

Universidad de Valladolid
E.T.S Ingeniería Informática
Grado en Ingeniería Informática [Tecnologías de la información]

Curso 2017/2018
Evaluación de Sistemas Informaticos
Práctica de Laboratorio 2
Evaluación del rendimiento de un servidor web.

Grupo de Laboratorio 07
Esteban Pellejero, Sergio
González Bravo, Miguel
González Pérez, Daniel
Rabadán Martín, Borja

Índice

1. Introducción y objetivos	3
2. Información disponible	3
3. Herramientas utilizadas	4
3.1. Configuración de AWSTATS	4
4. Análisis de utilización	5
4.1. Respuestas del servidor	5
5. Bibliografía	8
Appendices	9
ANEXO ANEXO I.	9
ANEXO ANEXO II.	9
ANEXO ANEXO III.	10

1. Introducción y objetivos

En la práctica anterior determinábamos el funcionamiento de un sistema informático en base a las pruebas y cargas que nosotros realizábamos contra él. A partir de esas pruebas extraíamos unos datos y unas conclusiones que nos permitían saber donde estaban los cuellos de botella, los recursos más demandados, etc. Sin embargo el objetivo de esta práctica es caracterizar y analizar el comportamiento de un servidor web en base a los ficheros de registro que han quedado almacenados en él. Con estos datos podremos saber a que recursos se accede más frecuentemente, a que horas tenemos que evitar por ejemplo hacer el mantenimiento del sistema (ya que hay mucha demanda de usuarios), etc.

En resumen, vamos a caracterizar la carga de nuestro servidor web, pero siguiendo un método diferente. También haremos un modelo teórico de la carga que ha soportado ese servidor web durante el periodo de tiempo suministrado por los archivos de log.

2. Información disponible

Para conseguir evaluar nuestro sistema informático, el servidor de la escuela de Ingeniería Informática, y hacer el modelo de carga, únicamente vamos a disponer de un fichero de log generado por el propio servidor Apache, durante las fechas 18 de Octubre de 2017 y 3 de Diciembre de 2017. En total vamos a analizar 47 días de funcionamiento \simeq 1 mes y medio.

Este fichero log de apache tiene dos características diferenciadas de los ficheros autogenerados por los servidores convencionales. las diferencias son:

- **Direcciones ip:** como veremos más adelante, en el log se reflejan las direcciones desde las que se ha accedido al servidor. ya que esto es un estudio teórico para una asignatura, se va a mantener la confidencialidad y las direcciones están anonimizadas.

La estructura que tienen los logs de apache la podemos encontrar en el enunciado de la práctica, por lo tanto no vamos a definir aquí lo que es cada campo. Pero para entenderlo mejor sí que vamos a poner un pequeño ejemplo mostrando en el fichero de log, los diferentes campos, separados por colores. Los campos de los ficheros de log, y la estructura del fichero que vamos a analizar se puede ver en la **Figura 1**.



```
ip1601 - [18/Oct/2017:09:35:40 +0200] "GET / HTTP/1.1" 200 27974 39393
```

Figura 1: Formato del fichero de log de Apache2

Cada uno de los colores corresponde a un campo diferente:

- **Cliente remoto** → verde claro.
- **Autenticación** → azul claro.
- **Autenticación** → rosa claro.
- **Fecha y hora de la petición** → amarillo.

- **Petición** → gris.
- **Status** → verde.
- **Volumen transferido** → azul.
- **Elapsed time** → rosa.

Como hemos dicho el significado de los campos lo podemos encontrar en el enunciado de la práctica o sino en el manual de referencia oficial proporcionado por Apache¹

3. Herramientas utilizadas

Para la parte de análisis de los ficheros de log de apache hemos decidido utilizar un programa llamado AWSTATS, que nos proporciona de manera automatizada la mayoría de la información requerida en la práctica. La parte de información que quizá no nos proporciona será sacada por programas propios o por algún otro programa como web log analyzer.

3.1. Configuración de awstats

AWSTATS es un software que nos permite ver de manera gráfica en una página HTML los resultados de un análisis automatizado de un fichero de log. Para poder utilizar de manera correcta Awstats hay que instalarlo en un servidor web ya configurado. Nosotros en la asignatura disponemos uno, pero en el momento de instalarlo, por mayor comodidad de acceso, el programa se ha instalado en un servidor propio, siendo. Es decir, podemos acceder a la información desde cualquier punto sin depender de las máquinas de la universidad. La URL es: *http://serverandroid.ddns.net:47000/cgi-bin/awstats.pl?config=uva.es*.

La configuración e instalación de este software es bastante simple. Tenemos que instalar los paquetes en el servidor y adaptar la configuración de AWSTATS para que coja los ficheros de log que nosotros le pasamos. Una vez hecho eso, configuramos el servidor web para que muestre los resultados en Internet y ya lo tenemos. La configuración exacta que se ha seguido, se puede ver en Internet².

Como podemos ver en la **Figura 2**, AWSTATS nos muestra la información de una manera muy amigable.

Algo a tener en cuenta durante el desarrollo de esta primera parte de la práctica son los tráficos de datos generados por los bots, web spiders o crawlers. Estos tráficos no se deberían de tener en cuenta para hacer el análisis del rendimiento del servidor ni para el modelado de la carga, ya que no son, por así decirlo, uso real del sistema por parte de usuarios finales. Por ejemplo los bots o los web spiders se caracterizan por conexión al servidor de muy poco periodo de tiempo, pero de gran frecuencia.

Empezamos con la primera parte de esta práctica que es el análisis de la utilización del servidor.

¹http://httpd.apache.org/docs/current/mod/mod_log_config.html

²<https://www.maketecheasier.com/set-up-awstats-ubuntu/>

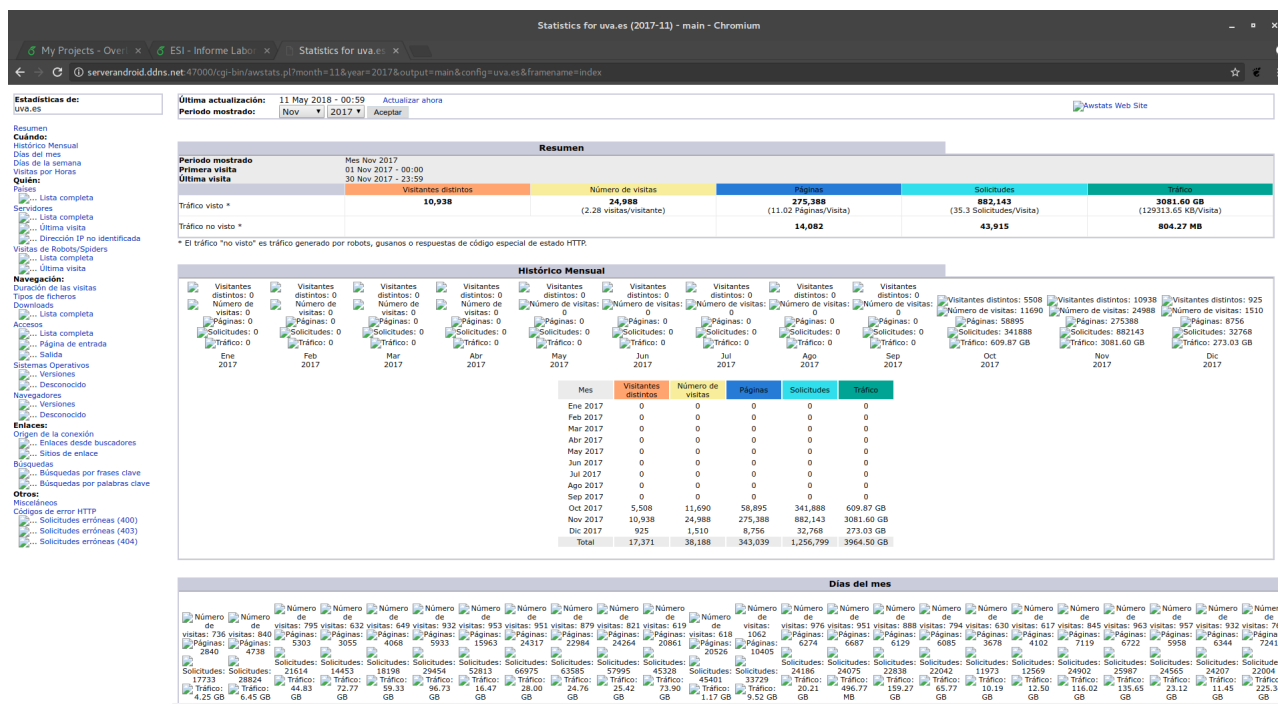


Figura 2: Visualización de la información con AWSTATS

4. Análisis de utilización

El análisis de utilización tiene implicaciones en la detección de problemas y ajuste del rendimiento del servidor. En este sentido el análisis de los datos de registro de accesos puede servir para abordar estudios relacionados con la decisión de qué documentos poner en caché, ajuste de sistemas relacionados con la gestión de la memoria caché, etc.

4.1. Respuestas del servidor

El análisis de las respuestas del servidor nos permite, por una parte, la identificación de las métricas a elegir en el estudio del rendimiento, y por parte, evaluar la calidad de servicio (QoS) ofrecida por nuestro sistema informático. Tenemos que saber que las posibles respuestas que se obtiene de un servidor web cuando se hace una petición http son las siguientes:

- **Éxito:** el usuario obtiene el documento, por lo que se consideran métricas de productividad, velocidad y utilización. No modificado, el usuario ya tiene una copia del documento y no se transfieren bytes.
- **Error:** documento encontrado pro en otro sitio.
- **Error:** documento no encontrado o no se dispone de permisos o privilegios suficientes.

Primero tenemos que especificar que la información ha sido proporcionada y extraída con AWSTATS y se ha hecho un resumen de las figuras que se encuentran en el ANEXO II (**Figura 5** y **Figura 12**). Tenemos que especificar que la **Figura 5** nos muestra la cantidad de peticiones que se han hecho al servidor y la cantidad de páginas. Hemos interpretado que la cantidad de páginas es la cantidad de respuestas con código 200, ya que aunque se han hecho muchas más peticiones,

algunas pueden haber devuelto códigos de error. Sin embargo las páginas servidas son las que respuestas correctas. A partir de estos datos hemos elaborado la siguiente figura (**Figura 3**) de resultados.

	Éxito (200)	Error (301+302)	Error (404)	Error (403)
Octubre	58.845	9.514	2.093	1.967
Noviembre	275.388	23.349	4.017	4.441
Diciembre	8.756	1.318	320	180

Figura 3: Tabla con las diferentes respuestas del servidor

Primero vamos a especificar los diferentes códigos de error:

- **200:** como hemos dicho anteriormente son las respuestas correctas, es decir que se ha servido la página correctamente.
- **301:** son las páginas movidas permanentemente, es decir, se han encontrado pero el sitio ha cambiado de manera permanente.
- **302:** son las páginas movidas temporalmente, es decir, se han encontrado pero el sitio ha cambiado de manera temporal.
- **404:** páginas no servidas porque no se ha encontrado la dirección.
- **403:** páginas no servidas porque no se disponía de los permisos necesarios para acceder a esa página.

Hemos decidido agrupar las métricas de los códigos 301 y 302 porque en el fondo son lo mismo y además las redirecciones temporales son minúsculas en comparación a las permanentes.

En nuestra opinión, lo importante es saber la distribución de como se realizan las peticiones correctas, el resto de peticiones, las erróneas suelen ser en momentos puntuales, que habrá que estudiar, pero creemos de mayor importancia saber las distribuciones más concretas de los códigos 200. Hemos elaborado la siguiente tabla, representada en la **Figura 4**. Esta tabla cuando se producen más visitas y menos visitas, filtrado por día de la semana y por horas del día.

	Días semana + páginas	Días semana - páginas	Horas día + páginas	Horas día - páginas
Octubre	X-V	S-D	9-11	4-6
Noviembre	J-V	S-D	8-19	1-6
Diciembre	V	D	10-16	2-8

Figura 4: Tabla con diferentes distribuciones de páginas servidas por el servidor

Como podemos observar, en general los días que más accesos hay al servidor son de mitad de semana hasta el viernes. Sin embargo los días que claramente hay menos accesos es durante el finde semana. Principalmente se debe al alma fiesterera de los estudiantes. También comentar las horas del día son lógicas, empiezan a servirse las peticiones entre las 8-9 de la mañana y su volumen es más o menos constante hasta las 18-19 de la tarde. A partir de esa hora parece que no hay demasiados accesos en comparación al resto de horas del día.

La tabla anterior ha sido confeccionada a partir de las figuras obtenidas por AWSTATS, que se encuentran en el ANEXO II (**Figura , Figura , Figura , Figura , Figura y Figura)**

5. Bibliografía

- [1] J.M. Marqués. Transparencias de la asignatura Evaluación de Sistemas Informáticos.

ANEXO I.

ANEXO II.

Mes	Visitantes distintos	Número de visitas	Páginas	Solicitudes	Tráfico
Ene 2017	0	0	0	0	0
Feb 2017	0	0	0	0	0
Mar 2017	0	0	0	0	0
Abr 2017	0	0	0	0	0
May 2017	0	0	0	0	0
Jun 2017	0	0	0	0	0
Jul 2017	0	0	0	0	0
Ago 2017	0	0	0	0	0
Sep 2017	0	0	0	0	0
Oct 2017	5,508	11,690	58,895	341,888	609.87 GB
Nov 2017	10,938	24,988	275,388	882,143	3081.60 GB
Dic 2017	925	1,510	8,756	32,768	273.03 GB
Total	17,371	38,188	343,039	1,256,799	3964.50 GB

Figura 5: Páginas servidas de manera correcta por el servidor

Códigos de error HTTP				
Códigos de error HTTP*		Solicitudes	Porcentaje	Tráfico
301	Moved permanently (redirect)	8,939	63.8 %	16.03 MB
404	Document Not Found (hits on favicon excluded)	2,093	14.9 %	122.96 MB
403	Forbidden	1,967	14 %	1.97 MB
302	Moved temporarily (redirect)	575	4.1 %	0
206	Partial Content	197	1.4 %	9.69 MB
500	Internal server Error	196	1.4 %	579.68 KB
421	Unknown error	16	0.1 %	6.05 KB
405	Method not allowed	10	0 %	10.89 KB
503	Server busy	2	0 %	618 Bytes

Figura 6: Páginas servidas de manera incorrecta por el servidor

Códigos de error HTTP				
		Códigos de error HTTP*		
		Solicitudes	Porcentaje	Tráfico
301	Moved permanently (redirect)	8,939	63.8 %	16.03 MB
404	Document Not Found (hits on favicon excluded)	2,093	14.9 %	122.96 MB
403	Forbidden	1,967	14 %	1.97 MB
302	Moved temporarily (redirect)	575	4.1 %	0
206	Partial Content	197	1.4 %	9.69 MB
500	Internal server Error	196	1.4 %	579.68 KB
421	Unknown error	16	0.1 %	6.05 KB
405	Method not allowed	10	0 %	10.89 KB
503	Server busy	2	0 %	618 Bytes

Figura 7: Páginas servidas de manera incorrecta por el servidor

Códigos de error HTTP				
Códigos de error HTTP*		Solicitudes	Porcentaje	Tráfico
301	Moved permanently (redirect)	8,939	63.8 %	16.03 MB
404	Document Not Found (hits on favicon excluded)	2,093	14.9 %	122.96 MB
403	Forbidden	1,967	14 %	1.97 MB
302	Moved temporarily (redirect)	575	4.1 %	0
206	Partial Content	197	1.4 %	9.69 MB
500	Internal server Error	196	1.4 %	579.68 KB
421	Unknown error	16	0.1 %	6.05 KB
405	Method not allowed	10	0 %	10.89 KB
503	Server busy	2	0 %	618 Bytes

Figura 8: Páginas servidas de manera incorrecta por el servidor

Códigos de error HTTP				
Códigos de error HTTP*		Solicitudes	Porcentaje	Tráfico
301	Moved permanently (redirect)	8,939	63.8 %	16.03 MB
404	Document Not Found (hits on favicon excluded)	2,093	14.9 %	122.96 MB
403	Forbidden	1,967	14 %	1.97 MB
302	Moved temporarily (redirect)	575	4.1 %	0
206	Partial Content	197	1.4 %	9.69 MB
500	Internal server Error	196	1.4 %	579.68 KB
421	Unknown error	16	0.1 %	6.05 KB
405	Method not allowed	10	0 %	10.89 KB
503	Server busy	2	0 %	618 Bytes

Figura 9: Páginas servidas de manera incorrecta por el servidor

Códigos de error HTTP				
Códigos de error HTTP*		Solicitudes	Porcentaje	Tráfico
301	Moved permanently (redirect)	8,939	63.8 %	16.03 MB
404	Document Not Found (hits on favicon excluded)	2,093	14.9 %	122.96 MB
403	Forbidden	1,967	14 %	1.97 MB
302	Moved temporarily (redirect)	575	4.1 %	0
206	Partial Content	197	1.4 %	9.69 MB
500	Internal server Error	196	1.4 %	579.68 KB
421	Unknown error	16	0.1 %	6.05 KB
405	Method not allowed	10	0 %	10.89 KB
503	Server busy	2	0 %	618 Bytes

Figura 10: Páginas servidas de manera incorrecta por el servidor

Códigos de error HTTP				
Códigos de error HTTP*		Solicitudes	Porcentaje	Tráfico
301	Moved permanently (redirect)	8,939	63.8 %	16.03 MB
404	Document Not Found (hits on favicon excluded)	2,093	14.9 %	122.96 MB
403	Forbidden	1,967	14 %	1.97 MB
302	Moved temporarily (redirect)	575	4.1 %	0
206	Partial Content	197	1.4 %	9.69 MB
500	Internal server Error	196	1.4 %	579.68 KB
421	Unknown error	16	0.1 %	6.05 KB
405	Method not allowed	10	0 %	10.89 KB
503	Server busy	2	0 %	618 Bytes

Figura 11: Páginas servidas de manera incorrecta por el servidor

Códigos de error HTTP				
Códigos de error HTTP*		Solicitudes	Porcentaje	Tráfico
301	Moved permanently (redirect)	8,939	63.8 %	16.03 MB
404	Document Not Found (hits on favicon excluded)	2,093	14.9 %	122.96 MB
403	Forbidden	1,967	14 %	1.97 MB
302	Moved temporarily (redirect)	575	4.1 %	0
206	Partial Content	197	1.4 %	9.69 MB
500	Internal server Error	196	1.4 %	579.68 KB
421	Unknown error	16	0.1 %	6.05 KB
405	Method not allowed	10	0 %	10.89 KB
503	Server busy	2	0 %	618 Bytes

Figura 12: Páginas servidas de manera incorrecta por el servidor

ANEXO III.