**Estadísticas web con webalizer**.

Introducción

Cuando tenemos varios servidores web y deseamos hacer estadísticas del uso de los mismos una buena opción es usar el paquete webalizer. Esta aplicación genera estadísticas gráficas en formato html que se pueden publicar en un servidor web.

Instalación

Bueno, esto dependerá de nuestra distribución de linux. Si usamos RedHat, por ejemplo, podemos encontrar el paquete en los CD´s de instalación. Para instalar el paquete usamos el comando:

```
# rpm -Uvh webalizer-xx.rpm
```

En cambio, si usamos debian, podemos utilizar la herramienta apt-get para instalar el paquete con la última versión desde internet. El comando sería:

```
# apt-get install webalizer
```

Modo de uso

Si tenemos varios servidores web, es posible que tambien tengamos diferentes virtualhost en ellos. Para hacer estadísticas del uso del servidor en si mismo y de los diferentes virtualhosts, webalizer utiliza el fichero de log que guarda los accesos del tipo http al servidor. Si para cada

virtualhost tenemos definido un fichero de log diferente, deberemos usarlo también.

Esto ficheros de log van rotando cada cierto tiempo (generalmente cada semana), manteniendo el sistema un histórico de accesos de, generalmente, cuatro ficheros por cada log de accesos. Esto quiere decir, que si un servidor guarda sus accesos en el fichero access_log, cada cierto tiempo este fichero se guardará con el nombre access_log.1 (se incrementa en uno el nº de log) y se creará un nuevo fichero access.log vacio. Lo mismo ocurrirá para los ya existentes ficheros de logs históricos que pudiera haber, llamados access_log.1, access_log.2, access_log.3, access_log.4. Con cada rotación de los logs, se incrementa en uno el nº del log, excepto para el access_log.4, que se elimina (si en la configuración de la rotación de logs está puesto para que se mantengan 4 ficheros).

Básicamente el funcionamiento de esta aplicación es el siguiente:

Se ejecuta el comando webalizer pasándole como parámetro el fichero con los accesos log del servidor o virtualhost del cual queramos generar estadísticas (por ejemplo access_log.1). Para un mayor control del proceso, webalizer utiliza un fichero de configuración donde se guardan parámetros como el directorio donde almacenar las páginas web generadas tras el procesado del fichero de logs.

Vamos a ver un ejemplo simple:

Supongamos que tenemos un servidor web llamado "Homer" y definidos dos virtualhosts, "Bart" y "Lisa". Pues bien, debemos fijarnos que ficheros de logs utiliza el apache para cada uno de ellos. Para ellos editamos el fichero /etc/hhttpd/conf/httpd.conf y buscamos las siguientes lineas en la sección general y en cada uno de los virtualhosts:

customlog.....

Podriamos encontrar algo como esto:

Para la sección general el fichero de log se llama: access_homer.log

Para el primer virtualhost el fichero de log usado es: access_bart.log

Para el segundo virtualhost el fichero de log usado es: access_lisa.log

Nota: Hay que tener en cuenta de que se debe usar el fichero de log una vez ya ha sido rotado, es decir, que si, por ejemplo, Homer guarda sus accesos en access_homer.log, el fichero que realmente se debe usar es access_homer_1.log

Una vez que ya sabemos que ficheros son los que utiliza apache para guardar los logs de accesos al servidor web, vamos a preparar la ejecución de webalizer sobre dichos ficheros.

FICHEROS DE CONFIGURACIÓN:

Vamos a crear varios directorios de trabajo lo primero.

# mkdir -p /estadisticas/web/conf

Aquí van los fichero de configuración de webalizer para cada uno de los logs a analizar

# mkdir -p /estadisticas/web/hist

Aquí van los ficheros de histórico que usa webalizer para ir incrementando las estadisticas de cada mes.

# mkdir -p /var/www/html/estadisticas/homer

# mkdir -p /var/www/html/estadisticas/bart

# mkdir -p /var/www/html/estadisticas/lisa

Aquí iran las paginas web con las estadisticas generadas por webalizer. Este directorio debe ser accesible en el servidor web, para poder ver las páginas html contenidas en ellos.

Copiamos ahora el fichero /etc/webalizer.conf con un nombre diferente para cada log a analizar en el directorio de configuración:

# cp /etc/webalizer.conf /estadisticas/web/conf/homer.conf

# cp /etc/webalizer.conf /estadisticas/web/conf/bart.conf

# cp /etc/webalizer.conf /estadisticas/web/conf/lisa.conf

Dentro de cada uno de ellos se deben hacer los siguientes cambios:

LogFile /var/log/httpd/access_log

por

LogFile /var/log/httpd/

OutputDir /var/www/html/usage

por

OutputDir /var/www/html/estadisticas/

HistoryName /var/lib/webalizer/webalizer.hist

por

HistoryName /estadisticas/web/hist/.hist

IncrementalName /var/lib/webalizer/webalizer.current

por

IncrementalName /estadisticas/web/hist/.current

Descomentamos las lineas:

#ReportTitle Usage Statistics for

#HostName localhost

y cambiamos localhost por el nombre de la máquina o virtualhost que queramos analizar.

De resto podemos dejar el fichero tal y como está.

**EJECUCIÓN DE WEBALIZER.**

Teniendo en cuenta que los logs generalmente rotan 1 vez por semana, a no se ser que se haya modificado expresamente, se tiene que ejecutar webalizer una vez por semana tambien, mejor, si es justo después de esta rotación de los logs, para que en las gráficas nos aparezcan los datos más actualizados.

Lo más sencillo es usar un pequeño script que nos ejecute webalizer para cada una de las máquinas o virtualhosts que queramos. Lo creamos es: /estadisticas/web/conf/estadisticas.sh. Quedaría así:

#!/bin/bash

webalizer -c /estadisticas/web/conf/homer.conf

webalizer -c /estadisticas/web/conf/bart.conf

webalizer -c /estadisticas/web/conf/lisa.conf

Y lo metemos en el cron de la máquina para que se ejecute semanalmente:

#crontab -e

00 01 * * 3 /estadisticas/web/conf/estadisticas.sh

Con esto conseguimos que las estadísticas se generen cada miercoles a la 1:00 de la mañana. Los resultados podemos consultarlos a través de las páginas que genera webalizer y que publica a través del servidor web apache en la ruta donde le especificamos en el fichero de configuración.