

Rastreo real: Análisis de los archivos LOGS

- A. Introducción al mundo de los logs
- B. Los Logs: El gran amigo del SEO
- C. Entendiendo el funcionamiento y sentando las bases
- D. Metodología de Análisis de Logs
- E. SFLA: Herramienta de Análisis de Logs
- F. Realizando análisis más profundos con Logs
- G. Cruzando otras fuentes de Datos con Logs

1. Introducción al mundo de los logs

Introducción al tema de los logs

La historia sobre el mundo en el que vivimos se compone de documentos que registran los eventos importantes de cada época. Gracias a estos registros podemos saber, por ejemplo, qué guerra aconteció en qué lugar, quién la ganó, o qué país se alió con qué país. Desde que el ser humano aprendió a escribir, comenzó a dejar constancia de los sucesos a su alrededor, dando lugar a los primeros diarios. Todos hemos escrito algún tipo de agenda en nuestra vida, dejando documentado nuestro día a día y plasmando nuestras reflexiones. Esto es precisamente lo que son los logs en el mundo de la informática: registros.

Los logs informáticos son documentos que registran las actividades de un sistema. Gracias a ellos podemos ser conscientes de los eventos relativos al funcionamiento de ese servicio o sistema. Estos logs suelen ser archivos de tipo .txt, .csv, JSON..., y suelen contener una serie de datos comunes:

- Fecha y hora del evento
- Dirección IP/identificación del usuario o dispositivo
- Acciones/solicitudes realizadas
- Estado del proceso o su resultado
- Mensajes de error, si los hay

Existen logs de diferentes tipos dependiendo del tipo de sistema que los genere, ya que cada uno cumple una función específica en la monitorización, diagnóstico o análisis de un proceso. Algunos de los tipos de logs más comunes según el sistema son:

- **Logs de servidor web** (como *Apache* ó *Nginx*): generan registros sobre las solicitudes de páginas web. Contienen información sobre la dirección IP del usuario, las URLs solicitadas, el código de estado HTTP (por ejemplo, 200 para éxito o 404 para no encontrado), la fecha y hora de la solicitud, y el User-Agent (que indica si la solicitud proviene de un navegador o un bot). Ayudan a rastrear el tráfico de un sitio web, a detectar errores o problemas de rendimiento y a analizar el comportamiento de los usuarios.
- **Logs de aplicaciones/sistemas de software:** registran las actividades dentro del propio sistema, y contienen información sobre el rendimiento, las transacciones realizadas, las operaciones realizadas por el usuario, y cualquier error o excepción. Estos logs son fundamentales para el diagnóstico de problemas en el software, la monitorización de su rendimiento y la mejora de la experiencia de usuario.
- **Logs del SO** (*Linux*, *Windows*, *macOS*...): son registros que documentan actividades relacionadas con el funcionamiento del propio sistema. Esto incluye el arranque del sistema, el uso de la CPU, los procesos en ejecución, la memoria y cualquier fallo o advertencia importante. Estos logs son esenciales para la administración de sistemas, la resolución de problemas y la seguridad, ya que permiten a los administradores identificar fallos, intrusiones o vulnerabilidades.
- **Logs de seguridad:** registran eventos relacionados con el acceso y la actividad en sistemas sensibles o redes. Pueden incluir registros de intentos de inicio de sesión, accesos no autorizados, modificaciones de archivos críticos o actividades sospechosas, siendo cruciales para la ciberseguridad.

- **Logs de BBDD:** registran eventos como consultas, transacciones, errores o intentos de acceso. Son útiles para **optimizar el rendimiento**, analizar el uso de la base de datos y resolver problemas relacionados con el almacenamiento y la integridad de los datos.
- **Logs de redes:** capturan información sobre el tráfico y las comunicaciones entre dispositivos en una red. Registran datos como la dirección IP, puertos de comunicación, protocolos utilizados y cualquier fallo en la conectividad o intentos de acceso no autorizados. Son útiles para monitorear el tráfico de red, detectar problemas de conectividad y asegurar la integridad de la red frente a posibles amenazas externas.

2. Los Logs: El gran amigo del SEO

Los logs: el gran amigo del SEO

Los **logs referentes al SEO** registran eventos y actividades referentes a un sitio web generados a medida que estos eventos ocurren. Estos archivos son generados por los servidores web para mantener un historial de todo lo referente al sitio web, y concretamente contienen información relativa a las solicitudes realizadas al sitio, tanto por usuarios como por bots de motores de búsqueda (algunos de los datos que incluyen estos registros son URLs rastreadas, códigos de respuesta HTTP, el agente de usuario y las marcas de tiempo de cada visita). Por lo tanto, **los logs de servidores web son especialmente relevantes en el mundo del SEO, ya que contienen información sobre cómo los bots y los usuarios interactúan con un sitio web.**

Estudiar estos logs permite obtener una visión única y profunda de cómo los motores de búsqueda perciben un sitio, qué partes están siendo rastreadas, y si hay problemas técnicos (como problemas de indexación, por ejemplo) que puedan estar afectando el rendimiento del sitio en los resultados de búsqueda.

Analizar los logs puede revelar:

1. **Cómo optimizar el presupuesto de rastreo:** Identificando si los bots están malgastando recursos en páginas no relevantes.
2. **Problemas de indexación:** Detectando errores recurrentes o contenido no accesible para los rastreadores.
3. **Rendimiento del sitio:** Monitoreando tiempos de carga y posibles cuellos de botella.
4. **Anomalías de tráfico:** Como accesos inesperados o patrones de uso sospechosos.

3. Funcionamiento y bases

Entendiendo el funcionamiento y sentando las bases

Para aprovechar al máximo el análisis de logs, es esencial comprender sus componentes principales y su funcionamiento. A continuación, se detallan los elementos clave con ejemplos y buenas prácticas:

- **IP:** Es la dirección desde donde se realiza la solicitud al servidor. Esto permite identificar el origen del tráfico, diferenciando entre usuarios reales y bots.
 - *Ejemplo:* Si una misma IP genera miles de solicitudes en pocos minutos, podría tratarse de un bot no deseado o un ataque de fuerza bruta.
- **Fecha y hora:** Indica el momento exacto en que ocurre cada interacción con el servidor. Esto es fundamental para analizar patrones temporales.
 - *Ejemplo:* Si notas un pico de actividad en horarios específicos, puedes correlacionarlo con tus campañas de marketing o con comportamientos típicos de los bots.
- **Método de solicitud:** Define el tipo de operación realizada. Los métodos más comunes son GET (obtener información) y POST (enviar datos).
 - *Ejemplo:* Si observas un número alto de solicitudes POST inesperadas, podrías estar frente a un intento de abuso en formularios.
- **URL:** Es la dirección del recurso solicitado. Aquí puedes identificar cuáles páginas están siendo accedidas y cuáles están siendo ignoradas.
 - *Ejemplo:* Si las páginas clave de tu sitio (como las de producto o servicios) tienen pocas solicitudes de rastreo, esto podría indicar problemas de indexabilidad.
- **Código de estado HTTP:** Indica el resultado de la solicitud (por ejemplo, 200 para éxito, 301 para redirección y 404 para no encontrado). Esto permite detectar problemas técnicos.
 - *Ejemplo:* Una alta cantidad de códigos 404 puede indicar enlaces rotos que deben ser corregidos para mejorar la experiencia del usuario y la eficiencia del rastreo.

También es importante utilizar herramientas para procesar estos datos, ya que los archivos logs pueden ser masivos y complejos.

4. Metodología de Análisis de Logs

Metodología de Análisis de Logs

Un análisis efectivo de logs sigue los siguientes pasos:

1. **Recolectar y almacenar logs:** Descargar los archivos directamente del servidor o configurar sistemas automáticos. Algunos servidores generan los logs de forma predeterminada, y estos se pueden descargar mediante FTP, paneles de control como cPanel o usando comandos en terminal (e.g., `scp` o `rsync`).
2. **Preparar los datos:** Antes de realizar un análisis, es necesario limpiar y estructurar los datos. Esto incluye eliminar entradas irrelevantes (como solicitudes de recursos estáticos: CSS, JS, imágenes) y asegurarse de que el formato sea consistente.
3. **Filtrar datos relevantes:** Identificar solicitudes realizadas por bots de búsqueda como Googlebot, Bingbot o Yandex. Esto puede lograrse filtrando por los user-agents registrados en los logs. Para verificar su autenticidad, se pueden comparar las IPs con las listas oficiales publicadas por los motores de búsqueda.
4. **Analizar el rastreo:** Determinar qué páginas reciben más atención y cuáles son ignoradas. Por ejemplo, ¿están los bots gastando tiempo en páginas irrelevantes? Esto puede implicar revisar la frecuencia de rastreo por URL, identificar patrones y observar si hay páginas importantes que no están siendo rastreadas.
5. **Detectar errores y patrones:** Identificar códigos de error (404, 500) o secciones del sitio que están generando demasiados errores. Además, observar patrones como excesivas solicitudes de una IP específica, lo que podría indicar actividad sospechosa o un mal uso del crawl budget.
6. **Segmentación por secciones o categorías:** Dividir el sitio en secciones clave (productos, blog, landing pages) y analizar cómo cada una está siendo rastreada. Esto permite priorizar esfuerzos en las áreas de mayor impacto.
7. **Visualizar los resultados:** Usar herramientas como Excel, Tableau o Python para crear gráficos que representen visualmente la actividad de los bots. Por ejemplo, un gráfico de barras para mostrar los códigos de estado más comunes o un diagrama de flujo del recorrido de los bots.
8. **Tomar decisiones basadas en datos:** Implementar soluciones como:
 - Bloquear páginas innecesarias en el archivo robots.txt.
 - Crear redirecciones 301 para URLs con errores recurrentes.
 - Optimizar enlaces internos para mejorar el rastreo de páginas clave.
 - Priorizar mejoras en las secciones con mayor actividad de rastreo.

Ejemplo: Si un bot está gastando gran parte del presupuesto en una página obsoleta, puedes bloquear su acceso mediante el archivo robots.txt o eliminar la URL si ya no es relevante.

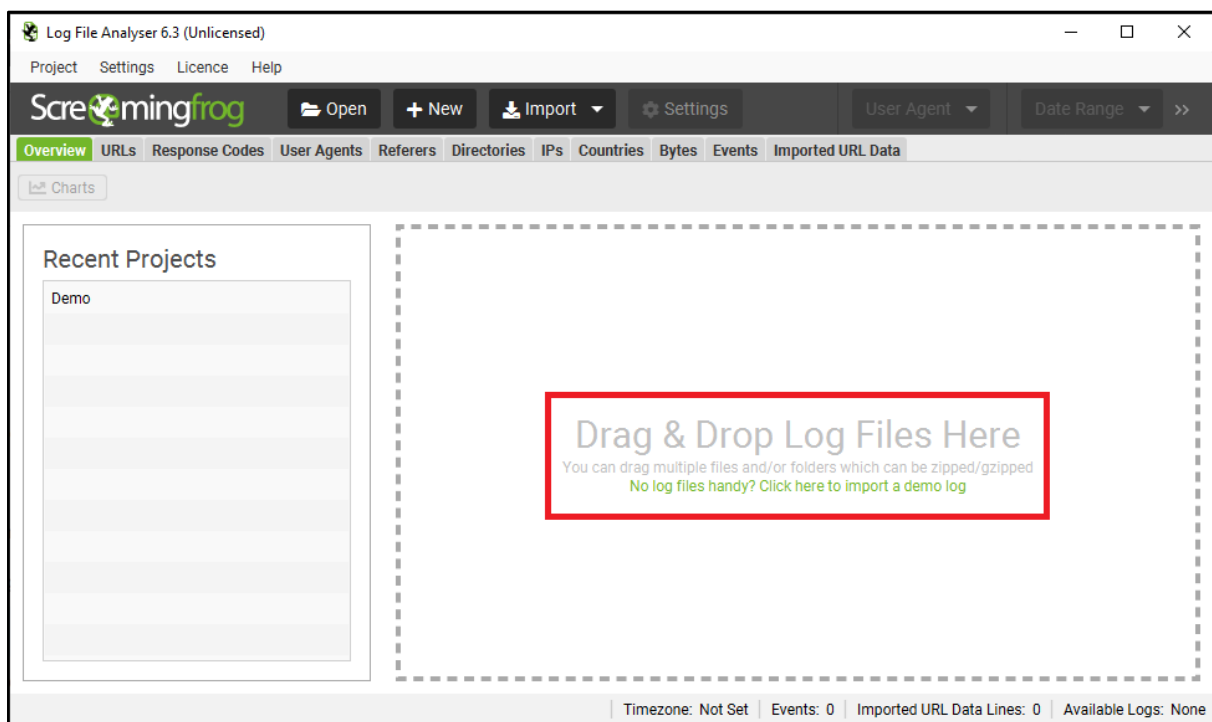
5. SFLA: Herramienta de Análisis de Logs

SFLA: Herramienta de Análisis de Logs

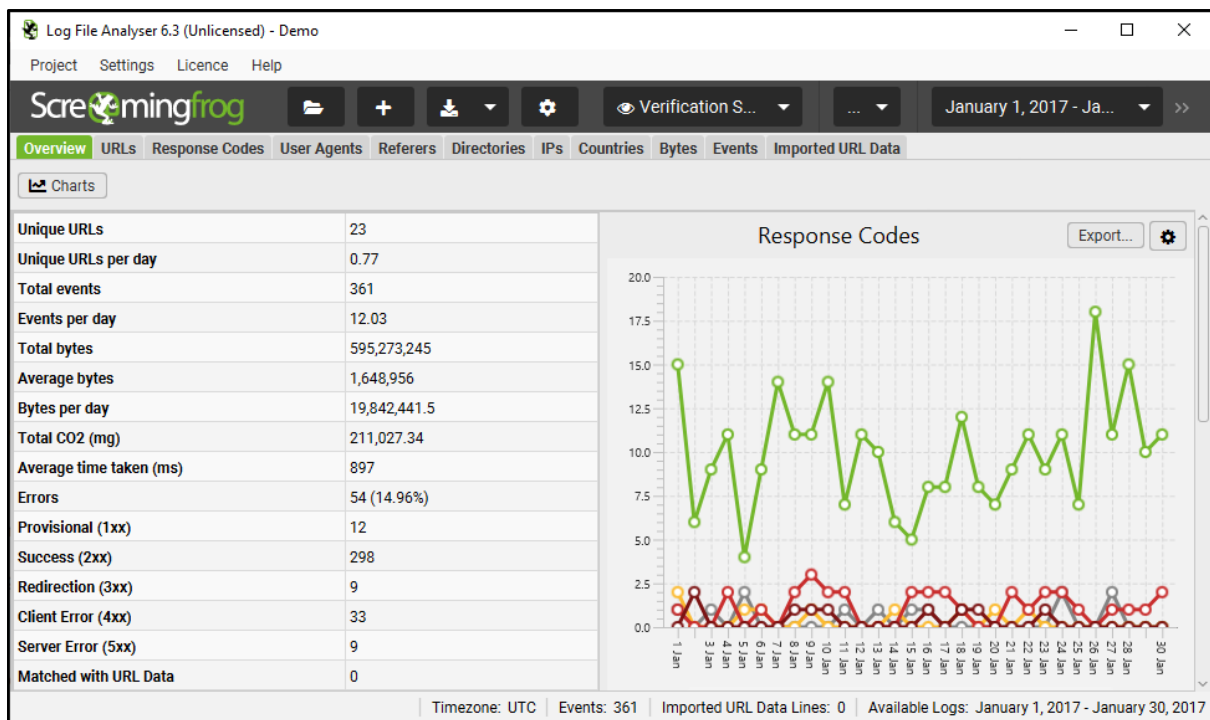
SFLA (*SEO Log File Analyser*) es una **herramienta específicamente diseñada para analizar estos archivos de registro desde una perspectiva del posicionamiento web**. Su objetivo es ayudar a los especialistas SEO y a los administradores web a identificar indicios bajos de optimización a la hora de la visibilidad de un sitio/aplicación web.

Cabe mencionar que este programa no es gratuito, aunque sí que cuenta con una capa gratuita que nos permite analizar 1000 líneas de registros y tener hasta 1 proyecto activo. Descargaremos el programa para el sistema operativo pertinente y seguiremos los pasos de instalación correspondientes.

Para poder ver el funcionamiento de dicho programa, importaremos un archivo de log ya existente y analizaremos su contenido.



Al importar dicho archivo de log, se nos creará un proyecto automáticamente con el nombre "Demo". Si lo abrimos, nos aparecerá una ventana con información resumida del log.



Este programa cuenta con varias pestañas a las que podemos acceder e inspeccionar con más detenimiento el contenido del archivo de registro. Empezaremos por la segunda pestaña, denominada “URLs”.

La pestaña *URL*

En la pestaña URLs, se pueden observar aquellas que se encuentran en el fichero de registro, junto con información relevante, como la petición de respuesta y la fecha cuando se hizo la petición. Si hacemos clic en una URL, veremos más detalles sobre la misma (número de bytes total, total de emisiones CO2, etcétera).

Row	URL	Last Response Code	Time Of Last Response
1	http://example.com/	200	Jan 30, 2017, 10:28:44 AM
2	http://example.com/used_to_work.php	404	Jan 30, 2017, 3:28:26 PM
3	http://example.com/used_to_work	200	Jan 30, 2017, 2:40:32 PM
4	http://example.com/slow_page.php	200	Jan 28, 2017, 1:56:13 PM
5	http://example.com/download/press.pdf	200	Jan 30, 2017, 4:56:03 AM
6	http://example.com/intro.swf	200	Jan 30, 2017, 12:01:52 PM

Filter Total: 23

Export

Name	Value
URL	http://example.com/
Last Response Code	200
Time Of Last Response	Jan 30, 2017, 10:28:44 AM
Days Since Last Crawl	0
Content Type	text/html

URL Details | Events | Referers | Chart

Timezone: UTC | Events: 361 | Imported URL Data Lines: 0 | Available Logs: January 1, 2017 - January 30, 2017

La pestaña *Response Codes*

Aquí podremos consultar información sobre los diferentes códigos de estado que se han hecho a las diferentes páginas (como 200, 301, 400, etcétera).

The screenshot shows the Screamingfrog Log File Analyser 6.3 (Unlicensed) - Demo interface. The 'Response Codes' tab is selected, displaying a table of log entries. The table has columns for Row, URL, Last Response Code, and Time Of Last Response. The first row is highlighted in green. Below the table, there is an 'Export' button and a summary table with fields like Name, URL, Last Response Code, Time Of Last Response, Days Since Last Crawl, and Num Events. At the bottom, there are tabs for 'URL Details', 'Events', 'Referers', and 'Chart', and a status bar showing 'Timezone: UTC', 'Events: 361', 'Imported URL Data Lines: 0', and 'Available Logs: January 1, 2017 - January 30, 2017'.

Row	URL	Last Response Code	Time Of Last Response
1	http://example.com/	200	Jan 30, 2017, 10:28:44 AM
2	http://example.com/used_to_work.php	404	Jan 30, 2017, 3:28:26 PM
3	http://example.com/used_to_work	200	Jan 30, 2017, 2:40:32 PM
4	http://example.com/slow_page.php	200	Jan 28, 2017, 1:56:13 PM
5	http://example.com/download/press.pdf	200	Jan 30, 2017, 4:56:03 AM
6	http://example.com/intro.swf	200	Jan 30, 2017, 12:01:52 PM

Name	Value
URL	http://example.com/
Last Response Code	200
Time Of Last Response	Jan 30, 2017, 10:28:44 AM
Days Since Last Crawl	0
Num Events	70

Filter Total: 23

Export

URL Details Events Referers Chart

Timezone: UTC | Events: 361 | Imported URL Data Lines: 0 | Available Logs: January 1, 2017 - January 30, 2017

La pestaña *User Agents*

Esta pestaña nos indica quién está accediendo a nuestra página y nos muestra algunas características que veremos en la siguiente sección.

Log File Analyser 6.3 (Unlicensed) - Demo

Project Settings Licence Help

Screamingfrog Open + New Import Settings Verification Status Show All

Overview URLs Response Codes **User Agents** Referers Directories IPs Countries Bytes Events Imported URL Data

Export Search User Agent

Row	User Agent	Unique URLs	Num Events	Total Bytes	Average Bytes	Tot +
1	Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.h...	20	99	150,177,522	1,516,944	53,
2	Mozilla/5.0 (iPhone; CPU iPhone OS 7_0 like Mac OS X) AppleWebKit/5...	22	73	88,777,032	1,216,123	31,
3	Mozilla/5.0 (Windows Phone 8.1; ARM; Trident/7.0; Touch; rv:11.0; iEMo...	21	60	128,134,032	2,135,567	45,
4	Mozilla/5.0 (compatible; bingbot/2.0; +http://www.bing.com/bingbot.ht...	18	60	122,605,545	2,043,425	43,
5	Mozilla/5.0 (Linux; Android 6.0.1; Nexus 5X Build/MMB29P) AppleWebK...	16	42	61,193,896	1,456,997	21,
6	Mozilla/5.0 (iPhone; CPU iPhone OS 8_3 like Mac OS X) AppleWebKit/6...	15	27	44,385,218	1,643,896	15,

User Agents Total: 6

Export

Name	Value
Row	1
User Agent	Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)
Unique URLs	20
Num Events	99
Total Bytes	150,177,522

User Agents Info Events Chart

Timezone: UTC Events: 361 Imported URL Data Lines: 0 Available Logs: January 1, 2017 - January 30, 2017

La pestaña *Referers*

Esta pestaña nos indica los enlaces que han referido a nuestra página (incluyen también los enlaces internos). El guión o la ausencia del identificador de un *referrer* indica que no ha habido enlazado interno/externo, el usuario ha introducido la URL manualmente o mediante un marcador.

Log File Analyser 6.3 (Unlicensed) - Demo

Project Settings Licence Help

Screamingfrog Open + New Import Settings Verification S... January 1, 2017 - Ja... >>

Overview URLs Response Codes User Agents **Referers** Directories IPs Countries Bytes Events Imported URL Data

Export Search Referrer

Row	Referrer	Unique URLs	Num Events	Total Bytes	Average Bytes	Tot +
1	-	12	162	338,000,157	2,086,420	119.8
2	http://www.example.com	6	85	462,442	5,440	163.9
3	http://example.com	5	59	254,267,655	4,309,621	90,13
4		5	55	2,542,991	46,236	901.5

Referers Total: 4

Export

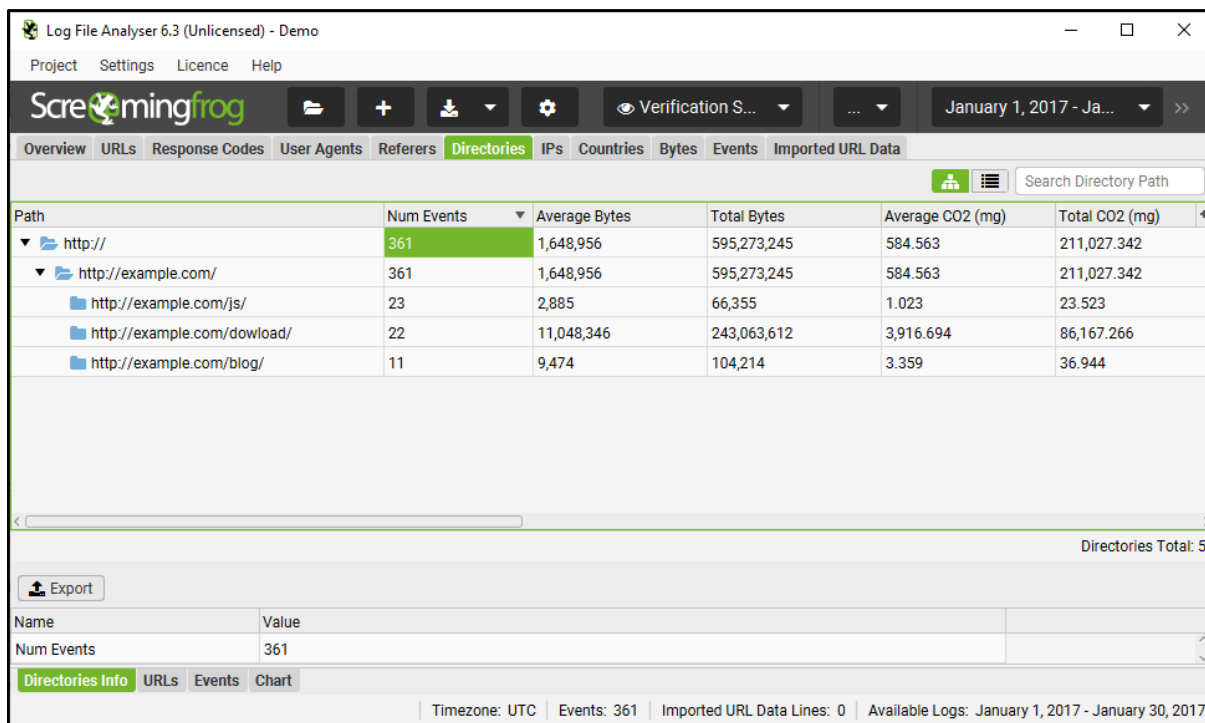
Name	Value
No Referrer selected	

Referrer Info Events Chart

Timezone: UTC Events: 361 Imported URL Data Lines: 0 Available Logs: January 1, 2017 - January 30, 2017

La pestaña *Directories*

Agrupar la información obtenida de manera general por directorios del sitio/aplicación web. Como podemos observar, la página principal es a la que más se ha accedido (<https://example.com>) dado el número de eventos.



The screenshot shows the Screamingfrog Log File Analyser 6.3 (Unlicensed) - Demo interface. The 'Directories' tab is selected, displaying a table of directory paths and their associated statistics. The table has columns for Path, Num Events, Average Bytes, Total Bytes, Average CO2 (mg), and Total CO2 (mg). The data shows that the root path 'http://' has the highest number of events (361) and the highest total bytes (595,273,245). Other paths include 'http://example.com/', 'http://example.com/js/', 'http://example.com/download/', and 'http://example.com/blog/'.

Path	Num Events	Average Bytes	Total Bytes	Average CO2 (mg)	Total CO2 (mg)
http://	361	1,648,956	595,273,245	584.563	211,027.342
http://example.com/	361	1,648,956	595,273,245	584.563	211,027.342
http://example.com/js/	23	2,885	66,355	1.023	23.523
http://example.com/download/	22	11,048,346	243,063,612	3,916.694	86,167.266
http://example.com/blog/	11	9,474	104,214	3.359	36.944

Directories Total: 5

Export

Name Value

Num Events 361

Directories Info URLs Events Chart

Timezone: UTC | Events: 361 | Imported URL Data Lines: 0 | Available Logs: January 1, 2017 - January 30, 2017

La pestaña *IP*

Esta pestaña nos muestra cada una de las direcciones IP encontradas en el registro (como GoogleBot o actividad maliciosa).

Log File Analyser 6.3 (Unlicensed) - Demo

Project Settings Licence Help

Screamingfrog

Verification S...

January 1, 2017 - Ja...

Overview URLs Response Codes User Agents Referers Directories **IPs** Countries Bytes Events Imported URL Data

Export Search IP Address

Row	Remote Host	Unique URLs	Num Events	Total Bytes	Average Bytes	Total CO2 (mg)	Average CO2 (mg)	A +
3	207.46.13.00	17	40	91,110,449	1,328,024	21,000.193	471.004	8
4	74.125.145.96	17	46	83,404,728	1,813,146	29,567.393	642.769	8
5	157.55.108.202	18	44	78,262,167	1,778,685	27,744.33	630.553	8
6	207.46.13.35	15	39	105,514,144	2,705,490	37,405.292	959.11	1
7	72.14.192.15	15	38	22,464,968	591,183	7,963.943	209.577	8
8	216.239.32.0	12	18	66,438,701	3,691,038	23,552.852	1,308.492	9

IPs Total: 10

Export

Name Value

No Remote Host selected

IP Info Events Chart

Timezone: UTC Events: 361 Imported URL Data Lines: 0 Available Logs: January 1, 2017 - January 30, 2017

La pestaña *Countries*

Esta pestaña nos da información sobre cada país al que ha accedido a nuestro sitio/aplicación web.

Log File Analyser 6.3 (Unlicensed) - Demo

Project Settings Licence Help

Screamingfrog

Verification S...

January 1, 2017 - Ja...

Overview URLs Response Codes User Agents Referers Directories **IPs** **Countries** Bytes Events Imported URL Data

Row	Country	Num Events	Rank
1	United States o...	361	1

1 91 181 271 361

Save HTML

Total: 1

Export Search URLs

Row	URL	Timestamp	Remote Host
2	http://example.com/user_to_work	Jan 30, 2017, 2:40:32 PM	157.55.112.243
3	http://example.com/sitemap.xml	Jan 30, 2017, 1:33:57 PM	157.55.108.202

Total Events: 361

Events Chart

Timezone: UTC Events: 361 Imported URL Data Lines: 0 Available Logs: January 1, 2017 - January 30, 2017

La pestaña *Bytes*

Nos muestra información sobre el número total de bytes que se han enviado/recibido desde una URL en concreto, junto con la media de bytes enviados/recibidos.

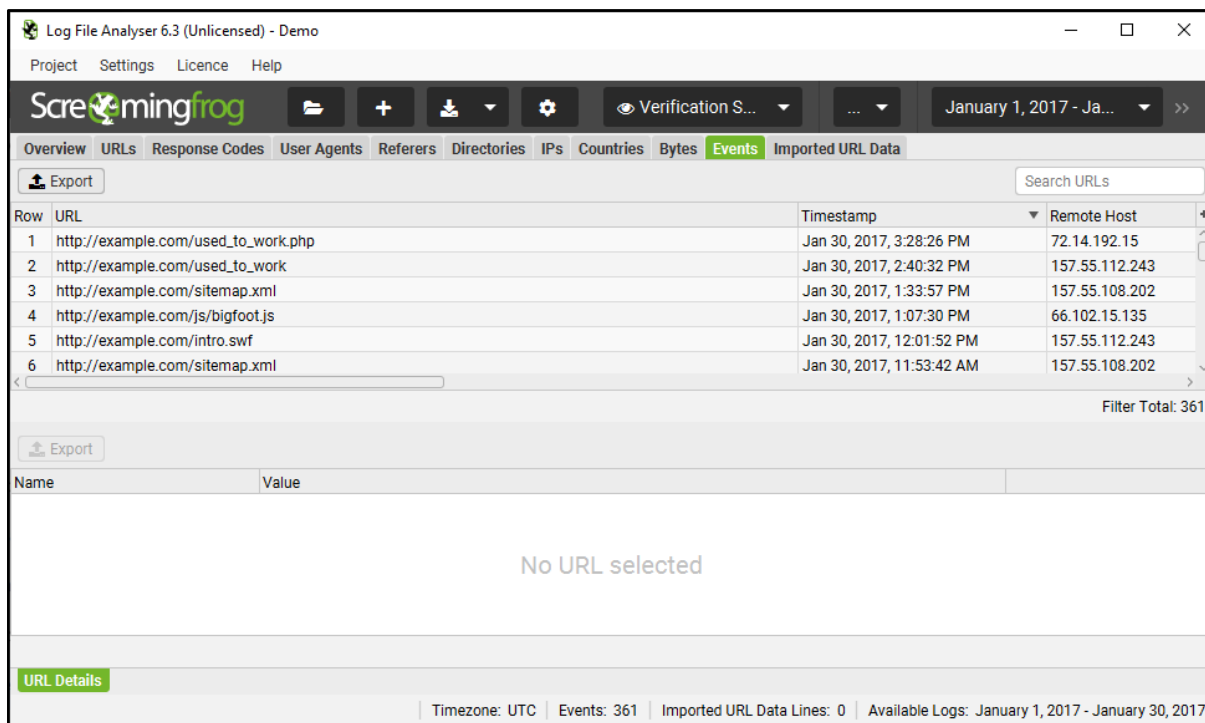
The screenshot displays the 'Bytes' tab in the Screaming Frog Log File Analyser 6.3 (Unlicensed) - Demo. The interface includes a top menu bar with 'Project', 'Settings', 'Licence', and 'Help'. Below this is a toolbar with icons for file operations and a search bar. The main navigation bar shows tabs for 'Overview', 'URLs', 'Response Codes', 'User Agents', 'Referers', 'Directories', 'IPs', 'Countries', 'Bytes' (selected), 'Events', and 'Imported URL Data'. A search bar labeled 'Search URLs' is present. The main content area displays a table with the following data:

Row	URL	Num Events	Total Bytes	Average Bytes
1	http://example.com/	70	386,457	5,520
2	http://example.com/used_to_work.php	27	164,691	6,099
3	http://example.com/used_to_work	26	119,158	4,583
4	http://example.com/slow_page.php	23	254,111,958	11,048,346
5	http://example.com/download/press.pdf	22	243,063,612	11,048,346
6	http://example.com/intro.swf	17	93,029,202	5,472,306

Below the table, there is a 'Filter Total: 23' indicator. An 'Export' button is located at the bottom left of the table area. The bottom section of the interface shows a 'Name' and 'Value' header, followed by a large empty area with the text 'No URL selected'. At the very bottom, there is a status bar with the following information: 'URL Details', 'Events', 'Chart', 'Timezone: UTC', 'Events: 361', 'Imported URL Data Lines: 0', and 'Available Logs: January 1, 2017 - January 30, 2017'.

La pestaña *Events*

Esta pestaña se corresponde con la información encontrada en el archivo de log (todos los eventos que han ocurrido). Hay que tener en cuenta que, si hay 10 millones de eventos en el archivo de registro, los mostrará todos (lo que puede ralentizar el programa).



The screenshot shows the Screamingfrog Log File Analyser 6.3 (Unlicensed) - Demo interface. The 'Events' tab is selected, displaying a table of log entries. The table has columns for Row, URL, Timestamp, and Remote Host. Below the table, there is a section for 'No URL selected' and a 'URL Details' section. The bottom status bar shows 'Timezone: UTC', 'Events: 361', 'Imported URL Data Lines: 0', and 'Available Logs: January 1, 2017 - January 30, 2017'.

Row	URL	Timestamp	Remote Host
1	http://example.com/used_to_work.php	Jan 30, 2017, 3:28:26 PM	72.14.192.15
2	http://example.com/used_to_work	Jan 30, 2017, 2:40:32 PM	157.55.112.243
3	http://example.com/sitemap.xml	Jan 30, 2017, 1:33:57 PM	157.55.108.202
4	http://example.com/js/bigfoot.js	Jan 30, 2017, 1:07:30 PM	66.102.15.135
5	http://example.com/intro.swf	Jan 30, 2017, 12:01:52 PM	157.55.112.243
6	http://example.com/sitemap.xml	Jan 30, 2017, 11:53:42 AM	157.55.108.202

Filter Total: 361

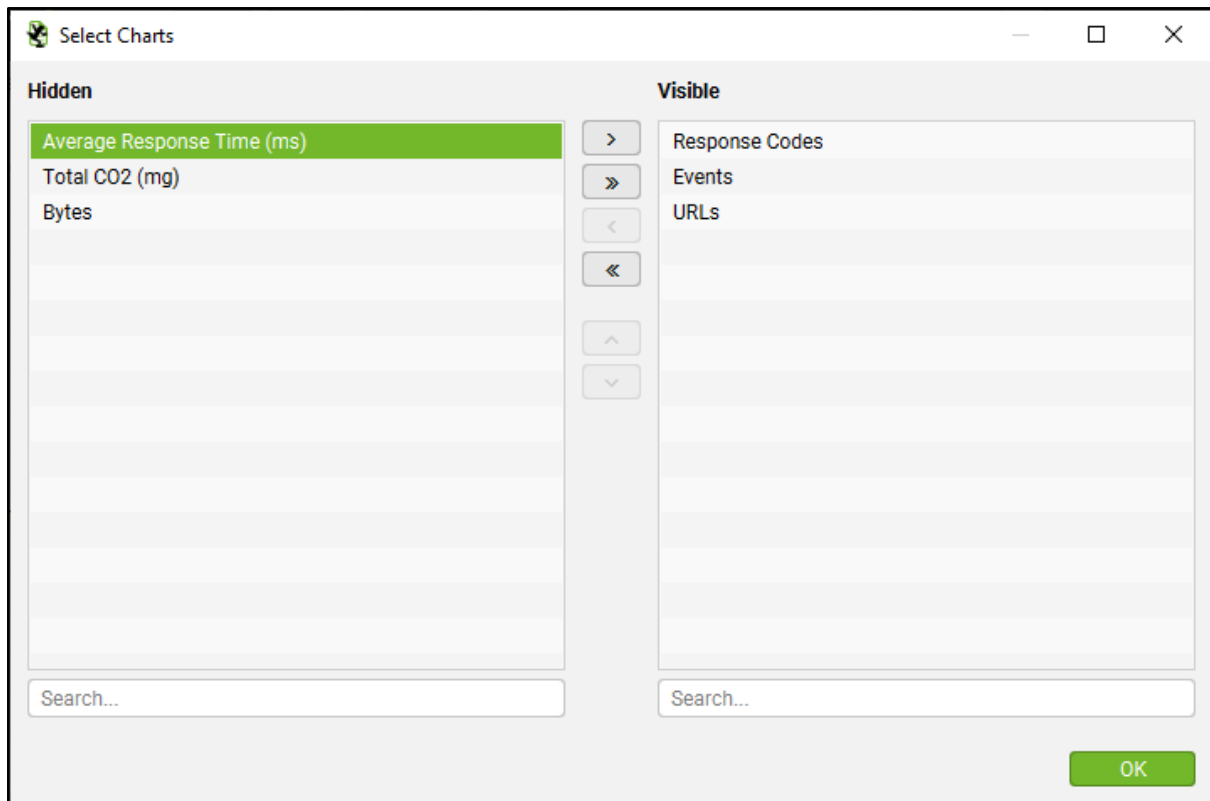
No URL selected

URL Details

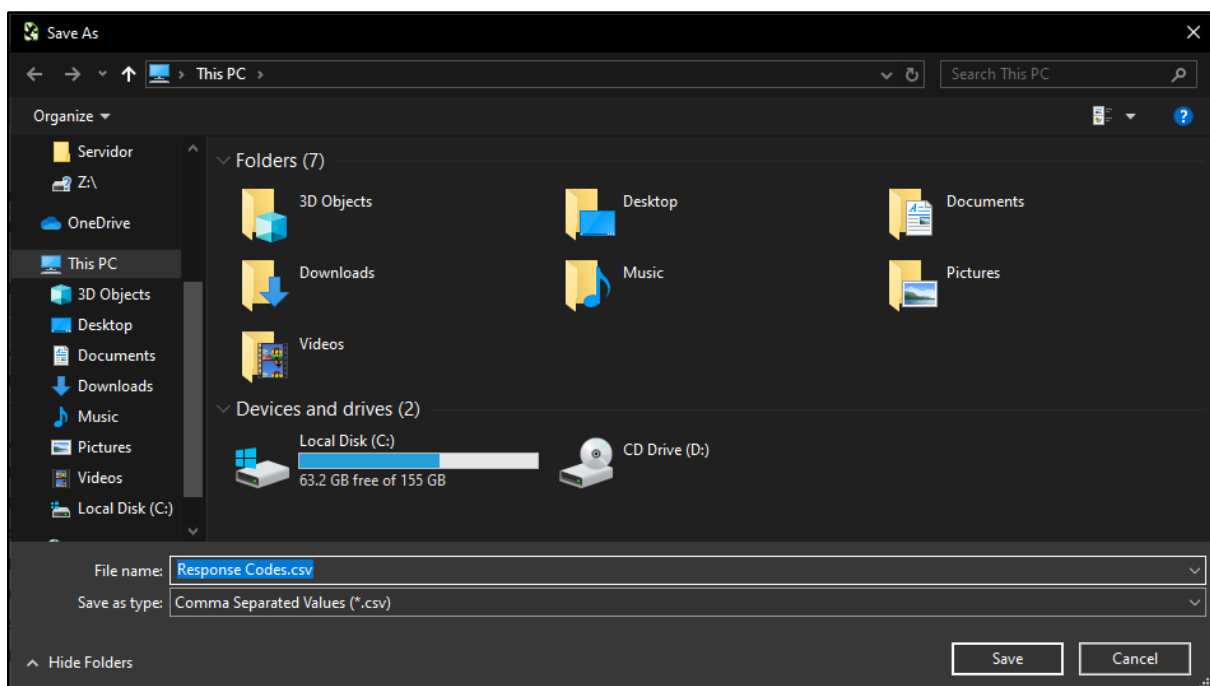
Timezone: UTC | Events: 361 | Imported URL Data Lines: 0 | Available Logs: January 1, 2017 - January 30, 2017

Personalización y exportación

También se puede personalizar la interfaz de usuario de este programa. Por ejemplo, a lo mejor nos interesa mostrar/ocultar un gráfico. Para ello, haremos clic en el botón *Charts* y añadiremos/eliminaremos los que necesitemos.

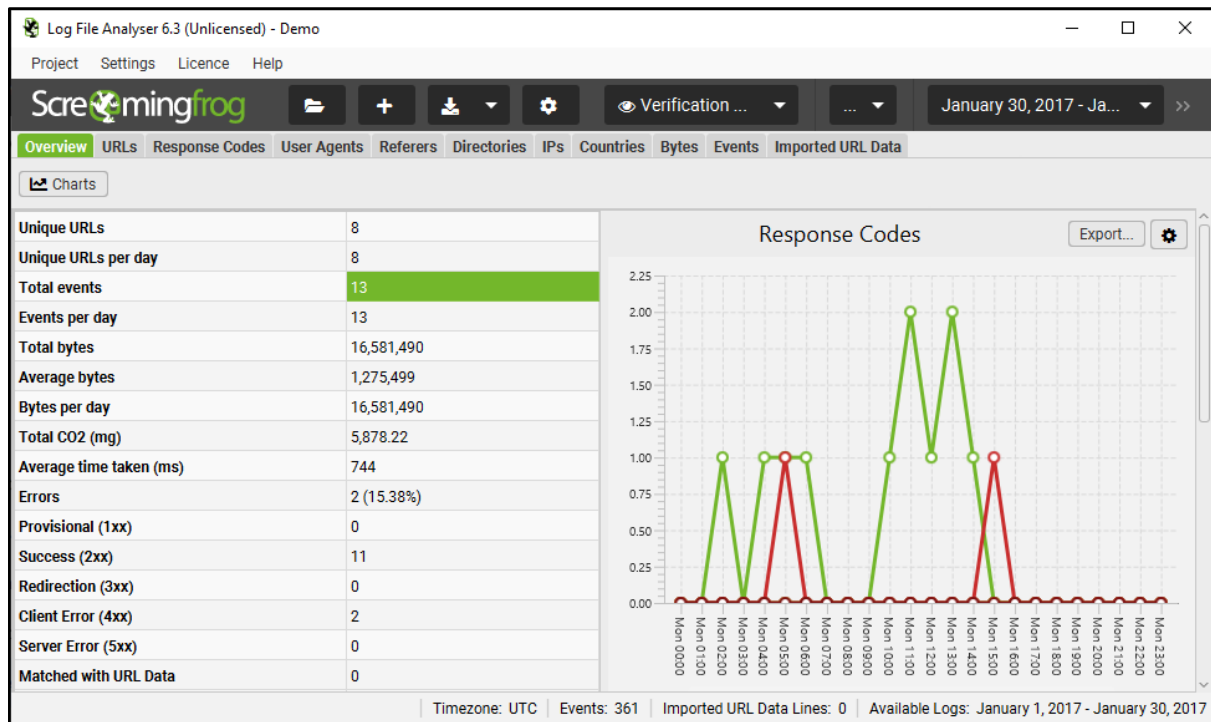


También nos puede interesar exportar determinada información. Exportaremos la información del código de estado de las peticiones, lo que generará un fichero CSV que luego podremos importar a otros programas (como Excel).



Asimismo, también podemos exportar los datos de cada pestaña (excepto la de *Directories*, ya que es más un resumen de lo encontrado en otras pestañas) con el botón *Export*.

Es posible filtrar también por la cantidad de tiempo que queremos ver los datos de registros. Si escojo los de más de un día, me saldrá una cantidad inferior de registros, como se puede observar en la siguiente imagen.



6. Realizando análisis más profundos con logs

Realizando análisis más profundos con logs

Para hacer un análisis más profundo es importante conocer el significado de los campos de un fichero de registro, tal y como hemos visto en el programa anteriormente.

- **Response Code:** el código de estado de la petición a esa URL (200, 404, etcétera).
 - 1XX, 2XX, 3XX, 4XX y 5XX
 - La cantidad de peticiones de código de estado 1XX, 2XX, 3XX, 4XX y 5XX a una URL en concreto. Por ejemplo, si se han hecho 70 peticiones con código 200, en 2XX aparecerá el número 70.
- **Last Response Code:** indica cuándo se devolvió dicha respuesta.
- **Days Since Last Crawled:** indica la última vez que un bot u otro *User Agent* accedió a dicha URL. Un número bajo indica que están siguen accediendo a dicha URL, lo que es bueno para el SEO.
- **Content Type:** indica el tipo de contenido del recurso (`text/html`, `application/pdf`, `application/xml`, etcétera).
- **Total Bytes:** el número de bytes que se han enviado a una URL en concreto desde/hacia el servidor.
- **Average Bytes:** la media de bytes que se han enviado dado el número de peticiones entre el número total de bytes. Por ejemplo, si tenemos 19.124 bytes totales y contamos con 2 eventos, tendríamos una media de 9.562 bytes ($19.124 / 2$)
- **Total CO2:** el número total de emisiones de dióxido de carbono emitidas al acceder a esa URL.
- **Average CO2:** la media de emisiones de dióxido de carbono emitidas según el número de eventos. Por ejemplo, si ha habido 6.78 y existen 2 eventos, la media sería 3.39 ($6.78 / 2$).
- **Num Events:** número de eventos generado por dicha URL.
- **Remote Host:** dirección IP del equipo de donde se generó el evento.
- **Inconsistent:** se establece a `true` si los códigos de estado generados por los eventos son diferentes (uno 200 y otro 404); de lo contrario (si ambos son 200), se establece a `false`.

7. Cruzando otras fuentes de Datos con Logs

Cruzando otras fuentes de Datos con Logs:

Bases de datos: Los logs a menudo contienen referencias a bases de datos. Al cruzar logs con información de bases de datos, puedes rastrear el origen de los errores en consultas, tiempos de respuesta, o identificar cuellos de botella en el acceso a la base de datos.

Monitoreo de sistemas (Métricas de servidores, CPU, memoria): Los logs de servidores y aplicaciones pueden cruzarse con métricas de rendimiento de hardware y software. Esto es útil para entender si un error o un descenso en el rendimiento está relacionado con problemas de infraestructura (por ejemplo, alta utilización de CPU, memoria insuficiente).

Datos de red (Firewall, tráfico de red): Los logs de firewall, tráfico de red o proxies pueden ayudar a detectar patrones inusuales o posibles intrusiones al correlacionar eventos de acceso, autenticación y tráfico con logs de aplicaciones o sistemas.

Sistemas de monitoreo de aplicaciones (APM): Las herramientas de monitoreo de aplicaciones pueden proporcionar información detallada sobre el rendimiento de la aplicación (por ejemplo, tiempos de respuesta, tasas de error). Al correlacionar estos datos con logs de errores, puedes identificar áreas de la aplicación que requieren optimización.

Datos de usuarios o autenticación (Active Directory, LDAP, etc.): Los registros de inicio de sesión, cambios de contraseña o accesos no autorizados pueden cruzarse con logs de acceso a la red o sistemas. Esto es útil para realizar auditorías de seguridad o rastrear posibles incidentes de acceso no autorizado.

Alertas y eventos de seguridad (SIEM): Sistemas como un SIEM (Security Information and Event Management) permiten correlacionar logs con datos de eventos de seguridad, como intentos de intrusión, accesos sospechosos, y otros incidentes. Esto ayuda a construir un panorama de la seguridad de la infraestructura.

¿Cómo funcionan?

IP:

192.168.22.5 — La dirección IP del cliente que realizó la solicitud.

Fecha y hora:

[18/Dec/2024:11:14:22 +0000] — La fecha y hora en que se realizó la solicitud.

Método de solicitud:

"GET" — El método HTTP utilizado para la solicitud (en este caso, GET).

URL solicitada:

"/" — La URL a la que se accedió (la raíz del servidor en este ejemplo).

Código de estado:

200 — El código de estado HTTP que indica el resultado de la solicitud (200 significa éxito).

Información adicional

977 es el tamaño de la respuesta en bytes.

"-" es el referente (referer) que no está especificado.

La cadena final es el **user agent que identifica el navegador utilizado:**

"Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:132.0) Gecko/20100101 Firefox/132.0"

Este tipo de logs suele provenir de servidores como Apache o Nginx.