DISTRIBUTED DIRECTORY SYSTEM FOR LIBRARY NETWORK

Dr. S.Upendra R   &   E.Rama reddy

INTRODUCTION :

With the advances in computer network technology, efficient and economical transfer of data among computers becomes feasible. The advantages of linking all the remote library databases together include not only sharing of data bases, but also provide distribution of large quantity of information. Such information handling capability over a geographically separated library not only provides economical service but also provide instant information handle that was previously unachievable.

For example,a reader can instantaneously find out the current periodical available at geographically separate area. In organizing a distributed database, we first allocate the individual data files to the sites in the network. This allocation is based on :

- file usage rates
- response time requirements
- communication and storage costs and
- reliability requirements

In organizing and planning a distributed database system, there arise many problems in need of solutions such as file allocation policy, directory design and distribution, integrity and consistency. After the data has been created and allocated in a network, a directory is needed to specify the physical, logical, operational and security characteristics the data bank so that the user/system can locate and access a shared file in the distributed data base.

DIRECTORY SYSTEM :

A directory is a listing of information about data to users of the distributed database in a network. Such a directory is similar to a card catalog in a library. There are several ways to design the directory system.

- Central directory system
- Multiple master directory system
- Local directory system and
- Distributed directory system

Based on the computer network topology communication cost, storage cost, the physical location of the data has to be decided.

Central Directory System :

In a centralized directory system, a master directory is located at one of the computers. Should a user require a data that is not listed by his local directory, he consults the master directory to find out the location and contents of the requested file.

Further, the centralized directory must be updated when a new version of a data or a change in storage location is required.
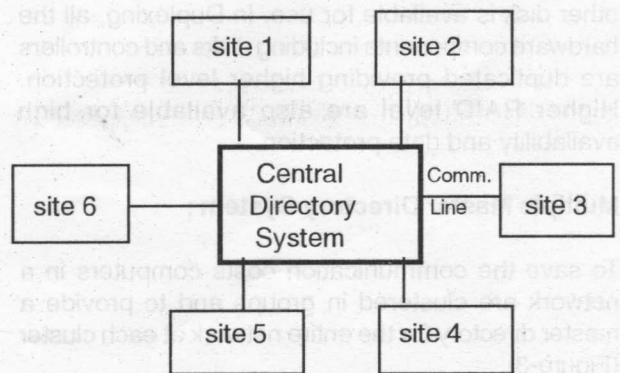

Fig. 1 :

Consider a schematic library network shown in Figure - 1.

It has a number of sites and maintains a central acquisition data loging and distribution center. A central catalog contains the title and a detailed

description of each item. However, each site maintains a local catalog and has access to the central catalog as well as catalog at other sites. In a manual system, the index card for items are duplicated at the central site and sent to each site where they are stored in their local catalogs. Access to the central catalog or the catalog of another site can only be had by calling on these locations. An alternate solution to this problem would be to include in each index card a list of all the branches at which a copy is maintained, and have a copy of the entire catalog stored at all branches. The advantages of centralized directory system is its ease in updating and disadvantage is the high communication cost.

When directory update rate increases, the centralized directory system yields lower operating cost than the distributed directory system. Even with the availability of high speed communication channels the centralized system may pose lower disk access times and reliability issues. In this case RAID technology may be used. The RAID (Redundant Array of Inexpensive Disks) not only provides fast update rate but also high availability rate with striping, mirroring and duplexing (Figure-2).

Striping spreads data across multiple disk drives. It is intended to provide high input/out rate without providing data protection. Striping increases performance by allowing commands directed to the hard disk drives to be overlapped. By striping the data across several disks, the workload is automatically balanced among disks resulting in multiple concurrent accesses. In Mirroring , two drives have the same data and are connected to a single controller. So, if one drive fails, the other from other disk is available for use. In Duplexing, all the hardware components including disks and controllers are duplicated providing higher level protection. Higher RAID level are also available for high availability and data protection.

Multiple Master Directory System :

To save the communication costs computers in a network are clustered in groups and to provide a master directory for the entire network at each cluster (Figure-3).

Local Directory System :

When directory updates are limited to local area, local directory system may be maintained. Local directories contain the directory information useful within the organization. These are rarely accessed by the other sites. The organization may maintain two different directories, one is local and the other is distributed for performance considerations.

Distributed Directory System :

In a computerized distributed database system, the catalog is fragmented and maintained in a database at each site. The data is distributed at different sites and the distribution is based on the access patterns and costs (Figure-4). When a communication cost is higher than the storage cost, the distributed directory system yields lower operating cost than the centralized directory system.

Distributed system is suitable for wide-area network of a corporate sector. Libraries are autonomous bodies for whom there is no central authority. Each library maintain its own data base and directory systems. Therefore, Distributed directory system is more suitable for wide-area Library network.

Other important considerations in the Distributed Data base design are :

- Site autonomy
- Load balancing among different systems
- Processing cost and communication cost
- Concurrency control.

PUBLICLY ACCESSIBLE LIBRARIES ON INTERNET :

Internet is the wide area distributed data bases across the globe. It allows searching of libraries, research papers and other background material. There is no central authority on the Internet. It consists of number of database servers. There databases can be accessed in following ways :

- sending and receiving mail
- Anonymous FTP
- Telnet

Over the last several years, most university libraries have switched from a manual catalog system to computerized library catalogs. The automated systems provide users with easily accessible and up-to-date information about books available in these libraries. Internet Resource Guide (IRG) describes a few libraries that are accessible.

For announcements of new libraries being available can be known by consulting Usenet newsgroup comp.internet.library. The NSF Network Service Center (NNSC)compiles and makes available the IRG. Resources listed are grouped by types into sections. Current sections include descriptions of online library catalogs, data archives, on line white page directory services, network information centers and computational resources. The list is distributed electronically by the NNSC. To receive a guide, or

to get on a mailing list, send a message to resource-guide-request@nnsc.nst.net. The current edition of the IRG is available via anonymous FTP from nnsc.nst.net in the directory '/resource-guide'.

For the convenience of the users two best-known and most comprehensive directories of Internet accessible libraries are listed below:

1. Internet - accessible Library Catalogs and Databases prepared and maintained by Art St. Geoge and Ron Larson. This directory is available by sending an E-Mail message to

LISTSERVE @ UNMVM.BITNET

with the following text in the body of the message:

GET LIBRARY PACKAGE

2. UNT's Accessing On-line Bibliographic databases is produced by Billy Barron, University of North Texas and is known as "Barron List".

CONCLUSION :

Library functioning is an autonomous nature within each organization. Generally the growth of a Library is in line with the organizational requirements where

it is located. There is no central authority on different Libraries, therefore, Distributed Directory System is recommended for wide-area Library network. If the communication cost is more, depending on the access rate, the directory may be down loaded into local computer. The frequency and time of down loading is depending on the access statistics.

The goal of distributed directory system is to get, for the least cost, the performance that the user requires and is clearly environment dependent.

REFERENCES :

1] Bipin C. desai
An introduction to Database Systems, Galgotia Pub.,1991.

2] P.Kehoe
Zen and the art of the Internet, 1992.

3] W.W.Chu
Distributed data base systems, Handbook of Software Engg, Ed. by C.R.Vivk and C.V.Ramamoorty, CBS Pub., 1986.

4] E. Lane and C. Summerhill
Internet Primer, London, Meckeler, 1993.