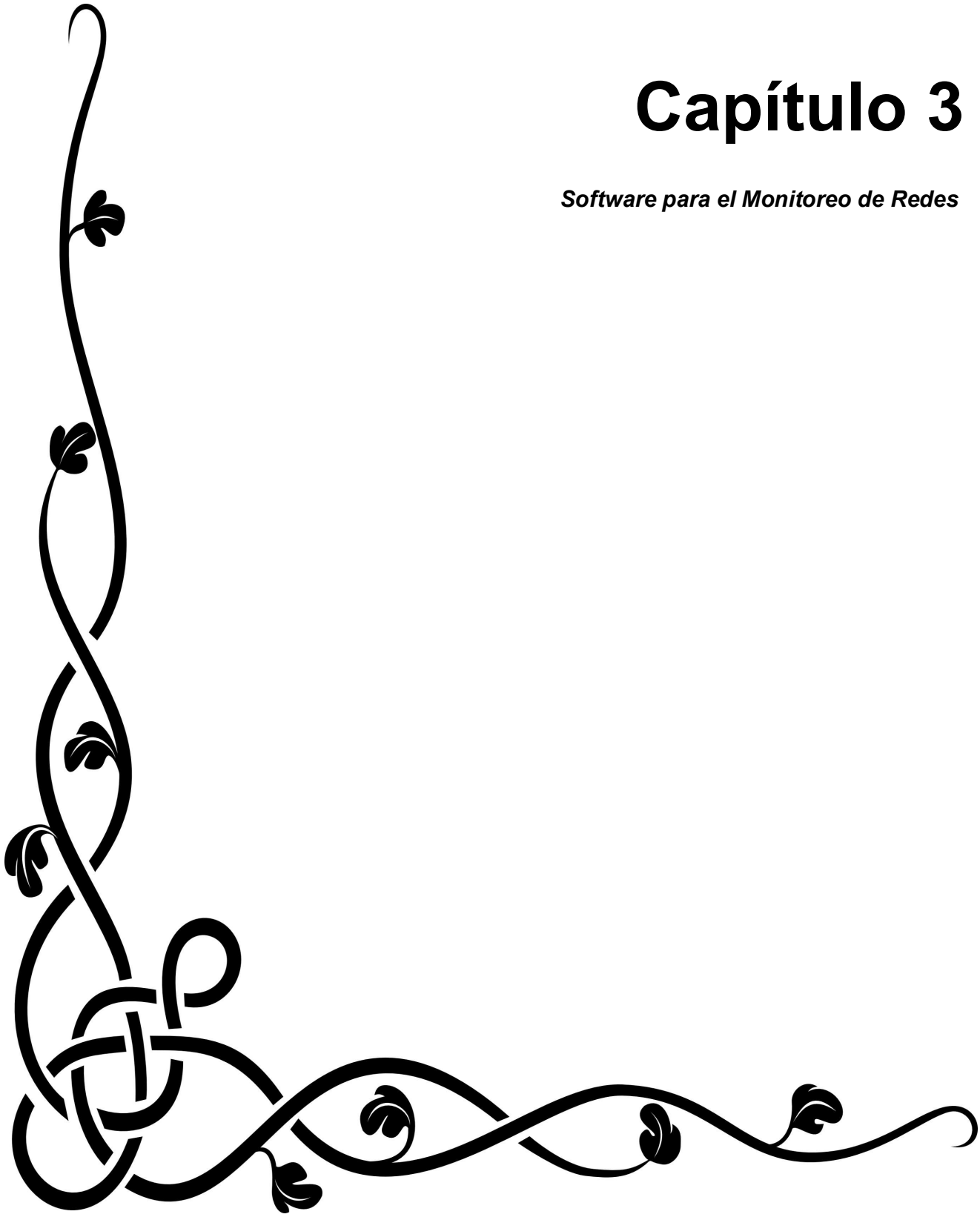

Capítulo 3

Software para el Monitoreo de Redes

"No basta saber, se debe también aplicar. No es suficiente querer, se debe también hacer."
Johann Wolfgang Goethe

Software para el Monitoreo de Redes

El estilo de trabajo en la actualidad ha encausado a muchas de sus funciones en el uso de redes de computadoras, por mencionar algunos casos están las instituciones bancarias, los centros de investigación, el comercio electrónico, la educación a distancia, los buscadores de información, entre muchos otros.

Conforme las redes de computadoras han incrementado en extensión y en demanda de servicios, el monitoreo de las mismas a adquirido mucha importancia ya que es necesario mantener un control de las actividades que se realizan y de la cantidad de información que circula en ella para ofrecer un servicio confiable de comunicación, debido a que en la actualidad muchas personas dependen de una red para poder desarrollar su trabajo o incluso existen redes de apoyo a equipo médico de soporte vital.

Es por esto que el tener un apoyo visual de los datos recolectados durante el monitoreo y vigilancia de la red es de gran utilidad para determinar las causas de los problemas que se presenten como las caídas de red, los cuellos de botella y algunas otras cuestiones.

El monitoreo consiste en observar y recolectar información referente al comportamiento de la red y para el IFUNAM es importante considerar los siguientes aspectos.

La utilización de enlaces
Obtener la cantidad de ancho de banda utilizada por cada una de las interfaces de red de las computadoras personales, servidores, estaciones de trabajo y otros elementos que forman parte de la red de cómputo del IFUNAM.

Porcentaje de transmisión y recepción de información

Identificar los equipos de cómputo personales, servidores, estaciones de trabajo y otros elementos de la red que mas solicitudes hacen y atienden.

3.1 Elección del software

Antes de encontrar un software que se adecuara a nuestras necesidades fue necesario realizar un análisis de las aplicaciones que ofrecen actualmente los sistemas disponibles para el monitoreo de redes.

Durante este procedimiento se buscaron aplicaciones consideradas como software libre que se distribuyeran bajo la Licencia Pública General de GNU y, principalmente, que pudieran ejecutarse en sistemas operativos GNU/Linux.

Es importante mencionar que hay una gran variedad de opciones en software que pueden adecuarse a cada usuario, pero solo se mencionaran los programas que se investigaron debido a sus características.

3.1.1 CACTI

Software desarrollado con PHP que proporciona una interfaz de usuario muy completa para aprovechar las funcionalidades de las herramientas RRDTool. Es capaz de almacenar información para generar gráficas con datos obtenidos de una base de datos MySQL, además cuenta con soporte SNMP (Simple Network Management Protocol)[18] para la recolección de datos del tráfico de red.

Cacti recolecta los datos que el usuario necesita cada determinado tiempo para almacenarlos en la base MySQL y en los archivos de *planificación Round Robin*[19], este procedimiento permite que las fuentes de datos se actualicen constantemente y permita generar gráficas

[18] El Protocolo Simple de Administración de Red es un conjunto de estándares que facilita el intercambio de información de tráfico entre diferentes dispositivos de red.

[19] La planificación Round Robin es un algoritmo utilizado para seleccionar un grupo de datos, que coinciden en tiempo y lugar, de forma concurrente.

con datos confiables.

Una característica muy importante de Cacti es que ofrece la funcionalidad de manejo de usuarios, por lo que se podrán dar de alta a usuarios con diferentes privilegios. Esto permite tener usuarios con permisos para modificar parámetros en las gráficas y otros que solo puedan consultarlas.

3.1.2 CRICKET

Software de alto rendimiento encargado de dar seguimiento a los datos y desarrollado para ayudar a los administradores a visualizar y entender el tráfico de la red. Cuenta con un sistema de configuración jerárquico, lo que evita que la información se repita. Fue escrito en su totalidad con Perl.

Compuesto por un colector y un graficador. El colector de datos se ejecuta cada determinado tiempo y almacena la información en Bases de Datos Circulares (RRD - Round Robin Database), posteriormente con ayuda de las herramientas RRDTool y una Interfaz de Entrada Común (CGI - Common Gateway Interface)[20] puede visualizar las gráficas de los datos recolectados.

3.1.3 MRTG

Herramienta escrita en C y en Perl para supervisar la carga de tráfico en determinadas interfaces de red. MRTG (Multi Router Traffic Grapher) se diseñó para recolectar contadores de tráfico SNMP y el registro de datos para convertirlos en gráficas, que posteriormente serán embebidas en una página web.

Se pueden generar gráficas que muestren el tráfico diario, semanal y anual de prácticamente cualquier dispositivo conectado a la red. Esto ayuda a tener una visualización rápida cuando hay problemas en la red.

[20] Mecanismo de comunicación entre el servidor web y una aplicación externa. Las aplicaciones CGI permiten crear contenido dinámico en páginas web.

3.1.4 IPTRAF

Programa que nos sirve para generar estadísticas de red, es una aplicación que se ejecuta en la terminal utilizando una interfaz de usuario desarrollada con la librería *curse*[21]. IPTRAF es un monitor que muestra información del tráfico IP que pasa a través de la red.

Permite elegir entre diferentes servicios como el monitoreo local, especificación de dispositivos que se desean controlar, filtrado de paquetes de diferentes protocolos, generar archivos de registros, búsqueda de DNS inversa para ofrecer mayor información.

Cada una de los programas anteriores ofrecía opciones que solucionarían parte del problema expuesto en dicho trabajo de Tesis, pero la Secretaría Técnica de Cómputo del IFUNAM buscaba un software más específico, por lo que se llegó a la decisión de utilizar dos herramientas:
RRDTool como el programa graficador.
IP Flow Meter como el programa recolector de información.

A continuación se hablará de cada uno de ellos más detalladamente.

3.2 IP Flow Meter (IPFM)

IPFM es una herramienta encargada de analizar el ancho de banda, es decir que mide la cantidad de datos que son enviados y recibidos a través de un enlace de internet para determinados *hosts* dentro de una red.

IPFM fue desarrollado usando la librería libpcap, la cual fue escrita en C para facilitar su portabilidad y proporciona al usuario una interfaz que facilita la captura, mediante funciones, de paquetes en la capa de red. Este sistema es utilizado para la monitorización de redes de bajo nivel.

[21] Librería de programación para el control de terminales en sistemas UNIX.

Ventajas de IPFM

- IPFM genera archivos de texto cada determinado tiempo, los cuales contienen la información en bytes del consumo de ancho de banda por cada *host* configurado. A continuación se muestra un ejemplo del formato de cada archivo generado:

```
# IPFMv0.11.5 yyyy/mm/dd 00:00:00 (local time) -- dump every 0d00:00:00 -- listening on eth0
# Host                  In (bytes)    Out (bytes)    Total (bytes)
host1.dominio.com       2575494             0         2575494
host2.dominio.com             0       2407417         2407417
host3.dominio.com             0         69312           69312
# end of dump yyyy/mm/dd 00:00:00
```

- Se definen los *hosts* de los que deseamos generar un registro de actividad.
- Podemos definir el intervalo de tiempo de salida, es decir cada cuánto determinado tiempo generaremos un archivo de registro.
- Se definen los nombres de los archivos de registro, en los cuales se incluye la fecha para tener una mejor referencia.
- Se puede configura la *resolución inversa de DNS*[22].
- Se realiza la clasificación de bytes en IN (la cantidad de paquetes recibidos), OUT (la cantidad de paquetes enviados) y TOTAL (la suma de IN y OUT es la cantidad total de datos transferidos).

Desventajas de IPFM

- IPFM se diseñó para realizar el registro de actividades de un solo dispositivo de red.
- Hay poca documentación actualizada.

[22] Traducción de una dirección IP a su nombre de dominio.

3.3 Herramientas de Base de Datos Circulares (RRDTool)

RRDTool es un programa de alto rendimiento utilizado para el registro de datos (tráfico de red, carga de servidores, temperatura, etc.) y su representación mediante gráficas en diferentes intervalos de tiempo. Los datos son almacenados en bases de datos compactas que tienen la característica de no crecer con el tiempo.

3.3.1 Bases de Datos Circulares (RRD – Round Robin Database)

En la actualidad el uso de las bases se datos se ha convertido en una actividad muy común, ya que la independencia física de la información ha favorecido el desarrollo de sistemas encargados de administrar los datos almacenados.

Una base de datos es un conjunto de archivos o datos relacionados y existe una colección de programas diseñado para crear y administrar bases de datos llamados sistema de bases de datos. Un sistema de base de datos es básicamente un sistema computarizado cuya finalidad general es almacenar información y permitir que los usuarios puedan recuperar y actualizar dicha información con base en peticiones.[23]

En la Figura 3.1 podemos apreciar la relación que existe entre una base de datos y un sistema de bases de datos.

Las bases de datos pueden clasificarse en diversos rubros, de acuerdo a su función o de acuerdo a su modelo de administración, aunque al final el objetivo es el mismo, almacenar información. Pero debido a la naturaleza del problema planteado en la presente tesis fue necesaria la búsqueda de soluciones diferentes a las que conocemos generalmente. Es así como se optó por utilizar las Round Robin Database - Bases de Datos Circulares.

[23] Fuente: C. J. Date “Introducción s los Sistema de Bases de Datos”

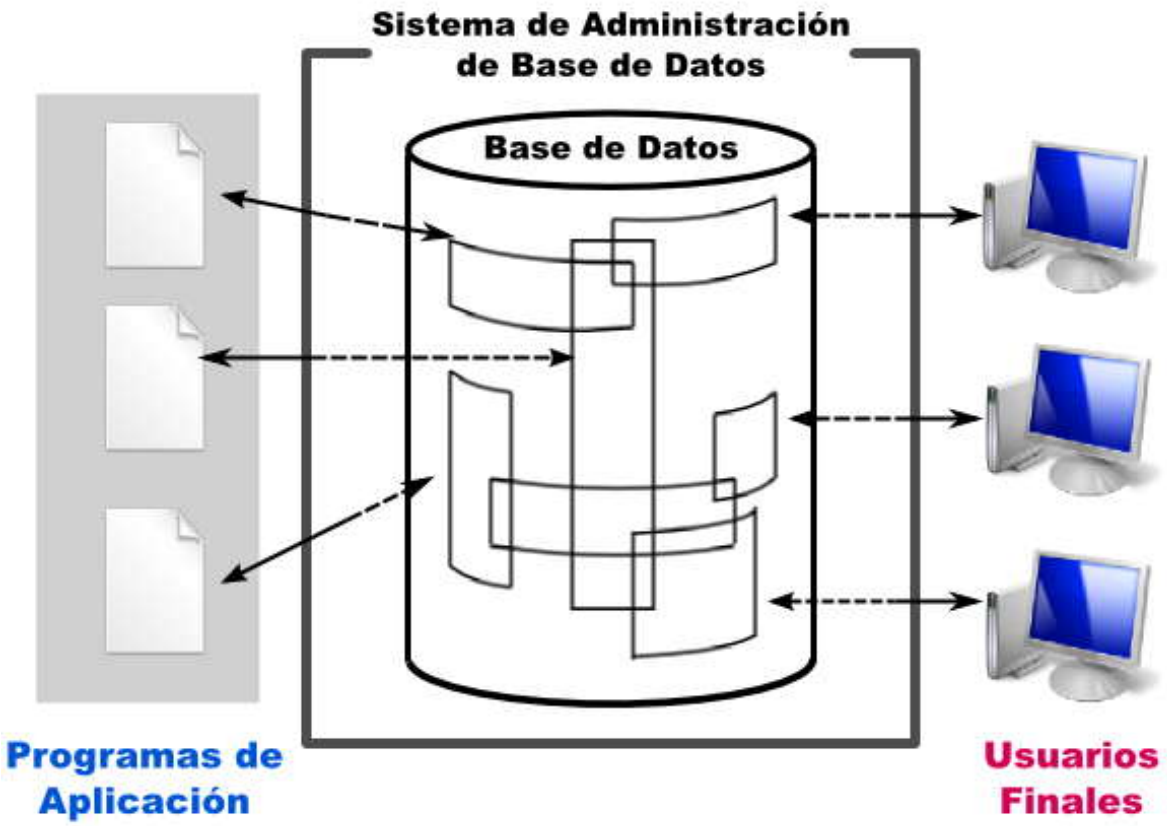

Figura 3.1 *Sistema de Base de Datos* [24]

Base de Datos Circular

Es un tipo de base de datos muy específica, orientada al almacenamiento de secuencias de datos medidos en determinados periodos de tiempo, espaciadas de forma uniforme y ordenadas cronológicamente. Trabaja con una cantidad fija de datos que es especificada en el momento de su creación y con un apuntador al dato más actual. Debido a su estructura sencilla son usadas principalmente como herramientas de monitoreo.

Razones para usar una Base de Datos Circular

Una RRD se caracteriza porque siempre contendrá la misma cantidad de datos, esto se logra gracias a que al llegar al límite máximo de almacenamiento, los datos se sobrescribirán en los más antiguos.

Para comprender mejor el concepto imaginemos un círculo dividido en varios sectores como

[24] Fuente: C. J. Date "Introducción s los Sistema de Bases de Datos"

se muestra en la Figura 3.2, dichos sectores representan la cantidad máxima de datos que se podrán almacenar en la RRD, cada determinado tiempo se ingresará información en uno de los sectores, la flecha nos sirve como apuntador para indicarnos cual es el dato más reciente, conforme la base vaya llenándose de datos llegará el momento en que se complete una vuelta del círculo. Al llegar nuevamente al inicio, los datos más nuevos comenzarán a sobrescribirse en los sectores más antiguos para continuar el siguiente ciclo.

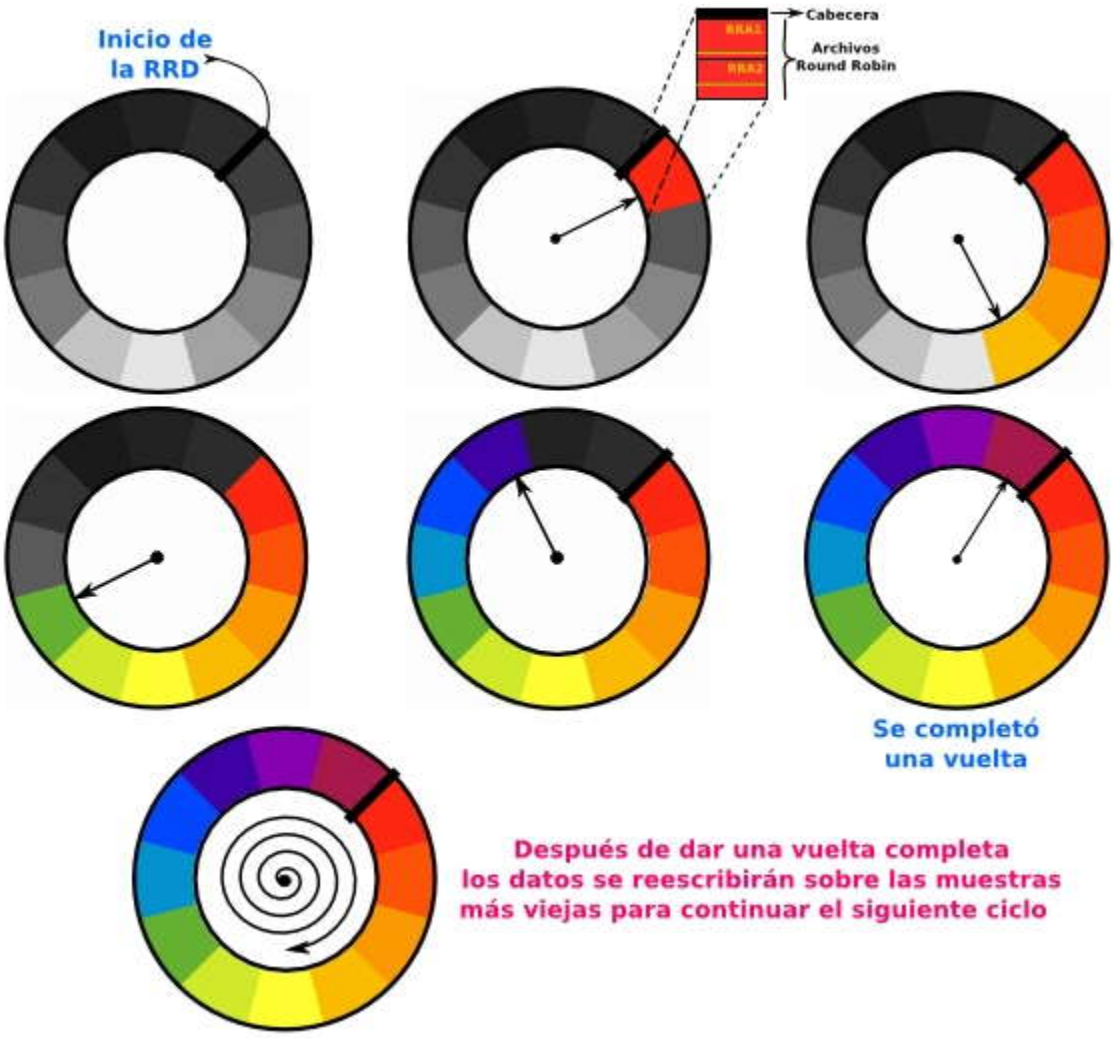

Figura 3.2 *Base de Datos Circular*

Este principio es la razón por la que se decidió usar las bases de datos circulares ya que al mantener un número fijo de datos (logrando que la base no crezca con el tiempo), evitamos depender físicamente del espacio de almacenamiento necesario para dicha información.

3.3.2 Componentes de una RRD

Las bases de datos circulares están organizadas de tal forma que cada componente cuenta con acciones específicas para su creación, a continuación se comentarán las características principales de cada uno de ellos.

Fuentes de Datos (DS)

Las fuentes de datos son las "muestras" o el origen de nuestros datos, cada RRD puede aceptar aportaciones de diversas fuentes de datos. En la Tabla 3.1 se describen los cuatro tipos de DS existentes:

Tipo DS	Descripción
GAUGE	Indicador de diferentes valores en el tiempo, registra los valores tal como lo medimos
COUNTER	Contador con incremento constante, se encarga de registrar el *incremento / intervalo de tiempo*
DERIVE	Contador que acepta valores negativos
ABSOLUTE	Contador que realiza un reset después de su lectura

Tabla 3.1 *Tipos de Fuentes de Datos*

Cuando configuramos una fuente de datos podemos definir las propiedades básicas de cada DS que deseamos almacenar en nuestra RRD. Dichas propiedades son el nombre, el tipo, el tiempo máximo que puede pasar antes de considerar los datos como desconocidos y los valores máximos y mínimos que pueden registrarse.

Funciones de Consolidación (CF)

Una función de consolidación consiste en un cálculo matemático o una selección lógica de los datos recolectados, se define en el momento de crear una RRD. Es necesario especificar el intervalo de tiempo en que ocurrirá dicha consolidación y que CF debe utilizarse para realizar la cuenta de los valores recolectados.

El uso de diferentes CF nos permite almacenar exactamente el tipo de dato que nos interesa, por lo que se cuenta con 4 tipos (Tabla 3.2):

CF	Tipo	Descripción
AVERAGE	Promedio	Se toma el promedio aritmético de los datos recolectados
LAST	Último valor leído	Se toma el último valor recolectado
MIN	Mínimo valor leído	Se toma el valor más pequeño recolectado
MAX	Máximo valor leído	Se toma el valor más grande recolectado

***Tabla 3.2** Tipos de Funciones de Consolidación*

Al crear una RRD se puede realizar una combinación de diferentes configuraciones de consolidación, por ejemplo, para el caso del Instituto de Física de la UNAM se recolectan datos cada 5 minutos y se definieron intervalos de consolidación para construir la cuenta de los valores cada 5 minutos, 24 horas, 1 mes y 1 año mediante las funciones AVERAGE, MIN y MAX para cada uno de ellos. Los datos consolidados son almacenados en los llamados Archivos Round Robin.

En la Figura 3.3 podemos observar un ejemplo sencillo de cómo se realiza el muestreo y la consolidación de los datos.

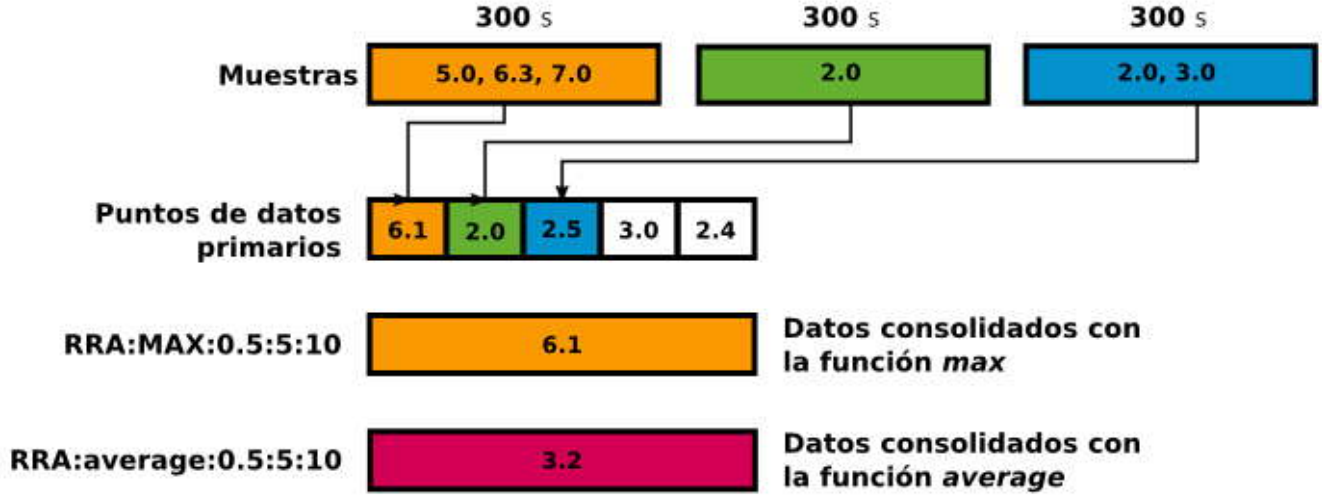

***Figura 3.3** Muestreo y Consolidación de Datos*

Cada 300 segundos (5 minutos) se recolectan datos y se establecen los puntos de datos primarios calculando el promedio de las muestras, por ejemplo, la primera muestra de la Figura 3.3 recolectó tres datos los cuales son sumados y divididos entre el número total de datos de la muestra [(5.0 + 6.3 + 7.0) / 3 = 6.1], así obtenemos el primer punto de datos

primario, posteriormente se calcula el promedio de la siguiente muestra [2.0 / 1 = 2.0] para el segundo punto de datos primario y así continuamente con todas las muestras obtenidas. A continuación se consolidan los datos de acuerdo a las funciones de consolidación que fueron definidas, para seguir con el ejemplo se puede observar en la figura dos tipos de funciones, la primera es la función MAX, la cual ontendrá de los puntos de datos primarios el máximo valor leído (6.1), la segunda es la función AVERAGE la cual dará como resulta el promedio de los puntos de datos primarios (3.2). Estos datos consolidados serán almacenados en los *Archivos Round Robin.*

Archivos Round Robin (RRA)

Los RRA son archivos configurados para indicar que tipo de datos queremos registrar, por lo que es necesario usar una función de consolidación. Estos archivos contienen un número limitado de datos consolidados para cada una de las fuentes de datos (DS) definidas.

Los datos adquiridos durante la recolección de información deben ser ingresados en una RRD cada determinado intervalo de tiempo, dichos intervalos son definidos en segundos. Estos se conocen como los puntos de datos primarios. Los datos deben pasar por el proceso de consolidación mediante las funciones que fueron especificadas en el momento de su creación.

En la figura 3.4 se observa un ejemplo sencillo de cómo se genera un RRA. Cuando se obtienen las muestras y se establecen los puntos de datos primarios, es necesario consolidar los datos, este proceso genera un RRA el cual contiene un número fijo de muestras durante un periodo de tiempo establecido. Cada RRA es almacenado en la base de datos circular, la cual también tiene un tamaño fijo, si ésta se encuentra en su límite el RRA se escribirá sobre el RRA más viejo, es pocas palabras, se descartará la muestra más vieja.

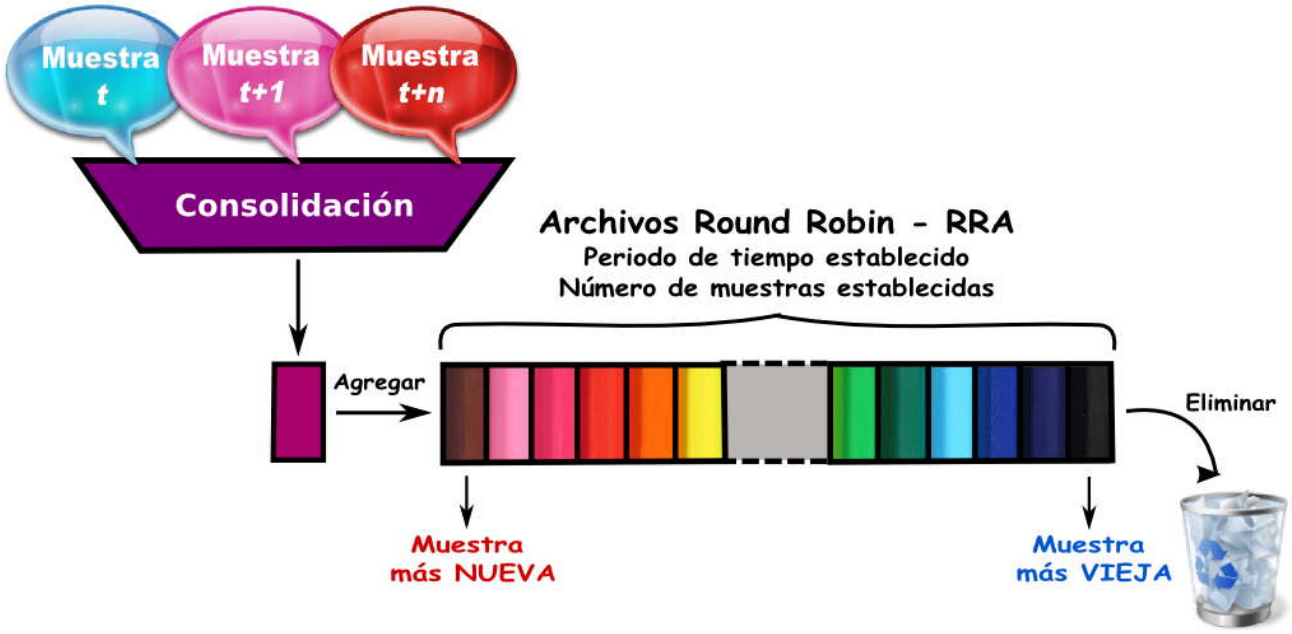

Figura 3.4 *Archivos Round Robin*

3.3.3 Herramientas más útiles

Las herramientas proporcionadas por RRDTool nos servirán principalmente para interactuar con la base de datos. Las funciones más utilizadas son:

rrdtool create

Esta función nos permite crear un nuevo archivo RRD. El archivo es creado con su tamaño completo y llenado con datos tipo UNKNOWN (desconocido).

rrdtool update

Lleva a cabo la actualización de los archivos RRD almacenando un conjunto de datos nuevos. Los datos son procesados de acuerdo a la configuración de las bases de datos.

rrdtool graph

Genera gráficos a partir de los datos de uno o más archivos RRD para mostrar la información en una forma que sea fácil de entender para las personas. Su principal objetivo es crear gráficas pero también puede generar reportes numéricos.

rrdtool fetch

Permite extraer información de un determinado periodo de tiempo. Esta instrucción es utilizada internamente por la función graph para obtener los datos de los archivos RRD, analiza dichos archivos e intenta recuperar la información en el tiempo solicitado.

rrdtool info

Nos proporciona la información de la cabecera de los archivos RRD en un formato amigable.

rrdtool rdcgi

Genera páginas web dinámicas en las que se incluyen las gráficas RRD. Su objetivo es ejecutar un programa CGI y generar una plantilla web que contenga etiquetas especiales, es decir, la impresión de una página web que incluya las cabeceras CGI necesarias.

Ventajas de RRDTool

- Se requiere poco espacio de almacenamiento ya que las bases de datos tienen un tamaño fijo y no crecen con el tiempo.
- Se pueden generar gráficas (.png y .gif) con atributos que son definidos por el usuario.
- Las gráficas pueden presentar reportes numéricos.
- Cada gráfica puede presentar información de una o varias fuentes de datos.
- Debido a que diversos programas administradores de red utilizan dicha herramienta, es fácil encontrar documentación que nos enseñe a utilizarla.

Desventajas de RRDTool

- Es necesario utilizar procesos externos para recolectar la información e insertar los datos en las RRDs correspondientes.
- Si no se configuran adecuadamente los intervalos de tiempo para la recolección de datos habrá información que se perderá.