▪ Màster en Seguretat de les Tecnologies
de la Informació i de les Comunicacions (MISTIC) ▪

Detecció d'anomalies amb Elastic Stack

TFM-Ad hoc (aplicació professional)

Autor: Miguel Ángel Flores Terrón
Tutor: Jordi Guijarro Olivares
Professor: Pau del Canto Rodrigo

Juny, 2017

Índex

Llista de Figures

Llista de Taules

Capítol 1
Introducció

L'objectiu del projecte es disposar d'un quadre de comandament o Dashboard que permeti identificar ràpidament anomalies (malware) o infraccions a la xarxa analitzant les dades dels diferents esdeveniments (logs). El processament de les dades es realitzaria amb el stack ELK [1] (Elasticsearch, Logstash i Kibana) això permetrà disposar d'una eina per descobrir amenaces emergents, campanyes APT, ransomware i DNS tunneling.

1.1 Origen

Uns dels pilars en el qual es basa la gestió de riscos de seguretat de la informació es, sense cap mena dubte, la anàlisis i la gestió de logs i la correlació d'esdeveniments, el que habitualment s'anomena com SIEM. La confluència d'aquest pilar de la seguretat amb d'altres permetrà al responsable de seguretat TIC assolir l'objectiu específic de saber en temps real què està succeint en els seus sistemes d'informació i quines coses poden ser rellevants per a la seguretat de les seves dades i per tant, en definitiva, per al seu negoci i activitats de la seva entitat.

Figura 1 Cicle de vida de la gestió de logs

1.2 Descripció del projecte

Aquest projecte té com a objectiu principal l'estudi, anàlisis i posterior implementació d'una plataforma per monitoritzar i centralitzar logs amb programari open source, concretament es vol utilitzar el stack ELK (Elasticsearch, Logstash i Kibana) amb la finalitat de ser capaç de detectar anomalies en els sistemes informàtics, aquesta plataforma es vol centrar en detectar anomalies especifiques a través dels esdeveniments en les consultes DNS, donat són objectius estratègics per als ciberdelinqüents, si a això li sumem que són una peça imprescindible en el funcionament correcte de les comunicacions, arribem a la conclusió que la seva seguretat i disponibilitat són essencials.

1.3 Motivació

El projecte pretén desenvolupar una plataforma o eina que sigui d'utilitat per l'Equip de Resposta a Incidents de l'Anella Científica (CSUC-CSIRT) del Consorci de Serveis Universitaris de Catalunya (CSUC) [2] en qual treballo com Administrador de Sistemes dins la unitat d'Operacions i Seguretat.

L'Equip de Resposta a Incidents de l'Anella Científica (CSUC-CSIRT) ajuda les institucions a millorar la seguretat a les seves xarxes, tant detectant possibles incidents com ajudant un cop s'han produït.

Coordina i gestiona la resolució d'incidents de seguretat TIC a l'Anella Científica i proporciona un punt de contacte per reportar, identificar, analitzar l'impacte i les amenaces del què succeeix, així com proposar solucions i estratègies de mitigació.

CSUC-CSIRT també difon les notificacions crítiques d'alerta davant d'amenaces imminents a través de les llistes de distribució i dóna suport tècnic en tecnologies de la seguretat informàtica (anàlisi de tràfic, seguretat al perímetre, etc.),

1.4 Definició d'objectius

Els principals objectius plantejats per aquest projecte son els descrits a continuació:

- Formació i conceptes teòrics en gestió de logs i DNS.
- Coneixement de l'estat de l'art de les plataformes SIEM.
- Dissenyar e implementar una plataforma de gestió de logs amb el Stack ELK.
- Descripció del funcionament i proves funcionals de la plataforma.
- Redacció de la documentació i memòria del treball.

1.5 Planificació

El projecte s'ha planificat en quatre etapes, cadascuna finalitza amb una fita que es correspon amb el lliurament de la PAC corresponent: la primera fase, (**Fase I. Planificació i Estudi),** ha consistit en realitzar la planificació inicial i la realització del pla treball, també s'ha realitzat un estudi sobre les tecnologies a emprar i l'estat de l'art per poder realitzar-lo satisfactòriament, principalment aprendre conceptes sobre la gestió de logs i la plataforma ELK en concret.
La segona fase (**Fase II. Anàlisis i Especificació)** on s'ha de especificar i dissenyar la solució i l'arquitectura que tindrà la plataforma que s'implementarà en la Fase III, en aquesta fase també s'analitzarà el cas d'ús específic del CSUC on es vol implementar la solució.

Durant la tercera fase (**Fase III. Implementació de la solució)** es durà a terme la implementació de la plataforma, amb la parametrització correcta al cas d'us específic estudiat i es duran a terme les primeres proves funcionals de la plataforma.
Finalment, la quarta fase (**Fase IV. Entrega i Documentació**) es dedicarà a l'entrega i lliurament de la documentació corresponent i la posterior defensa del projecte.

En la següent Taula 1 , es mostra un llistat detallat de totes les tasques definides en la planificació inicial prevista per a la realització del projecte.

	Nombre de tarea	Duraci	Comienzo	Fin	Predecesoras
1	**FASE I. Planificació i Estudi**	**14 días**	**mié 22/02/17**	**lun 13/03/17**	
2	101. Definició del problema	2 días	mié 22/02/17	jue 23/02/17	
3	102. Establiment dels objectius projecte	2 días	vie 24/02/17	lun 27/02/17	2
4	103. Elaboració PAC1 (Pla de treball)	7 días	mar 28/02/17	mié 08/03/17	3
5	104. Formació i conceptes teòrics en gestió de logs i DNS	10 días	mar 28/02/17	lun 13/03/17	3
6	105. Estudi d'estat de l'art dels sistemes SIEM	10 días	mar 28/02/17	lun 13/03/17	3
7	**F01. Lliurament PAC1 (Pla de Treball)**	**0 días**	**lun 13/03/17**	**lun 13/03/17**	
8	**FASE II. Anàlisis i Especificació**	**20 días**	**mar 14/03/17**	**lun 10/04/17**	**1**
9	201. Estudi del cas d'ús del CSUC	7 días	mar 14/03/17	mié 22/03/17	6
10	202. Especificació de l'arquitectura	7 días	jue 23/03/17	vie 31/03/17	9
11	203. Disseny de la infraestructura de la plataforma	5 días	lun 03/04/17	vie 07/04/17	10
12	205. Revisió de planificació	1 día	lun 10/04/17	lun 10/04/17	11
13	**F02. Lliurament PAC2**	**0 días**	**lun 10/04/17**	**lun 10/04/17**	
14	**FASE III. Implementació de la solució**	**25 días**	**mar 11/04/17**	**lun 15/05/17**	**8**
15	301. Instal·lació i configuració de la plataforma ELK	7 días	mar 11/04/17	mié 19/04/17	12
16	302. Parametrització i adaptacions de la solució	5 días	jue 20/04/17	mié 26/04/17	15
17	303. Aprenentatge funcional de la plataforma	5 días	jue 27/04/17	mié 03/05/17	16
18	304. Proves funcionals	3 días	jue 04/05/17	lun 08/05/17	17
19	**F03. Lliurament PAC3**	**0 días**	**lun 08/05/17**	**lun 08/05/17**	
20	305. Posada en producció	5 días	mar 09/05/17	lun 15/05/17	18
21	**FASE IV. Entrega i Documentació**	**35 días**	**lun 08/05/17**	**vie 23/06/17**	**19**
22	401. Redacció de la memoria	21 días	lun 08/05/17	lun 05/06/17	16
23	402. Elaboració de la presentació final	4 días	mar 06/06/17	vie 09/06/17	22
24	403. Gravació de la presentació final	1 día	lun 12/06/17	lun 12/06/17	22;23
25	**F04. Lliurament final**	**0 días**	**lun 05/06/17**	**lun 05/06/17**	
26	404. Tribunal de TFM	5 días	lun 19/06/17	vie 23/06/17	24
27					

Taula 1 Planificació inicial prevista del projecte

En l'actualitat, el projecte es troba situat a l'inici de la segona fase (**Fase II. Anàlisis i Especificació**), on caldrà definir i especificar l'arquitectura escollida per donar resposta al problema plantejat i a l'espera de la aprovació definitiva del Pla de treball per tirar endavant el projecte. Dins del projecte, s'han definit quatre fites a assolir, cadascuna es correspon als lliuraments de les diferents proves d'avaluació continuada (PAC2, PAC3 i PAC4) i per últim el lliurament final del projecte, amb totes aquestes dades, la previsió es defensar el projecte al mes de Juny.

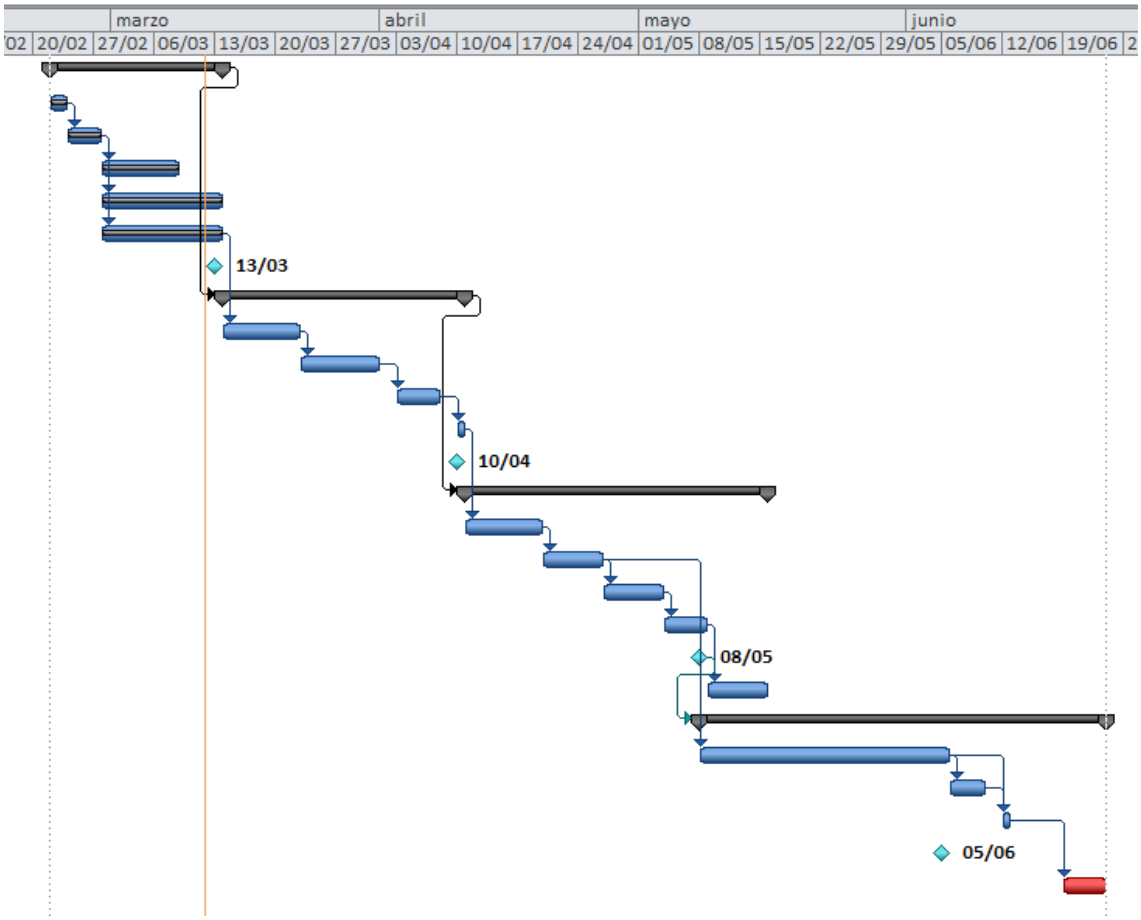


Figura 2 Diagrama de Gantt del projecte

1.6 Estat de l'art de les eines de gestió de logs

En aquest apartat analitzarem les eines o programaris més populars i utilitzats que existeixen en l'actualitat per a la gestió de logs.

Splunk [3]

Programari per cercar, monitoritzar i analitzar dades generades per màquines (Big Data) d'aplicacions, sistemes i infraestructura IT a través d'una interfície web. Splunk captura, indexa i correla en temps real, emmagatzemant-ho tot en un repositori on cerca per generar gràfics, alertes i panells fàcilment definibles per l'usuari.

Figura 3 Logo plataforma Splunk

ELK [1]

El stack ELK és un paquet de tres eines Open Source de l'empresa Elastic. Les eines són Elasticsearch, Logstash i Kibana. Aquestes tres eines són projectes independents i poden ser usades per separat.

Elasticsearch és un servidor de recerca basat en Lucene. Proveeix un motor de cerca de text complet (full-text), a través d'una interfície web RESTful. Mitjançant peticions HTTP podem emmagatzemar informació de forma estructurada en Elasticsearch perquè aquest la indexi, i posteriorment poder fer cerques sobre ella.

Logstash és una eina per a l'administració de logs. S'encarrega de recol · lectar, parsejar i filtrar els logs per posteriorment donar-los alguna sortida com, emmagatzemar-los en MongoDB, enviar-los per correu electrònic o guardar-los en Elasticsearch. Aquests logs li poden arribar a Losgstash des del mateix servidor o des d'un servidor extern, de manera que podríem tenir un servidor exclusiu per al stack ELK. L'aplicació es troba basada en JRuby i requereix de Java Virtual Machine per executar-la.

Kibana és una eina analítica de codi obert (llicència Apache) que ens permetrà interactuar amb la informació emmagatzemada (per Logstash) en Elasticsearch i monitoritzar-la.

Figura 4 Logos stack ELK

LogTrust [4]

Logtrust s'especialitza en oferir solucions en temps real per a grans volums de dades que permetin la integració, la gestió i la fàcil visualització de totes les dades generades. Logtrust tracta tota la informació com una base de dades, permet analitzar tots els registres, però sense normalització.

Figura 5 Logo plataforma LogTrust

Graylog [5]

Graylog (anteriorment coneguda com Graylog2) és una plataforma de gestió de syslog de codi obert, que ajuda a recollir, analitzar, indexar syslog en una ubicació centralitzada.

Figura 6 Logo plataforma graylog

Nagios Log Server [6]

És un programari que simplifica el procés de cerca. Aquesta eina permet ajustar les alertes per notificar quan ocorren amenaces potencials, o simplement fer consultes a registre de dades per ràpidament auditar el sistema. Ofereix un

entorn centralitzat on es tindran tots els logs de dades, amb alta disponibilitat i suport de *fail-over*. La versió lliure té la limitació de 500 MB/dia.

Figura 7 Logo plataforma Nagios Log Serve

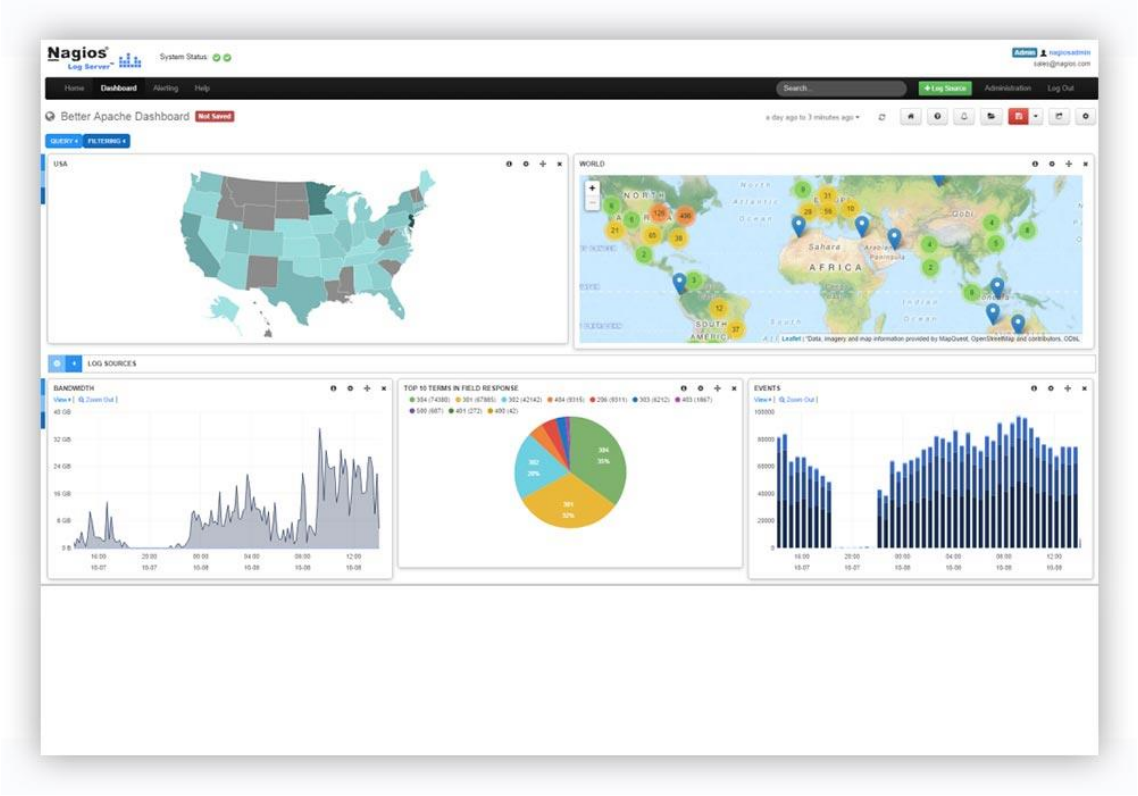


Figura 8 Dashboard plataforma Nagios Log Server

Capítol 2
Fonaments del DNS

En aquest capítol, es descriuen els elements que integren una infraestructura DNS [7]:

2.1 Qué es DNS?

Domain Name System (DNS) és un sistema globalment distribuït, escalable i jeràrquic. Ofereix una base de dades dinàmica associant adreces IP d'equips, serveis o qualsevol recurs connectat a internet o xarxa privada amb informació de divers tipus. Suporta tant IPv4 com IPv6, i la informació s'emmagatzema en forma de registres Resource Records (RR) de diferents tipus els quals poden emmagatzemar adreces IP o un altre tipus d'informació. Aquesta informació s'agrupa en zones, que corresponen a un espai de noms o domini i que són mantingudes pel servidor DNS autoritari de la mateixa.

Fonamentalment, DNS s'encarrega de traduir adreces IP de recursos de xarxa a noms fàcilment llegibles i memoritzables per les persones, i viceversa. Aquesta acció és coneix amb el nom de "resolució DNS".

DNS utilitza per a les comunicacions el port 53, tant per a datagrames UDP com per a paquets TCP. Generalment, en l'activitat DNS s'usen datagrames UDP ja que requereixen menys recursos de procés i de xarxa.

2.2 Elements integrants d'un DNS

- **Espai de dominis de noms**: Consisteix en un estructura jeràrquica d'arbre on cada node conté zero o més registres (Resource Records, o RR) amb informació del domini. Del node arrel, situat en el nivell més alt, parteixen les branques que conformen les esmentades zones. Aquestes, al seu torn, poden contenir un o més nodes o dominis que al seu torn poden dividir-se en subdominis segons es baixa en la jerarquia.

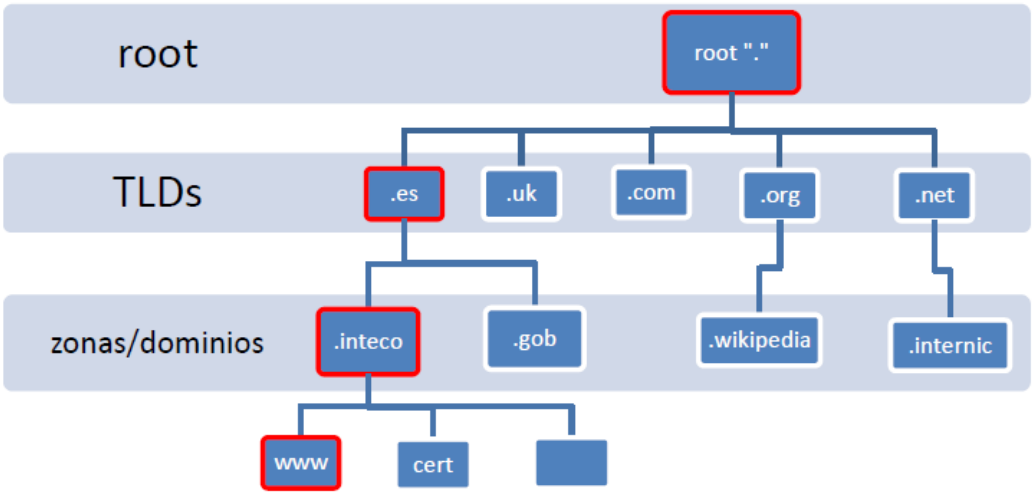


Figura 9 Jerarquia de l'espai de noms

- **Servidors de Noms**: Són servidors encarregats de mantenir i proporcionar informació de l'espai de noms o dominis. D'una banda, existeixen servidors que emmagatzemen informació completa per a un o diversos conjunts de l'espai de noms (dominis) i de les quals és responsable. Es diu que són servidors autoritatius d'aquestes zones/dominis en qüestió.

- **Resolvers**: Són servidors caché o programes client els quals s'encarreguen de generar les consultes necessàries i obtenir la informació sol · licitada per oferir-la a l'usuari que la sol · licita.

2.3. Registres DNS. Format i tipus

Un nom de domini s'identifica amb un node en la jerarquia DNS. Cada node conté un conjunt d'informació conegut com a registres (Resource Registers, RR) dels quals és responsable o autoritat.

Aquesta informació és formatada en un registre que es compon de 6 camps, que s'utilitza en transmetre aquesta informació en els missatges DNS. En la següent taula es descriuen els 6 possibles camps en un missatge DNS

Camp	Descripció	Longitud (bytes)
NAME	Nom del domini al qual pertany el registre	Cadena variable
TYPE	Codi del tipus de registre	2 bytes
CLASS	Codi de la classe del registre	2 bytes
TTL	Temps en segons durant el qual el registre és	4 bytes
RDLENGTH	Indica la longitud en bytes del camp RDATA	4 bytes
RDATA	Cadena de longitud variable que descriu el registre d'acord al tipus i classe del mateix	Cadena variable

Taula 2 Format de registre. Resource Record (RR)

El camp **TYPE** conté un codi que identifica de quin tipus de registre es tracta. Existeixen multitud de tipus de registres definits en diferents RFCs per cobrir d'altres funcionalitats. Alguns dels tipus més comuns es mostren en la següent taula:

Tipus (valor camp TYPE)	Funció
A = Address	Tradueix (resol) noms de recursos a adreces IPv4
AAAA = Address	Tradueix (resol) noms de recursos a adreces IPv6
CNAME = Canonical Name	Crear noms addicionals, o àlies, per al recurs
NS = Name Server	Indica quin servidor/s emmagatzema la informació del domini consultat
MX = Mail Exchange	Associa un nom de domini a una llista de servidors d'intercanvi de correu per a aquest domini.
PTR = Pointer	Inversa del registre A, traduint IPs en noms de domini.
SOA = Start of authority	Indica el servidor DNS primari de la zona, responsable del manteniment de la informació de la mateixa.
HINFO = Host INFOrmation	Descripció de la CPU i sistema operatiu que emmagatzema la informació d'un domini.
TXT = TeXT	Permet als dominis proporcionar dades addicionals.
LOC = LOCalización	Permet indicar les coordenades geogràfiques del domini.
SRV = SeRVicios	Informació sobre els serveis que oferts
SPF = Sender Policy Framework	Ajuda a combatre l'Spam. En aquest registre s'especifica quin o quins hosts estan autoritzats a enviar correu des del domini donat.

ANY = Tots	Per sol · licitar tots els registres disponibles

Taula 3 Valors més habituals camp TYPE

El camp **CLASS** és comunament fixat al valor IN (Internet) per a registres DNS relacionats amb hostnames, servidors o, en resolució inversa, adrecis IP. Existeixen a més les classes CH (Chaos) i HeSiod (HS) per a altres sistemes menys comuns.

En el camp **TTL**, un valor numèric que indica el temps en segons que s'escorcollarà el registre. Un valor 0 indica validesa només per a la transacció en curs i el registre associat no serà emmagatzemat en caché. Els registres SOA sempre tenen TTL igual a 0.

En el camp **RDATA** es descriu el contingut del registre segons el tipus indicat en el camp TYPE: SOA, A, NS, MX, etc. La mida d'aquesta informació s'indica en el camp **RDLENGTH**

2.4 Mecanisme de resolució d'una consulta DNS

El procés que se segueix en una resolució DNS és el següent. El client (resoldre) fa arribar la consulta al servidor DNS:

a) Si el servidor DNS està configurat com autoritativo i rep una consulta DNS sobre un domini sobre el qual ell és autoritativo, retornarà la resposta consultant els registres emmagatzemats en la seva configuració i retornant la resposta marcada com Authoritative Answer en la secció "ANSWER" del missatge de resposta. Si no té la informació, respon amb el missatge NXDOMAIN (Senar-Existent-Domain).

b) Si el servidor DNS és autoritativo i no configurat com recursivo i rep una consulta sobre un domini sobre el qual no és autoritativo, respondrà amb un missatge contenint registres en la secció "AUTHORITY" i en la secció ADDITIONAL informant en resoldre que no proporciona recursión i on ha de dirigir la seva consulta per obtenir informació autoritativa del domini sol · licitat. Es coneix com Referral Response.

c) Si el servidor DNS no és autoritativo, però està configurat com recursivo i rep una consulta, aquest inicia consultes iteratives (recursión) per trobar el servidor autoritativo del domini. Una vegada obté resposta retorna el registre al client (resoldre) indicant que es tracta d'una resposta no autoritativa. La informació la

guarda en caché, de manera que si torna a ser preguntat pel mateix recurs i el temps amb que el registre està marcat per "caducar" (TTL, o Estafi To Live) no ha passat, contestarà consultat aquesta caché.

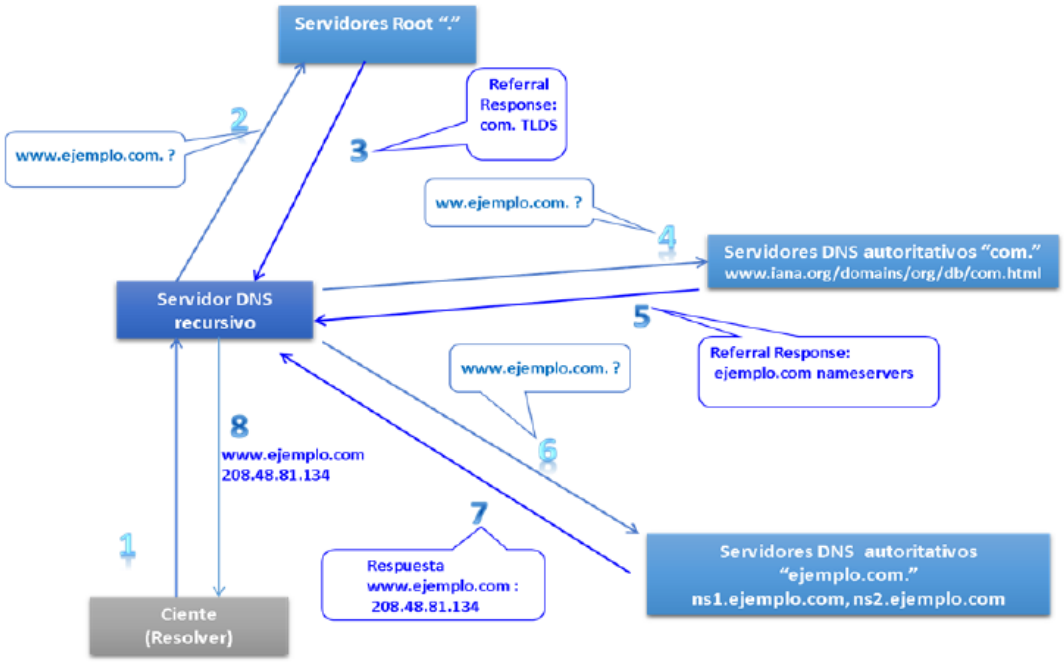


Figura 10 Successió de consultes en una resolució recursiva

2.5 Amenaces i vulnerabilitats del DNS

En un entorn DNS s'identifiquen diversos punts on possibles atacs poden desenvolupar-se. Aquests punts o "vectors d'atac" se situen tant localment en el propi servidor DNS i xarxa local, com en les comunicacions entre servidors i clients.

- **DNS CACHE POISONING i DNS SPOOFING**: Sota UDP i sense usar cap altre mecanisme de control, un atacant pot enviar multitud de respostes (flooding) amb diferents ID fins a aconseguir encertar amb l'ID generat en la consulta. Si és així, i s'aconsegueix fer arribar la resposta falsa abans que arribi la legítima (condició de carrera), el servidor que ha iniciat la consulta l'acceptarà i l'emmagatzemarà en la seva memòria cau. D'aquesta manera, és possible "enverinar" la memòria cau d'un servidor DNS recursiu amb un registre manipulat. A partir d'aquest moment, durant el temps que el registre queda emmagatzemat a la memòria cau (TTL), el servidor víctima redirigirà a una IP il·legítima totes les sol·licituds d'un resoldre que li consulti pel recurs manipulat.

El protocol DNS causa de la seva vulnerabilitat intrínseca a spoofing IP, es converteix un poderós aliat a l'hora d'implementar atacs de denegació de servei. Això, unit a la seva àmplia distribució i accés a nivell mundial fan d'aquest tipus d'atac un dels més eficaços i utilitzats.

- **Atac d'amplificació DNS**: L'ús del protocol UDP en el transport de missatges DNS, així com l'enorme quantitat de servidors recursius accessibles a internet (open resolvers) possibilita l'ús del servei per establir atacs distribuïts de denegació de servei cap a altres servidors. En un atac d'amplificació DNS es pretén desbordar la capacitat de resposta d'un servidor fent-li arribar una gran quantitat de dades DNS. El procediment consisteix a llançar consultes DNS a un open resolver falsejant la IP d'origen amb la IP del servidor / host a atacar.

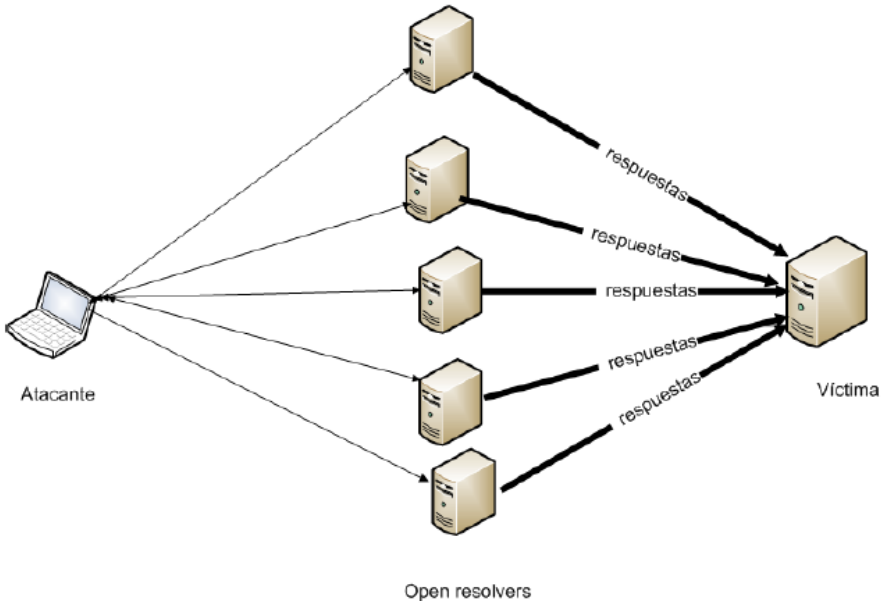


Figura 11 Atac d'amplificació DNS

- **Denegació de servei DOS**: Els atacs de denegació de servei són realment difícils d'evitar. En el cas del DNS, les seves característiques i la debilitat intrínseca del transport UDP en què es basa, fan que el mateix servei sigui una víctima en si mateixa i no un mer element col · laborador, com ocorre en el cas d'atacs d'amplificació. Donada la dificultat de localitzar i bloquejar un atac sobre UDP amb adreces IP falsificades, és important comptar amb mecanismes reactius per defensar-se d'un atac de denegació de servei quan s'és víctima final del mateix.

Capítol 3
Especificació

En aquest capítol, es descriu el cas d'ús que ens ocupa, es presenta una descripció detallada de la plataforma ELK [1] estudiada i el disseny de l'arquitectura de la plataforma que es vol implementar.

3.1 Estudi del cas d'ús del CSUC

En la actualitat el Consorci de Serveis Universitaris de Catalunya (CSUC) fa ús de la plataforma Splunk, aquesta plataforma s'utilitza fonamentalment com a eina per:

- Emmagatzemar, tractar i consultar logs. Per exemple, dels equips de xarxa (routers, commutadors), dels servidors (Supercomputació), o bé dels diferents serveis (Radius Eduroam, Correu electrònic).
- Generar gràfiques i estadístiques
- Rebre alertes en temps real, com per exemple, detectar la caiguda d'una interfície, un problema hardware o detectar connexions amb credencials incorrectes.
- Enviar informes periòdics
- Com a eina de seguretat, incorporació de fonts de dades preconfigurades de fabricants com F5 o Paloalto, incorporació de dades (feeds) d'altres fonts: Flows flowsonar o OSSEC.

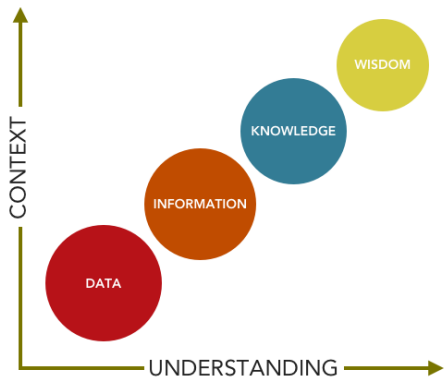


Figura 12 Piràmide del Coneixement

Aquesta solució presenta una sèrie d'avantatges i desavantatges si la comparem amb algunes plataformes similars. Aquestes es descriuen a continuació:

Avantatges

- Fàcil de posar en marxa.
- Plataforma intuïtiva
- Indexació en lectura (tradicionalment, en escriptura)
- Versió fins a 500 MB/dia gratuïta
- Permet generar informes periòdics i alarmes en temps real
- Permet incorporar fonts de dades prefabricades de fabricants
- Ofereix una solució en el Cloud.

Desavantatges

- Es de pagament (Llicencia anual d'un 1 GB/dia = 2.070 $).
- Oferir vistes a diferents usuaris no es trivial
- La configuració d'alarmes en temps real carrega la plataforma.
- Necessita un servidor potent

Actualment la plataforma corre de forma virtualitzada amb KVM dins d'una màquina virtual amb 4 vCPU i 8 GB de memòria RAM, orquestrada amb OpenNebula. Tot i això la plataforma recomanada requereix d'una màquina amb 16 GB de memòria RAM amb 12 cores i ha de permetre unes 1200 ops/s de Input/Output.

3.2 Plataforma ELK

ElasticSearch, Logstash i Kibana són projectes Open Source que ajuden a l'usuari a obtenir les dades de qualsevol font de dades, amb qualsevol format i fer una cerca i un anàlisi de les mateixes i poder visualitzar-les en temps real.
La companyia que ha desenvolupat el programaris s'anomena Elastic i va ser fundada l'any 2012, fins a l'actualitat ha tingut un gran creixement, fet que li ha permès expandir-se per les ciutats més importants del món. Els productes de llicència gratuïta que ofereix Elastic són ElasticSearch, Logstash i Kibana, però a banda d'aquests, ofereix altres eines que els hi donen suport, i que faciliten el

treball al usuari, per utilitzar aquestes altres eines ja cal disposar amb una llicència de pagament.

A continuació mostrarem quin es l'objectiu de cada component:

3.2.1 Logstash

Logstash és una aplicació Java de codi obert desenvolupada en JRuby amb l'objetiu de transportar, recol · lectar, filtrar e indexar logs. La arquitectura de LogStash està composada per tres components principals en forma de *plugins*: Input, Filter i Output.

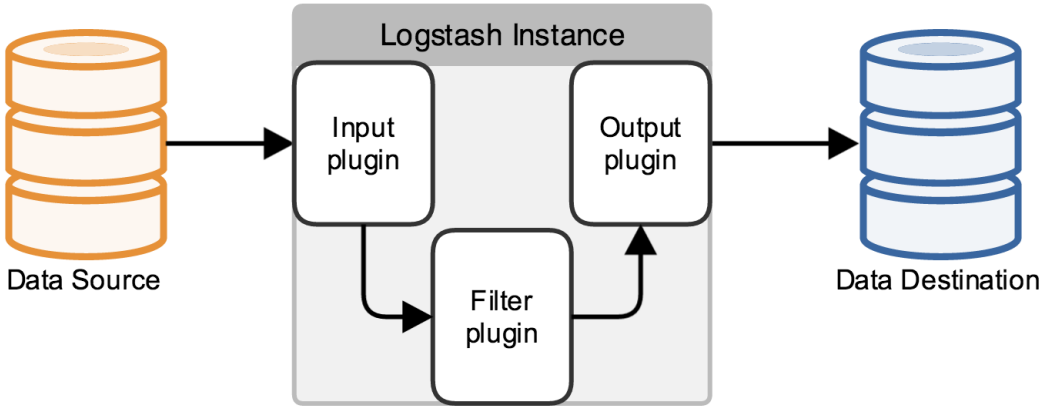


Figura 13 Estructura Logstash

3.2.1.1 Input

Existeix una gran quantitat de *plugins* que és poden utilitzar per recuperar les dades que l'usuari desitja, a continuació s'anomenaran una petita quantitat d'aquests i més endavant s'entrarà més en detall amb els *plugins* que s'ha utilitzat per dur a terme el projecte.

- **exec**: Aquest *plugin* captura el resultat d'una comanda executada des d'una *shell*.
- **file**: Aquest *plugin* ens permet obtenir les dades dels fitxers de text.
- **beats:** Rep esdeveniments des de Elastic Beats framework.
- **http:** Permet obtenir dades d'esdeveniments via http o https.
- **log4j**: Permet rebre esdeveniments sobre sockets TCP que utilitzen log4j.
- **pipe**: Amb aquest plugin podem rebre dades d'esdeveniments que estiguin executant-se utilitzant una pipe.
- **s3**: Transmet esdeveniments d'arxius en un *bucket* de S3.
- **syslog**: Llegeix els missatges de registre del sistema com esdeveniments

Es pot consultar tota la informació sobre els *plugins* d'Input a la documentació oficial de Logstash disponible en la següent URL:
https://www.elastic.co/guide/en/logstash/current/input-plugins.html

3.2.1.2 Filter

De la mateixa forma, existeixen una gran quantitat de *plugins* que és poden utilitzar per tractar les dades que rebem del Input, a continuació s'enumeren alguns dels *plugins* més interessants per la selecció de les dades:

- **aggregate**: Permet agregar diferents esdeveniments com si fos una única tasca.
- **checksum**: Crea un *checksum* dels esdeveniments per poder comprovar que no han estat modificats.
- **drop**: Eliminar esdeveniments que provinguin del input.
- **grok**: Ens permet analitzar gramaticalment el text rebut des del input i a més a més donar-li una estructura (és un dels *plugins* més utilitzats).
- **json**: Ens permet analitzar gramaticalment un text en format JSON.
- **mutate**: Permet canviar el nom de diversos paràmetres.
- **multiline**: Ens permet ajuntar diverses línies de text rebudes en un sol esdeveniment.
- **uuid**: Ens afegeix un ID únic per a cada esdeveniment.
- **xml**: Ens permet analitzar gramaticalment un text en format XML.

Tal i com ja hem comentat en l'apartat anterior existeixen molts altres *plugins* que és poden trobar detallats en la documentació oficial de Logstash en l'apartat dels *plugins* Filter en la URL:

https://www.elastic.co/guide/en/logstash/current/filter-plugins.html

3.2.1.3 Output

Finalitzarem esmentant els *plugins* que disposem per enviar les dades al seu destí final, un cop han estat recol · lectats i processats:

- **csv**: Escriu les dades en un fitxer amb format csv.
- **cloudwatch**:t Envia les mètriques cap a AWS CloudWatch.
- **elasticSearch**: Ens permet enviar les dades a ElasticSearch.
- **file**: Guarda les dades en un arxiu de text.

- **zabbix**: Permet enviar les dades a Zabbix.
- **s3**: Envia esdeveniments de Logstash a l'Amazon Simple Storage Service (S3).

La resta de *plugins* disponibles es pot trobar a la documentació oficial de Logstash dins la secció dels *plugins* Output, consultant el següent enllaç:

https://www.elastic.co/guide/en/logstash/current/output-plugins.html

3.2.2 ElasticSearch

ElasticSearch és una base de dades no relacional (NoSQL) d'emmagatzematge amb funcions incorporades de cerca de text i anàlisi de dades.

Les principals funcionalitats que ofereix ElasticSearch inclouen: accés i anàlisi a les dades en temps real, escalabilitat a través d'una arquitectura distribuïda, alta disponibilitat, múltiples índexs (Multitenancy), orientació a documents (JSON), interfície per al desenvolupament d'aplicacions (RESTful API).

El clúster és la unitat de servei en Elasticsearch, és a dir, és una base de dades completa a la qual es connecten els clients per emmagatzemar documents i realitzar cerques. Un clúster està estructurat en una part "lògica" i una part "física".

La part lògica està representada pels **índexs**, que no són més que particions lògiques de les dades basades en els criteris de les aplicacions que emmagatzemen la informació, una mateix dada (registre, document, objecte, etc...) no pot existir en dos índexs en el sentit d'identitat; però sí que poden existir dos documents iguals en contingut.

Per crear un índex podem fer servir la API REST, que segueix el següent patró:

```
http://localhost:9200/<index>/<tipus>/[<id>]
```

L'índex es crea si no existeix, així com el tipus. L'identificador és opcional, i si no es proporciona Elasticsearch assignarà un de forma automàtica.

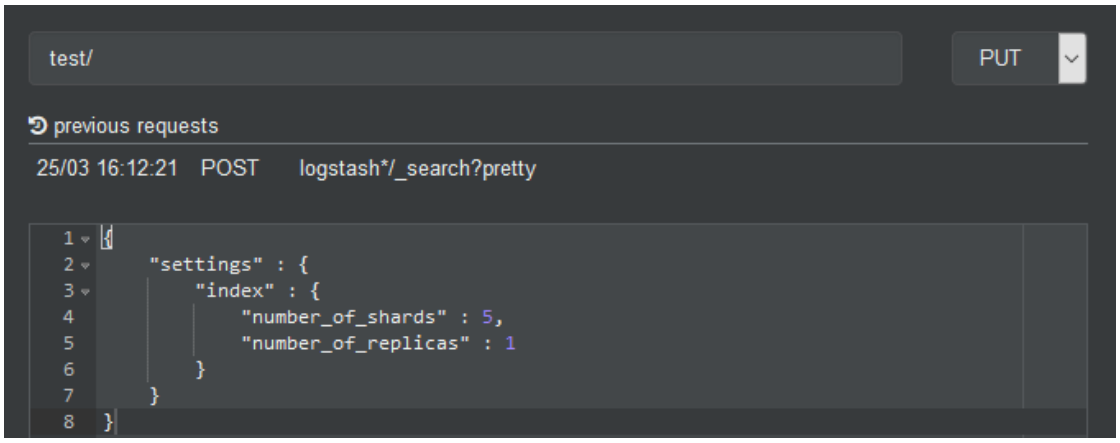


Figura 14 Exemple de creació d'un índex

La part física són els **nodes** del clúster, on cada node és una màquina virtual Java executant una instància del servei.

Elasticsearch distingeix diversos tipus de nodes:

- **master**, tenen com a responsabilitat gestionar el clúster i assegurar la seva integritat. Perquè un clúster pugui funcionar ha de tenir almenys `minimum_master_nodes` nodes d'aquest tipus (per defecte 1).
- **master-eligible** són nodes candidats a ser master en cas que sigui necessari, per exemple si cau algun master actiu o el clúster està inicialitzant-se i encara no s'han seleccionat què nodes van a actuar com a master.
- **data**, són els nodes normals que contenen les dades i executen les cerques.
- **client** és un node que ni és master ni conté dades, així que la seva única funció és enrutar peticions dins del clúster i com a molt agregar dades de consultes distribuïdes.
- **tribe**, compleixen una funció de façana, agregant diversos clústers de manera transparent.

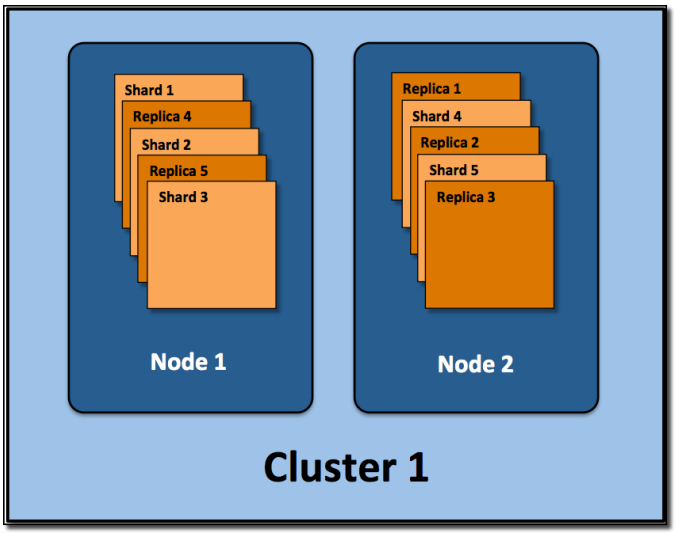


Figura 15 Estructura ElasticSearch

Els **shards** és on realment es realitzen les operacions d'indexació i cerca en Elasticsearch i és la unitat de distribució de treball en el clúster.

Internament és una instància d'Apache Lucene [7] amb les seves dades, metadades i índexs (no confondre amb el concepte d'índex de Elasticesarch).

Cada índex té un nombre fixat i predeterminat de shards primàries, que són les fonts que contenen la informació emmagatzemada i indexada en Elasticsearch.

L'única manera d'afegir o treure shards primàries és recrear l'índex (reindexar), que pot ser bastant costós. La configuració òptima és un shard primària per node i índex.

Els **shards de rèpliques**: són còpies que es distribueixen pels nodes del clúster per aconseguir major rendiment, alta disponibilitat i backup i que sí es poden afegir i treure en qualsevol moment.

3.2.2.1 Cerques

Exemples del cerques que podem realitzar en ElasticSearch utilitzant la seva API REST:

Figura 16 Cerca d'exemple en ElasticSearch

El resultat d'aquesta cerca tindria la següent sortida:


```
{ -
  "took": 27,
  "timed_out": false,
  "_shards": { -
    "total": 5,
    "successful": 5,
    "failed": 0
  },
  "hits": { -
    "total": 1,
    "max_score": 0.2876821,
    "hits": [ -
      { -
        "_index": "test",
        "_type": "prova",
        "_id": "1",
        "_score": 0.2876821,
        "_source": { -
          "username": "mflores2",
          "assignatura": "tfm",
          "prova": "pac2"
        }
      }
    ]
  }
}
```


Figura 17 Resultat d'una cerca en ElasticSearch

3.2.2.2 Monitorització del clúster

Cerebro [8] és l'evolució de l'anterior plugin d'Elasticsearch, **kopf** (https://github.com/lmenezes/elasticsearch-kopf) que no funciona en Elasticsearch 5.x o superior a causa de la eliminació dels plugins del lloc.

Amb aquesta eina podem obtenir dades per monitoritzar els recursos consumits per el clúster i els nodes.

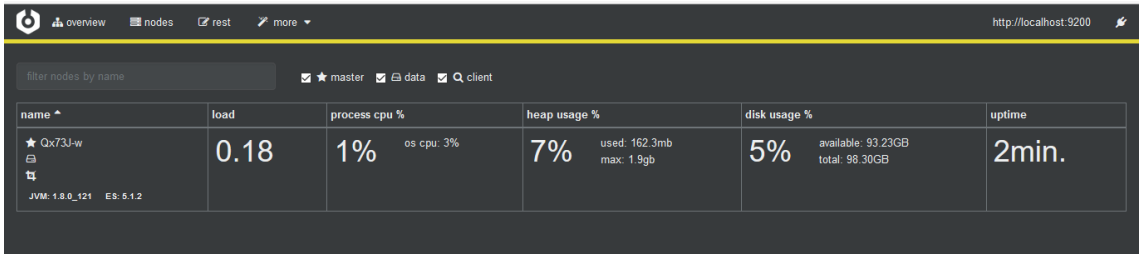


Figura 18 Plugin cerebro

3.2.3 Kibana

Kibana és una eina de visualització i exploració de dades. Entre les principals característiques que presenta es troben: integració completa amb ElasticSearch, vistes personalitzades, anàlisi incorporada i suport multiorigen.

Els usuaris podem escollir entre les següents opcions, per visualitzar les dades:

- **Area Chart**: Ens crea una gràfica d'àrea, basada en les dades que seleccionem.
- **Data Table**: Ens crea una taula amb les dades que facilitem.
- **Line Chart**: Ens crea una gràfica de línies, basada en les dades que seleccionem.
- **Markdown Widget**: Ens permet introduir qualsevol text.
- **Metric**: Ens permet incloure números de les dades que seleccionem.
- **Pie Chart**: Ens crea una gràfica de sectors basada en les dades que seleccionem.
- **Tag cloud**: Tag o núvols de paraules mostren una col · lecció de paraules, termes o frases petites, disposades totes adjacents entre si.
- **Tile Map**: Ens crea un gràfic situant les dades que seleccionem sobre el mapamundi.
- **Timeseries**: Calcula i combina dades de múltiples conjunts de dades de sèries de temps.
- **Vertical Bar Chart**: Ens crea una gràfica de barres, basada en les dades que seleccionem.

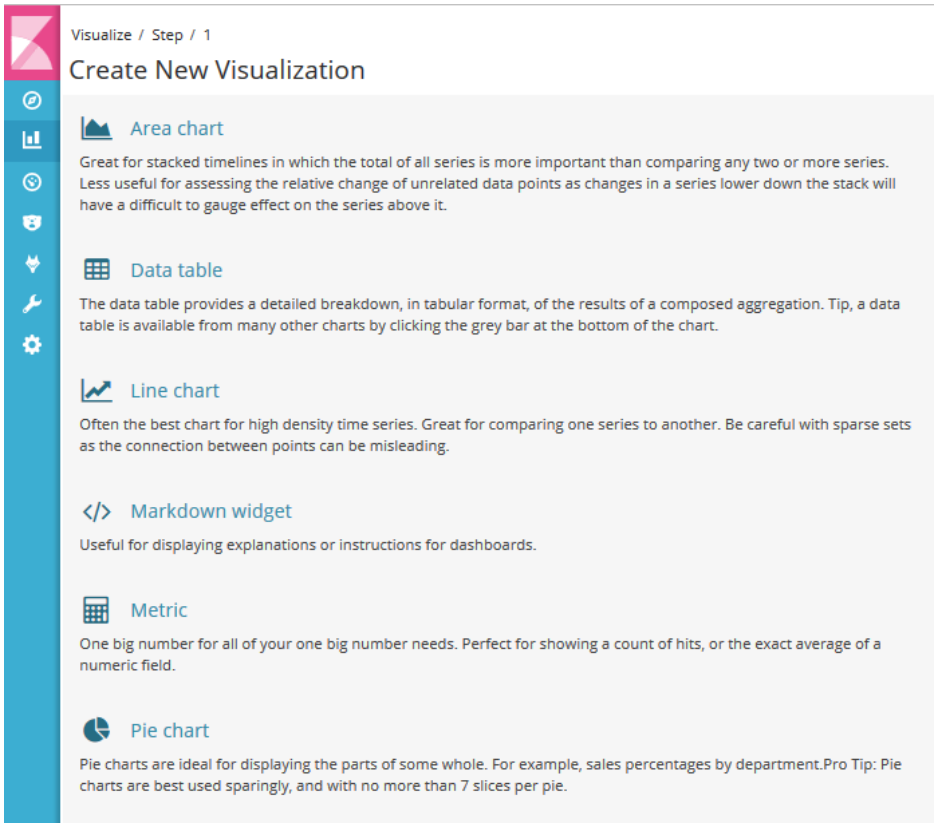


Figura 19 Tipus de visualització en Kibana

Kibana també permet muntar quadres de comandaments o *Dashboards* a mida segons les nostres necessitats.

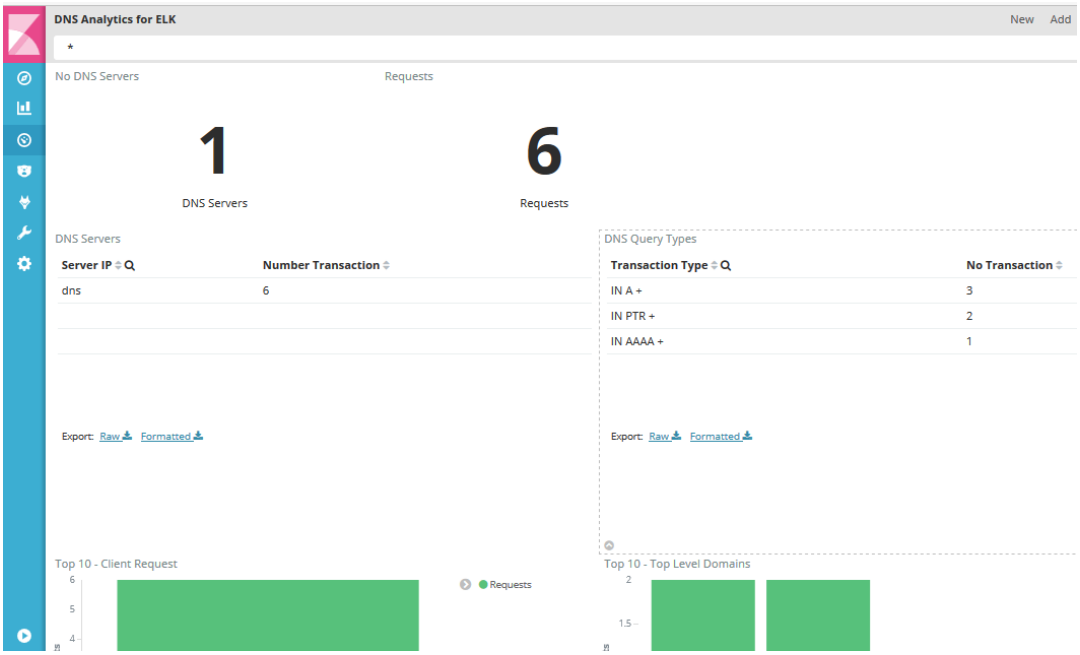


Figura 20 Exemple de Dashboard en Kibana

3.3 Disseny de l'arquitectura del laboratori

Inicialment el projecte el desplegarà en un entorn de laboratori amb màquines virtuals, orquestrat per la plataforma de IaaS: OpenNebula [9], aquest entorn servirà per establir les mètriques necessàries de la plataforma per un futur pas a producció.

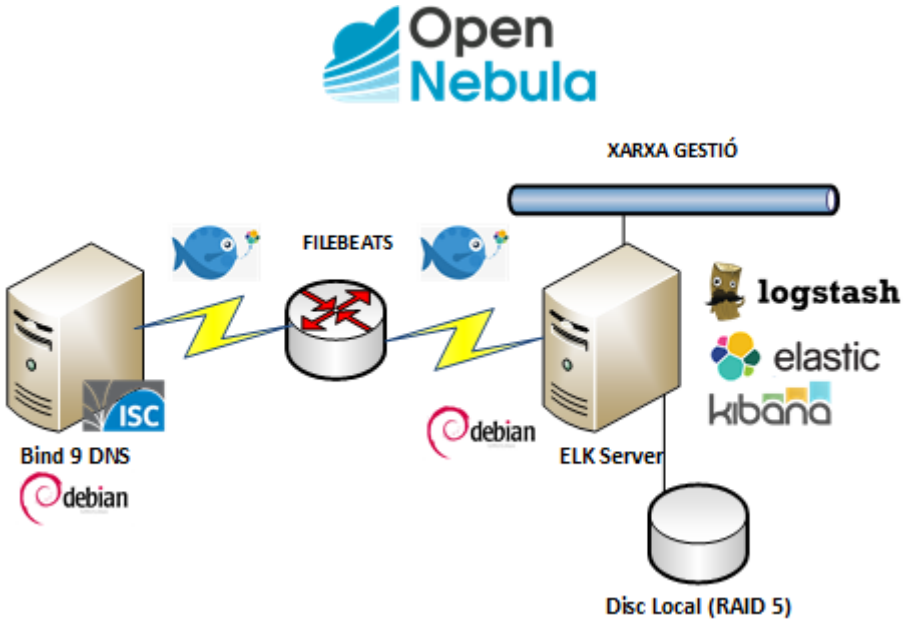


Figura 21 Esquema de l'arquitectura de la plataforma

La plataforma es desplegarà inicialment en una única màquina virtual amb 4 vCPU i 8 GB de RAM, mateixos requeriments hardware que l'actual plataforma de gestió de logs amb Splunk, cosa que permetrà fer una comparativa de rendiment entre les dues plataformes. El sistema operatiu escollit es la distribució Debian 8.0 " *jessie*", el servidor DNS utilitza com a programari ISC Bind 9 [11]

Inicialment s'havia plantejat s'emmagatzemar les dades en un volum NFS per disposar d'una protecció addicional i facilitar així la creació de copies de seguretat o instantànies (*snapshots*), finalment atenent als requeriments d'ElasticSearch es va descartar aquesta opció i es va obtar per utilitzar disc local de la pròpia màquina, amb una protecció de RAID 5 per obtenir un millor rendiment de la plataforma.

Donat el gran volum de dades diari que es preveu que emmagatzemi la plataforma, s'hauria de plantejar la possibilitat d'emmagatzemar la rotació del índexs d'ElasticSearch en un repositori extern, com per exemple un *bucket* de S3 (Simple Storage Service) fent servir l'eina de ElasticSearch: **Curator** [12].

Finalment, com podem veure en el diagrama de l'arquitectura només farem servir una única xarxa de gestió i per enviar les dades (logs) del servidor DNS cap a la plataforma ELK s'ha obtat per utilitzar **Filebeats** [13].

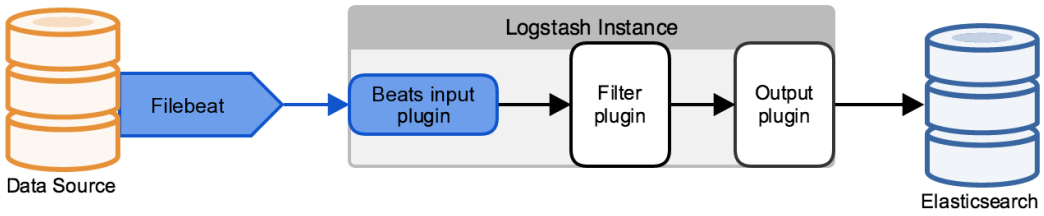


Figura 22 Pipeline bàsic de Logstash amb Filebeat

Filebeat és l'eina (també desenvolupada per Elastic) que s'utilitza en els servidors clients per constantment enviar els seus arxius de logs al servidor ELK.

Capítol 4
Implementació

En el capítol 4, es descriu la implementació de la plataforma ELK [1]

4.1 Implementació de la plataforma ELK

Un cop tenim definit el nostre objectiu, l'abast, requeriments del projecte i disseny, ja podem començar a implementar-lo. A continuació és començarà descriure el procés seguit per a realitzar el desplegament eficient de la plataforma ELK detallant les parts més significatives de la seva configuració.

4.1.1 Instal · lació i configuració Logstash

El primer pas per desplegar la plataforma ELK serà instal · lar els requeriments un dels més significatius és la instal · lació de Oracle JDK 8:

```
# apt-get install software-properties-common
# add-apt-repository "deb
http://ppa.launchpad.net/webupd8team/java/ubuntu xenial main"
# apt-get update
# apt-get install oracle-java8-installer
```

La forma més fàcil per dur a terme la instal · lació dels components necessaris per desplegar la plataforma és afegir els repositoris d'Elastic al nostre servidor, es poden afegir de la següent manera:

```
# wget -qO - https://artifacts.elastic.co/GPG-KEY-elasticsearch | apt-
key add -
# apt-get install apt-transport-https
# echo "deb https://artifacts.elastic.co/packages/5.x/apt stable main"
| tee /etc/apt/sources.list.d/elastic-5.x.list
# apt-get update
```

Ara només ens queda instal · lar el paquet corresponent:

```
# apt-get install logstash
```

4.1.1.1 Configuració dels logs ISC Bind

Farem sevir Filebeats [10] per supervisar els arxius de registre generats pel servidor ISC BIND. Així doncs caldrà configurar els canals de registres en el arxiu de configuració del servidor ISC BIND (`/etc/bind/named.conf.options`) en el servidor DNS.

```
logging{
        channel query_log {
            file "/var/log/named/query.log";
            severity info;
            print-time yes;
            print-severity yes;
            print-category yes;
        };

        category queries {
            query_log;
        };
};
```

Figura 23 Configuració logs ISC Bind

4.1.1.2 Instal · lació i configuració Filebeat

Filebeat és un recol · lector d'arxius de logs de codi obert, l'utilitzarem per enviar les dades i alimentar a Logstash. És el reemplaçament per *logstash-forwarder*. Per instal · lar Filebeats només cal descarregar el següent paquet:

```
# curl -L -O
https://artifacts.elastic.co/downloads/beats/filebeat/filebeat-5.1.2-
amd64.deb

# dpkg -i filebeat-5.1.2-amd64.deb
```

La configuració de Filebeat és troba a `/etc/filebeat/filebeat.yml`. Només cal afegir les rutes (*paths*) del arxius de logs que volem recol · lectar i especificar on volem enviar aquestes dades. En el nostres cas en concret:

```
filebeat.prospectors:

# Each - is a prospector. Most options can be set at the prospector level, so
# you can use different prospectors for various configurations.
# Below are the prospector specific configurations.

- input_type: log

  # Paths that should be crawled and fetched. Glob based paths.
  paths:
    - /var/log/named/*.log
    #- /var/log/*.log
    #- c:\programdata\elasticsearch\logs\*
```

```
#------------------------------ Logstash output -------------------------------
output.logstash:
  # The Logstash hosts
  hosts: ["elk.tfm.lab:5044"]

  # Optional SSL. By default is off.
  # List of root certificates for HTTPS server verifications
  #ssl.certificate_authorities: ["/etc/pki/root/ca.pem"]

  # Certificate for SSL client authentication
  #ssl.certificate: "/etc/pki/client/cert.pem"

  # Client Certificate Key
  #ssl.key: "/etc/pki/client/cert.key"

#================================ Logging =====================================
```

Figura 24 Configuració de Filebeat

4.1.2 Configuració Logstash per a ISC Bind

La següent Figura mostra el *Pipeline* bàsic d'una plataforma ELK, on a partir de les dades (*source*), es processaran amb Logstash per normalitzar les dades per posteriorment emmagatzemar-les a ElasticSearch, finalment un cop tenim indexades les dades les podem visualitzar amb Kibana.

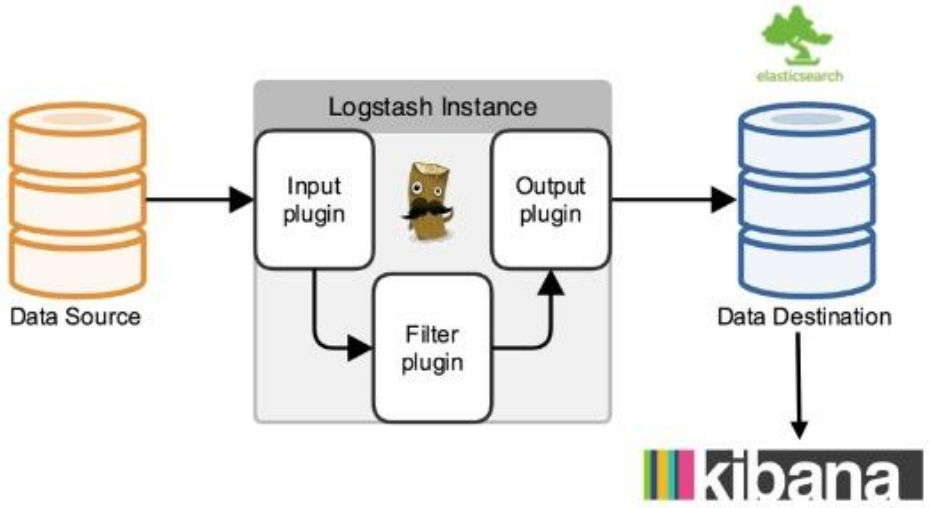


Figura 25 Pipeline bàsic de ELK

Entrada (Input)

El component o *plugin* d'entrada de LogStash que he escollit i que faré servir és *beats*, escoltant pel port 5044.

```
input {
    beats {
        port => 5044
        type => "isc_bind"
    }
}
```

Figura 26 Configuració plugin beats de Logstash

Filtre (Filter)

En el component de filtratge de LogStash s'utilitzarà el mòdul *grok* amb l'objectiu de normalitzar les entrades que provinguin del nostre servidor DNS ISC Bind. Ems podem ajudar de l'eina **Grok Debugger** [15] per verificar que efectivament el nostre filtre fa "*match*" amb el log que li estem enviant:

```
filter {

    grok {
        match => [ "message", "%{MONTHDAY:day}-%{MONTH:month}-%{YEAR:year} %{TIME:time} client %{IP:srcip}#%{DATA:srcport}%{SPACE}\(%{DATA:hostname}\): query:
        %{DATA:hostname2} %{DATA:querytype} \(%{IP:dstip}\)"]
    }

```

Figura 27 Configuració plugin grok de Logstash

Un altre dels *plugins* que he utilitzat ha estat el de *ruby* per obtenir el TLD, és a dir, el Top Level Domain d'una consulta DNS.

```
    ruby {
        code => "event.set('tld', event.get('hostname').split('.').last.downcase)"
    }
```

Figura 28 Configuració plugin ruby de Logstash

Un altre *plugin* que s'utilitza habitualment en Logstash és el filtre *date*, aquest s'utilitza per analitzar les dates dels camps, i després utilitzar aquesta data o marca de temps com la data i hora de Logstash per a l'esdeveniment.

```
    date {
        match => [ "timestamp" , "YYYY-MMM-dd HH:mm:ss.SSS" ]
        target => "@timestamp"
    }
```

Figura 29 Configuració filtre date de Logstash

Finalment, els filtres més interesants que he utilitzat han estat: *translate* i el *plugin geoip*, el primer s'utilitza per esbrinar si la consulta DNS ha estat afegida en alguna llista negra i el segon per obtenir la informació de la GEO localització de la consulta DNS respectivament.

```
    translate {
        field => "[hostname]"
        dictionary_path => "/etc/logstash/blacklist.yaml"
        add_tag => ["MALICIOUS"]
    }

    geoip {
        source => "srcip"
        target => "geoip"
        database => "/etc/logstash/GeoLite2-City.mmdb"
        add_field => [ "[geoip][coordinates]", "%{[geoip][longitude]}" ]
        add_field => [ "[geoip][coordinates]", "%{[geoip][latitude]}"  ]
    }

    mutate {
        convert => [ "[geoip][coordinates]", "float" ]
    }
```

Figura 30 Configuració dels plugins translate i geoip de Logstash

Per crear el diccionari amb la llista negra de llocs catalogats com a maliciosos s'ha desenvolupat un petit script en *Python* que consulta diverses llistes de dominis amb *malware* i genera un arxiu en format YAML per ser utilitzat amb el filtre *translate* de LogStash.

```
#!/usr/bin/env python

import urllib2

request = urllib2.Request('http://mirror1.malwaredomains.com/files/domains.txt')

outfile = open('blacklist.yaml', 'w')

for line in urllib2.urlopen(request):
    content = line.strip().split('\t')
    outfile.write( content[0]  + " : " +  content[1] + "\n")

request = urllib2.Request('http://www.malwaredomainlist.com/hostslist/domains.txt')

for line in urllib2.urlopen(request):
    content = line.strip().split('\t')
    outfile.write( content[0]  + " : " + "Malware" + "\n")

outfile.close()

```

Figura 31 Script blacklist.py

D'altre banda per obtenir les dades de la GEO localització de les IPs s'ha fet servir la següent base de dades:
https://dev.maxmind.com/geoip/geoip2/geolite2/

Sortida (Output)

Pel que fa al component de sortida o Output tots els resultats s'indexaran i s'emmagatzemaran en el clúster de ElasticSearch (`localhost:9200`), també cal indicar l'usuari i contrasenya del nostre clúster:

```
output {
    elasticsearch {
        hosts => ["localhost:9200"]
        user => elastic
        password => changeme
    }
}
```

Figura 32 Configuració de la secció output de Logstash

4.1.3 Instal · lació i configuració Elasticsearch

En aquest cas com ja tenim instal · lats els requeriments i el servidor també té configurat correctament els repositoris d'Elastic, per realitzar la instal · lació només es queda instal · lar el paquet:

```
# apt-get install elasticsearch
```

La configuració del clúster d'Elasticsearch la podem trobar a `/etc/elasticsearch/elasticsearch.yml`

```
# ---------------------------------- Cluster -----------------------------------
#
# Use a descriptive name for your cluster:
#
cluster.name: uoc-tfm
#
# ------------------------------------ Node ------------------------------------
#
# Use a descriptive name for the node:
#
node.name: tfm-1
#
# Add custom attributes to the node:
#
#node.attr.rack: r1
#
# ----------------------------------- Paths ------------------------------------
#
# Path to directory where to store the data (separate multiple locations by comma):
#
path.data: /data/elastic
#
# Path to log files:
#
#path.logs: /path/to/logs
#
```

Figura 33 Configuració de ElasticSearch

Com no volem utilitzar la configuració de *shards* i rèpliques (5 shard i 1 rèplica per *shard*) que té ElasticSearch per defecte, crearem un *template* personalitzat per emmagatzemar les nostres dades. Podem utilitzar la següent comanda per crear un nou *template*:

```
# curl -XPUT 'http://localhost:9200/_template/logstash' -d '{
 "template": "logstash*",
 "settings": {
  "index.refresh_interval": "5s",
  "index.codec": "best_compression",
  "number_of_shards": 3,
  "number_of_replicas": 0
 },
 "mappings": {
  "_default_": {
   "properties": {
    "geoip": {
     "properties": {
      "location": {
       "type": "geo_point"
      }
     }
    }
   },
   "_all": {
    "enabled": false
   }
  }
 }
}'
```

En aquest cas concret, hem reduït el nombre de shards a 3 i hem eliminats les rèpliques doncs inicialment només tindrem un node, també hem indicat que volem comprimir el nostre índex (*best_compression*) i hem definit un *mapping* per a la localització de la IP del tipus *geo_point*. Per acabar hem deshabilitat el camp _all, per reduir l'espai que ocupa el nostre índex, aquest camp permet buscar els valors en els documents sense saber quin camp conté el valor.

Per consultar el *template* generat podem utilitzar la comanda:

```
# curl -XGET 'http://localhost:9200/_template/logstash'
```

4.1.4 Instal · lació i configuració Kibana

De la mateixa forma que passava amb l'ElasticSearch com ja tenim configurat els requeriments i els repositoris d'Elastic, només es queda instal · lar el paquet:

```
# apt-get install kibana
```

4.1.4.1 Configuració dels índexs de ElasticSearch

Els esdeveniments de logs capturats mitjançant LogStash es troben emmagatzemats en una instància de base de dades no relacional. ElasticSearch organitza les dades recol · lectades en índexs amb el format `logstash-YYYY.MM.DD`.

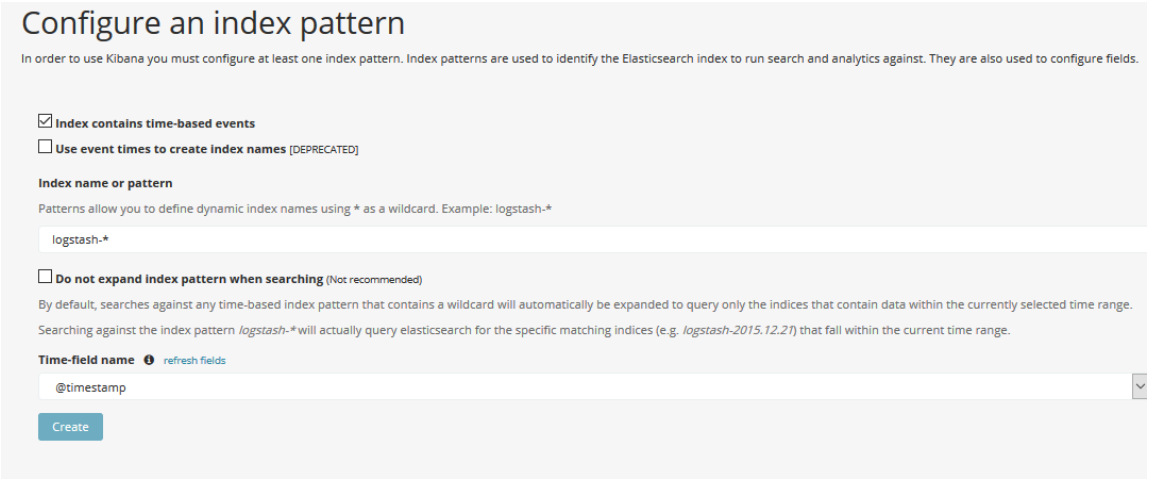


Figura 34 Configuració d'índexs en Kibana

Un cop tenim configurat com a mínim un índex, es pot iniciar l'exploració de les dades emmagatzemades en ElasticSearch mitjançant l'opció "*Discover*":

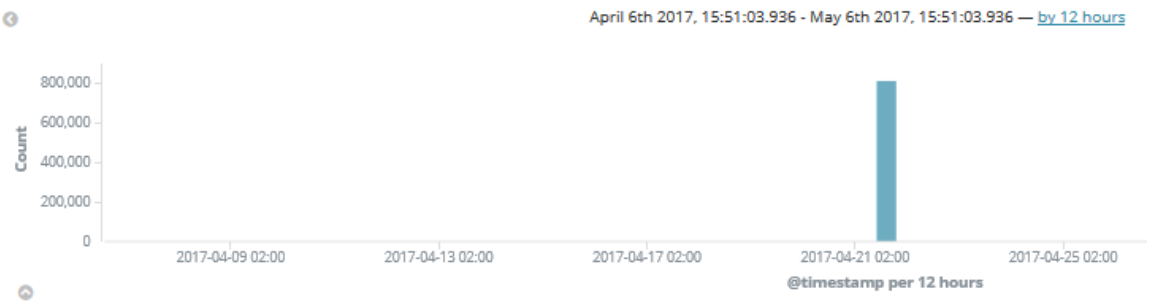


Figura 35 Timeline de la secció *Discover* de Kibana

En aquesta secció s'exploren els esdeveniments recol · lectats i emmagatzemats en un determinat període de temps, aquest període es pot seleccionar a l'opció "*Time Range*" a la zona superior dreta:

Figura 36 Filtre Time Range de Kibana

4.1.5 Instal · lació i configuració Nginx

Utilitzarem Nginx com a proxy invers per permetre l'accés extern a Kibana, degut a que deixarem configurat Kibana només escoltant per localhost. Per instal · lar nginx només ens caldrà el següent paquet:

```
# apt-get install nginx
```

Farem servir **openssl** per crear un usuari administrador, anomenat *"kibanaadmin"*, que podrà accedir a la interfície web de Kibana:

```
echo "kibanaadmin:`openssl passwd -apr1`" | tee -a
/etc/nginx/htpasswd.users
```

Finalment, definirem un nou Virtual Server (site), modificant l'arxiu: `/etc/nginx/sites-available/default`

```
server {
    listen 80;
    server_name elk.csuc.cat;

    auth_basic "Restricted Access";
    auth_basic_user_file /etc/nginx/htpasswd.users;

    location / {
        proxy_pass http://localhost:5601;
        proxy_http_version 1.1;
        proxy_set_header Upgrade $http_upgrade;
        proxy_set_header Connection 'upgrade';
        proxy_set_header Host $host;
        proxy_cache_bypass $http_upgrade;
    }
}
```

Capítol 5
Proves funcionals

En aquest capítol es presenten el primers resultats del quadre de comandament o Dashboard desenvolupat amb Kibana i les proves funcionals realitzades amb la plataforma ELK.

2 Anàlisis d'esdeveniments (logs) amb Kibana

Un dels objectius principals del projecte era crear un quadre de comandament, que pogués detectar anomalies o incidents de seguretat a través de les consultes DNS. Per aquesta tasca s'han estudiats diferents aplicacions disponibles en altres plataformes con ara Splunk, que actualment ja realitzen aquestes funcionalitats, concretament s'han analitzat les següents aplicacions:

- DNS Analytics for Splunk [11]
- DDST DNS Analytics for Splunk [12]

2.1 DNS Analytics for Splunk

Per identificar ràpidament el *malware* i la violació de polítiques mitjançant l'anàlisi de les dades d'esdeveniments de consulta DNS. El procés es pot dur a terme en les instal · lacions o en el núvol per descobrir noves amenaces, com ara campanyes de APT, ransomware, i exfiltración a través d'un túnel de DNS.

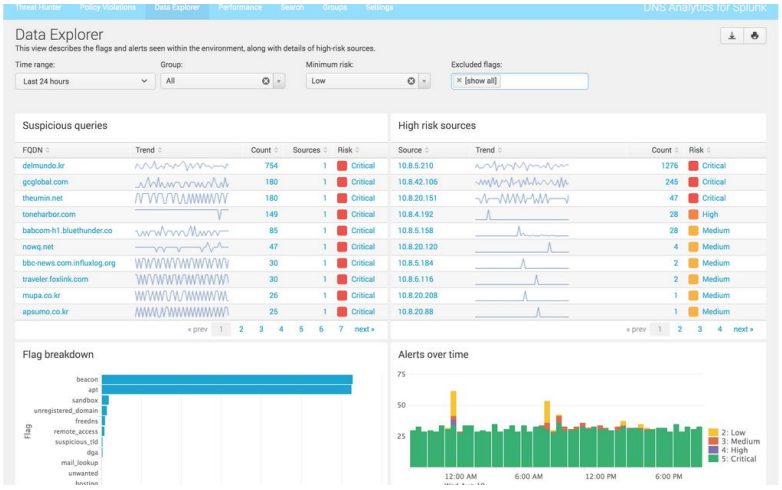


Figura 37 Dashboard de l'aplicació DNS Analytics for Splunk

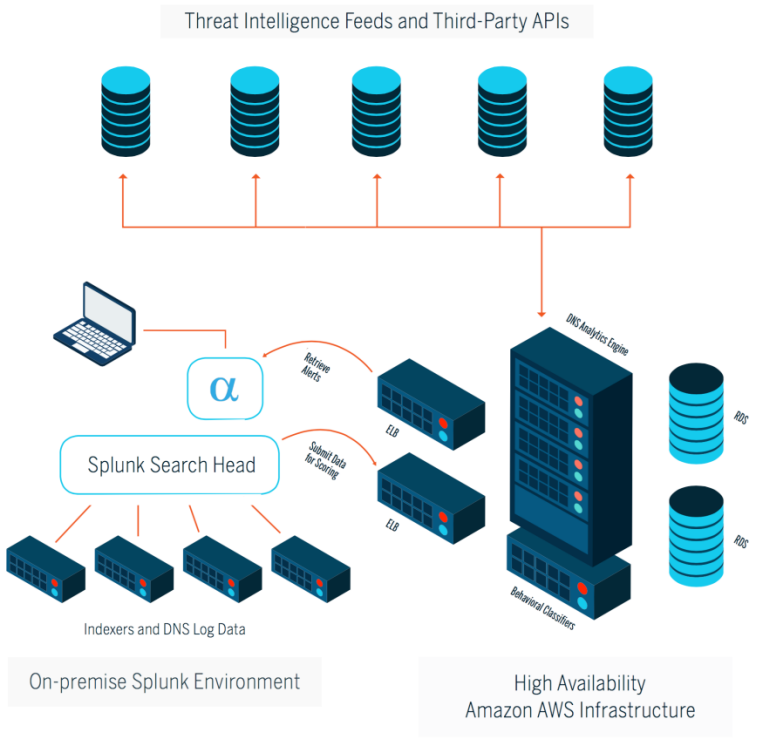


Figura 38 Arquitectura Aoo DNS Analytics for Splunk

El principal inconvenient d'aquesta aplicació o *app* de Splunk és que l'accés a la seva API de DNS Analytics és realitza de forma llicenciada per volum de consultes, segons la següent taula:

Band	Daily Limit	Cost
A	1M events	$165 / month
B	2M events	$320 / month
C	5M events	$790 / month
D	10M events	$1,560 / month
E	20M events	$3,080 / month

Figura 39 Taula de costos de la API

2.2 DDST DNS Analytics for Splunk

La segona aplicació estudiada té per objectiu permetre que l'administrador o l'analista de seguretat pugui veure ràpidament el que està passant en les peticions DNS que estan utilitzant els seus usuaris o bé els seus sistemes d'informació. L'aplicació DNS s'actualitzà per ajudar a identificar peticions DNS malicioses (*malware*).

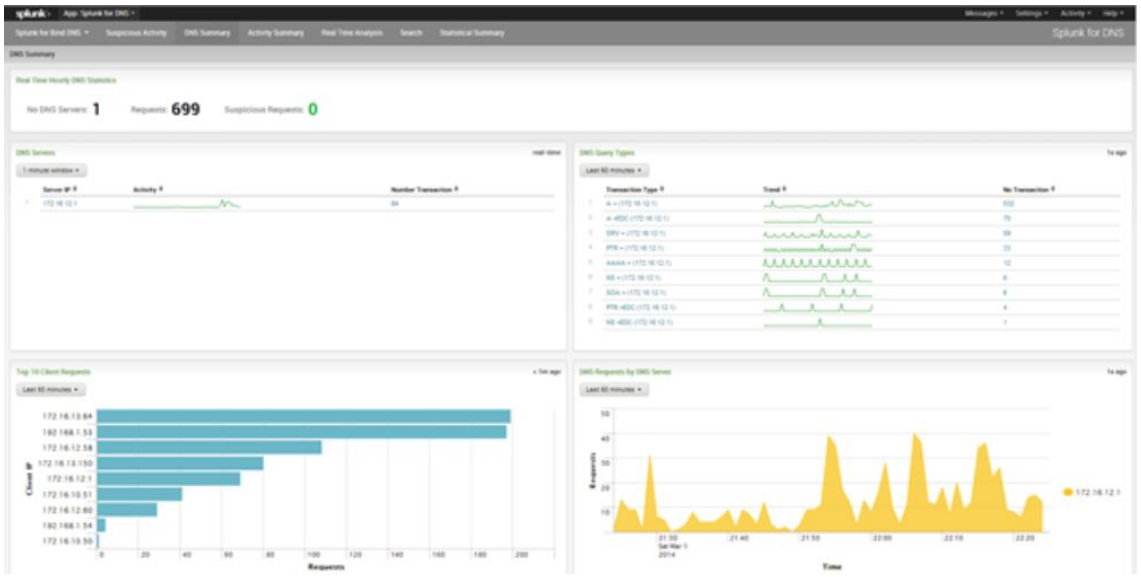
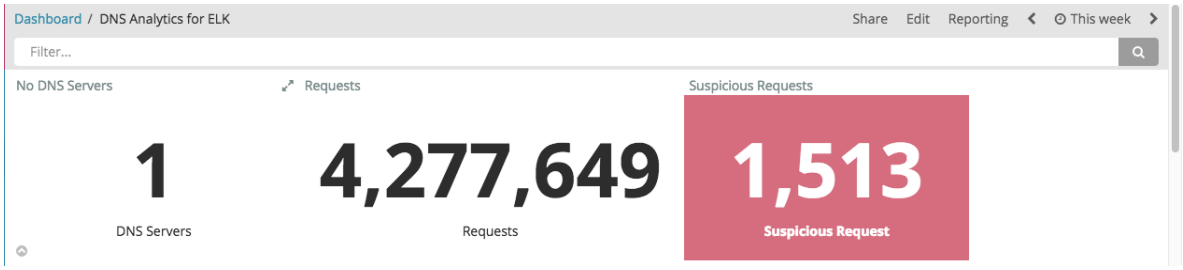


Figura 40 Dashboard de l'aplicació DDST DNS Analytics for Splunk

2.3 Dashboard desenvolupat amb Kibana

El quadre de comandament o Dashboard desenvolupat pel projecte, esta basant en la aplicació o *app* DDST DNS Analytics for Splunk descrita anteriorment, a continuació es mostra les seves funcionalitats, realitzat a partir d'analitzar aproximadament unes ~4.300.000 peticions DNS.

En primer lloc, podem veure el número de servidors DNS diferents que estan enviant peticions a la plataforma, així com el nombre total de peticions (documents a ElasticSearch), també podem detectar ràpidament quina d'aquestes peticions s'ha catalogat com a sospitoses.



En la següent taula es mostra el nombre total de transaccions rebudes per IP, en aquest cas 3.875.903 per IPv4 i 401.746 per IPv6, per motius de confidencialitat en totes les imatges del Dashboard les adreces IPs apareixeran ocultes.

DNS Servers

Server IP	Number Transaction
	3,875,903
	401,746

També s'ha creat una taula per identificar el tipus de consulta DNS que s'ha realitzat.

DNS Query Types

Transaction Type	No Transaction
IN A -EDC	891,440
IN A -E	489,046
IN A -	455,102
IN AAAA -EDC	415,237
IN A -ED	356,809

Export: Raw Formatted

Així per exemple, +EDC en una consulta indica que és:

- **Recursiva** (+) - ha vingut d'un client o d'un servidor que està reenviant consultes al seu servidor
- El remitent utilitza **EDNS0** (utilitzant mides de paquets UDP més grans i de senyalització de la mida que pot ser acceptat)
- El remitent entén **DNSSEC** (D) - es tracta d'una petició al servidor per incloure qualsevol material DNSSEC associada amb la resposta de la consulta.
- DNSSEC comprovació de la validació es desactiva (C) - el remitent desitja que la resposta de totes maneres, fins i tot si fallen les proves de validació.

En la següent taula, apareixen aquelles peticions que han estat identificades com a sospitoses i el tipus d'atac del que es tracta. En la taula es mostra informació sobre la IP del client que ha realitzat la consulta, la consulta en qüestió, el tipus d'atac així com el nombre de peticions que s'han rebut per aquest lloc.

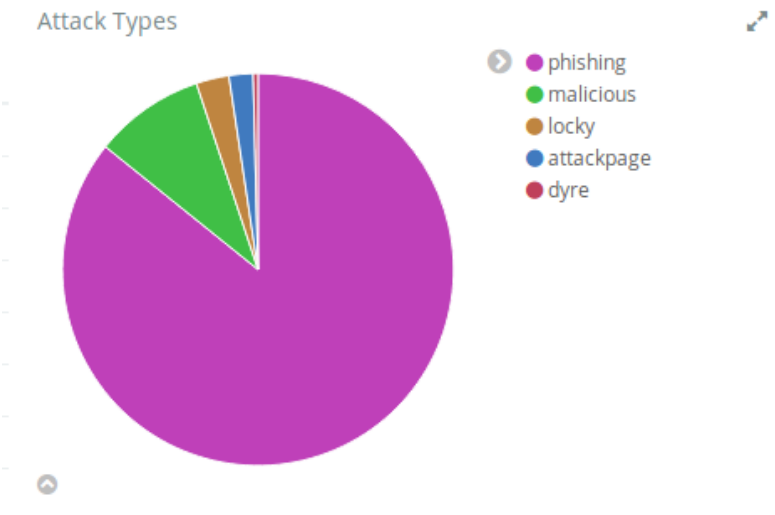
Suspicious Requests Table

Client IP	Query	Attack Type	Requests
	aquapuremultiservicios.es	malicious	2
	i.nfil.es	dyre	1
	meliurbis.es	phishing	1
	sutaxivigo.es	phishing	1
	energiasolarcanarias.es	phishing	1
	sybaristravel.es	phishing	1
	term-servicest01.esy.es	phishing	3
	ads-team-safety.esy.es	phishing	1
	clipsexx.esy.es	phishing	1
	ssl-unlock-pages.esy.es	phishing	1

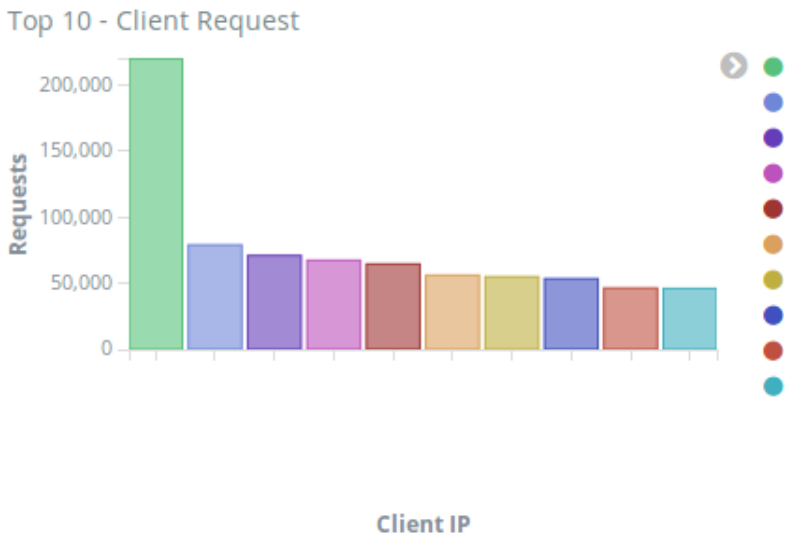
Export: Raw Formatted

D'aquesta forma podem identificar fàcilment quin ha estat el client (adreça IP) que ha realitzat la consulta maliciosa, també per temes de confidencialitat s'ha ocultat totes les adreces IP.

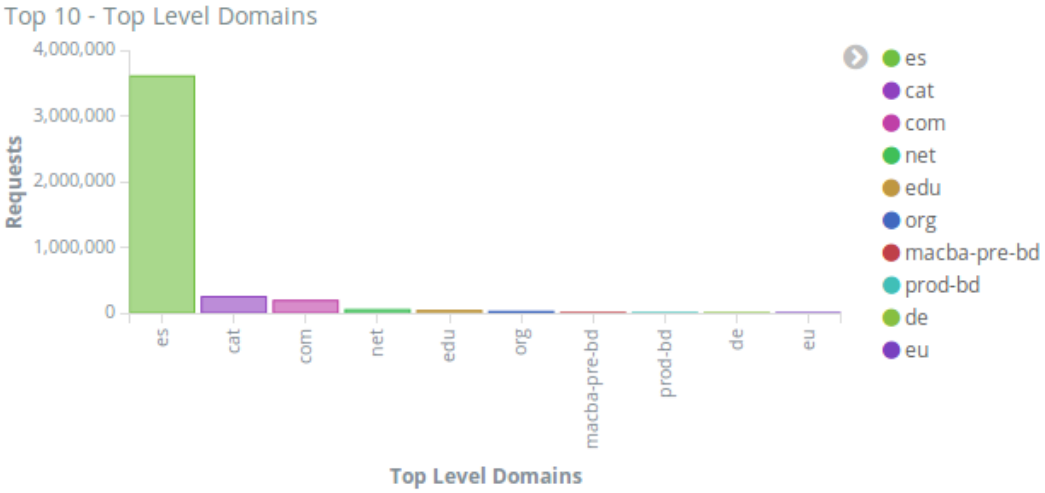
També s'ha creat un gràfic de tipus **Pie Chart** per classificar el tipus d'atacs com es pot observar a la Figura, la gran majoria de les peticions DNS sospitosos analitzades es tracta de llocs de *phishing*.

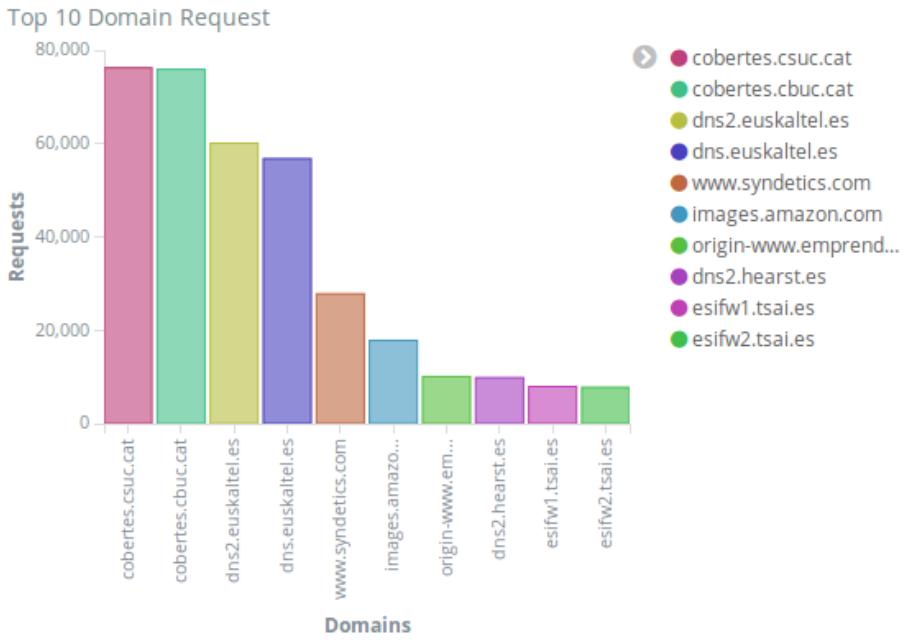


A continuació les següents gràfiques mostren el Top 10 de peticions per client, el Top 10 de *Top level domains* i el Top 10 per dominis, un cop més les adreces IP s'han ocultat:

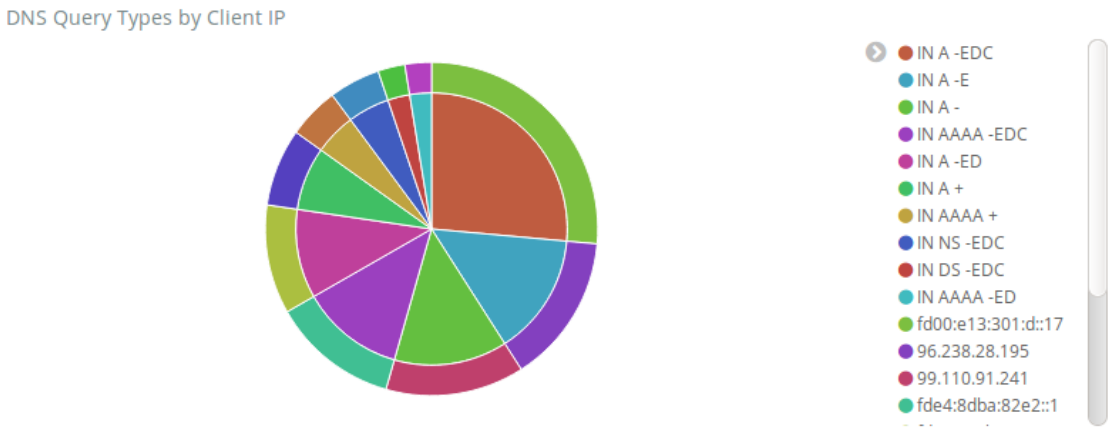


En la gràfica es pot apreciar que el domini **.es** el clarament el més consultat o sol · licitat seguit del **.cat** i el **.com**

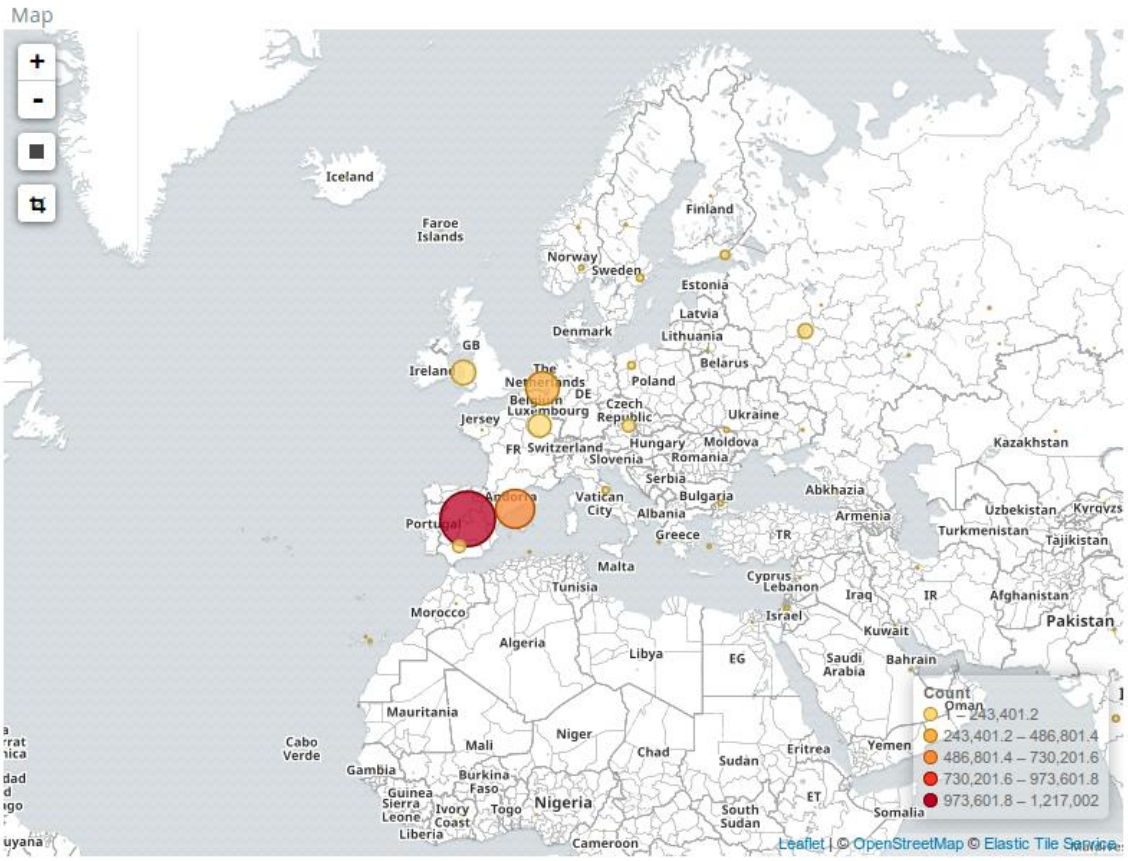




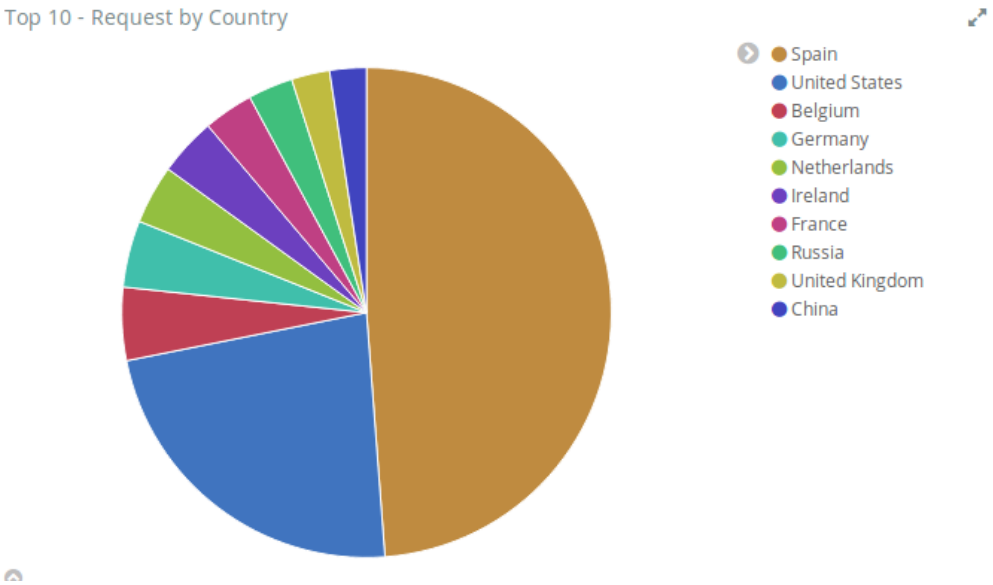
Kibana també ens permet crear gràfics combinant diferent camps com ara, per exemple, en aquest cas, hem combinat els tipus de consulta DNS amb el client que realitza més consultes d'aquest tipus concret, el resultat és el que es mostra en la següent Figura:



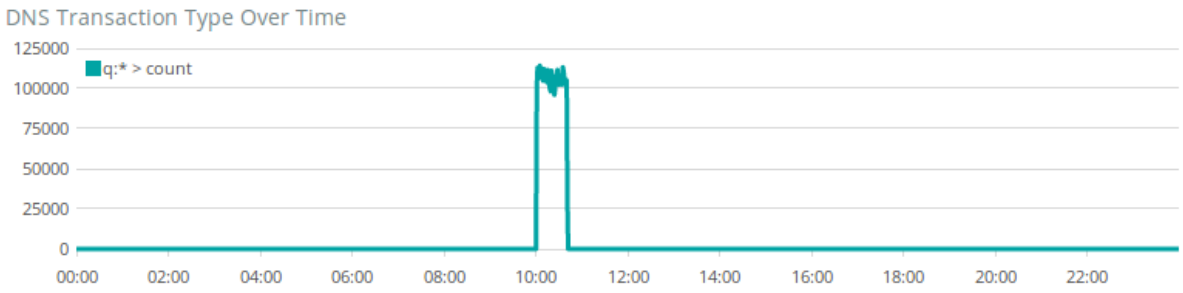
Finalment, utilitzant la informació obtinguda a partir del filtre **geoip** de Logstash, s'han situat en un mapa totes les peticions DNS analitzades, obtenint un mapa com el que es mostra a continuació:



En el següent gràfic de tipus **Pie Chart** es mostra les peticions analitzades per país:



Utilitzant el *plugin* Timelion de Kibana podem crear un gràfic per veure les transaccions DNS que es van processant per la plataforma al llarg del temps.



Finalment, per acabar el Dashboard s'ha creat una cerca de Kibana que mostra les últimes entrades DNS que s'han processat per la plataforma mostrant el Timestamp, la adreça IP del servidor DNS (dstip), l'adreça IP del client (srcip) i la query demanada (hostname):

DNS Request Activity

1–50 of 4,277,649

Time	dstip	srcip	hostname
May 29th 2017, 10:41:16.087			juanotero.es
May 29th 2017, 10:41:16.087			www.fragrantica.es
May 29th 2017, 10:41:16.087			WWW.fETOC.ES
May 29th 2017, 10:41:16.086			dramacool.es
May 29th 2017, 10:41:16.086			bicicletasvalladolid.es
May 29th 2017, 10:41:16.085			cdn.modalia.es
May 29th 2017, 10:41:16.085			www.artimedia.es
May 29th 2017, 10:41:16.085			cdn.modalia.es

2.4 Integració Logstash amb Zabbix

La opció escollida per rebré alertes de la plataforma ELK ha estat fer servir el *plugin* de sortida (output) de Zabbix per Logstash. Aquest *plugin* s'utilitza per enviar dades (parells clau / valor) a un servidor Zabbix.

El protocol de Zabbix Sender es descriu detalladament a:
https://www.zabbix.org/wiki/Docs/protocols/zabbix_sender/2.0

Aquest *plugin* juntament amb el filtre ***metrics*** de Logstash m'ha permès anar enviant dades al servidor Zabbix cada cop que la plataforma ELK processava una petició maliciosa del DNS.

A continuació es detalla la configuració aplicada al filtre de Logstash per detectar el comportament esmentat:

```
if "MALICIOUS" in [tags] {
        metrics {
                meter => "events"
                add_field => {"[@metadata][zabbix_key]" => "request_malicious"}
                add_field => {"[@metadata][zabbix_host]" => "elk"}
                add_tag => "metric"
                clear_interval => 3600
        }
    }
}

```

Figura 41 Configuració del filtre metrics de Logstash

Finalment, també s'ha hagut de modificar la sortida (output) per enviar a les dades a Zabbix, com es pot observar només aquells events que contenen el tag mètric són enviats al nostre servidor Zabbix la resta s'emmagatzemen en el clúster d'ElasticSearch:

```
output {

        if "metric" in [tags] {
            zabbix {
              zabbix_server_host => FQDN O IP SERVIDOR ZABBIX
              zabbix_host => "[@metadata][zabbix_host]"
              zabbix_key => "[@metadata][zabbix_key]"
              zabbix_value => "[events][count]"
            }
        }
        else {
          elasticsearch {
              hosts => ["localhost:9200"]
              user => elastic
              password => changeme
          }
        }
}

```

Figura 42 Configuració de la sortida Zabbix de Logstash

A continuació es mostren algunes captures de pantalla on es pot observar el resultat de la integració amb Zabbix. En la següent Figura es pot observar com en el host elk ha saltat una alarma classificada com a Warning, degut a que s'han rebut més de 5 peticions DNS malicioses.

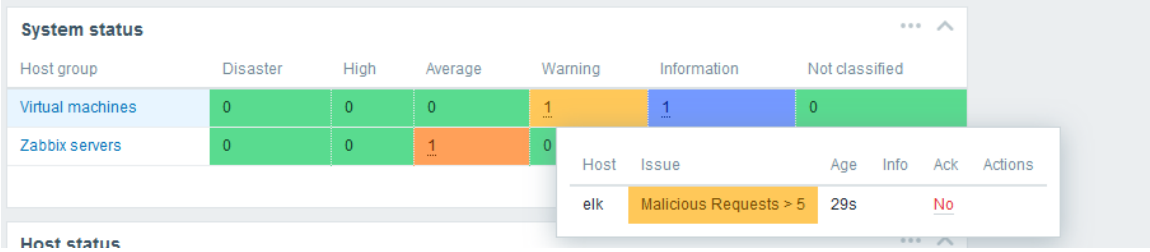


Figura 43 Dashboard de Zabbix amb una alerta de Warning

Aquesta gràfica va mostrant les darreres dades que va reben el servidor Zabbix de Logstash, les lines horitzontals representen els Triggers que faran saltar les alertes al Dashboard de monitoratge:

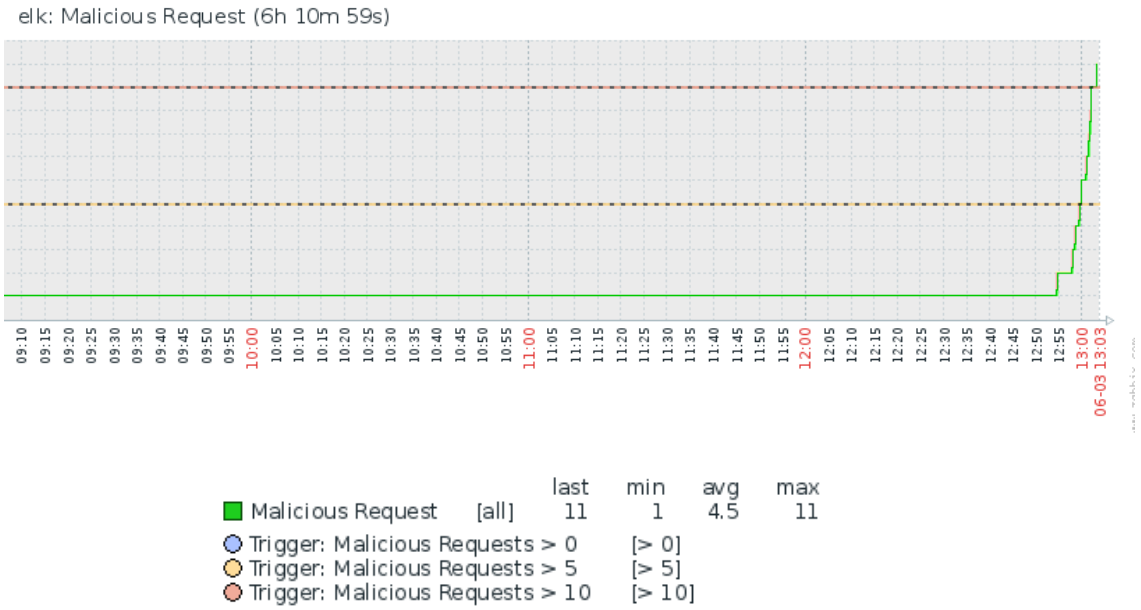


Figura 44 Gràfic amb les darredes dades de Zabbix

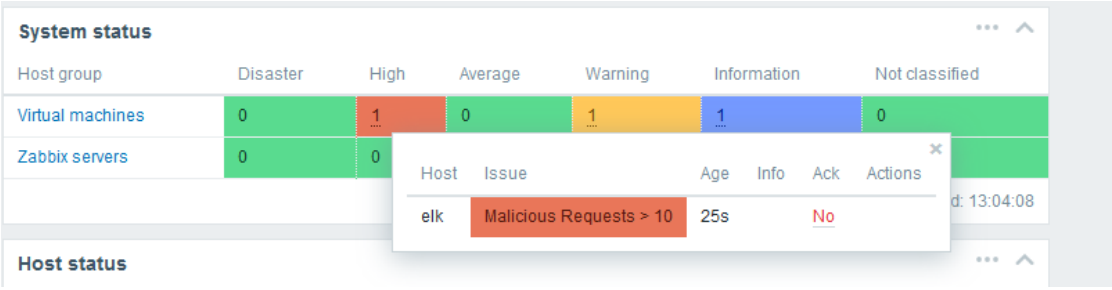


Figura 45 Dashboard de Zabbix amb una alerta de tipus High

2.5 Monitoratge de la plataforma ELK amb X-Pack

X-Pack és una extensió del stack d'Elastic que agrupa diverses funcionalitat com ara seguretat, alertes, monitoratge, *reporting* i capacitats de generació de gràfics en un paquet fàcil d'instal · lar (*plugin*).

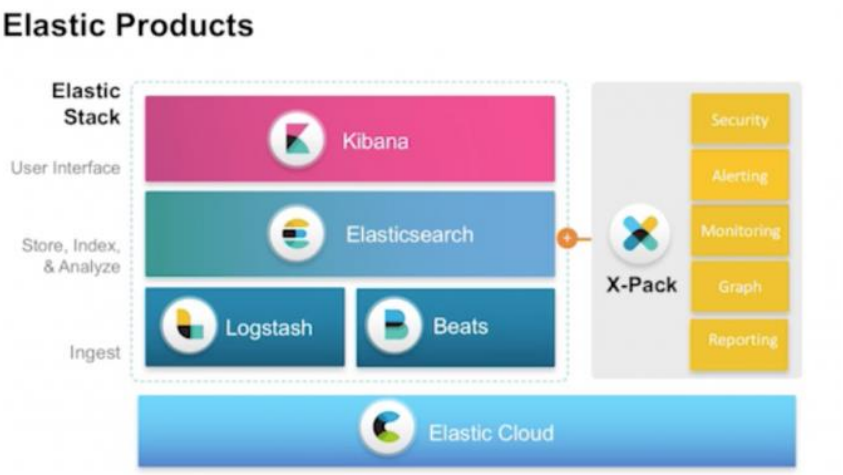


Figura 46 Productes del Stack d'Elastic

Per instal · lar el *plugin* a la plataforma ELK només cal executar les següents comandes:

```
# bin/elasticsearch-plugin install x-pack
# bin/kibana-plugin install x-pack
# bin/logstash-plugin install x-pack
```

Un cop instal · lat el *plugin* al accedir a la interfície web de Kibana, veurem una nova pestanya amb el nom de **Monitoring**, que es proporciona informació sobre el nostre clúster, per exemple informació sobre l'estat del clúster de ElasticSearch tal i com es pot veure a la Figura:

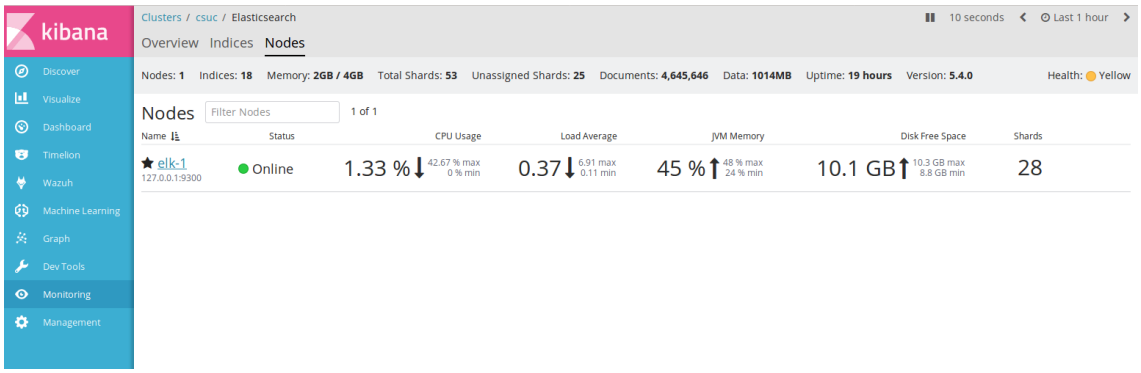


Figura 47 Monitoratge del clúster d'ElasticSearch

Amb aquest *plugin*, també podem obtenir informació detallada sobre qualsevol del índexs del nostre clúster, com per exemple:

- Mida del índex a memòria (en MB)
- Mida o espai en disc que ocupa el nostre índex (en MB)
- Taxa de cerca (/s)
- Taxa d'indexació (/s)
- Nombre total de documents indexats.
- Etc...

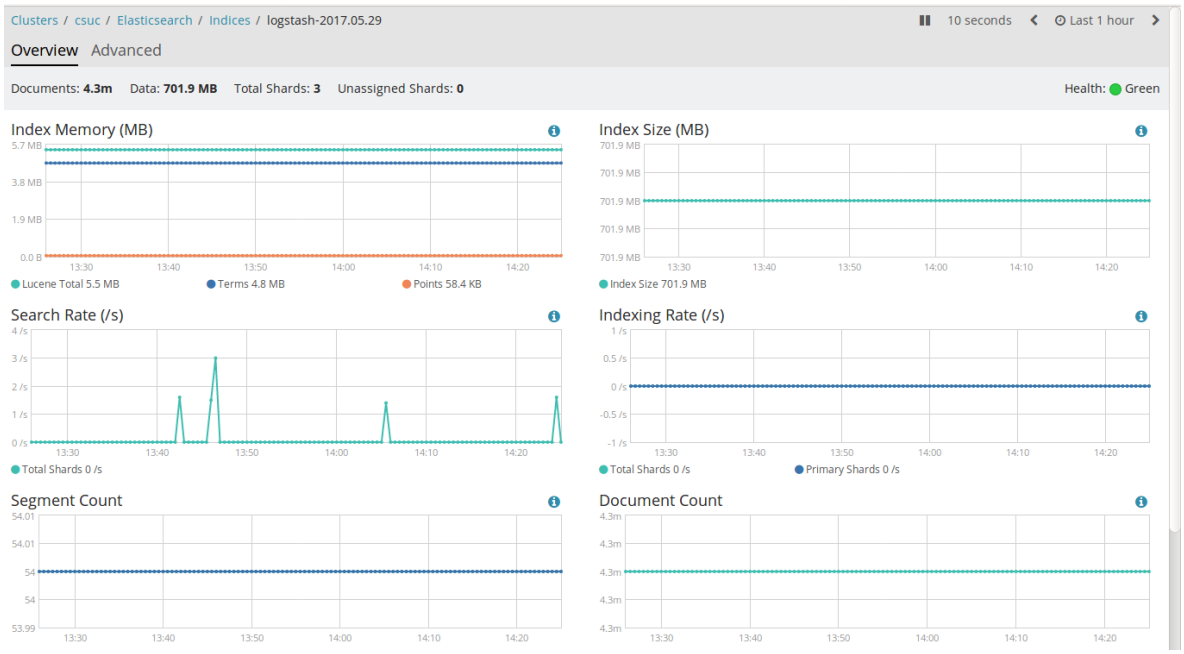


Figura 48 Estadístiques sobre un índex amb X-Pack

Capítol 6
Conclusions

En el capítol final s'exposaran les conclusions respecte els objectius inicialment plantejats per aquest projecte, també es presenta una estimació econòmica del cost que suposa dur a terme el projecte i finalment unes línies sobre actuacions futures o millores que es poden realitzar.

6.1 Quant als objectius

Els principals objectius plantejats a l'inici del projecte s'ha satisfet satisfactòriament.

Respecte a la planificació inicial s'ha donat més rellevància, per tant més dedicació, a la part de recerca i estudi d'una infraestructura DNS i com analitzar i gestionar els seus logs, per tant, la part d'integració amb l'eina OSSEC Wazuh [17] s'ha deixat com a treball futur.

- El primer objectiu assolit ha sigut estudiar i realitzar la recerca necessària sobre les tecnologies emprades per duu a terme el treball.
- S'ha realitzat un estudi en profunditat sobre la plataforma ELK, en comparació amb altres plataformes i s'ha adaptat al cas d'us concret del projecte.
- S'ha aconseguit l'objectiu de desplegar la plataforma ELK en un entorn de laboratori basat en la plataforma de Cloud OpenNebula.
- S'han completat amb èxit un seguit de proves funcionals per verificar el correcte funcionament de la plataforma desenvolupada.
- Finalment, s'ha documentat tot el treball realitzat elaborant el present document, que conforma la memòria final del projecte.

6.2 Gestió econòmica

En aquest apartat, farem una estimació del cost que suposaria la realització d'aquest projecte. diferenciant dues parts els costos dels recursos humans i els costos en recursos no humans.

6.2.1 Pressupost del personal

Per calcular el cost del personal s'han definit dos rols diferenciats, que són el Cap de projecte i l'Enginyer de Seguretat, cadascun d'aquests rols desenvolupa diverses tasques dins del projecte.

El Cap de projecte és la persona encarregada de l'estudi i la anàlisis de la plataforma, elaborar l'especificació de la plataforma, detectant els requisits que ha de satisfer i vetllar pel compliment de la planificació de les tasques de tot el projecte, mentre que l' Enginyer de Seguretat s'encarregarà de la implementació de la plataforma a partir de l'especificació del Cap de projecte i serà l'encarregat de realitzar les proves funcionals i el manteniment posterior de la plataforma.

Respecte al nombre d'hores de dedicació, s'ha pres com a referència el Pla Docent de la UOC que estableix una dedicació de 20 hores per crèdit, llavors un projecte com a aquest de 9 crèdits, tindrà una nombre total d'hores aproximat de: $20 \cdot 9 = 180$ hores. D'aquesta manera, un projecte de 9 crèdits es pot realitzar durant mig semestre a temps complert o durant un semestre a temps parcial (20 hores a la setmana), veure la següent Taula 4.

Perfil	Preu/hora	Hores	Total
Cap de projecte	45 €	50	2250 €
Enginyer de Seguretat	30 €	130	3900 €
Totals		**180**	**6150 €**

Taula 4 Pressupost del personal

6.2.2 Pressupost en materials i maquinari

Per elaborar el pressupost en maquinari, s'ha calculat el cost que tindria desplegar la infraestructura en modalitat de pagament per ús (IaaS). En aquest cas concret, el host físic on s'ha desplegat la màquina virtual és un servidor HP ProLiant DL370 G6 com el de la Figura.

Figura 49 Servidor HP ProLiant DL370 G6

Així, una màquina virtual de tipus H4 amb les següents prestacions 4 vCPU, 4 GB RAM, 50 GB disc, tindria un quota per hora de 0,125 € , el cost addicional d'afegir +1 vCPU o afegir +1 GB RAM tindria una quota per hora de 0,030 €

Pel que fa al emmagatzematge de dades cada bloc de 10 GB (disc) té una quota mensual de 0,722 €, inicialment s'ha pressupostat disposar d'un 1 TB d'espai d'emmagatzematge. Llavors, el cost mensual de la màquina virtual estaria al voltant dels 248.6 €/mes.

Maquinari	Valor/Unitari	Unitats	Total
Màquina virtual H4	0,125 €	720	90 €
+ 4 GB RAM addicionals	0,12	720	86,4€
1 TB de disc	0.722€	100	72.2 €
Totals			**248.6 €**

Taula 5 Pressupost del maquinari

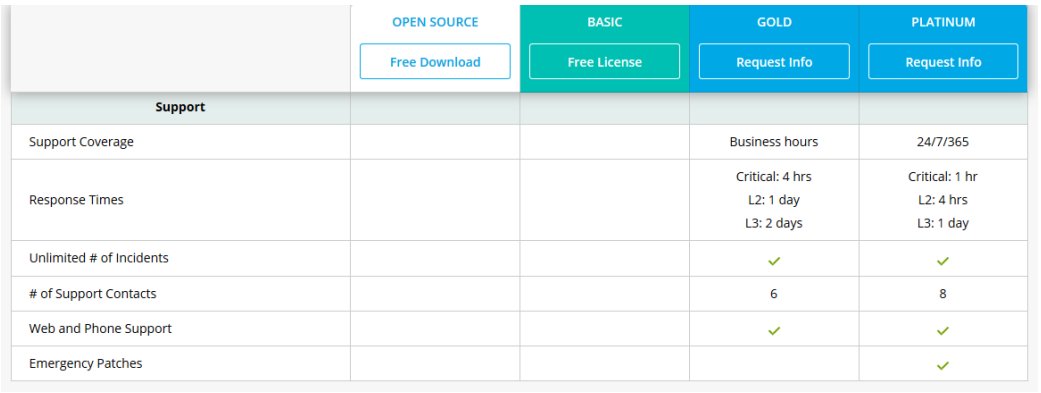
6.2.3 Pressupost en llicències i programari

Una de les característiques destacables del nostre projecte i que s'havia establert com a requeriment alhora d'implementar la plataforma era que el cost en llicencies fos reduït donant prioritat a la utilització de programari lliure (*Open Source*) sempre que fos possible, cosa que finalment s'ha pogut satisfer.

Programari	Valor/Unitari	Unitats	Total
Debian 8.0	0 €	2	0 €
ISC bind	0 €	1	0 €
ElasticSearcn	0 €	1	0 €
Logstash	0 €	1	0 €
Kibana	0 €	1	0 €
Totals			**0 €**

Taula 6 Pressupost del programari

Tot i que la majoria dels productes d'Elastic son *Open Source* algunes de les característiques addicionals o bé si es vol disposar d'algun serveis de suport, cal realitzar una subscripció amb Elastic per adquirir una llicencia del producte. El model de llicenciament es per node, es pot obtenir informació addicional a la següent URL: https://www.elastic.co/subscriptions



	OPEN SOURCE Free Download	BASIC Free License	GOLD Request Info	PLATINUM Request Info
Support				
Support Coverage			Business hours	24/7/365
Response Times			Critical: 4 hrs L2: 1 day L3: 2 days	Critical: 1 hr L2: 4 hrs L3: 1 day
Unlimited # of Incidents			✓	✓
# of Support Contacts			6	8
Web and Phone Support			✓	✓
Emergency Patches				✓

Figura 50 Taula amb els diferents tipus de subscripcions d'Elastic

6.2.4 Pressupost total del projecte

El pressupost total del projecte, s'obté de la suma dels pressupostos parcials (personal, material i maquinari, llicències i programari) calculats anteriorment, llavors la xifra total estaria al voltant d'uns 6398.6 € aproximadament.

	Total
Pressupost del personal	6150 €
Pressupost del maquinari	248.6 €/mes
Pressupost del programari	0 €
Total	**6398.6€**

Taula 7 Pressupost total del projecte

6.3 Treball futur

Tot i que el projecte desenvolupat cobreix gran part dels objectius plantejats inicialment, resulta interessant oferir alguns aspectes de millora i/o ampliacions que es podrien realitzar en un futur. Alguns aspectes interessants que es podrien treballar en desenvolupament futur podrien ser els següents:

- Posada en producció de la plataforma al CSUC.
- Millorar el sistema de *feeds* per disposar d'una informació més actualitzada i precisa d'aquelles consultes que puguin ser malicioses.
- Estudiar com realitzar una optimització del rendiment (*performance*) del clúster d'ElasticSearch.
- Dotar a la plataforma de més intel·ligència per exemple afegint informació relativa al **Passive DNS** i així poder detectar altres tipus d'anomalies.
- Estudiar altres *plugins* de filter per afegir noves funcionalitats com per exemple el plugin de **virustotal** [19]

Bibliografia

[1] Elastic, «Elastic,» 2017. [En línea]. Available: https://www.elastic.co/.

[2] «Consorci de Serveis Universitaris de Catalunya,» 2017. [En línea]. Available: http://csuc.cat/.

[3] «Splunk,» 2017. [En línea]. Available: https://www.splunk.com/es_es.

[4] «Logtrust,» 2017. [En línea]. Available: https://www.logtrust.com/en/.

[5] «Graylog,» 2017. [En línea]. Available: https://www.graylog.org/.

[6] N. L. Server, 2017. [En línea]. Available: https://www.nagios.com/products/nagios-log-server/.

[7] A. L. Padilla, «Guía de seguridad en servicios DNS,» 2017. [En línea]. Available: https://www.incibe.es/.../ManualesGuias/guia_de_seguridad_en_servicios_dns.pdf.

[8] «Apache Lucene,» 2017. [En línea]. Available: https://lucene.apache.org/core/.

[9] «Cerebro,» 2017. [En línea]. Available: https://github.com/lmenezes/cerebro.

[10] «OpenNebula,» [En línea]. Available: https://opennebula.org/. [Último acceso: 2017].

[11] «BIND,» 2017. [En línea]. Available: https://www.isc.org/downloads/bind/.

[12] Elastic, «Curator,» 2017. [En línea]. Available: https://www.elastic.co/guide/en/elasticsearch/client/curator/5.0/index.html.

[13] Elastic, «Filebeats,» 2017. [En línea]. Available: https://www.elastic.co/products/beats/filebeat.

[14] «Filebeat,» 2017. [En línea]. Available: https://www.elastic.co/guide/en/beats/filebeat/current/filebeat-getting-started.html.

[15] R. A. Heinlein, «Grok Debugger,» 2017. [En línea]. Available: https://grokdebug.herokuapp.com/.

[16] «DNS Analytics for Splunk,» 2017. [En línea]. Available: https://splunkbase.splunk.com/app/1657/.

[17] «DDST DNS Analytics for Splunk,» 2017. [En línea]. Available: https://splunkbase.splunk.com/app/1090/.

[18] Wazuh, «Open Source Host and Endpoint Security,» 2017. [En línea]. Available: https://wazuh.com/.

[19] J. Kendall, «Logstash Plugin Virustotal,» 2017. [En línea]. Available: https://github.com/coolacid/logstash-filter-virustotal.

[20] «Open Source Host and Endpoint Security,» 2007. [En línea]. Available: https://wazuh.com/.

[21] M. I. Gandía, «Uso de la herramienta Splunk en el CSUC,» 16 Junio 2015. [En línea]. Available: https://es.slideshare.net/CSUC_info/1506-tecniris-splunk.

[22] R. Calzada, «Taller de ELK,» UC3M, 2017. [En línea]. Available: https://www.rediris.es/jt/jt2016/ponencias/?id=jt2016-gt-taller_elk-a16b0c3.pdf.