Administración de Sistemas - Curso 2023 / 2024

Laboratorio 11.- Pila ELK

Contenido:

Objetivos: Combinar Logstash y Kibana con Elasticsearch para crear soluciones de análisis y visualización de logs.

Administración de Sistemas - Curso 2023 / 2024

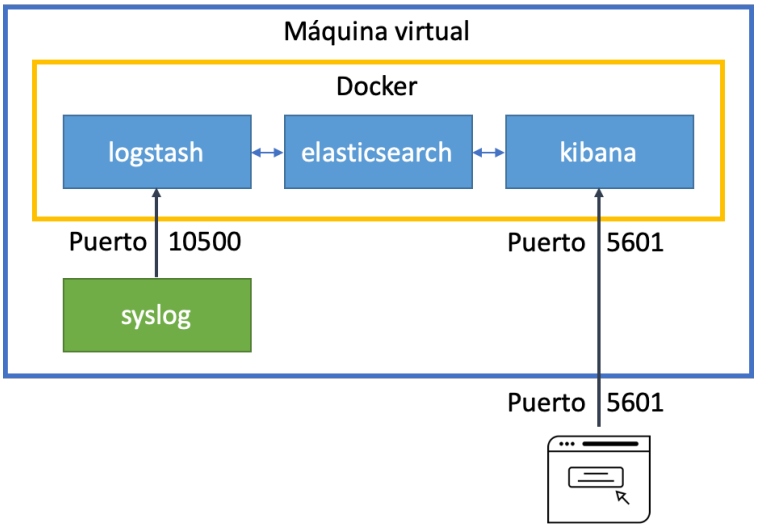
1 Preparar el entorno

Para realizar este laboratorio es necesario tener un despliegue funcional de Logstash, Elasticsearch y Kibana. Se recomienda utilizar Docker Compose. Las instrucciones se encuentran en las diapositivas de teoría.

2 Análisis de logs conectando con Syslog

En esta parte del laboratorio se propone trabajar con la pila ELK como herramienta para recogida y análisis de logs de un sistema Linux. El objetivo es recuperar los logs generados por syslog y visualizarlos en Kibana para explorar su contenido.

El sistema a desplegar se muestra en la siguiente imagen:



Utilizaremos una máquina virtual Ubuntu Server como sistema a monitorizar, que contendrá además Logstash, Elasticsearch y Kibana funcionando con Docker Compose. El proceso syslog de Ubuntu Server enviará los eventos de log a Logstash, que serán procesados y almacenados en Elasticsearch. La información de los logs se consultará desde el Dashboard Kibana utilizando un navegador, para lo cual será necesario configurar reenvíos de puertos.

El primer paso es configurar Logstash. Se deberá configurar una fuente de entrada tipo "conexión TCP" a la escucha en el puerto 10500, un filtro de Grok para parsear cada evento de log y Elasticsearch como destino al que reenviar los datos (a un índice llamado "logs-sistema"). A continuación, se muestra una línea aleatoria de syslog que puede servir como ejemplo para derivar el patrón Grok:

```
Dec  1 11:33:22 as-vm systemd[1]: systemd-tmpfiles-clean.service: Succeeded
```

Una vez que Logstash esté configurado, configurar el despliegue de la pila ELK con Docker Compose (incluyendo los reenvíos de puertos necesarios) y lanzar la pila ELK.

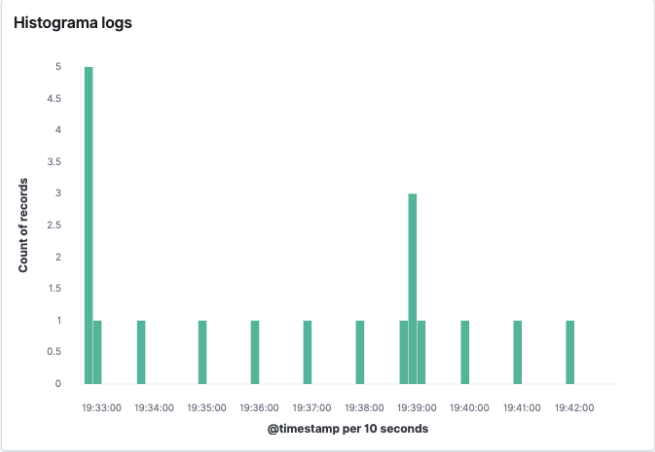
Lo siguiente es configurar syslog para que envíe cada nuevo evento a Logstash utilizando una conexión TCP. Editar el fichero /etc/rsyslog.d/50-default.conf de Ubuntu Server y añadir la siguiente línea:

```
*.*          @@127.0.0.1:10500
```

Esta línea indica a syslog que debe enviar todos los eventos de log a la IP de localhost y al puerto 10500. Los símbolos @@ indican que se debe enviar cada evento como una conexión TCP a la IP y puertos indicados. Si se utilizase un solo símbolo @ en lugar de dos se crearía una conexión UDP.

Una vez modificado el fichero de syslog, reiniciar el servicio de syslog. Realizar alguna operación en la máquina virtual para provocar que se genere un evento del sistema (p.e. enviar un mensaje con *logger*). Utilizando curl, verificar que los datos del nuevo evento se han almacenado correctamente en Elasticsearch.

Una vez que Elasticsearch reciba los eventos de log correctamente, utilizar Kibana Lens para crear un histograma del número de eventos recibidos en los últimos 15 minutos (se muestra un ejemplo en la figura inferior). Verificar que los datos se corresponden con el número de eventos en el log del sistema.



Finalmente, generar más eventos de log y utilizar el modo Discover de Kibana para explorar los datos que se han almacenado en el índice "logs-sistema".

3 Análisis de logs usando Beats

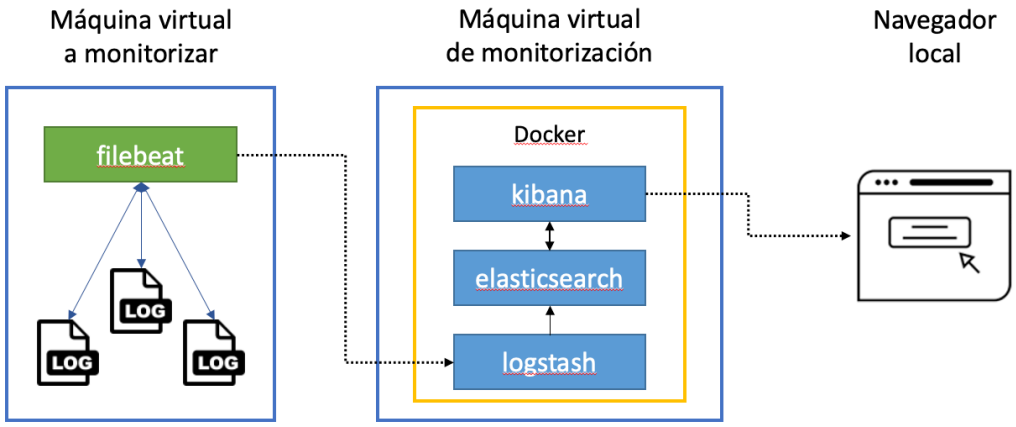
En la segunda parte del laboratorio se propone trabajar con Beats, un conjunto de aplicaciones desarrolladas por Elastic, diseñadas para facilitar la recolección de datos y pensadas para ser integradas con la pila ELK. Las aplicaciones Beats funcionan como sondas que se instalan en los sistemas a monitorizar, que envían datos a una instancia Logstash o Elasticsearch para que sean recopilados de forma centralizada.

Elastic proporciona 7 tipos Beats, cada uno con un propósito diferente[1]. P.e. MetricBeat está diseñado para recopilar métricas de rendimiento (uso de CPU, RAM, ...) o PacketBeat está orientado a monitorizar el estado del tráfico en una red. En este laboratorio se propone trabajar con FileBeat, el Beat dedicado a recopilación y agregación de logs.

[1] What are Beats?, Elastic reference: https://www.elastic.co/guide/en/beats/libbeat/current/beats-reference.html

Administración de Sistemas - Curso 2023 / 2024

En esta parte tendréis que construir un sistema con 2 máquinas virtuales: una de ellas contiene la pila ELK y puede ser la misma que se ha utilizado en el apartado anterior del laboratorio. La segunda máquina virtual será nueva y hará las veces de sistema a monitorizar, el cuál generará logs que serán capturados por Filebeat. Filebeat envíará la información de los logs a la máquina de monitorización para que, desde un navegador, se pueda consultar usando Kibana. Se muestra un esquema del sistema a continuación:



Para desarrollar este sistema, el primer paso es configurar la máquina virtual con la pila ELK para que reciba datos desde la sonda Filebeat. Logstash viene equipado con un tipo "input" llamado "Beats" que permite recoger datos de Beats. En el siguiente ejemplo, se muestra cómo configurar este "input" para que Logstash espere conexiones de Beats en el puerto 5044.

```
input {
  beats {
    port => 5044
  }
}
```

En este punto, se deben realizar 2 configuraciones en la máquina con la pila ELK:

- Configurar Logstash para recibir información de Beats en el puerto 5044, y que esta información se escriba en un índice elasticsearch llamado "logs-filebeat"
- Configurar Docker Compose para redirigir el puerto 5044 del contenedor Logstash al puerto 5044 del anfitrión. Además, abrir el puerto 5044 de la máquina virtual en el Firewall de Google Cloud.

Después, se debe instalar y configurar Filebeat en la máquina a monitorizar. La instalación se puede realizar usando un paquete deb con los siguientes comandos:

```
curl -L -O https://artifacts.elastic.co/downloads/beats/filebeat/filebeat-8.11.1-amd64.deb

sudo dpkg -i filebeat-8.11.1-amd64.deb
```

Una vez instalado, Filebeat se debe configurar para que envíe los datos recopilados a la máquina virtual con la pila ELK. Esta configuración se realiza en el fichero /etc/filebeat/filebeat.yml, en el cual se deben realizar los siguientes cambios:

- Comentar la línea "output.elasticsearch" y las indentadas inmediatamente debajo, para evitar conexiones directas a Elasticsearch.
- Descomentar la línea "output.logstash" y la línea "hosts" que se encuentra de inmediatamente después. Se debe editar la línea hosts para incluir la dirección del servicio Logstash en formato ‹IP›:‹PUERTO›.

El siguiente paso es indicar a Filebeat qué logs debe monitorizar y recopilar. Además de los logs del sistema, que se centralizan en syslog, Filebeat también permite gestionar logs de aplicaciones de 3ºs, como MySQL o MongoDB. Los formatos y aplicaciones de logs que soportan el Filbeat se pueden obtener usando el primero de los siguientes comandos – y podéis obtener más información sobre ellos en la documentación oficial[2].

```
filebeat modules list

filebeat modules enable system
```

En este laboratorio se propone monitorizar los logs del sistema y, para ello, se debe ejecutar el 2º comando del recuadro superior y después editar el fichero /etc/filebeat/modules.d/system.yml. Este fichero configura los logs del sistema que Filebeat debe recoger y enviar a Logstash. En este fichero, se deben cambiar los valores "enabled: false" a "enabled: true" para los apartados syslog y auth.

Con la configuración lista, Filebeat se pone en marcha con el siguiente comando.

```
sudo service filebeat start
```

Si todo ha ido bien tras realizar este último paso, Logstash comenzará a recibir información de Filebeat y el índice "logs-filebeat" de Elasticsearch recopilará documentos que representen las entradas de log del sistema monitorizado. Para verificarlo de forma visual, crear un Data View en Kibana asociado al índice y explorar con el modo Discover los contenidos del índice.

Si esto último no sucede, verificar las configuraciones realizadas anteriormente, paso por paso. En la máquina monitorizada, se puede utilizar el último comando con "status" en lugar de "start" para obtener la información de estado de Filebeat.

Como último paso del laboratorio, generar una nueva entrada de log en la máquina monitorizada usando *logger* y verificar que podéis encontrarla en el índice usando la interfaz Discover de Kibana. También se propone la siguiente pregunta: ¿qué diferencias hay entre monitorizar los logs de un sistema usando Beats frente a realizar una configuración manual de Syslog?

Tras finalizar el laboratorio, se recomienda revisar el saldo disponible en la cuenta Google Cloud y, si procede, eliminar las máquinas virtuales creadas.

[2] Filebeat modules, Elastic reference: https://www.elastic.co/guide/en/beats/filebeat/current/filebeat-modules.html