



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍAS  
INFORMÁTICA Y DE TELECOMUNICACIÓN

Departamento de Teoría de la Señal,  
Telemática y Comunicaciones

PROYECTO FIN DE CARRERA

## **Sensor para recopilación y visualización de información de seguridad en nodos de una red**

Moisés Gautier Gómez  
Director: Gabriel Maciá Fernández

Granada, 17 de julio de 2016



# Sensor para recopilación y visualización de información de seguridad en nodos de una red

Moisés Gautier Gómez

## Palabras clave

Iptables, Seguridad informática, Paquetes, Eventos, Python, Sonda, Django, Web, Visualización

## Resumen

El objetivo principal del proyecto es desarrollar un software que permita recopilar y visualizar la información generada por las aplicaciones de monitorización y control de seguridad que se ejecutan en una máquina.

La motivación del mismo surge fruto de la necesidad de monitorizar un red corporativa a través de un mecanismo de gestión automatizada de eventos (SIEM). Los pasos para la realización de este sistema se han modularizado y dividido en diferentes etapas que se desarrollaran como un todo dentro del proyecto de investigación VERITAS (<http://nesg.ugr.es/veritas/>) del Network Engineering & Security Group (NESG - <http://nesg.ugr.es/>) perteneciente al área de Ingeniería Telemática de la Universidad de Granada.

Para esta finalidad será necesario definir los pasos para la obtención de logs de una fuente de seguridad, configurar la instalación para dicha fuente, realizar un sistema de parseo de logs para extraer la información, almacenarla en un sistema persistente (base de datos) y visualizarla mediante una interfaz web en la sonda desplegada.

Por último, para comprobar la efectividad y analizar el funcionamiento de la solución software, se realizaría una demostración en directo con procesamiento real de eventos para la fuente de seguridad cuyo ámbito tiene este proyecto que será Iptables.



# Sensor for collecting and displaying information security network nodes

Moisés Gautier Gómez

## Keywords

Iptables, Informatic security, Packets, Events, Python, Probe, Django, Web, Visualization

## Abstract

The main objective of the project is to develop a software to collect and display information generated by monitoring applications and security control running on a machine.

The motivation arises because of the need to monitor a corporate network through a mechanism of automated event management (SIEM). The steps for the realization of this system are modularized and divided into different stages to be developed as a whole within the research project VERITAS (<http://nesg.ugr.es/veritas/>) fo Network Engineering & Security Group (NESG - <http://nesg.ugr.es/>) belonging to the area of Telematic Engineering of the University of Granada.

For this purpose it is necessary to define the steps to obtain logs of a security source, configure the installation for that source, perform system parsing logs to extract the information, stored in a persistent system (database) and view via a web interface on the probe deployed.

Finally, to test the efectiveness and analyze the performance of the software solution, a demonstration would be done with actual live event processing for security source project whose scope has to be Iptables.



Yo, *Moisés Gautier Gómez*, alumno de la titulación *Ingeniería en Informática* de la *Escuela Superior de Ingenierías Informática y de Telecomunicación* de la *Universidad de Granada*, con DNI, autorizo la ubicación de la siguiente copia de mi Proyecto Fin de Carrera en la biblioteca del centro, para que pueda ser consultada por las personas que lo deseen.

Fdo: Moisés Gautier Gómez

Granada, 15 de julio de 2016





**D. Gabriel Maciá Fernández**, Profesor del *Departamento de Teoría de la Señal, Telemática y Comunicaciones* de la *Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación* de la *Universidad de Granada*, como director del Proyecto Fin de Carrera de *Moisés Gautier Gómez*

**INFORMO:**

que el presente proyecto titulado “Sensor para recopilación y visualización de información de seguridad en nodos de una red” ha sido realizado por el mencionado alumno bajo mi supervisión, y que autorizo la defensa de dicho proyecto ante el tribunal que corresponda.

Y para que así conste, expido y firmo el presente informe en Granada a 15 de julio de 2016.

Fdo: D. Gabriel Maciá Fernández



# Agradecimientos

Este trabajo de fin de carrera significa la finalización de mis estudios de segundo ciclo, por lo que me gustaría agradecerlo a todas las personas que en menor o mayor medida han hecho posible que me encuentre hoy en esta situación.

En primer lugar me gustaría agradecerle a mi familia su ayuda y apoyo durante estos largos años, y el tesón que han demostrado hacia mí en los momentos más difíciles.

También quiero mencionar a mis compañeros de carrera, sin su inestimable ayuda no habría conseguido lograr mi meta con una sonrisa.

Por último quiero dedicarle este proyecto a todos los interesados en la seguridad informática que tantas horas de entretenimiento pueden darte.



# Licencia

Este documento ha sido liberado bajo Licencia GFDL 1.3 (GNU Free Documentation License). Se incluyen los términos de la licencia en inglés al final del mismo.

Copyright © 2016 Moisés Gautier Gómez. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".



# Índice general

Índice general	I
Índice de figuras	III
<b>1. Introducción</b>	<b>1</b>
1.1. Contexto: Seguridad informática	2
1.2. Motivación y objetivos del proyecto	2
1.3. Estructura de la memoria	3
<b>2. Arte</b>	<b>5</b>
2.1. Estado del arte	5
2.1.1. Lookwise	5
2.1.2. ELK Stack	7
2.1.3. Logstash	7
2.1.4. Elasticsearch	8
2.1.5. Kibana	8
2.1.6. Análisis de la herramienta	9
2.2. SIEM	9
2.2.1. Principales características	10
2.2.2. Análisis	11
2.2.3. Otros SIEM comerciales	11
2.3. Solución tecnológica de éste proyecto	12
2.4. Tecnologías utilizadas	13
2.4.1. Python	13
2.4.2. Procesos vs hilos	14
2.4.3. Django	16
2.4.4. C3js	18
2.4.5. D3js	19
2.4.6. <del>TeX</del> $\text{\LaTeX}$	19
2.4.7. Reactjs	20
2.4.8. SQLite	20
2.4.9. Rsyslog	21
2.4.10. Logrotate	21
2.4.11. Syslog	22
2.4.12. Iptables	22
2.4.13. Django-rest	23
2.4.14. JSON	23
2.4.15. PyCharm	24
2.4.16. Taiga	24

2.4.17. Bitbucket	25
2.4.18. Git	25
2.4.19. Digital Ocean	26
2.4.20. Nginx	26
2.4.21. Dependencias Python	27
<b>3. Requisitos</b>	<b>29</b>
3.1. Especificación	29
3.1.1. Requisitos funcionales	29
3.1.2. Requisitos no funcionales	30
<b>4. Diseño</b>	<b>31</b>
4.1. Diseño de la arquitectura	31
4.2. Diseño del software de cada módulo	32
4.3. Diagramas de Secuencia - Operaciones	41
4.3.1. Flujo de ejecución de la aplicación: BackEnd	42
4.3.2. Flujo de ejecución de la aplicación: FrontEnd	43
4.4. Diseño de la vista	44
<b>5. Implementación</b>	<b>45</b>
5.1. Flujo de ejecución BackEnd	45
5.2. Flujo de ejecución FrontEnd	46
5.3. Configuración de la aplicación	46
5.3.1. Chef	47
5.3.2. Ansible	47
5.3.3. Docker	48
5.4. Configuración local	50
5.4.1. Logs	52
5.4.2. Iptables	52
5.4.3. Parser	55
5.4.4. Workflow	57
5.4.5. Visualización de eventos	58
<b>6. Evaluación</b>	<b>59</b>
<b>7. Planificación temporal</b>	<b>61</b>
7.1. Diagrama de Gantt	61
<b>8. Conclusiones</b>	<b>65</b>
<b>Bibliografía</b>	<b>67</b>
<b>GNU Documentation Free License</b>	<b>73</b>



# Índice de figuras

2.1. Funcionalidades Lookwise . . . . .	6
2.2. ELK Stack [53] . . . . .	7
2.3. Logstash description . . . . .	7
2.4. Diagrama de flujo interno de logstash [53] . . . . .	8
2.5. Sistema SIEM multivariante distribuido - Proyecto VERITAS . . . . .	10
2.6. Ejemplo de D3js . . . . .	19
2.7. Esquema de Rsyslog . . . . .	21
2.8. Configuración de iptables para Logrotate . . . . .	22
2.9. Backlog del proyecto siguiendo un SCRUM . . . . .	25
2.10. Droplet desplegado en digital ocean . . . . .	26
2.11. Configuración de nginx en digital ocean . . . . .	27
4.1. Arquitectura interna del software . . . . .	31
4.2. Arquitectura: Assets . . . . .	32
4.3. Diagrama de clases: Assets . . . . .	33
4.4. Arquitectura: Monitor . . . . .	34
4.5. Diagrama de clases: Assets . . . . .	34
4.6. Arquitectura: Monitor . . . . .	35
4.7. Diagrama de clases para la BD (usando ORM) . . . . .	36
4.8. Ejemplo de clase ORM, en concreto PacketEventsInformation . . . . .	37
4.9. Arquitectura: Visualizaciones . . . . .	38
4.10. Clase Visualization para el paquete ReactJS . . . . .	39
4.11. Paquete Visualizations . . . . .	39
4.12. Paquete Events . . . . .	40
4.13. Paquete Info . . . . .	40
4.14. Diagrama de Secuencia para la parte BackEnd . . . . .	42
4.15. Diagrama de Secuencia para la parte FrontEnd . . . . .	43
5.1. Diagrama de clases BackEnd . . . . .	45
5.2. Flujo de ejecución del FrontEnd - [32] . . . . .	46
5.3. Arquitectura de un sistema bajo Chef . . . . .	47
5.4. Diagrama de la arquitectura Ansible . . . . .	48
5.5. Diagrama de la arquitectura general de Docker . . . . .	49
5.6. Configuración de iptables para Rsyslog . . . . .	50
5.7. Configuración de iptables para Logrotate . . . . .	51
5.8. Configuración de iptables.conf para Rsyslog.d . . . . .	51
5.9. Ejemplo de regla iptables . . . . .	52
5.10. Configuración reglas iptables . . . . .	52
5.11. Evento de ssh localhost en el sistema . . . . .	53

5.12. Log capturado y almacenado por rsyslog en /var/log/iptables.log . . . . .	54
5.13. Procesamiento del log capturado y almacenado en la bd interna de la aplicación .	54
5.14. Instancia de la clase Pygtail y lectura de las líneas del log . . . . .	55
5.15. Permisos de los archivos iptables.log e iptables.log.offset . . . . .	55
5.16. Uso del método split sobre la entrada de líneas de log . . . . .	55
5.17. Obtenemos la tupla key=>value para cada etiqueta del log . . . . .	56
5.18. Vamos asignando cada etiqueta y su valor a su asociado del ORM . . . . .	56
7.1. Panel de actividades - Taiga . . . . .	61
7.2. Diagrama de Gantt 1-05-15 a 3-11-15 . . . . .	62
7.3. Diagrama de Gantt 3-11-15 a 18-07-16 . . . . .	63

# Capítulo 1

## Introducción

*Estar preparado para la guerra  
es uno de los medios más eficaces  
para conservar la paz  
George Washington*

En el siglo XXI todo pasa por ser digital y si no, ya es que lo era antes de llegar a este punto. Quizás muchas de las tecnologías que hoy conocemos se basen en un sistema informatizado, ya sea en su formato de software o sistema embebido. Y es que todo pasa por ser una herramienta software diseñada para un propósito en concreto: un mecanismo de apertura de puertas mediante tarjetas RFID, procedimientos industriales o aplicados a alguna infraestructura crítica o de bien común, una aplicación del tiempo en tú terminal móvil, el propio sistema operativo con el que se puede éste el documento, etc.

Hay un sinnúmero de aplicaciones software que hacen nuestro día a día más llevadero y más fácil. Pero hay un punto que no todos conocen y es la necesidad de saber como se ha creado ése producto o como funciona realmente por su interior. En ése interior, a veces, podemos encontrar cosas que no estaban predestinadas a tener ese comportamiento y debido a ése comportamiento anómalo o imprevisto se generan situaciones de incertidumbre en las que el ser humano debe estar capacitado para afrontarlas. Dichas situaciones se suelen conocer con el término anglosajón de “bug” y sobre los bugs hay una especial categoría que se denominan fallas de seguridad o critical/several bugs.

Estas situaciones no previstas provocan que nuestro sistema, sea el que fuese, actúe de forma inesperada ante un input de información permitido o legítimo permitiendo un uso inadecuado de los recursos a los que se acceden mediante la aplicación. De este concepto o problema, surge en gran medida, el término de seguridad informática el cuál intenta abarcar y dar solución a estos problemas que pueden ir desde un simple fallo de desarrollo, a un fallo crítico que comprometa la seguridad o confidencialidad de los documentos de una empresa o gobierno.

Debido a esta problemática, surge la necesidad de analizar, monitorizar y generar sistemas de seguridad perimetral que permita a las empresas ver que tipo de tráfico interno se genera, que tipo de tráfico externo tiene y cómo se hace uso de él (navegación hacia el exterior, tuneles VPN, conexiones remotas a dispositivos, etc). Así pues, se podría decir que para obtener este tipo de eventos sobre protocolos, tráfico, DNS, IPs, VPNs,.. se tienen que configurar dispositivos de seguridad para la recolección de estos tipos de inputs o fuentes.

Una de las fuentes más conocidas dentro del mundo de la informática es el firewall, pero no así de las de un uso más extendido dentro del mundo doméstico sino el comercial o corporativo. Y dentro de los muchos tipos de software enfocados a tráfico (firewall) se encuentra el paquete de las distribuciones GNU/Linux: Iptables (software fruto del proyecto Netfilter para el kernel de GNU/Linux). Con esta herramienta se pueden definir políticas de filtrado de tráfico para cualquier tipo de protocolo TCP/UDP que queramos limitar entre el exterior y nuestra máquina y viceversa. Además, estas políticas nos permiten derivar dicho tráfico a archivos que podemos manipular obteniendo así los eventos que representan al tráfico generado por una máquina conectada a una red, que posteriormente podemos manipular para generar estadísticas o tipos de uso para una red.

## 1.1. Contexto: Seguridad informática

La seguridad informática es el proceso de mantener un aceptable nivel de percepción frente a un riesgo. Así pues ninguna organización se puede considerar “segura” en cualquier momento, más allá de la última comprobación que se realizó dentro de su política de seguridad.

Un proceso seguro se encuadra dentro de las siguientes 4 etapas: Evaluación, Prevención, Detección y Respuesta:

- **Evaluación:** es la preparación para las otras 3 etapas. Se considera cómo una acción separada porque se relaciona con las políticas, procedimientos, leyes, reglamentos, presupuestos y otras funciones de gestión, además de la propia evaluación técnica enfocada a la seguridad. No tener en consideración estos supuestos anteriores, podría dañar las etapas del diseño.
- **Prevención:** es la aplicación de contramedidas para reducir la probabilidad de tener una situación comprometida.
- **Detección:** es el proceso de identificación de intrusiones. Una intrusión se puede considerar como una violación de una política de seguridad o como un incidente de seguridad a nivel de software/dispositivo.
- **Respuesta:** es el proceso de validar los inputs recogidos por la detección para tomar medidas que solucionen las intrusiones. El primer enfoque que debemos realizar consiste en restaurar la funcionalidad dañada y seguir recopilando información para tener claras las evidencias del atacante sobre nuestro sistema y poder así emprender las acciones legales que correspondan.

De estas etapas anteriores haremos incapié en la etapa de detección que es en la que se enmarca el ámbito de este proyecto. Dado que el principal objetivo es facilitar la monitorización de la información de seguridad.

## 1.2. Motivación y objetivos del proyecto

El objetivo principal del proyecto es desarrollar un software que permita recopilar y visualizar la información generada por las aplicaciones de monitorización y control de seguridad que se

ejecutan en una máquina.

La motivación del mismo surge fruto de la necesidad de monitorizar un red corporativa a través de un mecanismo de gestión automatizada de eventos. Los pasos para la realización de este sistema se han modularizado y dividido en diferentes etapas que se acometerán como un todo dentro del proyecto de investigación VERITAS (<http://nesg.ugr.es/veritas/>) del Network Engineering & Security Group (NESG - <http://nesg.ugr.es/>) perteneciente al área de Ingeniería Telemática de la Universidad de Granada.

## 1.3. Estructura de la memoria

En cuanto a la estructura de esta memoria del proyecto final de carrera, tras éste capítulo dónde se presentan los objetivos y la visión en general del proyecto, se expone el estado del arte y el análisis de requisitos previos al desarrollo software.

En el capítulo siguiente, veremos la etapa del diseño de software así cómo posterior evaluación del mismo.

Finalmente, se presentan las conclusiones generales obtenidas una vez realizado el proyecto, así también la planificación del mismo y estimación de costes.

Además, se presentan las referencias bibliográficas dónde se incluyen las fuentes consultadas para la elaboración de este proyecto, un resumen que engloba las generalidades fundamentales de la aplicación, una guía de utilización (manual de usuario), una guía de instalación, un compendio del software utilizado para el desarrollo y otro de los lenguajes de programación, y finalmente, la licencia completa del documento.



# Capítulo 2

## Estado del arte y tecnologías utilizadas

### 2.1. Estado del arte de herramientas para monitorizar información

En primer lugar, se realiza un estado del arte de herramientas que pudieran realizar funcionalidades parecidas a las que se van a implementar en este proyecto.

Actualmente existen diversos tipos de herramientas que realizan tareas de recolección, filtrado y gestión de eventos dentro un sistema. Las que se han podido analizar y comprobar han sido las siguientes:

#### 2.1.1. Lookwise



Lookwise es una herramienta corporativa que hace las funciones de SIEM en materia de gestión de seguridad, Big Data y cumplimiento normativo (ISO/LOPD y PCI DSS).

#### Características

##### Gestión centralizada

- Interfaz gráfica de administración y operación centralizada.
- Creación y distribución de políticas de forma remota.
- Integración de alertas y explotación de resultados.
- Cuadro de mandos de seguridad.
- Visibilidad en base a roles y permisos.
- Integración con sistemas SIEM.

##### Comunicaciones

- Autenticadas y cifradas

- Ininterrumpidas
- Comprimidas

Plataformas soportadas

- Familia Windows XP (Windows Kernel 5)
- Familia Windows 7 (Windows Kernel 6)

Arquitectura

- Arquitectura Distribuida.
- Arquitectura Modular.
- Flexible y escalable.
- Balanceo de carga.
- Despliegue remoto de funcionalidades y actualizaciones.

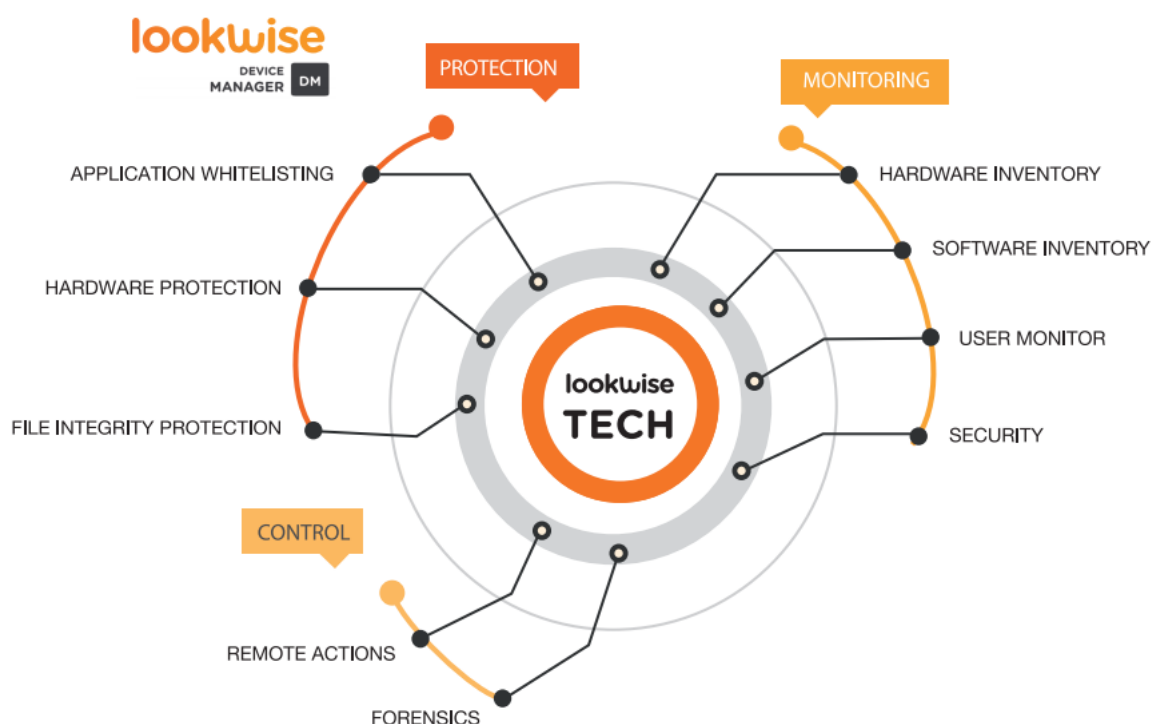


Figura 2.1: Funcionalidades Lookwise

### Análisis de la herramienta

La herramienta de análisis de incidencias y vulnerabilidades hace la mismas funcionalidades de un SIEM pero con una capa más enfocada al cumplimiento del estándar de seguridad de datos de cuentas bancarias - PCI DSS (Payment Card Industry Data Security Standard -



<https://es.pcisecuritystandards.org/minisite/en/>) y de infraestructuras críticas. Gestiona eventos del sistema y los correla con alertas predefinidas internamente o que se hayan incluido cómo especificación del cliente. Detecta fallos en el Active Directory y nos genera un sistema de informe con las incidencias más graves de cara a la parte de consultoría de incidentes por parte el equipo interno.

### 2.1.2. ELK Stack

La pila ELK se basa en una solución open-source de tres productos bien diferenciados que se relacionan entre sí (en resumidas cuentas se trata de un SEM) de la siguiente manera:

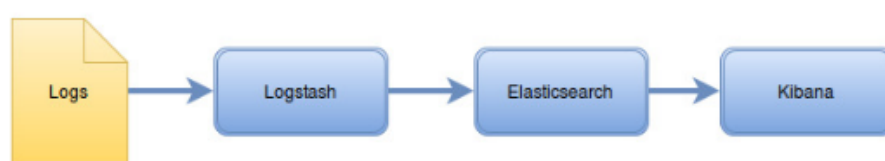


Figura 2.2: ELK Stack [53]

- Logstash para el registro centralizado de logs y su posterior normalización y enriquecimiento de datos.
- Elasticsearch para la búsqueda de datos y análisis en profundidad de un sistema.
- Kibana cómo herramienta de visualización de los datos recolectados y procesados anteriormente según las especificaciones que queramos para el filtrado.

### 2.1.3. Logstash

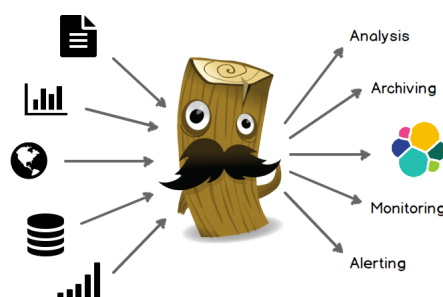


Figura 2.3: Logstash description

Logstash es un motor open-source de recopilación de datos con capacidad de multihebrado en tiempo real. Esta herramienta puede unificar dinámicamente datos de diferentes fuentes y normalizar dichos datos para los outputs de nuestra elección. Además nos permite filtrar y

discretizar todos los datos recolectados para obtener información analítica que pueda ser visualizada.

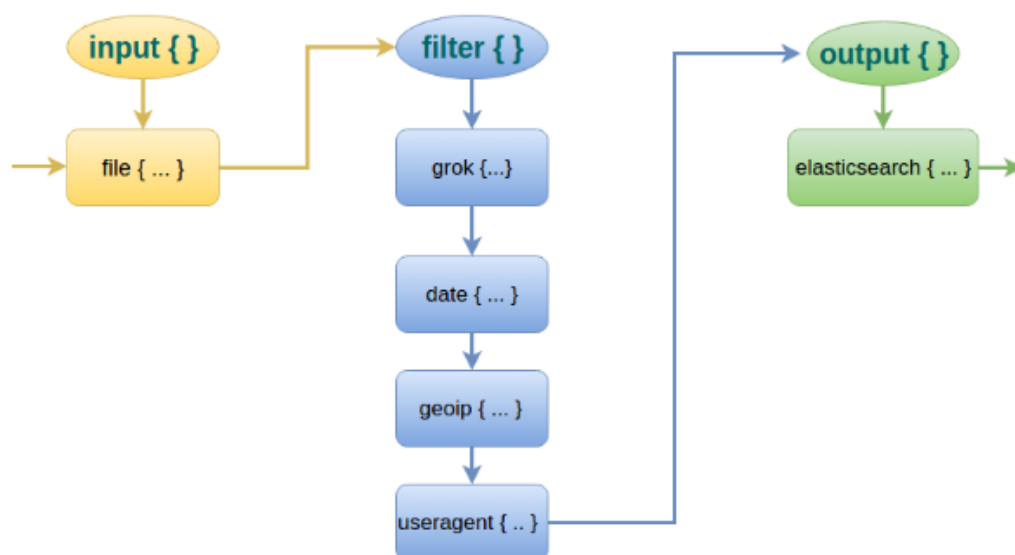


Figura 2.4: Diagrama de flujo interno de logstash [53]

Con logstash, cualquier tipo de evento puede ser enriquecido y transformado con un amplio repertorio de inputs/filtros y los datos de salida, se pueden modificar dependiendo del tipo de software sobre el que queramos introducir esos datos (plugins y códecs).

#### 2.1.4. Elasticsearch



Elasticsearch es un motor open-source de búsqueda y análisis de información de gran escalabilidad. Esta herramienta permite almacenar, buscar y analizar grandes volúmenes de datos de forma rápida y en tiempo real. Se suele utilizar como motor/tecnología subyacente de otras aplicaciones (wrappers), permitiendo así realizar funciones de búsqueda complejas de una manera más ágil.

#### 2.1.5. Kibana



Kibana es una plataforma open-source de análisis y visualización de datos diseñada para trabajar con Elasticsearch. Kibana se utiliza para buscar, ver e interactuar con datos almacenados en

los índices o base de datos de Elasticsearch. De esta forma se puede realizar fácilmente un análisis avanzado de datos que permita visualizarlo en una gran variedad de gráficas, tablas y mapas.

Kibana hace que sea fácil de entender y procesar grandes volúmenes de datos. Mediante su interfaz basada en un cliente web, nos permite crear de forma simple y ágil filtros sobre los datos extraídos mediante consultas a la bd de Elasticsearch en tiempo real.

### 2.1.6. Análisis de la herramienta

La pila ELK, cómo bien se ha comentado en los tres puntos anteriores nos permite recolectar, procesar y correlar cualquier tipo de logs que genere nuestro sistema de una manera visual, ágil y fácil de entender.

El motor de indexación de datos, Elasticsearch, nos permite obtener una ejecución en tiempo real de los logs del sistema así cómo poder escalar dicho volumen de datos dependiendo de la situación. El único inconveniente de esta herramienta es que el sistema de referenciación de documentos internos es mediante json. Es un formato muy versátil pero incapaz de tener funcionalidad por si sola.

Después tenemos Logstash que es el encargado de hacer de middleware entre elastic y kibana, es decir, la parte de la recogida de muestras/eventos y la parte dónde se visualizan esas muestras. Logstash hace de filtro y de motor de normalización de datos entre los diferentes activos que tiene asignados para así poder tener todos los datos unificados según la especificación que queramos dar a cada uno de ellos.

Cada parte de la pila esta desarrollada en su propio lenguaje de desarrollo, siendo Java (o Groovy) el lenguaje de desarrollo de elastic, Ruby el de logstash y Javascript para Kibana. El único inconveniente que se ha podido observar es que la solución ELK está diseñada para trabajar más optimizada en entornos de cloud computing y no de manera local en un servidor. Dado que tendría que usar recursos propios de la máquina y conforme se vayan escalando nuevos recursos estos irán aumentando los de la máquina. Si hay limitación de hardware por los mismos puede llegarse a experimentar un cuello de botella entre elastic y kibana, siendo la carga de datos muy lenta y con tiempos de refresco bastante altos.

## 2.2. Sistemas de gestión de información y eventos de seguridad: SIEM

Un sistema de gestión de información y eventos de seguridad, es un concepto que se suele usar para referirse a dos tecnologías que unifican el concepto como son SEM y SIM.

SEM (Security Event Management), tiene como principal funcionalidad la de recolectar, procesar y almacenar los eventos de seguridad que ocurren en una máquina; y SIM (Security Information Management) se suele usar para la correlación de eventos y obtención de una normalización de los mismos para un posterior análisis, procesamiento mediante algún tipo de técnica estadística o matemática, generado de informe sobre los datos que se han obtenido en

la etapa del SEM, etc.

Así pues, el término SIEM haría referencia a la recopilación de eventos de seguridad y otros relacionados para su documentación y análisis. La mayoría de estos sistemas trabajan mediante la implementación de varios agentes en una tipología jerárquica en la que cada nivel de la jerarquía cumple con un cometido a menor escala dentro del sistema de seguridad definido (máquinas de usuarios, servidores, equipos en red, firewalls, antivirus, etc). Los agentes reenvían dichos eventos a una consola de gestión centralizada, que realiza inspecciones sobre los mismos y asigna la criticidad correspondiente.

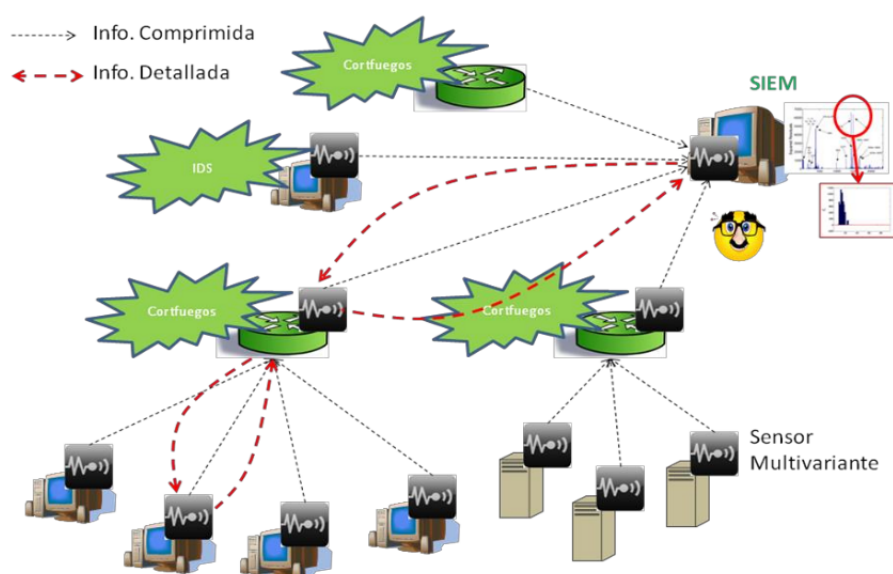


Figura 2.5: Sistema SIEM multivariante distribuido - Proyecto VERITAS

### 2.2.1. Principales características

Estos son los puntos que principalmente gestiona un SIEM:

- Gestión de parches (actualizaciones) del kernel o software de terceros como adobe, java, etc.
- Antivirus en máquinas de usuarios o en servidor.
- Gestión de cortafuegos.
- Integración con Active Directory (LDAP).
- Sistema de prevención de intrusiones (en red: NIPS / basado en hosts: HIPS)
- Proxy / Filtro de contenidos
- Email: anti-spam / anti-phishing

- Análisis de vulnerabilidades
- Herramientas de seguridad opcionales:
  - ◊ IPS para redes Wifi.
  - ◊ Control de firewall web.
  - ◊ Aplicaciones que monitoricen bases de datos.
  - ◊ Prevención de pérdida de datos.
  - ◊ Gestión de riesgos y herramienta de cumplimiento de políticas

### 2.2.2. Análisis

Aunque esta herramienta sea como un conglomerado de aplicaciones o herramientas de detección y análisis, su puesta en marcha no es así tan fácil como cabría esperarse dado que cada despliegue requiere de unos activos o fuentes diferentes. Además, cada solución final requerirá de unas especificaciones distintas, con lo que su implantación depende en gran medida de la facilidad de adaptación al entorno y también de que los responsables de dicha herramienta tengan un profundo conocimiento sobre ella.

Una de las palabras más comunes que definen la implementación de un SIEM es: desalentador. A menudo termina costando más de lo previsto, requiere una experiencia sobre la herramienta que a menudo suele ser externalizada (sobre el propio fabricante) y puede llevar un tiempo considerable antes de obtener resultados tangibles.

Los motivos para los que generalmente se introducen éste sistema en la red corporativa suelen ser varios, pero entre los más destacados suelen ser: el cumplimiento de una normativa industrial o de un gobierno, la gestión de incidentes recurrentes de seguridad o también que en la licitación de un contrato esté contemplado la puesta en marcha de este sistema para una mayor calidad del servicio prestado por un tercero o por la propia entidad. Y aún así, que la propia empresa ya gestione sus eventos internos o los monitorice, no implica que su migración al SIEM sea inmediata.

Además, la licitación o adquisición de un producto de estas características supone que el entorno sobre el que se va a aplicar contiene dispositivos de seguridad que monitorizar (sino se está realizando dicho control), que se dispongan de herramienta de centralización de datos (físicamente o en cloud) y que se quiera un sistema de gestión 24/7, incluido la monitorización fuera de horario laboral.

Son estas razones por las que a veces un sistema tan complejo y gigantesco puede que genere un sobre coste o cubra pocas áreas de la red en las que no se tiene necesidad de vigilar. Siendo esto un inconveniente finalmente, dado que hay otras herramientas (de las que se nutre el SIEM) que ya cumplen con dicha funcionalidad a coste de tener un sistema muy pesado que gestionar.

### 2.2.3. Otros SIEM comerciales

Existen otras muchas más herramientas, comerciales en su gran mayoría, que tratan de abordar el concepto de SIEM a un paso más allá de lo anteriormente analizado. A continuación se van a citar algunas de esas herramientas:

- **OSSIM** - Alienvault: Herramienta SIEM de código abierto.
- **HPE Arsight SIEM** - HP: Herramienta SIEM privativa.
- **IBM Security QRadar SIEM** - IBM
- **Splunk**
- **LogRhythm**

## 2.3. Solución tecnológica de éste proyecto

La principal diferencia que existe entre éste proyecto final de carrera y las herramientas anteriormente analizadas es que el desarrollo y gestión de la información no se hace de forma analítica sino que se gestionan cantidades de datos (en la fase de recolección de logs de dispositivos) y estos no reciben ningún tipo de filtro o procesamiento previo. Se transportan desde el dispositivo monitorizado a otro en dónde se correlan los distintos logs o recolecciones según patrones definidos por los ingenieros de seguridad de la compañía encargada de detectar anomalías en los mismos.

No existe ningún tipo de mecanismo analítico de esta información dado que el punto de inflexión o la inteligencia, está en las etapas finales y no en las previas del tratamiento de la información. Para dotar de una mayor funcionalidad a estos sistemas hay extensiones o módulos que desarrollan las compañías que amplían la capacidad analítica de la misma por una cantidad económica bastante alta y su uso suele estar supeditado a licencia.

Además, la visualización de los eventos almacenados no siempre es inmediata y necesita de otra aplicación externa o módulo (funcionalidad extra) que explote estos recursos y los haga fácil para la interacción humana.

## 2.4. Tecnologías utilizadas

A continuación se detallan las diferentes tecnologías/bibliotecas/lenguajes que se han empleado para la elaboración del proyecto y por qué se han escogido por encima de otras posibles soluciones.

### 2.4.1. Python



Web: <https://www.python.org/>

Python es un lenguaje multiparadigma cuya estructura principal está íntegramente basada en objetos con sus diferentes métodos y atributos internos. Además también posee tipado dinámico, recolector/administrador de la memoria interna y un sistema de referenciado interno de atributos.

Las principales características que han favorecido su elección fueron:

- Es un lenguaje que facilita la implementación de una forma muy ágil y dinámica.
- Su sistema de paquetes es muy intuitivo y ligero. Se puede instalar cualquier dependencia descargándola del repositorio de paquetes PyPi en el que se pueden encontrar infinidad de soluciones software dependiendo de lo que necesites. Sino, siempre se puede definir el propio paquete software y subirlo al repositorio.
- Es de licencia similar a la BSD (Python Software Foundation License) compatible con la GNU GPL. Por lo tanto se puede modificar cualquier aspecto del kernel del lenguaje o mejorar cualquier funcionalidad anteriormente definida.
- Permite la fácil portabilidad entre plataformas, ya que sólo requiere de un intérprete que traduzca dicho código fuente a lenguaje máquina por lo que no existe una pesada fase de compilación por parte del sistema.
- Tiene una amplia comunidad de desarrolladores y es muy fácil de aprender en un corto periodo de tiempo para alguien iniciado.
- Es ampliamente empleado en el mundo de la seguridad informática para ser usado en parsers, procesadores de eventos y usos con la web como principal reclamo (Protocolos, paquetes, tráfico, etc).

### Python frente a otras soluciones

En la fase de estudio de tecnologías, se planteó la posibilidad de desarrollar el proyecto con el lenguaje de programación Java pero se descartó por los siguientes aspectos:

- Precisa de unos conocimientos más avanzados sobre el control de procesos e hilos aunque sea más eficiente, también es más pesada la ejecución de los mismos en un sistema concurrente que puede que tenga muchos activos o fuentes que usar.
- Lenguaje fuertemente tipado y muy estructurado.
- Precisa de una compilación previa que pudiera incrementar los costes de empaquetar dicha solución y además solo se puede ejecutar en un entorno dónde exista una JVM o java virtual machine.
- Para adaptar la solución obtenida a un resultado web tendría que hacer uso de un framework o IDE que ayudase a la hora de la gestión y diseño de la interfaz web así como la comunicación. En Python frameworks como Django, facilitan la tarea del despliegue en formato web y lo más cercano que se le parece es el lenguaje Groovy sobre el que se basa el software Elasticsearch.

### 2.4.2. Procesos vs hilos

En el proceso de desarrollo del proyecto hubo varios momentos en los que se valoró y estudió las diferencias entre un desarrollo software basado en procesos, a la hora de cada nuevo input que llegase a la monitorización, frente hilos o threads. A continuación, se hará un breve resumen sobre las diferencias entre ambos:

#### Procesos

Un proceso, se basa en la parte de un programa que se ejecuta en nuestro sistema, es decir, un conjunto de recursos reservados del sistema.

#### Hilos

Un hilo o thread, es similar a un proceso dado que hace uso de unos recursos reservados del sistema. Pero con una gran diferencia, porque si los procesos ocupan diferentes espacios de memoria, los hilos comparten ese espacio entre ellos.

#### Problemas de los hilos

La normal general es que un conjunto de hilos o procesos tiendan a compartir información entre ellos. Por lo que la solución de los hilos a priori parece ser la más adecuada dado que compartir información será mucho más fácil. Sin embargo, la compartición de grandes cantidades de información y haciendo un uso amplio de tareas concurrentes, pueden producir dos situaciones delicadas: el bloqueo mutuo (deadblock) y la condición de carrera (race condition).

#### Hilos del kernel vs hilos del usuario

Diferentes hilos comparten un mismo PID mientras que diferentes procesos, poseen sus propios PID. Sin embargo, esto no sucede a nivel de kernel. Los hilos del kernel tienen sus propios PID, debido a la forma en la que el kernel es ejecutado.



El kernel (para sistemas GNU) en sí mismo no se ejecuta cómo un proceso sino que sus tareas se ejecutan cómo parte de otros procesos. Debido a la gran cantidad de tareas que se ejecutan, en el kernel se realiza una implementación o acción alternativa para operar de forma similar a los procesos (esto es a lo que se le conoce cómo demonios).

### Solución que aplica Python

A la hora de la implementación para comprobar que hilos se están ejecutando a través el hilo padre, se utiliza el método de la clase Thread enumerate.

Lo que se realiza en esa comprobación es que dentro de la pila o lista de thread que se han lanzado haya alguna coincidencia de objetos de la clase Iptables y si la hay ya existe un hilo previo en ejecución que hace las comprobaciones pertinentes.

```
1 for threads in threading.enumerate():
2
3     test = Iptables(
4         args=(1,),
5         source={'T': 'Firewall', 'M': 'iptables', 'P':
6                 '/var/log/iptables.log',
7                 'C': './secapp/kernel/conf/iptables-conf.conf'}
8     )
9     if type(threads) == type(test):
10         exist_thread = True
11
12 if not exist_thread:
13     thread_iptables = Iptables(
14         args=(1,),
15         source={'T': 'Firewall', 'M': 'iptables', 'P':
16                 '/var/log/iptables.log',
17                 'C': './secapp/kernel/conf/iptables-conf.conf'}
18     )
19     thread_iptables.start()
```

### 2.4.3. Django



Web: <https://www.djangoproject.com/>

Django es un framework web de alto nivel para el lenguaje de programación Python. Éste framework permite un despliegue de una aplicación de escritorio al entorno web de una forma sencilla, segura y fácilmente escalable.

Su metodología es MVT - Model View Template. A continuación, se explicarán en que consisten cada una de ellas:

#### Model

Model o Modelo es la parte encargada de la gestión con la base de datos y de abstraer su uso encapsulandola mediante clases que representan a cada una de las instancias de la BD (o tablas).

#### View

View o Vista es la parte encargada de la gestión entre la base de datos y el template. Puede llevar a confusión esta metodología de desarrollo web, ya que existe una similar: MVC. Pero en MVC (Model-View-Controller) el modelo se encarga de la gestión de la BD, la vista de representar los datos y el controlador de administrador o motor de contenidos entre la BD y la vista final de la aplicación web.

Así pues la **Vista** en el framework Django se representa al modelo MVC cómo el controlador.

#### Template

Template o Plantilla se refiere a la parte encargada de representar la información almacenada en BD, procesada y generada mediante la Vista (View). Al contrario de un modelo MVC, en dónde el controlador se encarga de generar las vistas, en este modelo se representan las mismas cómo una plantilla de muchas otras que se van enrutando a diferentes funciones generadas por la vista.

Dicho esto, siempre habra una plantilla Master o padre de la que heredarán todas las hijas que vayamos asociando a nuestra web. Estas heredan una estructura o skeleton en formato html enriquecido con lenguaje propio de Django en formato python. Veamos un ejemplo de plantilla Master y una plantilla que se nutre de éste skeleton:

```
1 <!DOCTYPE html>
2 {% load staticfiles %}
```

```
3 <html lang="en">
4 <head>
5   <meta charset="UTF-8">
6   <meta name="viewport" content="width=device-width,
7     initial-scale=1, maximum-scale=1"/>
8   <title>Security Sensor | {% block title %}{% endblock %}</title>
9
10  <link rel="stylesheet" href="{% static 'css/main.css' %}"
11    type="text/css"/>
12  <link rel="stylesheet" href="{% static
13    'bower_components/c3/c3.css' %}" type="text/css"/>
14
15 </head>
16 <body>
17   {% block content %}
18
19   <script src="{% static 'bower_components/react/react.js'
20     %}"></script>
```

En el fragmento de la plantilla Master incluimos los archivos estáticos del sistema (**load static-files**) y además especificamos dónde iría el cuerpo de las vistas o plantillas hijas que heredarán este skeleton mediante la especificación de un bloque (**block content** y **endblock**). Y ahora veremos un ejemplo de lo que podría ir dentro de éste bloque.

```

1 {% extends 'MasterPages/MasterPage.html' %}
2 {% block title %}Home{% endblock %}
3 {% block content %}
4
5 <div id="content">
6
7 </div>
8 <div id="chart">
9 <svg></svg>
10 </div>
11 <div id="sub-content">
12
13 </div>
14 <div id="events-info">
15
16 </div>
17 {% if latest_source_list %}
18 <ul>
19     {% for logsource in latest_source_list %}
20     <li><a href="{% url 'secapp:events' logsource.id %}">{{
21         logsource.Type }}</a></li>
22     {% endfor %}
23 </ul>
24 {% else %}
25 <p> No sources are available.</p>
26 {% endif %}
27 {% endblock %}

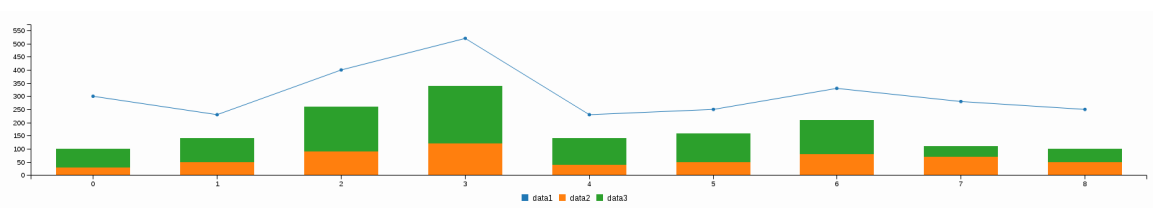
```

Aquí se observa perfectamente cómo se heredan las características de la plantilla Master y se define el bloque de contenido que en la plantilla Master ya se especificó (se sitúa dentro de las etiquetas **block** y **endblock**).

### 2.4.4. C3js

Web: <http://c3js.org/>

Es una biblioteca gráfica basada D3js, que es otra biblioteca visual en Javascript. En este caso, la biblioteca hace uso de la funcionalidad para gráficas de D3js adaptando su motor a otras posibles soluciones de implementación.



### 2.4.5. D3js



Web: <https://d3js.org/>

D3.js es una biblioteca javascript para la manipulación de documentos basado en datos. D3 nos ayuda a la hora de dar vida a los datos usando HTML, SVG y CSS de una forma sencilla y visualmente muy espectacular. Tiene su propio lenguaje que hace uso de las funcionalidades de Javascript por debajo y cuya comunidad es muy amplia en dónde podemos obtener multitud de ejemplos visuales que demuestran la capacidad de generación de gráficas y eventos de la biblioteca.

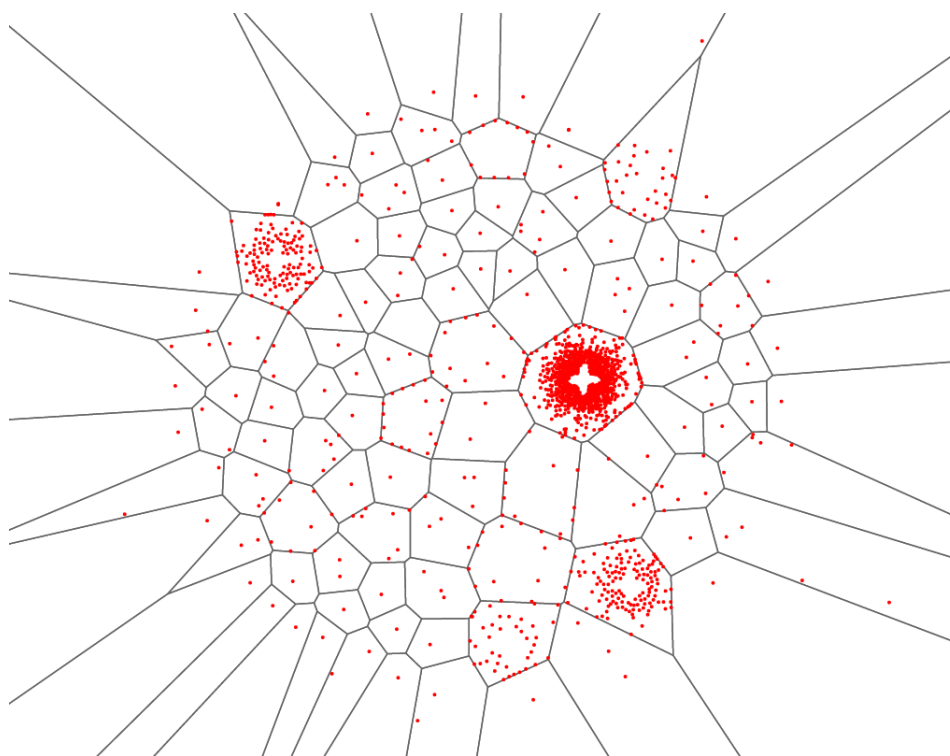


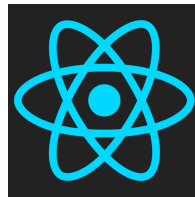
Figura 2.6: Ejemplo de D3js

### 2.4.6. $\text{\LaTeX}$

Web: <https://www.latex-project.org/>

$\text{\LaTeX}$  es un lenguaje de marcado que sirve para la redacción de documentos científicos o técnicos. Con esta herramienta o lenguaje se ha desarrollado la memoria actual del proyecto de final de carrera.

### 2.4.7. Reactjs



Web: <https://facebook.github.io/react/>

React es una biblioteca Javascript que permite construir interfaces de usuario en nuestra aplicación web. Existen soluciones similares a React cómo podría ser JQuery, pero en este caso el proceso de creación de componentes visuales se hace muy intuitiva mediante clases internas que encapsulan el contenido que posteriormente se traducirá a código Javascript nativo.

Para la traducción, React utiliza un componente llamada JSX, pero en nuestra solución se ha usado el paquete Babel (<https://babeljs.io/>) que es similar pero en formato Javascript y no NodeJS.

### 2.4.8. SQLite



Web: <https://www.sqlite.org/>

SQLite es un biblioteca software que implementa un motor para bases de datos SQL. Sus principales características son las siguientes:

- Las transacciones de datos son atómicas, consistentes, aisladas y duraderas (ACID) incluso después de que el sistema tengo un fallo o se apague inesperadamente.
- No necesita de una administración o configuración previa para su normal funcionamiento.
- Implementación de un sistema SQL íntegro con características más avanzadas cómo indexación parcial.
- La base de datos, en su totalidad, se almacena cómo un archivo normal en disco. Esto es útil para ser cargado directamente como un archivo en una aplicación.
- Soporta bases de datos de Terabytes de tamaño y Gigabytes de string o archivos.
- Facilidad de uso mediante su API interna.
- No tiene ninguna dependencia externa.
- Multiplataforma: Android, BSD, iOS, Linux, Mac, Solaris, VxWorks y Windows soportan este tipo de formato de base de datos.
- El código fuente está bajo dominio público para cualquier tipo de uso.

### 2.4.9. Rsyslog

Web: <http://www.rsyslog.com/>

Rsyslog (Rocket-Fast System for Log Processing), es un sistema de recogida de logs de sistemas UNIX (servidos mediante syslogd) que nos permite manipularlos y exportarlos a un formato más adecuado para su procesado.

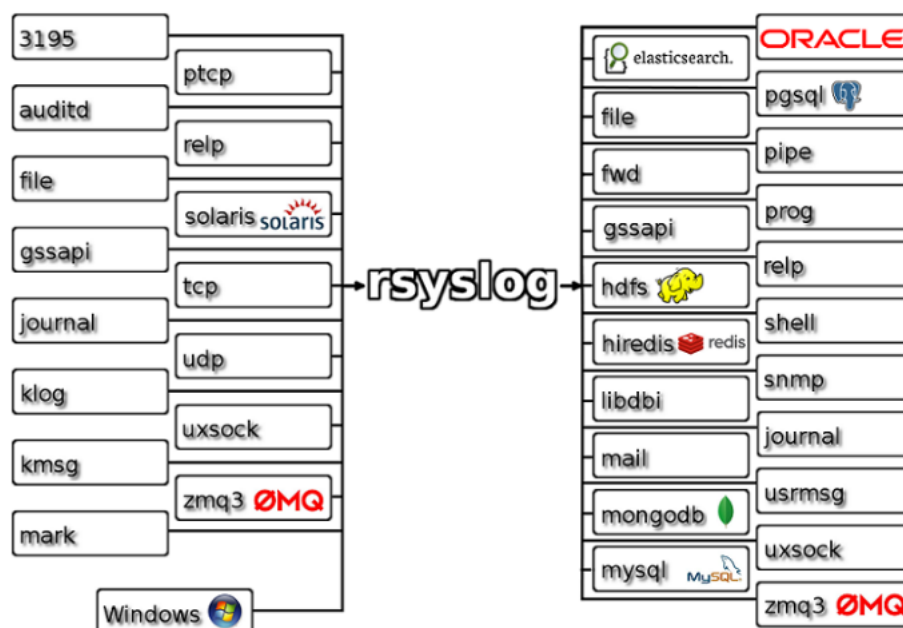


Figura 2.7: Esquema de Rsyslog

- Multihilo
- TCP, SSL, TLS RELP
- Se puede filtrar cualquier parte de los mensajes de syslog
- Se puede configurar el formato de salida de la recolección

### 2.4.10. Logrotate

Web: <https://github.com/logrotate/logrotate>  
 man: <http://linux.die.net/man/8/logrotate>

Es una utilidad de los sistemas UNIX que nos permite simplificar la administración de archivos de logs en un sistema en dónde haya logs de muchos tipos de fuentes. Nos permite automatizar la rotación, compresión, eliminación y envío por email de los archivos de logs del sistema. Logrotate se encuentra normalmente corriendo como un proceso cron diario.

```
1 /var/log/iptables.log
2 {
3     rotate 7
4     daily
5     missingok
6     notifempty
7     delaycompress
8     compress
9     postrotate
10         invoke-rc.d rsyslog restart > /dev/null
11     endscript
12 }
```

Figura 2.8: Configuración de iptables para Logrotate

Ahora se realizará una breve explicación de cada configuración que se ha establecido sobre el archivo `iptables.log`:

- **rotate <count>**: Los archivos de log son rotados una cantidad de <count> veces antes de ser eliminados o enviados por correo. Si <count> es 0, las versiones anteriores son eliminados antes de efectuarse la rotación.
- **daily**: Los archivos de log son rotados diariamente.
- **missingok**: Si el archivo de log no existe, ir al siguiente sin mostrar un mensaje de error.
- **notifempty**: No rotar el log si éste está vacío.
- **delaycompress**: Pospone la compresión de los logs previos para el siguiente ciclo de rotaciones. Esto sólo sucede cuando se combina con la opción **compress**.
- **compress**: Las versiones antiguas de logs son comprimidas con gzip por defecto.
- **postrotate-endscript**: Las líneas que se encuentran entre estas dos palabras en la configuración, son ejecutadas (usando `/bin/sh`) antes de que el archivo de log sea rotado y sólo si el log actual va a ser rotado.

En este caso, el comando que se ejecuta fuerza a rsyslog a reabrir el archivo de log para escribir en él. Se suele usar después que logrotate mueva los archivos de logs antiguos, entonces rsyslog comienza a escribir en los nuevos.

### 2.4.11. Syslog

Web (man): <http://linux.die.net/man/3/syslog>

Sirve para el envío de mensajes al sistema de logs interno. `syslog()` genera un mensaje de log, el cuál se distribuye mediante el demonio `syslogd`.

### 2.4.12. Iptables

Web: <http://www.netfilter.org/projects/iptables/index.html>

Manual: <http://www.netfilter.org/documentation/HOWTO/es/packet-filtering-HOWTO-7>.



html

man: <http://linux.die.net/man/8/iptables>

Herramienta de filtrado de paquetes IPv4/IPv6 y NAT. Iptables es usado para configurar, mantener e inspeccionar tablas de paquetes IP mediante filtros o reglas en el kernel de Linux. Se pueden definir multitud de tablas, en las que cada una pueda contener un número de reglas precargadas o que pueden ser redefinidas por el usuario.

### 2.4.13. Django-rest



Web: <http://www.django-rest-framework.org/>

Django REST es un framework potente y flexible que permite construir Web APIs. Se usa cómo un complemento al framework Django para el uso de API Restfull dentro de la propia aplicación.

### 2.4.14. JSON

Web: <http://www.json.org/>

JSON (Javascript Object Notation) es un formato de intercambio de datos ligero. Se usa para facilitar la legibilidad y escritura para los humanos y además es fácil de interpretar y parsear para una máquina. Está basado cómo un subconjunto de JavaScript y el Standard ECMA-262 3ª Edición.

JSON está construido sobre dos estructuras:

- Una colección de pares nombre/valores. En varios lenguajes, esto se realiza mediante objetos, registros, estructuras, diccionarios, tablas hash, listas de claves o arrays asociativos.
- Una lista de valores ordenados. En la mayoría de lenguajes, esto se realiza mediante un array, un vector, lista o secuencia.

### 2.4.15. PyCharm



Web: <https://www.jetbrains.com/pycharm/>

PyCharm es un IDE que permite el trabajo con aplicaciones Python en sus respectivos entornos virtuales (virtualenv) así cómo la integración con frameworks de desarrollo cómo es el caso de Django. La aplicación en sí fue desarrollada nativamente mediante un editor de texto (Emacs) pero a la hora de realizar un empaquetado e integración con otra herramientas se optó por éste IDE.

### 2.4.16. Taiga



Web: <https://taiga.io/>

Taiga es un gestor de proyectos que nos permite implementar una metodología SCRUM o Kanban sobre nuestro proyecto. Es una herramienta muy intuitiva y colaborativa, que permite definir hitos/tareas/wikis para solucionar cada punto de la fase de desarrollo de un proyecto. Además, permite la integración (WebHooks) con otras plataformas de repositorios de proyectos cómo GitHub, GitLab o BitBucket.

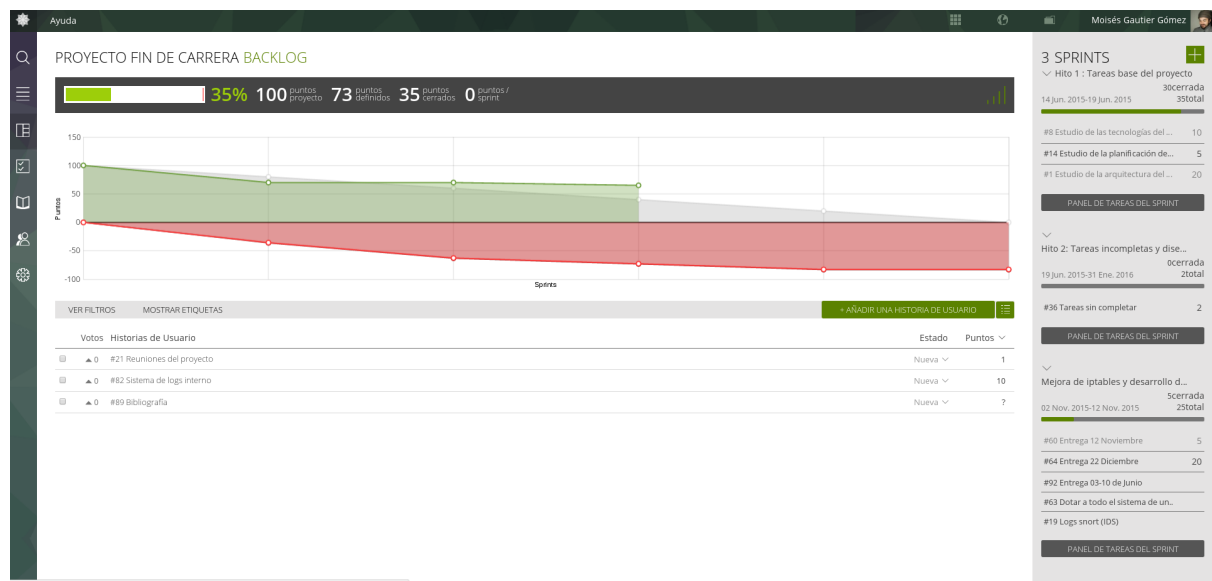


Figura 2.9: Backlog del proyecto siguiendo un SCRUM

### 2.4.17. Bitbucket



Web: <https://bitbucket.org/>

Repositorio: <https://bitbucket.org/MGautierGomez/securityproject>

Gestor de repositorios Git y Mercurial. Se optó por éste gestor para probar su funcionamiento, así como la posibilidad de tener repositorios privados para desarrollar partes del proyecto que necesitasen ser ocultadas. También se encuentra alojado en el repositorio de **GitHub**: <https://github.com/MGautier/security-sensor>

### 2.4.18. Git



Web: <https://git-scm.com/>

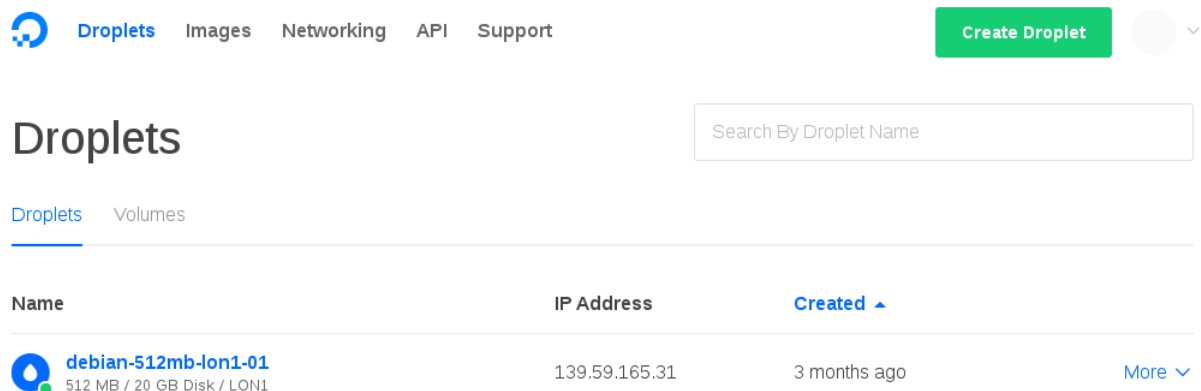
Git es un sistema open-source de control de versiones diseñado para manejar integralmente las fases de desarrollo de proyectos, simples y complejos, con velocidad y eficiencia.

### 2.4.19. Digital Ocean



Web: <https://www.digitalocean.com/>

Servidor web para alojar proyectos en cloud. La ventaja de este servicio de VPS es que te permite desplegar máquinas de cualquier tipo (siempre que sean software libre) de una manera muy fácil y rápida. Además tiene un punto fuerte y es que la información se almacena en discos SSD, con lo que el procesamiento se ve muy mejorado a la hora de computar (en este caso eventos de Iptables).

A screenshot of the DigitalOcean website's 'Droplets' page. At the top, there's a navigation bar with links for 'Droplets', 'Images', 'Networking', 'API', and 'Support'. A green 'Create Droplet' button is on the right. Below the navigation bar, the word 'Droplets' is prominently displayed. A search bar labeled 'Search By Droplet Name' is on the right. Underneath, there's a sub-navigation bar with 'Droplets' and 'Volumes'. A table lists the droplets. The first droplet is 'debian-512mb-lon1-01' with an IP address of '139.59.165.31' and was created '3 months ago'. The table has columns for 'Name', 'IP Address', and 'Created'. There's a 'More' link at the end of the row.

Name	IP Address	Created
debian-512mb-lon1-01 512 MB / 20 GB Disk / LON1	139.59.165.31	3 months ago

Figura 2.10: Droplet desplegado en digital ocean

### 2.4.20. Nginx



Web: <http://nginx.org/>

Características: <http://nginx.org/en/>

Nginx (engine x) es un servidor HTTP y proxy inverso, un servidor proxy de email y un servidor proxy genérico de TCP/UDP. Algunas de sus principales características son las siguientes:

- Sirve archivos estáticos, index y autoindexados.
- Acelera el proceso de proxy inverso con caché: Tolerancia a fallos y carga balanceada de datos.
- Soporta aceleración con cacheo de FastCGI, uwsgi, SCGI y servidores memcached: Tolerancia a fallos y carga balanceada de datos.
- Soporta SSL y TLS SNI.
- Soporta HTTP/2 con dependencia basada en prioridad y balanceo.

Configuración de nginx para el servidor en digital ocean:

```
1 server {
2
3     root /var/www/html;
4
5     # Tipos de archivos index de nuestro sistema
6
7     index index.html index.htm index.nginx-debian.html;
8
9     # Nombre del servidor en local
10
11     server_name 139.59.165.31;
12
13     location /static/ {
14         alias ~/trunk/version-1-0/webapp/secproject/static/;
15         expires 30d;
16     }
17
18     location / {
19         proxy_set_header X-Forwarded-For
20             $proxy_add_x_forwarded_for;
21         proxy_set_header Host $http_host;
22         proxy_redirect off;
23         proxy_pass http://127.0.0.1:8000;
24         proxy_pass_header Server;
25         proxy_set_header X-Real-IP $remote_addr;
26         proxy_connect_timeout 10;
27         proxy_read_timeout 10;
28
29     }
30 }
```

Figura 2.11: Configuración de nginx en digital ocean

### 2.4.21. Dependencias Python

Estas son las dependencias que se necesitan para la fase de desarrollo del proyecto. Para poder hacer uso de ellas, previamente tendremos que tener instalado un entorno virtual o virtualenv desde el cual hacer la instalación de estos paquetes.

### Virtualenv

Es una herramienta que nos permite crear entornos aislados de Python. De esta forma podemos realizar pruebas de dependencias o paquetes para un entorno de desarrollo, en dónde las instalaciones o pruebas no afecten a las dependencias internas del sistema.

- Django 1.9.2: <https://pypi.python.org/pypi/Django/1.9.2>
- argparse 1.4.0: <https://docs.python.org/2.7/library/argparse.html>
- dnspython 1.12.0: <https://pypi.python.org/pypi/dnspython/1.12.0> - Se usa para hacer resolución directa e inversa de DNS.
- gunicorn 19.4.5: <https://pypi.python.org/pypi/gunicorn> - Sirve para desplegar un servidor WSGI (Web Server Gateway Interface) HTTP en entornos Unix.
- optional-django 0.3.0: <https://pypi.python.org/pypi/optional-django/> - Otras utilidades del framework Django
- pygtail 0.6.1: <https://pypi.python.org/pypi/pygtail> - Sirve para leer archivos de log en los cuales haya registros del mismo sin leer. Similar al comando tail -f en sistemas Unix.
- python-dateutil 2.4.2: <https://pypi.python.org/pypi/python-dateutil/2.4.2> - Extensión al módulo principal de Python datetime.
- react 2.0.2: <https://pypi.python.org/pypi/react/2.0.2> - Módulo que ejecuta un servidor de datos para los componentes react de la aplicación en Python.
- requests 2.9.1: <https://pypi.python.org/pypi/requests/> - Biblioteca Python para poder consumir recursos HTTP de forma segura y controlada.
- setuptools 20.1.1: <https://pypi.python.org/pypi/setuptools> - Herramienta para la descarga, compilación, instalación, desinstalación y actualización de todos los paquetes Python. El resultado de la misma se puede usar mediante pip que se nutre el repositorio de paquetes PyPi.
- six 1.10.0: <https://pypi.python.org/pypi/six> - Es una biblioteca de compatibilidad entre Python 2 y 3.
- wheel 0.29.0: <https://pypi.python.org/pypi/wheel> - Es un gestor o generador de paquetes en Python.
- wsgiref 0.1.2: <https://pypi.python.org/pypi/wsgiref> - Es una biblioteca que sirve como soporte de validación para WSGI 1.0.1 para versiones de Python inferiores a la 3.2.
- configparser 3.5.0: <https://docs.python.org/2/library/configparser.html> - Es módulo o clase que implementa un lenguaje muy sencillo para el parseo de archivos de configuración, siguiendo una estructura similar a cómo se describen y procesan los archivos INI en sistemas Microsoft Windows.

# Capítulo 3

## Especificación y análisis de requisitos

Para el caso que nos concierne en el proyecto, dentro del marco de investigación que define la totalidad de la infraestructura, la funcionalidad principal del mismo será:

- Definir los pasos para obtener recolección de logs de una fuente de seguridad. Configuración de rsyslog, logrotate y supervisord (este último en el caso de que sea necesario).
- Una vez configurada la máquina, configurar la instalación para cada fuente en particular. El ámbito del proyecto se enfoca sobre iptables.
- Realizar un sistema de parseo de logs para extraer la información necesaria para cada evento que se registre en el sistema.
- Sistema de gestión de base de datos en dónde se encuentren los datos en crudo recogidos, los procesados y los dispuestos para su visualización.
- Panel de control dónde visualizar la información de ese nodo con total detalle de la información.

### 3.1. Especificación de los requisitos

En esta etapa del modelado de requisitos se captura el propósito general del sistema:

- Se analiza qué debe hacer el sistema.
- Se obtiene una versión contextualizada del sistema.
- Identifica y delimita el sistema.
- Se determinan las características, cualidades y restricciones que debe satisfacer el sistema.

#### 3.1.1. Requisitos funcionales

Los requisitos funcionales que se han obtenido en el sistema son los siguientes:

- Ser una herramienta multiplataforma y que permita a cualquier usuario definir sus propias interfaces de gestión de eventos.

- Dotar de funcionalidad gráfica que permita extraer información en tiempo real con gráficas o mecanismos visuales (en web) del sistema de base de datos que ha procesado los inputs de las fuentes para las que ha sido configurada.
- Dotar de una api interna que nos permita extraer información en tiempo real en un formato uniforme para la web o para que cualquier usuario pueda usar la funcionalidad del proyecto para su propio beneficio usando herramientas generadas en el back-end para otro tipo de aplicaciones.
- Ser parte de un todo, en el que el todo sea un SIEM capaz de obtener información de las diferentes sondas o módulos, que en este caso, sería la solución desarrollada.
- Desarrollar una sonda para procesar eventos logs de firewall.
- Extracción y parseo de eventos de firewall Iptables.

### 3.1.2. Requisitos no funcionales

Los requisitos no funcionales son aquellos que describen cualidades o restricciones del sistema que no se relacionan de forma directa con el comportamiento funcional del mismo. A continuación se especifican los más importantes del sistema:

- No requiere un conocimiento específico del sistema una vez puesto en funcionamiento.
- La aplicación tendrá manual de uso.
- La base de datos estará implementada en un lenguaje objeto relacional como SQLite.
- La aplicación estará realizada en el lenguaje de programación Python.
- La interfaz debe reflejar claramente la distinción entre las distintas partes del sistema.
- La documentación del código fuente será llevada a cabo mediante la aplicación Sphinx.
- El sistema se desplegará sobre una versión GNU Linux Debian 8 Jessie.
- El código fuente de la aplicación seguirá un estilo uniforme y normalizado para todos los módulos del mismo.
- El formato de las fechas será dd/mm/yyyy a excepción de la utiliza para la funcionalidad api/events/day/<source>/yyyy-mm-dd.



# Capítulo 4

## Diseño

A continuación pasaremos a detallar cada una de las partes del diseño del sistema.

### 4.1. Diseño de la arquitectura

La arquitectura sobre la cual se ha basado el proyecto se representa en el siguiente diagrama correspondiente con una sonda o nodo de recolección de información para el proyecto VERITAS:

#### Arquitectura del sensor PCA

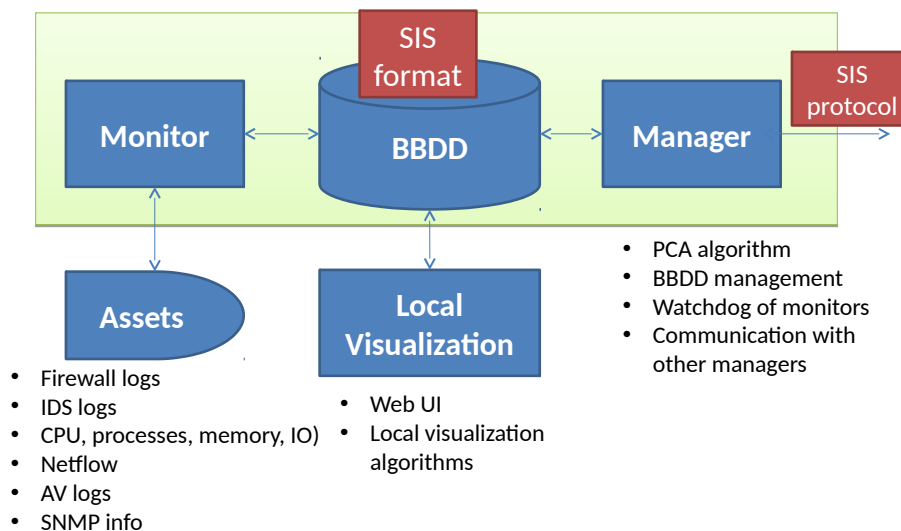


Figura 4.1: Arquitectura interna del software

En el esquema podemos ver toda la arquitectura que tendría una sonda para el proyecto VERITAS, aunque el ámbito de éste proyecto no se engloba en su totalidad, sino en unas partes en concreto del esquema que en el siguiente punto se explicarán en detalle. Puntualizar que la interacción de la sonda con la fuente de seguridad es independiente de la interacción base de

datos con la visualización de los datos. Si bien se hace uso del mismo ORM, proporcionado por el framework Django, estos se ejecutan por separado a la hora de procesar las fuentes.

## 4.2. Diseño del software de cada módulo

### Assets

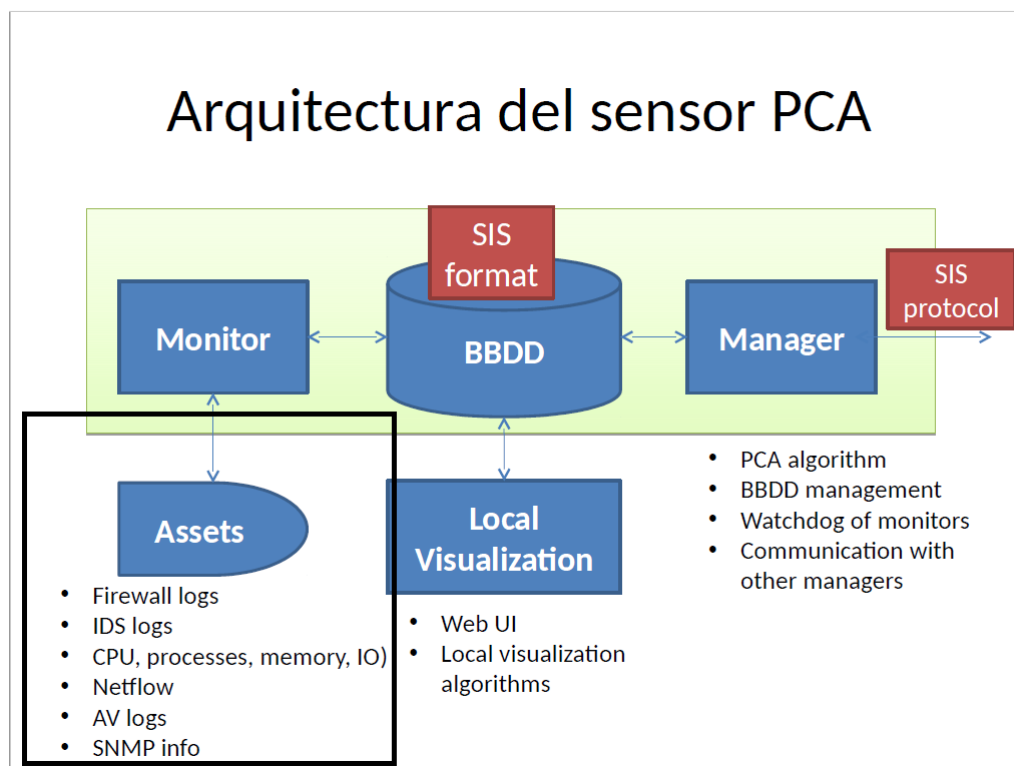


Figura 4.2: Arquitectura: Assets

En esta sección de la arquitectura es dónde se define la parte de las fuentes de seguridad que se van a gestionar desde la sonda, es decir, implementar, recolección y pasar el control final a su clase superior: Monitor. La jerarquía de clases sería de la siguiente manera:

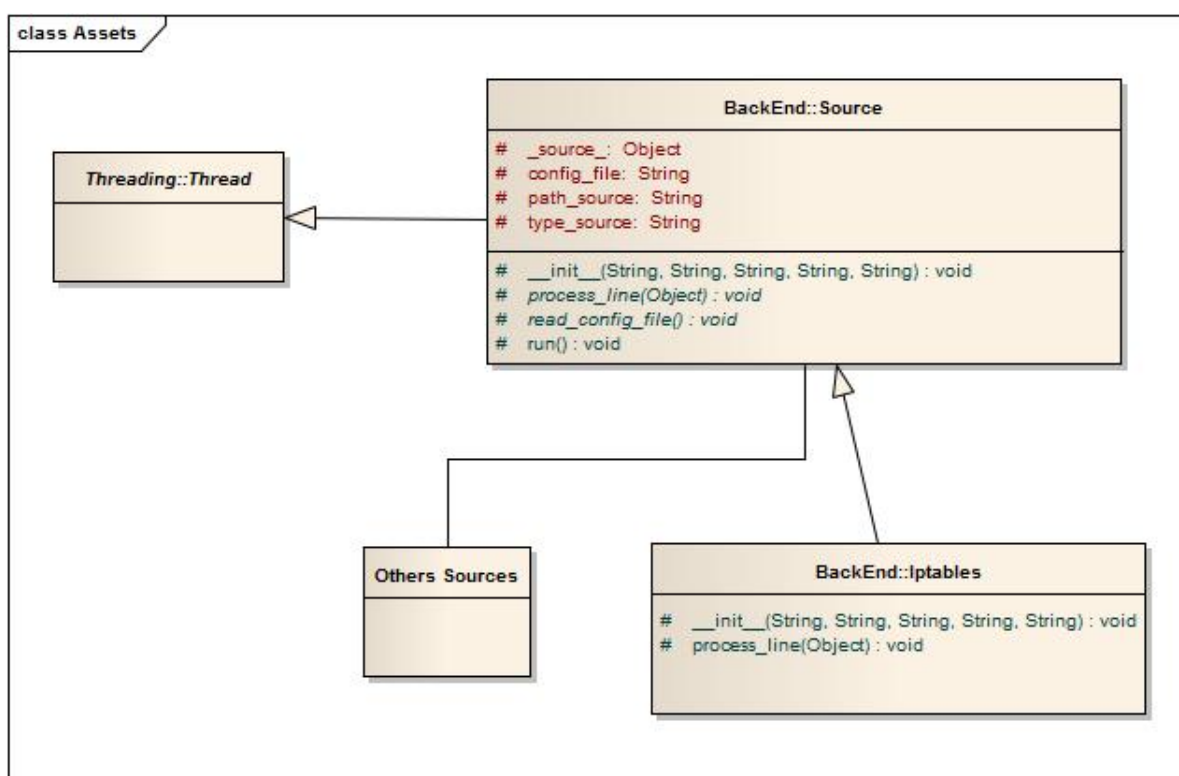


Figura 4.3: Diagrama de clases: Assets

La clase base que sería Source, hereda del comportamiento de la clase Thread (de la biblioteca `threading` de Python) y a su vez toda fuente de seguridad que se decida implementar heredaré de ella los métodos previstos.

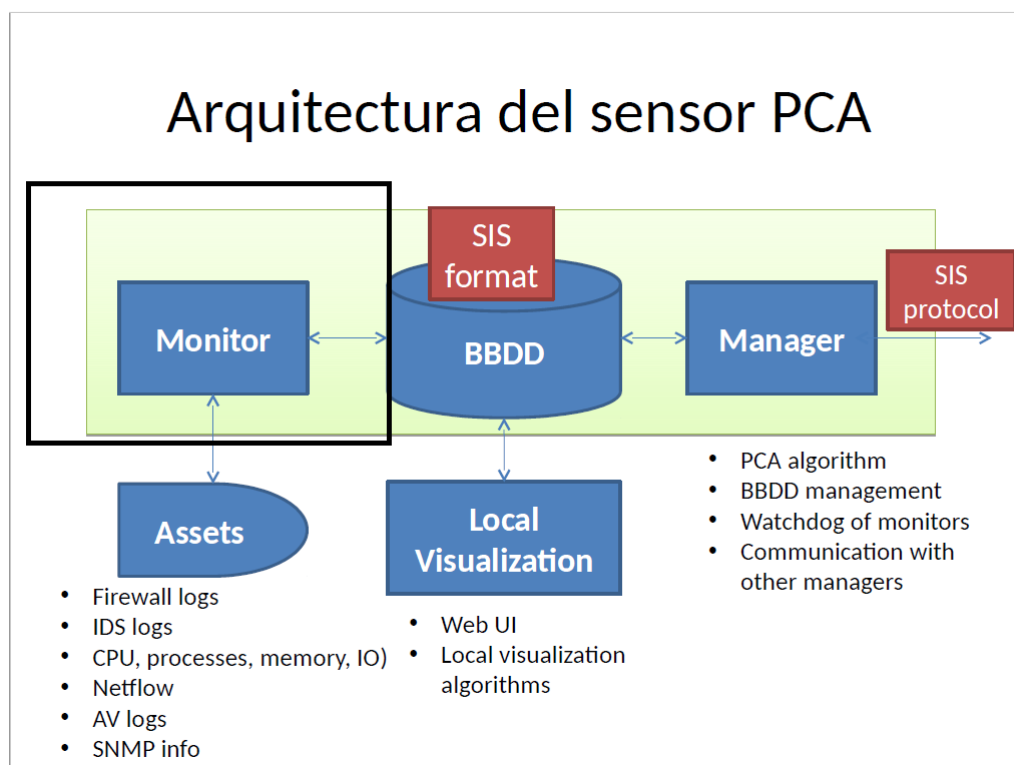
**Monitor**

Figura 4.4: Arquitectura: Monitor

En esta sección de la arquitectura es dónde se define la parte del control y ejecución (hilos) de las fuentes de seguridad implementadas. A su vez estas pasan dicha información a la base de datos usando el modelo ORM del framework Django. La jerarquía de clases sería de la siguiente manera:

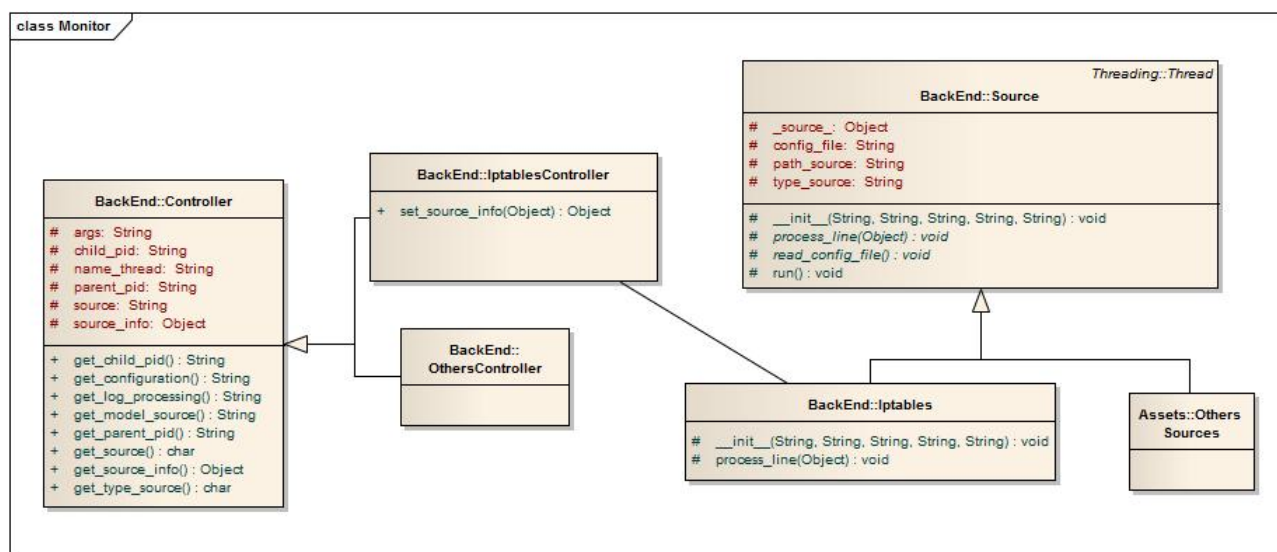


Figura 4.5: Diagrama de clases: Assets

Nuestra arquitectura monitor se representa con la clase Controller que se encarga de dotar de funcionalidad de ejecución de la fuente (hilo) para unos determinados parámetros de configuración.

### BBDD

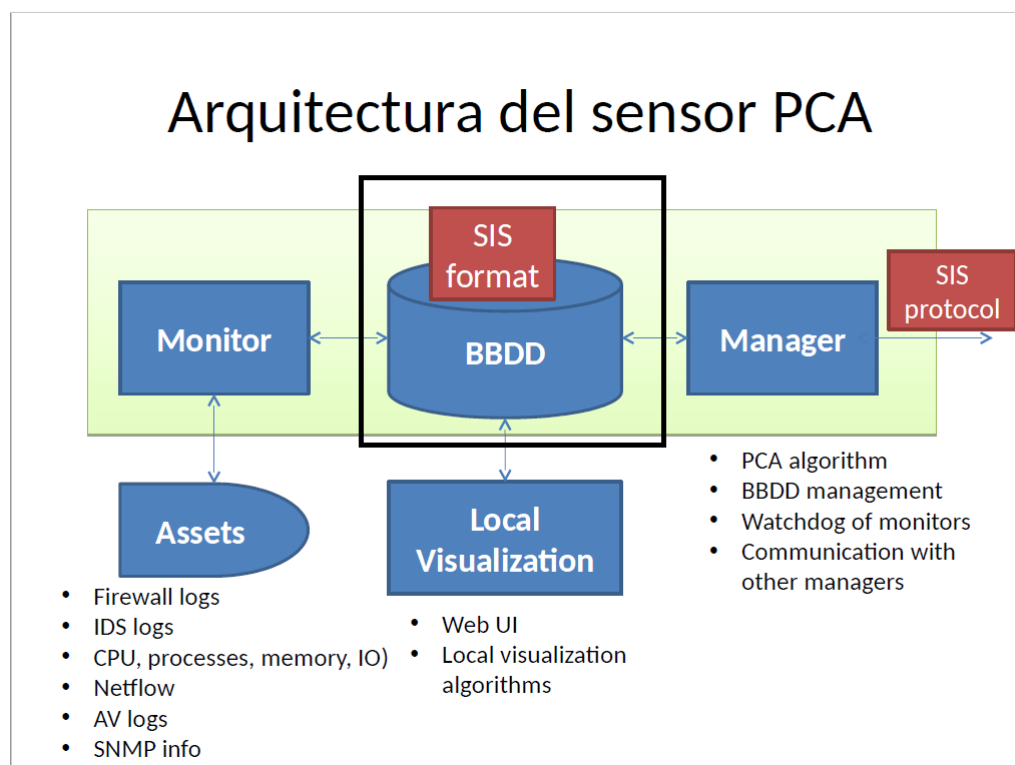


Figura 4.6: Arquitectura: Monitor

En esta sección de la arquitectura es dónde se define la parte de la interacción con la base de datos y que tablas (clases del modelo ORM) se han definido y con que relaciones. La jerarquía de clases sería de la siguiente manera:

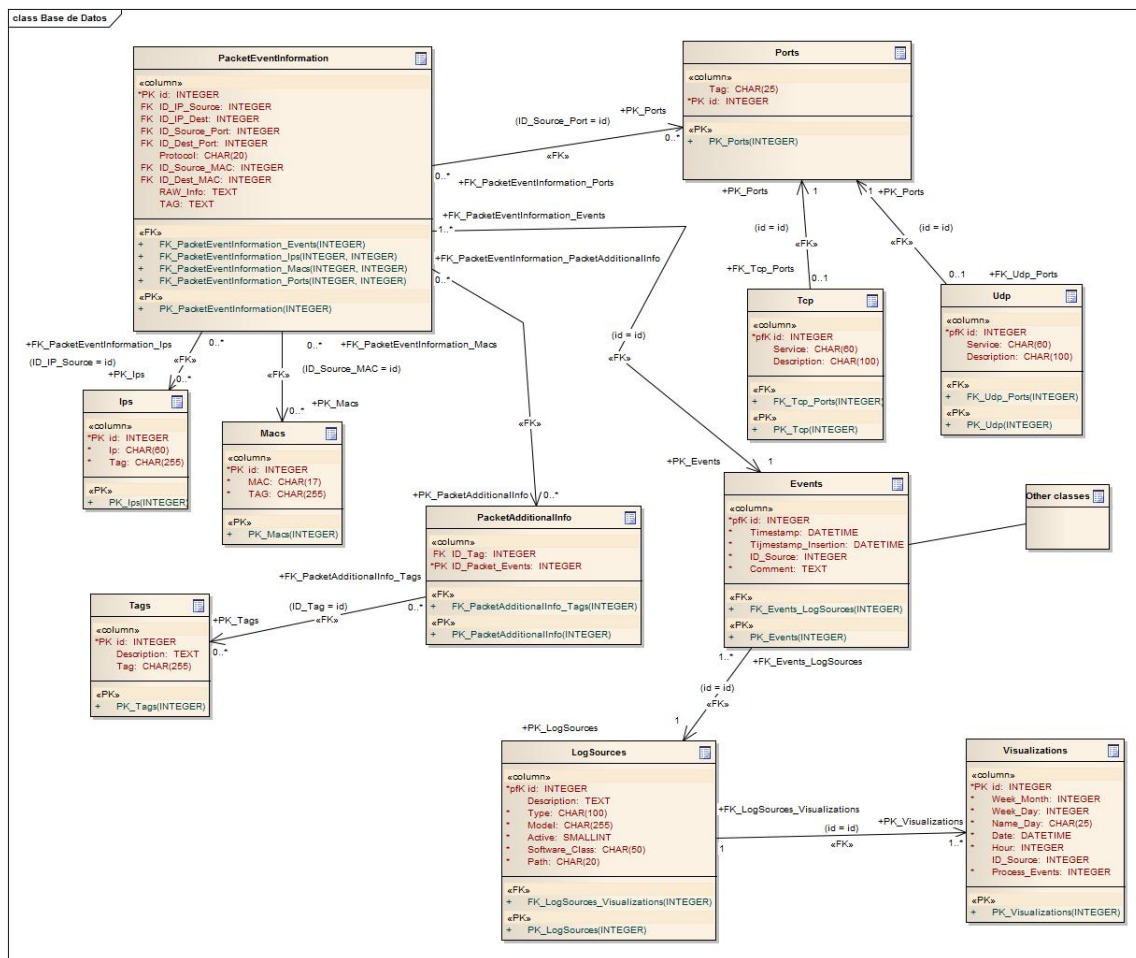


Figura 4.7: Diagrama de clases para la BD (usando ORM)

La clase o tabla que contendrá el peso de toda la jerarquía de base de datos, será **PacketEventsInformation**. Esta tabla sólo contendrá referencias externas o “foreign keys” a cada tabla que haga partícipe en su definición, es decir:

- Ips
- Macs
- Ports
- Events
- PacketAdditionalInfo

Cómo podemos observar en el diagrama anterior, por ejemplo, para la clase **PacketEventsInformation**, tenemos su traducción a formato ORM del motor proporcionado por Django:

```

1  # Clase que alberga la informacion relacionada con el paquete extraido
2  # mediante el log. La gran mayoria de los campos
3  # son identificadores o claves foraneas a otros
4  # objetos/instancias de la base de datos.
5
6  class PacketEventsInformation(models.Model):
7      ID_IP_Source = models.ForeignKey(Ips, models.SET_NULL, blank=True,
8                                      null=True,
9                                      related_name="ip_source")
10     ID_IP_Dest = models.ForeignKey(Ips, models.SET_NULL, blank=True,
11                                   null=True,
12                                   related_name="ip_dest")
13     ID_Source_Port = models.ForeignKey(Ports, models.SET_NULL,
14                                       blank=True,
15                                       null=True,
16                                       related_name="port_source")
17     ID_Dest_Port = models.ForeignKey(Ports, models.SET_NULL,
18                                     blank=True,
19                                     null=True,
20                                     related_name="port_dest")
21     Protocol = models.CharField(max_length=20, default='-')
22     ID_Source_MAC = models.ForeignKey(Macs, models.SET_NULL,
23                                     blank=True,
24                                     null=True,
25                                     related_name="mac_source")
26     ID_Dest_MAC = models.ForeignKey(Macs, models.SET_NULL, blank=True,
27                                    null=True,
28                                    related_name="mac_dest")
29     RAW_Info = models.TextField(default='-')
30     TAG = models.CharField(max_length=255, default='-')
31     id = models.OneToOneField(Events, on_delete=models.CASCADE,
32                              primary_key=True)
33
34     def __str__(self):
35         return '%s' % self.id

```

Figura 4.8: Ejemplo de clase ORM, en concreto PacketEventsInformation

El formato normalizado de la base de datos para visualizar dicha información será mediante JSON, ya que para hacer uso de la información la implementación consumirá dichos datos de la api que se proporciona con la aplicación.

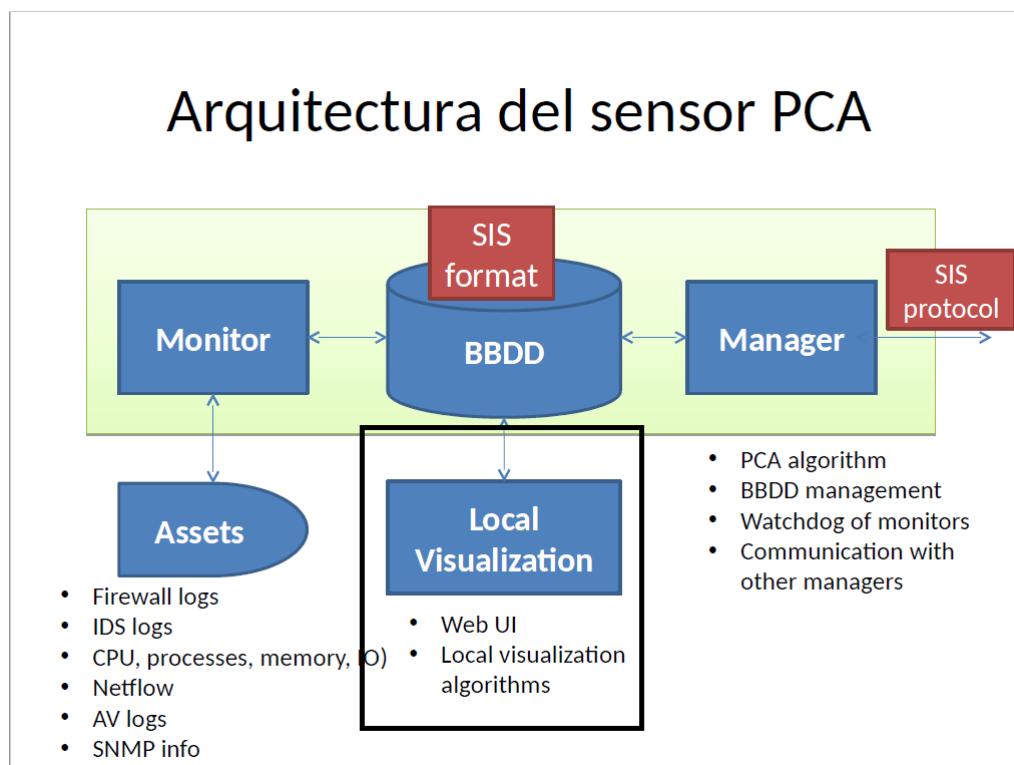
**Visualizations**

Figura 4.9: Arquitectura: Visualizaciones

En esta sección de la arquitectura es dónde se define la parte de la interacción con la base de datos y la visualización de los datos en la interfaz web. La jerarquía de clases sería de la siguiente manera:



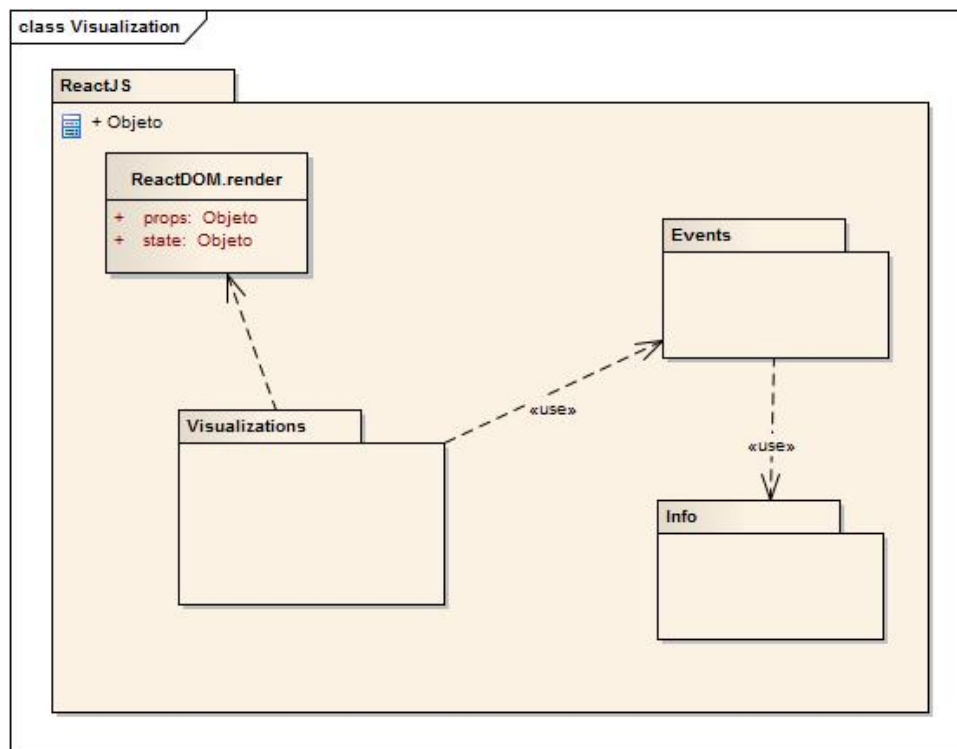


Figura 4.10: Clase Visualization para el paquete ReactJS

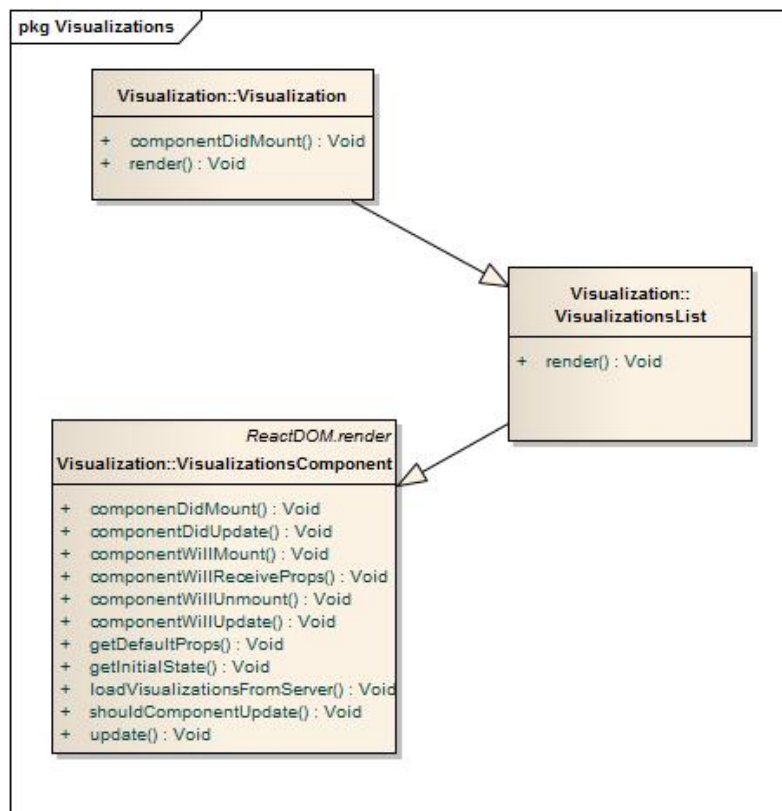


Figura 4.11: Paquete Visualizations

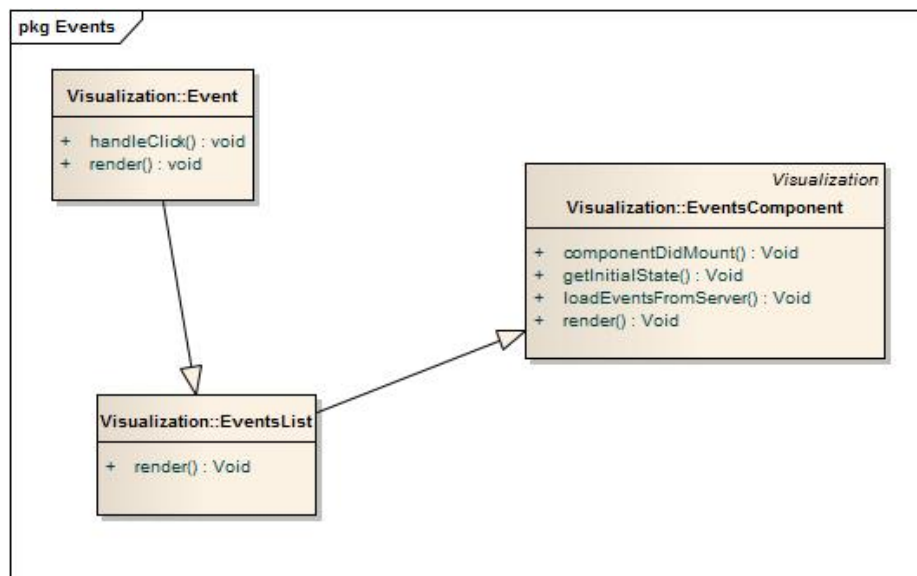


Figura 4.12: Paquete Events

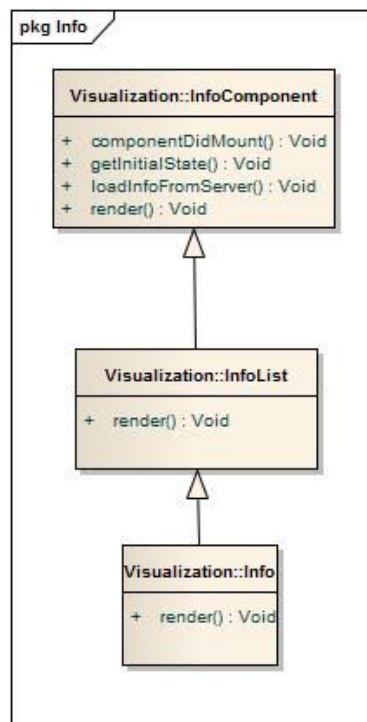


Figura 4.13: Paquete Info

### 4.3. Diagramas de Secuencia - Operaciones

En esta sección vamos a describir los diagramas de secuencia de operaciones tales como:

- Ejecución principal de la sonda.
- Ejecución principal del servidor web que sirve los datos a la interfaz web.

Ahora definiremos la parte de diagramas de secuencias para la interacción entre el usuario y el procesamiento de fuentes (Back); y el usuario y la visualización de los datos en la web de la aplicación (Front).

#### 4.3.1. Flujo de ejecución de la aplicación: BackEnd

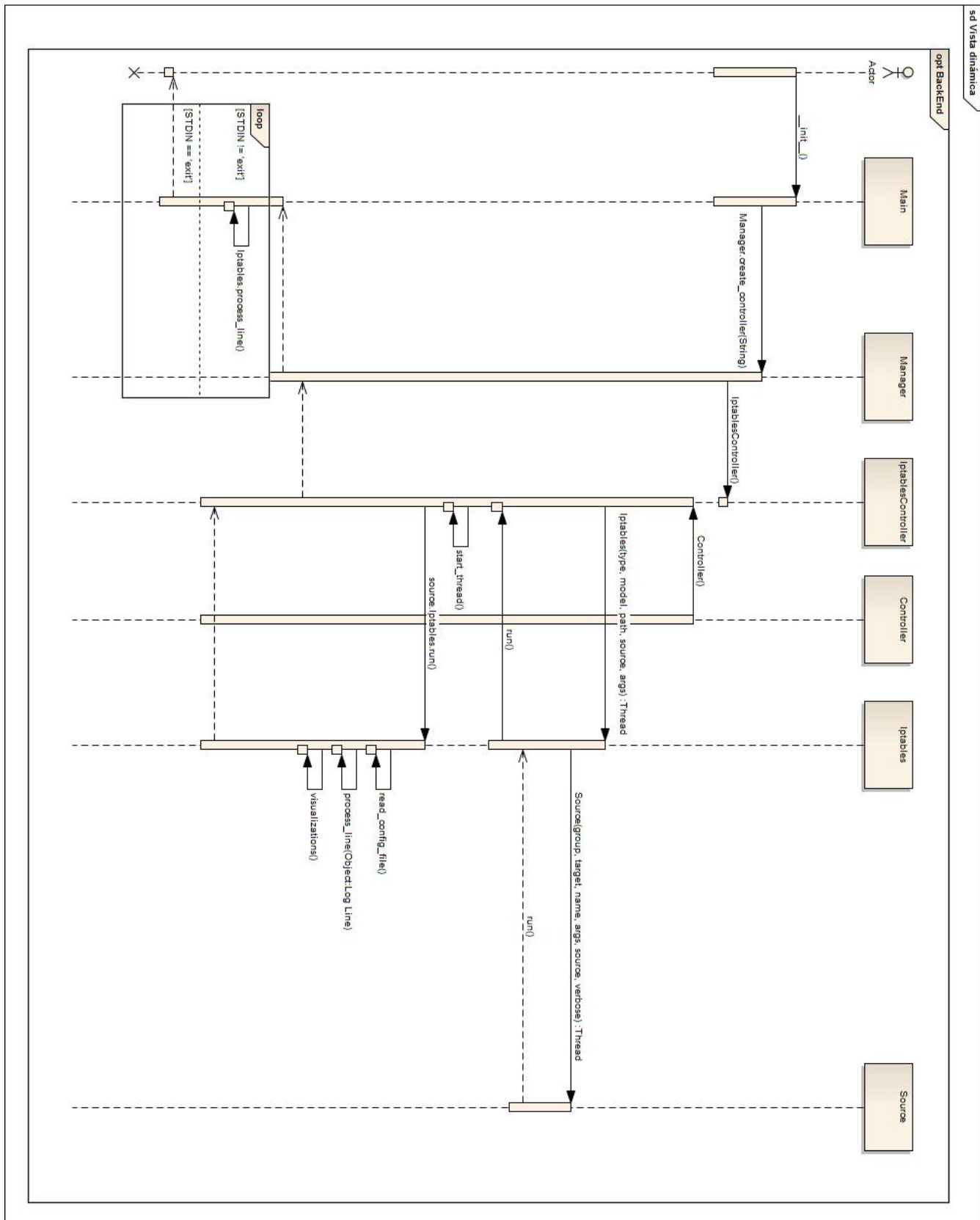


Figura 4.14: Diagrama de Secuencia para la parte BackEnd

## 4.3.2. Flujo de ejecución de la aplicación: FrontEnd

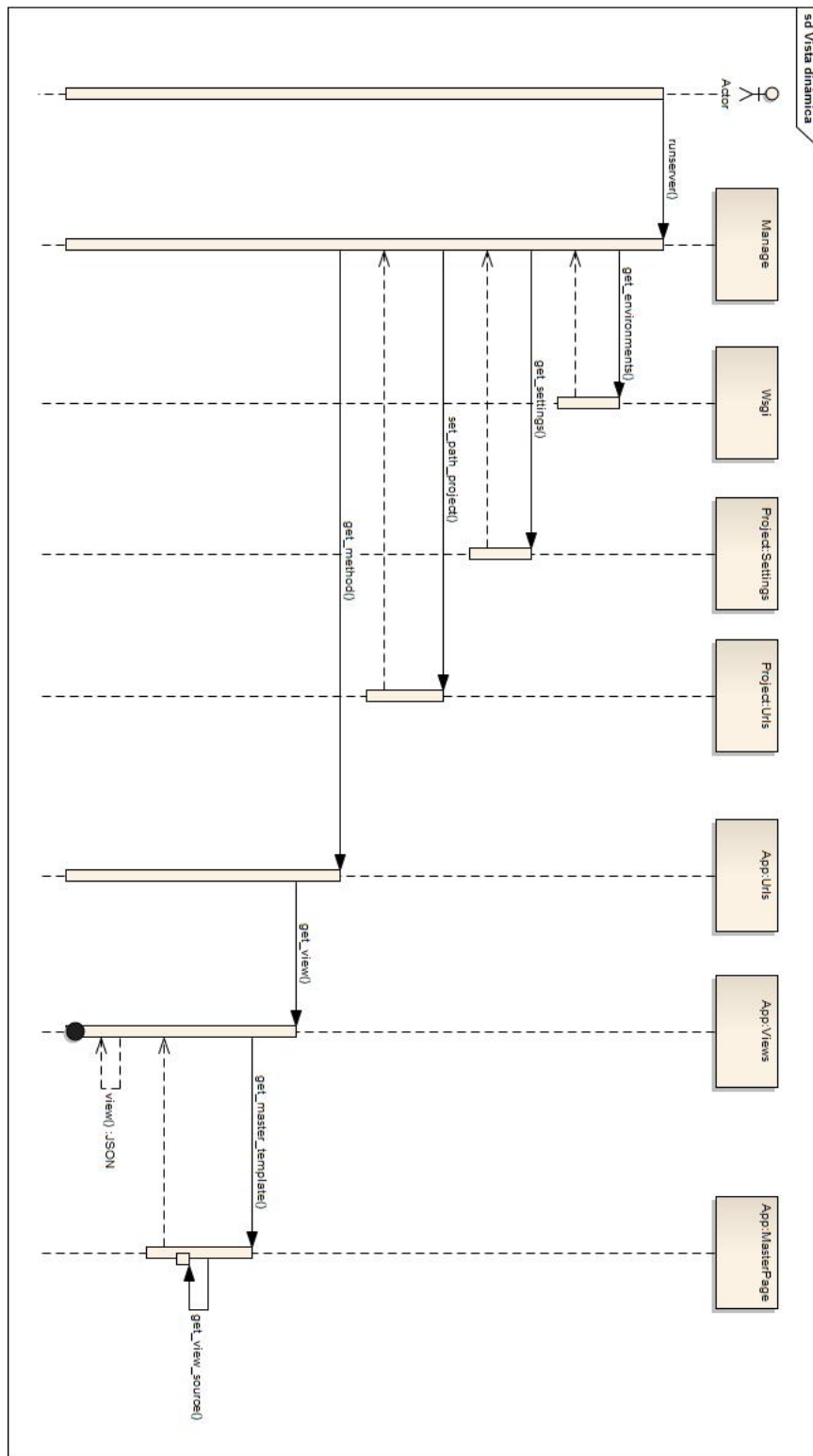


Figura 4.15: Diagrama de Secuencia para la parte FrontEnd

## **4.4. Diseño de la vista**

[TO DO]

# Capítulo 5

## Implementación

En este capítulo se definirán las configuraciones, diagramas e implementaciones que se han llevado a cabo en el proyecto. Primero, especificaremos los diagramas de clases para el BackEnd y FrontEnd (Flujo de ejecución Django) de la aplicación.

### 5.1. Flujo de ejecución BackEnd

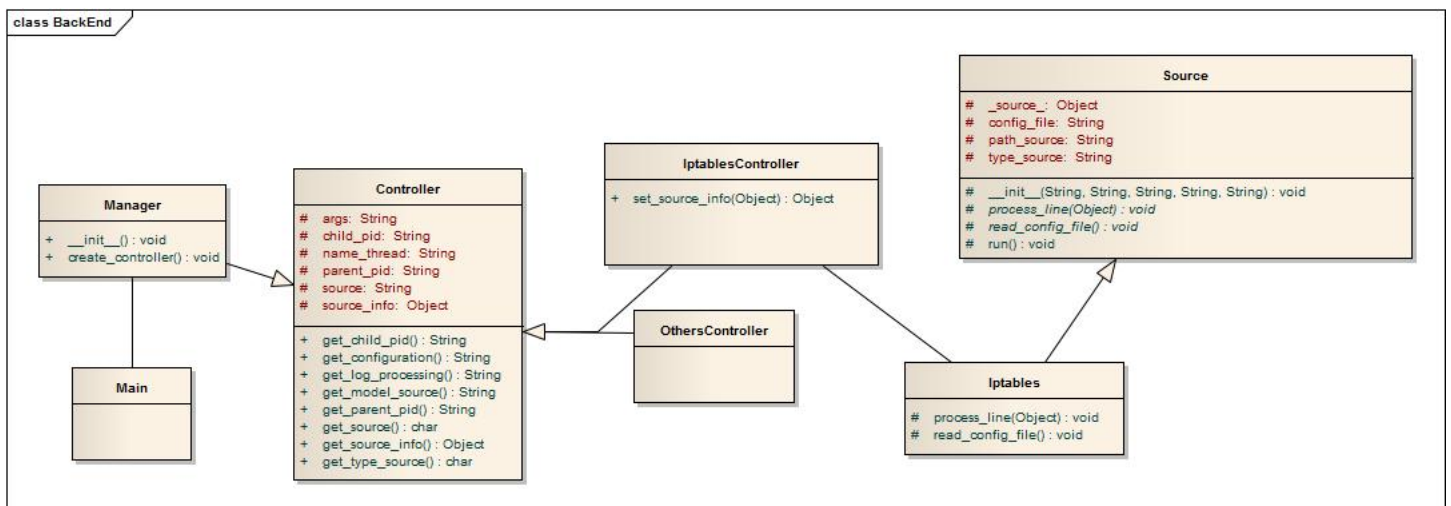


Figura 5.1: Diagrama de clases BackEnd

En el diagrama podemos ver de izquierda a derecha la herencia que hay entre las clases que se encargan de recoger la información de la sonda.

## 5.2. Flujo de ejecución FrontEnd

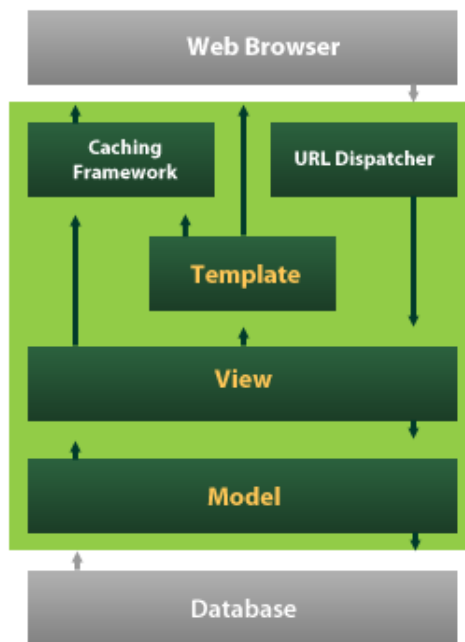


Figura 5.2: Flujo de ejecución del FrontEnd - [32]

Para este caso se ha decidido mostrar el flujo de ejecución que sigue una petición desde que llega al servidor de Django hasta que se sirve una vista asociada al navegador.

## 5.3. Configuración de la aplicación

En la fase de implementación del proyecto entran en juego la configuración previa de la máquina en la cuál se van a desplegar los componentes para su utilización. Hay ciertos factores a tener en cuenta.

Dado que se ha desarrollado una primera version funcional de la aplicación, en el futuro se podrían incluir ciertas herramientas software de automatización de despliegues y configuraciones como podrían ser:



### 5.3.1. Chef

Web: <https://www.chef.io/>



Chef es una plataforma de automatización de gran alcance que nos transforma el control de infraestructuras en código. Da igual sobre el sistema operativo o plataforma sobre el que se tenga que operar, Chef automatiza configuraciones de infraestructuras, desplegados y gestionado a través de la red, sin importar el tamaño de la solución software a controlar.

A continuación un diagrama explicando la arquitectura de un sistema bajo Chef:



Figura 5.3: Arquitectura de un sistema bajo Chef

### 5.3.2. Ansible

Web: <https://www.ansible.com/>



Ansible es una herramienta de código abierto que se utiliza para distribuir aplicaciones en nodos o servidores remotos de una manera automatizada. Nos proporciona un framework común a todas las aplicaciones, permitiendo así configurar cada una de ellas de una manera más fácil y

eficaz. Ansible se basa en “playbooks” (al contrario de Chef que se basa en Cookbooks) dónde se pueden definir una amplia variedad de sistemas para el despliegue de una aplicación.

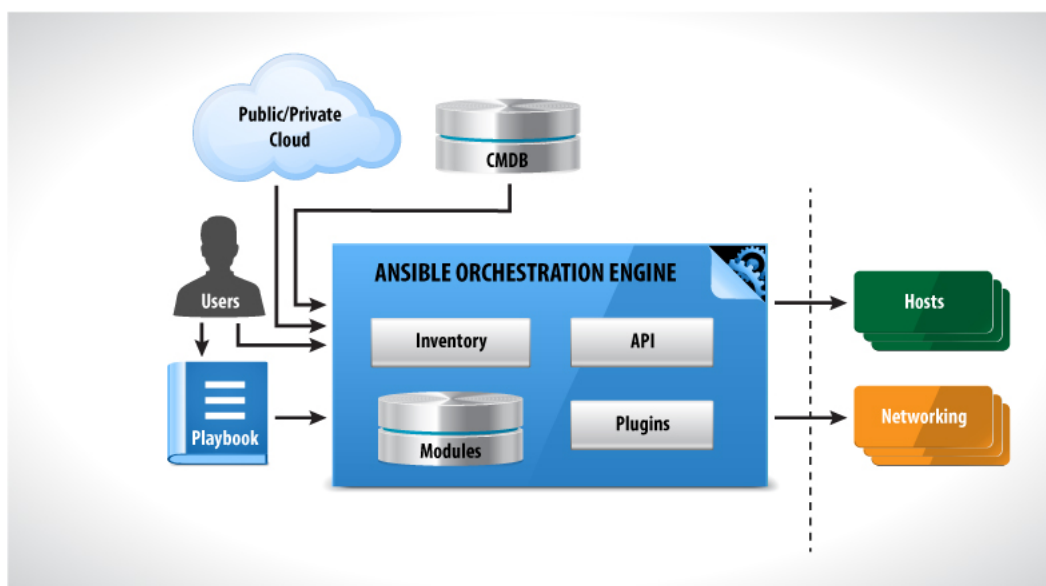
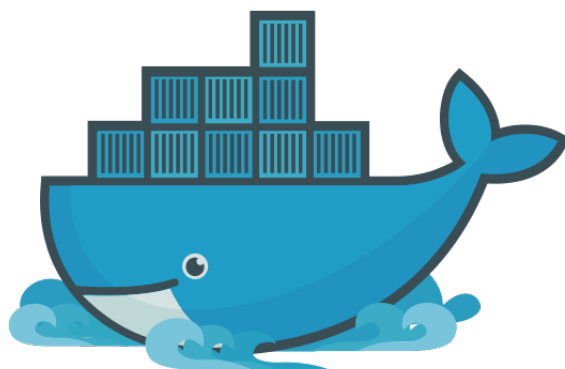


Figura 5.4: Diagrama de la arquitectura Ansible

### 5.3.3. Docker

Web: <https://www.docker.com/>



Docker es un sistema de desarrollo de sistemas completos, utilizando el concepto de contenedor (LXC) de los sistemas UNIX. Un contenedor de Docker, envuelve una fragmento de software en un sistema de archivos completo (imagen) que contiene todo lo necesario para funcionar: código, instancias de arranque, herramientas del sistema, bibliotecas del sistema... cualquier cosa que tendría un sistema operativo normal pero en una versión más optimizada y ligera (estos sistemas normalmente se obtienen de un Docker Repository y en su totalidad son sistemas open source). Esto garantiza que el software se ejecutará siempre de la misma forma, independientemente del medio en el que se despliegue.

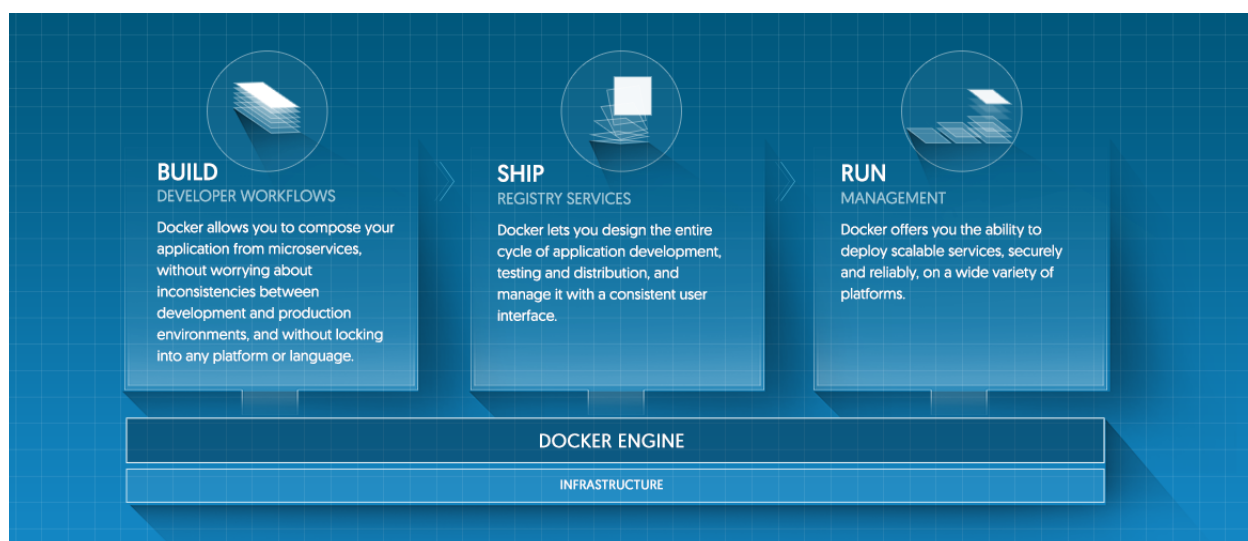


Figura 5.5: Diagrama de la arquitectura general de Docker

## 5.4. Configuración local para procesamiento

Los primeros pasos para la obtención de logs o eventos generados por la fuente de seguridad, iptables, serán los de configurar el sistema interno de correlación de logs rsyslog junto con el sistema de rotación de logs logrotate. Para el caso de rsyslog tenemos que definir un filtro para que cualquier evento que genere el sistema con un determinado mensaje definido en las reglas de iptables, sea capturado y almacenado en un determinado directorio. También hemos dotado a los logs del sistema de timestamp con mayor precisión para poder diferenciar eventos con mayor afinamiento y cambiado la tupla de permisos a la hora de crear un archivo con `FileCreateMode`.

```
1      #
2      # Use traditional timestamp format.
3      # To enable high precision timestamps, comment out the following
4      # line.
5      #
6      # $ActionFileDefaultTemplate RSYSLOG_TraditionalFileFormat
7      #
8      # Set the default permissions for all log files.
9      #
10     $FileOwner root
11     $FileGroup adm
12     $FileCreateMode 0644
13     $DirCreateMode 0755
14     $Umask 0022
15
16     # IPTABLES
17
18     :msg,contains,"IPTMSG= " -/var/log/iptables.log
19     :msg,regex,"^\[ *[0-9]*\.[0-9]*\] IPTMSG= " -/var/log/iptables.log
20     :msg,contains,"IPTMSG= " ~
```

Figura 5.6: Configuración de iptables para Rsyslog

Hemos de configurar Logrotate (más información sobre los campos visitar sección [2.4.10](#)):

```
1 /var/log/iptables.log
2 {
3     rotate 7
4     daily
5     missingok
6     notifempty
7     delaycompress
8     compress
9     postrotate
10         invoke-rc.d rsyslog restart > /dev/null
11     endscript
12 }
```

Figura 5.7: Configuración de iptables para Logrotate

Tenemos que definir una regla específica para el demonio de rsyslog en la cual se filtre también por los campos que queramos de iptables:

```
1 # into separate file and stop their further processing
2 if ($syslogfacility-text == 'kern') and \
3 ($msg contains 'IPTMSG=' and $msg contains 'IN=') \
4 then
5     & ~
6     -/var/log/iptables.log
```

Figura 5.8: Configuración de iptables.conf para Rsyslog.d

Estos tres pasos o configuraciones nos permiten redireccionar un log de iptables al directorio `/var/log/` y en concreto para el archivo **iptables.log**. Esta configuración tendrá efecto una vez hayamos reiniciado el servicio rsyslog o en el próximo inicio de sesión sobre la máquina.

### 5.4.1. Recogida y almacenamiento de logs

Para el caso que nos ocupa, iptables tiene una forma de definir reglas internas para el filtrado de paquetes según el tipo de comunicación que se establezca contra la máquina o desde la máquina hacia el exterior.

```
1 iptables -A INPUT -p tcp -m tcp --dport 22 -j LOG --log-prefix  
  "IPTMSG=Connection SSH "
```

Figura 5.9: Ejemplo de regla iptables

En la siguiente sección explicaremos en profundidad cada una de las opciones de la regla, pero a simple vista podemos observar que el mensaje asociado a la regla coincide con la palabra clave del filtro empleado en rsyslog: **IPTMSG=**. Así pues, una vez dicho evento se genere el sistema y syslog lo procese como un mensaje de un servicio determinado, en este caso iptables, rsyslog filtrará dicho mensaje según su configuración para almacenarlo posteriormente en `/var/log/iptables.log`.

### 5.4.2. Iptables

El servicio de firewall del kernel de GNU Linux, iptables, nos proporciona una interfaz de reglas y tablas en dónde podemos definir patrones o reglas que actúen sobre el tráfico que llega o sale desde nuestra máquina. Las reglas que se han definido por defecto son las siguientes (si queremos más filtros hay que implicar al protocolo y su puerto asociado con un mensaje):

```
1 # Generated by iptables-save v1.4.21 on Mon Jan 25 20:37:18 2016  
2 *filter  
3 :INPUT ACCEPT [0:0]  
4 :FORWARD ACCEPT [0:0]  
5 :OUTPUT ACCEPT [0:0]  
6 -A INPUT -d 127.0.0.1/32 -p icmp -m icmp --icmp-type 8 -m state  
  --state NEW,RELATED,ESTABLISHED -j LOG --log-prefix  
  "IPTMSG=Connection ICMP "  
7 -A INPUT -d 127.0.0.1/32 -p icmp -m icmp --icmp-type 8 -m state  
  --state NEW,RELATED,ESTABLISHED -j DROP  
8 -A INPUT -p tcp -m tcp --dport 22 -j LOG --log-prefix  
  "IPTMSG=Connection SSH "  
9 -A INPUT -p tcp -m tcp --dport 22 -j DROP  
10 COMMIT
```

Figura 5.10: Configuración reglas iptables

A continuación una breve explicación de cada opción o comando:

- -A: Añadir una nueva regla a una cadena de la tabla.
- -d: Especificación para la ip destino, en este caso localhost con una máscara de subred determinada.
- -dport: Especificación para el puerto destino al que se realizará la posible conexión o envío de paquetes.

- -p: Especificación del protocolo del paquete.
- -m: Especificación de matching, en este caso icmp o tcp dentro de la descripción del paquete.
- -icmp-type: Extensión del tipo de ping que se va a procesar desde la regla.
- -state: Tipo de paquete según conexión:
  - ◊ NEW: Paquete que crea una nueva conexión.
  - ◊ RELATED: Paquete que está relacionado a una conexión existente, pero que no es parte de ella, como un error ICMP o, un paquete que establece una conexión de datos FTP.
  - ◊ ESTABLISHED: Paquete que pertenece a una conexión existente (que tuvo paquetes de respuesta).
  - ◊ INVALID (no usado): Paquete que no pudo ser identificado por alguna razón: quedarse sin memoria o errores ICMP que no corresponden a ninguna conexión conocida. Normalmente estos paquetes deben ser descartados.
- -j: Acción de salto cuando encuentre dicha regla de paquetes. Se especifican dos acciones:
  - ◊ LOG -log-prefix "message" : Cuando se encuentre dicha regla se recolecta cómo log de la misma adjuntando un mensaje para diferenciarla del resto de reglas.
  - ◊ DROP : Cuando se encuentre dicha regla se descarta el paquete dentro del propio sistema. Al ir seguido LOG de un DROP el paquete se muestra en el registro de syslog para posteriormente ser eliminado del registro de almacenamiento de paquetes (Para esto usamos rsyslog que se encarga de almacenar dicho mensaje en un archivo de log).

Así pues una vez tengamos un evento o paquete o log de iptables en nuestro sistema generados mediante `ssh 127.0.0.1` o `ping 127.0.0.1` obtendremos lo siguiente:

```

1 [dom jul  3 17:04:03 2016] IPTMSG=Connection SSH IN=lo OUT=
  MAC=00:00:00:00:00:00:00:00:00:00:00:00:00:00:08:00 SRC=127.0.0.1
  DST=127.0.0.1 LEN=60 TOS=0x00 PREC=0x00 TTL=64 ID=39454 DF
  PROTO=TCP SPT=47706 DPT=22 WINDOW=43690 RES=0x00 SYN URGP=0
2 [dom jul  3 17:04:05 2016] IPTMSG=Connection SSH IN=lo OUT=
  MAC=00:00:00:00:00:00:00:00:00:00:00:00:00:00:08:00 SRC=127.0.0.1
  DST=127.0.0.1 LEN=60 TOS=0x00 PREC=0x00 TTL=64 ID=39455 DF
  PROTO=TCP SPT=47706 DPT=22 WINDOW=43690 RES=0x00 SYN URGP=0

```

Figura 5.11: Evento de ssh localhost en el sistema

Lo anterior correspondía a la salida del comando `$ dmesg -T` que muestra todos los mensajes del sistema que se han pasado al syslog. La opción -T se usa para especificar el timestamp de cada mensaje con una mayor precisión.

```

1 2016-07-03T17:03:35.664324+02:00 debian kernel: [23337.363387]
  IPTMSG=Connection SSH IN=lo OUT=
  MAC=00:00:00:00:00:00:00:00:00:00:00:00:08:00 SRC=127.0.0.1
  DST=127.0.0.1 LEN=60 TOS=0x00 PREC=0x00 TTL=64 ID=39454 DF
  PROTO=TCP SPT=47706 DPT=22 WINDOW=43690 RES=0x00 SYN URGP=0
2 2016-07-03T17:03:37.668326+02:00 debian kernel: [23339.369692]
  IPTMSG=Connection SSH IN=lo OUT=
  MAC=00:00:00:00:00:00:00:00:00:00:00:00:08:00 SRC=127.0.0.1
  DST=127.0.0.1 LEN=60 TOS=0x00 PREC=0x00 TTL=64 ID=39455 DF
  PROTO=TCP SPT=47706 DPT=22 WINDOW=43690 RES=0x00 SYN URGP=0

```

Figura 5.12: Log capturado y almacenado por rsyslog en /var/log/iptables.log

Lo anterior corresponde con la manipulación por parte de rsyslog del mensaje obtenido en syslog. Cómo podemos observar se añade un campo de timestamp de mayor precisión según la configuración que hemos establecido en rsyslog.conf para poder diferenciar con mayor exactitud eventos entre diferentes franjas de tiempo.

```

1 ++++++
2 -----
3
4 Procesando linea --> 2016-07-03T17:12:53.632264+02:00 debian kernel:
  [23896.003739] IPTMSG=Connection ICMP IN=lo OUT=
  MAC=00:00:00:00:00:00:00:00:00:00:00:00:08:00 SRC=127.0.0.1
  DST=127.0.0.1 LEN=84 TOS=0x00 PREC=0x00 TTL=64 ID=64157 DF
  PROTO=ICMP TYPE=8 CODE=0 ID=14177 SEQ=7
5
6 -----
7 ++++++
8 --> Insertado registro: {'TAG': 'Connection ICMP', 'ID_Source_PORT':
  None, 'Protocol': u'ICMP', 'RAW_Info':
  '2016-07-03T17:12:53.632264+02:00 debian kernel 23896.003739
  IPTMSG=Connection ICMP IN=lo OUT=
  MAC=00:00:00:00:00:00:00:00:00:00:00:00:08:00 SRC=127.0.0.1
  DST=127.0.0.1 LEN=84 TOS=0x00 PREC=0x00 TTL=64 ID=64157 DF
  PROTO=ICMP TYPE=8 CODE=0 ID=14177 SEQ=7 ', 'ID_Source_MAC': <Macs:
  00:00:00:00:00:00:00:00:00:00:00:00:00:08:00>, 'ID_Source_IP': <Ips:
  127.0.0.1>, 'ID_Dest_IP': <Ips: 127.0.0.1>, 'ID_Dest_PORT': None,
  'ID_Dest_MAC': <Macs: 00:00:00:00:00:00:00:00:00:00:00:00:00:08:00>}
9
10 ++++++
11 --> Fin de procesado de linea

```

Figura 5.13: Procesamiento del log capturado y almacenado en la bd interna de la aplicación



### 5.4.3. Transformación de log en información útil: Parser

Una vez hemos descrito los pasos a seguir para obtener un log para un evento iptables, llega la hora de procesar y hacer útil todos esos datos que tenemos. Para esta finalidad tenemos que usar expresiones regulares para generar un parser con el que ser capaz de traducir todos esos datos en información útil para nuestra aplicación.

```

1 obj = Pygtail("/var/log/iptables.log")
2
3 while True:
4     try:
5         for line in obj:
6             if len(line) > 1:
7                 self.process_line(line)
8     except Exception as ex:
9         print "Pygtail processing -> ", ex

```

Figura 5.14: Instancia de la clase Pygtail y lectura de las líneas del log

Hacemos uso del módulo o paquete **Pygtail** que nos permite leer archivos de log cuyos registros internos aún no han sido leídos. Es una especie de **\$ tail -f** a un archivo en concreto, pero usa un concepto de offset e inode para saber la última actualización y posición del archivo antes y después de ser abierto para así saber que parte de la última lectura se quedo en ejecución. Éste es otro punto importante dado que para hacer uso de esta funcionalidad debemos crear un archivo con el nombre del log, véase **/var/log/iptables.log** cuya extensión final sea offset (**/var/log/iptables.log.offset**) y sus permisos los siguientes:

```

1 -rw-r--r-- 1 root adm 10391 jul  3 17:12 /var/log/iptables.log
2 -rw-r--rw- 1 root root 14 jul  3 17:13 /var/log/iptables.log.offset

```

Figura 5.15: Permisos de los archivos iptables.log e iptables.log.offset

Una vez tengamos el archivo abierto y con sus líneas procesadas por Pygtail, es el momento de usar regex sobre los objetos string de cada línea de log. Para ello hacemos uso del método **split** para dividir por palabras, es decir, posiciones separadas en una lista a todas las coincidencias con una palabra que pudiera tener toda la cadena.

```

1 # Se trocea por palabras el log leído
2 line = re.split("\W? ", line)

```

Figura 5.16: Uso del método split sobre la entrada de líneas de log

Una vez separado por palabras toda la línea de log, pasamos a diferenciar entre las etiquetas que iptables pone a cada campo con su valor, es decir, una tupla **key=>value**.

```

1 tag_str = ((re.compile('^(.*)=')).search(str(line))).group(0)
2 tag_split = tag_str.split(',')

```

Figura 5.17: Obtenemos la tupla key=>value para cada etiqueta del log

Ya tenemos todas las etiquetas de los campos del log y ahora nos toca extraer su valor asociado para asignarlo al ORM de la base de datos, es decir, almacenar la información en la BD.

```

1 db_column = ['ID_Source_IP', 'ID_Dest_IP',
2             'ID_Source_PORT', 'ID_Dest_PORT',
3             'Protocol', 'ID_Source_MAC',
4             'ID_Dest_MAC']
5
6 # El nombre de las tags, segun el orden de la
7 # columnas en db_column, las extraigo del fichero
8 # de configuracion a traves del registro info_config_file
9
10 labels = [self.info_config_file["Source_Ip"],
11           self.info_config_file["Dest_Ip"],
12           self.info_config_file["Source_Port"],
13           self.info_config_file["Dest_Port"],
14           self.info_config_file["Protocol"]]
15
16 # Almacenamos las etiquetas o campos del log de iptables
17 for it in tag_split:
18     if len(it.split('=')) == 2:
19         self.tag_log.append((it.split('=')[0].strip('\ ')))
20
21 # Buscamos la correlacion entre los campos definidos en la
22 # configuracion con los extraidos del log de iptables
23
24 for label in labels:
25     if (re.compile(label)).search(tag_str):
26         if self.tag_log.index(label) > 0:
27             db_column_name = db_column[0]
28             register[db_column.pop(0)] = self.regexp(db_column_name,
29             label, str(line))
30             self.tag_log.remove(label)
31         else:
32             register[db_column.pop(0)] = None

```

Figura 5.18: Vamos asignando cada etiqueta y su valor a su asociado del ORM

Los siguientes pasos de comprobación de integridad de valores y demás se relegan a la visualización del interesado en el método `process_line` del archivo código fuente `iptables.py`

### 5.4.4. Workflow

Ahora vamos a detallar los pasos que seguirá la aplicación a la hora de la visualización de información en la web mediante el framework Django. Pero primero cómo se procesan y almacenan los datos de los logs:

- Primera instancia a ejecutar `main.py`: Se definen las llamadas al sistema Manager que se encargará a su vez de llamar al controlador que se asocia con la fuente encargada (iptables).
- Segunda instancia a ejecutar `manager.py`: Gestión de los controladores que hacen uso de la herramienta. Cada controlador tendrá asociada su fuente.
- Tercera instancia a ejecutar `controller.py`: Esta clase contiene unas características que herederán los controladores de cada fuente y a su vez lanza la ejecución de los hilos para cada una de las fuentes. Aquí se define o exporta la variable de entorno de Django para utilizar el ORM del mismo.
- Cuarta instancia a ejecutar `source.py`: Aunque se vaya directamente al siguiente punto, primero se pasa el control a esta clase de la cuál hereda iptables y se crean los fragmentos de código para la ejecución en iptables.
- Quinta instancia a ejecutar `iptables.py`: Ejecución y configuración de todo el entramado de iptables, de la insercción de datos en la BD y de la extracción de características de los logs.

Ahora corresponde el turno al flujo de ejecución para la visualización web:

- Primera instancia a ejecutar `manage.py`: Aquí se cargan las configuraciones por defecto del proyecto Django para la aplicación definida.
- Segunda instancia a ejecutar `wsgi.py`: Se usa la biblioteca wsgi para poder lanzar la aplicación por su nombre dentro del namespace definido por Django.
- Tercera instancia a ejecutar `settings.py`: Se cargan todas las configuraciones establecidas por el usuario para el proyecto y que a su vez usarán cómo configuración base para las aplicaciones derivadas del mismo.
- Cuarta instancia a ejecutar `urls.py` (archivo que pertenece al proyecto Django): Aquí se encamina la visualización entre la aplicación o la parte de administración del proyecto.
- Quinta instancia a ejecutar `urls.py` (archivo que pertenece a la aplicación): Aquí es dónde se define el router de la aplicación web y los métodos que se lanzarán una vez se hagan las peticiones sobre el servidor web Django.
- Sexta instancia a ejecutar `views.py`: Se ejecuta el método asociado a la renderización de la vista, que por defecto, sino hay ninguna ruta será `index`. Dependiendo del método a ejecutar sólo mostrará contenido estático o dinámico (JSON) obtenido de la base de datos.

- Séptima instancia a ejecutar aplicacion/index.html (Paso previo de MasterPage.html): Se carga el trozo de esta plantilla o template (que en nuestro caso es una vista) que contiene el esqueleto principal de la parte web para todas las vistas/plantillas que heredan de esta general.
- Octava instancia a ejecutar aplicacion/index.html (Carga de la plantilla index.html): Se carga el trozo de esta plantilla dentro del bloque definido en la plantilla MasterPage que anteriormente se proceso.

#### 5.4.5. Visualización de eventos

[TO DO]

# Capítulo 6

## Evaluación

- pruebas funcionales y no funcionales
- caja blanca
- caja negra

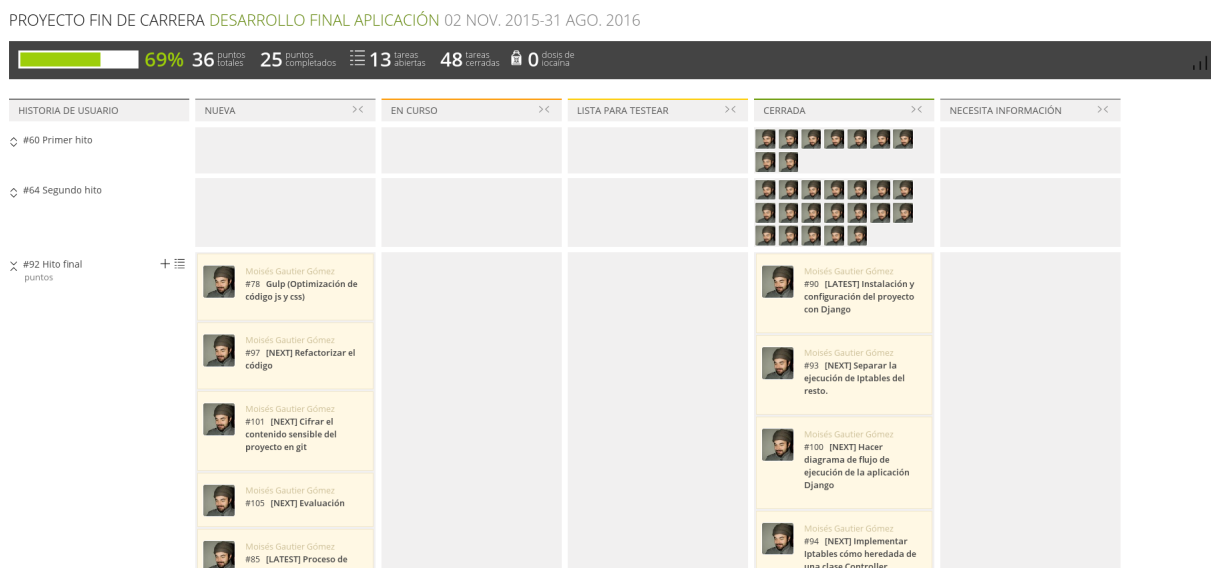


# Capítulo 7

## Planificación temporal

La planificación general del proyecto siguiendo un modelo SCRUM, para su aprendizaje, se encuentra en un panel de tareas en

Taiga <https://tree.taiga.io/project/mgautier-proyecto-fin-de-carrera/backlog>.



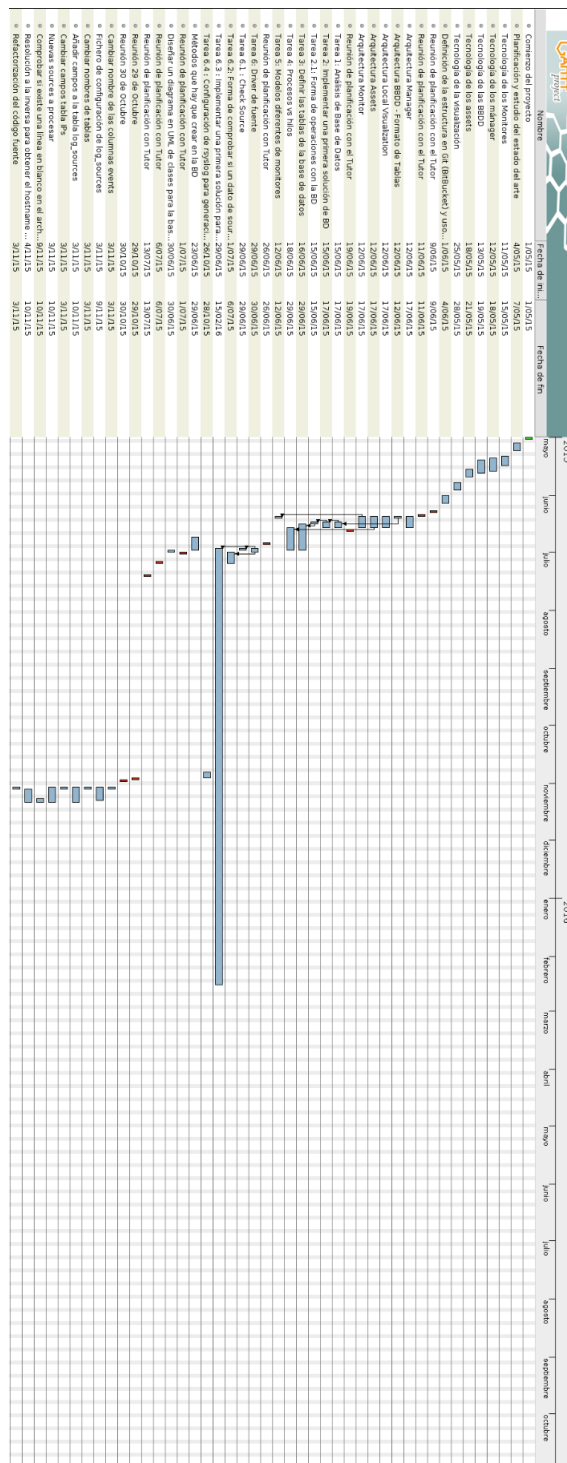


Figura 7.2: Diagrama de Gantt 1-05-15 a 3-11-15



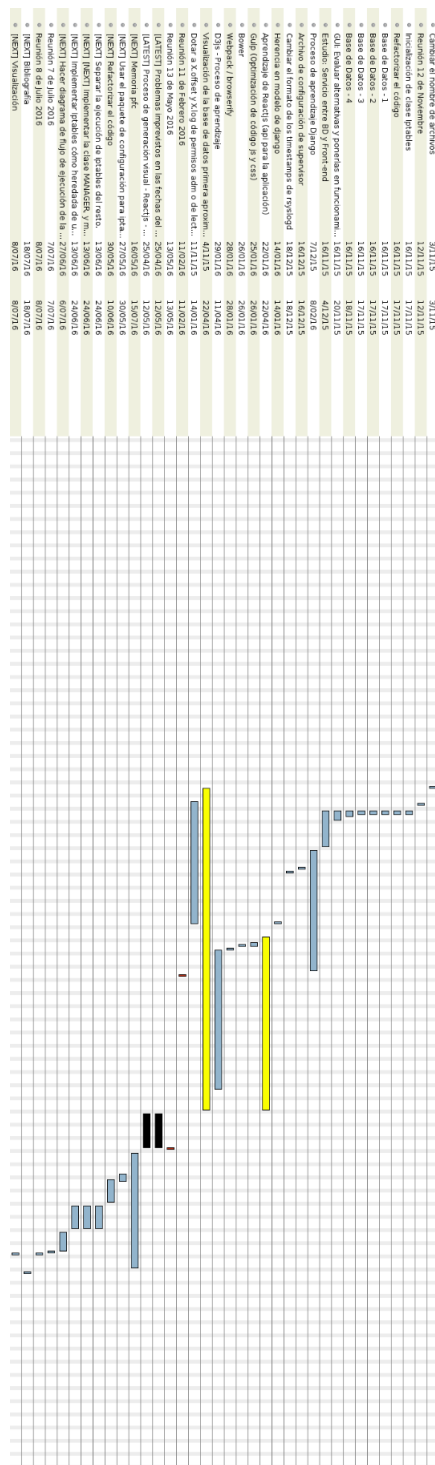


Figura 7.3: Diagrama de Gantt 3-11-15 a 18-07-16



# Capítulo 8

## Conclusiones

Durante la realización de éste proyecto, se han conseguido los siguientes resultados:

- Comprender el funcionamiento de los servicios del sistema: rsyslog, syslog, logrotate, nginx.
- Comprender el funcionamiento del firewall del kernel, Iptables.
- Desarrollar una solución software usando el lenguaje de programación Python.
- Desarrollar una solución software, para sistemas web, usando el framework Django.
- Manipulación de bases de datos relacionales usando un modelo orientado a objetos (ORM) proporcionado por el framework de desarrollo.
- Diseñar el prototipo base para los sensores de recopilación de información de seguridad.
- Diseñar una interfaz web dónde el usuario pueda visualizar la información que el sensor ha analizado.
- Despliegue de la aplicación en un servidor web en la nube para realizar demostraciones de la herramienta (Digital Ocean).
- Documentar todo el proceso de realización del proyecto.

A partir de los resultados obtenidos durante la realización de éste proyecto se han extraído las siguientes conclusiones:

- La aplicación *Sensor para recopilación y visualización de información de seguridad en nodos de una red* es una herramienta que permite el análisis de información de dispositivos de seguridad para una máquina en una red.
- El código fuente implementado, responde a los objetivos que debe cumplir el desarrollo del proyecto.
- Con el diseño de la interfaz web se ha conseguido que un usuario sin conocimientos sobre el sistema sea capaz de entender lo que se está registrando en él y que tipo de información de seguridad puede disponer a la hora de realizar un estudio de la misma.

# Bibliografía

- [1] Nosql vs sql: Sql is the new nonosql. Disponible en la web: <http://www.nosql-vs-sql.com/>.
- [2] App accelerator. Working with a sqlite database. Disponible en la web: <https://wiki.appcelerator.org/display/guides2/Working+with+a+SQLite+Database#WorkingwithaSQLiteDatabase-ClosingthedatabaseandResultSet>.
- [3] Shadypixel Blog. Log iptables messages to a separate file with rsyslog. Disponible en la web: <https://blog.shadypixel.com/log-iptables-messages-to-a-separate-file-with-rsyslog/>.
- [4] Takipi Blog. 5 reasons you should stop hosting your elk stack locally. Disponible en la web: <http://blog.takipi.com/hosted-elasticsearch-the-future-of-your-elk-stack/>.
- [5] Takipi Blog. Log management tools face-off: Splunk vs. logstash vs. sumo logic. Disponible en la web: <http://blog.takipi.com/log-management-tools-face-off-splunk-vs-logstash-vs-sumo-logic/>.
- [6] Xorl blog. Descripción del formato log para iptables. Disponible en la web: <https://xorl.wordpress.com/2009/02/03/logging-with-iptables-and-syslog/>.
- [7] Python Central. Python's sqlalchemy vs other orms. Disponible en la web: <http://pythoncentral.io/sqlalchemy-vs-orms/>.
- [8] Cisco. Cisco application networking manager 5.2. Disponible en la web: [http://www.cisco.com/c/en/us/products/collateral/application-networking-services/application-networking-manager/data\\_sheet\\_c78-706167.pdf](http://www.cisco.com/c/en/us/products/collateral/application-networking-services/application-networking-manager/data_sheet_c78-706167.pdf).
- [9] Cyberciti. Force iptables to log messages to a different log file. Disponible en la web: <http://www.cyberciti.biz/tips/force-iptables-to-log-messages-to-a-different-log-file.html>.
- [10] Maria DB. Maria db. Disponible en la web: <https://mariadb.com/>.
- [11] debianHackers. Threads, procesos y cómo estos afectan la seguridad de un sistema. Disponible en la web: <https://debianhackers.net/threads-procesos-y-como-estos-afectan-la-seguridad-de-un-sistema-explicado-de-forma-que>
- [12] django. Logging. Disponible en la web: <https://docs.djangoproject.com/en/1.9/topics/logging/>.
- [13] Elastic. Github repository: Elastic. Disponible en la web: <https://github.com/elastic>.
- [14] Firebird. Firebird. Disponible en la web: <http://www.firebirdsql.org/>.

- [15] Web Forefront. Set up logging for a django project. Disponible en la web: <https://www.webforefront.com/django/setupdjologging.html>.
- [16] Genbeta:dev. Multiprocesamiento en python: Threads a fondo, enumeración, herencia y temporizadores. Disponible en la web: <http://www.genbetadev.com/python/multiprocesamiento-en-python-threads-a-fondo-enumeracion-herencia-y-temporizadores>.
- [17] Genbeta:dev. Nuestra primera aplicación con google app engine (python). Disponible en la web: <http://www.genbetadev.com/programacion-en-la-nube/nuestra-primer-a-aplicacion-con-google-app-engine-python>.
- [18] Cambridge Intelligence. The keylines toolkit. Disponible en la web: <http://cambridge-intelligence.com/keylines/>.
- [19] ipswitch. Log and event management. Disponible en la web: <https://www.ipswitch.com/solutions/log-and-event-management>.
- [20] Jooq. Jooq. Disponible en la web: <http://www.jooq.org/>.
- [21] Michal Karzynski. Setting up django with nginx, gunicorn, virtualenv, supervisor and postgresql. Disponible en la web: <http://michal.karzynski.pl/blog/2013/06/09/django-nginx-gunicorn-virtualenv-supervisor/>.
- [22] Groovy lang. Groovy: Differences with java. Disponible en la web: <http://www.groovy-lang.org/differences.html>.
- [23] lansweeper. Network management. Disponible en la web: <http://www.lansweeper.com/>.
- [24] LogRhythm. Siem. Disponible en la web: <https://logrhythm.com/es/products/siem/>.
- [25] Macaron. Macaron: Python o/r mapper. Disponible en la web: <http://nobrin.github.io/macaron/>.
- [26] Markwingerd. Setting up django with ubuntu 12.10, sqlite, gunicorn, and nginx on a digital ocean droplet. Disponible en la web: <https://markwingerd.wordpress.com/2013/09/01/setting-up-django-with-ubuntu-12-10-sqlite-gunicorn-and-nginx-on-a-digital-ocean-droplet>.
- [27] McAfee. Información de seguridad y administración de eventos. Disponible en la web: <http://www.mcafee.com/es/products/siem/index.aspx>.
- [28] MGautier. Iptables log timestamp value is from 1970. Disponible en la web: <http://stackoverflow.com/questions/32914701/iptables-log-timestamp-value-is-from-1970/32914903#32914903>.
- [29] MGautier. Proyecto fin de carrera - backlog. Resto de referencias en: <https://tree.taiga.io/project/mgautier-proyecto-fin-de-carrera/backlog>.
- [30] MGautier. Read a file with root permissions from python script. Disponible en la web: <http://stackoverflow.com/questions/34533505/read-a-file-with-root-permissions-from-python-script>.
- [31] MongoDB. MongoDB. Disponible en la web: <https://www.mongodb.com/>.
- [32] MyTardis. Django workflow. Disponible en la web: [http://mytardis.readthedocs.io/en/3.5/\\_images/DjangoArchitecture-JeffCroft.png](http://mytardis.readthedocs.io/en/3.5/_images/DjangoArchitecture-JeffCroft.png).

- [33] Erick Navarro. Despliegue django con nginx, supervisor y gunicorn. Disponible en la web: <https://codeandoando.com/despliegue-django-con-nginx-supervisor-y-gunicorn/>.
- [34] Neo4j. Graph visualization for neo4j. Disponible en la web: <https://neo4j.com/developer/guide-data-visualization/>.
- [35] Netfilter. Uso de iptables. Disponible en la web: <http://www.netfilter.org/documentation/HOWTO/es/packet-filtering-HOWTO-7.html>.
- [36] Digital Ocean. A comparison of nosql database management systems and models. Disponible en la web: <https://www.digitalocean.com/community/tutorials/a-comparison-of-nosql-database-management-systems-and-models>.
- [37] Digital Ocean. How to install and manage supervisor on ubuntu and debian vps. Disponible en la web: <https://www.digitalocean.com/community/tutorials/how-to-install-and-manage-supervisor-on-ubuntu-and-debian-vps>.
- [38] Digital Ocean. How to install and manage supervisor on ubuntu and debian vps. Disponible en la web: <https://www.digitalocean.com/community/tutorials/how-to-install-and-manage-supervisor-on-ubuntu-and-debian-vps>.
- [39] Digital Ocean. How to set up django with postgres, nginx, and gunicorn on ubuntu 14.04. Disponible en la web: <https://www.digitalocean.com/community/tutorials/how-to-set-up-django-with-postgres-nginx-and-gunicorn-on-ubuntu-14-04>.
- [40] Digital Ocean. Sqlite vs mysql vs postgresql: A comparison of relational database management systems. Disponible en la web: <https://www.digitalocean.com/community/tutorials/sqlite-vs-mysql-vs-postgresql-a-comparison-of-relational-database-management-systems>.
- [41] Digital Ocean. Understanding sql and nosql databases and different database models. Disponible en la web: <https://www.digitalocean.com/community/tutorials/understanding-sql-and-nosql-databases-and-different-database-models>.
- [42] Digital Ocean. Understanding sql and nosql databases and different database models. Disponible en la web: <https://www.digitalocean.com/community/tutorials/understanding-sql-and-nosql-databases-and-different-database-models>.
- [43] Stack overflow. django.request logger not propagated to root? Disponible en la web: <http://stackoverflow.com/questions/20282521/django-request-logger-not-propagated-to-root/22336174#22336174>.
- [44] Stack overflow. Simple log to file example for django 1.3+. Disponible en la web: <http://stackoverflow.com/questions/5739830/simple-log-to-file-example-for-django-1-3>.
- [45] Philogb. Javascript infovis toolkit. Disponible en la web: <http://philogb.github.io/jit/>.
- [46] PostgreSQL. Postgresql. Disponible en la web: <https://www.postgresql.org/>.
- [47] Python. Pep 249 – python database api specification v2.0. Disponible en la web: <https://www.python.org/dev/peps/pep-0249/>.

- [48] Python. Python - data model. Disponible en la web: <https://docs.python.org/2/reference/datamodel.html>.
- [49] Python. sqlite3 — db-api 2.0 interface for sqlite databases. Disponible en la web: <https://docs.python.org/2/library/sqlite3.html>.
- [50] Python. Web frameworks for python. Disponible en la web: <https://wiki.python.org/moin/WebFrameworks/>.
- [51] Mail Python. [python-es] threads con operaciones i/o en python. Disponible en la web: <https://mail.python.org/pipermail/python-es/2010-June/027561.html>.
- [52] Qbox. Elk stack. Disponible en la web: <https://qbox.io/blog/tag/elk-stack>.
- [53] Qbox. Parsing logs using logstash. Disponible en la web: <https://qbox.io/blog/parsing-logs-using-logstash>.
- [54] redis. redis. Disponible en la web: <http://redis.io/>.
- [55] Rsyslog. omfile: File output module. Disponible en la web: <http://www.rsyslog.com/doc/v8-stable/configuration/modules/omfile.html>.
- [56] Rthalley. Dnso toolkit for python. Disponible en la web: <https://github.com/rthalley/dnspython>.
- [57] s21sec. Lookwise. Disponible en la web: <https://www.s21sec.com/es/productos/lookwise>.
- [58] s21sec. Lookwise: ficha técnica. Disponible en la web: <https://www.s21sec.com/es/docs-a-resources/category/14-bitacora?download=176%3Aficha-bitacora>.
- [59] Tom Sawyer Software. Tom sawer software. Disponible en la web: <https://www.tomsawyer.com/>.
- [60] SpiceWorks. Spiceworks. Disponible en la web: <http://www.spiceworks.com/free-network-monitoring-management-software/>.
- [61] SQLAlchemy. Sqlalchemy. Disponible en la web: [http://docs.sqlalchemy.org/en/rel\\_1\\_0/orm/tutorial.html](http://docs.sqlalchemy.org/en/rel_1_0/orm/tutorial.html).
- [62] SQLite. Command line shell for sqlite. Disponible en la web: <https://www.sqlite.org/cli.html>.
- [63] SQLite. Sqlite encryption extension. Disponible en la web: <http://www.sqlite.org/see/doc/trunk/www/readme.wiki>.
- [64] Sqlite. Sqlite foreign key support. Disponible en la web: [https://www.sqlite.org/foreignkeys.html#fk\\_schemacommands](https://www.sqlite.org/foreignkeys.html#fk_schemacommands).
- [65] Unix StackExchange. How can i configure syslog.conf file, to log iptables messages in a separate file? Disponible en la web: <http://unix.stackexchange.com/questions/96484/how-can-i-configure-syslog-conf-file-to-log-iptables-messages-in-a-separate-fil>.
- [66] Supervisor. How to configure python script to run as a daemon. Disponible en la web: <http://serverfault.com/questions/501071/how-to-configure-python-script-to-run-as-a-daemon>.



- 
- [67] Supervisor. Supervisor: A process control system. Disponible en la web: <http://supervisord.org/>.
  - [68] tom's IT PRO. A guide to security information and event management. Disponible en la web: <http://www.tomsitpro.com/articles/siem-solutions-guide,2-864.html>.
  - [69] tom's IT PRO. A guide to security information and event management. Disponible en la web: <http://www.tomsitpro.com/articles/siem-solutions-guide,2-864.html>.
  - [70] Libros web. Python - polimorfismo. Disponible en la web: [https://librosweb.es/libro/algoritmos-python/capitulo\\_15/polimorfismo.html/](https://librosweb.es/libro/algoritmos-python/capitulo_15/polimorfismo.html/).
  - [71] Tornado Web. Tornado web server. Disponible en la web: <http://www.tornadoweb.org/en/stable/>.
  - [72] Wikipedia. List of tcp and udp port numbers. Disponible en la web: [https://en.wikipedia.org/wiki/List\\_of\\_TCP\\_and\\_UDP\\_port\\_numbers](https://en.wikipedia.org/wiki/List_of_TCP_and_UDP_port_numbers).
  - [73] Zetcode. Inserting, updating, and deleting data in sqlite. Disponible en la web: <http://zetcode.com/db/sqlite/datamanipulation/>.



# GNU Documentation Free License

Version 1.3, 3 November 2008

Copyright © 2000, 2001, 2002, 2007, 2008 Free Software Foundation, Inc.

[<http://fsf.org/>](http://fsf.org/)

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

## Preamble

The purpose of this License is to make a manual, textbook, or other functional and useful document “free” in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondly, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of “copyleft”, which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

## 1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The “**Document**”, below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as “**you**”. You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A “**Modified Version**” of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A “**Secondary Section**” is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document’s overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a

Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The “**Invariant Sections**” are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The “**Cover Texts**” are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A “**Transparent**” copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not “Transparent” is called “**Opaque**”.

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The “**Title Page**” means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, “Title Page” means the text near the most prominent appearance of the work’s title, preceding the beginning of the body of the text. The “**publisher**” means any person or entity that distributes copies of the Document to the public.

A section “**Entitled XYZ**” means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as “**Acknowledgements**”, “**Dedications**”, “**Endorsements**”, or “**History**”.) To “**Preserve the Title**” of such a section when you modify the Document means that it remains a section “Entitled XYZ” according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

## 2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this

License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

### 3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

### 4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.

- B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
- C. State on the Title page the name of the publisher of the Modified Version, as the publisher.
- D. Preserve all the copyright notices of the Document.
- E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
- F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
- G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- H. Include an unaltered copy of this License.
- I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
- K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
- L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
- M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.
- N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.
- O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled “Endorsements”, provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

## **5. COMBINING DOCUMENTS**

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled “History” in the various original documents, forming one section Entitled “History”; likewise combine any sections Entitled “Acknowledgements”, and any sections Entitled “Dedications”. You must delete all sections Entitled “Endorsements”.

## **6. COLLECTIONS OF DOCUMENTS**

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

## **7. AGGREGATION WITH INDEPENDENT WORKS**

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an “aggregate” if the copyright resulting from the compilation is not used to limit the legal rights of the compilation’s users beyond what the individual works permit. When the Document is included in

an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

## 8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

## 9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License.

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it.

## 10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

See <http://www.gnu.org/copyleft/>.



Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License “or any later version” applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of this License can be used, that proxy’s public statement of acceptance of a version permanently authorizes you to choose that version for the Document.

## 11. RELICENSING

“Massive Multiauthor Collaboration Site” (or “MMC Site”) means any World Wide Web server that publishes copyrightable works and also provides prominent facilities for anybody to edit those works. A public wiki that anybody can edit is an example of such a server. A “Massive Multiauthor Collaboration” (or “MMC”) contained in the site means any set of copyrightable works thus published on the MMC site.

“CC-BY-SA” means the Creative Commons Attribution-Share Alike 3.0 license published by Creative Commons Corporation, a not-for-profit corporation with a principal place of business in San Francisco, California, as well as future copyleft versions of that license published by that same organization.

“Incorporate” means to publish or republish a Document, in whole or in part, as part of another Document.

An MMC is “eligible for relicensing” if it is licensed under this License, and if all works that were first published under this License somewhere other than this MMC, and subsequently incorporated in whole or in part into the MMC, (1) had no cover texts or invariant sections, and (2) were thus incorporated prior to November 1, 2008.

The operator of an MMC Site may republish an MMC contained in the site under CC-BY-SA on the same site at any time before August 1, 2009, provided the MMC is eligible for relicensing.

## ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

Copyright © YEAR YOUR NAME. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled “GNU Free Documentation License”.

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the “with . . . Texts.” line with this:

with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.