Capítulo 1

Marco Teórico

1.1 Redes

Una red de computadoras es un conjunto de computadoras interconectadas las cuales comparten información. [Blanco, 1993]. Las redes consisten de conexiones de hardware e interfaces de software, las cuales permiten transferir información a través de un medio de transmisión.

Algunos tipos de conexión que podemos encontrar son:

Punto a Punto.- El puerto de comunicación del equipo se conecta a un solo puerto del otro equipo.

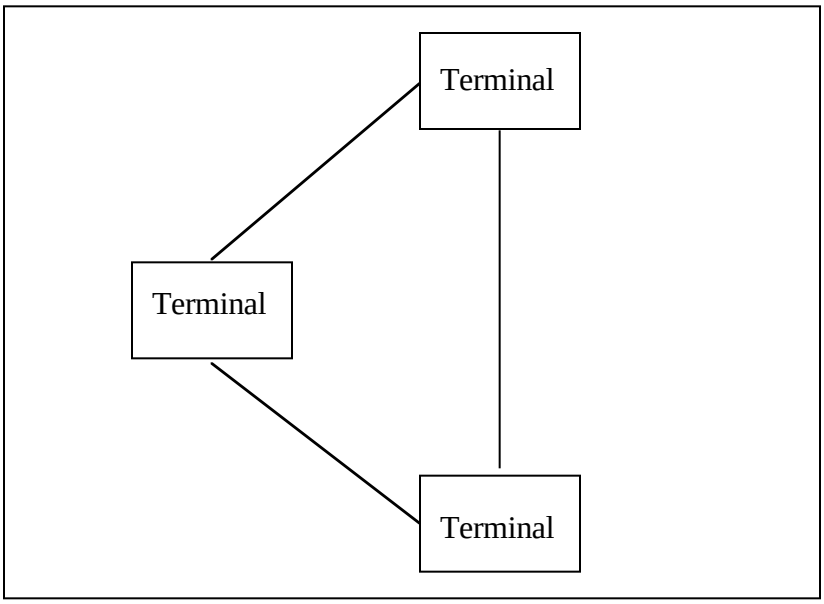

Fig. 1.1 Red Punto a Punto

Nube.- Cada equipo está conectado a una red

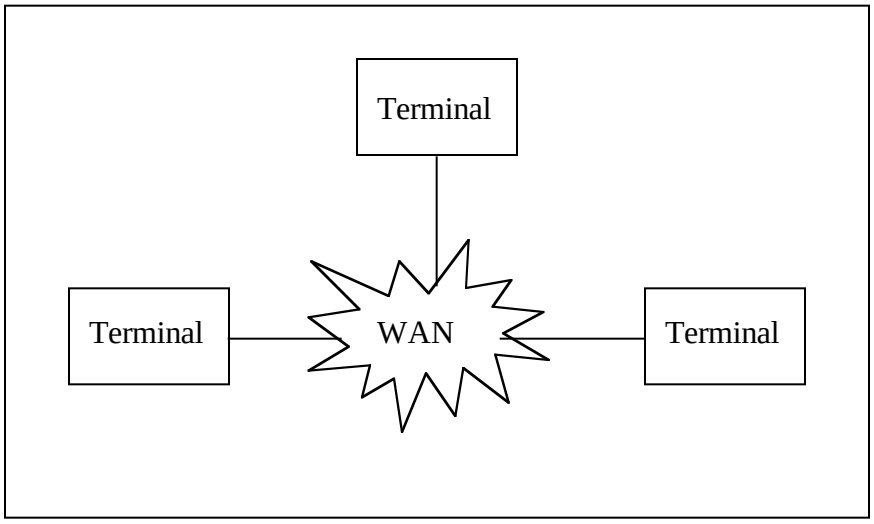

Fig. 1.2 Red tipo Nube

Multipunto.- Diversos equipos pueden estar conectados al mismo medio

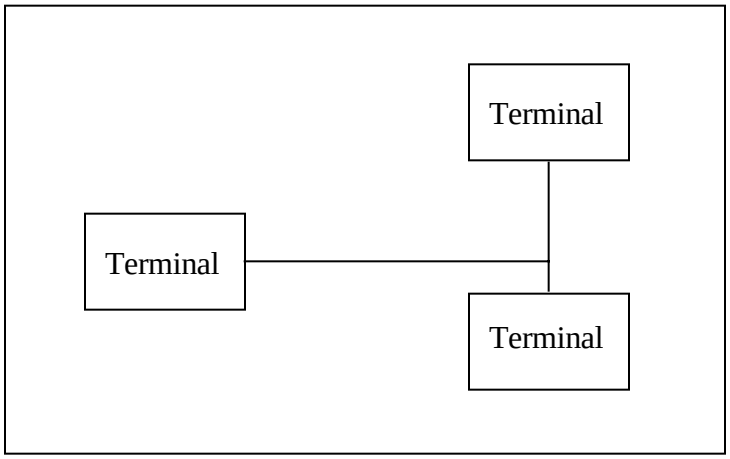

Fig 1.3 Red Multipunto

Daisy Chain.- Los equipos conectados uno después de otro

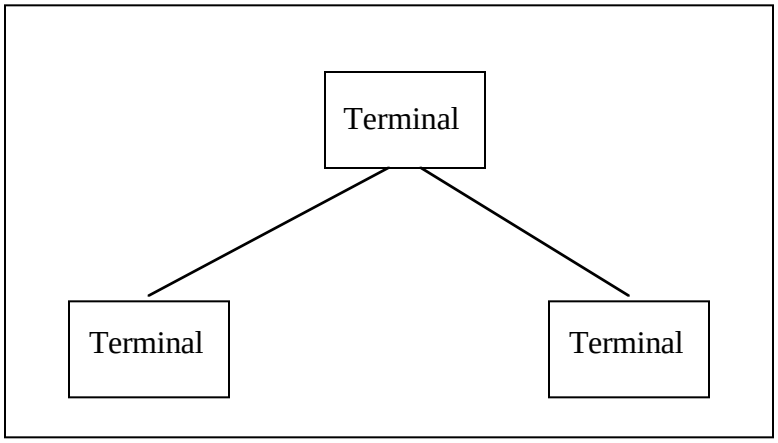

Fig. 1.4 Red Daisy Chain

Tal y como se explica en la tesis de Modelo Cliente-Servidor para la Administración de una Red bajo Unix [Blanco, 1993], en general, podemos clasificar a las redes de acuerdo a la distancia física que ocupan:

- WAN (Wide Area Network).- Es aquella que se expande a través de una zona geográfica de un país o continente [Gibbs, 1995]. Los medios de transmisión que se usan para cubrir estas distancias pueden ser microondas, cables de cobre, fibra óptica, satélite, etc.

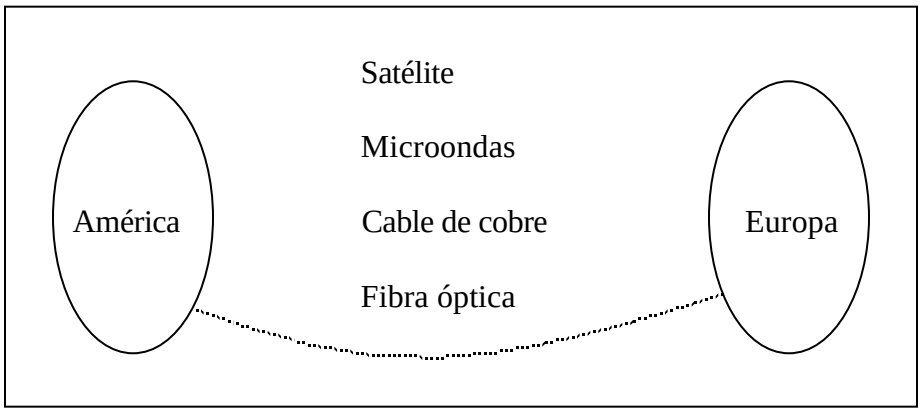

Fig. 1.5 Red WAN

- MAN (Metropolitan Area Network).- Es una versión más grande que la LAN en cuanto a topología, protocolos y medios de transmisión [Gibbs, 1995]. Esta red puede abarcara a un conjunto de empresas en una ciudad. Por ejemplo la conexión que ofrece Telmex

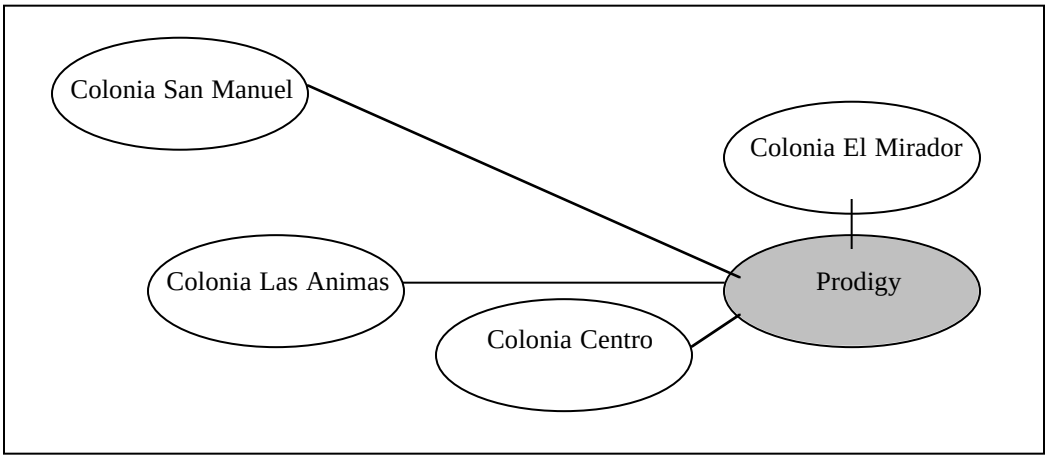

Fig. 1.6 Red MAN

- LAN (Local Area Network).- Es un sistema de comunicación de datos que permite conectar un conjunto de recursos y compartirlos. [Gibbs,1995] Su extensión va de unos cuantos metros a unos 5 kilómetros. El canal de comunicación es de capacidad media o alta, por ejemplo par trenzado, cable coaxial, fibra óptica, portadoras de rayo infrarojo o láser, microondas, etc. Las velocidades en estas redes van de 10 Megabits por segundo hasta 622 Mbps. Por ejemplo la conexión entre el edifico 2 y el edifico 4 de la UDLA.

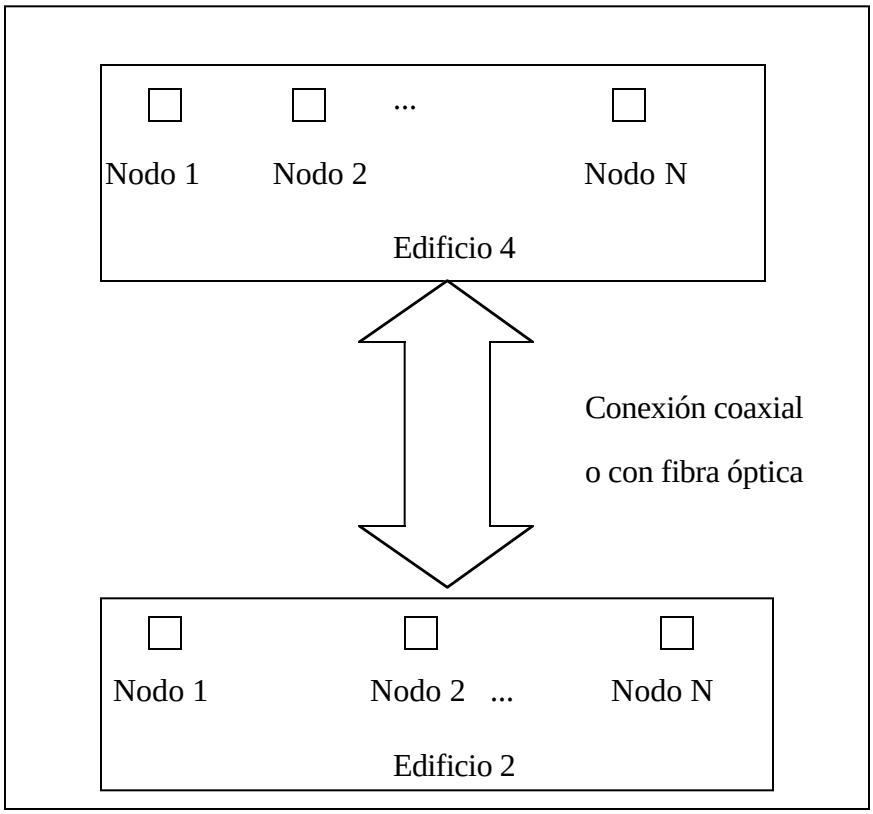

Fig. 1.7 Red LAN

Las estructura topológica de la red es un parámetro primario que condiciona fuertemente las prestaciones que se pueden obtener de la red. El acierto en la elección de una u otra estructura dependerá de su adaptación al tipo de tráfico que debe cruzar y de una valoración de la importancia relativa de las prestaciones que de la red se pretenden obtener. Las topologías generales son:

Estrella.- Todas las estaciones están unidas mediante medio bidireccionales a un módulo o nodo central que efectúa funciones de comunicación. El nodo central, asume además las labores de control y dispone de gran parte de los recursos informáticos comunes.

Esta topología requiere muchísimo cableado. Si la máquina central falla, falla toda la red. Una ventaja de esta configuración es que cada conexión no tiene que soportar múltiples máquinas en competencia por acceso, de manera que es posible lograr altas frecuencias de transferencia de datos.

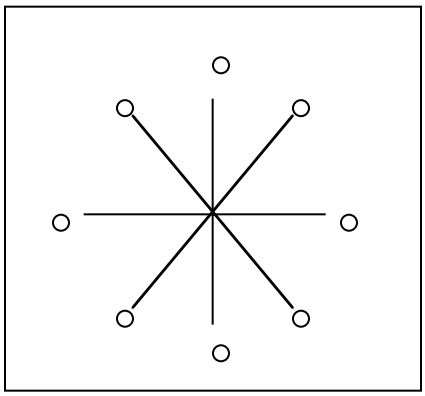

Fig. 1.8 Topología de Red tipo estrella

Árbol.- Es una extensión de la arquitectura de estrella. Por medio de la interconexión de varias se permite establecer una jerarquía clasificando las estaciones en grupos y niveles según el nodo al que están conectadas y su distancia jerárquica al nodo central.

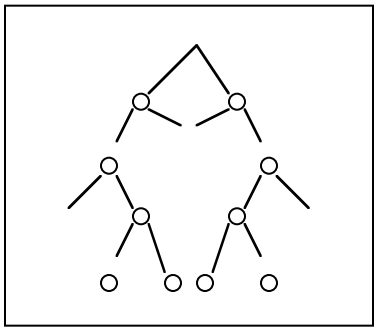

Fig. 1.9 Topología de red tipo árbol

Bus.- Los modos de comunicaciones están conectados de un único medio de comunicación que recorre todas las estaciones. La topología es de difusión y todas las estaciones reciben simultáneamente la información. La ventaja de esto es que si un punto falla, no basta para desactivar todo el sistema.

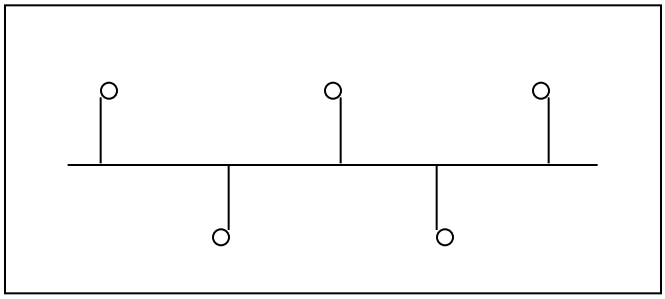

Fig. 1.10 Topología de red tipo bus

Anillo.- Las estaciones están conectadas en círculo y cada una de ellas necesita dos canales de comunicación, por ejemplo dos cables coaxiales, uno para recibir la información, la procesa, si le corresponde la toma y si no la pasa por el otro canal de comunicación a la siguiente estación. La desventaja es que con una estación que falle, puede fallar toda la red.

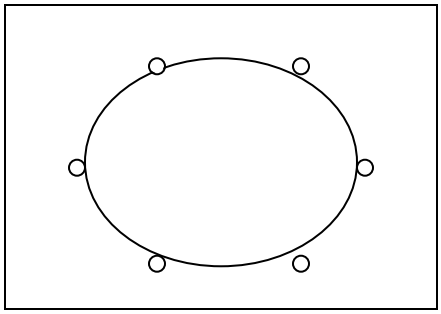

Fig. 1.11 Topología de red tipo anillo

El procesamiento de datos en la red juega un papel muy importante debido a que el desarrollo de las aplicaciones principales continúa la operación del sistema. La comunicación de datos ha sido responsabilidad de las operaciones del procesamiento de datos

Existe una gran variedad de hardware y software, el problema es que en algunas ocasiones el control llega a ser bastante complejo. Eso se refiere especialmente a de quién y de dónde vienen las decisiones del procesamiento de un sistema, qué equipo usar o dónde instalarlo, entre otros.

Tenemos dos tópicos:

- La centralización donde lo ideal es la coordinación para las necesidades del grupo
- La descentralización donde es necesario mantener la compatibilidad entre todas las terminales. La descentralización viene a ser una buena opción considerando que es un poco mas barata.

Algunas consideraciones que no debemos olvidar son:

- La capacidad de los procesos donde que deben ser interactivos
- Costo de procesamiento de patrones
- Demandas de la base de datos
- Accesos locales o remotos

1.2 Sistema Operativo Unix

El sistema operativo UNIX ha evolucionado durante los últimos veinte años, desde su invención como experimento informático hasta llegar a convertirse en uno de los entornos informáticos más populares e influyentes del mundo. [Todino, 1992]

Dentro de las características de este sistema operativo encontramos: reutilización de código, portabilidad, flexibilidad, potencia y elegancia. Además, está diseñado para manejar las necesidades múltiples y simultáneas de un usuario (multitarea), y al mismo tiempo soporta las actividades de más de una persona a la vez (multiusuario). Lo anterior, aunado a la enorme base de comunicaciones que posee, permite que la comunicación por red sea fácil de implantar [Coffin,1992].

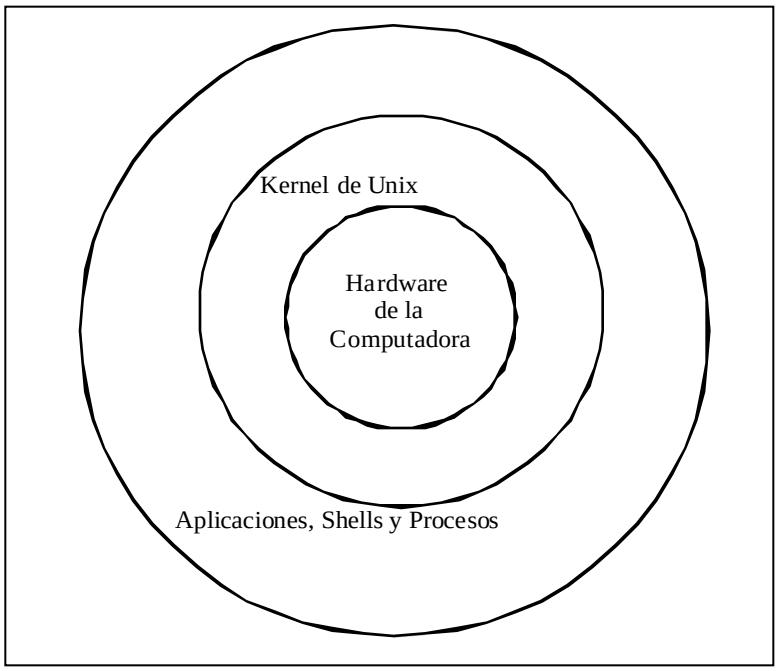

Fig. 1.12. Estructura del Sistema Operativo UNIX

En el sistema Unix los archivos se colocan en directorios. Como el sistema trata a cada archivo como una secuencia de bytes, se puede almacenar cualquier tipo de archivo en un directorio, de hecho, se puede colocar un directorio en otro directorio, esto ayuda a explicar el sistema de archivos globalmente, puesto que realmente el sistema es una jerarquía de directorios y subdirectorios.

SunOS consiste de un Kernel y cientos de otros archivos que contienen información y programas. El Kernel controla todas las funciones básicas de la computadora, incluyendo acceso a disco, a impresoras, manejo de archivos y servidores de red. Todas las WS Sun reciben al Kernel cuando se arrancan o se inicializan. El Kernel es llamado vmunix, éste, maneja todos los recursos físicos de una WSSun. El sistema operativo SunOS y Unix en general manejan todos los periféricos como archivos. [Winsor, 1993]

Las funciones más importantes del Kernel.:

- Permite crear procesos, leerlos, escribirlos y accesar archivos.
- Manejar guardianes. Son procesos que empiezan mientras se inicializa la máquina y deben correr constantemente hasta que se les llama. Los guardianes hacen la mayoría de las funciones del sistema, entre ellas encontramos: subir un archivo y rutear el correo electrónico.
- Mueve información del disco a la memoria principal (memoria virtual)
- Mantiene información sobre los procesos activos y decide que proceso continua.
- Maneja periféricos.
- Maneja software de red (rpcgen) que implanta servicios de red (RPC).

- Maneja facilidades para la comunicación entre procesos como sockets, los cuales están involucrados en actividades de red.
- Proveen facilidades para crear, examinar y modificar procesos [Leach, 1994]

1.3 Administración del Sistema Unix.

La administración del sistema es el mantenimiento rutinario necesario para que un sistema crezca y se modifique. El ingreso de identificaciones de presentación de usuario, la instalación de nuevo software, la eliminación de ficheros de registro, el formateo de discos flexibles y la realización de copias de seguridad son todas tareas que caen en la categoría de administración del sistema. Un usuario de una máquina Unix esta generalmente designado como administrador del sistema, y es tarea del administrador mantener el sistema activo y en ejecución de otros usuarios. Esta función es vital, incluso en una máquina de un solo usuario.

Aunque las tareas de administración son significativamente más complejas sobre un sistema multiusuario que sobre una máquina MS-DOS, las versiones más recientes del sistema Unix han introducido herramientas que pueden facilitar estas tareas. La mayoría de las implantaciones disponen de un agente de usuario para administración. Este agente es generalmente una herramienta orientada a menús la cual ayuda al administrador a realizar las funciones, y gran parte del trabajo complicado es efectuado por la herramienta sin requerir una atención muy detallada por parte del usuario [Coffin, 1992].

La administración de la red no es una tarea fácil, el sistema operativo Unix ha introducido herramientas que la hacen más sencilla. Dentro de las tareas que el administrador de la red debe cumplir encontramos las siguientes:

- Dar de alta nuevos usuarios.
- Instalar software (SW) y actualizarlo.
- Instalar hardware (HW) y mantenerlo.
- Mantener la seguridad e integridad del sistema y de la red.
- Diagnosticar y arreglar los problemas de SW y HW a medida que ocurran.
- Respaldar los sistemas de archivos. [Winsor, 1993]

Cabe mencionar que cada una de ellas contiene inscritas tareas menores, la administración de una red varía debido a que cada administrador engloba sus tareas o resultados de cierta manera.

Para realizar algunas de estas tareas es necesario conocer la ubicación (directorios) de los archivos, así como comandos y formatos específicos. Es necesario subir archivos de otras máquinas (montaje), enviar archivos a máquinas o modificar archivos. En un ambiente tradicional de Unix, el montaje es el proceso de tener acceso a los archivos del sistema.

Cuando la máquina arranca, automáticamente corre el programa SunOS mount y hace el montaje de los archivos que necesita para operar. También se puede correr el programa mount para tener acceso a los archivos adicionales como aplicaciones de software y

programas de usuarios. Cuando un máquina sube un directorio, el programa mount añade el directorio a un punto de montaje. Los puntos de montaje son localidades dentro de los archivos del sistema de la máquina local.

Todas las computadoras Sun tienen integrado, pero no configurado NFS por default. Cuando un archivo servidor de NFS anuncia que sus archivos están disponibles, se dice que los exporta. La máquina que quiere importarlos, hace un mount NFS para tener acceso a los archivos.

Como administrador del sistema es necesario modificar y consultar constantemente la base de datos del sistema administrativo y otros archivos locales para mantener a la máquina y/o a la red corriendo sin problemas.

Para que un cliente reciba archivos de un servidor apropiado debe pasar lo siguiente: el administrador del servidor de archivos debe exportar (a través de NFS) el directorio que tiene disponible. Los usuarios o administradores de la red (clientes), deben crear puntos de montaje para esos directorios, después, añadir entradas a su archivo local /etc/fstab por cada directorio que quiera montar. Las máquinas que tienen conectada una impresora pueden ser servidores de impresoras, siempre y cuando ésta tenga disco. El sistema operativo puede manejar una variedad de impresoras y configuraciones

1.4 Perl (Practical Extraction and Report Language)

Perl ha llegado a ser la herramienta de UNIX para muchas tareas de pequeña y mediana complejidad, donde la manipulación de texto, archivos y procesos es muy fácil. Perl es capaz de correr en múltiples plataformas por lo que realmente es muy útil. [Wall, 1992]

1.4.1 Historia de Perl

Perl fue creado originalmente para facilitar la administración de una red. Su inventor es Larry Wall y la primera versión se basó en una utilería de Unix denominada Awk que permite el uso de expresiones regulares para manipular y procesar texto. Se trataba de que Perl fuera un lenguaje para navegación entre varios archivos, teniendo el poder de buscar entre varios archivos de texto de una manera eficaz a través de la invocación de datos dinámicos e imprimiendo reportes basados en la información extraída. [Wall, 1992]

Realmente lo anterior sí se logró, pues muy a menudo Perl es usado para programas desarrollados en C. Con lo anterior, se pueden manipular archivos por sí mismos o por la información que contengan, permitiendo la creación y destrucción de procesos a través de los sockets.

1.4.2 Ventajas de Perl

- Perl es sumamente compatible entre versiones (cosa que no sucede en Java)
- Perl es gratis y es fácil de obtener e instalar.
- Perl no tiene limitaciones como:
    - Las líneas y los arreglos pueden ser tan grandes como uno desee

- La recursión puede ser tan profunda como uno desee
- Los nombres de las variables no tienen restricción en cuanto a longitud
- Los datos binarios no causan problemas
- Las tablas de hash usadas para los arreglos asociativos se expanden cuanto sea necesario para evitar la degradación del desarrollo

- Perl puede determinar cuáles datos son derivados de fuentes inseguras para prevenir operaciones peligrosas antes de que sucedan
- Perl tiene la habilidad de procesar múltiples operaciones con un sólo comando
- Tiene una manera muy sencilla de hacer las cosas, incluso es más claro que C
- Perl es un intérprete y no un compilador,
- Existe Perl en varias plataformas como Sun OS, AIX, Windows NT (WinPerl) y Macintosh. Debido a que son scripts no hay problema.
- Perl es portable, sobretodo con CGI, lo único que hay que hacer es correr el script
- Es posible crear libererías de Perl donde se crearían complejos scripts
- Combina la versatilidad y eficiencia de un shell con la capacidad de hacer llamadas de bajo nivel, tal y como lo hace C
- Perl trabaja con diferentes herramientas de texto como WAIS, Glimpse, bases de datos relacionales como Sybase, Oracle, bases de datos de objetos relacionales como Illustra. [Wall, 1992 & Srinivasan, 1997]

1.4.3 Desventaja de Perl:

- Perl no genera código binario

1.4.4 Razones para escoger Perl:

- Popularidad.- Cada vez mas gente conoce Perl y sus librerías crecen día con día. Lo anterior porque todo mundo hace aportaciones a través de la red
- Existen en Internet rutinas gratis
- Hay numerosas áreas de soporte

1.5 Sed, Awk y CShell

Sed y Awk son dos utilerías de Unix que contienen una serie de cosas en común, pues ambas son expresiones regulares con patrones. Sed y Awk son herramientas para administradores, usuarios y programadores, prácticamente para cualquier persona que trabaja con archivos de texto. [Dougherty, 1991]

Cshell es un programa que se ejecuta en dos modos. [Arick, 1992]

Modo Interactivo.- Es la manera en cómo interactúa con la terminal y ejecuta comandos desde el teclado.

Modo no Interactivo.- Se ejecuta un conjunto de comandos que son leídos desde un archivo (script). Las líneas son leídas una por una.

1.5.1 Sed (Stream Editor)

Sed es perfecto para editar un número determinado de archivos. Pero a la vez tiene la característica de no ser interactivo. Sed interpreta un script que controla las acciones a desarrollar. Las entradas son procesadas en el programa y son proporcionadas a través del

teclado o de un archivo, mismas que son direccionadas a una salida que puede ser desplegada en una terminal o almacenada en un archivo.

Algunas características de Sed son:

- Automatiza la edición de uno o más archivos
- Simplifica la tarea de desarrollar las mismas ediciones en diferentes archivos
- Escribe la conversión de los programas

Por ejemplo una sustitución global donde se cambia la palabra red por redes

$ sed ´s/red/redes´ archivo.txt

Los comandos más generales son:

a	append
i	insert
c	change
s	sustitución
>	direcciona la salida a un archivo de texto

1.5.2 Awk (Aho, Weinberger y Kernighan)

Awk, llamado así por sus autores, permite la fácil manipulación de estructuras de datos y de generación de reportes. Desde la edición número 17 de Unix, Sed y Awk son parte de él. La diferencia es que Awk ha sufrido una serie de modificaciones desde 1985 llegando a surgir nawk (nuevo awk). Awk permite el uso de estructuras de texto para escribir los procedimientos, capacidad que es muy útil.

Algunas características de Awk son:

- Uso de variables para la manipulación de datos
- Uso de operadores aritméticos y de cadena
- Uso de estructuras de datos como ciclos y condicionales
- Habilidad de generar reportes formateados

Por ejemplo La impresión de todas las líneas donde aparezca la cadena “ejemplo” del archivo tesis.txt

$ awk ‘/ejemplo/’ tesis.txt

Los comandos más generales son:

f logra que se trabaje de la misma manera que sed

{print} imprime el primer campo

F Cambia el separador existente entre los campos

1.5.3 C Shell

C Shell tiene la capacidad de proveer un número de métodos que manejan la salida de los comandos, además de poseer la capacidad de crear comandos en un lenguaje propio, es decir, se puede reemplazar algún comando existente por cualquiera que se desee. Algo muy útil en esto es reducir los comandos que son demasiado largos.

Los scripts que son creados, pueden manipularse de tal manera que es posible ordenar la ejecución de los mismos. Otra opción es ejecutar concurrentemente los programas y determinar sus estatus.

Algunos comandos usados son:

csh	invoca el C Shell
alias	permite renombrar algún comando existente
unalias	permite quitar alguna asignación previamente hecha a algún comando
history	permite recordar los comandos que se han ejecutado

En esta tesis se usó C Shell como leguaje auxiliar, debido a que su código es 100 % portable entre las diferentes plataformas que tiene Unix como son: HP, SUN, IBM, Apple, Compaq y otros.

1.6 NIS & NFS

NIS & NFS son servicios permiten la construcción de sistemas de cómputo distribuidos. Ambos son consistentes y transparentes, debido a la manera en como los datos y los archivos son compartidos, además de que son protocolos de red de alto nivel. [Stern, 1992].

El principal objetivo de NFS es soportar un sistema heterogéneo donde los clientes y servidores podrían ejecutar distintos sistemas operativos en hardware diverso, pare ello, es esencial que la interfaz entre servidor y cliente este bien definida. Las claves utilizadas para la autentificación, así como otra información, están contenidas en NIS que almacena las parejas (clave, valor). Cuando se proporciona un clave, regresa el valor correspondiente. Ello ayuda al esquema de protección de Unix. Pero NIS, también almacena la asociación de los nombres de los usuarios con las contraseñas, así como la asociación de los nombres con las máquinas con las direcciones en la red y otros elementos.

1.6.1 NIS (Network Information System)

NIS provee un sistema de base de datos distribuida para la configuración de los archivos. Los servidores de NIS administran copias de los archivos de la base de datos y los clientes piden la información a los servidores en lugar de usar la propia o la local.

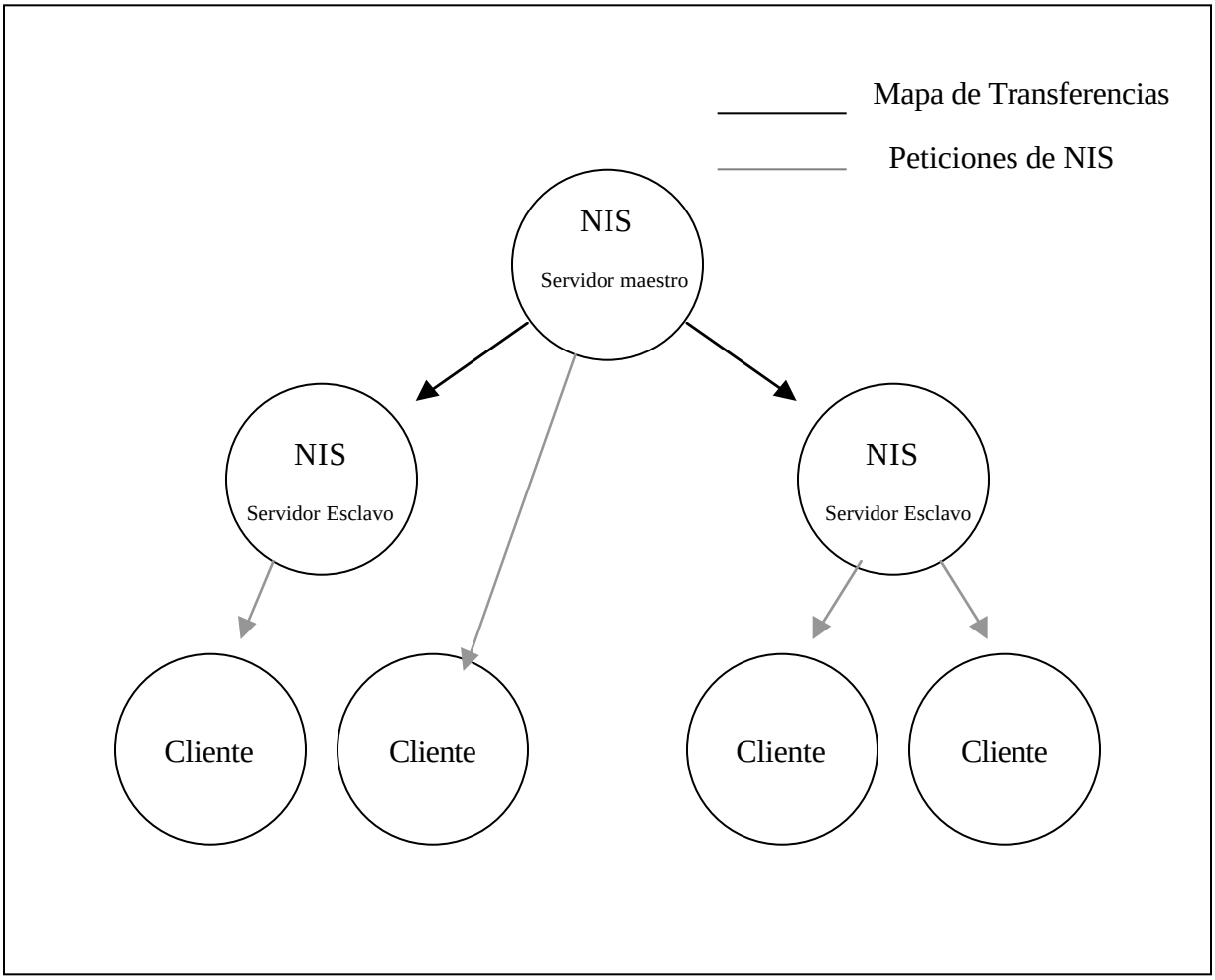

Fig. 1.13 Interacción de clientes y servidores de NIS

1.6.2 NFS (Network File Ssytem)

NFS tiene uno o más filesystems montados por clientes de NFS. Los discos remotos parecen discos locales y los filesystems son montados usando el comando mount de Unix. Todas las utilerías trabajan como si se estuviera en discos locales. NFS facilita la tarea del administrador porque elimina la necesidad de mantener múltiples copias de archivos en varias máquinas.

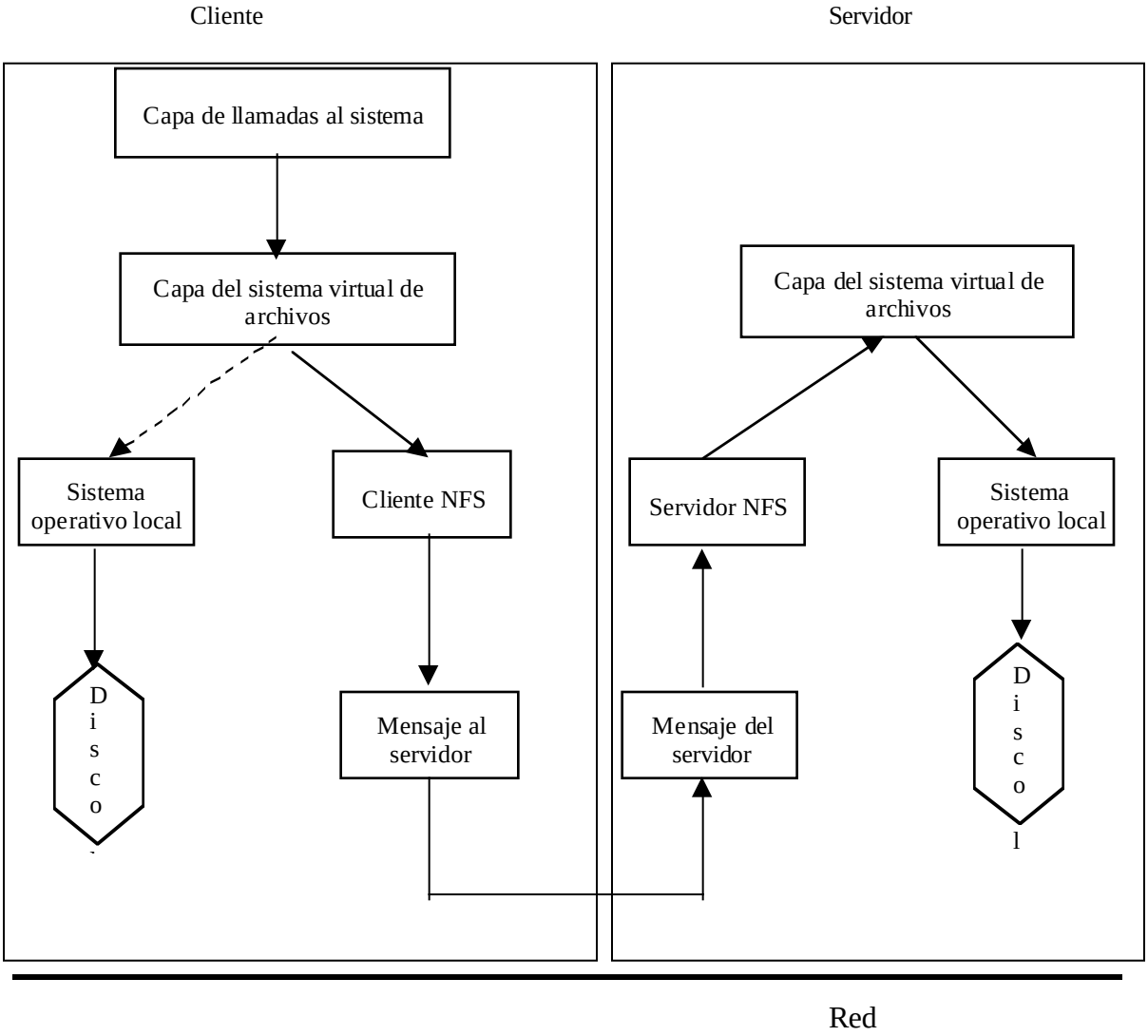

Fig. 1.14 Estructura de la capa NFS

1.7 RPC (Remote Procedure Call)

Este método permite que los programas llamen a procedimientos localizados en otras máquinas [Tanenbaum, 1993]. Cuando un proceso en la máquina A llama a un proceso en la máquina B, el proceso que realiza la llamada se suspende y la ejecución del procedimiento se realiza en B. La información se puede transportar de un lado a otro mediante los parámetros y puede egresar en el mensaje.

Lo descrito anteriormente parece sencillo; sin embargo, existen problemas como que las máquinas al ser diferentes, utilizan espacio de direcciones diferentes. La idea principal de RPC es que se parezca lo más posible a una llamada local, para lograr la transparencia. Las fallas más comunes que pueden ocurrir son:

- El cliente no puede localizar al servidor
- Se pierde el mensaje de solicitud del cliente al servidor
- Se pierde el mensaje de respuesta del servidor al cliente
- El servidor falla antes de recibir la solicitud
- El cliente falla después de enviar la solicitud

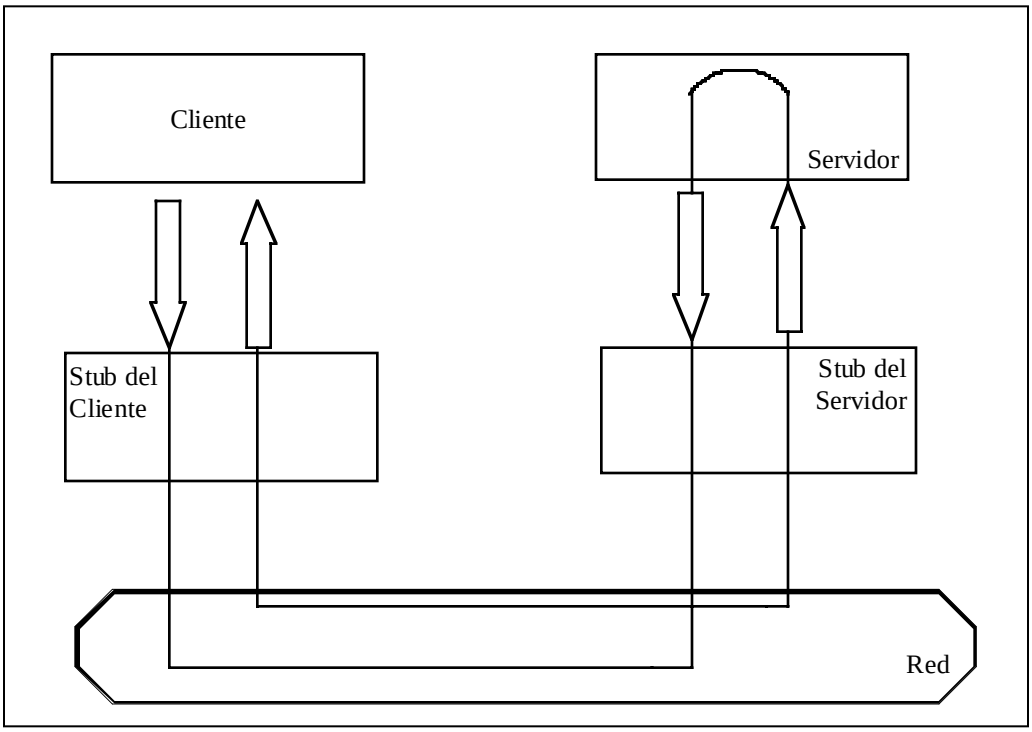

Fig. 1.15 Comunicación de RPC

1.8 RDIST (Remote File Distribution Client Program)

Es un programa que mantiene copias idénticas de los archivos sobre múltiples hosts. Este preserva los permisos de todos los archivos y de ser posible actualiza los programas que se están ejecutando. Rdist lee los comando desde distfile para actualizar los archivos y los directorios. En caso de que distfile tenga un "-", se usará el estándar de entrada. Algunas de las opciones más utilizadas son:

-A número	Indica el mínimo número de archivos libres en un filesystem que debe existir para que se ejecute el rdist
-a número	Indica el mínimo número de espacio libres en un filesystem que debe existir para que se ejecute el rdist

-F	No permite el fork
-n	Imprime los comandos sin ejecutarlos

1.9 Kerberos

El sistema de kerberos autentifica a los usuarios en la red. Kerberos es una de las muchas opciones que pueden usarse en la red, tal y como se hace con NFS [Fiedler, 1991]. Gracias a kerberos todo el tráfico de la red queda encriptado. En algunos ambientes es posible realizar procesos remotos, por lo que se puede usar rlogin, rcp y rsh.

Lo primero que se debe hace es registrarse en la base de kerberos con el comando kinit, mismo que pedirá un password para poder dar el acceso requerido. Una vez que el acceso fue autentificado, kerberos otorga un ticket que es un mensaje de protocolo encriptado que provee la autentificación. Los tickets privilegiados como los de root caducan en pocos minutos; sin embargo, los tickets de los usuarios comunes pueden estar varias horas o varios días, dependiendo de las políticas que se hayan usando en la instalación. Una vez que la sesión ha terminado, basta con ejecutar el comando kdestroy, mismo que eliminará el ticket.

1.10 Conceptos de Estadística

Las técnicas estadísticas se utilizan en casi todos los aspectos de la vida. Se diseñan encuestas para recabar información previa al día de las elecciones y así predecir el resultado de las mismas. Se seleccionan al azar consumidores para obtener información con el fin de predecir la preferencia con respecto de cierto producto. El ingeniero muestrea las

características de calidad de u producto, junto con otras variables controlables del proceso, para facilitar la información relacionada a la calidad, etc.

Los Ingenieros en Sistemas no podían se la excepción y es de mucha utilidad aplicar esta técnicas para lograr el total conocimiento de algunos campos que se estén investigando. En este tesis usó la Estadística para determinar la media y la desviación estándar del espacio que ocupan los estudiantes dependiendo de la carrera que estén cursando.

1.10.1 Media

La media es la tendencia central más común que su usa en la Estadística. La media de un conjunto de mediciones únicamente localiza el centro de la distribución de los datos; por sí misma, no ofrece una descripción adecuada de un conjunto de mediciones, pues las frecuencias podría ser muy diferentes. [Mendenhall, 1994]

La fórmula de la media es:

$$\mu = 1/n \Sigma y_i$$

donde:

μ = media

n = número de elementos de la muestra

y_i = cada uno de los elementos de la muestra

- **Varianza**

La medida más común de variabilidad usada en la Estadística es la varianza que es una función de las desviaciones (distancias) de las mediciones muestrales con respecto a la media [Mendenhall, 1994] . Su fórmula viene dada por la siguiente ecuación:

$$s'^2 = 1/n \, \Sigma \, (y_i - \mu)^2$$

donde:

s'^2 = varianza

n= número de elementos de la muestra

y_i = cada uno de los elementos de la muestra

μ = media de los elementos de la muestra

1.10.3 Desviación Estándar

La varianza es útil en la comparación relativa de dos conjuntos de mediciones, pero sólo aporta información al respecto de la variación de un solo conjunto cuando se interpreta en términos de la desviación estándar [Mendenhall, 1994] . Su fórmula es:

$$s' = \sqrt{s'^2}$$

donde:

s’ = desviación estándar

s'^2 = varianza

Estos conceptos se van a aplicar en la tesis en dos poblaciones:

a) En todos los estudiantes de esta Universidad divididos por las carreras que estudian

b) En todos los empleados de esta Universidad

1.11 Conclusiones

El sistema operativo Unix es muy poderoso y se requiere de su conocimiento a fondo para administrar de una manera eficaz una red y sus recursos. Para ello, existen una serie de herramientas y programas que interactúan de una manera muy natural, logrando que algunas tareas se simplifiquen. Perl es un lenguaje que tiene mucho auge y que nos ofrece una serie de ventajas que permiten el eficaz desarrollo de aplicaciones basadas en Unix.

Gracias al conocimiento de todos los conceptos descritos a lo largo de este capítulo, fue posible desarrollar la aplicación de una manera eficaz aprovechando al máximo los recursos que brindan. A continuación se muestra una gráfica de cómo interactúan las tareas de un administrador y de las herramientas que usa para llevarlas a cabo.

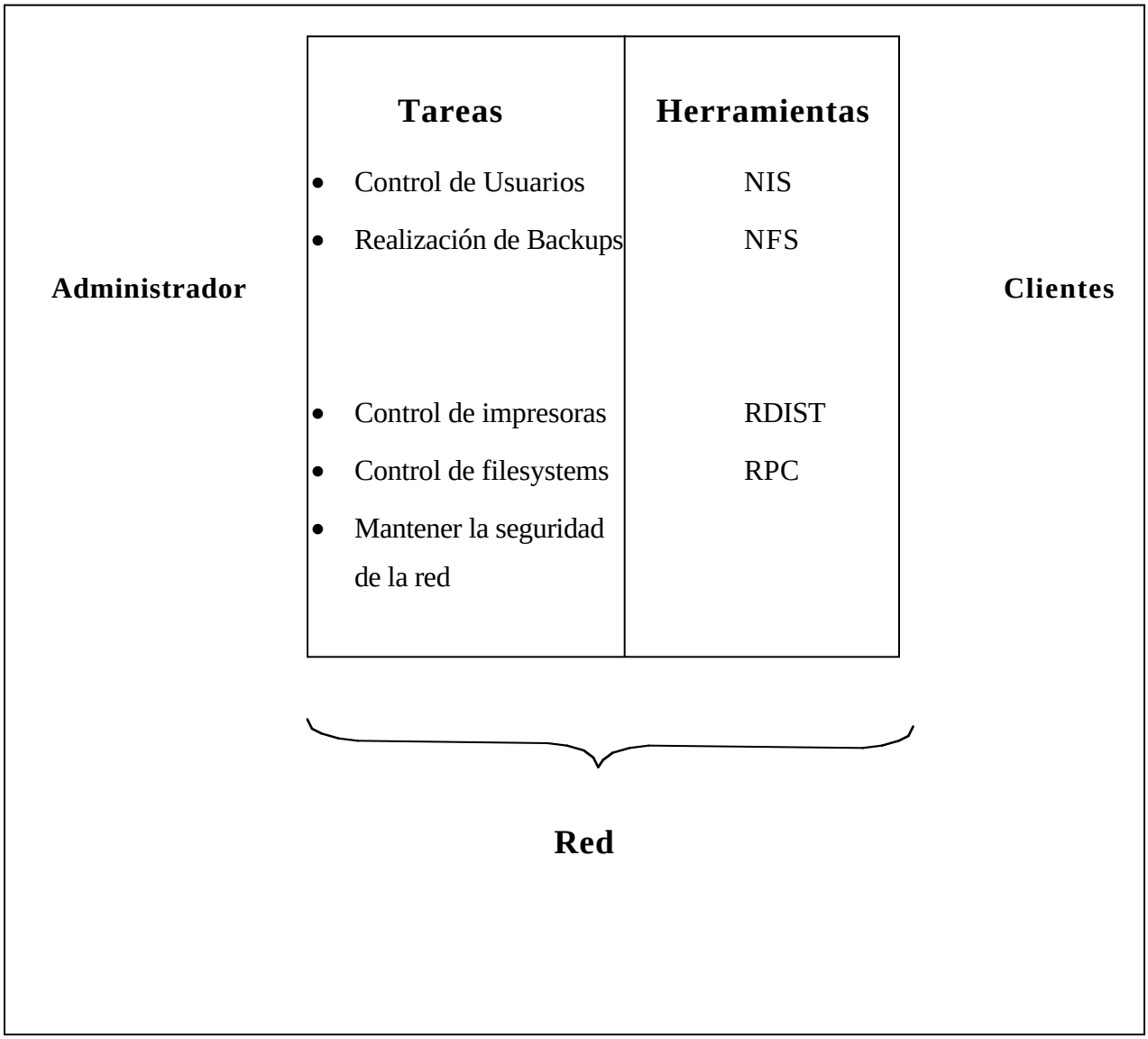

Fig. 1.16 Tareas del Administrador