the emergence of new types and methods of doing business. It is clear that expanding business and social awareness will continue to increase public demand for access to resources on the Internet.

There is a direct relationship between the value of the Internet and the number of sites connected to the Internet. As the Internet grows, the value of each site's connection to the Internet increases because it provides the organization with access to an ever-expanding user/customer population.

* http://cache.3com.com/www.3com.com/other/pdfs/solutions/en_US/50130201a.pdf

Internet Scaling Problems

Over the past few years, the Internet has experienced two major scaling issues as it has struggled to provide continuous and uninterrupted growth:

- The eventual exhaustion of the IPv4 address space
- The ability to route traffic between the ever increasing number of networks that comprise the Internet

The first problem is concerned with the eventual depletion of the IP address space. The current version of IP, IP version 4 (IPv4), defines a 32-bit address which means that there are only 2^{32} (4,294,967,296) IPv4 addresses available. This might seem like a large number of addresses, but as new markets open and a significant portion of the world's population becomes candidates for IP addresses, the finite number of IP addresses will eventually be exhausted.

The address shortage problem is aggravated by the fact that portions of the IP address space have not been efficiently allocated. Also, the traditional model of classful addressing does not allow the address space to be used to its maximum potential. **The Address Lifetime Expectancy (ALE) Working Group of the IETF has expressed concerns** that if the current address allocation policies are not modified, the Internet will experience a near to medium term exhaustion of its unallocated address pool. If the Internet's address supply problem is not solved, new users may be unable to connect to the global Internet!

The second problem is caused by the rapid growth in the size of the Internet routing tables. Internet backbone routers are required to maintain complete routing information for the Internet. Over recent years, routing tables have experienced exponential growth as increasing numbers of organizations connect to the Internet - in December 1990 there were 2,190 routes, in December 1992 there were 8,500 routes, and in December 1995 there were 30,000+ routes. Unfortunately, the routing problem cannot be solved by simply installing more router memory and increasing the size of the routing tables. Other factors related to the capacity problem include the growing demand for CPU horsepower to compute routing table/topology changes, the increasingly dynamic nature of WWW connections and their effect on router forwarding caches, and the sheer volume of information that needs to be managed by people and machines. If the number of entries in the global routing table is allowed to increase without bounds, core routers will be forced to drop routes and portions of the Internet will become unreachable!

The long-term solution to these problems can be found in the widespread deployment of IP Next Generation (IPng or IPv6) towards the turn of the century. However, while the Internet community waits for IPng, IPv4 will need to be patched and modified so that the Internet can continue to provide the universal connectivity we have come to expect. This patching process may cause a tremendous amount of pain and may alter some of our fundamental concepts about the Internet.

Classful IP Addressing

When IP was first standardized in September 1981, the specification required that each system attached to an IP-based Internet be assigned a unique, 32-bit Internet address value. Some systems, such as routers, which have interfaces to more than one network, must be assigned a unique IP address for each network interface.

The first part of an Internet address identifies the network on which the host resides, while the second part identifies the particular host on the given network. This created the two-level addressing hierarchy, which is illustrated in Figure 2.1

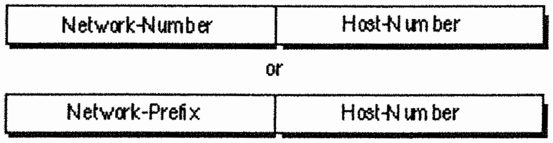


Figure 2.1 Two-Level Internet Address Structure

In recent years, the network-number field has been referred to as the "network-prefix" because the leading portion of each IP [1]address identifies the network number. All hosts on a given network share the same network-prefix but must have a unique host-number. Similarly, any two hosts on different networks must have different network-prefixes but may have the same host-number.

Primary Address Classes

In order to provide the flexibility required to support different size networks, the designers decided that the IP address space should be divided into three different address classes - Class A, Class B, and Class C. This is often referred to as "classful" addressing because the address space is split into three predefined classes, groupings, or categories. Each class fixes the boundary between the network-prefix and the host-number at a different point within the 32-bit address. The formats of the fundamental address classes are illustrated in Figure 2.2

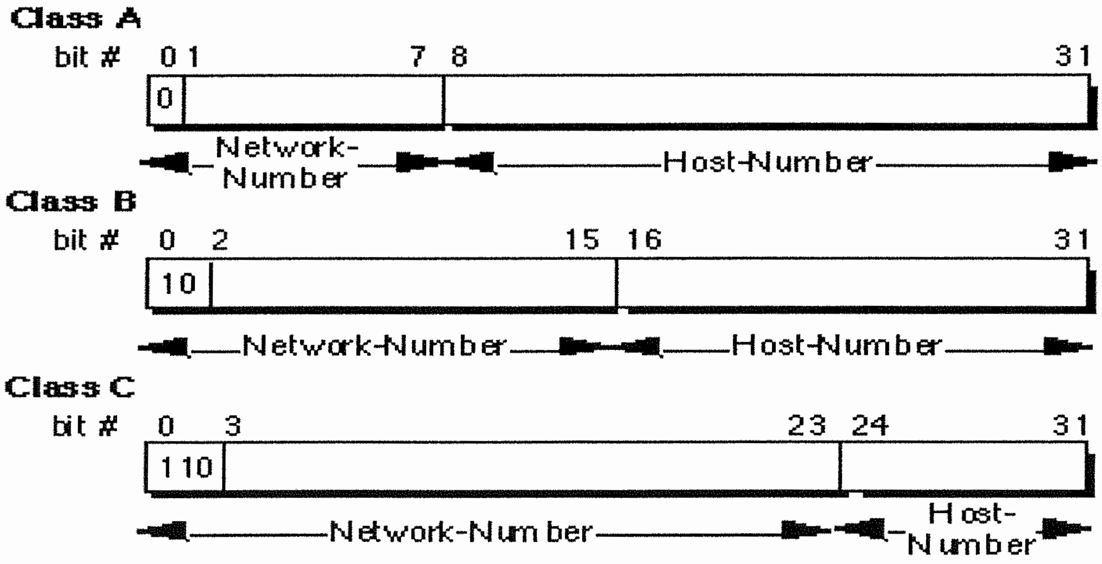


Figure 2.2 Principle Classful IP Address Formats*

One of the fundamental features of classful IP addressing is that each address contains a self-encoding key that identifies the dividing point between the network-prefix and the host-number. For example, if the first two bits of an IP address are 1-0, the dividing point falls between the 15th and 16th bits. This simplified the routing system during the early years of the Internet because the original routing protocols did not supply a "deciphering key" or "mask" with each route to identify the length of the network-prefix.

* http://cache.3com.com/www.3com.com/other/pdfs/solutions/en_US/50130201a.pdf

Class C Networks (/24 Prefixes)

Each Class C network address has a 24-bit network-prefix with the three highest order bits set to 1-1-0 and a 21-bit network number, followed by an 8-bit host-number. Class C networks are now referred to as "/24s" since they have a 24-bit network-prefix.

A maximum of 2,097,152 (2^{21}) /24 networks can be defined with up to 254 (2^8 -2) hosts per network. Since the entire /24 address block contains 2^{29} (536,870,912) addresses, it represents 12.5% (or 1/8th) of the total IPv4 unicast address space.

Subnet Design Considerations*

The deployment of an addressing plan requires careful thought on the part of the network administrator. There are four key questions that must be answered before any design should be undertaken:

1) How many total subnets does the organization need today?

2) How many total subnets will the organization need in the future?

3) How many hosts are there on the organization's largest subnet today?

4) How many hosts will there be on the organization's largest subnet in the future?

The first step in the planning process is to take the maximum number of subnets required and round up to the nearest power of two. For example, if an organization needs 9 subnets, 2^3 (or 8) will not provide enough subnet addressing space, so the network administrator will need to round up to 2^4 (or 16). When performing this assessment, it is critical that the network administrator always allow adequate room for future growth. For example, if 14 subnets are required today, then 16 subnets might not be enough in two years when the 17th subnet needs to be deployed. In this case, it might be wise to allow for more growth and select 2^5 (or 32) as the maximum number of subnets.

The second step is to make sure that there are enough host addresses for the organization's largest subnet. If the largest subnet needs to support 50 host addresses today, 2^5 (or 32) will not provide enough host address space so the network administrator will need to round up to 2^6 (or 64).

The final step is to make sure that the organization's address allocation provides enough bits to deploy the required subnet-addressing plan. For example, if the organization has a single /16, it could easily deploy 4-bits for the subnet-number and 6-bits for the host number. However, if the organization has several /24s and it needs to deploy 9 subnets, it may be required to subnet each of its /24s into four subnets (using 2 bits) and then build the Internet by combining the subnets of 3 different /24 network numbers. An alternative solution would be to deploy network numbers from the private address space (RFC 1918) for internal connectivity and use a Network Address Translator (NAT) to provide external Internet access.

It is a known fact that subnetting of a network is not an easy task, which involves a lot of calculation and logic to successfully implement the network addresses. To overcome this problem I have designed a program in JavaScript

* http://cache.3com.com/www.3com.com/other/pdfs/solutions/en_US/50130201a.pdf

which checks the IP address (I have taken into account only the class C addresses) and then after the user selects the number of subnets then displays the number and range of IP addresses and host addresses thus a person without the knowledge of the calculation process can see the subnet configuration details.

Subnetting an Intranet

When intranets are over a certain size, or are spread over several geographical locations, it becomes difficult to manage them as a single network. To solve the problem, the single intranet can be subdivided into several *subnets*; subsections of an intranet that make them easier to manage. To the outside world, the intranet still looks as if it's a single network.

If you're building an intranet and want it to be connected to the Internet, you'll need a unique IP address for your intranet network, which the InterNIC Registration Services will handle. There are three classes of intranet you can have: Class A, Class B, or Class C. Generally, a Class A rating is best for the largest networks, while a Class C is best for the smallest. A Class A network can be composed of 127 networks, and a total of 16,777,214 nodes on the network. A Class B network can be composed of 16,383 networks, and a total of 65,534 nodes. A Class C network can be composed of 2,097,151 networks, and 254 nodes.

When an intranet is assigned an address, it is assigned the first two IP numbers of the Internet numeric address (called the *netid* field) and the remaining two numbers (called the *hostid* field) are left blank, so that the intranet itself can assign them, such as 147.106.0.0. The hostid field consists of a number for a subnet and a host number.

When an intranet is connected to the Internet, a router handles the job of sending packets into the intranet from the Internet.

When intranets grow-for example, if there is a department located in another building, city, or country-there needs to be some way to manage network traffic. It may be impractical and physically impossible to route all the data necessary among many different computers spread across a building or the world. A second network-called a *subnetwork* or *subnet*-needs to be created.

In order to have a router handle all incoming traffic for a subnetted intranet, the first byte of the hostid field is used. The bits that are used to distinguish among subnets are called *subnet numbers*. In our example, there are two subnets on the intranet. To the outside world, there appears to be only one network.

Each computer on each subnet gets its own IP address, as in a normal intranet. The combination of the netid field, the subnet number, and then finally a host number, forms the IP address.

The router must be informed that the hostid field in subnets must be treated differently than non-subnetted hostid fields, otherwise it won't be able to properly route data. In order to do this, a *subnet mask* is used. A subnet mask is a 32-bit number such as 255.255.0.0 that is used in concert with the numbers in the hostid field. When a calculation is performed using the subnet mask and the IP address, the router knows where to route the mail. The subnet mask is put in people's network configuration files.

The first part of the project was to implement the idea of subnetting also known as VLAN(Virtual Lan). This concept assumes the fact that even if two computers are next to each other and are connected to a same hub they may not be able to communicate, as their network id may be different based upon the subnet masks.

It is not an easy task to determine the subnet mask and the corresponding number of networks needed to make a subnet. Often the network administrator needs to calculate the corresponding networks and the hosts per network. I have tried to implement this logic by designing a user interface to calculate the number of hosts and networks using a simple JavaScript program.

Features of this Program

Checks to see if the given IP address belongs to class C address.

Asks the user to input the number of networks to be determined.

Calculates the new IP address and the corresponding subnet mask and the number of networks and hosts that can be calculated.

2.2 Intranet Design

An Intranet is a network within an organization. It differs from its counterpart Internet in that Internet is a global issue with network of computers worldwide whereas the Intranet is concerned with only the computers within the organization. There are distinct advantages and disadvantages of the design and implications involved. An Intranet is supposed to have a less number of users and security may not be the prime concern whereas in the case of the Internet security is the only concern.

As far as the design for the MSIS lab is concerned, considering the type of the courses and the level of security, I have tried to implement an intranet first on the basis of dynamic IP address allocation. As long as the lab is in one room there is no need for a subnet. But considering the future prospects where the lab may be expanded then the concept of subnetting must be used.

Assigning domain names was one of the challenges and using DNS server as the intermediate, different names for the network were used. I had taken into account the case of the networking concentration and named the sites as www.courses.dsu.edu/infs750 ....../753 etc. A virtual directory for these courses were also created and named according to the course title. FTP sites for these courses were also created with different level of permission granted to the instructor and the students.

A NT 4.0 server was implemented to configure the groups and subgroups. I implemented the concept of local and global groups. A local group was used to give permission to a local group whereas a global group was used for logical grouping of the users within the domain. The main difference between these two is the fact that a local group cannot travel across other domains whereas the global groups can.

A security model within the NT server was implemented for the users to browse only the required folder. A student enrolled in a course INFS750 was given a read access to this folder and the instructor of this course was given full access so that he can put on new materials for the students to use.

A similar approach was also considered for the FTP site implementing the same logic. If a student registers for more than one course then a same password can be used to access all the courses he has registered. A FTP site is very helpful in downloading materials from the desired site by using FTP commands.

2.3 Web Server Analysis

The analysis of web server is an important topic as monitoring resources has always been a challenge.

Until now, Web servers did not focus on performance as an intrinsic design premise. This was acceptable during the early adopter phase of the Web, as most Web sites serviced a limited load over low-capacity links and less demand was placed on the servers. Today, the feature set for Web servers has stabilized, commercial implementations are available, and the importance of Web server performance is increasing. Several reasons that contribute to the increasing focus on performance include:

- Scalability: Customers require their Web servers to scale in many dimensions. Web servers must be easy to install and manage; yet powerful enough to deliver the next generation of dynamic, or "active" sites. Moreover, since Web servers may initially be installed as the only application, they must be efficient enough to leave system resources for database and mail server applications.

- Server Extensions: Web servers are now frequently used to access or coexist with other server-based applications that can run on the same hardware. Web applications for database publishing, content indexing, and collaboration are now common on many Internet servers.

- Enterprise Webs: Many Web servers will be installed on internal local area networks, and will therefore no longer be limited by the relatively low-bandwidth connections to the Internet to which most servers are presently attached. The typical Intranet server will be on a shared 10-100 Megabyte (MB) network.

- Multipurpose Servers: Internal and external Web servers are no longer being used as dedicated Web servers; other server applications are often run on the same server system. High-performance Web server software leaves more CPU and memory resources available for other applications and custom extensions.

- Dynamic Content: As the Internet moves to more dynamic content--3-D, video, audio, and custom-generated Web pages, the amount of required compute power to build and serve this next generation of content increases dramatically.

- Infrastructure Improvements: The performance of the Internet infrastructure itself is improving as more users move to higher speed modems and ISDN lines. As the speed of the communications infrastructure increases, and the total number of users with Web access expands, the requirement for high-speed servers grows proportionately.

Web based servers are characterized by a number of factors among them the following are very important in the enhancement of web servers used today.

- Throughput--in megabits per second. Measures the maximum rate at which the Web server transfers data to the client.

- Connections per Second: Depicts the sum of successful interactions between all clients and the Web server.

- Average Response Time in Seconds: The total amount of time required to complete an operation. This is also called *Latency* in some Web performance tests.

- API Performance: The relative performance of alternative APIs to execute a given amount of work.

- Cross-platform Tests: A comparison of previously published WebStone performance results with the results obtained by these set of tests.

- Errors per second: This is defined as any failure in attempting an interaction with the server from the client. This could be a "Connection Refused" error when attempting to create a TCP/IP connection, or a time-out on creating the connection, or receiving less or more data from the server than is expected. A lower number indicates better performance. Microsoft Internet Information Server and Netscape Communications Server for Windows NT did not report any errors during the tests. Both Netscape Communications Server for BSDI UNIX and the Novell NetWare Web Server reported a significant number of errors as the number of clients increased.

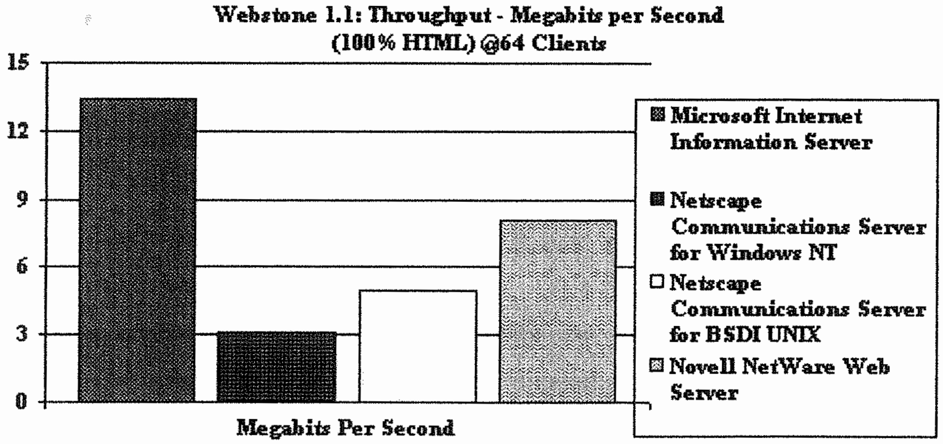
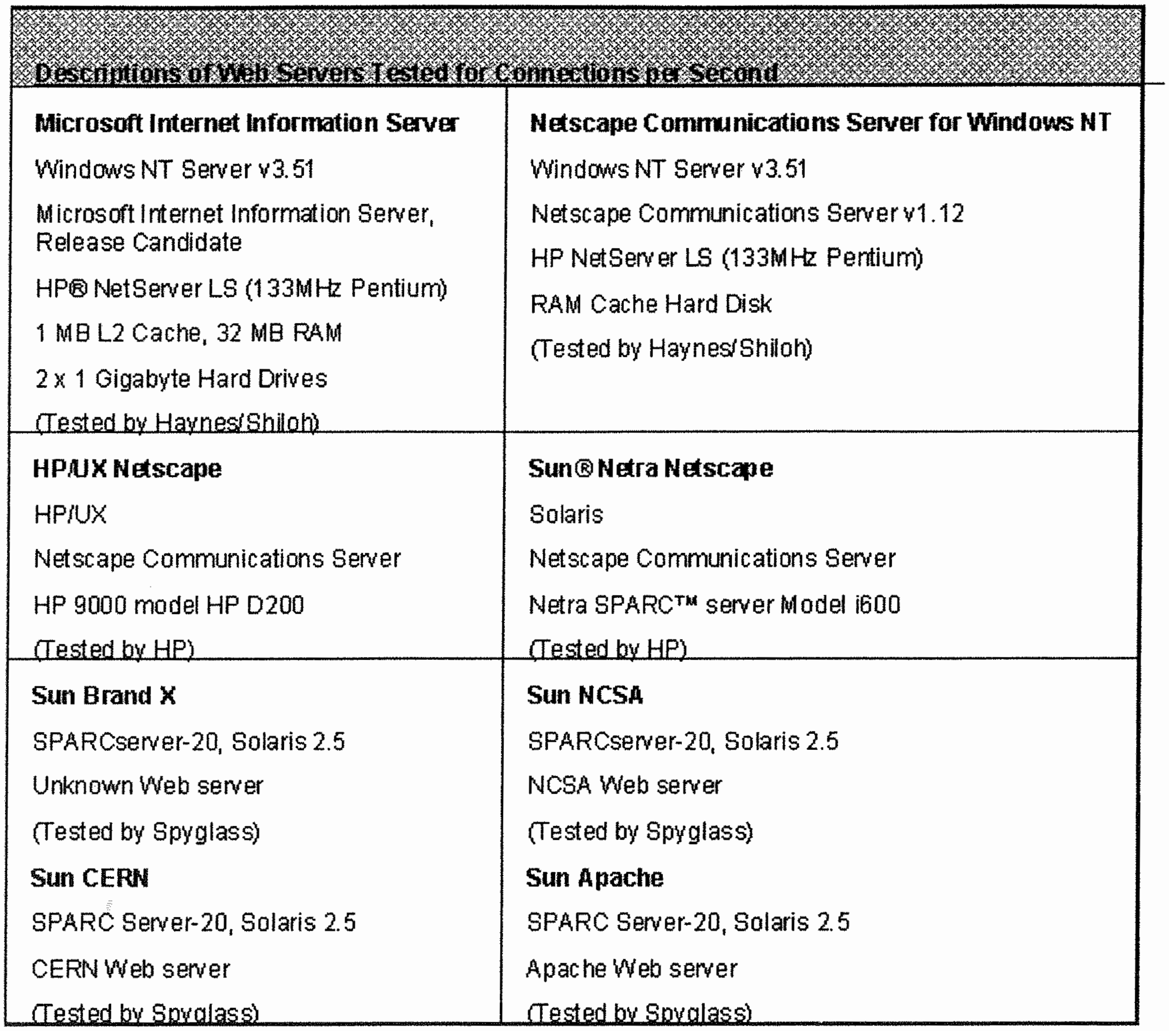


Figure 1.3 Throughputs[*]

I have included a test benchmark test that was conducted to justify why I chose IIS server for my project work.



Descriptions of Web Servers Tested for Connections per Second	
Microsoft Internet Information Server Windows NT Server v3.51 Microsoft Internet Information Server, Release Candidate HP® NetServer LS (133MHz Pentium) 1 MB L2 Cache, 32 MB RAM 2 x 1 Gigabyte Hard Drives (Tested by Haynes/Shiloh)	**Netscape Communications Server for Windows NT** Windows NT Server v3.51 Netscape Communications Server v1.12 HP NetServer LS (133MHz Pentium) RAM Cache Hard Disk (Tested by Haynes/Shiloh)
HP/UX Netscape HP/UX Netscape Communications Server HP 9000 model HP D200 (Tested by HP)	**Sun® Netra Netscape** Solaris Netscape Communications Server Netra SPARC™ server Model i600 (Tested by HP)
Sun Brand X SPARCserver-20, Solaris 2.5 Unknown Web server (Tested by Spyglass) **Sun CERN** SPARC Server-20, Solaris 2.5 CERN Web server (Tested by Spyglass)	**Sun NCSA** SPARCserver-20, Solaris 2.5 NCSA Web server (Tested by Spyglass) **Sun Apache** SPARC Server-20, Solaris 2.5 Apache Web server (Tested by Spyglass)

Figure 2.4[*]

Internet Information Server 4.0 Tuning Parameters for High-Volume Sites:

[*] National Software Testing Laboratory (NSTL) www.graphcomp.com/info/specs/isapi

[*] Microsoft and Michael Stephenson and the Windows NT Server Performance Team http://www.microsoft.com/ntserver/techresources/deployment/web/tuning.asp

A web server is mainly based upon various factors. From the OS level to the applications used. There are a number of parameters that can be set to maximize the performance of the web server. An example of the physical resources will be to increase the amount of RAM or use fast disk controllers like SCSI.

There are some parameters that can change in the existing NT system environment that can enhance the performance of the web server. The general tuning parameters are:

- Set NT as an application server so that NT trims the file cache more aggressively
- Remove irrelevant mappings from the IIS server and keep only the .asp extension
- Dropped packets on the receiving end cause TCP to retransmit we can set the NIC properties to maximum to avoid this problem
- Set tcpip parameters in the registry
- Set time to live for web pages to or about 30 secs
- Set open file in cache to a value of about 1000 for 32 mb of memory
- Minimize the session time out value

2.3 Web Server Analysis*

Although you can use the Windows NT Event Viewer to look at the security and system logs, these logs do not contain as much information about who is doing what with your Internet services as the Internet Information Server logs. In a busy site it is quite impossible to generate all the required information

* Best OS Professor Reference MS Back Office Administrator's Guide

Your IIS logs contain the following fields: A SQL server database is created to host these field values and a ODBC connection is established to map these entities to the database and generating query for the output.

- Client host: Specifies the IP address of the connected client.

- Username: Specifies the name supplied by the user to log on to the specified service.

- Log time: Specifies when the user started the connection in MM/DD/YY HH:MM:SS time format.

- Service: Specifies the service to which the user connected. This will be MSFTPSVC for the FTP Publishing Service, W3SVC for the WWW Publishing Service, or GopherSvc for the Gopher Publishing Service.

- Machine: Specifies the IIS server on which the service is executing.

- ServerIP: Specifies the IP address of the server on which the service is executing.

- Processing time: Specifies the time the client was connected.

- Bytes sent: Specifies the number of bytes sent to the client.

- Bytes received: Specifies the number of bytes sent by the client.

- Service status: Specifies a service-specific status code.

- Win32Status: Specifies a Windows NT 32-bit API-specific status code.
- Operation: Specifies the service operation type for the client request.
- Target: Specifies the data object name requested by the client.
- Parameter: Specifies any additional command parameter.

By saving these logs and reviewing them (at least once a week), you can obtain performance characteristics such as how many users visited your site. You can also use these logs to determine if someone was trying to access data on your site that they should not have access to by looking in user logon failures field. If you see several of these from the same IP address, then you may want to prevent that IP address from logging on to your site by configuring the advanced options for the service. I have created a database in SQL server and mapped the text based log files to this database with ODBC configuration. Some queries that can be generated to find some useful information would be:

Select Max(logtime) from ftplog ⇒ to see the last logged on user time

Select username, clienthost from inetlog ⇒ to see the client IP address and the username who have logged on to the system.

Chapter 3 Database

This section covers the key issues in the server side of database systems and some guidelines in the backup and recovery of ORACLE server 8i. This section also covers the performance tuning of an ORACLE 8i server and a general comparison in the performance analysis (tuning) of a DSS and an OLTP system in designing an ORACLE database. This section is based on a manually created database named anuplab with arbitrary defined memory parameters. This database consists of manually created table spaces to hold specified types of data including temporary Tablespaces, Tablespaces for the users, indexes and rollback back segments. The other key features of this database is that each user is given some quota in the Tablespaces and a profile is set up to limit the disk usage of the users in the system. The users will be monitored in password checking features like failed login attempts, password life time, password grace time while the profile will be used to limit the sessions per user, connection time, idle time.

The tuning guidelines part of this section mainly tries to compare the difference between a DSS/OLAP and OLTP systems and some considerations regarding the tuning parameters of such type of systems. The concentration in this part was to analyze the key resources used by such systems and making changes accordingly.

3.1 Backup and Recovery[*]

In developing a backup-and-recovery strategy, you must first look at the parameters of your system. Many variables affect the backup strategy and the backup process, including but not limited to

- Time allotted for the backup
- Amount of data to be backed up
- Speed of the backup device(s)
- System downtime allotted for backups (if any)
- Off hours (if no downtime allotted)
- Performance degradation allowed

These factors and others will affect the backup strategy. They must be measured against the main goals of the backup process, which include

- Protecting the database by having a current backup.
- Reducing the time necessary to recover.
- Affecting the performance of the system as little as possible.
- By putting together all these factors, you should be able to come up with a backup strategy that meets as many of these requirements as possible.

When and What to Back Up

Scheduling backups is usually not a very easy task. Backups should be taken often enough to protect the system's data from losses, yet should not interfere with normal business operations. Depending on the size of the data that needs to be backed up and the operational schedule of your business, you might

[*] http://pcsupport.about.com/cs/backuprecovery/

choose different options. The options vary, depending on which type of operation your business runs:

- 5x8 Operation: This term designates a business whose corporate data must be available 5 days a week, 8 hours a day.
- 7x24 Operation: This term designates a business whose corporate data must be available 7 days a week, 24 hours a day. No downtime is allowed.

Each of these types of operation has different backup needs, depending on the type of data and the frequency of modifications to that data. Let's look at those cases.

Scheduling Backups in a 5x8 Shop

Scheduling backups is a little easier in a 5x8 shop because there is plenty of time when the system need not be available to the users and can be backed up without disturbing them. The frequency of the backups in this type of operation can depend in some part on the activity in the database. Here are some guidelines:

Backup Database

- A Small database

For a small database, perform a full offline backup every night. This will offer the most protection possible for your data.

- A Large database

If the database is very large, perform a full offline backup every weekend and incremental backups during the week.

- An Active Tablespace

If certain Tablespaces are very active, we should ensure that these Tablespace are backed up as often as possible using online Tablespace backup methods.

Structural changes

Any time a structural change is made to the database, you should perform a full offline backup.

Unrecoverable operations

Any time unrecoverable operations are performed on the database, a full offline backup should be done.

Reset logs

Anytime that you have to reset the redo logs with the RESETLOGS command, you should perform a full offline backup.

Recover until

Any time you have recovered with the Recover until option, you should perform a full offline backup.

Archive log files

Every night or every few nights, one should back up the archive log files. The frequency depends on the number of archive log files created every day.

Backup methods in ORACLE

Backup and recovery are the two most important topics in ORACLE as data and other configuration parameters are very essential. As ORACLE consists of various important configuration parameters and set of files, we should ensure that the appropriate files are stored in a secure environment.

The basic important files used in ORACLE are:

- Parameter file⇒stores the database information
- Password file⇒stores the password including that of internal
- Control file⇒contains the physical properties and information about data and redo log files
- Data files⇒actual physical files that store data
- Redo log files⇒ files that contain the changed information
- Archive files⇒it is the backup of the redo log files

These files are binary files so that the structure will be in a non-readable format. It should be ensured that these files are backed up in regular intervals especially the control files and the data files which change according to the system changes.

Closed database backup (Cold backup)

The dictionary files are V$ data file, V$ table space, dba data files, V$ control file, before performing the backup we should check and make sure the physical location of these files.

Closed database backup is performed by copying all the files required to a backup destination(in OS level). Examples might be copying c:\ORACLE\...... to d:\ORACLE...\...... for Microsoft supported operating systems and cp /orahome/bin/.. /usr/home.... for an UNIX operating system.

Required files

Archive is a redo log file's backup file.If the database is not in an archive made there is no need backup this file.

Open backup

It is quite easy and simple to back up all the files when the database is closed but most of the times the database is in constant use and may not have any time to shutdown then in this case we have to do an online backup:

The key point in doing the online backup mode is that:

- Tablespace wise backup (one by one) should be performed.
- The database should be running in the archive log mode

Syntax:

- Alter Tablespace DATA begin backup;
- $ Copy D:\ Loc data file D:\ Loc (backup the data file containing the Tablespace into your own directory)
- Alter Tablespace DATA end backup;

We should repeat these steps for all the Tablespaces that are in the database.

The online backup is used for the data file but the control files are also very important in the database

Control file backup

- Alter database backup control file to
- ' D:\loc……… '; --→Location or we can use the command
- Alter database backup control file to TRACE (udump destination);

The udump destination file is specified in the parameter file

Backup of a read only Tablespace

If a database is made read only then one backup in needed.

Syntax: alter Tablespace <t.bname> read only →then backup is needed once.

Recovery

There may be certain instances when we have a backup and some unwanted incidence occurs then in that case we have to be familiar with a certain criteria for recovery.

Recovery without archive

If a database is not in archive mode then and we might lose our files or cannot start up the database then the following procedure should be adopted:

- Restore all the files from backup (data files, control files and redo log)
- Startup the instance

Archive log mode

If the database is in archive log mode then there may be two possibilities:

- Closed database recovery
- Open database recovery

Here too we can classify into two subtypes i) Initially open ii) Initially close

Closed

- Restore only the damaged file from the backup
- Mount the database
- Recover data files 'D:\loc.'

Open database recovery (initially open)

- Alter database data files 'D:\loc '___' offline
- Alter Tablespace (backup files) < > offline
- Restore manually
- Recover data file D: \loc…
- Alter Tablespace user_ data online;

Open database recovery initially closed

- Restore (from the dos prompt then copy D:\)
- Mount the database
- Offline the data file (alter database datafile < location > offline
- Alter database open
- Recover datafile <location>
- Bring the datafile online <alter database datafiles>

Scenarios:

Performing backups are not very complicated and recovery can be as easy as mentioned as above but the database is volatile meaning that we should never expect the database will be up and running all the time and we can follow the guidelines in most general cases but some times unexpected errors might occur and we should be prepared to be able to restore the data in such type of situations:

Instance crash

If there is an instance crash then

- Shutdown abort (like normally closed)
- Startup (SMON process automatically builds up the database and detects any failure)
- Smon also checks for fragmentation in Tablespaces. So we do not need to perform any recovery as the processs itself fragments the corrupt instance

Dead lock

When there is a deadlock in the database then:

- Shutdown abort then
- Startup restrict mode
- Shutdown normal

Disk cannot be replaced

If there is a crashed disk and it can't be replaced,

Process:

- Restore
- Mount
- Rename
- Offline (table space)
- Open

Data file deleted

If a Database is in archive log mode, data file is deleted which has no backup.

Then the following command can be issued to recover:

- Alter database create data file D:\ORACLE\....

Data file crashed which has no backup

Data file in D:\loc.. is crashed , which has no backup. Then the following command can be issued to recover the datafile.

- Alter database create data file ….as …< > new location

Power failure while backing up

When we are trying to backup the database and we find out that a power failure has occurred then the following steps should be followed:

- Startup
- Recover
- Mount
- Alter database end backup
- Alter database open

DB belonging to system Tablespace is shutdown

If a Database belonging to a system Tablespace is shutdown. The following process should be followed.

- Restore
- Mount
- Offline data file
- Open
- Recovery

RBS crashed

If a Data file is crashed which contains a rollback segments then the following process should be followed.

- Restore
- Mount
- Recovery
- Offline
- Open
- Recover
- Online

Incomplete recovery

Suppose there is a incomplete recovery that needs to be recoverd:

If we happen to delete the EMP table at wed 11:45 and we had backup till Monday then the following process should be followed

- Restore data file
- Database must be in archive log mode
- Mount the database
- Recover database until time ' 2000:04:21:12:30:00(+ 12 if pm)

OR recover database until cancel;

- Alter database open reset logs (synchronization)
- Backup (after rest log is applied)

A compared to the physical backup and recovery process there are a certain number of backup and recovery tools available in ORACLE they are mainly export-import

Logical Backup / Recovery

- Tables indexes
- Export, import

Export (modes)

- Table
- User
- Full

3.2 Performance Tuning:*

Why do we need performance tuning?

As the organization grows in size the need for more resources becomes a necessity. ORACLE consumes a lot of memory and when the instance is started then the some of the OS memory is allocated to the ORACLE application.

The two most important reasons why we should tune a system is:

- For faster access
- Better resource management

General tuning guidelines:

- First tune application and then memory, disk and IO
- Do not use screen savers as CPU resources are utilized
- Set virtual memory twice the physical memory

* http://www.csee.umbc.edu/help/oracle8.bak/server803/A54638_01/toc.htm

- Have separate page files on dedicated disks for NT
- Use SCSI disk as NT can multitask with SCSI at once

We can tune the ORACLE database by considering the following four areas:

- Application tuning --- deals with the design (normalization) phase, SQL
- Memory tuning--- allocating the shared pool, db buffers
- I/O tuning----physical disks consideration
- Contention tuning—deadlocks, locks and latches

Guidelines For Application Tuning

- Denormalize for better application tuning
- Write SQL using index
- Use Indexing in where clause columns
- Use Indexing on foreign key column or DML statements that are frequently used.

Memory Tuning

There are certain parameters that can be monitored while analyzing the memory parameter: Query the v$library cache and if the ratio of reload: pin <1% then we have to increase the size of the shared pool.

V$ row cache

- Get misses – are records that are not found
- Gets- records that are found

Ratio get misses/ gets < 50%

If not then increase area of shared pool size

I/O Tuning

- Disk reads should be made minimum
- Distribution of disk IO load should be among all the disks

Contention

- Prevent deadlock
- Use locks, as there are two types in ORACLE: -row shared and row exclusive

3.3 Comparison between an OLAP and OLTP

The last part of the project tries to formulate the differences between a DSS/OLAP and OLTP systems and to recommend the tuning parameters for such types of systems:

The characteristics of DSS systems:

- Multi dimensional conceptual view
- Analysis capabilities
- Shared access
- More data as compared to OLTP
- More joins and aggregation is involved
- Generate extensive use of server resources
- Data are historical and summarized
- Queries are extensive
- Performance metric in terms of query throughput and response times

- Example: sales analysis, promotion planning, costing in manufacturing, inventories and warehousing

Characteristics of OLTP systems:

- Many users access data together
- Few queries
- High number of updates, inserts
- Contention should be monitored
- Data should be up-to-date and details
- Tasks are repetitive
- Performance metric transaction throughput
- Examples: airlines reservation systems, banking systems

The general tuning parameters for these systems are based upon their nature of operation. OLTP requires a heavy transaction processing whereas DSS does not require such processing. There are some guidelines for these systems:

OLTP

- Frequent DML
- Db- block size = maximum (as per O/S 64kb)
- Rollback segment should be more and of large size
- Shared pool size should be made higher
- Checkpoint processing should be made at regular intervals
- Online backups daily

DSS/OLAP

- Db_block_size=minimum 2kb

- Rollback segment smaller size but many
- Few updates so no need to keep in achieve log mode
- Increase sort area size can be done through the temporary Tablespaces
- Fast controllers for data access
- Indexes in separate tablespaces

Chapter 4 Conclusion

4.1 Conclusion

The prototype of the model was built successfully and was demonstrated. A NT server was configured to host the WWW and the FTP sites. The project was completed by testing all the sites and verifying that the databases were running. The final outcome of this project was to study and recommend some guidelines to the current issues in networking and database. The general guidelines and recommendations were implemented in designing a separate lab for MSIS students in DSU. The next part covers the design and implementation of the project.

4.2 Project Implementation Report: MSIS Lab Proposal

MSIS Program Overview

The MSIS program in DSU is now almost two years old. Over these years there has been a significant increase in the number of students. Being in the program for over twelve months and with a wide range of interactions with my colleagues and faculties I have a proposal that analyzes the need for more IT resources to increase the IT skills of the students. I propose to set up a MSIS lab dedicated specifically to MSIS students and I have built a prototype of such a proposed system. Since the MSIS program admits students from different undergraduate programs there can be some difficulties for the students from non computer related discipline to cope with the practical aspects of IS areas. The need for more hands on experience is a must for all the students to succeed in the job market.

Current IT Resources Available

Although there are some resources available for MSIS students, I feel that those are not adequate in graduate level to conduct some research-oriented activities. The only lab that can be used for high level programming and research is located in East Hall and is open from 5-11 in the evenings 4 days a week. The timing of the lab coincides with the timing of most of the classes from beginning from 6-10. The other labs do not offer any facilities for research and new tests. The Sun lab that we have access 24 hours has some limitations in manipulating the server and workstations. The objectives of my paper and work will be to establish a MSIS lab strictly for MSIS students open 24 hour a day 7 days a week where students have full rights to implement their logic and ideas.

Our program concentrates mostly on three areas Ecommerce, Networking and Database. With the changing technology and systems for these fields we need constant upgrading of our servers and workstations. These changes cannot be implemented through the college wide network as these changes may conflict with the whole college network and may bring the network down. If a new lab is established which is isolated from the college network then upgrading of the server and computers will be easier.

Need for a MSIS Lab

- Faculties can also use this lab for their research.
- Students are not fully exposed to the products in the market e.g. SQL Server, NOVELL as we have to depend on the installed products only.
- Students will have all the necessary tools to broaden their skills.
- Can help students conduct research.
- Will be instrumental in the project work. There is a problem for students in doing their project due to the lack of resources so a new MSIS lab will be the basic platform for the students to start project work.
- Students can install new feature, implement their ideas effectively.

Structure of the New Lab

- A server with standard network operating systems, web server, database in ORACLE and SQL server. I have installed a server with all these features installed and have tested all the functionality aspects.

- An Intranet builds up with a web URL and necessary security measures.
- 10-15 workstations hooked up to the intranet with latest client side features installed.
- Switches are available.
- A stand by server for students to test their creative innovations and for students to learn practical knowledge in ecommerce, database and networking fields. When the products are tested then can be implemented in the main server.

Expectations of the New Lab

- If the proposed lab is ready then 2 Graduate Assistants can be given overall responsibility of the lab. They should ensure that the lab is functional. One day of the week should be reserved for teaching other interested students of the whole operation of the system. Each semester these Graduate Assistants can be changed so that most students get feel of a real system before they search for jobs.
- Professors should involve their Graduate Assistants in more research work and this lab can be used for their research work too.
- Students can contact the responsible Graduate Assistants to check if the resources they want can be made accessible.
- The Two Graduate Assistants should maintain contact with Computing Services, faculties for any resources needed.
- Suitable System security measures must be implemented for smooth operations of the lab.

- Students are finding it very difficult to find project topics the new lab can be used for research topics such as VLAN, Subnetting, Super netting, Web server analysis, parallel server, database design analysis security of web and many more research areas.

Some Questions and Answers Regarding the Lab

Q1. What is the initial cost?

Practically zero, only a room is needed as we have most of the resources available to us.

Q2. What is the duration for the completion of the lab?

As half of the work is already done will require another 2-3 months to fully implement all the features.

Q3. What will be the benefits of this lab?

Can be used by students for research, project work; faculties can use the web server as a course web site for important announcements and relevant course materials.

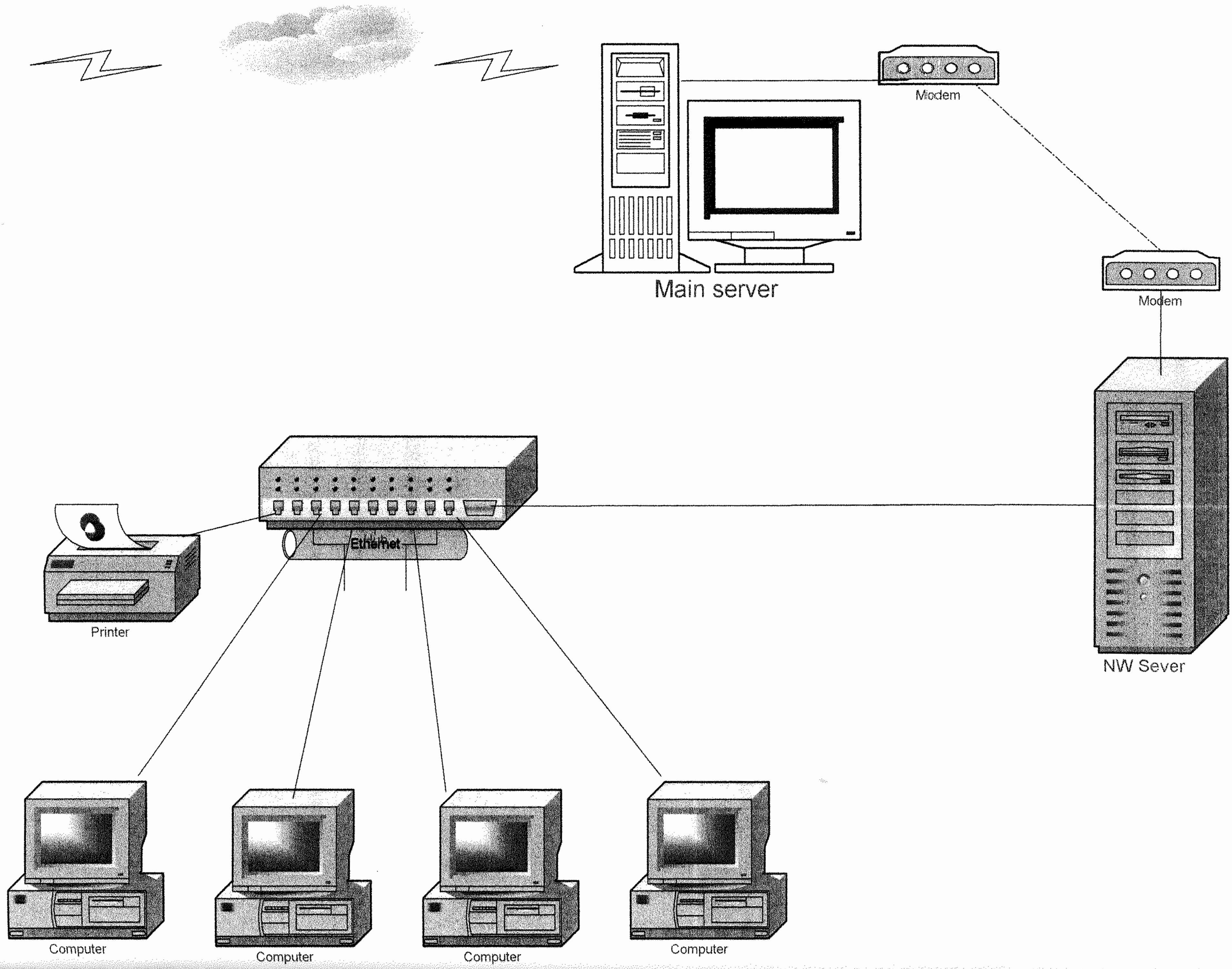
Q4.Will off campus students access the network?

Yes.

Operation Of The Lab

There will be a main a server with an Ethernet based network. This server consists of a database, web server. There needs to be a modem that can contact our main server so that all the users can use the Internet through a proxy server. A firewall can also be installed so that certain IP address range can be blocked.

Since the local server validates all the local computers, there are fewer security risks for the main server.


Main server
Modem
Modem
NW Sever
Ethernet
Printer
Computer
Computer
Computer
Computer

References

There were numerous references involved in this project. The work breakdown structure and WBS concepts were a combination of the class materials from the INFS755 class and with the help of the Internet links below:

- http://appl.nasa.gov/tools/tools_wbs.htm
- http://criticaltools.com/pertmain.htm

The ideas and concepts of subnetting were used from 3com Corporation based on an article written and compiled by Chuck Semeria.

http://cache.3com.com/www.3com.com/other/pdfs/solutions/en_US/50130201a.pdf

Textbooks

- Comer, Douglas E. Internetworking with TCP/IP Volume 1 Principles, Protocols, and Architecture Second Edition, Prentice Hall, Inc. Englewood Cliffs, New Jersey, 1991
- Huitema, Christian. Routing in the Internet, Prentice Hall, Inc. Englewood Cliffs, New Jersey, 1995

The concepts for the web server analysis part were drawn from

- Special Edition of Windows NT 4.0 web site: http://sunsite.net.edu.cn/tutorials/winnt4/
- Best OS Professor Reference MS Back Office Administrator's Guide

- National Software Testing Laboratory (NSTL)

  www.graphcomp.com/info/specs/isapi

The database part references were based on the technical resources from ORACLE corporation and other online learning sites.

- http://www.orafaq.com/faqdba.htm
- http://pcsupport.about.com/cs/backuprecovery/
- www.orafaq.org/faqdbabr.htm

Course Work

- INFS 751 INFS 754 networking courses for the design and implementation of the network lab.
- INFS 755 and INFS 780 for the project management concepts.
- INFS 684 and INFS 634 for the system analysis and the database design.
- INFS 730 and INFS 740 for the programming in Java Scripts.

There were a number of professional certified course materials on the web that I studied for the practical implementation of the project for networking.

- http://www.mcmcse.com/
- http://www.mcsedirectory.com/win2k.shtml
- http://www.arlanatech.com/70210.html
- http://www.mcseguide.com/networkplus.htm

There were also some references on the ORACLE from these certification guide sites.

- http://www.oracle.com/oramag/oracle/01-jan/o11educ.html

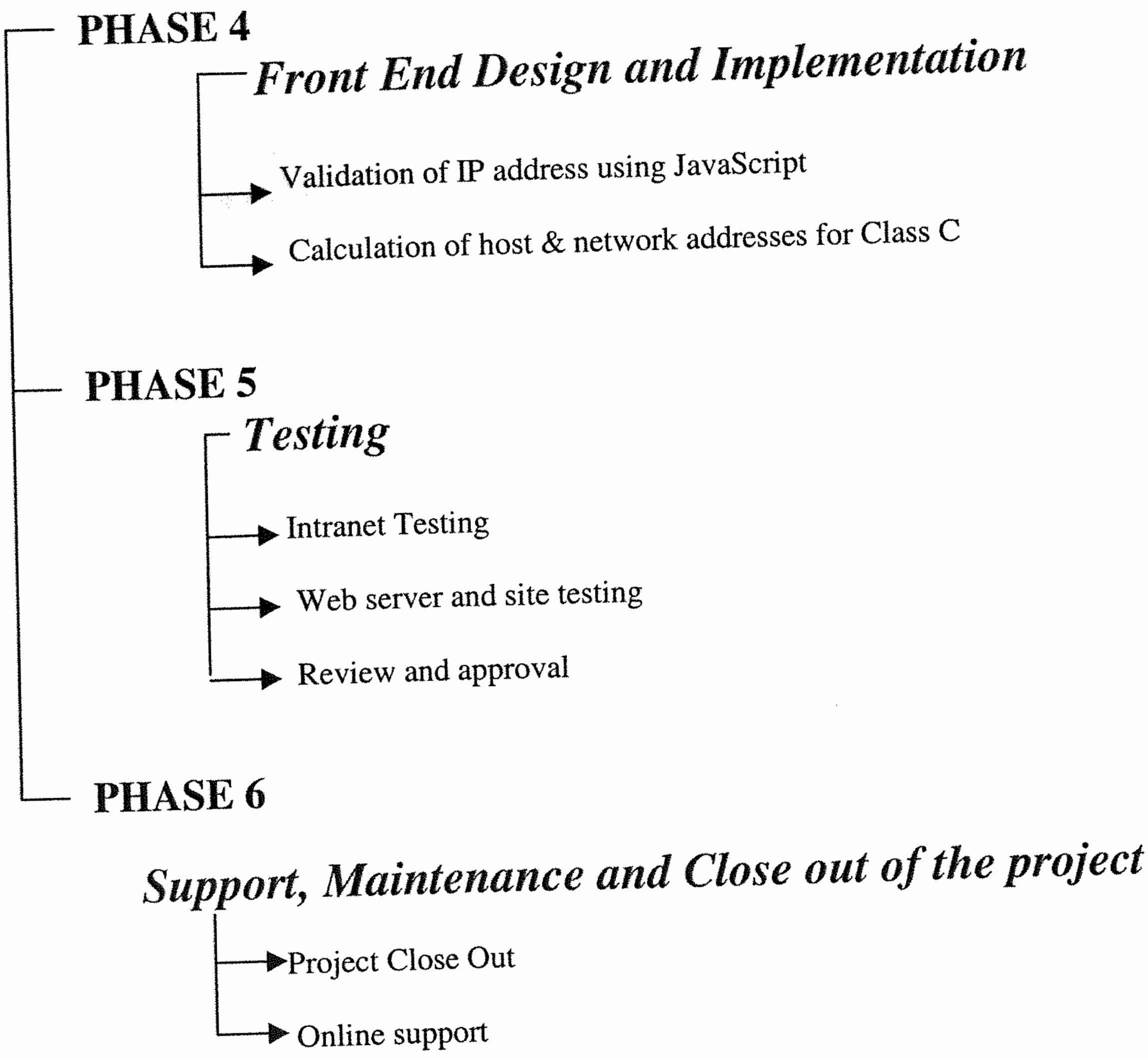
- http://certcities.com/editorial/features/story.asp?EditorialsID=7

APPENDIX A

Work Breakdown Structure

- **PHASE 1**
  - ***Planning and Project Inception***
    - Define the Project objective statement and get the approval of the MSIS Committee.
    - Estimate Resources available on hand and set up the resources.
    - Review and get the approval.
- **PHASE 2**
  - ***Network Design and Implementation***
    - Analysis of subnetting
    - Intranet design
      - Setting up of domain name
      - Security mode design
    - Web Server Implementation
      - Hosting of multiple FTP, WWW sites
      - Performance & Tuning analysis
      - Mapping of log files to SQL server
- **PHASE 3**
  - ***Database design and Implementation***
    - Creating of database without using the Database assistant.
    - Configuration of database
    - Design of database and recovery strategies
    - Performance tuning of Database with emphasis on OLAP/OLTP

APPENDIX A



PHASE 4

Front End Design and Implementation

- Validation of IP address using JavaScript
- Calculation of host & network addresses for Class C

PHASE 5

Testing

- Intranet Testing
- Web server and site testing
- Review and approval

PHASE 6

Support, Maintenance and Close out of the project

- Project Close Out
- Online support

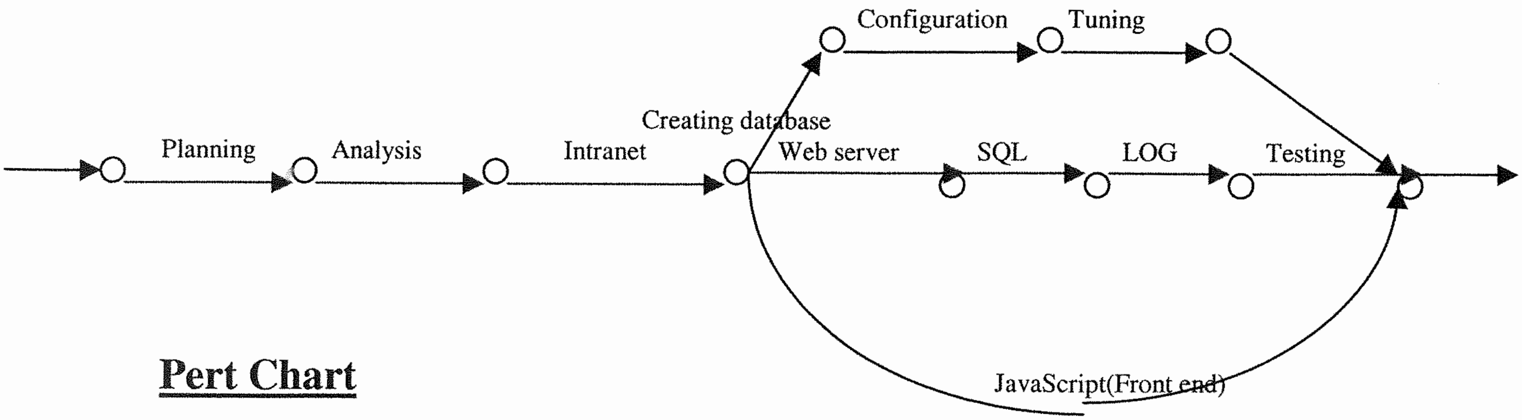
APPENDIX B

Pert Chart

Project Start Date: Tue 12/5/00
Project Finish Date: Tue 4/17/01

Tasks

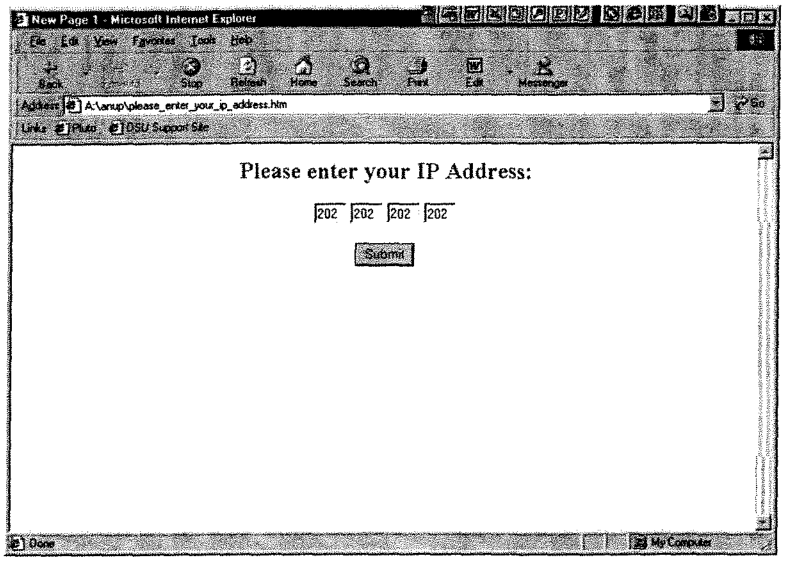
Resources and Assignments	Start	Finish	Work
Planning	Tue 12/5/00	Mon 1/15/01	30 days
Analysis	Tue 1/16/01	Wed 1/24/01	7 days
Intranet	Thu 1/25/01	Wed 2/14/01	15 days
Creating database	Thu 2/15/01	Mon 2/19/01	3 days
Web server setup	Thu 2/15/01	Wed 3/7/01	15 days
JavaScript program	Thu 2/15/01	Tue 2/20/01	4 days
Configuration of database	Tue 2/20/01	Mon 2/26/01	5 days
Performance tuning of database	Tue 2/27/01	Mon 3/5/01	5 days
SQL server configuration	Thu 3/8/01	Fri 3/16/01	7 days
Log Files	Mon 3/19/01	Tue 3/27/01	7 days
Testing and Implementation	Wed 3/28/01	Tue 4/17/01	15 days



Pert Chart

Critical path ________

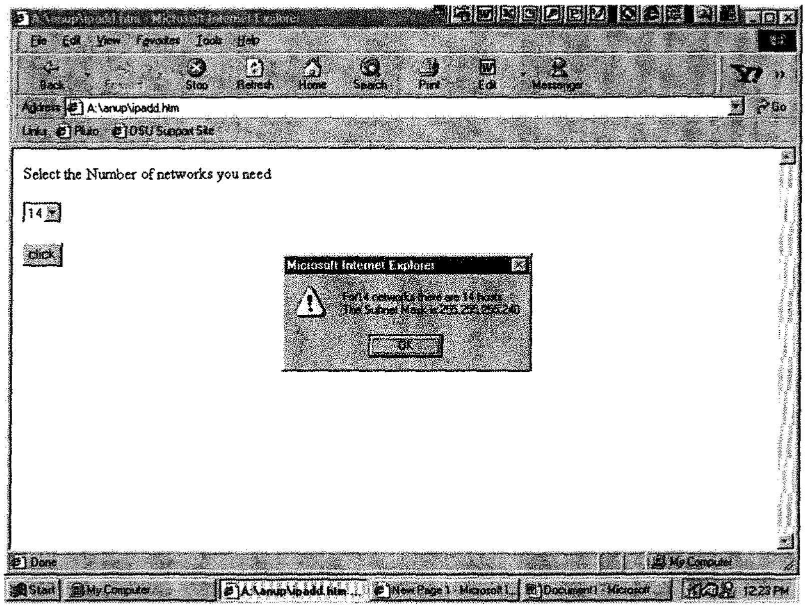
APPENDIX C


New Page 1 - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Forward Stop Refresh Home Search Print Edit Messenger
Address A:\anup\please_enter_your_ip_address.htm
Go
Links Pluto DSU Support Site
Submit
Done
My Computer

Please enter your IP Address:

202 202 202 202

Fig: Program that lets user specify a class C address


A:\anup\ipadd.htm - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Forward Stop Refresh Home Search Print Edit Messenger
Address A:\anup\ipadd.htm
Go
Links Pluto DSU Support Site
click
Microsoft Internet Explorer
OK
Done
My Computer
Start
My Computer
A:\anup\ipadd.htm ...
New Page 1 - Microsoft I...
Document1 - Microsoft
12:23 PM

Select the Number of networks you need

14

For14 networks there are 14 hosts
The Subnet Mask is 255.255.255.240

Fig: Program that calculates the network and host address